Estimating latent processes on a graph from indirect data

Edo Airoldi

Department of Statistics
Harvard University

Joint with Alex Blocker, Bertrand Haas

Graph Exploitation Symposium, August 9th 2011, MIT Lincoln Laboratory

Overview

- Inference problem

  Estimating origin-destination flows form aggregate traffic

- Technical challenges

  A sequence of ill-posed linear inverse problems

  Extremely constrained estimands space

- In detail

  Deconvolution of mixing time series

  Distributions on polytopes and sampling strategies

Statistical issues in network analysis

- Representation and compressed sensing

  How to smoothly represent the space of all graph structures?

  Motifs, metrics, spectral, …, semi-parametric

- Population models

  Sample size? Notions of variability? *(See survey paper)*

- Diffusion of information on a network

  How to infer who talks to whom from aggregate traffic?

Statistical issues in network analysis

- Goodness-of-fit, CIs, tests, selection

  How to establish confidence sets for network structure?

- Causal inference with interference

  How to separate peer-influence effects from homophily?

- Diffusion of information on a network

  How to combine local cascades to reveal global structure?

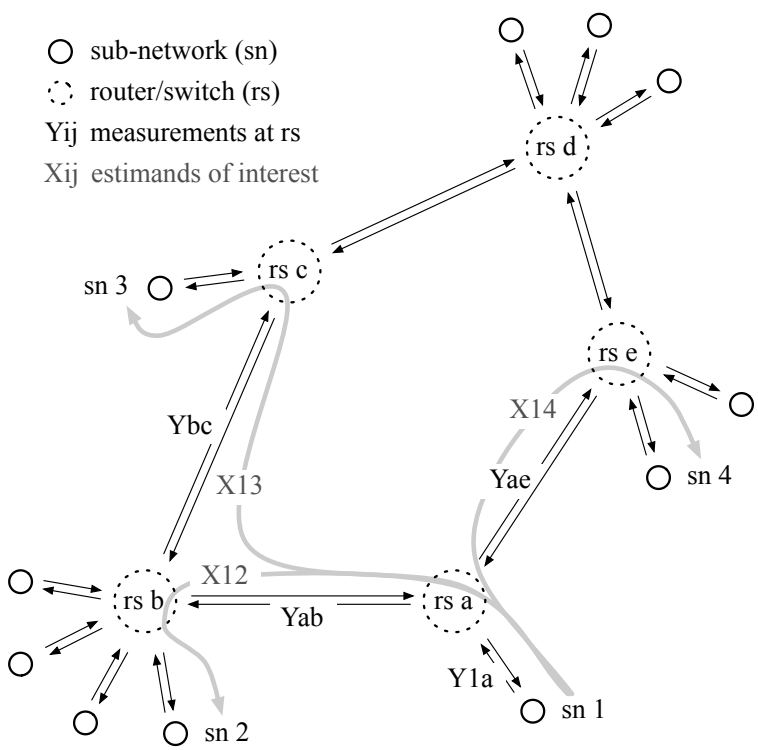
  How to infer who talks to whom from aggregate traffic?

Agenda

- The inference problem and its relevance
- Deconvolution of mixing time series on a graph
- Sampling extremely constrained estimands
- Empirical results
- Concluding remarks

Inferring origin-destination flows

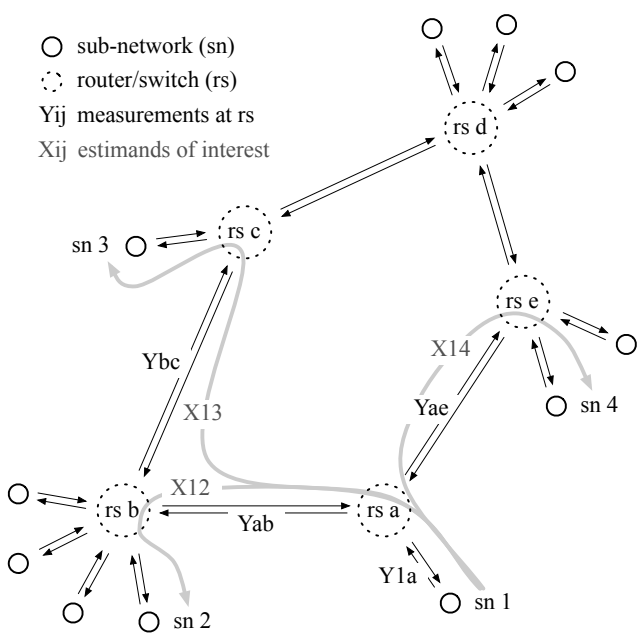
A large ISP network has 100s of nodes, 1000s of links, 10000s routes, and over 1 petabyte traffic/day



1. Reliability analysis: predict loads | failures
2. Traffic engineering: minimize congestion
3. Capacity planning: forecast requirements

Mathematical formulation

Given a collection of link loads $Y_{(m \times t)}$ and the routing matrix $A_{(m \times n)}$, find non-observable OD traffic flows, $X_{(n \times t)}$, such that $Y = A \cdot X$, where $n > m$.



$$\begin{bmatrix} y(1,t) \\ y(2,t) \\ y(3,t) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x(1,t) \\ x(2,t) \\ x(3,t) \\ x(4,t) \end{bmatrix}$$

Broader context and approaches

- Other popular problems can be formulated as ill-posed linear systems with a TUM matrix A
  1. Sampling multi-way contingency tables given margins
  2. Sampling networks with given degree sequence
- Two main approaches
  1. Algebraic geometric methods: Diaconis, Sturmfels, Fienberg, Dobra, Sullivant, …
  2. Sequential MCMC: Chen, Liu, Diaconis, Holmes; Importance sampling: Blitzstein, Diaconis, …

Traffic matrix estimation literature

Recover order N^2 OD flows from order N sums.

Classical statistical problem: Vanderbei & Iannone, 1994; Vardi, 1996; Tebaldi & West, 1998; Cao et al., 2000; Zhang et al., 2003; Airoldi & Faloutsos, 2004, Michailidis et al., 2006, 2007; Liang & Yu, 2008; …

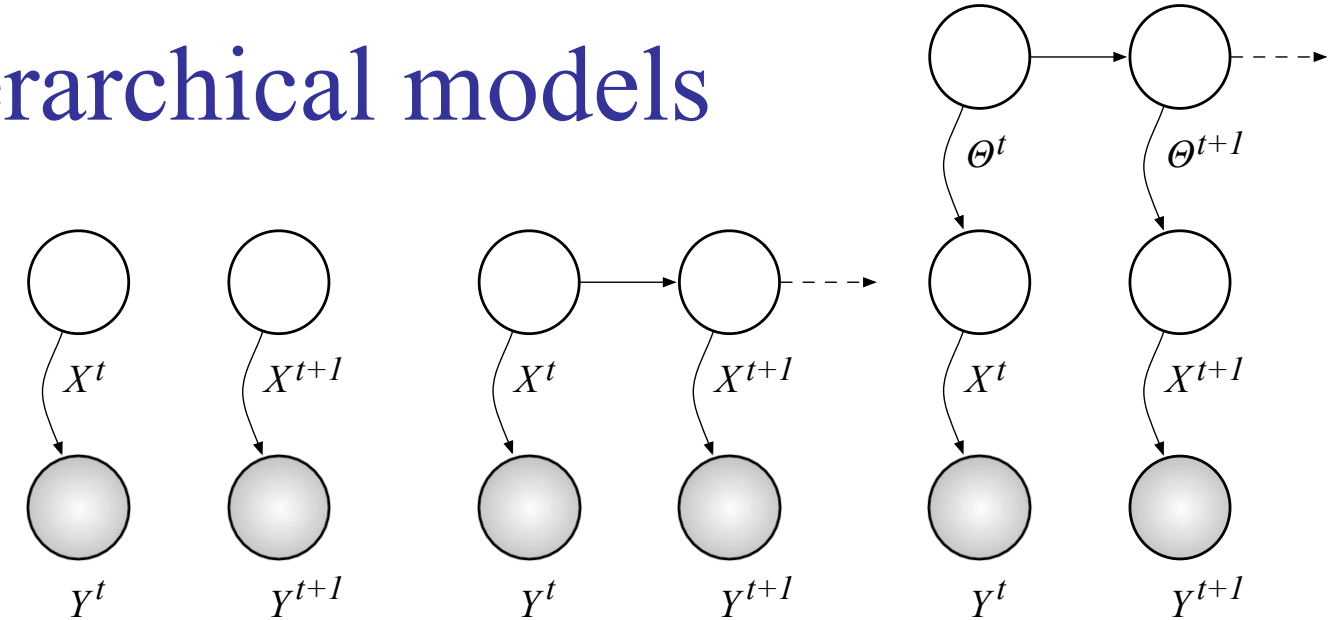
We extend existing models by (1) introducing explicit time dependence; (2) modeling the dynamic as a multiplicative process; and (3) positing non-Gaussian marginals for the OD flows.

Model	**Time Dependence**	**Online Estimation**	**Skewed Marginals**	
Local Likelihood	No	No	No	Cao et al. (2000)
Augmented Gaussian SSM	Yes	Yes	No	This work
Inverse Allocation (static)	No	No	Yes	Tebaldi & West (1998)
Inverse Allocation (dynamic)	Yes	Yes	Yes	This work

Agenda

- The inference problem and its estimands
- Deconvolution of mixing time series on a graph
- Sampling extremely constrained estimands
- Empirical results
- Concluding remarks

An illustration of three hierarchical models



A model with no explicit time dependence (left, factor analysis); a model that introduces an explicit dynamical behavior (center, augmented linear Gaussian state-space model); a model where the explicit time dependence is moved one layer up in the hierarchy, thus allowing for the OD flows to be more variable (right, inverse allocation models).

Modeling sparse, bursty time series

Two modeling strategies with two latent processes $\{\lambda_t : t > 0\}$ and $\{x_t : t > 0\}$

1. Assume λ_t smooth, $x_t|\lambda_t$ independent bursts

   Captures spikes but not sparsity, intractable inference

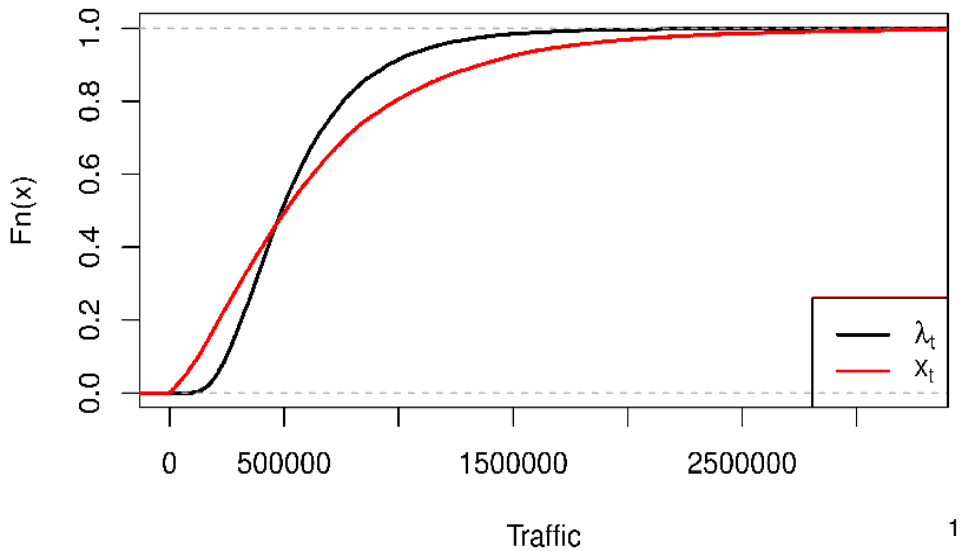
2. Assume λ_t bursty, $x_t|\lambda_t$ error with point mass at 0

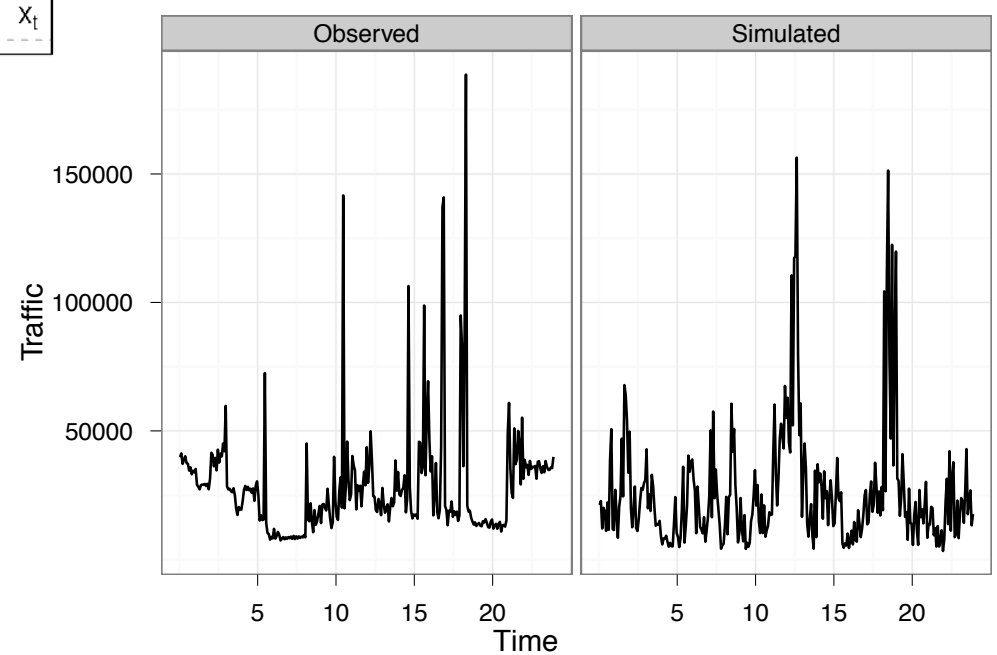
   Captures spikes and sparsity, tractable inference

$$\log \lambda_{i,t} = \rho \log \lambda_{i,t-1} + \varepsilon_{i,t}$$

$$x_{i,t}|\lambda_{i,t}, \phi_t \sim \mathrm{TruncN}_{(0,\infty)}\left(\lambda_{i,t},\ \lambda_{i,t}^{\tau}(\exp(\phi_t) - 1)\right)$$

Qualitative model checking





Inference: Issues and solutions

- Use SIRM filter. Looses efficiency quickly as the dimensionality of the solution space (n-m) grows

  Intuitively: key quantity is the ratio between the volumes of the solution space and the support of the proposal in R^n

- Develop a 2-stage estimation strategy to increase efficiency of the SIRM filter

  Use $E[x_t|y_{1:t}]$ from a Gaussian SSM to calibrate prior for λ_t

  Not EB; in the spirit of EB as in Clogg et al. (JASA '91)

Agenda

- The inference problem and its estimands
- Deconvolution of mixing time series on a graph
- Sampling extremely constrained estimands
- Empirical results
- Concluding remarks

Geometry of ill-posed inverse problems

Consider $Y_{(m\times 1)} = A_{(m\times n)} X_{(n\times 1)}$

Given Y,A we need to find X

- Rewrite $A = [A_1|A_2]$ with $r(A_1) = m$, and $x = [x_1|x_2]$
- The posterior is $p(x \mid y,\lambda) \propto p(x_1 \mid y,\lambda) \cdot I_{y - A_2 x_2}(x_1)$
- Part of the estimand is $x_1 = A_1^{-1}(y - A_2 x_2)$

Solutions x_2 lie in the intersection of a linear space of dim. n-m with the positive orthant: *a convex polytope*

Implications for inference

- Suppose we have a predictive distribution $p(\boldsymbol{x}_t|\cdots)$ that is quasi-concave (uni-modal): all super-level sets are convex, so the set of modes is convex

- Then the posterior update
  $p(\boldsymbol{x}_t|\cdots) \to p(\boldsymbol{x}_t|\boldsymbol{y}_t,\cdots) \propto p(\boldsymbol{y}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\cdots)$
  preserves quasi-concavity (uni-modality).

The posterior remains uni-modal despite reported multi-modality: disconnected modes are not possible

New idea: Polytope samplers

Strategy

1. Leverage HNF to find first vertex
2. Greedily move along the edges to find all vertices (via HNF pivoting)
3. Place a distribution on the polytope; we develop three strategies to do this using Dirichlet dist.

Polytope samplers provide a new exact sampling strategy to address ill-posed inverse problems

Hermite normal form

- Hermite Normal Form of integer matrix A: B=AQ
- Columns of Q_2 generates null-space

$$
\begin{array}{c|c|c|}
\cline{2-3}
m & A_1 & A_2 \\
\cline{2-3}
 & \multicolumn{2}{c}{n}
\end{array}
\quad
\begin{array}{c|c|c|}
\multicolumn{1}{c}{Q_1} & \multicolumn{1}{c}{Q_2} \\
\cline{1-2}
A_1^{-1} & -A_1^{-1}A_2 \\
\cline{1-2}
0 & I \\
\cline{1-2}
\multicolumn{1}{c}{m} & \multicolumn{1}{c}{n-m}
\end{array}
\quad = \quad
\begin{array}{c|c|c|}
\multicolumn{1}{c}{B_1} & \multicolumn{1}{c}{} \\
\cline{1-2}
I & 0 \\
\cline{1-2}
\multicolumn{1}{c}{m} & \multicolumn{1}{c}{n-m}
\end{array}
$$

Almost finding the first vertex

Start from $y = Ax$, with A of size $m \times n$

Define $x' = Q^{-1}x$

Then rewrite $y = AQQ^{-1}x = AQx'$

Notice that $AQ = [I_m \mid 0]$

So $x' = [y \mid 0]$ is a solution!

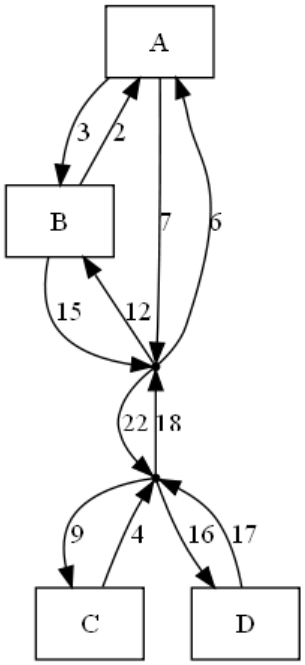
Caveats apply, but it turns out this is a good start

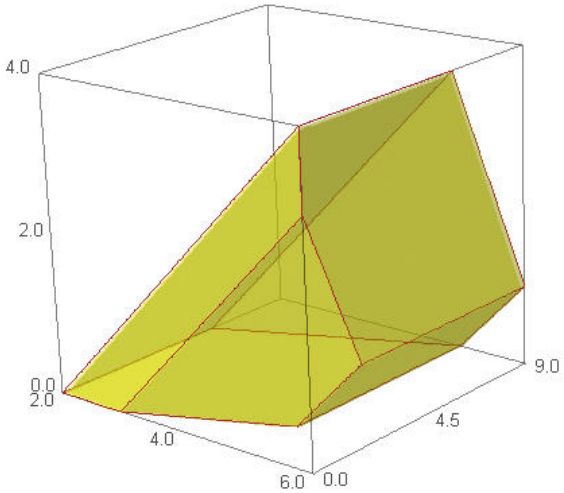
Distributions on a polytope

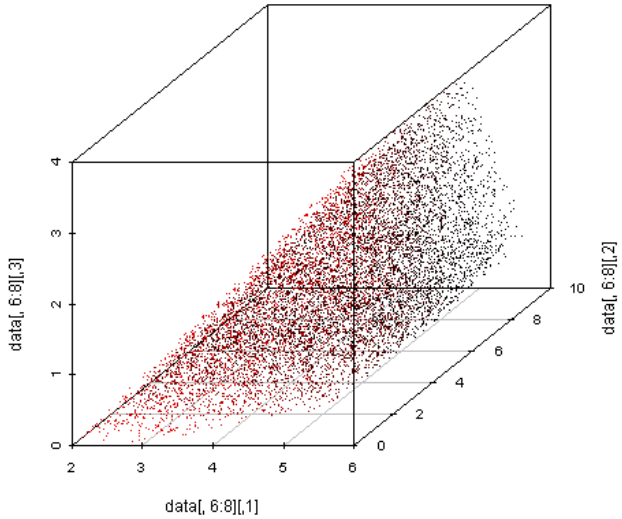
1. Lift the polytope into a higher dimensional simplex, posit a Dirichlet, project back
2. Triangulate the convex polytope into simplices, posit a collection of Dirichlet's weighted by their volumes
3. Direct generalization of Dirichlet that leverages moment map and projective geometry

Illustrative example

Matrix A is 9x12 and leads to a 3D solution space







Agenda

- The inference problem and its estimands
- Deconvolution of mixing time series on a graph
- Sampling extremely constrained estimands
- Empirical results
- Concluding remarks

Two case studies

- Bell Labs: 7 link loads, 16 origin-destination flows
- CMU: 26 link loads, 144 origin-destinations flows

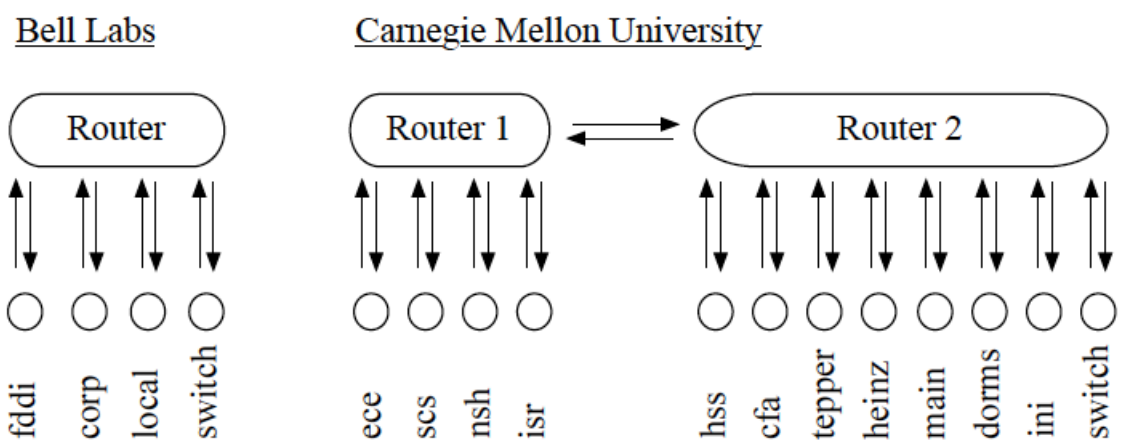


Figure 5: Bell Labs and CMU network structures.

Empirical evaluation

Table 1: Performance comparison (Bell Labs results in Kb, CMU results in Mb. * Denotes our own improvement on the original algorithm by Tebaldi & West)

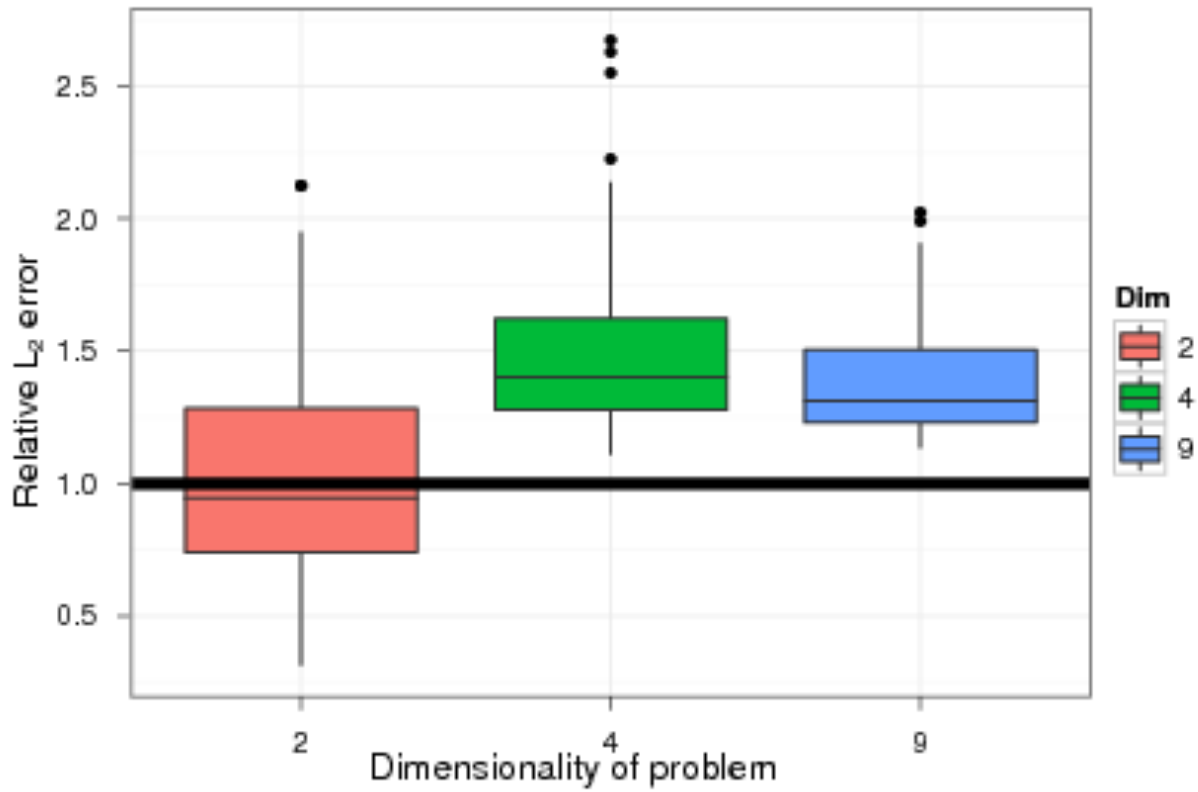
	BELL LABS				CMU		
Method	L_2 **Error**	**SE**	L_1 **Error**	**SE**	L_2 **Error**	**SE**	L_1 **Error**
Locally IID model	104.59	5.54	160.24	6.53	592.49	9.91	1169.15
Smoothed locally IID	104.25	5.52	157.87	6.48	—	—	—
Tebaldi & West (uniform prior)	76.60	4.91	173.94	7.49	—	—	—
Tebaldi & West (joint proposal)*	49.43	2.58	147.66	6.18	167.94	4.42	712.37
Dynamic multilevel model (naïve prior)	63.29	3.35	178.43	8.09	311.21	6.25	1109.68
Calibration model (stage 1)	19.35	0.72	57.66	2.06	110.47	6.19	389.14
Dynamic multilevel model (stage 2)	19.93	0.87	58.20	2.39	93.42	5.20	334.74

Two simulation studies

1. Simulated data for range of network sizes from our model, then evaluated effect of regularization strategy on results

2. Built simulated networks from real CMU traffic data and compared methods' overall performance

Results – regularization strategy



Here, data was simulated from the model. As such, there is no sparsity effect. The plotted variable is the the ratio of average L_2 error for the naïve prior model to that for our two-stage, regularized procedure.

Results – relative performance

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.23	0.12	50.59	0.00
Size = 4	0.003	0.12	0.02	0.98
Size = 5	0.82	0.12	6.71	0.00
Size = 9	1.88	0.12	15.40	0.00
Locally IID	1.76	0.13	13.20	0.00
Gaussian SSM	-0.26	0.13	-1.93	0.06
Naiive prior	-0.02	0.13	-0.16	0.87
Multilevel model	-0.35	0.13	-1.90	0.06
log10(sparsity)	-8.28	3.59	-2.31	0.02

Dependent variable in ANCOVA is $\log_{10}$ of mean (time-averaged) L_1 error in $\boldsymbol{x}_t$.

Agenda

- The inference problem and its estimands
- Deconvolution of mixing time series on a graph
- Sampling extremely constrained estimands
- Empirical results
- Concluding remarks

Concluding remarks

- Deconvolution of mixing time series on a network
- A sequence of ill-posed inverse problems: project low-dim. data to hi-dim./constr. parameter space
- Polytope samplers provide a new strategy to address the hardness of the inference problem

Acknowledgements and pointers

CDC, Facebook, Bell Labs, AT&T. J Blitztein, J Cao, C Carley, C Faloutsos, J Liu, C Marlow, XL Meng, D Rubin, PJ Wolfe, B Yu.

1. Deconvolution of mixing time series on a graph. *Uncertainty in Artificial Intelligence*, 2011. Blocker & EA.

2. Polytope samplers for inference in ill-posed inverse problems. *AI & Statistics*, 2011. EA & Haas.

3. A survey of statistical network models. *Foundations & Trends in Machine Learning*, 2009. Goldenberg, Zheng, Fienberg & EA

4. Network sampling and classification: An investigation of network model representations. *Decision Support Systems*, 2011. EA, Bai & Carley.

BACK UP

Naïve inference strategy

Sample-Importance-Resample-Move algorithm

for $t \leftarrow t$ **to** T **do**

Sample step:

for $j \leftarrow 1$ **to** m **do**

Draw a proposal $\log \lambda_{i,t}^{(j)*} \sim \mathrm{N}(\theta_{1\,i,t} + \log \lambda_{i,t-1}^{(j)}, \theta_{1\,i,t})$

Draw $\phi_t^{(j)} \sim \mathrm{Gamma}(\alpha, \hat{\phi}_t/\alpha)$

Draw $\boldsymbol{x}_t^{(j)*}$ from a truncated Normal distribution with mean $\boldsymbol{\mu}^* = \boldsymbol{\theta}_{1\,t} + \rho/m \sum_{j=1}^{m} \boldsymbol{\lambda}_{t-1}^{(j)}$ and covariance matrix $\Sigma^* = (\exp(\hat{\phi}_t) - 1)\mathrm{diag}\left(\boldsymbol{\mu}^{*2}\right)$ on the feasible region given by $\boldsymbol{x}_t^{(j)*} \geq 0$, $y_t = A\boldsymbol{x}_t^{(j)*}$ using Algorithm 2

Resample our particles $(\boldsymbol{\lambda}_t^{(j)*}, \phi_t^{(j)*}, \boldsymbol{x}_t^{(j)*})$ with probabilities proportional to our weights $w_t^{(j)}$

Move each of our resampled particles $(\boldsymbol{\lambda}_t^{(j)}, \phi_t^{(j)}, \boldsymbol{x}_t^{(j)})$ using a MCMC algorithm (Metropolis-Hastings within Gibbs, with proposal on $\boldsymbol{x}_t$ given by Algorithm 2)

return $(\lambda_t^{(j)}, \phi_t^{(j)}, \boldsymbol{x}_t^{(j)})$ for $j \leftarrow 1$ **to** m, $t \leftarrow 1$ **to** T

Algorithm 1: SIRM algorithm for inference with multilevel state-space model

Extremely constrained solution space

Random Directions Algorithm

Initialization

begin

Decompose A into $[A_1\ A_2]$, A_1 $(r \times r)$ full-rank as in Tebaldi and West (1998)

Store $B := A_1^{-1}$; $C := A_1^{-1} A_2$

Metropolis step

given $\boldsymbol{x}_t$

begin

Draw $\boldsymbol{z} \sim N(0, I)$, $\boldsymbol{z} \in \mathbb{R}^{c-r}$

Set $\boldsymbol{d} := \boldsymbol{z}/\|\boldsymbol{z}\|$

Calculate $\boldsymbol{w} := C \cdot d$

Set $h_1 := \max\{\min_{k:w_k>0}(\boldsymbol{x}_{1,t})_k/w_k, 0\}$

Set $h_2 := \max\{\min_{k:d_k<0} -(\boldsymbol{x}_{2,t})_k/d_k, 0\}$

Set $h := \min\{h_1, h_2\}$

Set $l_1 := \max\{\max_{k:w_k<0}(\boldsymbol{x}_{1,t})_k/w_k, 0\}$

Set $l_2 := \max\{\max_{k:d_k>0} -(\boldsymbol{x}_{2,t})_k/d_k, 0\}$

Set $l := \max\{l_1, l_2\}$

Draw $u \sim \mathrm{Unif}(l, h)$

Set $\boldsymbol{x}^*_{2,t} := \boldsymbol{x}_{2,t} + u \cdot \boldsymbol{d}$; $\boldsymbol{x}^*_{1,t} = \boldsymbol{x}_{1,t} - u \cdot w$; $\boldsymbol{x}^*_t = (\boldsymbol{x}^*_{1,t}, \boldsymbol{x}^*_{2,t})$

Set $\boldsymbol{x}_t := \boldsymbol{x}^*_t$ with probability $\min\{f(\boldsymbol{x}^*_t)/f(\boldsymbol{x}_t), 1\}$

return $\boldsymbol{x}_t$

Algorithm 2: RDA algorithm for sampling from $f(\boldsymbol{x}_t)$, truncated to the feasible region given by $A \cdot \boldsymbol{x}_t = \boldsymbol{y}_t$