College of Science - Department of Statistics
Statistical Bioinformatics Center

Reverse Engineering Gene Networks Using Approximate Bayesian Computation (ABC)

Rencontre de statistique autour des modèles hiérarchiques: Université de Strasbourg

Andrea Rau

January 14, 2011

Outline

1. Introduction
   - Gene regulatory networks
2. Approximate Bayesian Computation
   - Background and motivation
   - Monte Carlo approaches
3. ABC-MCMC for Networks
   - Simulation studies
   - Real data analysis: SOS DNA repair system in *E. coli*
4. Discussion

Gene Expression

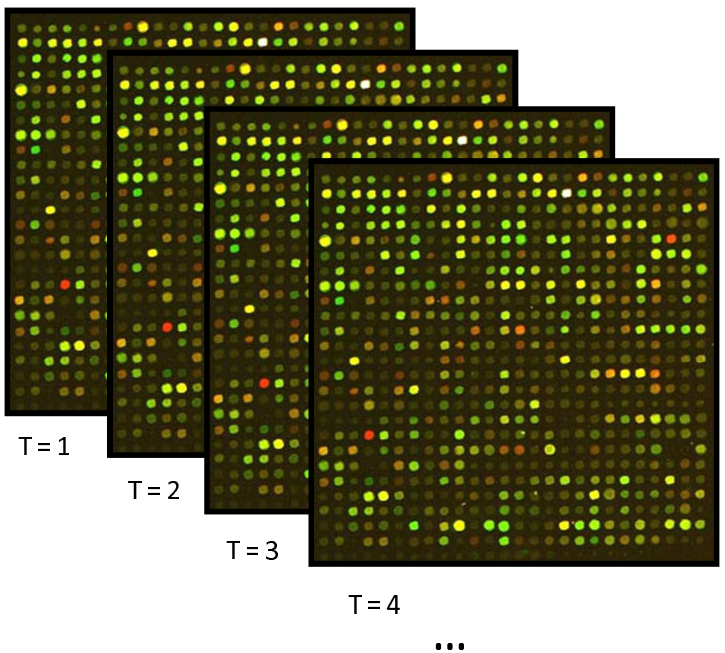
- Genes: Functional regions of DNA that encode proteins and RNA molecules

Gene Expression

- Genes: Functional regions of DNA that encode proteins and RNA molecules
- Expression levels of thousands of genes can be measured using "high-throughput" technologies (e.g., microarrays, serial analysis of gene expression, next-generation sequencing)

Time-Course Gene Expression

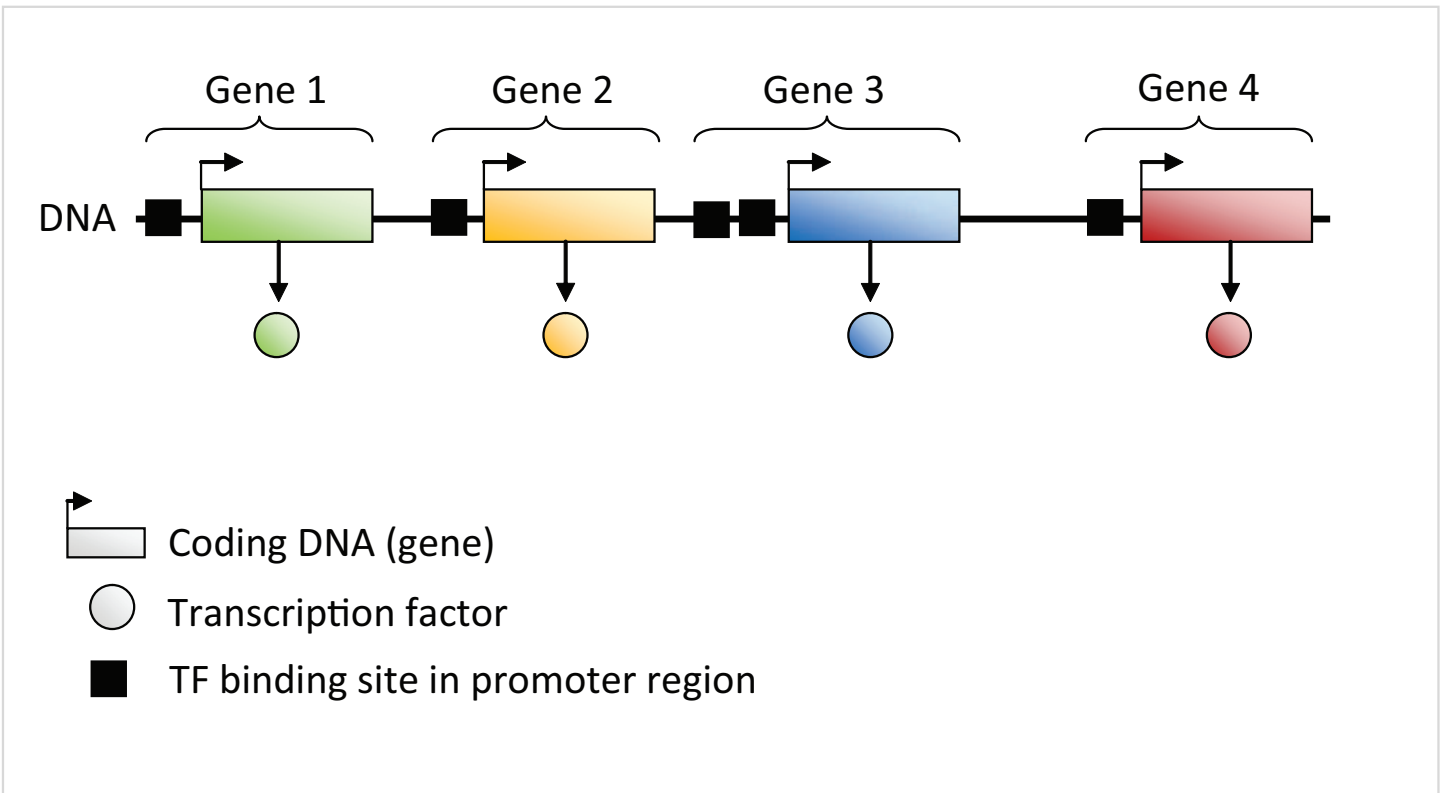
- Time-course gene expression data can elucidate information about *patterns* of relationships of gene expression in a cell



- Large number of genes, few biological replicates or time points... $N \ll P$ paradigm

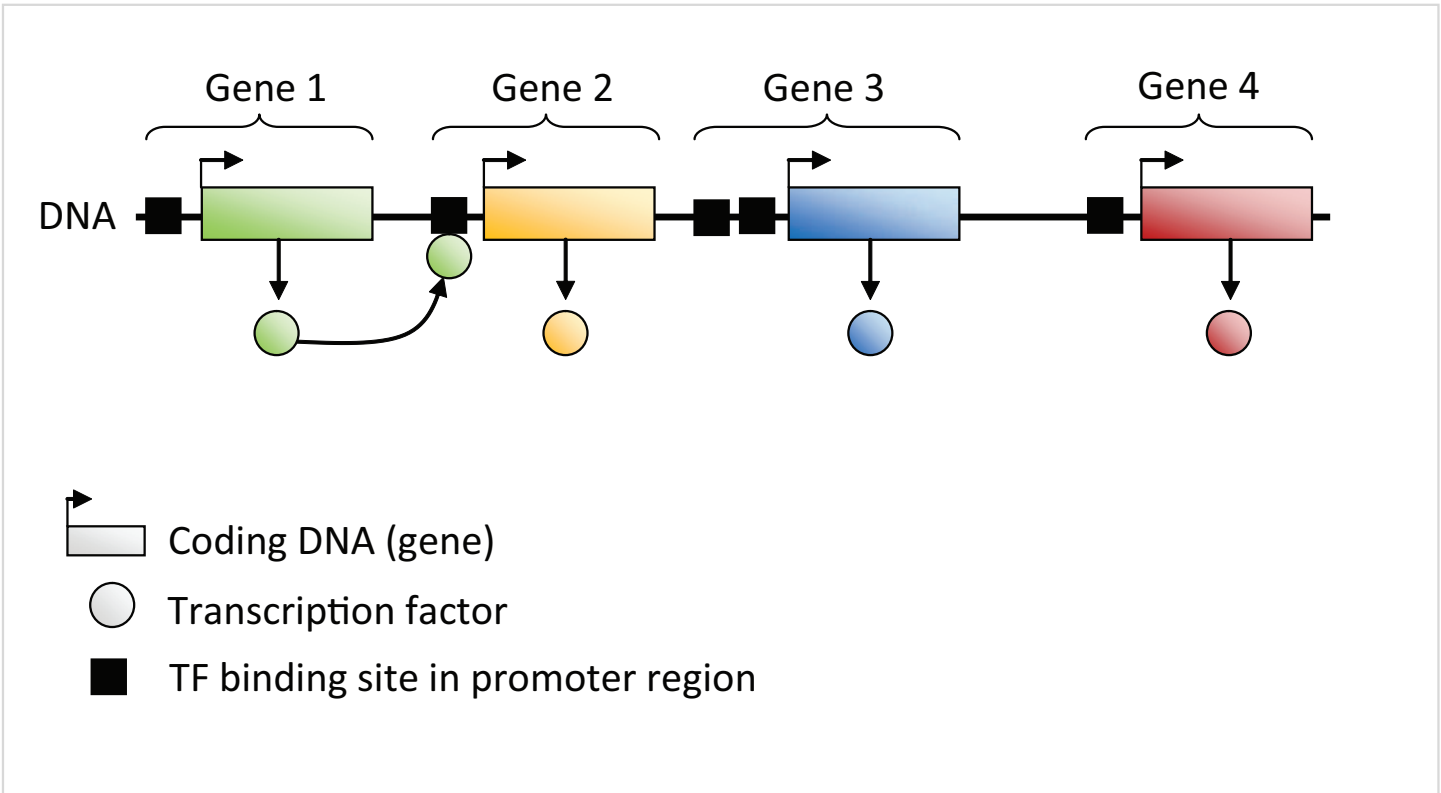
Gene Regulatory Networks

- *Gene regulatory networks*: set of genes that interact indirectly with one another through proteins called transcription factors (TF)



Gene Regulatory Networks

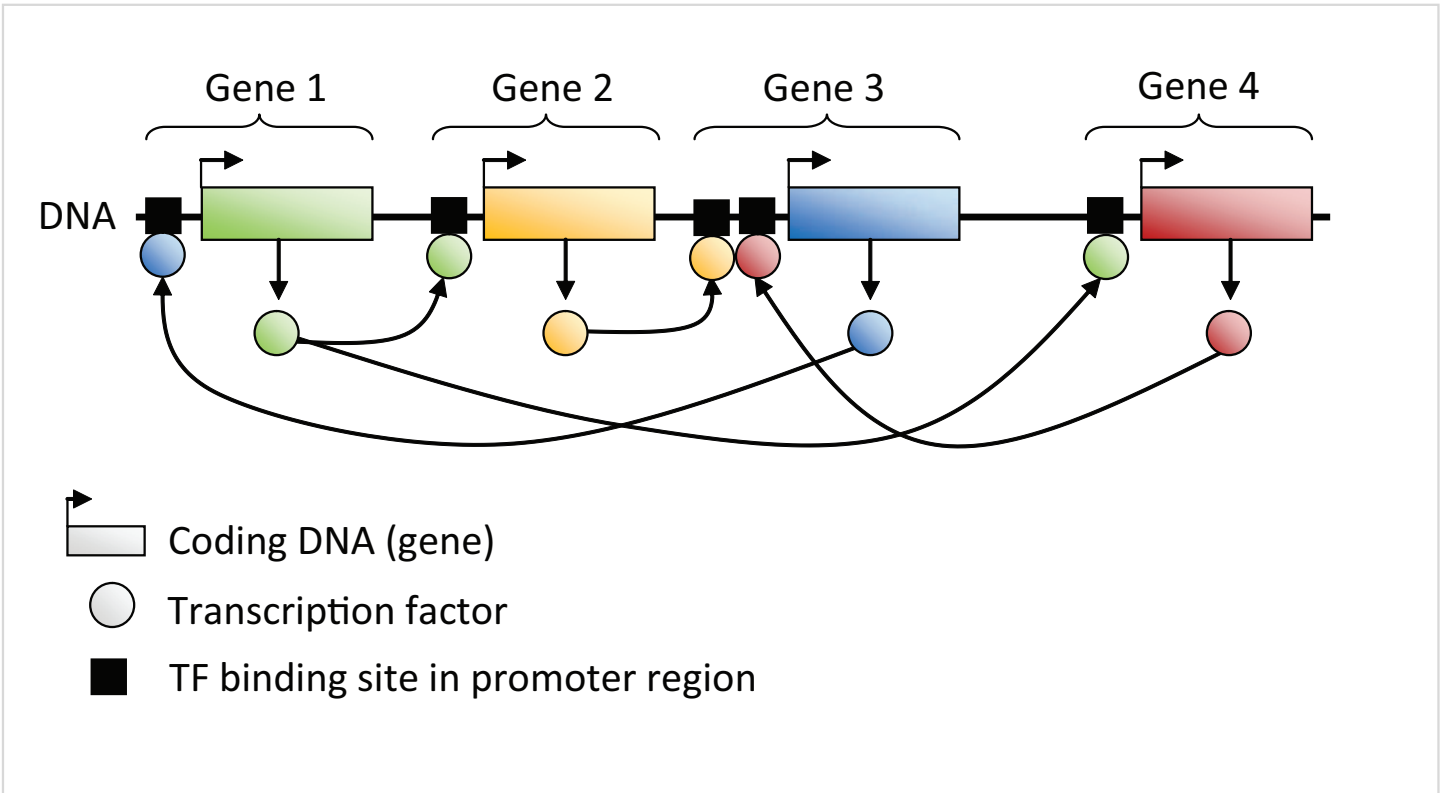
- *Gene regulatory networks*: set of genes that interact indirectly with one another through proteins called transcription factors (TF)



- Abundance of TF is difficult to measure $\Rightarrow$ expression levels of corresponding genes usually used as proxy

Gene Regulatory Networks

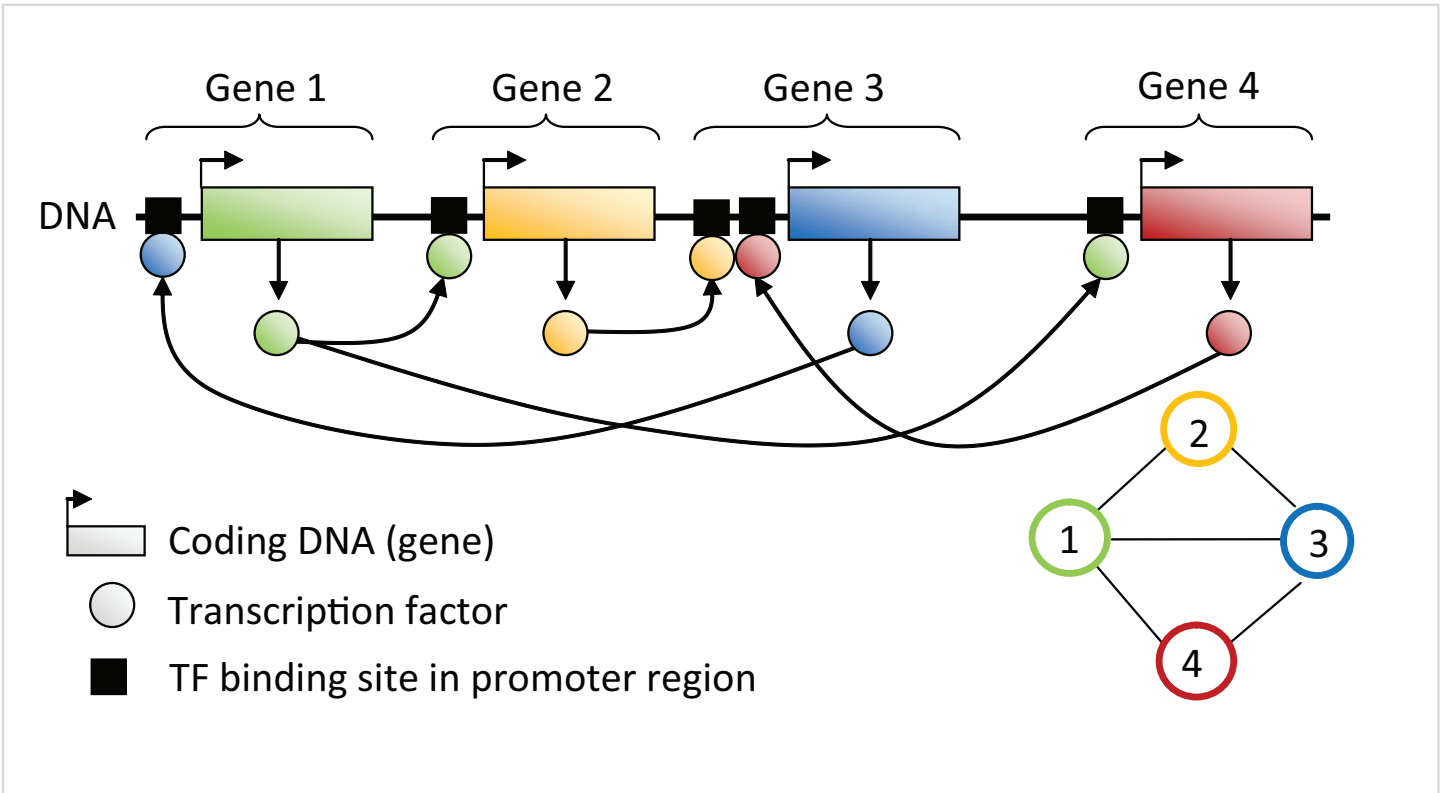
- *Gene regulatory networks*: set of genes that interact indirectly with one another through proteins called transcription factors (TF)



- Abundance of TF is difficult to measure $\Rightarrow$ expression levels of corresponding genes usually used as proxy

Gene Regulatory Networks

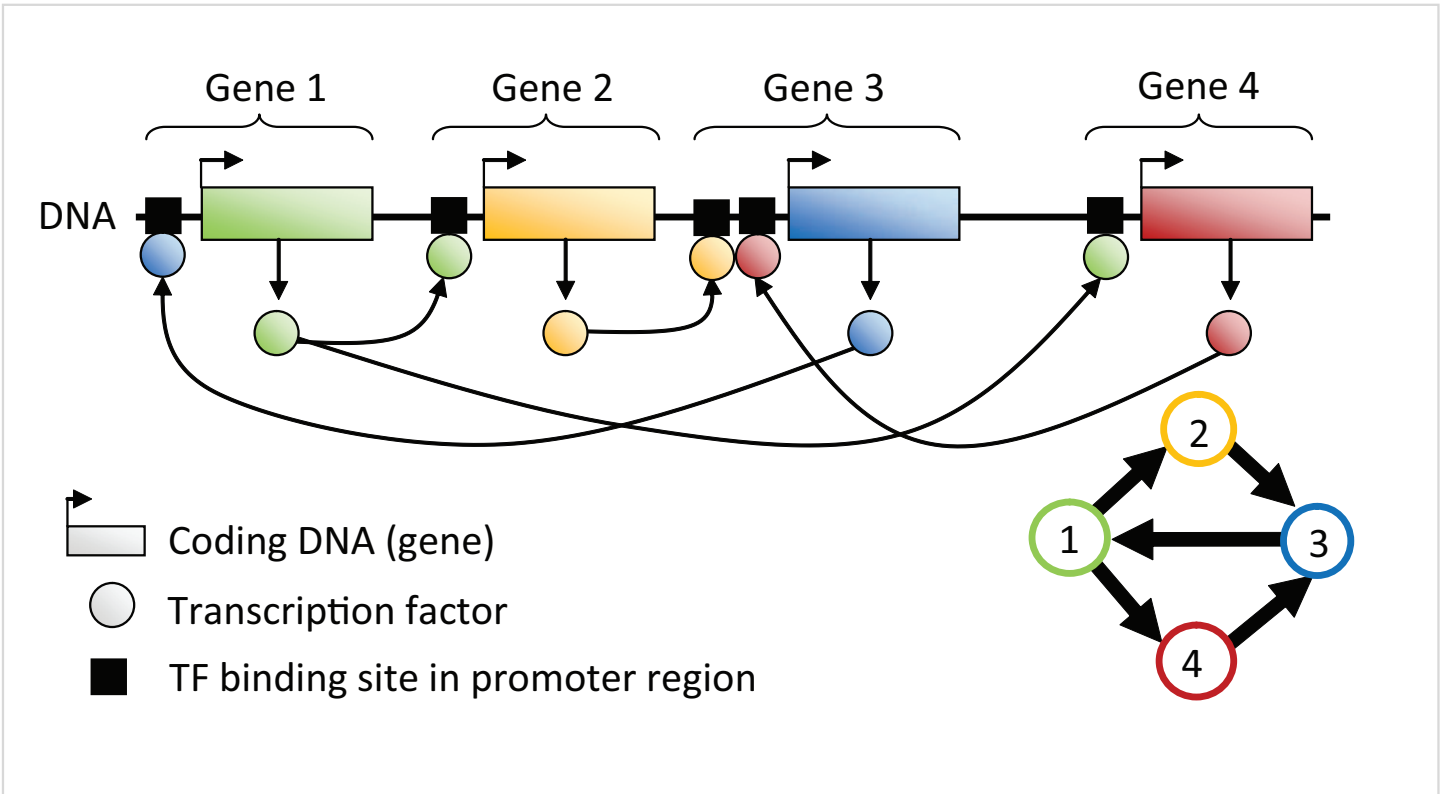
- *Gene regulatory networks*: set of genes that interact indirectly with one another through proteins called transcription factors (TF)



- Abundance of TF is difficult to measure $\Rightarrow$ expression levels of corresponding genes usually used as proxy

Gene Regulatory Networks

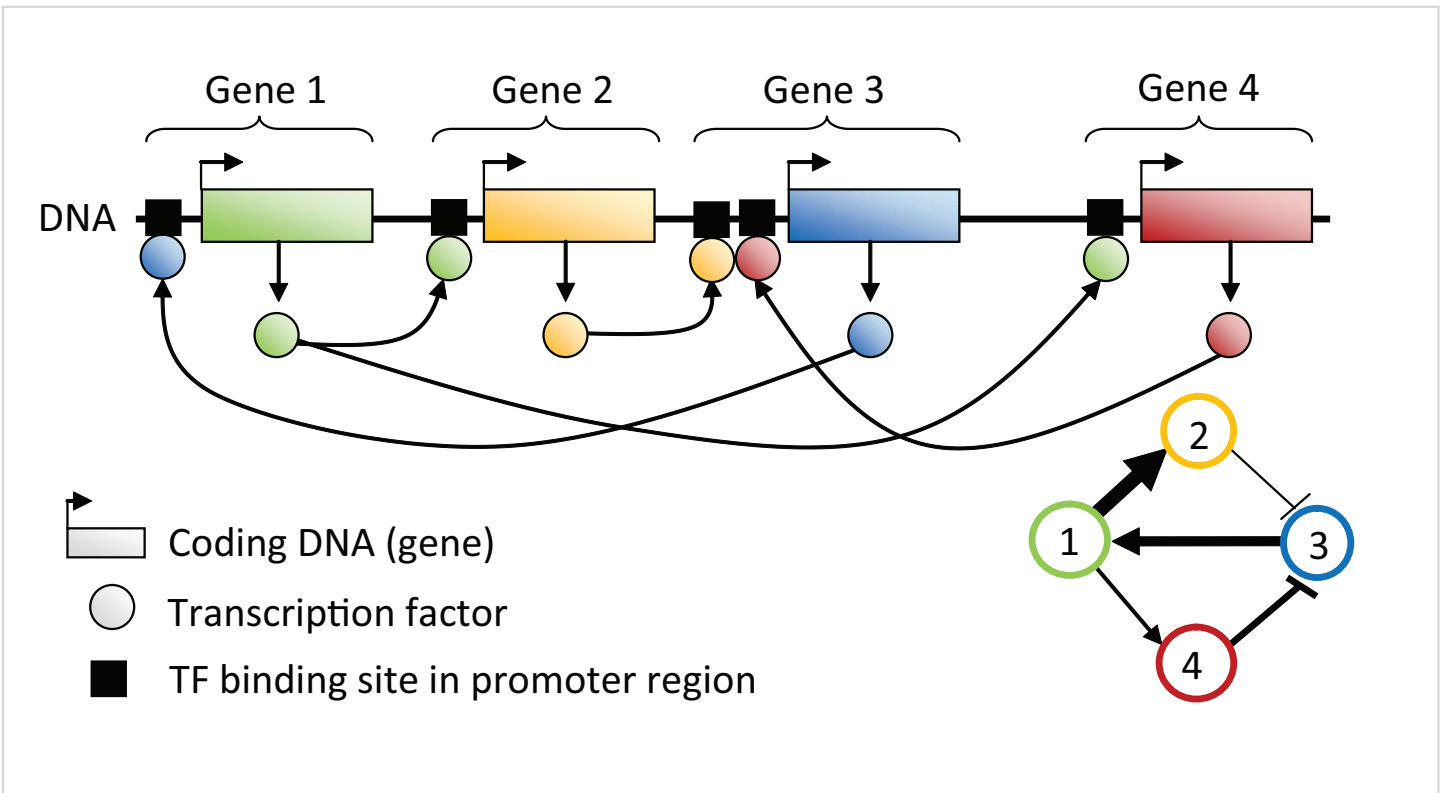
- *Gene regulatory networks*: set of genes that interact indirectly with one another through proteins called transcription factors (TF)



- Abundance of TF is difficult to measure $\Rightarrow$ expression levels of corresponding genes usually used as proxy

Gene Regulatory Networks

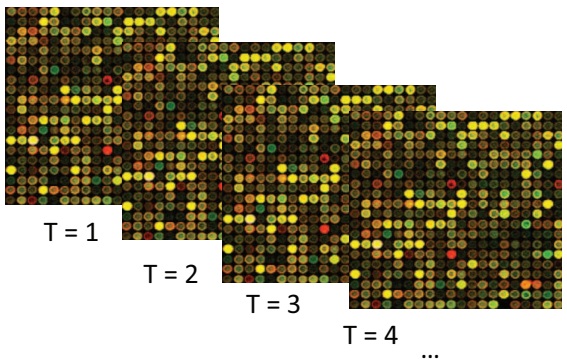
- *Gene regulatory networks*: set of genes that interact indirectly with one another through proteins called transcription factors (TF)



- Abundance of TF is difficult to measure $\Rightarrow$ expression levels of corresponding genes usually used as proxy

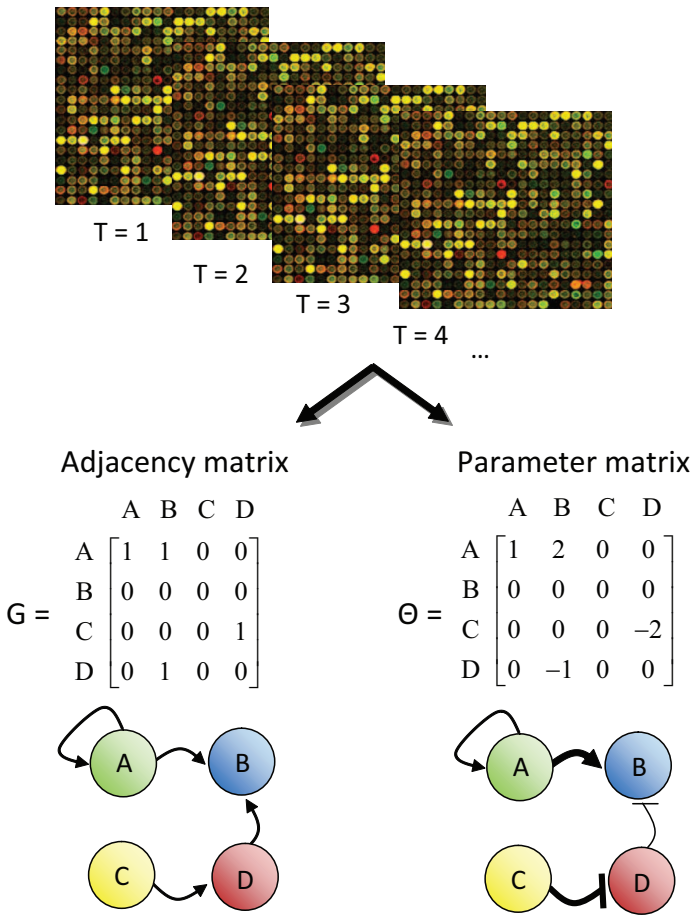
Reverse Engineering Gene Regulatory Networks

- Expression levels of thousands of genes can be measured using "high-throughput" technologies



Reverse Engineering Gene Regulatory Networks

- Expression levels of thousands of genes can be measured using "high-throughput" technologies
- Time-course gene expression data can elucidate information about *patterns* of relationships of gene expression



Bayesian Framework

- Let observed time-course gene expression data be $\mathbf{y} = \{\mathbf{y}_t : t = 1, \ldots, T\}$, where $\mathbf{y}_t = (y_{t1}, ..., y_{tP})^T$.

Bayesian Framework

- Let observed time-course gene expression data be $\mathbf{y} = \{\mathbf{y}_t : t = 1, \ldots, T\}$, where $\mathbf{y}_t = (y_{t1}, ..., y_{tP})^T$.
- Reverse engineering gene networks is a high dimensional problem: many possible gene-to-gene interactions, few time points and replicates ($P \ll N$)

Bayesian Framework

- Let observed time-course gene expression data be $\mathbf{y} = \{\mathbf{y}_t : t = 1, \ldots, T\}$, where $\mathbf{y}_t = (y_{t1}, ..., y_{tP})^T$.
- Reverse engineering gene networks is a high dimensional problem: many possible gene-to-gene interactions, few time points and replicates ($P \ll N$)
- Many network structures may yield similarly high likelihoods, so posterior distributions may be more informative about particular gene-to-gene interactions

Bayesian Framework

- Let observed time-course gene expression data be $\mathbf{y} = \{\mathbf{y}_t : t = 1, \ldots, T\}$, where $\mathbf{y}_t = (y_{t1}, ..., y_{tP})^T$.
- Reverse engineering gene networks is a high dimensional problem: many possible gene-to-gene interactions, few time points and replicates ($P \ll N$)
- Many network structures may yield similarly high likelihoods, so posterior distributions may be more informative about particular gene-to-gene interactions
- *A priori* biological information may be encoded into the prior distributions (network topology, sparsity, information about pathways from bioinformatics databases, ...)

Bayesian Framework

- Let observed time-course gene expression data be $\mathbf{y} = \{\mathbf{y}_t : t = 1, \ldots, T\}$, where $\mathbf{y}_t = (y_{t1}, ..., y_{tP})^T$.
- Reverse engineering gene networks is a high dimensional problem: many possible gene-to-gene interactions, few time points and replicates ($P \ll N$)
- Many network structures may yield similarly high likelihoods, so posterior distributions may be more informative about particular gene-to-gene interactions
- *A priori* biological information may be encoded into the prior distributions (network topology, sparsity, information about pathways from bioinformatics databases, ...)
- $\Rightarrow$ Fit model $f(\mathbf{y}|\theta)$ to observed data $\mathbf{y}$, where parameters are also random variables following $\pi(\theta)$.
  - Conditional distribution of network edges given observed data is $\pi(\theta|\mathbf{y}) \propto f(\mathbf{y}|\theta)\pi(\theta)$

Approximate Bayesan Computation (ABC)

- Without restrictive distributional assumptions on model parameters, likelihood may be difficult to calculate

Approximate Bayesan Computation (ABC)

- Without restrictive distributional assumptions on model parameters, likelihood may be difficult to calculate
- *Approximate Bayesian Computation*: Sampling-based Bayesian approach to infer approximate posterior distribution $\pi(\Theta|\rho(\mathbf{y}^\star, \mathbf{y}) \leq \epsilon)$ using simulated data $\mathbf{y}^\star$, a distance function ρ, and tolerance ϵ

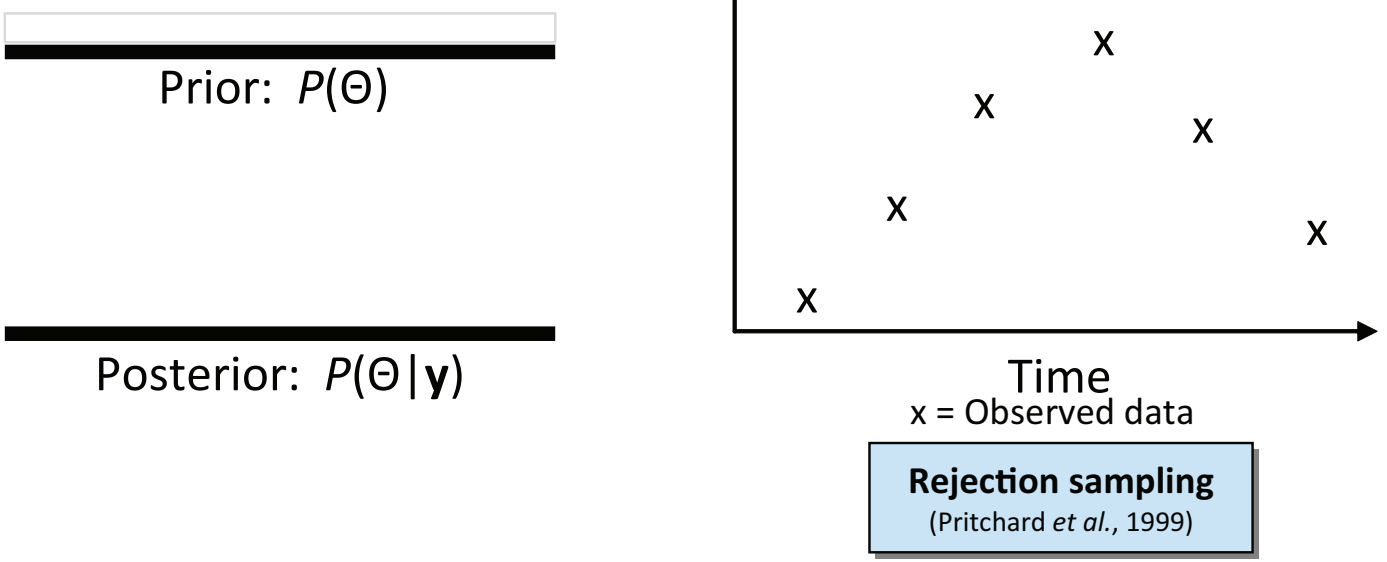
Approximate Bayesan Computation (ABC)

- Without restrictive distributional assumptions on model parameters, likelihood may be difficult to calculate
- *Approximate Bayesian Computation*: Sampling-based Bayesian approach to infer approximate posterior distribution $\pi(\Theta|\rho(\mathbf{y}^\star, \mathbf{y}) \leq \epsilon)$ using simulated data $\mathbf{y}^\star$, a distance function ρ, and tolerance ϵ
  - First applied in population genetics problems (e.g., Pritchard *et al.*, 1999; Beaumont *et al.*, 2002)
  - Some approaches for biological networks (Ratmann *et al.*, 2007; Toni *et al.*, 2010)
  - Novel (and non-standard?) adaptation to reverse engineering gene regulatory networks

Approximate Bayesan Computation (ABC)

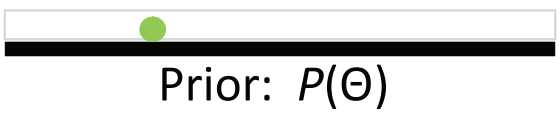
- Without restrictive distributional assumptions on model parameters, likelihood may be difficult to calculate
- *Approximate Bayesian Computation*: Sampling-based Bayesian approach to infer approximate posterior distribution $\pi(\Theta|\rho(\mathbf{y}^\star, \mathbf{y}) \leq \epsilon)$ using simulated data $\mathbf{y}^\star$, a distance function ρ, and tolerance ϵ
  - First applied in population genetics problems (e.g., Pritchard *et al.*, 1999; Beaumont *et al.*, 2002)
  - Some approaches for biological networks (Ratmann *et al.*, 2007; Toni *et al.*, 2010)
  - Novel (and non-standard?) adaptation to reverse engineering gene regulatory networks
- Approximate when $\epsilon > 0$ and equivalent to simulating from the prior when $\epsilon \to \infty$

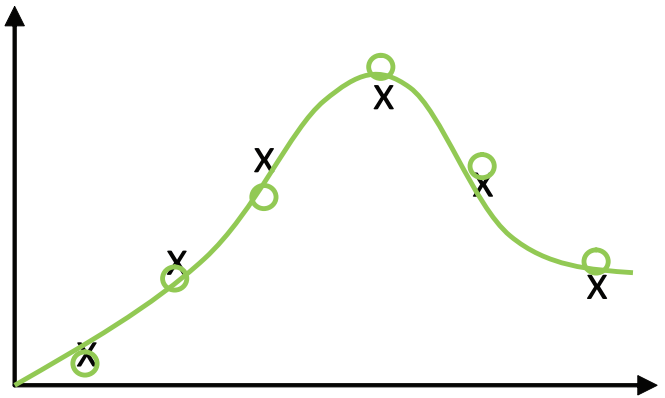
ABC Motivation



(Slide idea from Tina Toni)

ABC Motivation

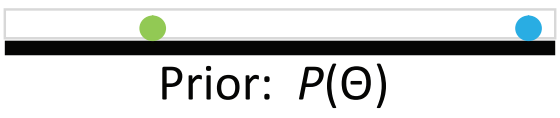


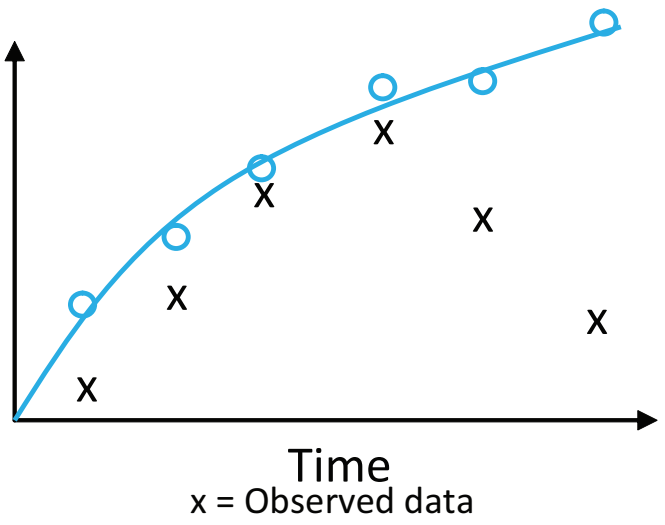


Rejection sampling
(Pritchard *et al.*, 1999)

(Slide idea from Tina Toni)

ABC Motivation

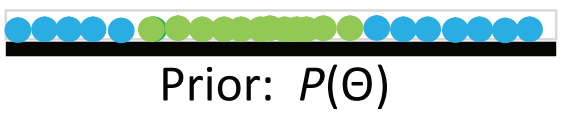




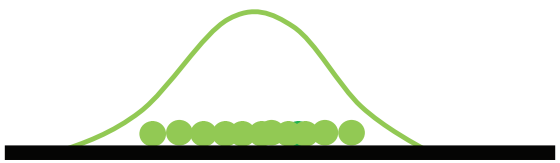
Rejection sampling
(Pritchard *et al.*, 1999)

(Slide idea from Tina Toni)

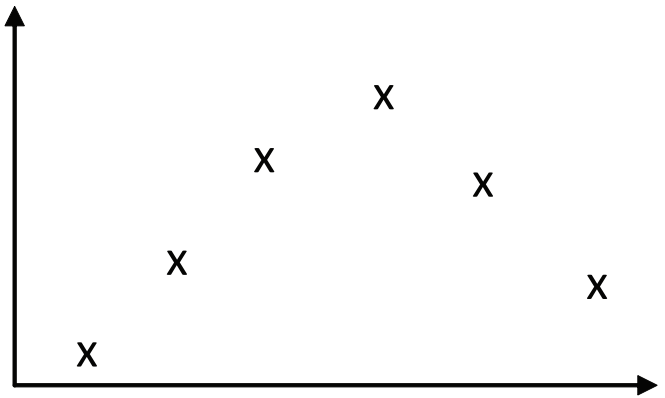
ABC Motivation



Prior: $P(\Theta)$



Posterior: $P(\Theta|\mathbf{y})$

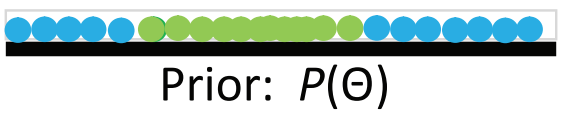


x = Observed data

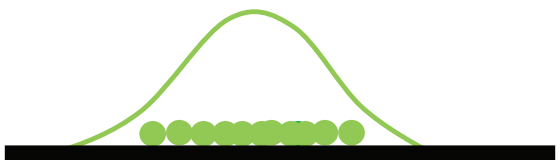
Rejection sampling
(Pritchard *et al.*, 1999)

(Slide idea from Tina Toni)

ABC Motivation

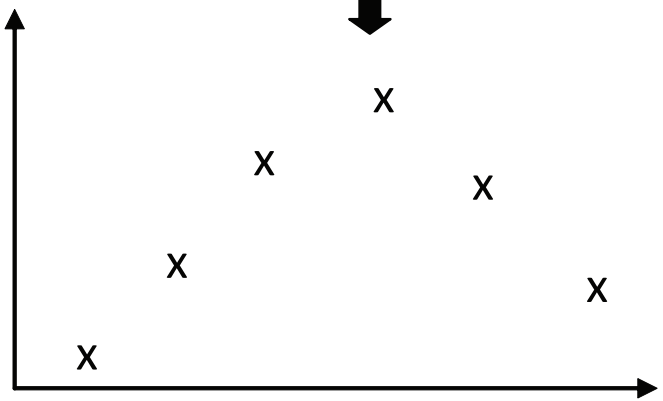


Prior: $P(\Theta)$



Posterior: $P(\Theta|\mathbf{y})$

- Tolerance/distance criteria
- Summary statistics
  (Beaumont *et al.*, 2002)

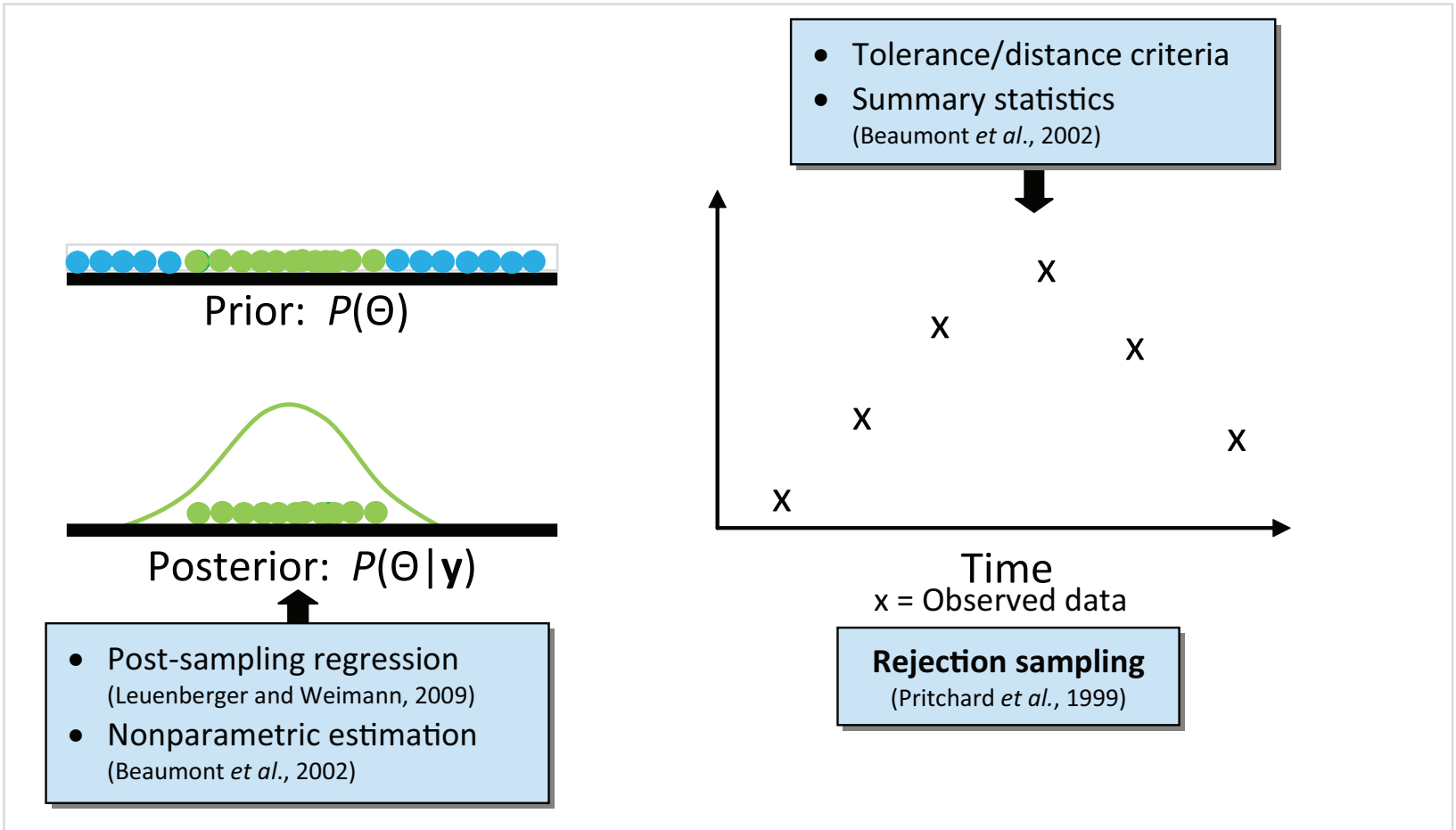


x = Observed data

Rejection sampling
(Pritchard *et al.*, 1999)

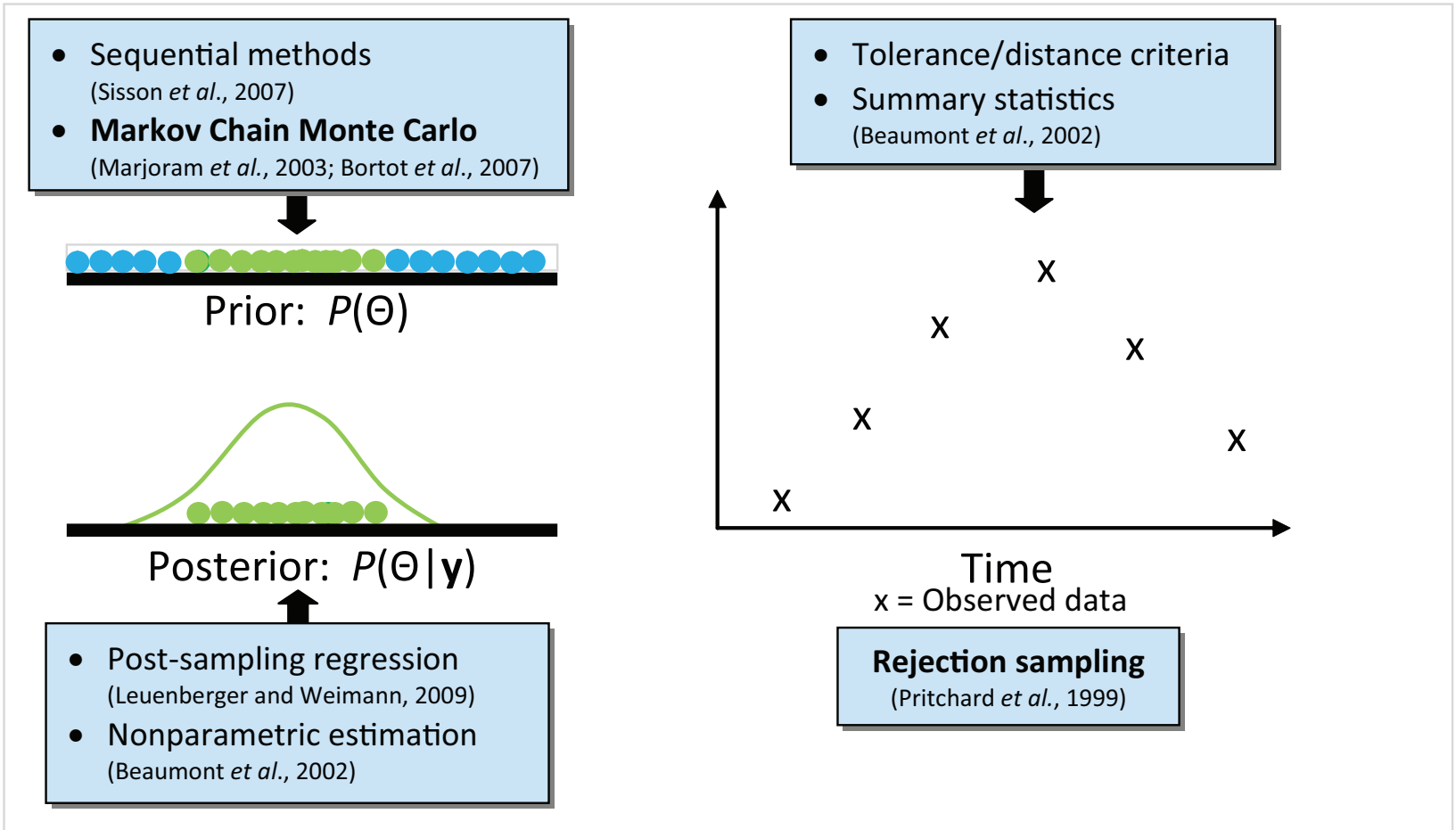
(Slide idea from Tina Toni)

ABC Motivation



(Slide idea from Tina Toni)

ABC Motivation



(Slide idea from Tina Toni)

ABC-MCMC (Marjoram et al., 2003)

- ABC-Markov chain Monte Carlo (MCMC): Construct a Markov chain (e.g., using Metropolis-Hastings algorithm) with approximate posterior distribution $\pi(\Theta|\rho(\mathbf{y}^\star, \mathbf{y}) \leq \epsilon)$ as equilibrium distribution (Marjoram et al., 2003) ▸ Details

ABC-MCMC (Marjoram et al., 2003)

- ABC-Markov chain Monte Carlo (MCMC): Construct a Markov chain (e.g., using Metropolis-Hastings algorithm) with approximate posterior distribution $\pi(\Theta|\rho(\mathbf{y}^\star, \mathbf{y}) \leq \epsilon)$ as equilibrium distribution (Marjoram et al., 2003) ▸ Details
- Let $q(\cdot|\cdot)$ and $\pi(\cdot)$ be the transition and prior distributions, respectively.
- Given previous Θ^i, a proposed $\Theta^\star$ is accepted at the iterations with $(i+1)^{\text{st}}$ probability

$$\alpha = \min\left\{1, \frac{\pi(\Theta^\star)q(\Theta^i|\Theta^\star)}{\pi(\Theta^i)q(\Theta^\star|\Theta^i)}\mathbf{1}(\rho(\mathbf{y}^\star, \mathbf{y}) < \epsilon)\right\}$$

Adapting ABC-MCMC to Networks

- Several adaptations must be made to the ABC-MCMC method of Marjoram et al. (2003) for reverse engineering gene regulatory networks:

Adapting ABC-MCMC to Networks

- Several adaptations must be made to the ABC-MCMC method of Marjoram et al. (2003) for reverse engineering gene regulatory networks:
  1. Computationally efficient way to simulate expression data $\mathbf{y}^\star$ from a known regulatory network $\Theta^\star$

Adapting ABC-MCMC to Networks

- Several adaptations must be made to the ABC-MCMC method of Marjoram et al. (2003) for reverse engineering gene regulatory networks:
  1. Computationally efficient way to simulate expression data $\mathbf{y}^\star$ from a known regulatory network $\Theta^\star$
  2. Appropriate distance function ρ and tolerance ϵ to compare simulated ($\mathbf{y}^\star$) and observed ($\mathbf{y}$) data

Adapting ABC-MCMC to Networks

- Several adaptations must be made to the ABC-MCMC method of Marjoram et al. (2003) for reverse engineering gene regulatory networks:
  1. Computationally efficient way to simulate expression data $\mathbf{y}^\star$ from a known regulatory network $\Theta^\star$
  2. Appropriate distance function ρ and tolerance ϵ to compare simulated ($\mathbf{y}^\star$) and observed ($\mathbf{y}$) data
  3. Prior and proposal distributions for network structures

Simulating $\mathbf{y}^\star$ for Network $\Theta^\star$ (continuous)

Generally, we simulate gene expression at time t as a function of the gene expression at the previous time point:

$$\mathbf{y}_t^\star = f_t(\mathbf{y}_{t-1}, \Theta^\star)$$

Simulating $\mathbf{y}^\star$ for Network $\Theta^\star$ (continuous)

Generally, we simulate gene expression at time t as a function of the gene expression at the previous time point:

$$\mathbf{y}_t^\star = f_t(\mathbf{y}_{t-1}, \Theta^\star)$$

In practice, for continuous data (e.g., microarrays):

- Set $\mathbf{y}_1^\star = \mathbf{y}_1$.
- Generate one-step-ahead predictors based on first-order VAR model on gene expression for $t = 2, \ldots, T$:

$$\mathbf{y}_t^\star = \Theta^\star \mathbf{y}_{t-1}$$

▸ Details

Simulating $\mathbf{y}^\star$ for Network $\Theta^\star$ (discrete)

For count data (e.g., serial analysis of gene expression, RNA sequencing):

- Set $\mathbf{y}_1^\star = \mathbf{y}_1$.
- $\mathbf{y}_t \sim \text{Poisson}(\boldsymbol{\lambda}_t)$, where $\boldsymbol{\lambda}_t = \boldsymbol{\pi}_t y_{\cdot t}$, $\sum_{i=1}^{P} \pi_{it} = 1$, and $y_{\cdot t} = \sum_{i=1}^{P} y_{it}$.

Simulating $\mathbf{y}^\star$ for Network $\Theta^\star$ (discrete)

For count data (e.g., serial analysis of gene expression, RNA sequencing):

- Set $\mathbf{y}_1^\star = \mathbf{y}_1$.
- $\mathbf{y}_t \sim \text{Poisson}(\boldsymbol{\lambda}_t)$, where $\boldsymbol{\lambda}_t = \boldsymbol{\pi}_t y_{\cdot t}$, $\sum_{i=1}^P \pi_{it} = 1$, and $y_{\cdot t} = \sum_{i=1}^P y_{it}$.
- Generate one-step-ahead predictors based on first-order VAR model on the *level* of gene expression for $t = 2, \ldots, T$:

$$\tilde{\boldsymbol{\pi}}_t^\star = \exp\left\{\frac{1}{y_{\cdot t-1}}\Theta^\star \mathbf{y}_{t-1}\right\} \text{ and } \boldsymbol{\pi}_t^\star = \frac{1}{\sum_{i=1}^P \tilde{\pi}_{it}^\star}\tilde{\boldsymbol{\pi}}_t^\star$$

$$\mathbf{y}_t^\star \sim \text{Poisson}(\boldsymbol{\pi}_t^\star y_{\cdot t})$$

▸ Details

Distance Function and Tolerance

Distance functions (ρ):

- Canberra: $\rho(\mathbf{y}^\star, \mathbf{y}) = \sum_{t=1}^{T} \sum_{i=1}^{P} \frac{|y_{it}^\star - y_{it}|}{|y_{it}^\star + y_{it}|}$
- Euclidean: $\rho(\mathbf{y}^\star, \mathbf{y}) = \sqrt{\sum_{t=1}^{T} \sum_{i=1}^{P} (y_{it}^\star - y_{it})^2}$
- Manhattan: $\rho(\mathbf{y}^\star, \mathbf{y}) = \sum_{t=1}^{T} \sum_{i=1}^{P} |y_{it}^\star - y_{it}|$
- Multivariate Time Series (MVT):

$$\rho(\mathbf{y}^\star, \mathbf{y}) = \frac{1}{T} \sum_{t=1}^{T} [(\mathbf{y}_t - \mathbf{y}_t^\star) - (\hat{\mathbf{y}}_t - \hat{\mathbf{y}}_t^\star)]' \, \hat{\Sigma}^{-1} \, [(\mathbf{y}_t - \mathbf{y}_t^\star) - (\hat{\mathbf{y}}_t - \hat{\mathbf{y}}_t^\star)]$$

Distance Function and Tolerance

Distance functions (ρ):

- Canberra: $\rho(\mathbf{y}^\star, \mathbf{y}) = \sum_{t=1}^{T} \sum_{i=1}^{P} \frac{|y_{it}^\star - y_{it}|}{|y_{it}^\star + y_{it}|}$
- Euclidean: $\rho(\mathbf{y}^\star, \mathbf{y}) = \sqrt{\sum_{t=1}^{T} \sum_{i=1}^{P} (y_{it}^\star - y_{it})^2}$
- Manhattan: $\rho(\mathbf{y}^\star, \mathbf{y}) = \sum_{t=1}^{T} \sum_{i=1}^{P} |y_{it}^\star - y_{it}|$
- Multivariate Time Series (MVT):

$$\rho(\mathbf{y}^\star, \mathbf{y}) = \frac{1}{T} \sum_{t=1}^{T} [(\mathbf{y}_t - \mathbf{y}_t^\star) - (\hat{\mathbf{y}}_t - \hat{\mathbf{y}}_t^\star)]' \, \hat{\Sigma}^{-1} \, [(\mathbf{y}_t - \mathbf{y}_t^\star) - (\hat{\mathbf{y}}_t - \hat{\mathbf{y}}_t^\star)]$$

Tolerance (ϵ):

- $\epsilon = 1\%$ quantile of distances ρ from 5000 random networks

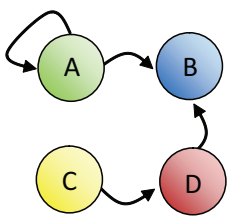
Network Proposals

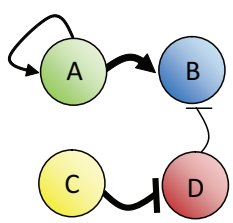
- With networks, we must propose both a new structure and a new set of parameters

Network Proposals

- With networks, we must propose both a new structure and a new set of parameters
- To facilitate simulation, we introduce as an auxiliary variable a $P \times P$ adjacency matrix G, where $G_{ij} = 1$ if gene j regulates gene i, and $G_{ij} = 0$ otherwise.
- Note that $G_{ij} = 0 \Leftrightarrow \Theta_{ij} = 0$ and $G_{ij} = 1 \Leftrightarrow \Theta_{ij} \neq 0$

$$G = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array} \qquad \Theta = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 \\ 0 & -1 & 0 & 0 \end{bmatrix} \end{array}$$

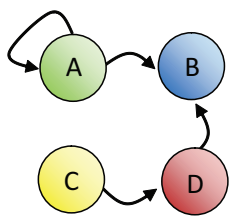


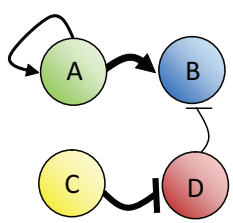


Network Proposals

- With networks, we must propose both a new structure and a new set of parameters
- To facilitate simulation, we introduce as an auxiliary variable a $P \times P$ adjacency matrix G, where $G_{ij} = 1$ if gene j regulates gene i, and $G_{ij} = 0$ otherwise.
- Note that $G_{ij} = 0 \Leftrightarrow \Theta_{ij} = 0$ and $G_{ij} = 1 \Leftrightarrow \Theta_{ij} \neq 0$

$$G = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array} \qquad \Theta = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 \\ 0 & -1 & 0 & 0 \end{bmatrix} \end{array}$$





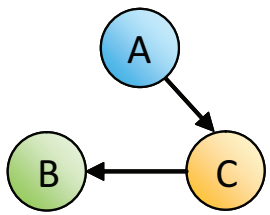
- Joint distribution of G and Θ may be seen as a completion to the marginal density of Θ

Two-Step Proposal Distribution

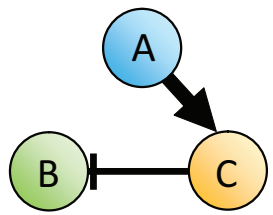
- Two-step proposal distribution: $q(G^\star|G^i)$ and $q(\Theta^\star|\Theta^i, G^\star)$:

Adjacency matrix

$$G^i = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{array}$$



Parameter matrix



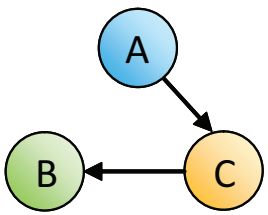
$$\Theta^i = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -2 \\ 3 & 0 & 0 \end{pmatrix} \end{array}$$

Two-Step Proposal Distribution

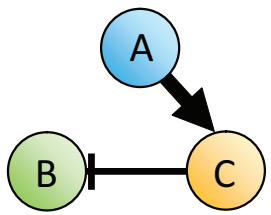
- Two-step proposal distribution: $q(G^\star|G^i)$ and $q(\Theta^\star|\Theta^i, G^\star)$:

Adjacency matrix

$$G^i = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{array}$$



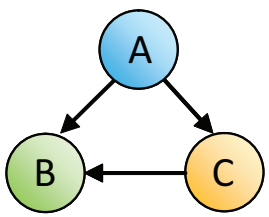
Parameter matrix



$$\Theta^i = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -2 \\ 3 & 0 & 0 \end{pmatrix} \end{array}$$

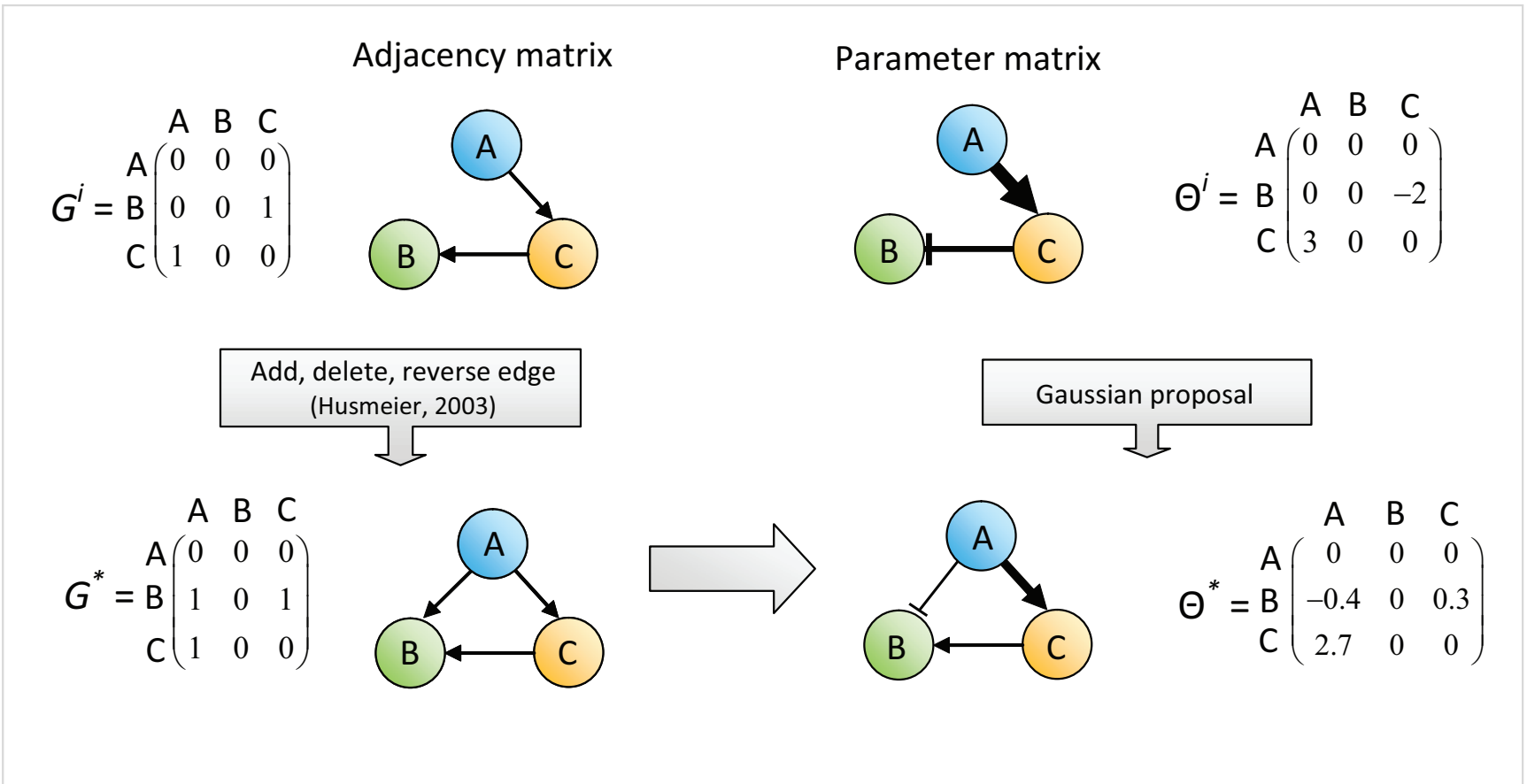
Add, delete, reverse edge
(Husmeier, 2003)

$$G^* = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{array}$$



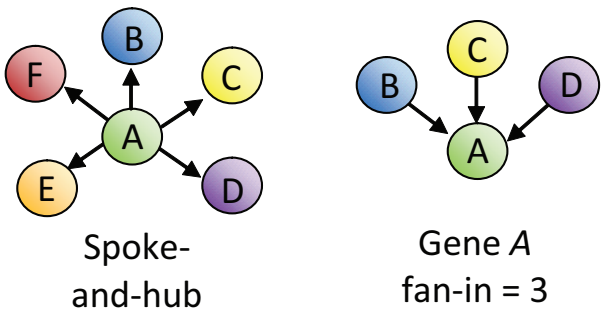
Two-Step Proposal Distribution

- Two-step proposal distribution: $q(G^\star|G^i)$ and $q(\Theta^\star|\Theta^i, G^\star)$:



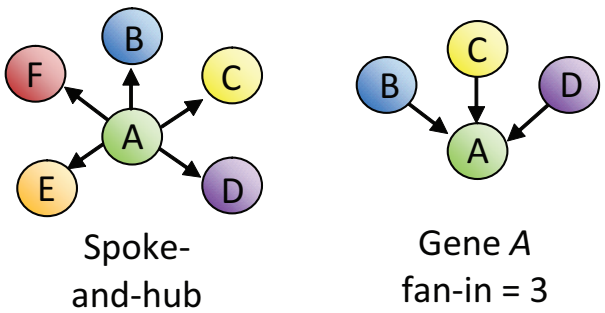
Prior Distributions

- Gene regulatory networks typically sparse with spoke-and-hub structure and few regulators per gene (fan-in)



Prior Distributions

- Gene regulatory networks typically sparse with spoke-and-hub structure and few regulators per gene (fan-in)



Prior distributions:

- $\pi(G)$ is uniform over all structures, with maximum fan-in of 5 or less
- $\pi(\Theta|G)$ is uniform

ABC-MCMC Network Method

ABC-Net Algorithm:

0. Initialize Θ^i, G^i, $i = 0$.

ABC-MCMC Network Method

ABC-Net Algorithm:

0. Initialize Θ^i, G^i, $i = 0$.
1. (a) Propose $G^\star$ according to $q(G|G^i)$.
   (b) Propose $\Theta^\star$ according to $q(\Theta|\Theta^i, G^\star)$.

ABC-MCMC Network Method

ABC-Net Algorithm:

0. Initialize Θ^i, G^i, $i = 0$.
1. (a) Propose $G^\star$ according to $q(G|G^i)$.
   (b) Propose $\Theta^\star$ according to $q(\Theta|\Theta^i, G^\star)$.
2. Simulate $\mathbf{y}^\star$ from $f(\cdot|\Theta^\star, G^\star)$.

ABC-MCMC Network Method

ABC-Net Algorithm:

0. Initialize Θ^i, G^i, $i = 0$.
1. (a) Propose $G^\star$ according to $q(G|G^i)$.
   (b) Propose $\Theta^\star$ according to $q(\Theta|\Theta^i, G^\star)$.
2. Simulate $\mathbf{y}^\star$ from $f(\cdot|\Theta^\star, G^\star)$.
3. Set $\{G^{i+1}, \Theta^{i+1}\} = \{G^\star, \Theta^\star\}$ with probability
   $\alpha = \min\{1, \frac{\pi(G^\star)\pi(\Theta^\star|G^\star)q(G^i|G^\star)q(\Theta^i|\Theta^\star)}{\pi(G^i)\pi(\Theta^i|G^i)q(G^\star|G^i)q(\Theta^\star|\Theta^i)}\mathbf{1}\left[\rho(\mathbf{y}^\star, \mathbf{y}) \leq \epsilon\right]\}$
   and $\{G^{i+1}, \Theta^{i+1}\} = \{G^i, \Theta^i\}$ with probability $1 - \alpha$.

ABC-MCMC Network Method

ABC-Net Algorithm:

0. Initialize Θ^i, G^i, $i = 0$.
1. (a) Propose $G^\star$ according to $q(G|G^i)$.
   (b) Propose $\Theta^\star$ according to $q(\Theta|\Theta^i, G^\star)$.
2. Simulate $\mathbf{y}^\star$ from $f(\cdot|\Theta^\star, G^\star)$.
3. Set $\{G^{i+1}, \Theta^{i+1}\} = \{G^\star, \Theta^\star\}$ with probability
   $\alpha = \min\{1, \frac{\pi(G^\star)\pi(\Theta^\star|G^\star)q(G^i|G^\star)q(\Theta^i|\Theta^\star)}{\pi(G^i)\pi(\Theta^i|G^i)q(G^\star|G^i)q(\Theta^\star|\Theta^i)}\mathbf{1}\left[\rho(\mathbf{y}^\star, \mathbf{y}) \leq \epsilon\right]\}$
   and $\{G^{i+1}, \Theta^{i+1}\} = \{G^i, \Theta^i\}$ with probability $1 - \alpha$.
4. Set $i = i + 1$. If $i < N$ (a pre-set number of iterations), return to 1.

ABC-MCMC Network Method

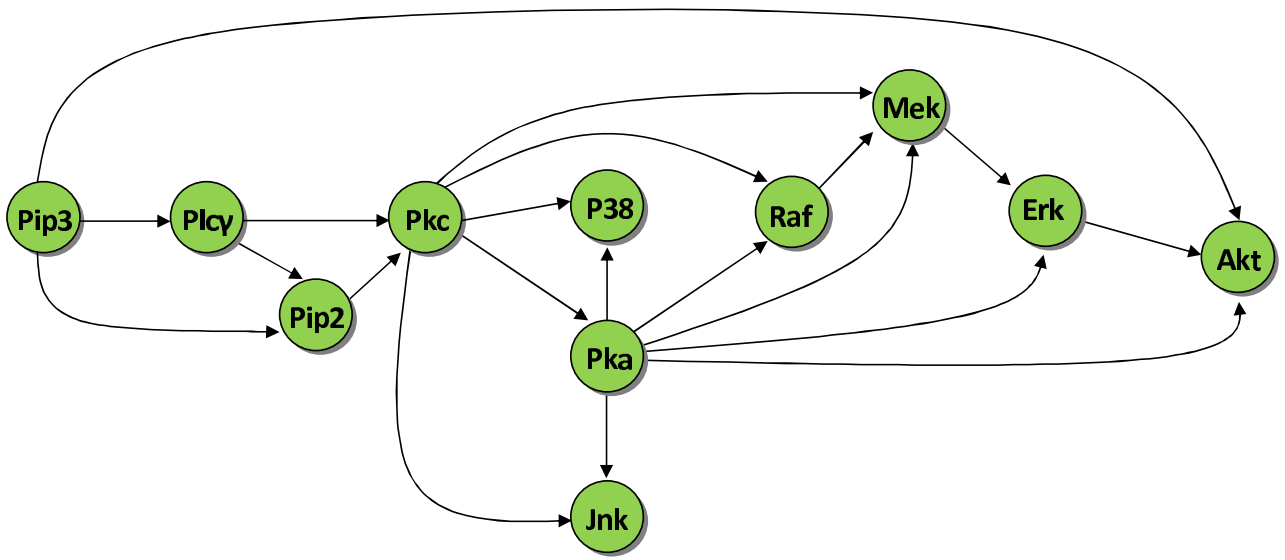
ABC-Net Algorithm:

0. Initialize Θ^i, G^i, $i = 0$.
1. (a) Propose $G^\star$ according to $q(G|G^i)$.
   (b) Propose $\Theta^\star$ according to $q(\Theta|\Theta^i, G^\star)$.
2. Simulate $\mathbf{y}^\star$ from $f(\cdot|\Theta^\star, G^\star)$.
3. Set $\{G^{i+1}, \Theta^{i+1}\} = \{G^\star, \Theta^\star\}$ with probability
   $\alpha = \min\{1, \frac{\pi(G^\star)\pi(\Theta^\star|G^\star)q(G^i|G^\star)q(\Theta^i|\Theta^\star)}{\pi(G^i)\pi(\Theta^i|G^i)q(G^\star|G^i)q(\Theta^\star|\Theta^i)}\mathbf{1}[\rho(\mathbf{y}^\star, \mathbf{y}) \leq \epsilon]\}$
   and $\{G^{i+1}, \Theta^{i+1}\} = \{G^i, \Theta^i\}$ with probability $1 - \alpha$.
4. Set $i = i + 1$. If $i < N$ (a pre-set number of iterations), return to 1.

- Burn-in period, number of iterations, chain thinning, ... ▸ Details

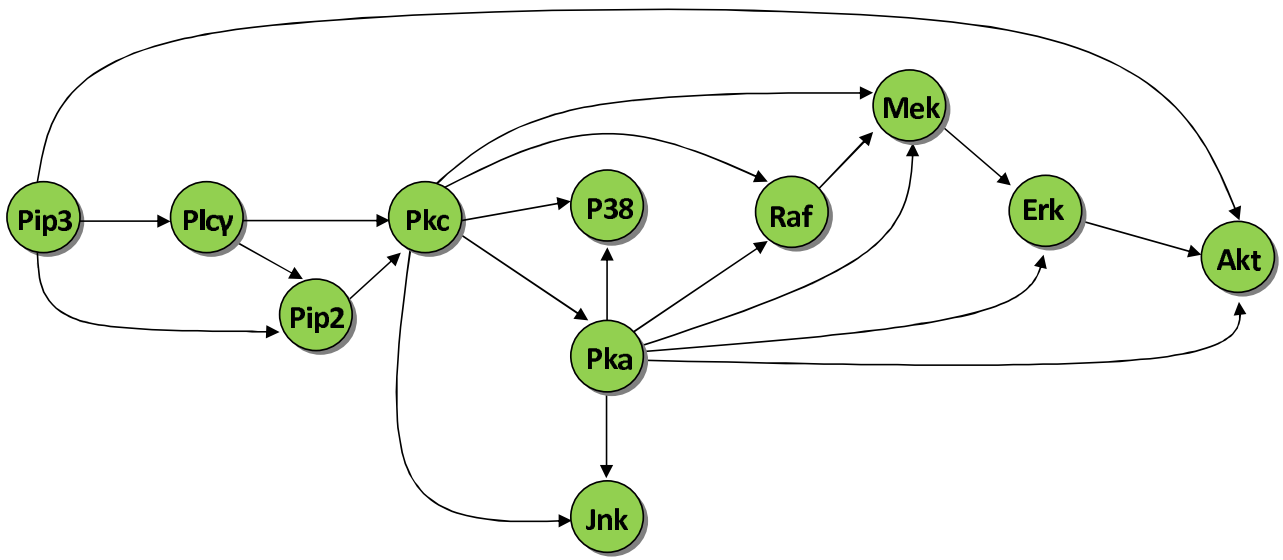
Simulations: Raf Signalling Protein Pathway

- Simulations based on currently accepted gold-standard Raf signalling pathway (Sachs et al., 2005) in human immune system cells for 11 genes (20 total edges)



Simulations: Raf Signalling Protein Pathway

- Simulations based on currently accepted gold-standard Raf signalling pathway (Sachs et al., 2005) in human immune system cells for 11 genes (20 total edges)



- Simulate $T = 20$ time points, $R = 1$ replicate using VAR model
- Run ABC-Net algorithm for 10 independent chains of length 1×10^6 with thinning interval of 50
- Use Gelman-Rubin statistic to assess convergence across chains

ABC-Net Simulations

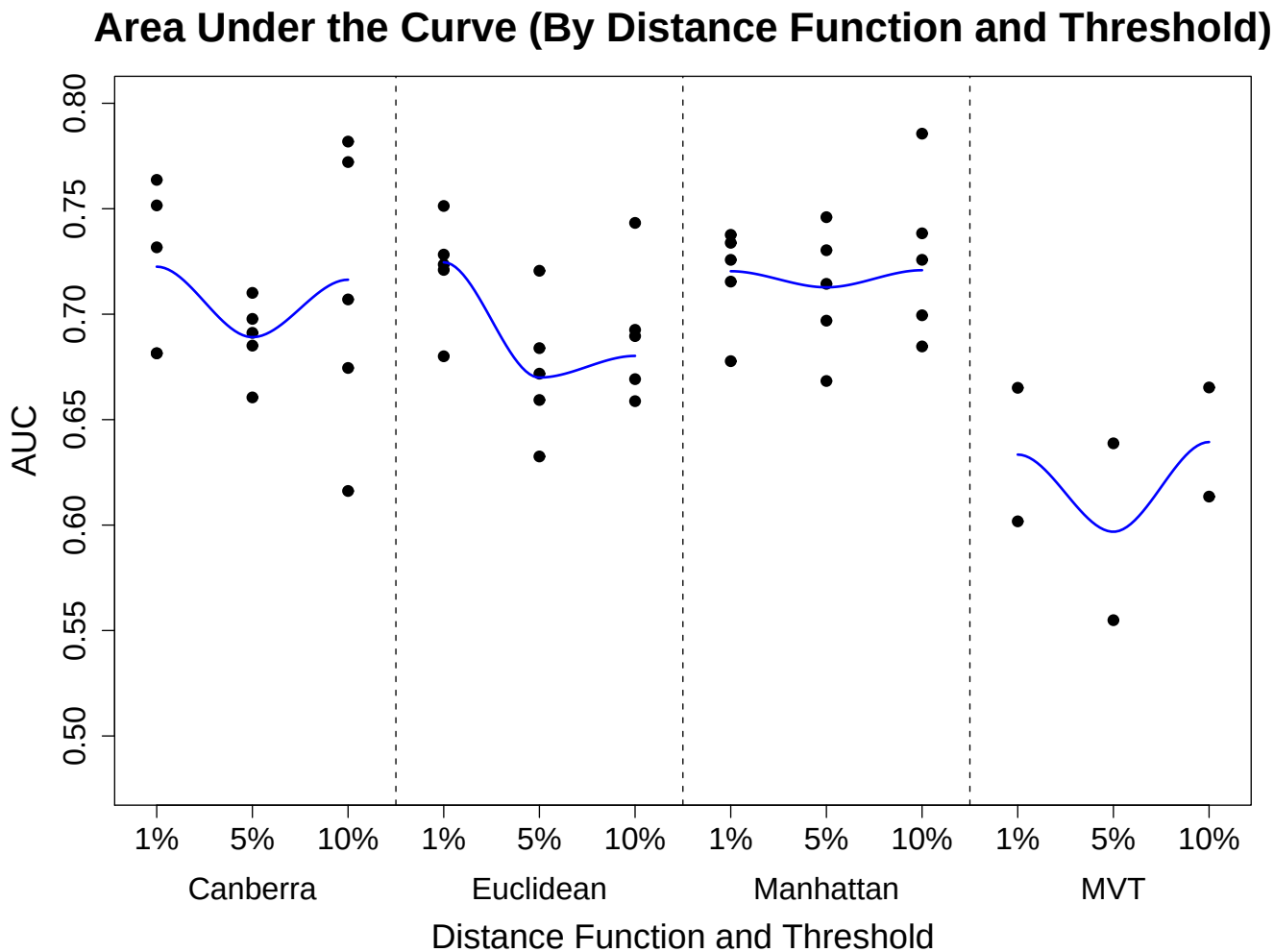
1. Choice of distance function ρ and tolerance ϵ

2. Suitability of VAR simulator for data generated with alternative models (nonlinear models, second-order models, and ordinary differential equations)

3. Sensitivity to prior distribution bounds

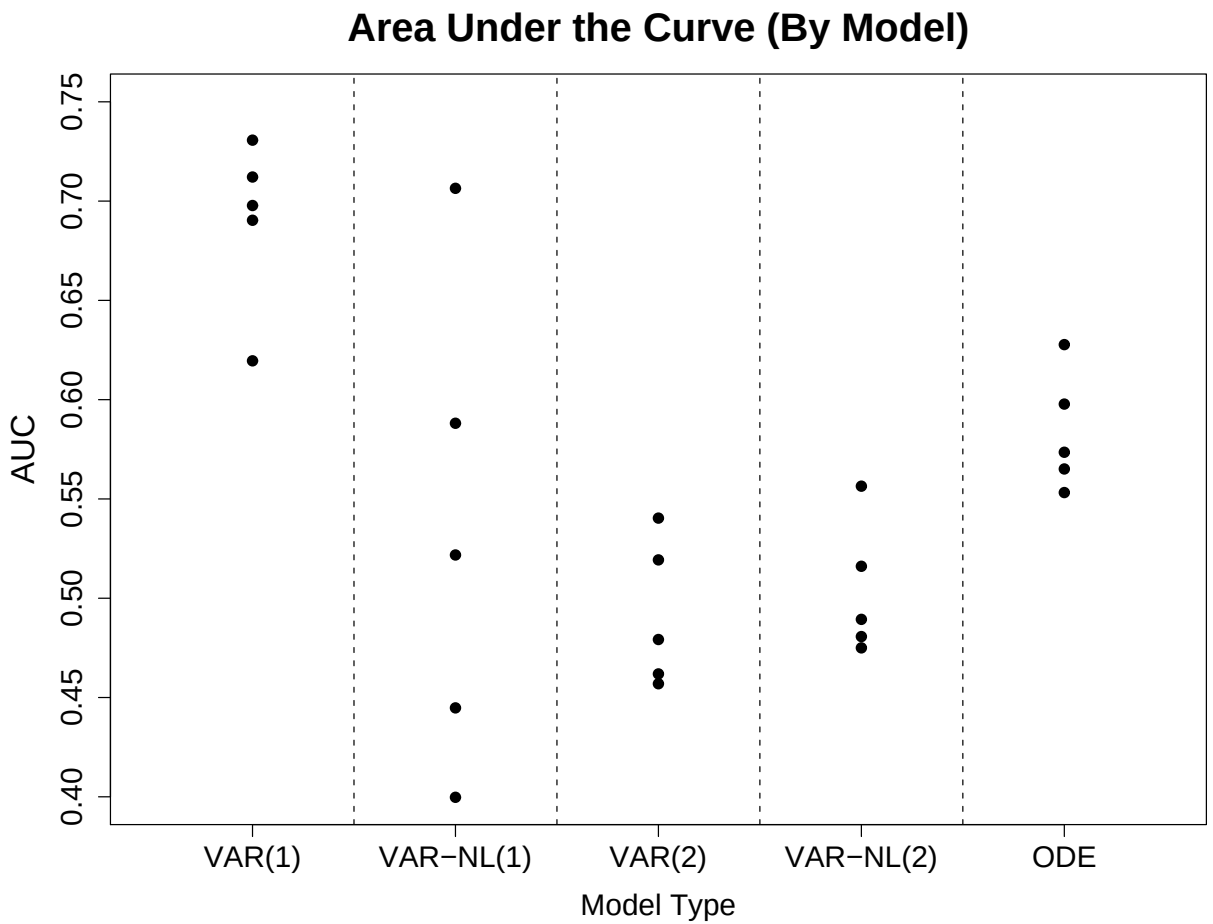
Simulations I: Choice of ρ and ϵ

- Set ϵ to be the 1%, 5%, or 10% quantile of distances ρ from 5000 random networks



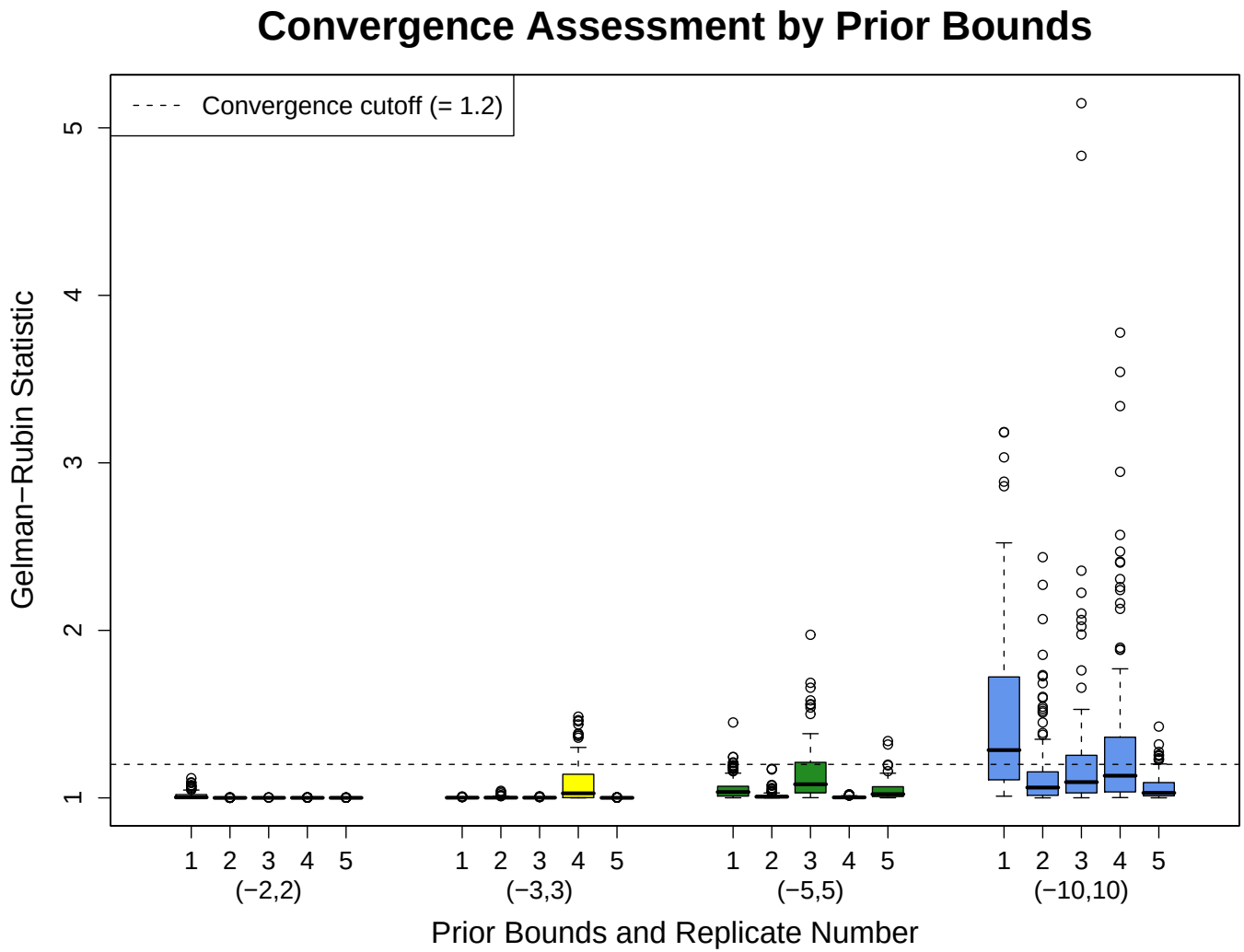
Simulations II: Suitability of VAR Simulator

- Alternative models: first-order nonlinear VAR (VAR-NL(1)), second-order VAR (VAR(2)), second-order nonlinear VAR (VAR-NL(2)), and ordinary differential equation (ODE)

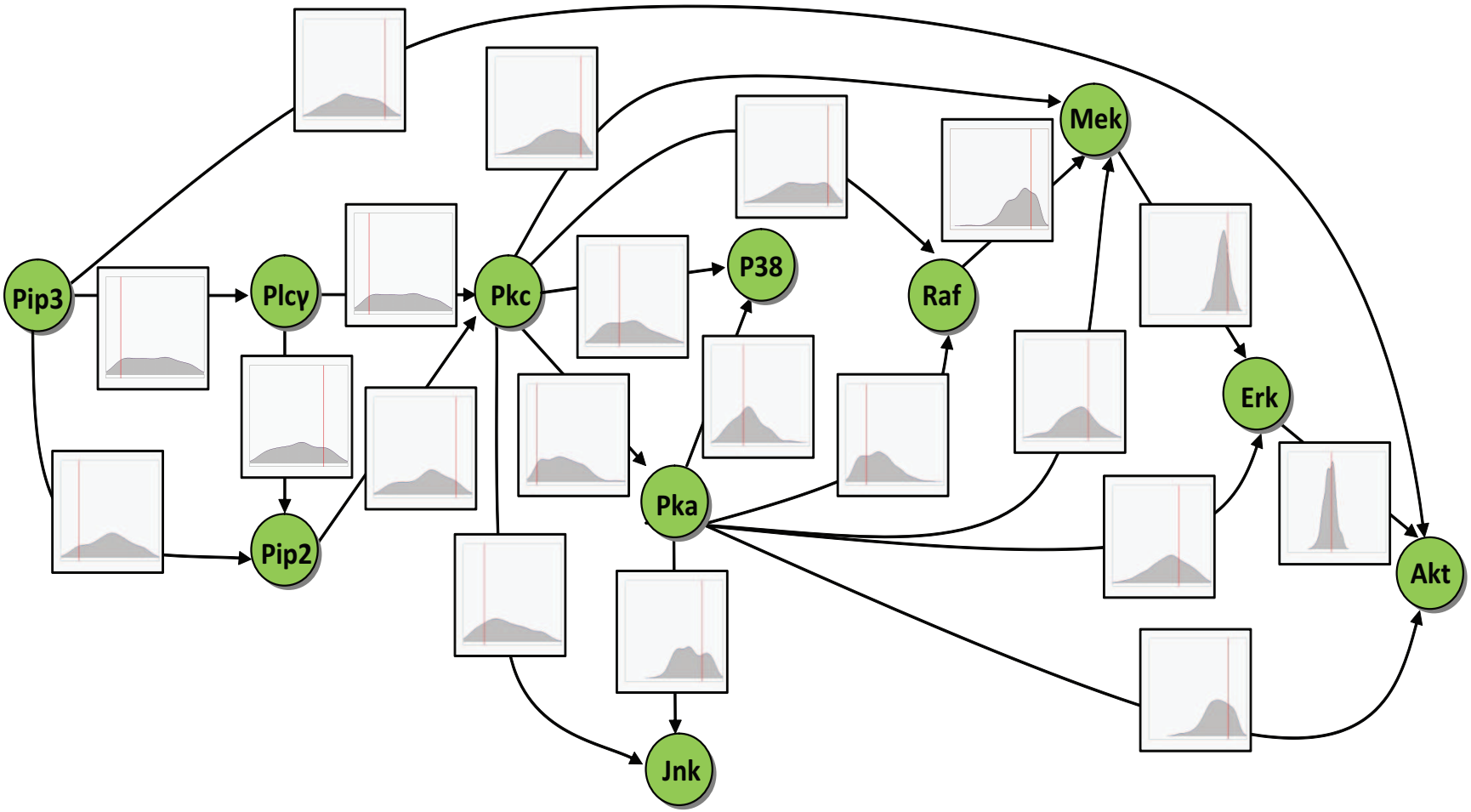


Simulations III: Sensitivity to prior distribution bounds

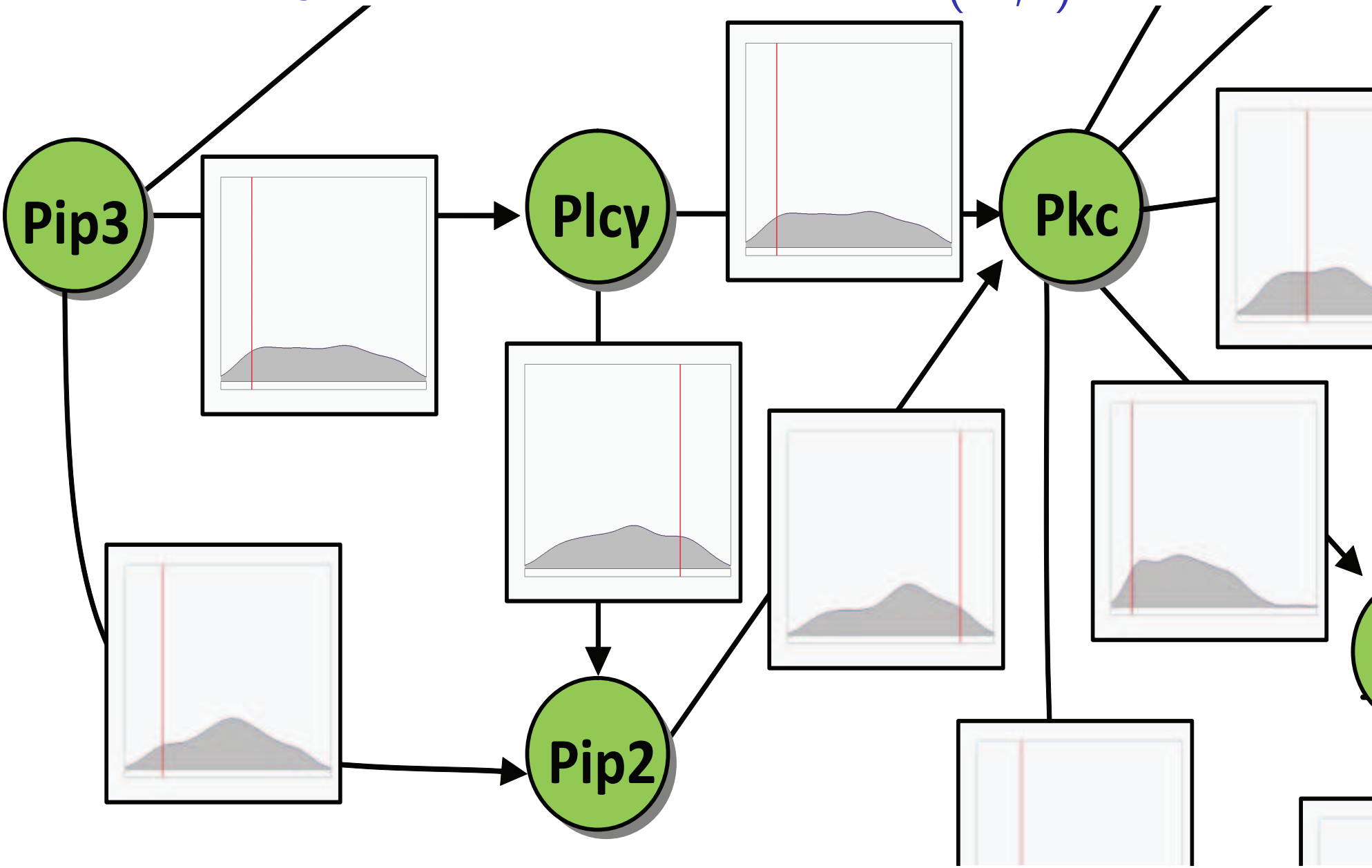
- Vary prior bounds $\pi(\Theta|G)$ between (-2,2), (-3,3), (-5,5) and (-10,10)



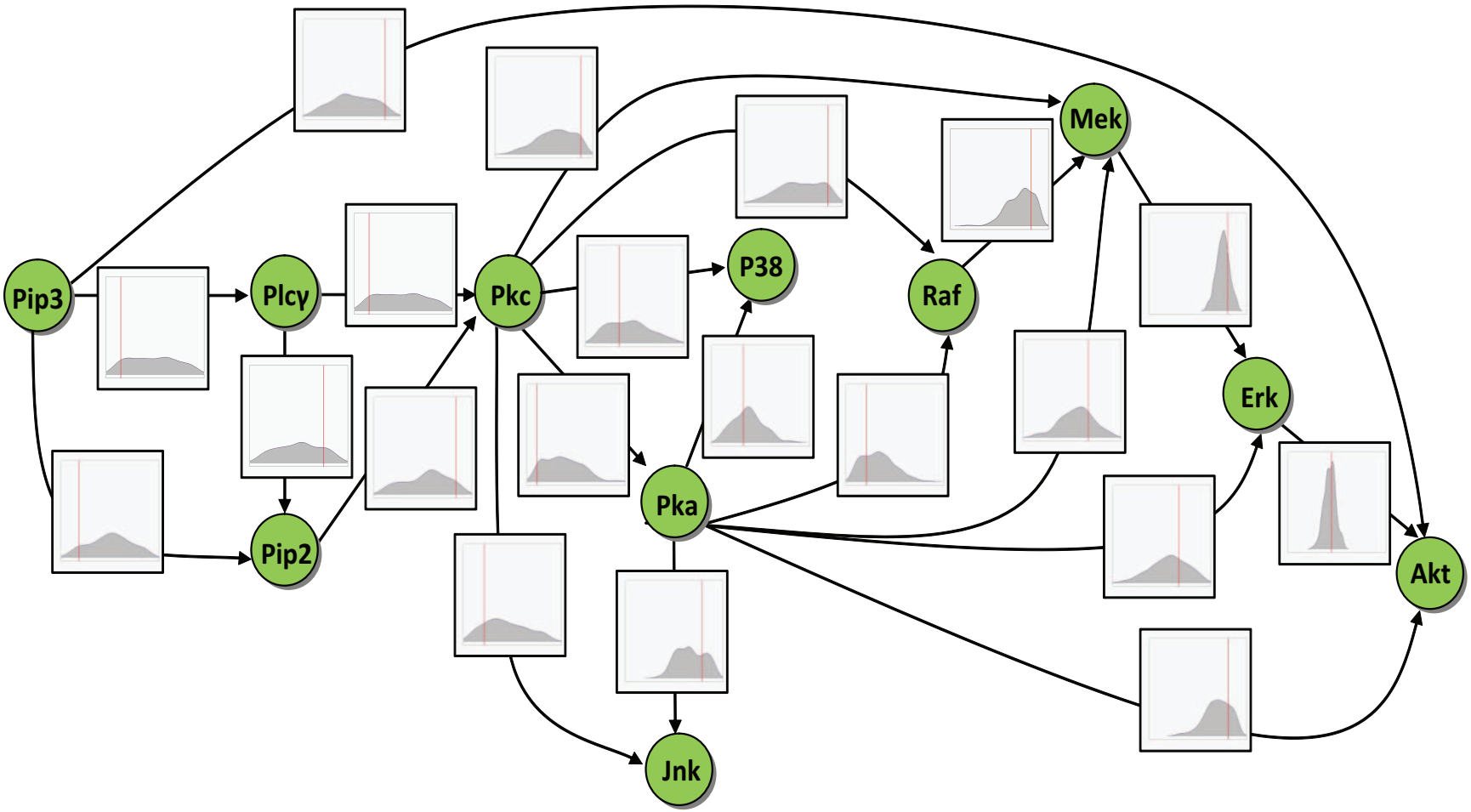
Simulations III: Prior bounds (-2,2)



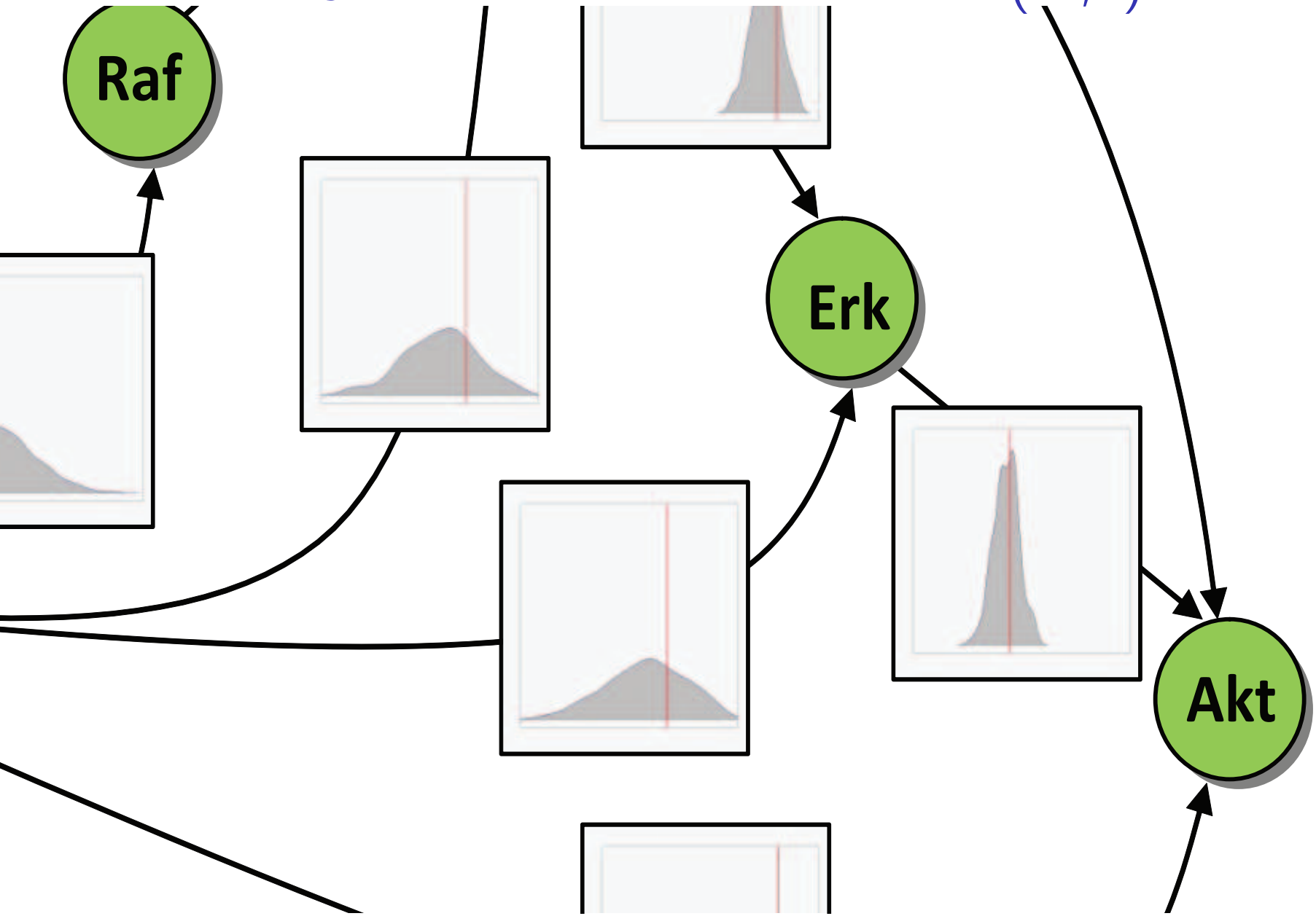
Simulations III: Prior bounds (-2,2)



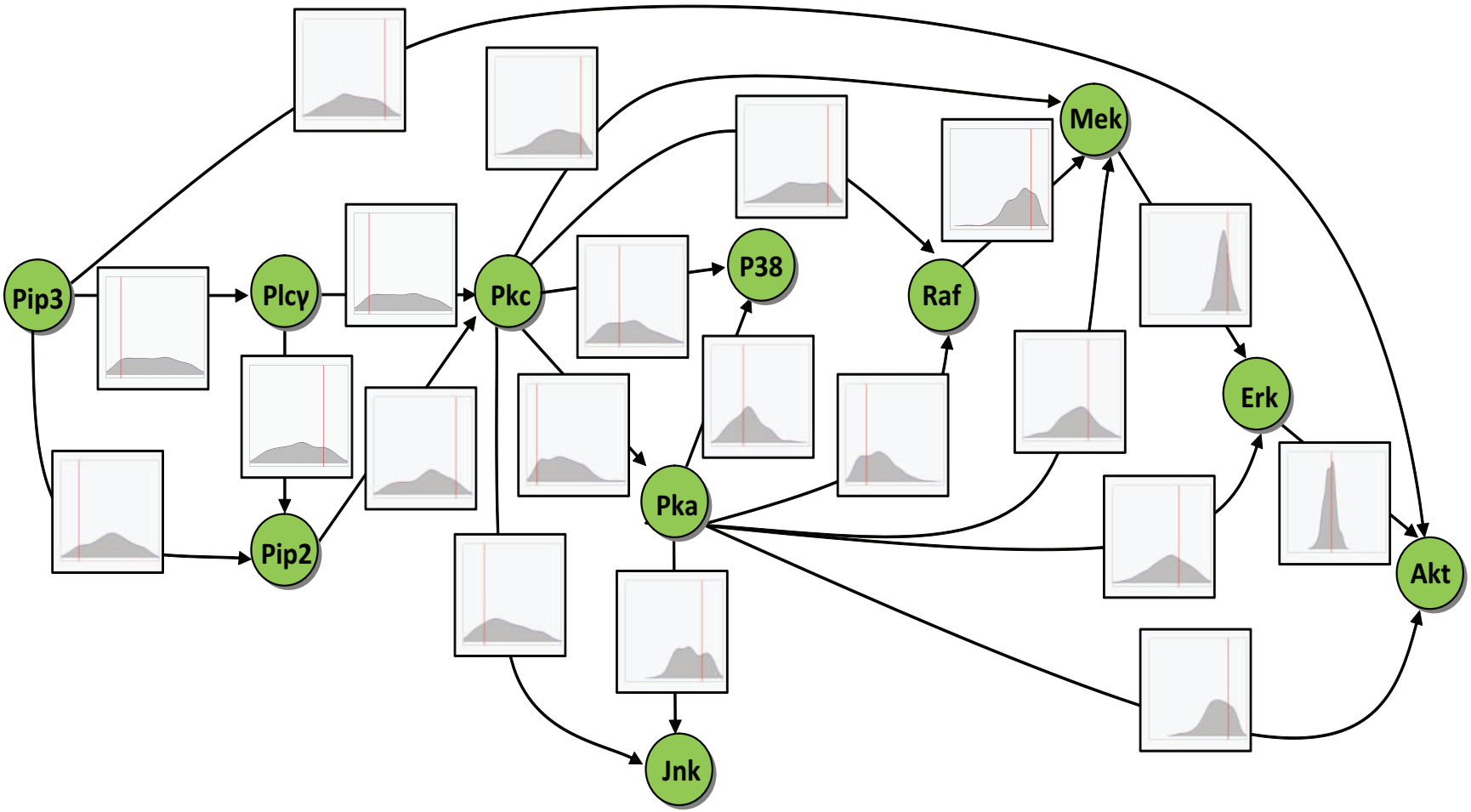
Simulations III: Prior bounds (-2,2)

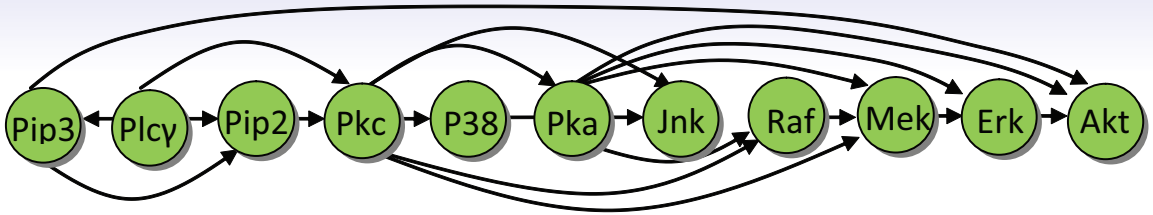


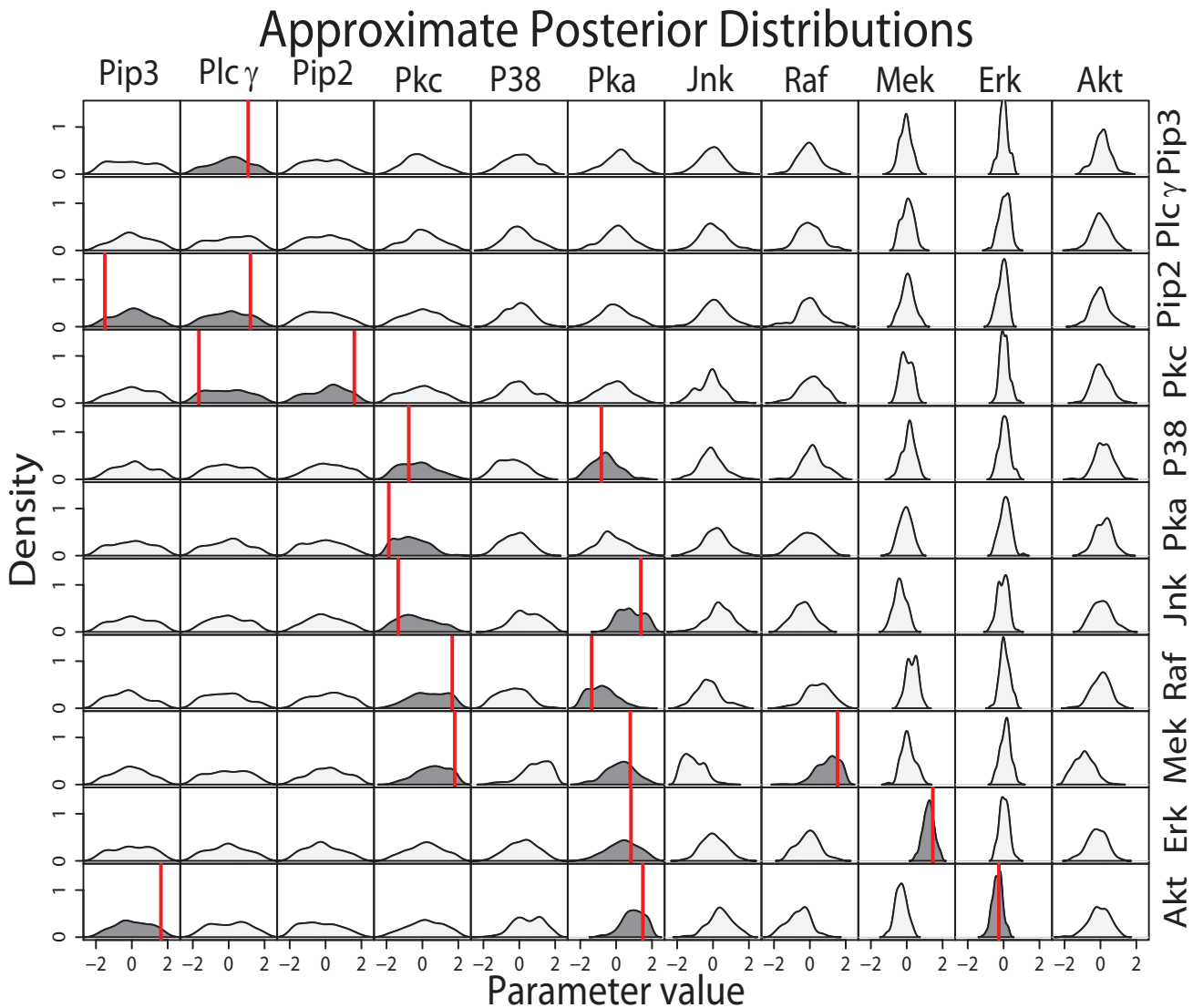
Simulations III: Prior bounds (-2,2)



Simulations III: Prior bounds (-2,2)




Pip3
Plcγ
Pip2
Pkc
P38
Pka
Jnk
Raf
Mek
Erk
Akt


Approximate Posterior Distributions
Pip3
Plc γ
Pip2
Pkc
P38
Pka
Jnk
Raf
Mek
Erk
Akt
Density
Parameter value

Simulations: Discussion

- "Flexible" and "rigid" edges yield additional information about the dynamics of the network
  - Rigidity and flexibility are closely linked to the network dynamics, robustness, and sensitivity

Simulations: Discussion

- "Flexible" and "rigid" edges yield additional information about the dynamics of the network
  - Rigidity and flexibility are closely linked to the network dynamics, robustness, and sensitivity
- Canberra, Euclidean, and Manhattan distances perform similarly in terms of AUC; MVT distance does not perform as well

Simulations: Discussion

- "Flexible" and "rigid" edges yield additional information about the dynamics of the network
  - Rigidity and flexibility are closely linked to the network dynamics, robustness, and sensitivity
- Canberra, Euclidean, and Manhattan distances perform similarly in terms of AUC; MVT distance does not perform as well
- Performance of ABC-Net deteriorates for alternative models when a VAR simulator is used
  - Alternative simulators may be used in situations where other models are known to be more appropriate

Simulations: Discussion

- "Flexible" and "rigid" edges yield additional information about the dynamics of the network
  - Rigidity and flexibility are closely linked to the network dynamics, robustness, and sensitivity
- Canberra, Euclidean, and Manhattan distances perform similarly in terms of AUC; MVT distance does not perform as well
- Performance of ABC-Net deteriorates for alternative models when a VAR simulator is used
  - Alternative simulators may be used in situations where other models are known to be more appropriate
- Wider prior bounds lead to convergence problems and may require more iterations

Data Analysis

- Different inference methods are better suited to different tasks:
  - Empirical Bayes Dynamic Bayesian Network (EBDBN) method (Rau et al. (2010)) is a hierarchical (empirical) Bayesian method for moderately sized networks (e.g., 50 - 100 genes):

$$\pi(\Theta|\mathbf{y}, \hat{\psi}) \propto f(\mathbf{y}|\Theta)\pi(\Theta|\hat{\psi})\pi(\hat{\psi})$$

▸ Details

  - ABC-Net algorithm may be used for detailed analyses of small, well-characterized networks (e.g., 10 - 20 genes)

Data Analysis

- Different inference methods are better suited to different tasks:
  - Empirical Bayes Dynamic Bayesian Network (EBDBN) method (Rau et al. (2010)) is a hierarchical (empirical) Bayesian method for moderately sized networks (e.g., 50 - 100 genes):

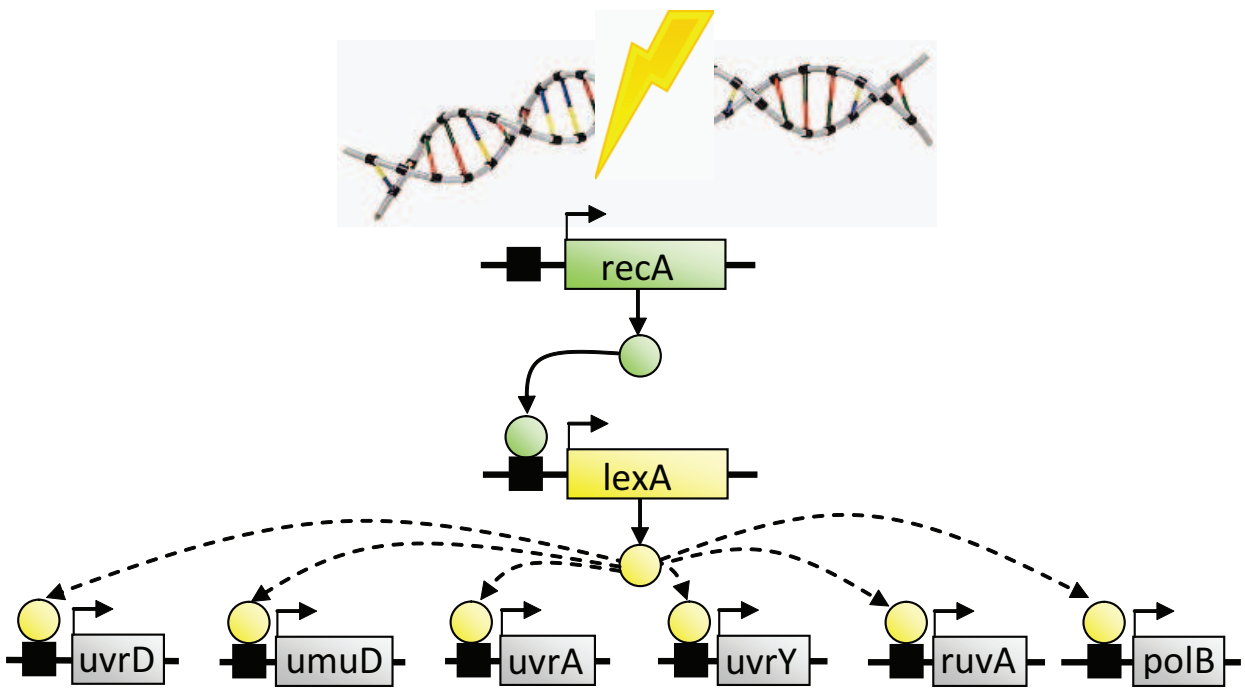
$$\pi(\Theta|\mathbf{y}, \hat{\boldsymbol{\psi}}) \propto f(\mathbf{y}|\Theta)\pi(\Theta|\hat{\boldsymbol{\psi}})\pi(\hat{\boldsymbol{\psi}})$$

▸ Details

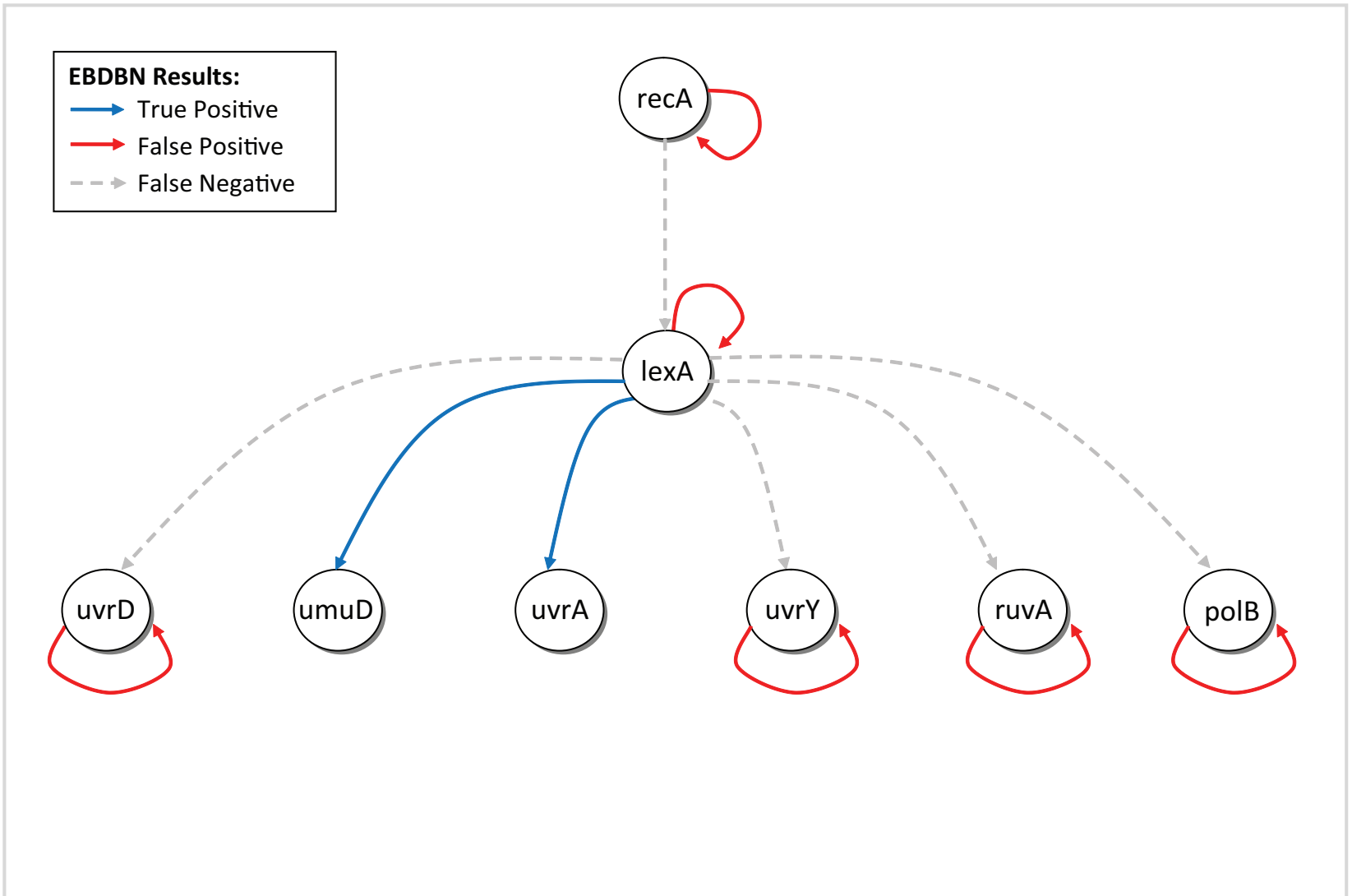
  - ABC-Net algorithm may be used for detailed analyses of small, well-characterized networks (e.g., 10 - 20 genes)
- Using two algorithms on a common task can help elucidate the strengths and weaknesses of each one:
  - S.O.S. DNA repair system in *Escherichia coli*

Data Analysis: S.O.S. DNA Repair System in *E. coli*

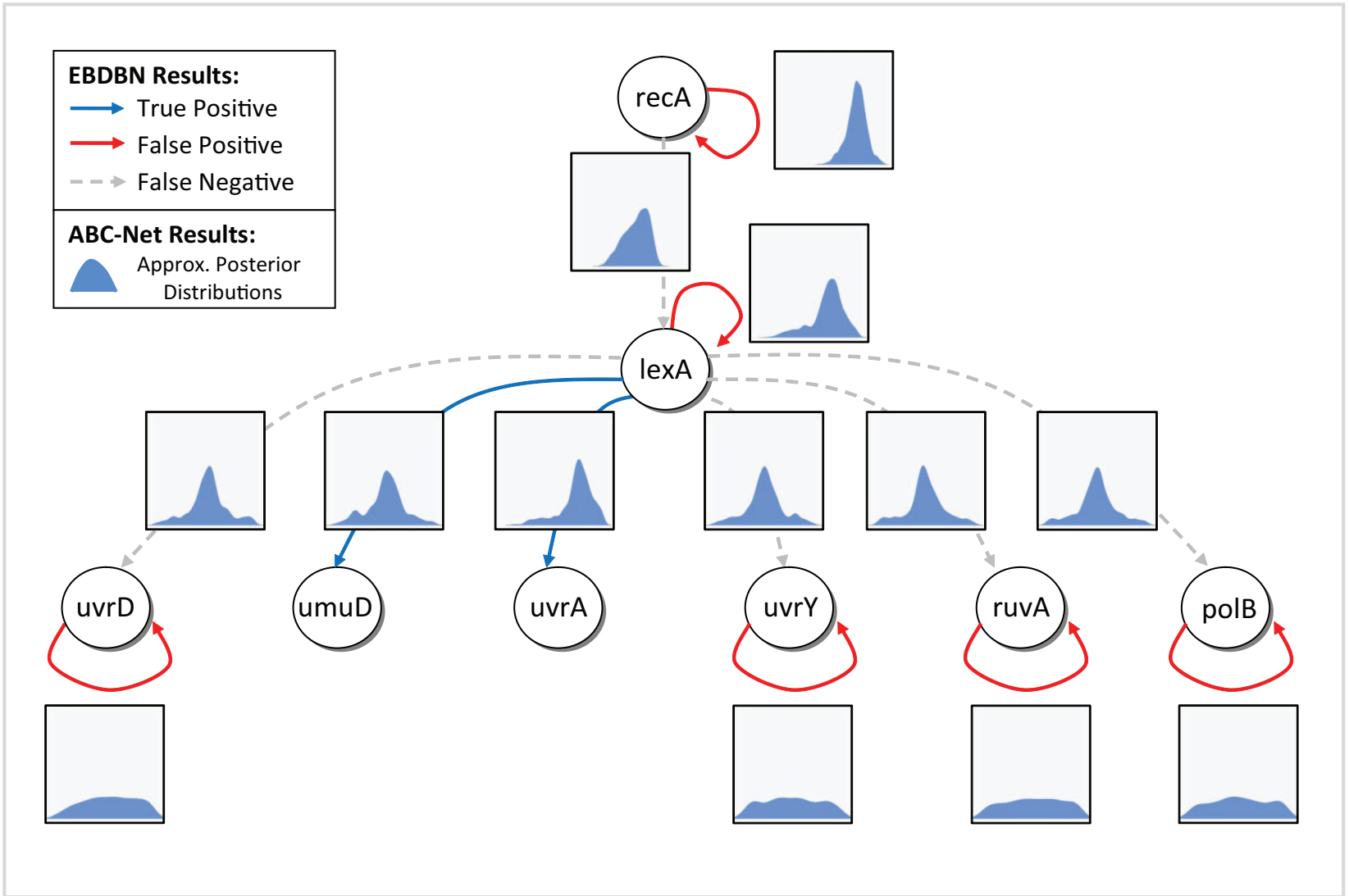
- S.O.S. DNA repair system of *Escherichia coli* (Ronen et al., 2002)
- 8 genes, with lexA as a master regulator that inhibits S.O.S. genes under normal conditions but activates them when DNA damage is sensed by recA ("single-input" module architecture)
- 50 time points, 1 replicate
- Maximum fan-in for ABC-Net method constrained to 2



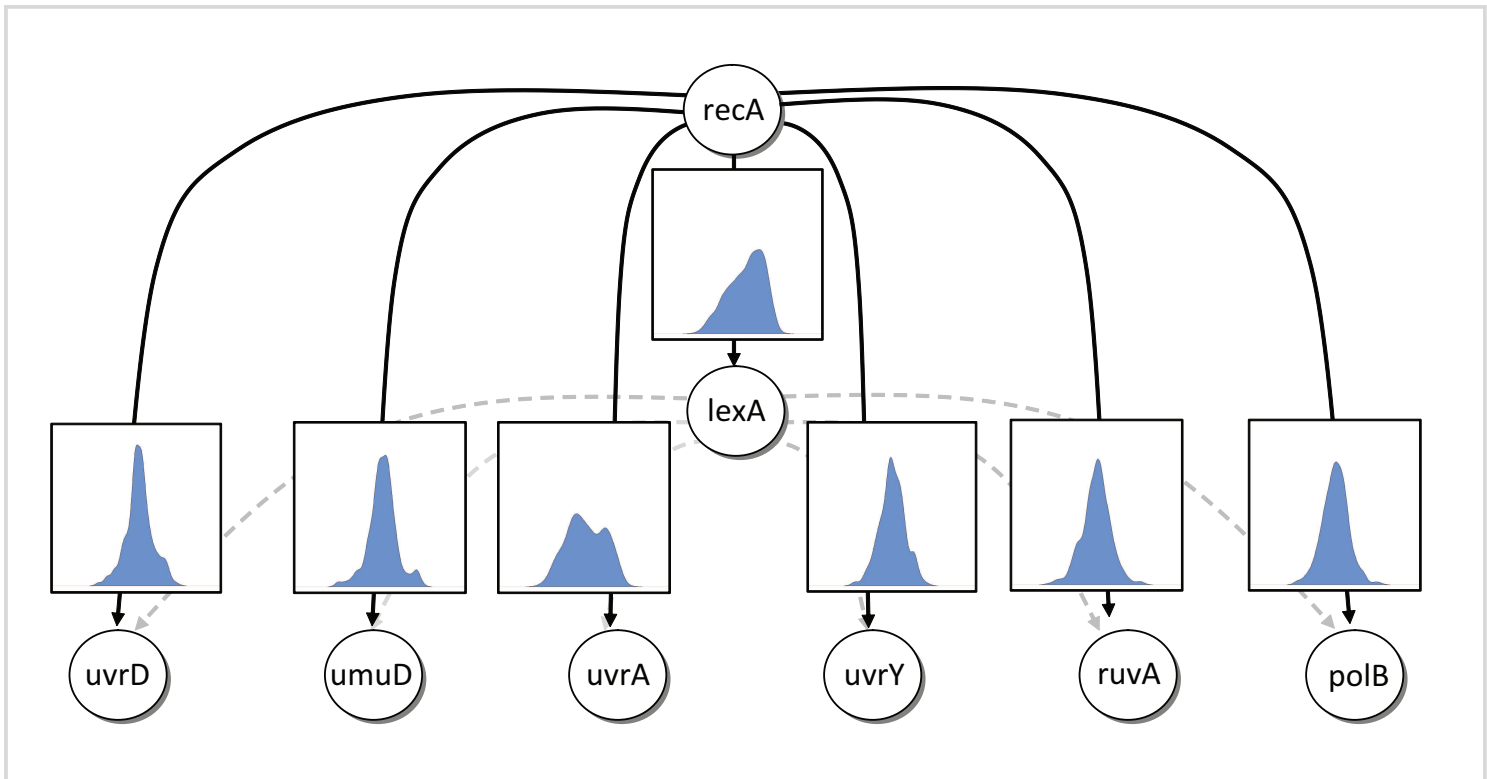
Results: S.O.S. DNA Repair System



Results: S.O.S. DNA Repair System



Results: S.O.S. DNA Repair System



Discussion: S.O.S. DNA Repair System

- In S.O.S. system, lexA decreases very rapidly, so S.O.S. genes turn on at about the same time
  - Time-delay models (e.g., autoregressive models) show stronger link between recA and S.O.S. genes

Discussion: S.O.S. DNA Repair System

- In S.O.S. system, lexA decreases very rapidly, so S.O.S. genes turn on at about the same time
  - Time-delay models (e.g., autoregressive models) show stronger link between recA and S.O.S. genes
- S.O.S. DNA repair is a simple, yet sophisticated network $\Rightarrow$ network is reacting to conditions within the cell

Discussion: S.O.S. DNA Repair System

- In S.O.S. system, lexA decreases very rapidly, so S.O.S. genes turn on at about the same time
  - Time-delay models (e.g., autoregressive models) show stronger link between recA and S.O.S. genes
- S.O.S. DNA repair is a simple, yet sophisticated network $\Rightarrow$ network is reacting to conditions within the cell
- "Rigid" and "flexible" edges identified by the ABC-Net algorithm can help clarify results from other inference methods

Summary

- Inferring gene regulatory networks is intrinsically difficult: complex network topology, small number of replicates and time points, noise in expression measurements

Summary

- Inferring gene regulatory networks is intrinsically difficult: complex network topology, small number of replicates and time points, noise in expression measurements
- Approximate Bayesian Computation methods can reveal information about the dynamics of biological systems from time-series gene expression data

Summary

- Inferring gene regulatory networks is intrinsically difficult: complex network topology, small number of replicates and time points, noise in expression measurements
- Approximate Bayesian Computation methods can reveal information about the dynamics of biological systems from time-series gene expression data
- ABC-MCMC Network (ABC-Net) approach uses a simulation-based Bayesian method with few distributional assumptions to infer approximate posterior distributions in small networks

Future Work

- Further examine components of ABC-Net method:
  - More sophisticated data simulators and techniques to identify optimal simulators for real data
  - Alternative and efficient network structure proposal schemes
  - Objective criterion to characterize approximate posterior distributions (e.g., introduce hierarchical prior on latent indicator variable G in ABC-Net method, and use local Bayes factor to quantitatively examine evidence of network edges)

Future Work

- Further examine components of ABC-Net method:
  - More sophisticated data simulators and techniques to identify optimal simulators for real data
  - Alternative and efficient network structure proposal schemes
  - Objective criterion to characterize approximate posterior distributions (e.g., introduce hierarchical prior on latent indicator variable G in ABC-Net method, and use local Bayes factor to quantitatively examine evidence of network edges)
- Examine alternative simulators and distance functions for time series digital gene expression measures (e.g., RNA sequencing data)

Future Work

- Further examine components of ABC-Net method:
  - More sophisticated data simulators and techniques to identify optimal simulators for real data
  - Alternative and efficient network structure proposal schemes
  - Objective criterion to characterize approximate posterior distributions (e.g., introduce hierarchical prior on latent indicator variable G in ABC-Net method, and use local Bayes factor to quantitatively examine evidence of network edges)
- Examine alternative simulators and distance functions for time series digital gene expression measures (e.g., RNA sequencing data)
- Develop statistical methods to combine results from multiple inference methods (i.e., consensus networks or model averaging)

Acknowledgements

Ph.D. advisor:
Rebecca W. Doerge

Ph.D. committee members:
Bruce Craig
Jayanta Ghosh
Alan Qi
Florence Jaffrézic (INRA-GABI)

Jean-Louis Foulley
RWD research group
My Truong
Doug Crabill

References

- Beaumont, M. *et al.* (2002) Approximate Bayesian computation in population genetics. *Genetics* 162, 2025-2035.
- Bortot, P. *et al.* (2007) Inference for stereological extremes. *JASA* March 2007, 84-92.
- Husmeier, D. (2003) Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics* 19, 2271-2282.
- Leuenberger, C. and Wegmann, D. (2009) Bayesian computation and model selection without likelihoods. *Genetics* 183, 1-10.
- Marjoram, P. *et al.* (2003) Markov chain Monte Carlo without likelihoods. *PNAS* 100, 15324-15328.
- Pritchard, J. *et al.* (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution* 16, 1791-1798.
- Ratmann, O. *et al.* (2007) Using likelihood-free inference to compare evolutionary dynamics of the protein networks of *H. pylorii* and *P. falciparum*. *PLoS Computational Biology* 3, 2266-2278.
- Rau, A. *et al.* (2010) An empirical Bayesian method for estimating biological networks from temporal microarray data. *SAGMB* 9:1, Article 9.
- Ronen, M. *et al.* (2002) Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics.*PNAS* 99, 10555-10560.
- Sisson, S. *et al.* (2007) Sequential Monte Carlo without likelihoods. *PNAS* 104, 1760-1765.
- Toni, T. and Stumpf, M. (2010) Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* 26, 104-110.

Recent reviews on ABC:

- Beaumont, M. *et al.* (2010) Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Sytematics* 41, 379-405.
- Csilléry, K. *et al.* (2010) Approximate Bayesian computation (ABC) in practice. *Trends in Ecology and Evolution*, 25, 410-418.

Likelihood-Free MCMC

Theorem

Under suitable regularity conditions, $\pi(\Theta|\rho(\mathbf{y}, \mathbf{y}^\star) \leq \epsilon)$ *is the stationary distribution of the chain.*

Likelihood-Free MCMC

Theorem

Under suitable regularity conditions, $\pi(\Theta|\rho(\mathbf{y}, \mathbf{y}^\star) \leq \epsilon)$ *is the stationary distribution of the chain.*

Let $r(\Theta \to \Theta^\star)$ be the transition mechanism of the chain. We must check whether $f(\Theta|\rho(\mathbf{y}^\star, \mathbf{y}) \leq \epsilon) r(\Theta \to \Theta^\star) = f(\Theta^\star|\rho(\mathbf{y}^\star, \mathbf{y}) \leq \epsilon) r(\Theta^\star \to \Theta)$.

Proof.

Without loss of generality, choose $\Theta^\star \neq \Theta$ such that

$$\frac{\pi(\Theta^\star)q(\Theta|\Theta^\star)}{\pi(\Theta)q(\Theta^\star|\Theta)} \leq 1$$

Detailed balance equation:

Back

Proof.

Without loss of generality, choose $\Theta^\star \neq \Theta$ such that

$$\frac{\pi(\Theta^\star)q(\Theta|\Theta^\star)}{\pi(\Theta)q(\Theta^\star|\Theta)} \leq 1$$

Detailed balance equation:

$f(\Theta|\rho(\mathbf{y}^\star, \mathbf{y}) \leq \epsilon)r(\Theta \to \Theta^\star) =$

Proof.

Without loss of generality, choose $\Theta^\star \neq \Theta$ such that

$$\frac{\pi(\Theta^\star)q(\Theta|\Theta^\star)}{\pi(\Theta)q(\Theta^\star|\Theta)} \leq 1$$

Detailed balance equation:

$$f(\Theta|\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon)r(\Theta \to \Theta^\star) =$$
$$= f(\Theta|\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon)q(\Theta^\star|\Theta)\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta^\star\right]\alpha(\Theta,\Theta^\star)$$

Proof.

Without loss of generality, choose $\Theta^\star \neq \Theta$ such that

$$\frac{\pi(\Theta^\star)q(\Theta|\Theta^\star)}{\pi(\Theta)q(\Theta^\star|\Theta)} \leq 1$$

Detailed balance equation:

$$\begin{aligned} f(\Theta|\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon) r(\Theta \to \Theta^\star) = \\ &= f(\Theta|\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon) q(\Theta^\star|\Theta) \mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta^\star\right] \alpha(\Theta,\Theta^\star) \\ &= \frac{\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta\right]\pi(\Theta)}{\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon\right]} \left\{ q(\Theta^\star|\Theta)\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta^\star\right] \times \frac{\pi(\Theta^\star)q(\Theta|\Theta^\star)}{\pi(\Theta)q(\Theta^\star|\Theta)} \right\} \end{aligned}$$

Back

Proof.

Without loss of generality, choose $\Theta^\star \neq \Theta$ such that

$$\frac{\pi(\Theta^\star)q(\Theta|\Theta^\star)}{\pi(\Theta)q(\Theta^\star|\Theta)} \leq 1$$

Detailed balance equation:

$$\begin{aligned}
f(\Theta|\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon) r(\Theta \to \Theta^\star) &= \\
&= f(\Theta|\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon) q(\Theta^\star|\Theta) \mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta^\star\right] \alpha(\Theta,\Theta^\star) \\
&= \frac{\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta\right]\pi(\Theta)}{\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon\right]} \left\{ q(\Theta^\star|\Theta)\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta^\star\right] \times \frac{\pi(\Theta^\star)q(\Theta|\Theta^\star)}{\pi(\Theta)q(\Theta^\star|\Theta)} \right\} \\
&= \frac{\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta^\star\right]\pi(\Theta^\star)}{\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon\right]} \left\{ q(\Theta|\Theta^\star)\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta\right] \right\}
\end{aligned}$$

Proof.

Without loss of generality, choose $\Theta^\star \neq \Theta$ such that

$$\frac{\pi(\Theta^\star)q(\Theta|\Theta^\star)}{\pi(\Theta)q(\Theta^\star|\Theta)} \leq 1$$

Detailed balance equation:

$$\begin{aligned}
f(\Theta|\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon) r(\Theta \to \Theta^\star) &= \\
&= f(\Theta|\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon) q(\Theta^\star|\Theta) \mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta^\star\right] \alpha(\Theta,\Theta^\star) \\
&= \frac{\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta\right]\pi(\Theta)}{\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon\right]} \left\{ q(\Theta^\star|\Theta)\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta^\star\right] \times \frac{\pi(\Theta^\star)q(\Theta|\Theta^\star)}{\pi(\Theta)q(\Theta^\star|\Theta)} \right\} \\
&= \frac{\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta^\star\right]\pi(\Theta^\star)}{\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon\right]} \left\{ q(\Theta|\Theta^\star)\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta\right] \right\} \\
&= f(\Theta^\star|\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon) q(\Theta|\Theta^\star) \mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta\right] \alpha(\Theta^\star,\Theta)
\end{aligned}$$

Proof.

Without loss of generality, choose $\Theta^\star \neq \Theta$ such that

$$\frac{\pi(\Theta^\star)q(\Theta|\Theta^\star)}{\pi(\Theta)q(\Theta^\star|\Theta)} \leq 1$$

Detailed balance equation:

$$\begin{aligned}
f(\Theta|\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon) r(\Theta \to \Theta^\star) &= \\
&= f(\Theta|\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon) q(\Theta^\star|\Theta) \mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta^\star\right] \alpha(\Theta,\Theta^\star) \\
&= \frac{\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta\right]\pi(\Theta)}{\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon\right]} \left\{ q(\Theta^\star|\Theta)\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta^\star\right] \times \frac{\pi(\Theta^\star)q(\Theta|\Theta^\star)}{\pi(\Theta)q(\Theta^\star|\Theta)} \right\} \\
&= \frac{\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta^\star\right]\pi(\Theta^\star)}{\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon\right]} \left\{ q(\Theta|\Theta^\star)\mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta\right] \right\} \\
&= f(\Theta^\star|\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon) q(\Theta|\Theta^\star) \mathbb{P}\left[\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon|\Theta\right] \alpha(\Theta^\star,\Theta) \\
&= f(\Theta^\star|\rho(\mathbf{y}^\star,\mathbf{y}) \leq \epsilon) r(\Theta^\star \to \Theta)
\end{aligned}$$

Back

Simulation Procedure

- One-step-ahead predictors: $\mathbf{y}^\star = \Theta^\star \mathbf{y}_{t-1}$
  - Suppose we have complete knowledge from the past ($\mathbf{y}_{t-1}$) and we want to predict (simulate) expression at time t based on the current network $\Theta^\star$: $\mathbf{y}_t^\star = \hat{\mathbf{y}}_t = E(\mathbf{y}_t | \mathbf{y}_{t-1}, \Theta^\star)$.
  - Data from previous time point used to determine forecasted (simulated) values at current time point

Simulation Procedure

- One-step-ahead predictors: $\mathbf{y}^\star = \Theta^\star \mathbf{y}_{t-1}$
  - Suppose we have complete knowledge from the past ($\mathbf{y}_{t-1}$) and we want to predict (simulate) expression at time t based on the current network $\Theta^\star$: $\mathbf{y}_t^\star = \hat{\mathbf{y}}_t = E(\mathbf{y}_t | \mathbf{y}_{t-1}, \Theta^\star)$.
  - Data from previous time point used to determine forecasted (simulated) values at current time point
- Toni and Stumpf (2010): Numerically solved ordinary differential equations and added noise to obtain simulated time-course data

Simulation Procedure

- One-step-ahead predictors: $\mathbf{y}^\star = \Theta^\star \mathbf{y}_{t-1}$
  - Suppose we have complete knowledge from the past ($\mathbf{y}_{t-1}$) and we want to predict (simulate) expression at time t based on the current network $\Theta^\star$: $\mathbf{y}_t^\star = \hat{\mathbf{y}}_t = E(\mathbf{y}_t | \mathbf{y}_{t-1}, \Theta^\star)$.
  - Data from previous time point used to determine forecasted (simulated) values at current time point
- Toni and Stumpf (2010): Numerically solved ordinary differential equations and added noise to obtain simulated time-course data
- Ratmann *et al*. (2007): Evolutionary history of protein interaction networks simulated using mixture evolution model (average summary statistics over 50 generated networks)

Simulation Procedure

- One-step-ahead predictors: $\mathbf{y}^\star = \Theta^\star \mathbf{y}_{t-1}$
  - Suppose we have complete knowledge from the past ($\mathbf{y}_{t-1}$) and we want to predict (simulate) expression at time t based on the current network $\Theta^\star$: $\mathbf{y}_t^\star = \hat{\mathbf{y}}_t = E(\mathbf{y}_t | \mathbf{y}_{t-1}, \Theta^\star)$.
  - Data from previous time point used to determine forecasted (simulated) values at current time point
- Toni and Stumpf (2010): Numerically solved ordinary differential equations and added noise to obtain simulated time-course data
- Ratmann *et al*. (2007): Evolutionary history of protein interaction networks simulated using mixture evolution model (average summary statistics over 50 generated networks)
- Marjoram *et al*. (2003): DNA sequences simulated using coalescent trees

Back

LFN Implementation Details

- Burn-in period
  - Cooling procedure: Temper acceptance with exponential cooling scheme, starting at some initial temperature ϵ_0 and cooling to $\epsilon_{i+1} = \lambda\epsilon_i$ until the minimal temperature $\epsilon_{\text{min}} = \epsilon$ is reached. We use $\lambda = 0.90$ and set $\epsilon_0 = \epsilon\lambda^{-10}$.
  - Use each ϵ_i for 200 iterations, then cool to next value.
  - If ϵ_{min} is reached and the acceptance rate for the chain $\leq 1\%$, the burn-in period is reinitialized.

LFN Implementation Details

- Burn-in period
  - Cooling procedure: Temper acceptance with exponential cooling scheme, starting at some initial temperature ϵ_0 and cooling to $\epsilon_{i+1} = \lambda\epsilon_i$ until the minimal temperature $\epsilon_{\text{min}} = \epsilon$ is reached. We use $\lambda = 0.90$ and set $\epsilon_0 = \epsilon\lambda^{-10}$.
  - Use each ϵ_i for 200 iterations, then cool to next value.
  - If ϵ_{min} is reached and the acceptance rate for the chain $\leq 1\%$, the burn-in period is reinitialized.
- Chain length:
  - 10 chains for 1×10^6 iterations each (1×10^7 iterations total)
  - Thinning interval of 50 (2×10^5 remaining iterations)
  - Inference made on samples corresponding to smallest 1% of $\rho(\mathbf{y}^\star, \mathbf{y})$ (2000 iterations)

Back

Approximate Bayesian Methods I

Empirical Bayes Dynamic Bayesian Network (EBDBN) Algorithm

- Parameters Θ (i.e., network edges) may be related to one another

Approximate Bayesian Methods I

Empirical Bayes Dynamic Bayesian Network (EBDBN) Algorithm

- Parameters Θ (i.e., network edges) may be related to one another $\Rightarrow$ Hierarchical Bayes model samples parameters from a common distribution: $\pi(\Theta|\mathbf{y}, \boldsymbol{\psi}) \propto f(\mathbf{y}|\Theta)\pi(\Theta|\boldsymbol{\psi})\pi(\boldsymbol{\psi})$

Approximate Bayesian Methods I

Empirical Bayes Dynamic Bayesian Network (EBDBN) Algorithm

- Parameters Θ (i.e., network edges) may be related to one another $\Rightarrow$ Hierarchical Bayes model samples parameters from a common distribution: $\pi(\Theta|\mathbf{y}, \boldsymbol{\psi}) \propto f(\mathbf{y}|\Theta)\pi(\Theta|\boldsymbol{\psi})\pi(\boldsymbol{\psi})$
- Idea: Use observed data $\mathbf{y}$ to estimate $\boldsymbol{\psi}$
- Common parametric EB models (based on conjugate distributions) include Poisson-Gamma, Beta-binomial, multinomial-Dirichlet, and **Gaussian-Gaussian**

Approximate Bayesian Methods I

Empirical Bayes Dynamic Bayesian Network (EBDBN) Algorithm

- Parameters Θ (i.e., network edges) may be related to one another $\Rightarrow$ Hierarchical Bayes model samples parameters from a common distribution: $\pi(\Theta|\mathbf{y}, \boldsymbol{\psi}) \propto f(\mathbf{y}|\Theta)\pi(\Theta|\boldsymbol{\psi})\pi(\boldsymbol{\psi})$
- Idea: Use observed data $\mathbf{y}$ to estimate $\boldsymbol{\psi}$
- Common parametric EB models (based on conjugate distributions) include Poisson-Gamma, Beta-binomial, multinomial-Dirichlet, and **Gaussian-Gaussian**
- Let $\mathbf{y}_t$ and $\mathbf{x}_t$ be observed gene expression and unobserved hidden states measured at time t:

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{y}_{t-1} + \mathbf{w}_t$$
$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{\Theta}\mathbf{y}_{t-1} + \mathbf{v}_t$$
$$\mathbf{w}_t \sim N(0, I), \mathbf{v}_t \sim N(0, V^{-1}))$$

Empirical Bayes Dynamic Bayesian Network (EBDBN): Rau *et al.*, 2010

Let $\mathbf{y} = \{\mathbf{y}_t\}_{t=1,\ldots,T}$ be observed gene expression and $\mathbf{x} = \{\mathbf{x}_t\}_{t=1,\ldots,T}$ be unobserved hidden states measured at the same time points.

Empirical Bayes Dynamic Bayesian Network (EBDBN): Rau *et al.*, 2010

Let $\mathbf{y} = \{\mathbf{y}_t\}_{t=1,\ldots,T}$ be observed gene expression and $\mathbf{x} = \{\mathbf{x}_t\}_{t=1,\ldots,T}$ be unobserved hidden states measured at the same time points.

- Linear Gaussian state space model:

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{y}_{t-1} + \mathbf{w}_t$$
$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{\Theta}\mathbf{y}_{t-1} + \mathbf{v}_t$$
$$\mathbf{w}_t \sim N(0, I), \mathbf{v}_t \sim N(0, V^{-1}))$$

Empirical Bayes Dynamic Bayesian Network (EBDBN): Rau *et al.*, 2010

Let $\mathbf{y} = \{\mathbf{y}_t\}_{t=1,\ldots,T}$ be observed gene expression and $\mathbf{x} = \{\mathbf{x}_t\}_{t=1,\ldots,T}$ be unobserved hidden states measured at the same time points.

- Linear Gaussian state space model:

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{y}_{t-1} + \mathbf{w}_t$$
$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{\Theta}\mathbf{y}_{t-1} + \mathbf{v}_t$$
$$\mathbf{w}_t \sim N(0, I), \mathbf{v}_t \sim N(0, V^{-1}))$$

$$\mathbf{A}_{\text{rows}} \sim N(0, \alpha^{-1}) \qquad \mathbf{B}_{\text{rows}} \sim N(0, \beta^{-1})$$
$$\mathbf{C}_{\text{rows}} \sim N(0, (V\gamma)^{-1}) \qquad \mathbf{\Theta}_{\text{rows}} \sim N(0, (V\delta)^{-1})$$

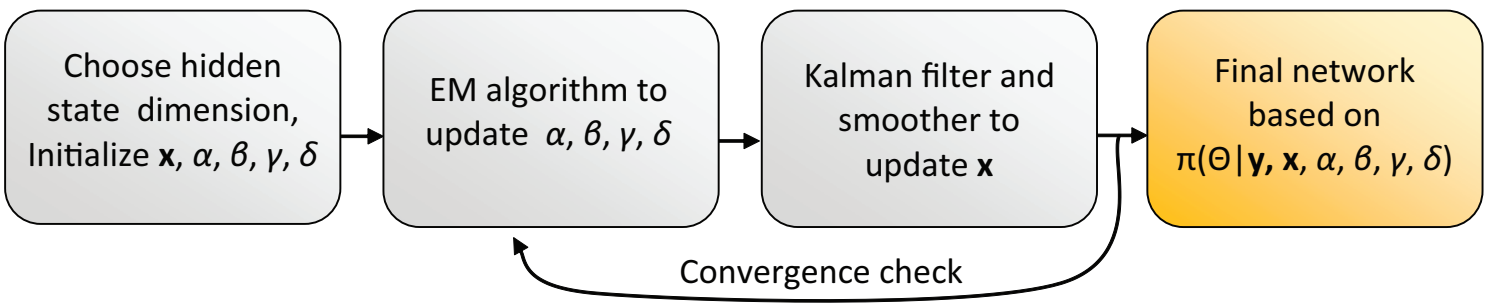
Empirical Bayes Dynamic Bayesian Network (EBDBN): Rau *et al.*, 2010

Let $\mathbf{y} = \{\mathbf{y}_t\}_{t=1,\ldots,T}$ be observed gene expression and $\mathbf{x} = \{\mathbf{x}_t\}_{t=1,\ldots,T}$ be unobserved hidden states measured at the same time points.

- Linear Gaussian state space model:

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{y}_{t-1} + \mathbf{w}_t$$
$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{\Theta}\mathbf{y}_{t-1} + \mathbf{v}_t$$
$$\mathbf{w}_t \sim N(0, I), \mathbf{v}_t \sim N(0, V^{-1}))$$

$$\mathbf{A}_{\text{rows}} \sim N(0, \alpha^{-1}) \qquad \mathbf{B}_{\text{rows}} \sim N(0, \beta^{-1})$$
$$\mathbf{C}_{\text{rows}} \sim N(0, (V\gamma)^{-1}) \qquad \mathbf{\Theta}_{\text{rows}} \sim N(0, (V\delta)^{-1})$$



Back