MLSB[12]

Network Inference Using Steady-State Data and Goldbeter-Koshland Kinetics

C.J.Oates[1,2], B.T.Hennessy[3], Y.Lu[4], G.B.Mills[4] and S.Mukherjee[2,1]

[1] Dept. Biochemistry, Nederlands Kanker Instituut
[2] Depts. Statistics & Complexity, Univ. Warwick
[3] Dept. Med. Oncol., Beaumont Hospital, Dublin
[4] Dept. Sys. Bio., Tex. M.D.Anderson Cancer Ctr.

c.oates@nki.nl; s.mukherjee@nki.nl

9th September, 2012

Outline

- ~~What is network inference?~~

Outline

- ~~What is network inference?~~
- Problems with the linear model

Outline

- ~~What is network inference?~~
- Problems with the linear model
- Problems with steady-state data

Outline

- ~~What is network inference?~~
- Problems with the linear model
- Problems with steady-state data
- "Kinetics-driven" inference

Outline

- ~~What is network inference?~~
- Problems with the linear model
- Problems with steady-state data
- "Kinetics-driven" inference
- Preliminary results

Problems with the linear model

Problems with the linear model

The linear model in regression notation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

Problems with the linear model

The linear model in regression notation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

As usual we consider the following quantities:

Problems with the linear model

The linear model in regression notation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

As usual we consider the following quantities:

- An (observed) response y (phosphoprotein)

Problems with the linear model

The linear model in regression notation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

As usual we consider the following quantities:

- An (observed) response y (phosphoprotein)
- (Known) covariates $x_1, x_2, \ldots, x_p$ (other phosphoproteins)

Problems with the linear model

The linear model in regression notation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

As usual we consider the following quantities:

- An (observed) response y (phosphoprotein)
- (Known) covariates $x_1, x_2, \ldots, x_p$ (other phosphoproteins)
- (Unknown) regression coefficients $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$

Problems with the linear model

The linear model in regression notation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

As usual we consider the following quantities:

- An (observed) response y (phosphoprotein)
- (Known) covariates $x_1, x_2, \ldots, x_p$ (other phosphoproteins)
- (Unknown) regression coefficients $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$

There are two subsets of covariates which may be of interest:

Problems with the linear model

The linear model in regression notation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

As usual we consider the following quantities:

- An (observed) response y (phosphoprotein)
- (Known) covariates $x_1, x_2, \ldots, x_p$ (other phosphoproteins)
- (Unknown) regression coefficients $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$

There are two subsets of covariates which may be of interest:

- Covariates $\mathcal{P}$ which, together, are (in some sense) optimal for prediction of response

Problems with the linear model

The linear model in regression notation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

As usual we consider the following quantities:

- An (observed) response y (phosphoprotein)
- (Known) covariates $x_1, x_2, \ldots, x_p$ (other phosphoproteins)
- (Unknown) regression coefficients $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$

There are two subsets of covariates which may be of interest:

- Covariates $\mathcal{P}$ which, together, are (in some sense) optimal for prediction of response
- Covariates $\mathcal{C}$ such that each member of $\mathcal{C}$ directly **causes** the response, in the interventional sense (Pearl, 2009)

Problems with the linear model

The linear model in regression notation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

As usual we consider the following quantities:

- An (observed) response y (phosphoprotein)
- (Known) covariates $x_1, x_2, \ldots, x_p$ (other phosphoproteins)
- (Unknown) regression coefficients $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$

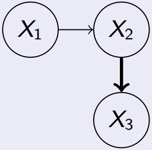
There are two subsets of covariates which may be of interest:

- Covariates $\mathcal{P}$ which, together, are (in some sense) optimal for prediction of response
- Covariates $\mathcal{C}$ such that each member of $\mathcal{C}$ directly **causes** the response, in the interventional sense (Pearl, 2009)

In **signalling network inference** we aim to find $\mathcal{C}$.

Example: $\mathcal{P} \neq \mathcal{C}$

Consider the simple system of structural equations:



$$\begin{aligned} X_1 &:= \epsilon_1 \\ X_2 &:= X_1 + 0.1\epsilon_2 \\ X_3 &:= X_2 + 0.01\epsilon_3 \end{aligned}$$

where $\epsilon_j \sim N(0,1)$ are i.i.d. Then X_3 is a much better predictor of X_2 than X_1, even though X_3 does not drive the variation in X_2.

- Biological goals are often the prioritisation of interventional experiments (e.g. knock down, knock out, etc.) rather than prediction per se.

- Biological goals are often the prioritisation of interventional experiments (e.g. knock down, knock out, etc.) rather than prediction per se.
- However, mainstream bioinformatics approaches rely on variable selection techniques developed for prediction.

- Biological goals are often the prioritisation of interventional experiments (e.g. knock down, knock out, etc.) rather than prediction per se.
- However, mainstream bioinformatics approaches rely on variable selection techniques developed for prediction.
- Commonly used procedures are
    - Bayesian variable selection (median model is predictive optimal)
    - L_1 penalisation (lower MSE than MLE)

- Biological goals are often the prioritisation of interventional experiments (e.g. knock down, knock out, etc.) rather than prediction per se.
- However, mainstream bioinformatics approaches rely on variable selection techniques developed for prediction.
- Commonly used procedures are
  - Bayesian variable selection (median model is predictive optimal)
  - L_1 penalisation (lower MSE than MLE)
- Linear regression based approaches have drawbacks in this setting:

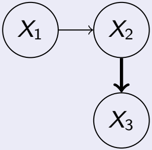
- Biological goals are often the prioritisation of interventional experiments (e.g. knock down, knock out, etc.) rather than prediction per se.
- However, mainstream bioinformatics approaches rely on variable selection techniques developed for prediction.
- Commonly used procedures are
  - Bayesian variable selection (median model is predictive optimal)
  - L_1 penalisation (lower MSE than MLE)
- Linear regression based approaches have drawbacks in this setting:
  - Model misspecification may lead to inefficient or inconsistent estimation (Heagerty and Kurland, 2001; Lv and Liu, 2010).

- Biological goals are often the prioritisation of interventional experiments (e.g. knock down, knock out, etc.) rather than prediction per se.
- However, mainstream bioinformatics approaches rely on variable selection techniques developed for prediction.
- Commonly used procedures are
  - Bayesian variable selection (median model is predictive optimal)
  - L_1 penalisation (lower MSE than MLE)
- Linear regression based approaches have drawbacks in this setting:
  - Model misspecification may lead to inefficient or inconsistent estimation (Heagerty and Kurland, 2001; Lv and Liu, 2010).
  - Variates may be highly correlated.

- Biological goals are often the prioritisation of interventional experiments (e.g. knock down, knock out, etc.) rather than prediction per se.
- However, mainstream bioinformatics approaches rely on variable selection techniques developed for prediction.
- Commonly used procedures are
  - Bayesian variable selection (median model is predictive optimal)
  - L_1 penalisation (lower MSE than MLE)
- Linear regression based approaches have drawbacks in this setting:
  - Model misspecification may lead to inefficient or inconsistent estimation (Heagerty and Kurland, 2001; Lv and Liu, 2010).
  - Variates may be highly correlated.
  - Symmetry of the linear equivalence precludes identification of underlying causal relationships (Pearl, 2009).

Example: Symmetry of the linear equivalence

Consider the same system of structural equations:

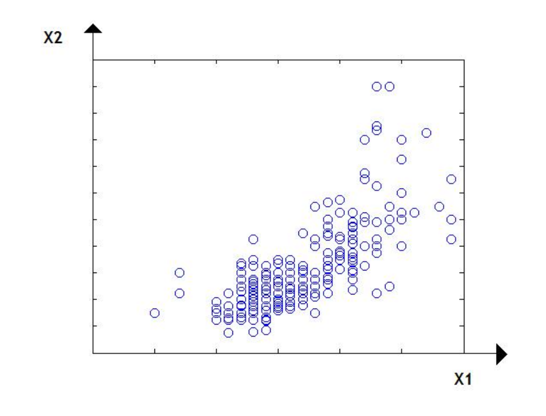


$$\begin{aligned} X_1 &:= \epsilon_1 \\ X_2 &:= X_1 + 0.1\epsilon_2 \\ X_3 &:= X_2 + 0.01\epsilon_3 \end{aligned}$$

where $\epsilon_j \sim N(0, 1)$ are i.i.d.

- X_3 highly correlated with X_2
- Only the undirected graph is recoverable from observation.

Example: Nonlinearity aids causal inference



$$\{X_1, X_2\} = \{\text{weight}, \text{height}\}$$

A bit more generally

A bit more generally

- Any nonlinearity will do, as long as the inverse isn't physically plausible.

A bit more generally

- Any nonlinearity will do, as long as the inverse isn't physically plausible.
- Consider functional relationships $X_i = f_i(\mathbf{X}_{Pa(i)}, \epsilon_i)$ belonging to a **functional model class** $f_i \in \mathcal{F}$.

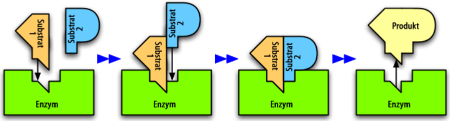
A bit more generally

- Any nonlinearity will do, as long as the inverse isn't physically plausible.
- Consider functional relationships $X_i = f_i(\mathbf{X}_{Pa(i)}, \epsilon_i)$ belonging to a **functional model class** $f_i \in \mathcal{F}$.
- The underlying causal structure is **identifiable** if, for each i, each $j \in Pa(i)$ and each $\mathbf{X}_{Pa(i)\setminus\{j\}}$, the functional

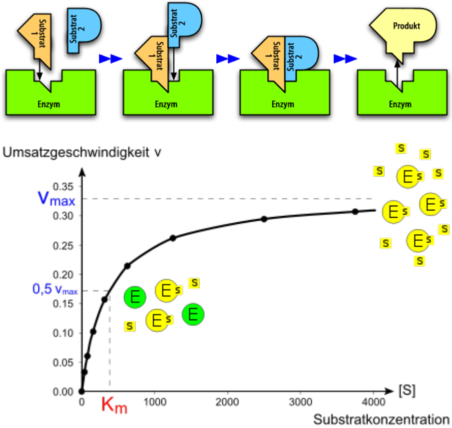
$$f_i(\mathbf{X}_{Pa(i)\setminus\{j\}}, \underbrace{\cdot}_{X_j}, \underbrace{\cdot}_{\epsilon_i})$$

is bivariate identifiable in $\mathcal{F}$ (Peters *et al.*, 2011).

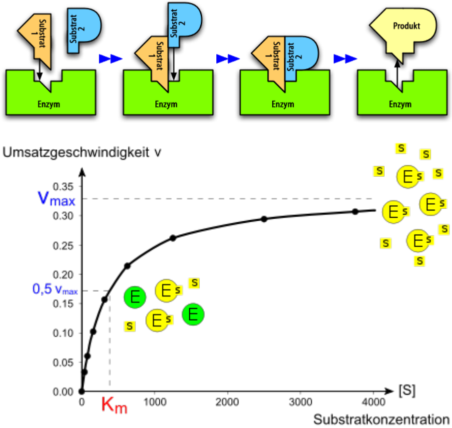
Any nonlinearity will do...



Any nonlinearity will do...



Any nonlinearity will do...



This nonlinearity is captured by the **Michaelis-Menten** functional

$$\frac{d[\text{product}]}{dt} = \frac{V \times [\text{enzyme}] \times [\text{substrate}]}{K + [\text{substrate}]}$$

Problems with steady-state data

Problems with steady-state data

In the dynamical setting, define **causality** as a nontrivial dependence

$$\frac{dX_i}{dt} = f_i(\mathbf{X}_{Pa(i)}, \mathbf{U}; \boldsymbol{\theta}_i)$$

where $\mathbf{X}_{\mathcal{C}} = (X_c)_{c \in \mathcal{C}}$.

Problems with steady-state data

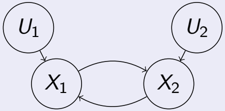
In the dynamical setting, define **causality** as a nontrivial dependence

$$\frac{dX_i}{dt} = f_i(\mathbf{X}_{Pa(i)}, \mathbf{U}; \boldsymbol{\theta}_i)$$

where $\mathbf{X}_{\mathcal{C}} = (X_c)_{c \in \mathcal{C}}$.

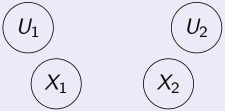
Example: Nonidentifiability at equilibrium

Consider the following dynamical system:



G_{dyn}

$$\begin{aligned} \dot{X}_1 &:= -X_2 \\ \dot{X}_2 &:= -X_1 \end{aligned}$$



G_{equ}

At equilibrium $x_2 := 0$, $x_1 := 0$, so it is not possible to infer causal structure at equilibrium.

Faithfulness at equilibrium

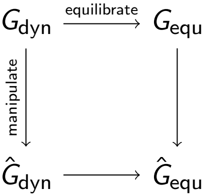
Faithfulness at equilibrium

PROBLEM: The equilibrium distribution may be **unfaithful** to the causal structure derived from dynamics.

Faithfulness at equilibrium

PROBLEM: The equilibrium distribution may be **unfaithful** to the causal structure derived from dynamics.

- Intervention in an equilibrium model G_{equ} forces the system out of equilibrium, into G_{dyn} (Dash, 2003).



Faithfulness at equilibrium

PROBLEM: The equilibrium distribution may be **unfaithful** to the causal structure derived from dynamics.

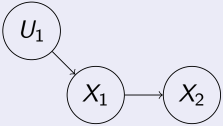
- Intervention in an equilibrium model G_{equ} forces the system out of equilibrium, into G_{dyn} (Dash, 2003).

$$\begin{array}{ccc} G_{\text{dyn}} & \xrightarrow{\text{equilibrate}} & G_{\text{equ}} \\ \Big\downarrow{\scriptstyle\text{manipulate}} & & \Big\downarrow \\ \hat{G}_{\text{dyn}} & \longrightarrow & \hat{G}_{\text{equ}} \end{array}$$

- This must be a commutative diagram. i.e. the operators "manipulation" and "equilibrate" must commute.

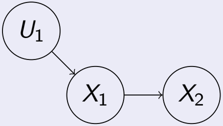
Example: Faithfulness at equilibrium

Consider the following dynamical system:



G_{dyn}

$$\begin{aligned}\dot{X}_1 &:= -X_1 \\ \dot{X}_2 &:= -X_2 - X_1\end{aligned}$$

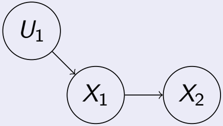


G_{equ}

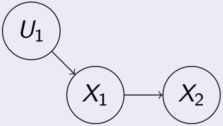
Here the equilibrium relations are $x_1 := 0$, $x_2 := -x_1$, so G_{equ} is faithful.

Example: Faithfulness at equilibrium

Consider the following dynamical system:



$$\begin{aligned} \dot{X}_1 &:= -X_1 \\ \dot{X}_2 &:= -X_2 - X_1 \end{aligned}$$



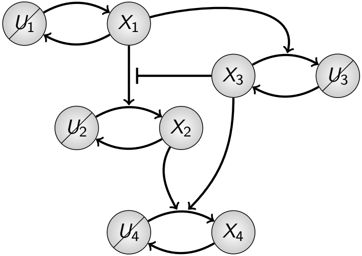
Here the equilibrium relations are $x_1 := 0$, $x_2 := -x_1$, so G_{equ} is faithful.

- More generally, for systems of the form

$$\begin{aligned} \dot{X}_1 &:= -X_1 + g_1(\mathbf{X}_{Pa(1)}, U_1) \\ &\vdots \\ \dot{X}_p &:= -X_p + g_p(\mathbf{X}_{Pa(p)}, U_p), \end{aligned}$$

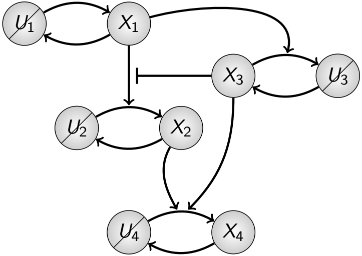
G_{equ} will be faithful to the equilibrium distribution.

A model for protein phosphorylation



U_i = concentration of unphosphorylated i
X_i = concentration of phosphorylated i

A model for protein phosphorylation



U_i = concentration of unphosphorylated i
X_i = concentration of phosphorylated i

The rate of phosphorylation of protein p depends on the concentration of enzymes X_E, $E \in \mathcal{E}$, and inhibitors X_I, $I \in \mathcal{I}_E$, of those enzymes:

$$\frac{dX_i}{dt} := -V_0 X_i + \sum_{E \in \mathcal{E}} \frac{V_E X_E U_i}{U_i + K_E \left(1 + \sum_{I \in \mathcal{I}_E} \frac{X_I}{K_I}\right)}$$

Such equations result in the equilibrium relations

$$x_i := \sum_{E \in \mathcal{E}} \frac{(V_E/V_0) x_E U_i}{U_i + K_E \left(1 + \sum_{I \in \mathcal{I}_E} \frac{x_I}{K_I}\right)}.$$

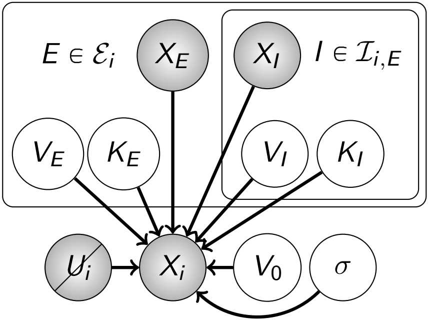
Such equations result in the equilibrium relations

$$x_i := \sum_{E \in \mathcal{E}} \frac{(V_E/V_0) x_E U_i}{U_i + K_E \left(1 + \sum_{I \in \mathcal{I}_E} \frac{x_I}{K_I}\right)}.$$

Introducing uncertainty, we arrive at the conditional probability model

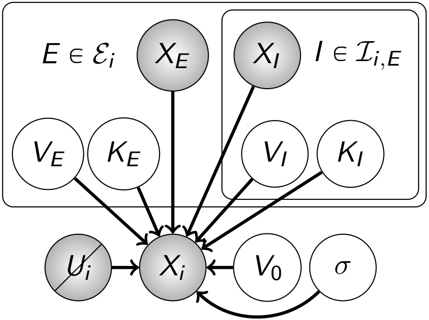
$$\log(x_i) | x_{(i)}, U_i, \boldsymbol{\theta}_i \sim N\left(\log\left(\sum_{E \in \mathcal{E}} \frac{(V_E/V_0) x_E U_i}{U_i + K_E \left(1 + \sum_{I \in \mathcal{I}_E} \frac{x_I}{K_I}\right)}\right), \sigma^2\right).$$

Inference by MCMC sampling



$$
\begin{aligned}
V|K,\sigma,M &\sim \text{Gamma}(2,1/2) \\
K|\sigma,M &\sim \text{Gamma}(2,1/2) \\
\sigma|M &\sim \text{InvGamma}(6,1) \\
M &\sim \text{Uniform}
\end{aligned}
$$

Inference by MCMC sampling



$$\begin{aligned} V \mid K, \sigma, M &\sim \text{Gamma}(2, 1/2) \\ K \mid \sigma, M &\sim \text{Gamma}(2, 1/2) \\ \sigma \mid M &\sim \text{InvGamma}(6, 1) \\ M &\sim \text{Uniform} \end{aligned}$$

A Reversible Jump sampler was constructed over the joint space of structure and parameters:

$$\mathcal{S} = \bigcup_{k \in \mathcal{K}} (\{k\} \times \Theta_k), \quad k = \bigtimes_{E \in \mathcal{E}^{M^{(k)}}} (\{E\} \times \mathcal{I}_E^{M^{(k)}})$$

Metropolis-Hastings proposal mechanisms:

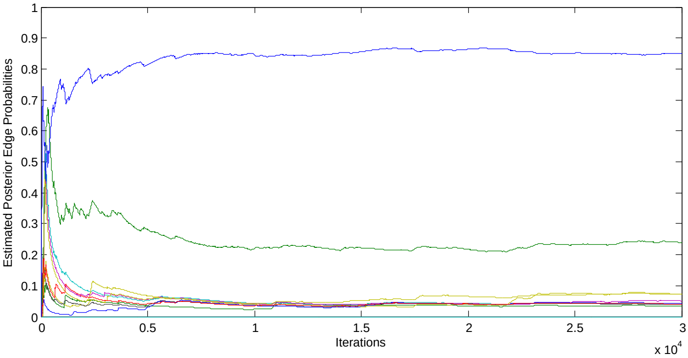
1. Update parameters V, K, σ
2. Update structure
   (a) Add/remove a kinase E to $\mathcal{E}_i$
   (b) Add/remove an inhibitor $I \in \mathcal{I}_{i,E}$
   (c) Swap a kinase
   (d) Swap an inhibitor

Move	Accept. Rate
1	55%
2(a)	7%
2(b)	10%
2(c)	7%
2(d)	9%

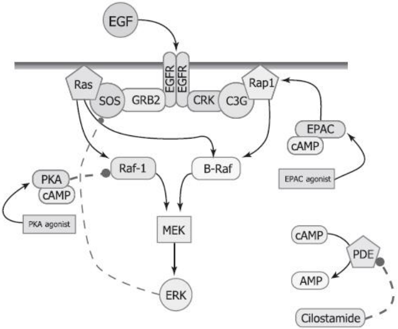
Metropolis-Hastings proposal mechanisms:

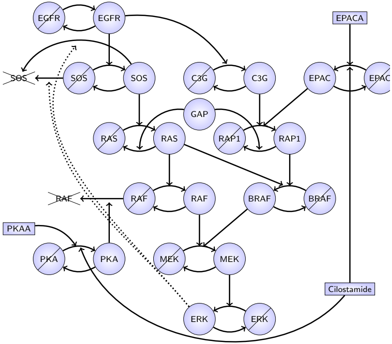
1. Update parameters V, K, σ
2. Update structure
   (a) Add/remove a kinase E to $\mathcal{E}_i$
   (b) Add/remove an inhibitor $I \in \mathcal{I}_{i,E}$
   (c) Swap a kinase
   (d) Swap an inhibitor

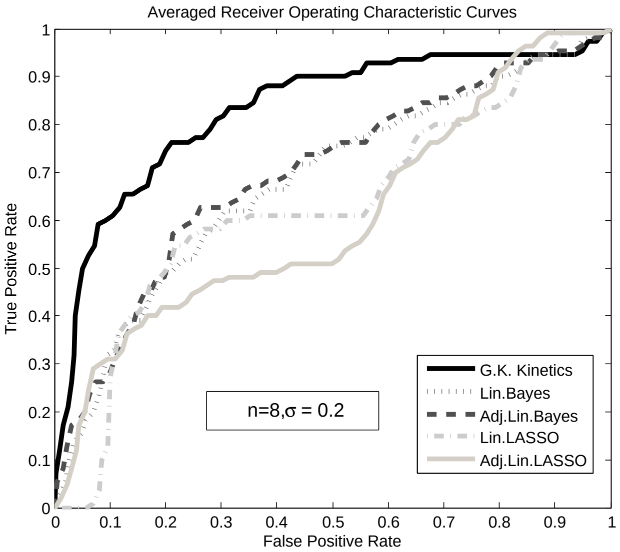
Move	Accept. Rate
1	55%
2(a)	7%
2(b)	10%
2(c)	7%
2(d)	9%



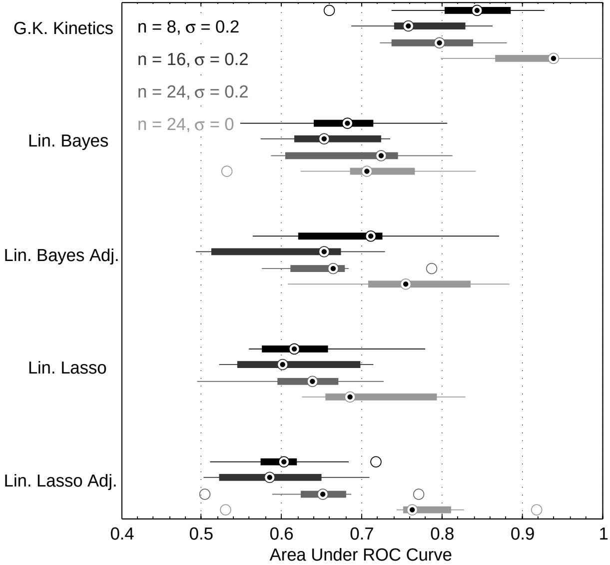
Empirical results: Simulation study (Xu *et al.*, 2010)



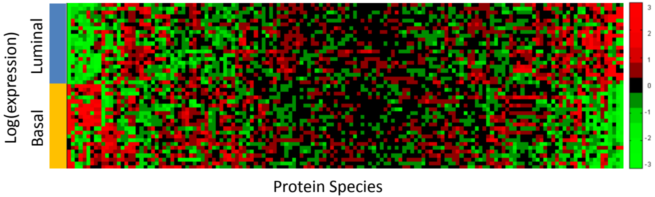




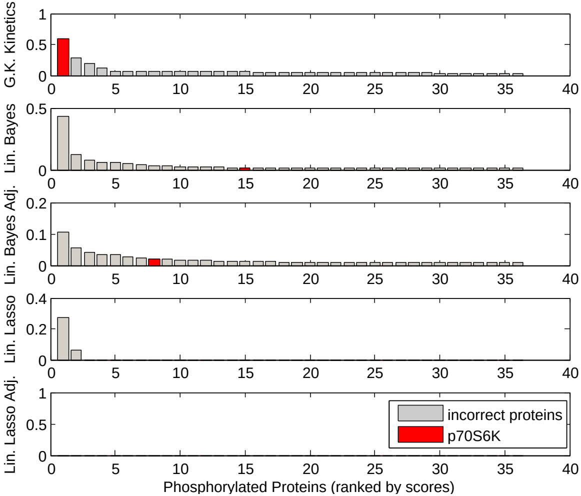
Averaged Receiver Operating Characteristic Curves
True Positive Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
n=8,σ = 0.2
G.K. Kinetics
Lin.Bayes
Adj.Lin.Bayes
Lin.LASSO
Adj.Lin.LASSO

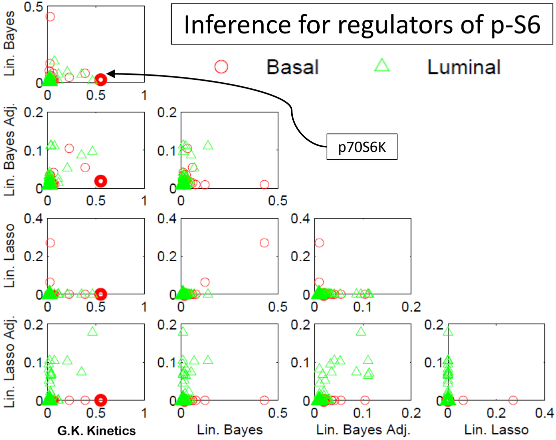

G.K. Kinetics
Lin. Bayes
Lin. Bayes Adj.
Lin. Lasso
Lin. Lasso Adj.
n = 8, σ = 0.2
n = 16, σ = 0.2
n = 24, σ = 0.2
n = 24, σ = 0
0.4
0.5
0.6
0.7
0.8
0.9
1
Area Under ROC Curve

Empirical results: Real data



Inference for p-S6




Inference for regulators of p-S6
Basal
Luminal
p70S6K
Lin. Bayes
Lin. Bayes Adj.
Lin. Lasso
Lin. Lasso Adj.
G.K. Kinetics
Lin. Bayes
Lin. Bayes Adj.
Lin. Lasso

Target:	Akt	p70S6K	S6	p53
G.K. Kinetics	**4**	**3**	**1**	**8**
Lin. Bayes	10	9	15	32
Lin. Bayes Adj.	14	8	8	14
Lin. Lasso	NA	8	NA	NA
Lin. Lasso Adj.	NA	12	NA	NA
Total # Candidates:	36	37	36	37

Summary

In this talk we have

Summary

In this talk we have

- seen theoretically how kinetics may be exploited to infer causes rather than correlations

Summary

In this talk we have

- seen theoretically how kinetics may be exploited to infer causes rather than correlations
- discussed why, for signalling, network structure ought to be identifiable from steady-state data

Summary

In this talk we have

- seen theoretically how kinetics may be exploited to infer causes rather than correlations
- discussed why, for signalling, network structure ought to be identifiable from steady-state data
- seen empirical evidence in favour of "kinetics-driven" inference.

Summary

In this talk we have

- seen theoretically how kinetics may be exploited to infer causes rather than correlations
- discussed why, for signalling, network structure ought to be identifiable from steady-state data
- seen empirical evidence in favour of "kinetics-driven" inference.

Interesting directions for future research include

Summary

In this talk we have

- seen theoretically how kinetics may be exploited to infer causes rather than correlations
- discussed why, for signalling, network structure ought to be identifiable from steady-state data
- seen empirical evidence in favour of “kinetics-driven” inference.

Interesting directions for future research include

- systematic experimental validation in the mammalian setting

Summary

In this talk we have

- seen theoretically how kinetics may be exploited to infer causes rather than correlations
- discussed why, for signalling, network structure ought to be identifiable from steady-state data
- seen empirical evidence in favour of "kinetics-driven" inference.

Interesting directions for future research include

- systematic experimental validation in the mammalian setting
- expanding the repertoire of kinetic equations.

References and Acknowledgments

1. Cho, H., Fryzlewicz, P. (2012) High dimensional variable selection via tilting, Journal of the Royal Statistical Society, Series B, to appear.
2. Dash, D. (2003) Caveats for Causal Reasoning with Equilibrium Models, PhD thesis, Intelligent Systems Program, University of Pittsburgh.
3. Heagerty, P.J., Kurland, B.F. (2001) Misspecified Maximum Likelihood Estimates and Generalised Linear Mixed Models, Biometrika, 88(4), 973-985.
4. Lv, J., Liu, J.S. (2010) Model Selection Principles in Misspecified Models, Technical Report, arXiv:1005.5483v1.
5. Pearl, J. (2009) Causal inference in statistics: An overview, Statistics Surveys, 3, 96-146.
6. Peters, J. et al. (2011) Identifiability of Causal Graphs using Functional Models, Proceedings of the 27th Annual Conference: Uncertainty in Artificial Intelligence (UAI-11), 589-598.
7. Xu, T. et al. (2010) Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species, Science Signaling, 3(113), ra20.

Financial support was provided by NCI CCSG support grant CA016672, NIH U54 CA112970, UK EPSRC EP/E501311/1 and the Cancer Systems Biology Center grant from the Netherlands Organisation for Scientific Research.

Sister paper:

Causal Variable Selection Using Equilibrium Relations from Nonlinear Dynamics, Workshop on Causal Structure Learning, UAI2012.