Probabilistic Graphical Models for Cellular Pathways

Florian Markowetz
florian.markowetz@molgen.mpg.de
Max Planck Institute for Molecular Genetics
Computational Diagnostics Group
Berlin, Germany

IPM workshop
Tehran, 2005 April

Cellular networks

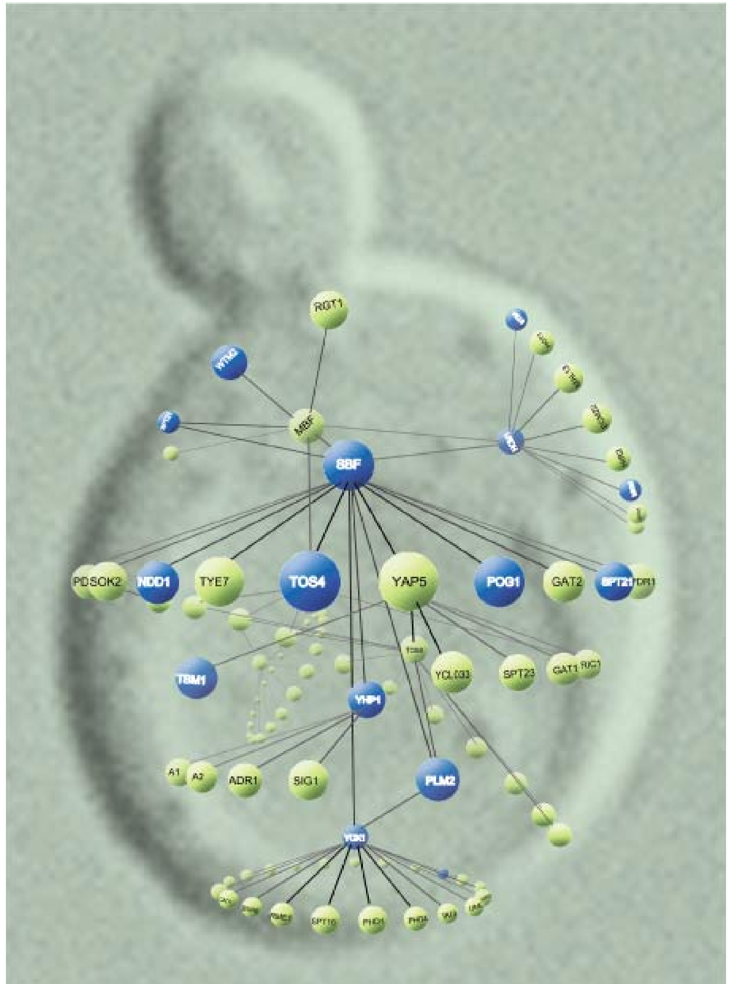
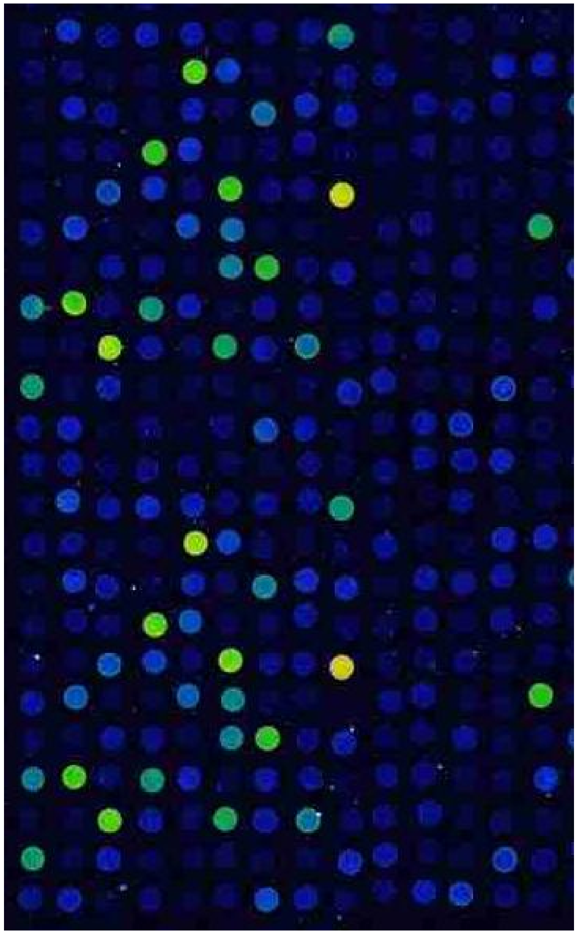


Figure from `http://array.mbb.yale.edu/yeast/transcription/`

Modelling networks



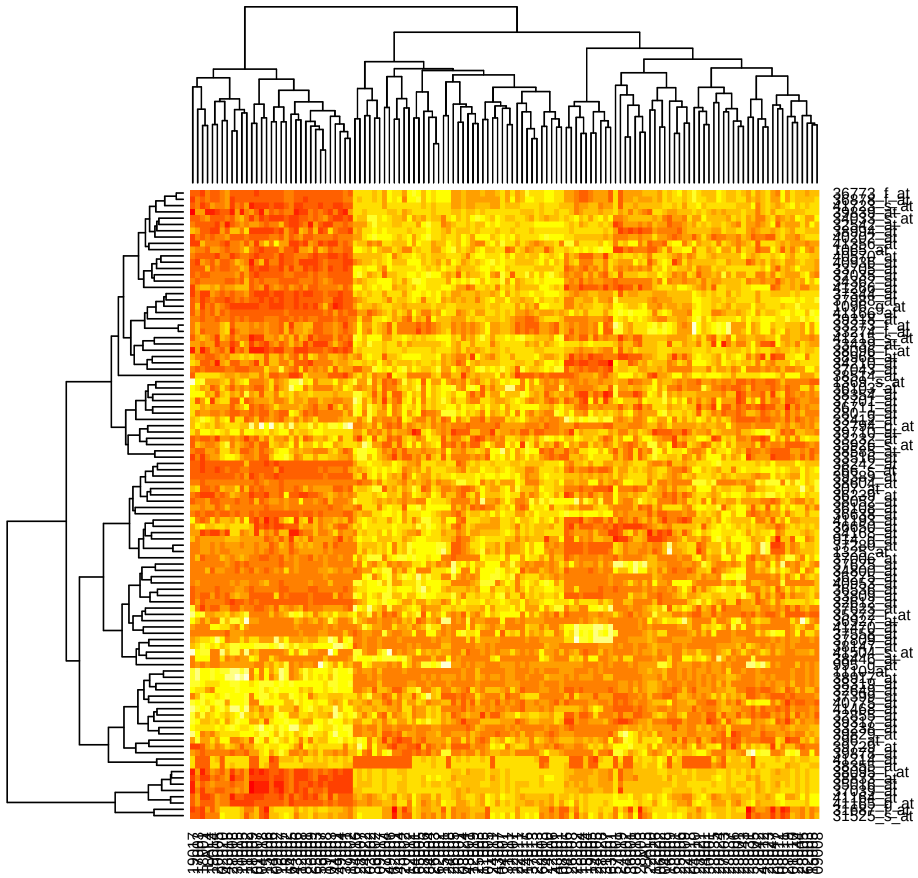
High-throughput assays can probe cells at a genome-wide scale.

Very prominent: **microarrays** that measure mRNA transcript quantitites.

Need to use **probabilistic models**, which account for

- measurement noise,
- variability in the biological system, and
- aspects of the system not captured by the model.

Clustering by coexpression



Assumption:

Coexpression $\sim$ coregulation

If genes show the same expression profiles they follow the same regulatory regimes [7, 25].

Correlation graphs

An expression profile is a random vector $\mathbf{X} = (X_1, \ldots, X_p)$.

Correlation graph: Depict genes as vertices of a graph and draw an edge (i, j) iff the correlation coefficient $\rho_{ij} \neq 0$.

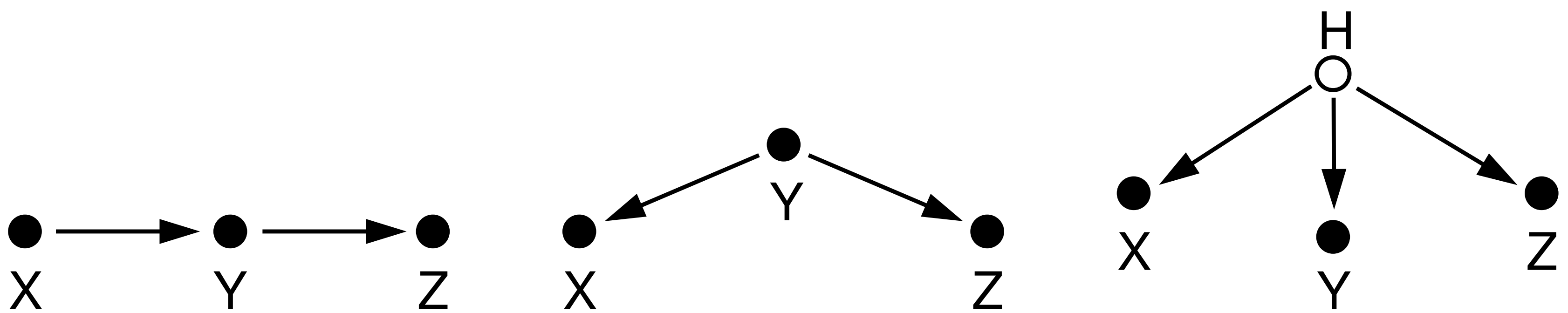
Advantage: This representation of the marginal dependence structure is easy to interpret and can be accurately estimated even if $p \gg N$.

Application: Stuart *et. al* [28] build a graph from coexpression across multiple organisms.

Problems of correlation based approaches

We cannot distinguish direct from indirect dependencies!

Three reasons, why X, Y, and Z are highly correlated:



As a cure:

search for correlations which cannot be explained by other variables.

Overview

1. Gaussian graphical models
- conditional independence
- partial correlations

2. Bayesian networks
- d-separation
- PC algorithm
- equivalence of networks

3. Bayesian structure learning
- marginal likelihood
- search strategies

Part I.

Gaussian graphical models

Conditional independence

Be X, Y, Z random variables with joint distribution P.

X is conditionally independent of Y given Z

$$X \perp\!\!\!\perp Y \mid Z \quad \Leftrightarrow$$

$$\begin{aligned}
P(X = x, Y = y | Z = z) &= P(X = x | Z = z) \cdot P(Y = y | Z = z) \\
P(X = x | Y = y, Z = z) &= P(X = x | Z = z)
\end{aligned}$$

Conditional independence: interpretation

Interpret random variables as **abstract pieces of knowledge** obtained from, say, reading books [16].

Then $X \perp\!\!\!\perp Y \mid Z$ means

Knowing Z, reading Y is irrelevant for reading X

If I already know Z,
then Y offers me no new information
to understand X.

Conditional independence in Gaussian models

- Consider a random vector $\mathbf{X} = (X_1, \ldots, X_p)$.
- Assume that $\mathbf{X} \sim \mathsf{N}(\mu, \Sigma)$, where Σ is regular.
- Let $K = \Sigma^{-1}$ be the *concentration matrix* of the distribution (aka *precision matrix*).

Then it holds for $i, j \in \{1, \ldots, p\}$ with $i \neq j$ that

$$\boxed{X_i \perp\!\!\!\perp X_j \mid \mathbf{X}_{\text{rest}} \;\Leftrightarrow\; k_{ij} = 0,}$$

where $\text{rest} = \{1, \ldots, p\} \setminus \{i, j\}$ [16].

Gaussian Graphical models (GGM)

Given a random vector $\mathbf{X} = (X_1, \ldots, X_p)$.

A Gaussian graphical model [16, 6] is an **undirected graph** on vertex set V, with $|V| = p$.

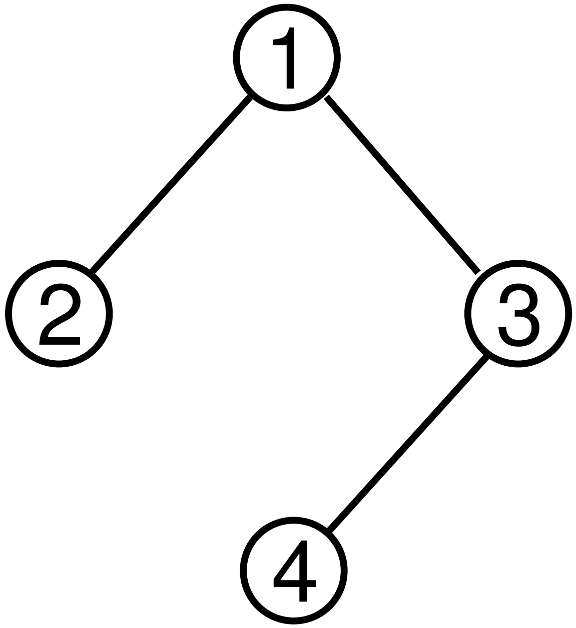
To each vertex $i \in V$ corresponds a **random variable** $X_i \in \mathbf{X}$.

Draw an **edge** between vertices i and j if and only if $k_{ij} \neq 0$.

Note:

In correlation graphs we modeled via Σ, in GGMs we use $K = \Sigma^{-1}$.

Example of a GGM



Missing edges indicate independencies:

$$X_i \perp\!\!\!\perp X_j \mid \mathbf{X}_{\text{rest}}$$

$$X_1 \perp\!\!\!\perp X_4 \mid \{X_2, X_3\}$$

$$X_2 \perp\!\!\!\perp X_3 \mid \{X_1, X_4\}$$

$$X_2 \perp\!\!\!\perp X_4 \mid \{X_1, X_3\}$$

Estimation from data

Likelihood

$$n(\mathbf{x}; K) \;=\; (2\pi)^{-\frac{p}{2}}\;|K|^{\frac{1}{2}}\;\exp\left\{-\frac{1}{2}\mathbf{x}^T K \mathbf{x}\right\}$$

Test **Null-Hypothesis** $k_{ij} = 0$ versus **Alternative** $k_{ij} \neq 0$.

- The Null-Hypothesis constrains the precision matrix K,
- the alternative leaves K unconstrained.

Likelihood ratio test statistic is asymptotically χ^2 distributed [16].

What if $p \gg N$?

Full conditional relationships can only be accurately estimated if the number of samples N is relatively large compared to the number of variables p.

Thus, if $p \gg N$, you can ...

either improve your estimators of partial correlations (*e.g.* Schäfer and Strimmer [23] use the Moore-Penrose pseudoinverse and bootstrap aggregation (bagging) to stabilize the estimator.)

or resort to a simpler model.

Sparse graphical Gaussian modeling

Do not condition on the complete rest as in GGMs. Instead explore dependency of **two variables conditioned on a third** [30, 31, 17, 5].

Draw an edge between vertices i and j $(i \neq j)$ if and only if the correlation coefficient

$$\rho_{ij} \neq 0$$

and no third variable can explain the correlation:

$$\boxed{X_i \not\perp\!\!\!\perp X_j \mid X_k \quad \text{for all } k \in \text{rest},}$$

whrere again rest $= \{1, \ldots, p\} \setminus \{i, j\}$.

Summary of part I

We have seen methods to build graphs from

1. marginal independencies

$$X_i \perp\!\!\!\perp X_j,$$

2. full conditional independence

$$X_i \perp\!\!\!\perp X_j \mid X_{\{1,\dots,p\}\setminus\{i,j\}},$$

3. first order independencies

$$X_i \perp\!\!\!\perp X_j \mid X_k \quad \forall k \in \{1,\dots,p\} \setminus \{i,j\}.$$

Summary of part I

We have seen methods to build graphs from

1. marginal independencies

$$X_i \perp\!\!\!\perp X_j,$$

2. full conditional independence

$$X_i \perp\!\!\!\perp X_j \mid X_{\{1,\ldots,p\}\setminus\{i,j\}},$$

3. first order independencies

$$X_i \perp\!\!\!\perp X_j \mid X_k \quad \forall k \in \{1,\ldots,p\} \setminus \{i,j\}.$$

Where does this lead us?

Include all higher order dependencies

Draw an edge between vertices i and j if

$$X_i \not\perp\!\!\!\perp X_j \mid \mathbf{X}_S \quad \text{for all } S \subseteq \{1, \ldots, p\} \setminus \{i, j\}.$$

This includes testing marginal, first order and full conditional independencies.

Include **all** higher order dependencies

Draw an edge between vertices i and j if

$$X_i \not\perp\!\!\!\perp X_j \mid \mathbf{X}_S \quad \text{for all } S \subseteq \{1, \ldots, p\} \setminus \{i, j\}.$$

This includes testing marginal, first order and full conditional independencies.

In the next part we will see:

- It will be possible to direct some of the edges.
- The resulting probabilistic model is a **Bayesian network**.
- Causation instead of just correlation [21, 26].

Part II.

Bayesian networks

Factorization of joint distribution

Given random vector $\mathbf{X} = (X_1, \ldots, X_p)$ we can always decompose

$$\begin{aligned} p(\mathbf{x}) &= p(x_1, \ldots, x_p) \\ &= p(x_1, \ldots, x_{p-1})\; p(x_p | x_1, \ldots, x_{p-1}) \end{aligned}$$

Factorization of joint distribution

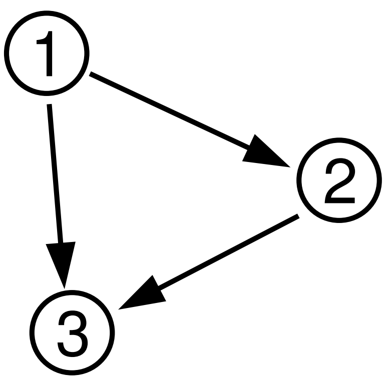
Given random vector $\mathbf{X} = (X_1, \ldots, X_p)$ we can always decompose

$$\begin{aligned} p(\mathbf{x}) &= p(x_1, \ldots, x_p) \\ &= p(x_1, \ldots, x_{p-1})\; p(x_p | x_1, \ldots, x_{p-1}) \\ &= p(x_1) \prod_{v=2}^{p} p(x_v | x_1, \ldots, x_{v-1}) \end{aligned}$$

Factorization of joint distribution

Given random vector $\mathbf{X} = (X_1, \ldots, X_p)$ we can always decompose

$$\begin{aligned} p(\mathbf{x}) &= p(x_1, \ldots, x_p) \\ &= p(x_1, \ldots, x_{p-1})\ p(x_p | x_1, \ldots, x_{p-1}) \\ &= p(x_1)\ \prod_{v=2}^{p} p(x_v | x_1, \ldots, x_{v-1}) \end{aligned}$$



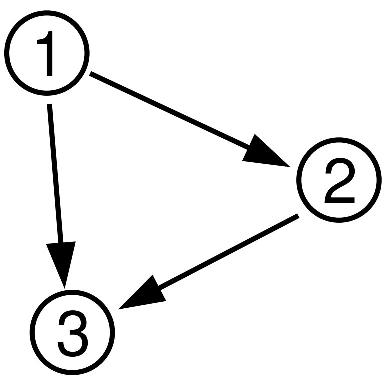
Example:

$$p(x_1, x_2, x_3) = p(x_1)\ p(x_2 | x_1)\ p(x_3 | x_1, x_2)$$

Factorization of joint distribution

Given random vector $\mathbf{X} = (X_1, \ldots, X_p)$ we can always decompose

$$\begin{aligned} p(\mathbf{x}) &= p(x_1, \ldots, x_p) \\ &= p(x_1, \ldots, x_{p-1})\ p(x_p | x_1, \ldots, x_{p-1}) \\ &= p(x_1) \prod_{v=2}^{p} p(x_v | x_1, \ldots, x_{v-1}) \end{aligned}$$



Example:

$p(x_1, x_2, x_3) = p(x_1)\ p(x_2|x_1)\ p(x_3|x_1, x_2)$

$\Rightarrow$ **completely connected directed acyclic graph**

Bayesian network

A Bayesian Network for a random vector $\mathbf{X}$ consists of

1. a **network structure**
   - directed acyclic graph (DAG) on vertex set V,
   - node v corresponds to variable X_v,

2. a set of **probability distributions**
   - locally: conditional distribution of a gene given its parents.
   - such that globally

$$p(\mathbf{x}) = \prod_{v \in V} p(x_v \mid \mathbf{x}_{pa(v)}, \theta_v)$$

Questions

1. How do the **local probability distributions** look like?
   $\longrightarrow$ Conditional Gaussian networks

Questions

1. How do the **local probability distributions** look like?
   $\longrightarrow$ Conditional Gaussian networks

2. How is **conditional independence** defined for directed models?
   $\longrightarrow$ Global Directed Markov Property

Questions

1. How do the **local probability distributions** look like?
$\longrightarrow$ Conditional Gaussian networks

2. How is **conditional independence** defined for directed models?
$\longrightarrow$ Global Directed Markov Property

3. How can we learn a Bayesian network **structure** from data?
$\longrightarrow$ Constraint-based algorithm (and a Bayesian in Part III)

Questions

1. How do the **local probability distributions** look like?
$\longrightarrow$ Conditional Gaussian networks

2. How is **conditional independence** defined for directed models?
$\longrightarrow$ Global Directed Markov Property

3. How can we learn a Bayesian network **structure** from data?
$\longrightarrow$ Constraint-based algorithm (and a Bayesian in Part III)

4. Are there **natural limits** in structure learning?
$\longrightarrow$ equivalence of network structures

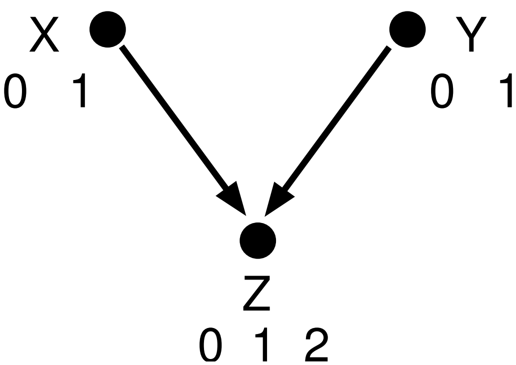
Children depend on parents

The DAG defines families.

Relationships are further characterized by local probability distributions:

Children depend on parents

The DAG defines families.
Relationships are further characterized by local probability distributions:



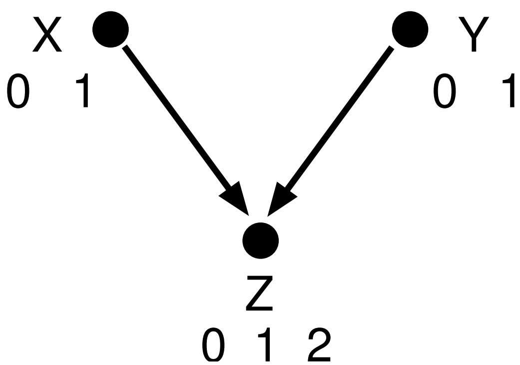
$$p(x) = (0.6 \quad 0.4)$$
$$p(y) = (0.2 \quad 0.8)$$

$$p(z|x,y) =$$

Children depend on parents

The DAG defines families.

Relationships are further characterized by local probability distributions:



$$p(x) = (0.6 \quad 0.4)$$

$$p(y) = (0.2 \quad 0.8)$$

$$p(z|x,y) = \begin{cases} (0.8 \quad 0.1 \quad 0.1) & \text{if } (X,Y) = (0,0) \\ (0.1 \quad 0.8 \quad 0.1) & \text{if } (X,Y) = (0,1) \\ (0.1 \quad 0.8 \quad 0.1) & \text{if } (X,Y) = (1,0) \\ (0.1 \quad 0.1 \quad 0.8) & \text{if } (X,Y) = (1,1) \end{cases}$$

Local probability distributions I

Discrete node with discrete parents

$$X_v \mid \mathbf{x}_{pa(v)}, \theta_v \;\sim\; \mathsf{Multin}(1, \theta_{v|\mathbf{x}_{pa(v)}})$$

Parametrization: $\theta_v = \{\theta_{v|\mathbf{x}_{pa(v)}}\}$ is a set of probability vectors – one for each configuration $\mathbf{x}_{pa(v)}$ of parents of v.

Density: [12]

$$p(x_v \mid \mathbf{x}_{pa(v)}, \theta_v) \;=\; \prod_{x'_v} \theta_{x'_v|\mathbf{x}_{pa(v)}}^{\mathbf{1}(x'_v = x_v)}$$

Local probability distributions II

Continous node with continuous parents

$$X_v \mid \mathbf{x}_{pa(v)}, \theta_v \;\sim\; \mathsf{N}(\mu_v, \sigma_v^2),$$

where $\mu_v = \beta_v^{(0)} + \sum_{i \in pa(v)} \beta_v^{(i)} x_i$.

Parametrization: $\theta_v = (\beta_v, \sigma_v^2)$ contains a vector of regression coefficients and a variance for node v.

Density:

$$p(x_v \mid \mathbf{x}_{pa(v)}, \theta_v) \;=\; \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_v - \mu_v)^2}{2\sigma_v^2}\right\}$$

Local probability distributions III

Continous node with mixed parents

Calling continous variables Y and discrete variables I [16], we can write

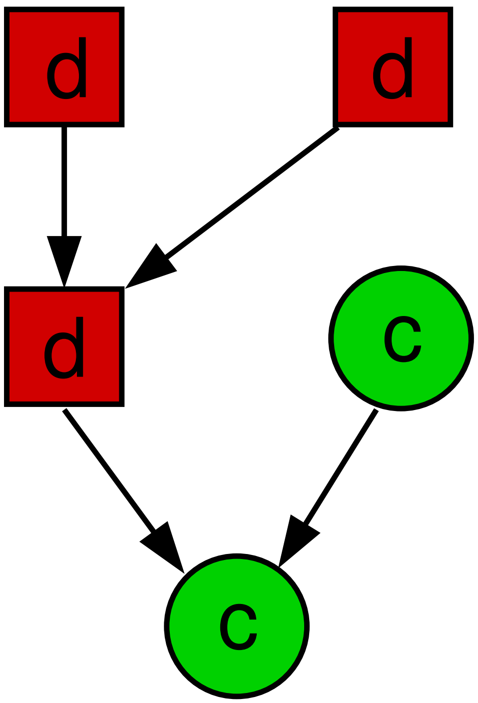
$$Y_v \mid \mathbf{i}_{pa(v)}, \mathbf{y}_{pa(v)}, \theta_v \;\sim\; \mathsf{N}(\mu_{v|\mathbf{i}_{pa(v)}}, \sigma^2_{v|\mathbf{i}_{pa(v)}}),$$

where $\mu_{v|\mathbf{i}_{pa(v)}} = \beta^{(0)}_{\mathbf{i}_{pa(v)}} + \sum_{i \in pa(v)} \beta^{(i)}_{\mathbf{i}_{pa(v)}} x_i$.

Parametrization: $\theta_v = (\beta_{v|\mathbf{i}_{pa(v)}}, \sigma^2_{v|\mathbf{i}_{pa(v)}})$ contains a vector of regression coefficients and a variance for node v, which depend on the state of the discrete parents [1].

Conditional Gaussian networks

We can combine the different LPDs in the framework of CG networks:

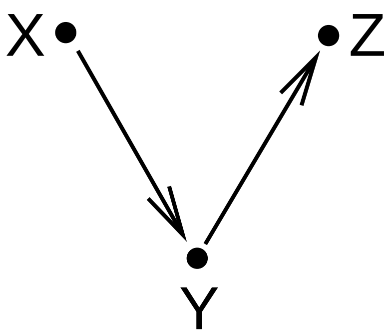


The random vector $\mathbf{X}$ has a discrete part $\mathbf{I}$ and a continuous part $\mathbf{Y}$ and the distribution decomposes as

$$p(\mathbf{x}) = p(\mathbf{i}, \mathbf{y}) = p(\mathbf{i})\, p(\mathbf{y}|\mathbf{i}).$$

These are the general parametric networks used in statistics [16].

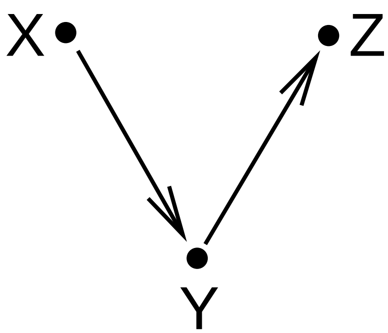
Conditional Independence I



Chain/linear

$$X \perp\!\!\!\perp Z \mid Y \quad \text{and} \quad X \not\!\perp\!\!\!\perp Z \mid \emptyset$$

Conditional Independence I

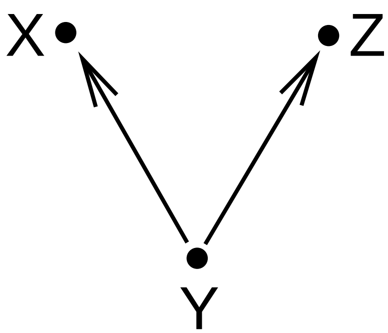


Chain/linear

$$X \perp\!\!\!\perp Z \mid Y \quad \text{and} \quad X \not\!\perp\!\!\!\perp Z \mid \emptyset$$

$$p(x,z|y) = \frac{p(x,y,z)}{p(y)} = \frac{p(x)\ p(y|x)\ p(z|y)}{p(y)} = p(x|y)\ p(z|y)$$

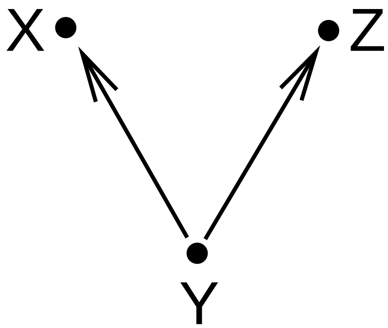
Conditional Independence II



Fork/diverging

$$X \perp\!\!\!\perp Z \mid Y \quad \text{and} \quad X \not\!\perp\!\!\!\perp Z \mid \emptyset$$

Conditional Independence II

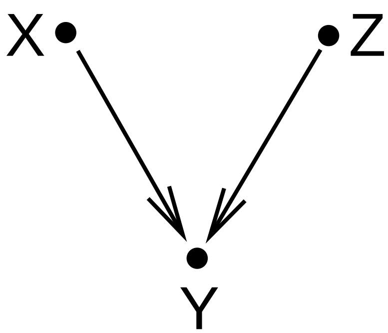


Fork/diverging

$$X \perp\!\!\!\perp Z \mid Y \quad \text{and} \quad X \not\perp\!\!\!\perp Z \mid \emptyset$$

$$p(x,z|y) = \frac{p(x,y,z)}{p(y)} = \frac{p(x|y)\ p(y)\ p(z|y)}{p(y)} = p(x|y)\ p(z|y)$$

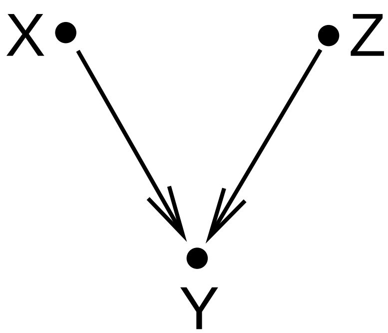
Conditional Independence III



Collider/converging

$$X \perp\!\!\!\perp Z \mid \emptyset \quad \text{and} \quad X \not\perp\!\!\!\perp Z \mid Y$$

Conditional Independence III



Collider/converging

$$X \perp\!\!\!\perp Z \mid \emptyset \quad \text{and} \quad X \not\perp\!\!\!\perp Z \mid Y$$

$$p(x,y,z) = p(x)\ p(y|x,z)\ p(z) = p(x)\ p(z)\ \frac{p(x,y,z)}{p(x,z)}$$

PC algorithm, part 1

How to find the skeleton of a Bayesian network [26, 21]

Form the complete undirected graph on node set $\{1, \ldots, p\}$. For each pair of variables X_i and X_j:

1. Remove the edge $i \sim j$ iff there exists a subset $S \subseteq \{1, \ldots, p\} \setminus \{i, j\}$ such that $X_i \perp\!\!\!\perp X_j \mid \mathbf{X}_S$.

2. Start with $S = \emptyset$, then continue for increasing $|S|$.

3. This includes testing marginal, first order and full conditional independencies.

PC algorithm, part 2

How to direct the edges [26, 21]

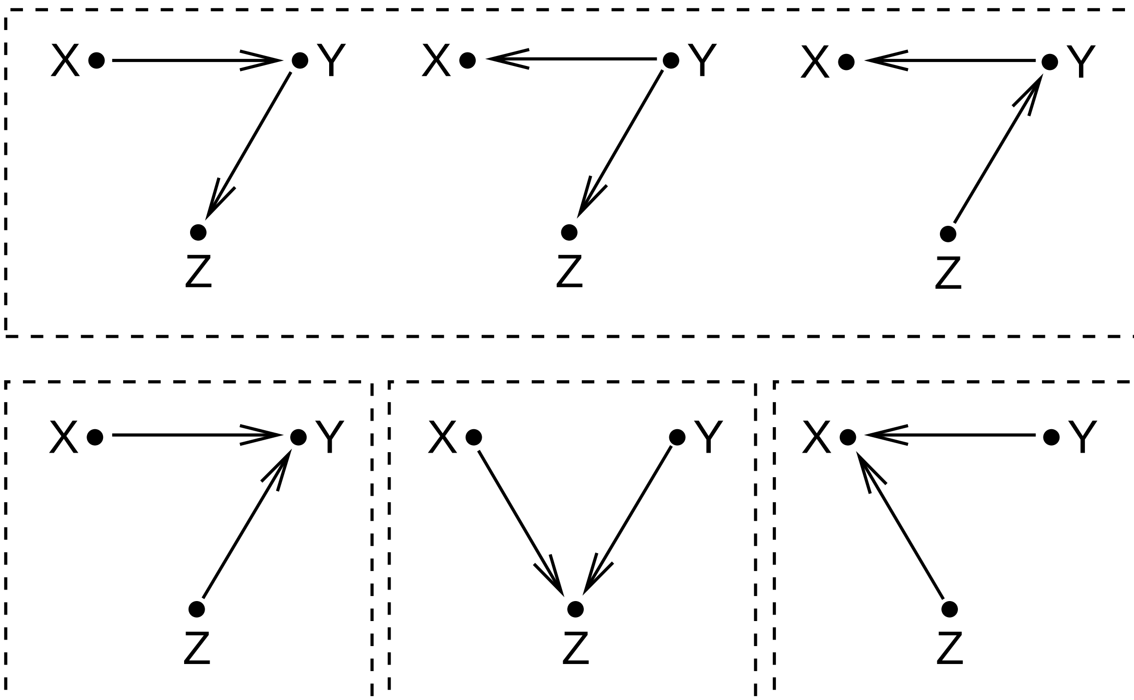
Once we have the skeleton, we can start putting directions on the edges.

First identify v-structures: Orient X—Y—Z into $X \longrightarrow Y \longleftarrow Z$ whenever $X \not\perp\!\!\!\perp Z \mid Y$.

Second direct as many edges as possible while respecting acyclicity and the independence constraints from step 1.

Equivalence of Networks

Two structures and are equivalent if both represent the same set of independence assertions.



Part III.

Bayesian structure learning

Situation

Model: We assume that the dependency structure of a random vector $\mathbf{X}$ follows an unknown DAG D.

The distribution $p(\mathbf{x})$ is Conditional Gaussian and factors according to D.

Data: We observe independent and identically distributed data $d = \{\mathbf{x}^1, \ldots, \mathbf{x}^N\}$. Each observation is a realization of $\mathbf{X}$.

Goal: Estimate D from d.

Being Bayesian about structure learning

1. **Score model**
   devise a scoring function that evaluates each network with respect to the training data.
2. **Search for best model**
   search for the optimal network according to this score.
3. **Assess model uncertainty**
   use MCMC or Bootstrap.

Scoring metric for networks

The **posterior distribution** of structure and parameters given data is

$$p(D, \theta \mid d) \;\propto\; p(d \mid D, \theta) \cdot p(\theta | D) \cdot p(D).$$

Scoring metric for networks

The **posterior distribution** of structure and parameters given data is

$$p(D, \theta \mid d) \;\propto\; p(d \mid D, \theta) \cdot p(\theta|D) \cdot p(D).$$

Integrating out **nuisance parameters** yields

$$p(D \mid d) \;\propto\; p(D) \cdot \int p(d \mid D, \theta)\; p(\theta|D)\; \mathrm{d}\theta.$$

Scoring metric for networks

The **posterior distribution** of structure and parameters given data is

$$p(D, \theta \mid d) \;\propto\; p(d \mid D, \theta) \cdot p(\theta|D) \cdot p(D).$$

Integrating out **nuisance parameters** yields

$$p(D \mid d) \;\propto\; p(D) \cdot \int p(d \mid D, \theta)\; p(\theta|D)\; \mathrm{d}\theta.$$

The righthand side will be our score for network fitness. It consists of a **structure prior** $p(D)$ and a **marginal likelihood** $p(d \mid D)$.

A local view of marginal likelihood

We zoom in on **one discrete family** of nodes with a **fixed configuration of parents**.

Assuming **parameter independence** [13] we will solve the integral

$$p(\text{batch} \mid D) = \int p(\text{batch} \mid D, \theta)\; p(\theta|D)\; \mathrm{d}\theta,$$

where “batch” means the part of data d corresponding to this one family.

To solve it analytically, we need priors, which **fit** to the likelihood.

Conjugate priors

Discrete part: Multinomial likelihood with *Dirichlet* prior:

$$p(\text{batch} \mid D, \theta) = \prod_k \theta_k^{n_k} \qquad\qquad p(\theta \mid D) = \frac{\Gamma(\alpha_+)}{\prod_k \Gamma(\alpha_k)} \prod_k \theta_k^{\alpha_k - 1}.$$

Mixed part: Gaussian likelihood with *Normal-inverse-*χ^2 prior.

Data likelihood is multivariate Normal, vector of regression β coefficients has Normal prior, variance σ^2 has inverse-χ^2 prior [1, 18].

Marginal likelihood of discrete family

$$
\begin{aligned}
p(\text{batch} \mid D) &= \int p(\text{batch} \mid D, \theta)\, p(\theta|D)\, \mathrm{d}\theta \\
&= \frac{\Gamma(\alpha_+)}{\prod_k \Gamma(\alpha_k)} \int_\Theta \prod_k \theta_k^{n_k+\alpha_k-1}\, \mathrm{d}\theta_v \\
&= \frac{\Gamma(\alpha_+)}{\prod_k \Gamma(\alpha_k)} \cdot \frac{\prod_k \Gamma(\alpha_k + n_k)}{\Gamma(\alpha_+ + n_+)}
\end{aligned}
$$

with counts n_k and Dirichlet parameters α_k. For the marginal likelihood of the complete network, you have to multiply terms like this for all nodes and all configurations of discrete parents [3, 13].

Where are we?

We learned in the case of discrete networks, how to compute the **marginal likelihood** $p(d \mid D)$. This is the right part of the score:

$$p(D \mid d) \;\propto\; p(D) \cdot \int p(d \mid D, \theta)\; p(\theta | D)\; \mathrm{d}\theta.$$

To complete the score, we need a **structure prior** $p(D)$.

And after that, we have to come up with a smart strategy to find high-scoring network structures.

Search for high scores

Exhaustive search: Infeasible for more than 5 nodes! [22]

If topological order of nodes is known

Start with empty network and iteratively add parents [3].

Hillclimbing (with random restarts)

- Start at randomly chosen network D.
- Score all neighbors (single edge deletions, insertions, inversions).
- Repeat for highest scoring neighbor.
- Runs into next local maximum.

Simulated annealing

Choose suboptimal neighbor with decreasing probability.

On true models

A quote from Edwards [6]:

"*Any method (or statistician) that takes a complex multivariate dataset and, from it, claims to identify one true model, is both naive and misleading.*"

What we have found is just a simple model **consistent with the data** — nothing more, nothing less.

Assessing uncertainty

Predicting **the best** network tells us nothing about the robustness of the solution.

MCMC: Use Markov Chain Monte Carlo to sample from the posterior distribution [14, 10].

Bootstrap: Computationally efficient approach to address confidence in network features [9, 11].

Biased-corrected bootstrap: Graphical models learned from bootstrap samples are biased towards too complex models. Steck and Jaakkola [27] suggest a bootstrap procedure corrected for this bias.

A caveat [8]

If the expression of gene A is regulated by proteins B and C, then A's expression level is a function of the joint activity levels of B and C. We treat the expression of A as a stochastic function of its regulators.

A caveat [8]

If the expression of gene A is regulated by proteins B and C, then A's expression level is a function of the joint activity levels of B and C. We treat the expression of A as a stochastic function of its regulators.

Problem 1: In most current biological data sets, however, we do not have access to measurements of protein activity levels.

A caveat [8]

If the expression of gene A is regulated by proteins B and C, then A's expression level is a function of the joint activity levels of B and C. We treat the expression of A as a stochastic function of its regulators.

Problem 1: In most current biological data sets, however, we do not have access to measurements of protein activity levels.

Resort: Expression levels of genes as a proxy for the activity level of the proteins they encode.

A caveat [8]

If the expression of gene A is regulated by proteins B and C, then A's expression level is a function of the joint activity levels of B and C. We treat the expression of A as a stochastic function of its regulators.

Problem 1: In most current biological data sets, however, we do not have access to measurements of protein activity levels.

Resort: Expression levels of genes as a proxy for the activity level of the proteins they encode.

Problem 2: There are numerous examples where an activation or silencing of a regulator is carried out by posttranscriptional protein modifications.

Books on Graphical models

1. Lauritzen: Graphical Models [16]
2. Edwards: Introduction to Graphical Modelling [6]
3. Pearl: Probabilistic Reasoning in Intelligent Systems [20]
4. Cowell *et al.*: Probabilistic Networks and Expert Systems [4]
5. Jordan: Learning in Graphical Models [15]

Software on Graphical models

1. BNT [19] `http://www.cs.ubc.ca/~murphyk/Software/BNT/bnt.html`
2. MGraph [29] `http://folk.uio.no/junbaiw/mgraph/mgraph.html`
3. PNL `https://sourceforge.net/projects/openpnl/`
4. GeneTS [23] `http://www.stat.uni-muenchen.de/~strimmer/genets/`
5. DEAL [2] `http://www.math.aau.dk/~dethlef/novo/deal/`
6. MIM [6] `http://www.hypergraph.dk/`
7. TETRAD [26] `http://www.phil.cmu.edu/projects/tetrad/`

Much more on `http://www.cs.ubc.ca/~murphyk/Software/BNT/bnsoft.html`.

Summary

1. Increasing order of resolution:
   Clustering, Graphical Gaussian models, Bayesian networks;

2. Central concept: Conditional independence;

3. Learning structure:
   Constraint-based approach and Bayesian scoring.

Summary

1. Increasing order of resolution:
   Clustering, Graphical Gaussian models, Bayesian networks;
2. Central concept: Conditional independence;
3. Learning structure:
   Constraint-based approach and Bayesian scoring.

Thank you! Questions?

References

[1] Susanne Gammelgaard Bøttcher. *Learning Bayesian Networks with Mixed Variables*. PhD thesis, Aalborg University, Denmark, 2004.

[2] Susanne Gammelgaard Bøttcher and Claus Dethlefsen. deal: A package for learning bayesian networks. *Journal of Statistical Software*, 8(20), 2003.

[3] Gregory F. Cooper and Edward Herskovits. A Bayesian Method for the Induction of Probabilistic Networks from Data. *Machine Learning*, 9:309–347, 1992.

[4] R.G. Cowell, A.P. Dawid, S.L. Lauritzen, and D.J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer-Verlag, New York, 1999.

[5] Alberto de la Fuente, Nan Bing, Ina Hoeschele, and Pedro Mendes. Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics*, 20(18):3565–3574, 2004.

[6] David Edwards. *Introduction to Graphical Modelling*. Springer, 2000.

[7] MB Eisen, PT Spellman, PO Brown, and D Botstein. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*, 95(25):14863–8, Dec 1998.

[8] Nir Friedman. Inferring Cellular Networks Using Probabilistic Graphical Models. *Science*, 303(5659):799–805, 2004.

[9] Nir Friedman, Moises Goldszmidt, and Abraham Wyner. Data analysis with Bayesian networks: A bootstrap approach. In *Uncertainty in Artificial Intelligence: Proceedings of the Fifteenth Conference (UAI-1999)*, pages 196–205, San Francisco, CA, 1999. Morgan Kaufmann Publishers.

[10] Nir Friedman and Daphne Koller. Being Bayesian about network structure: A Bayesian approach to structure discovery in Bayesian networks. *Machine Learning*, 50:95–126, 2003.

[11] Nir Friedman, Michal Linial, Iftach Nachman, and Dana Pe'er. Using Bayesian networks to analyze expression data. *Journal of Computational Biology*, 7(3):601–620, August 2000.

[12] A. Gelman, J. B. Carlin, H.S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall-CRC, 1996.

[13] David Heckerman, Dan Geiger, and David Maxwell Chickering. Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Machine Learning*, 20(3):197–243, Sep. 1995.

[14] Dirk Husmeier. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics*, 19(17):2271–2282, 2003.

[15] Michael I. Jordan, editor. *Learning in Graphical Models*. MIT Press, Cambridge, MA, 1999.

[16] Steffen L. Lauritzen. *Graphical Models*. Clarendon Press, Oxford, 1996.

[17] Paul M Magwene and Junhyong Kim. Estimating genomic coexpression networks using first-order conditional independence. *Genome Biol*, 5(12):R100, 2004.

[18] Florian Markowetz, Steffen Grossmann, and Rainer Spang. Probabilistic soft interventions in conditional gaussian networks. In Robert Cowell and Zoubin Ghahramani, editors, *Proc. Tenth International Workshop on Artificial Intelligence and Statistics*, Jan 2005.

[19] Kevin Murphy. The Bayes Net Toolbox for Matlab. *Computing Science and Statistics*, 33, 2001.

[20] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[21] Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge, 2000.

[22] Robert W. Robinson. Counting labeled acyclic digraphs. In F. Harary, editor, *New Directions in the Theory of Graphs*, pages 239–273. Academic Press, New York, 1973.

[23] Juliane Schfer and Korbinian Strimmer. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics*, 21(6):754–64, Mar 2005.

[24] Peter W. F. Smith and Joe Whittaker. Edge exclusion tests for graphical Gaussian models. In Michael Jordan, editor, *Learning in Graphical Models*, pages 555 – 574. MIT Press, 1999.

[25] PT Spellman, G Sherlock, MQ Zhang, VR Iyer, K Anders, MB Eisen, PO Brown, D Botstein, and B Futcher. Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell*, 9(12):3273–97, Dec 1998.

[26] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT Press, Cambridge, MA, second edition, 2000.

[27] Harald Steck and Tommi S. Jaakkola. Bias-corrected bootstrap and model uncertainty. In Sebastian Thrun, Lawrence Saul, and Bernhard Schölkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.

[28] Joshua M Stuart, Eran Segal, Daphne Koller, and Stuart K Kim. A gene-coexpression network for global discovery of conserved genetic modules. *Science*, 302(5643):249–55, Oct 2003.

[29] Junbai Wang, Ola Myklebost, and Eivind Hovig. Mgraph: graphical models for microarray data analysis. *Bioinformatics*, 19(17):2210–2211, 2003.

[30] Anja Wille and Peter Bühlmann. Tri-graph: a novel graphical model with application to genetic regulatory networks. Technical report, Seminar for Statistics, ETH Zrich, 2004.

[31] Anja Wille, Philip Zimmermann, Eva Vranová, Andreas Fürholz, Oliver Laule, Stefan Bleuler, Lars Hennig, Amela Prelic, Peter von Rohr, Lothar Thiele, Eckart Zitzler, Wilhelm Gruissem, and Peter Bühlmann. Sparse graphical Gaussian modeling of the isoprenoid gene network in Arabidopsis thaliana. *Genome Biol*, 5(11):R92, 2004.