Thompson sampling for large decision problems

Eric B. Laber

Department of Statistics, North Carolina State University

CRM 2018

Acknowledgments

Thanks to

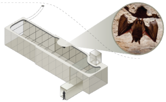
- Erica Moodie
- NCSU and UNC colleagues
- National Science Foundation
- National Institutes of Health
- Department of Natural Resources

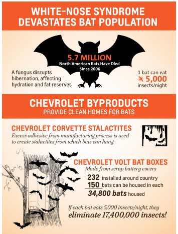
Joint work with

- John Drake
- Tao Hu
- Nick Meyer
- Krishna Pacifici
- Brian Reich

Example: white-nose syndrome

- Headlines
  - *Fungus that's killing millions of bats isn't going away*
    -Los Angeles Times, Nov. 1, 2013
  - *The secret bataclysm: white nose syndrome and extinction*
    -Wired, Aug. 12, 2014
  - *Relentless bat plague not abating*
    -Athens Herald, July 27, 2017





Bat images from AP photo, poster from 3blmedia.com

Example: white-nose syndrome cont'd

USGS
science for a changing world

SCIENCE PRODUCTS NEWS CONNECT ABOUT

Search

Deadly Fungus Affecting Hibernating Bats Could Spread During Summer

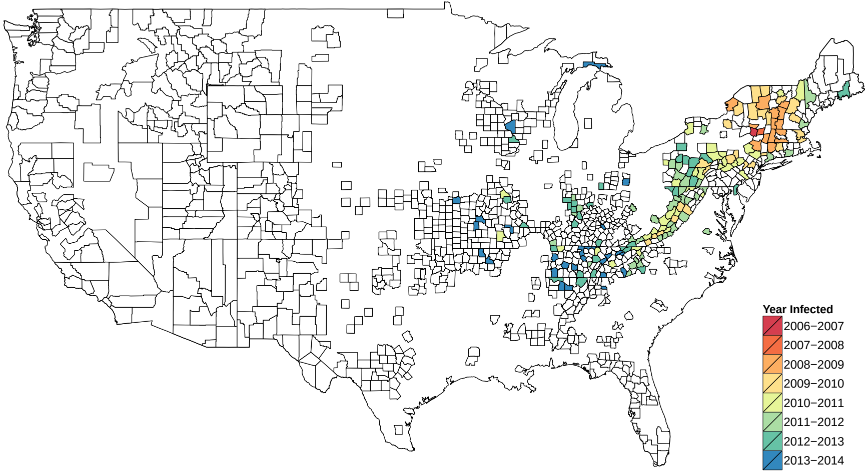
Release Date: AUGUST 1, 2017

The cold-loving fungus (Pseudogymnoascus destructans, or Pd) that causes white-nose syndrome, a disease that has killed millions of North American bats during hibernation, could also spread in summer months. Bats and humans visiting contaminated caves and mines can

Contacts

Department of the Interior,
U.S. Geological Survey

Example: white-nose syndrome cont'd



- Allocation strategy: map from current epidemic status to subset of counties marked 'high priority for treatment'

Additional motivating examples

- Other spatio-temporal allocation problems:
  - Spread of human disease
  - Ecological management
  - Precision agriculture
  - ...

Major challenges

- Data
  - No data on intervention effectiveness at $t = 1$
  - Noisy, incomplete, and sparse measurements
  - High-dimensional, in WNS, $p = 11,370$ per time point
- Estimation and inference
  - Online estimation means allocations are selected from a continually changing (data-dependent) strategy $\Rightarrow$ sampling distns are complex
- Computation
  - There are 2^L possible allocations at each time point
  - Enumeration is not possible

Outline

- Defining an optimal allocation strategy
- Estimation with a parametric dynamics model
- Estimation without a dynamics model
- Simulation experiments

Outline

- Defining an optimal allocation strategy
- Estimation with a parametric dynamics model
- Estimation without a dynamics model
- Simulation experiments

Setup and notation

- Allocation problem evolves over
  - Locations $\mathcal{L} = \{1, 2, \ldots, L\}$
  - Time points $\mathcal{T} = \{1, 2, \ldots\}$
- At each time t and location ℓ
  - Observe state $\boldsymbol{S}_\ell^t \in \mathbb{R}^p$
  - Observe outcome $Y_\ell^t \in \mathbb{R}$
  - Select allocation $A_\ell^t \in \{0, 1\}$
  - Define $\boldsymbol{S}^t = \left\{\boldsymbol{S}_\ell^t\right\}_{\ell \in \mathcal{L}}$, $\boldsymbol{A}^t = \{A_\ell^t\}_{\ell \in \mathcal{L}}$, and $\boldsymbol{Y}^t = \{Y_\ell^t\}_{\ell \in \mathcal{L}}$
- Assume allocation based on $\boldsymbol{S}^t$ at time t

Stochastic allocation strategy

- Random allocations needed to learn in online setting
- Let $\mathcal{B}_L$ denote all distributions over $\{0,1\}^L$ and $\mathcal{S} = \operatorname{dom} \boldsymbol{S}^t$
- Allocation strategy π is a map

$$\pi : \mathcal{S} \to \mathcal{B}_L,$$

under π, a decision maker presented with $\boldsymbol{S}^t = \boldsymbol{s}^t$ will select allocation $\boldsymbol{a}^t$ with probability $\pi(\boldsymbol{a}^t; \boldsymbol{s}^t)$

Class of potential strategies

- Focus on estimation within class of strategies Π
  - Enforce feasibility and cost constraints
  - Add'l considerations: parsimony, interpretability, and scalability
- Ex. mixture over deterministic class of strategies

$$\pi(\boldsymbol{a}; \boldsymbol{s}) = \int_{e \in \Delta} 1\left\{e(\boldsymbol{s}) = \boldsymbol{a}\right\} dP(e),$$

where Δ set of maps from $\mathcal{S}$ into $\{0, 1\}^{L}$ and P a distn over Δ

Optimal strategy via potential outcomes

- Potential outcomes under $\overline{\boldsymbol{a}}^t = (\boldsymbol{a}^1, \ldots, \boldsymbol{a}^t)$

$$\boldsymbol{W}^* = \left\{ \boldsymbol{Y}^{*t}(\overline{\boldsymbol{a}}^t), \boldsymbol{S}^{*(t+1)}(\overline{\boldsymbol{a}}^t) \,:\, \overline{\boldsymbol{a}}^t \in \{0,1\}^{t \times L} \right\}_{t \in \mathcal{T}}$$

- Potential outcome under strategy π

$$\boldsymbol{Y}^{*t}(\pi) = \sum_{\overline{\boldsymbol{a}}^t} \boldsymbol{Y}^{*t}(\overline{\boldsymbol{a}}^t) \prod_{v=1}^{t} \mathcal{I}\left[\xi_\pi^v \left\{ \boldsymbol{S}^{*v}(\overline{a}^{v-1}) \right\} = \boldsymbol{a}^v \right],$$

where $\{\xi_\pi^t(\boldsymbol{s})\}_{t \in \mathcal{T}, \boldsymbol{s} \in \mathcal{S}}$ collection independent r.v.s with $P\left\{\xi_\pi^v(\boldsymbol{s}^v) = \boldsymbol{a}^v\right\} = \pi(\boldsymbol{a}^v; \boldsymbol{s}^v)$

Optimal strategy

- Discounted marginal mean outcome under π is

$$V(\pi) = \mathbb{E}\left[\sum_{t \in \mathcal{T}} \gamma^{t-1} u\left\{\boldsymbol{Y}^{*t}(\pi)\right\}\right],$$

where $\gamma \in (0, 1)$ and $u(\cdot)$ a utility fn

- Optimal strategy satisfies $V(\pi^{\mathrm{opt}}) \geq V(\pi)$ for all $\pi \in \Pi$

Optimal strategy via generative model

- Let $\boldsymbol{H}^t$ be history at time t, i.e., $\boldsymbol{H}^t = (\overline{\boldsymbol{S}}^t, \overline{\boldsymbol{A}}^{t-1}, \overline{\boldsymbol{Y}}^{t-1})$
- Standard assumptions:
  - (C1) Sequential ignorability: $\boldsymbol{A}^t \perp\!\!\!\perp \boldsymbol{W}^* \mid \boldsymbol{H}^t$
  - (C2) Consistency: $\boldsymbol{Y}^t = \boldsymbol{Y}^{*t}(\overline{\boldsymbol{A}}^t)$, $\boldsymbol{S}^t = \boldsymbol{S}^{*t}(\overline{\boldsymbol{A}}^{t-1})$
  - (C3) Positivity: there exists $\epsilon > 0$ s.t. $P\left\{\boldsymbol{A}^t = \boldsymbol{a}^t \mid \boldsymbol{H}^t\right\} \geq \epsilon$ with probability one for all Π-feasible $\boldsymbol{a}^t$

Optimal strategy via generative model

- Let $\boldsymbol{H}^t$ be history at time t, i.e., $\boldsymbol{H}^t = (\overline{\boldsymbol{S}}^t, \overline{\boldsymbol{A}}^{t-1}, \overline{\boldsymbol{Y}}^{t-1})$
- Standard assumptions:

(C1) Sequential ignorability: $\boldsymbol{A}^t \perp\!\!\!\perp \boldsymbol{W}^* \mid \boldsymbol{H}^t$

(C2) Consistency: $\boldsymbol{Y}^t = \boldsymbol{Y}^{*t}(\overline{\boldsymbol{A}}^t)$, $\boldsymbol{S}^t = \boldsymbol{S}^{*t}(\overline{\boldsymbol{A}}^{t-1})$

(C3) Positivity: there exists $\epsilon > 0$ s.t. $P\left\{\boldsymbol{A}^t = \boldsymbol{a}^t \mid \boldsymbol{H}^t\right\} \geq \epsilon$ with probability one for all Π-feasible $\boldsymbol{a}^t$

Spillover effects due to spatial proximity violate SUTVA $\Rightarrow$ experimental unit is the entire collection of locations $\mathcal{L}$

Optimal strategy via generative model cont'd

- Under std causal assumptions, $\mathbb{E}\left[u\left\{\boldsymbol{Y}^{*t}(\pi)\right\}\right]$ is

$$\int u(\boldsymbol{y}^t)\prod_{v=1}^{t}\left[f_v\left(\boldsymbol{y}^v \mid \boldsymbol{h}^v, \boldsymbol{a}^v\right)\pi(\boldsymbol{a}^v; \boldsymbol{s}^v) g_v\left(\boldsymbol{h}^v \mid \boldsymbol{h}^{v-1}\right)\right] d\lambda(\overline{\boldsymbol{y}}^t, \overline{\boldsymbol{a}}^t, \overline{\boldsymbol{h}}^t),$$

where f_v is cond'l density of $\boldsymbol{Y}^v$, g_v is cond'l density of $\boldsymbol{H}^v$

  - One observation per time point $\Rightarrow$ cannot identify f_v, g_v
  - Even if $f_v \equiv f$ and $g_v \equiv g$ nonparametric estimation is essentially impossible because $\boldsymbol{a}^t \in \{0,1\}^L$
  - Need assumptions that allow pooling over locations

Outline

- Defining an optimal allocation strategy
- Estimation with a parametric dynamics model
- Estimation without a dynamics model
- Simulation experiments

Parametric estimation

- Assume Markovian, low-dimensional parametric models for system dynamics
  - Conditional density for outcome $f_v\left(\boldsymbol{y}^v \mid \boldsymbol{h}^v, \boldsymbol{a}^v\right) = f\left(\boldsymbol{y}^v \mid \boldsymbol{s}^v, \boldsymbol{a}^v; \beta\right)$
  - Conditional density for <u>state</u> $g_v\left(\boldsymbol{s}^v \mid \boldsymbol{h}^{v-1}\right) = g(\boldsymbol{s}^v \mid \boldsymbol{s}^{v-1}, \boldsymbol{a}^{v-1}; \theta)$
- Use simulation-optimization to estimate π^{opt}

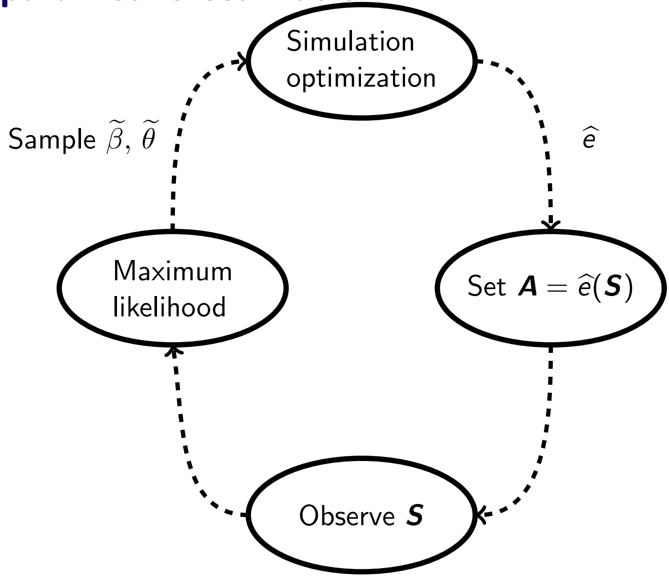
(S1) Constructed estimators $\widehat{\beta}$, $\widehat{\theta}$ of β, θ

(S2) Draw $\widetilde{\beta}, \widetilde{\theta}$ from sampling distn of $\widehat{\beta}, \widehat{\theta}$

(S3) Simulate process under $\widetilde{\beta}, \widetilde{\theta}$ to form estimator $\widehat{V}(\cdot)$ of $V(\cdot)$

(S4) Stochastic programming to compute $\widehat{e} = \arg\max_{e \in \Delta} \widehat{V}(e)$ (recall Δ is class of deterministic strategies)

Online parametric estimation



Note: this is Thompson sampling (Thompson, 1933).

Convergence of online parametric estimation

- Overview of technical assumptions:

  (A1) Parametric model correctly specified

  (A2) π^{opt} is in the strategy space Π

  (A3) Parameter space compact

  (A4) Likelihood is sufficiently smooth

  (A5) Equicontinuity condition (Heijmans and Magnus, 1986)

Convergence of online parametric estimation

- Overview of technical assumptions:

  (A1) Parametric model correctly specified

  (A2) π^{opt} is in the strategy space Π

  (A3) Parameter space compact

  (A4) Likelihood is sufficiently smooth

  (A5) Equicontinuity condition (Heijmans and Magnus, 1986)

Theorem

Let $\widehat{\pi}^t$ be the Thompson sampling estimator at time t and assume (A1)-(A5). If η^t satisfies $\eta^t \to \infty$ and $\eta^t = o(t)$ then

$$\left|V(\widehat{\pi}^t) - V(\pi^{\mathrm{opt}})\right| = O_P\left\{\sqrt{\frac{\eta^t}{t}} + \gamma^{\eta^t}\right\}.$$

Probability bound

Definition

Let $\pi^{\text{opt}}_{\beta,\theta} = \arg\max_{\pi \in \Pi} V_{\beta,\theta}(\pi)$ and

$$R_{\beta^*,\theta^*}(\epsilon) = \sup_{(\beta,\theta)\in\mathcal{N}_\epsilon(\beta^*,\theta^*)} V_{\beta^*,\theta^*}(\pi^{\text{opt}}) - V_{\beta^*,\theta^*}(\pi^{\text{opt}}_{\beta,\theta}).$$

Theorem

Assume (A1)-(A5). Then for $\delta > 0$ *and* $\eta^t \to \infty$ *s.t.* $\eta^t = o(t)$

$$P\left\{V(\pi^{\text{opt}}) - V(\widehat{\pi}^t) > \delta\right\} \le k_1 \exp\left[-k_2 t \left\{R^{-}_{\beta^*,\theta^*}(\delta/3)\right\}^2\right]$$
$$+ k_3 \exp\left\{-\frac{t}{(\eta^t)^2}\left[(1-\gamma^{\eta^t})\left\{\frac{1}{\sqrt{\eta^t}} + k_4 R^{-}_{\beta^*,\theta^*}(\delta/3)\right\}\right]^{-2}\delta^2 k_5\right\}$$

where $k_1, \ldots, k_5$ *are positive constants.*

Parametric estimation discussion

- Positives
  - Intuitive
  - Leverage existing scientific theory
  - Low variance, applies in data impoverished setting
  - Extensions to non-stationary setting via state-space models
  - Asymptotic optimality (under strong conditions)
- Drawbacks: the devil is in the details
  - Potentially high bias
  - Computation is non-trivial, requires climbing up-hill in the strategy space Δ using stochastic programming
  - Induced Π is a mixture over Δ

Outline

- Defining an optimal allocation strategy
- Estimation with a parametric dynamics model
- Estimation without a dynamics model
- Simulation experiments

Is a system dynamics model necessary?

- Parametric model suited for large, data-impoverished setting
  - As data accumulate can weaken dependence on model
  - Complete dynamics not needed to determine π^{opt}
  - Idea: apply Thompson sampling with model until sufficient data then 'switch' to semi-parametric estimator of π^{opt}

Semi-parametric estimation

- For any $e \in \Delta$ define Q-function

$$Q^e(\boldsymbol{s}, \boldsymbol{a}) = \mathbb{E}\left[\sum_{k \geq 1} \gamma^{k-1} u\left\{\boldsymbol{Y}^{*(t+k-1)}(e)\right\} \mid \boldsymbol{S}^t = \boldsymbol{s}, \boldsymbol{A}^t = \boldsymbol{a}\right],$$

- Define $V_R(e) = \int Q^e\left\{\boldsymbol{s}, e(\boldsymbol{s})\right\} dR(s)$
  - R is reference distribution
  - If R is distn of S^1 then $V_R(e) = V(e)$

Building an estimating equation

- Under previous causal assumptions, for any $z(s, a)$

$$\mathbb{E}\left[\left\{u(Y^t) + \gamma Q^e\left\{\boldsymbol{S}^{t+1}, e(\boldsymbol{S}^{t+1})\right\} - Q^e(\boldsymbol{S}^t, \boldsymbol{A}^t)\right\} z(\boldsymbol{S}^t, \boldsymbol{A}^t)\right] = 0,$$

  where expectation is wrt to data-generating model

- Above forms basis for estimating equation[1]
  - Postulate parametric model $Q^e(\boldsymbol{s}, \boldsymbol{a}) = Q(\boldsymbol{s}, \boldsymbol{a}; \lambda^e)$
  - Set $z(s, a) = \nabla_{\lambda^e} Q(s, a; \lambda^e)$

[1] Maei et al., 2010; Ertefaie, 2016; Luckett et al., 2017

Semi-parametric estimation cont'd

- Let $M_1^t, \ldots, M_{t-1}^t$ be triangular-array of bootstrap weights
- For each $e \in \Delta$ define

$$\Lambda^{t(M)}(\lambda, e) = \sum_{v=1}^{t-1} M_v^t \Big\{ u\left(\boldsymbol{Y}^v\right) + \gamma Q\left\{\boldsymbol{S}^{v+1}, e(\boldsymbol{S}^{v+1}); \lambda\right\} - Q(\boldsymbol{S}^v, \boldsymbol{A}^v; \lambda) \Big\} \nabla_\lambda Q_t(\boldsymbol{S}^v, \boldsymbol{A}^v; \lambda)$$

then $\Lambda^{t(M)}(\lambda, e) = 0$ is a perturbed estimating equation for λ^e

Semi-parametric estimation cont'd

- Penalized estimator

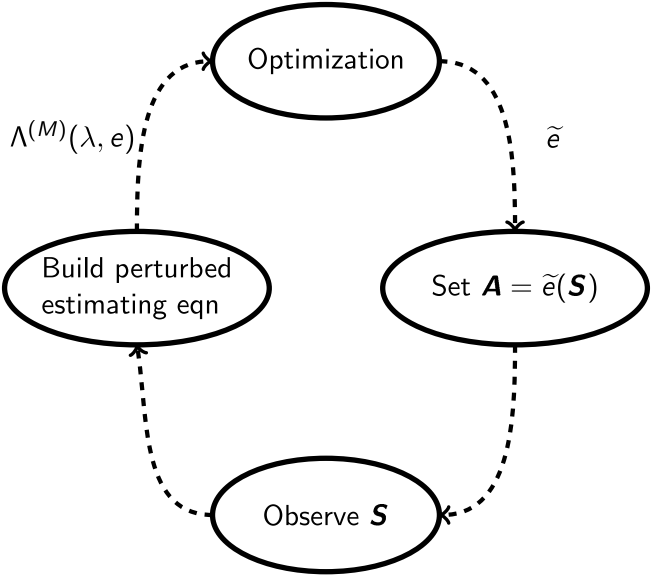
$$\widehat{\lambda}^{t,e} = \arg\min_{\lambda} \Lambda^{t(M)}(\lambda, e)^{\intercal} \Sigma^{t} \Lambda^{t(M)}(\lambda, e) + \tau \mathcal{P}(\lambda),$$

where $\tau > 0$ is a tuning parameter and Σ^t p.d. matrix

- Semi-parametric estimator of π^{opt}

$$\widetilde{e}^{t} = \arg\max_{e \in \Delta} \int Q \left\{ \boldsymbol{s}, e(\boldsymbol{s}); \widehat{\lambda}^{t,e} \right\} dR(s)$$

Online semi-parametric estimation



Note: this is a semi-parametric variant of Thompson sampling.

Semi-parametric estimation discussion

- Positives
  - Low bias
  - Fewer underlying assumptions than parametric model
  - Generally much faster than simulation optimization
  - Extensions to non-stationary setting via state-space models
  - Penalty can be used to impose structure
  - Also applies to V-learning and related methods
- Drawbacks
  - High variance, requires more data
  - Incorporation of scientific theory more limited
  - More complex

Anchoring the semi-parametric model

- Positives and drawbacks of the parametric and semi-parametric models are complimentary
- Idea: use parametric model until sufficient data accrues then switch to semi-parametric model
  - Trade-off bias and variance
  - We call this anchoring the semi-parametric model

Outline

- Defining an optimal allocation strategy
- Estimation with a parametric dynamics model
- Estimation without a dynamics model
- Simulation experiments

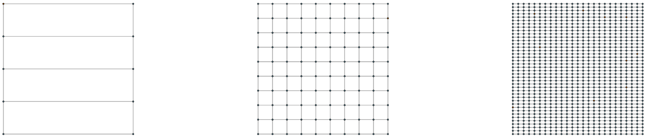
Simulation experiments: overview

- Simulate spread of disease across nodes in a network
  - Randomly select 1% of nodes to infect at baseline
  - No interventions for 6 time steps
  - Select 6% of nodes for treatment at $t = 7, \ldots, 15$
- Generative model built from white-nose syndrome data
  - Variant of gravity model (Maher et al., 2012), probability that node i infects node j is linear on logit scale
  - Coefficients estimated using white-nose data then scaled s.t. spread to 70% of the network at $t = 15$ under no treatment
  - Additional initial state values at each node generated from Normal$(\mathbf{0}_{10}, \mathbf{I}_{10})$
- Goal: minimize number infected nodes at $t = 15$

Simulation experiments: setup

- Competing allocation strategies
  - NoTxt: no treatment
  - Myopic: allocate to nodes with highest predicted probability of infection at next time point (Ihlo and Baker, 2013)
  - Proximal: ad hoc strategy proposed by USFWS that treats locations on the 'border' of a spreading infection
  - SimOpt: simulation optimization with linear ranks
  - Anchored: switch at halfway point, spatial linear model
- Consider correct and incorrect dynamics models
  - Correct: time-dependent SI model (Maher et al., 2012)
  - Incorrect: estimated dynamics only depend on distance
- Use probit rank class of strategies with linear features constructed from postulated dynamics model

Simulation: grid example



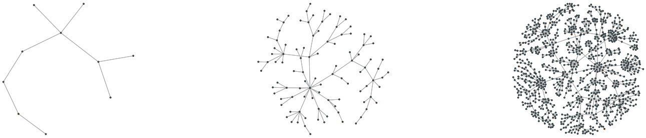
Correctly specified dynamics model

Nodes	NoTxt	Myopic	Proximal	SimOpt	Anchored
100	0.70	0.54	0.54	0.31	0.33
500	0.70	0.56	0.59	0.41	0.42
1000	0.70	0.56	0.60	0.43	0.44

Incorrectly specified dynamics model

Nodes	NoTxt	Myopic	Proximal	SimOpt	Anchored
100	0.69	0.53	0.54	0.44	0.35
500	0.70	0.62	0.59	0.55	0.44
1000	0.70	0.65	0.60	0.55	0.45

Simulation: scale-free example



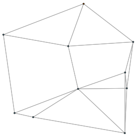
Correctly specified dynamics model

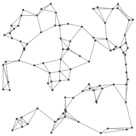
Nodes	NoTxt	Myopic	Proximal	SimOpt	Anchored
100	0.70	0.64	0.56	0.51	0.52
500	0.70	0.61	0.62	0.47	0.47
1000	0.70	0.63	0.64	0.52	0.53

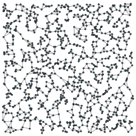
Incorrectly specified dynamics model

Nodes	NoTxt	Myopic	Proximal	SimOpt	Anchored
100	0.70	0.65	0.56	0.52	0.51
500	0.70	0.63	0.62	0.53	0.50
1000	0.70	0.66	0.64	0.55	0.53

Simulation: random 3-nearest neighbors example







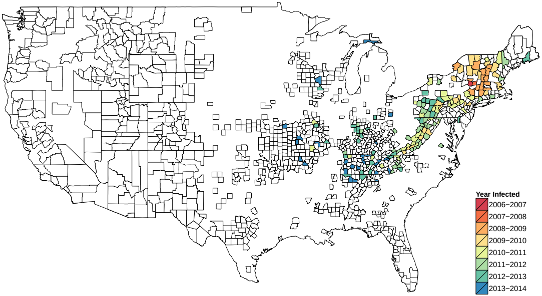
Correctly specified dynamics model

Nodes	NoTxt	Myopic	Proximal	SimOpt	Anchored
100	0.70	0.36	0.31	0.27	0.29
500	0.70	0.45	0.37	0.31	0.32
1000	0.70	0.53	0.48	0.46	0.46

Incorrectly specified dynamics model

Nodes	NoTxt	Myopic	Proximal	SimOpt	Anchored
100	0.70	0.45	0.31	0.41	0.29
500	0.70	0.55	0.37	0.45	0.31
1000	0.70	0.58	0.48	0.54	0.46

Simulation: white-nose syndrome



Correctly specified dynamics model

Nodes	NoTxt	Myopic	Proximal	SimOpt	Anchored
1137	0.60	0.40	0.45	0.30	0.35

Incorrectly specified dynamics model

Nodes	NoTxt	Myopic	Proximal	SimOpt	Anchored
1137	0.60	0.54	0.45	0.49	0.34

Discussion

- Management of epidemics can be posed as spatio-temporal allocation problems
- An effective allocation strategy:
  - Accounts for spillover effects (no SUTVA)
  - Implemented online
  - Computationally feasible
  - Accommodates evolving logistical constraints
- We examined parametric and semi-parametric estimators

Discussion cont'd

- Many open and exciting problems
  - Imperfect measurement/detection
  - Scaling to networks with millions or billions of nodes (we currently have heuristics that apply to $\sim$10 million nodes)
  - Sieves for non-parametric estimation
  - Choose amount of stochasticity in allocation strategy to optimize exploration/exploitation trade-off
  - Unknown and/or evolving network structure

Thank you.

laber@stat.ncsu.edu

laber-labs.com