# A SINful approach to Gaussian graphical model selection

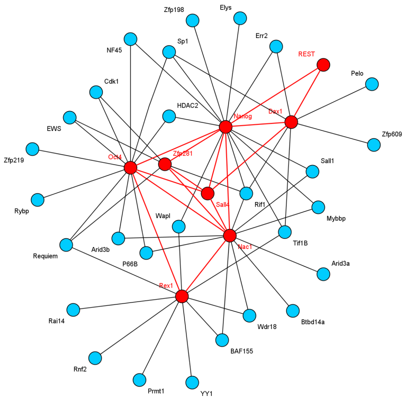
Mathias Drton and Michael Perlman

Adam Gustafson

Stat 572: Methodology Project
University of Washington

April 16, 2013

- In many data sets, characterizing the conditional independencies is of great interest.
- *Gene Regulatory Networks*: collection of DNA segments which interact with each other indirectly through RNA and protein expression; wish to learn network structure.




- Authors' methods: infer structure for Gaussian data.

# Graphical Models

- A *graphical model* is a probabilistic model in which conditional independencies are encoded via graphical properties.
- Formally, we have
  - A Graph $G = (V, E)$ where $V$ is set of nodes and $E \subset V \times V$ is a set of edges.
  - If $i, j \in V$, then $(i, j) \in E$ means that nodes $i$ and $j$ are connected (may or may not be directed edge).
  - One-to-one correspondence between nodes and a set of random variables.
  - A graphical model describes a family of distributions $p(x_V)$ over $X_V$.
  - A rule $r \in R$ is a predicate on a graph: $r(p, G) \in \{\text{true, false}\}$.
  - The set of distributions which a graphical model describes is

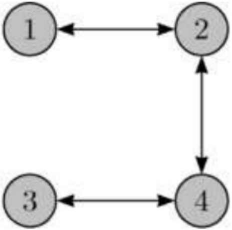
$$\mathcal{F}(G, R) = \{p : p \text{ is a distribution over } X_V \text{ and } r(p, G) = \text{true}, \forall r \in R\}$$

# Bidirected Graphs

- Encode *marginal* independence properties.
- The *pairwise bidirected Markov property*:

$$\mathcal{F}(G, R^{bg}) = \{p : X_u \perp\!\!\!\perp X_v \text{ for all non-adjacent pairs } u, v \in V(G)\}$$

- Example: $X_1 \perp\!\!\!\perp X_3$, $X_1 \perp\!\!\!\perp X_4$, $X_2 \perp\!\!\!\perp X_3$



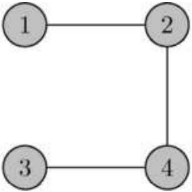

- *Gaussian Case*: $X_u \perp\!\!\!\perp X_v \iff \Sigma_{u,v} = 0$, where $\Sigma$ is the covariance matrix of $X_V$.

# Undirected Graphical Models

- Encode *conditional* independence properties.
- Authors use *pairwise Markov property* rule:

$$\mathcal{F}(G, R^p) = \{p : X_u \perp\!\!\!\perp X_v | X_{V \setminus \{u,v\}} \text{ for all non-adjacent pairs } u, v \in V(G)\}$$

- Ex: $X_1 \perp\!\!\!\perp X_4 | (X_2, X_3)$, $X_1 \perp\!\!\!\perp X_3 | (X_2, X_4)$, $X_2 \perp\!\!\!\perp X_3 | (X_1, X_4)$, etc.



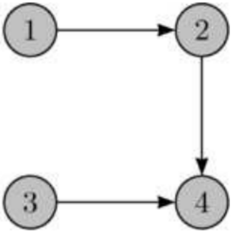

- *Gaussian Case*: $X_u \perp\!\!\!\perp X_v | X_{V \setminus \{u,v\}} \iff K_{uv} = 0$, where $K = \Sigma^{-1}$.

# Directed Acyclic Graphical Models

- Conditional independence properties for *directed acyclic graphs* (DAGs) can be stated in terms of conditioning on parents, or directed factorization:

$$\mathcal{F}(G, R^{df}) = \{p : p(x) = \prod_{v \in V} p(x_v | x_{pa(v)})\}$$

- Ex: $p(x_{1:4}) = p(x_1)p(x_2|x_1)p(x_4|x_2, x_3)p(x_3)$



# Directed Acyclic Graphical Models 2

- Create *partial order*: renumber nodes such that $u \preceq v$ if $u = v$ or there is a directed path $u \to \cdots \to v$. Not all nodes comparable.
- Create *total order*: *Well-number* partial order by setting $i \leq j$ whenever $i \preceq j$ – may need to renumber again. Not unique.
- The *well-numbered pairwise directed Markov property* states:

$$\mathcal{F}(G, R^{wn}) = \{p : V \text{ is well-numbered }, X_u \perp\!\!\!\perp X_v | X_{\{1,\ldots,v\}\setminus\{u,v\}} \iff \text{no directed path from } u \text{ to } v,\ u \leq v\}$$

- Equivalent to directed factorization, but allows identifiability of structure. Simplest Markov chain:

$$X_1 \to X_2 : p(x_{1:2}) = p(x_2|x_1)p(x_1) = p(x_1|x_2)p(x_2) \implies X_2 \to X_1$$

- Ex: $X_1 \perp\!\!\!\perp X_3 | X_2$, $X_2 \perp\!\!\!\perp X_3 | X_1$, $X_1 \perp\!\!\!\perp X_4 | (X_2, X_3)$.
- Ex: swapping nodes 2 and 3 leads to well-numbering.

# Chain Graphical Models

- For each $v \in V$, we now have a vector $U(v)$ of variables.
- First factorization is directed factorization over $U$; second factorization is clique factorization representation of undirected graphical model:

$$\mathcal{F}(G, R^{cg}) = \{p : p(x_U) = \prod_{v \in V} p(x_{U(v)} | x_{pa(v)})\}$$

where for each $v \in V$ we may write:

$$p(x_{U(v)} | x_{U(pa(v))}) \propto \prod_{c \in C(v)} \phi_c(x_c).$$

- Generalizes undirected and directed graphical models – more specificity.
- TODO: Two different versions of chain graphs: LWF and AMP. Reading paper on these rules.

# Gaussian Graphical Models

- Gaussian data can be viewed as any type of the preceeding graphical models.
- 0's in covariance matrix $\Sigma$ correspond to missing links in bidrected graphical model.
- 0's in the precision matrix $K = \Sigma^{-1}$ correspond to missing links in the undirected graphical model.
- Taking the Cholesky decomposition of $K$ yields $U'DU$, where $U$ is upper triangular with ones on the diagonal, and $D$ is diagonal.
- Directed factorization: Let $B = (I - U)$. $B_{ij} = 0$ for $j > i$ implies no edge from node $j$ to node $i$. Can well-number.
- Previous methods: backwards selection starting with full graph, controlling error at each stage.
- *Problem*: Overall error rate for false edge inclusion not controlled.

# Authors' Method: Asymptotics

- *Gaussian Data*: $n$ samples, $p$-dimensional likelihood:

$$L(\mu, K) \propto (\det K)^{n/2} \exp\{-\mathrm{tr}(KX'X/2) + \mu' KX' 1_p - n\mu' K\mu/2\}$$

- *Exponential Family*: canonical paramter $\theta = (K, K\mu)$, canonical statistic $T(X) = (-X'X/2, X'1_p)$.

$$L(\mu, K) \propto \exp\{\langle T(X), \theta\rangle - [n\theta' K^{-1}\theta/2 - (n/2)\log\det K]\},$$

  where $\langle (S_1, s_1), (S_2, s_2)\rangle = \mathrm{tr}(S_1 S_2) + s_1' s_2$.
- *Maximum Likelihood*: For $n > p$, we have

$$\widehat{\mu} = X' 1_p / n \qquad \widehat{\Sigma} = \frac{1}{n}(X - \bar{x}' \otimes 1_n)'(X - \bar{x}' \otimes 1_n)$$

- *Delta Method*: MLE asymptotically normal implies that $\widehat{K} = \widehat{\Sigma}^{-1}$ also asymptotically normal. Depends on unknown population covariance.

# Authors' Method: Model Selection

- *Rescale*: Sample partial correlations $r_{ij\cdot C(i,j)}$ asymptotically normal, and Fisher's $\mathcal{Z}$-transform improves asymptotics and is free of population partial correlations. $r_{ij\cdot C(i,j)}$ includes bias correction depending on conditioning set $C(i,j)$.

$$z_{ij\cdot C(i,j)} = \frac{1}{2}\log\left(\frac{1+r_{ij\cdot C(i,j)}}{1-r_{ij\cdot C(i,j)}}\right)$$

- *Model Selection*: $p(p-1)/2$ testing problems:

$$H_{ij} : \rho_{ij\cdot C(i,j)} = 0 \quad \text{vs.} \quad H_{ij} : \rho_{ij\cdot C(i,j)} \neq 0$$

  Note: $\mathcal{Z}(\rho_{ij}) = 0 \iff \rho_{ij} = 0$.

- *Bidirected Graphical Model*: $C(i,j) = \emptyset$.
- *Undirected Graphical Model*: $C(i,j) = V \setminus \{i,j\}$.
- *Directed Acyclic Graphical Model*: $C(i,j) = \{1,\ldots,j\}$.

# Authors' Method: Model Selection Continued

- Authors derive $p$-values which control overall error of false edge inclusion at level $\alpha$. Undirected case:

$$\pi_{ij} = 2\left(1 - \Phi(\sqrt{n_p}|z_{ij \cdot C(i,j)}|)\right)^{p(p-1)/2}$$

  Similar expressions for bidirected and DAG cases.

- Estimation procedure for estimating presence of edge $e_{ij}$:

$$\widehat{e}_{ij}(\alpha) = \begin{cases} 0, & \pi_{ij} \geq \alpha \\ 1, & \pi_{ij} < \alpha. \end{cases}$$

- Authors state estimation procedure is conservative: can be improved with Holm's step-down procedure to form adjusted $p$-values while controlling false edge inclusion at $\alpha$.

# Authors' Method: Guarantees

- $1 - \alpha$ 'consistency' result:

$$\liminf_{n\to\infty} \Pr(\widehat{G}(\alpha) = G_{\text{faithful}}) \geq 1 - \alpha$$

*Faithful* represents to distributions that are precisely encoded by the graph. In general, we have distributions which have additional conditional independencies than what is encoded by the graph.

- False edge inclusion result:

$$\limsup_{n\to\infty} \Pr(\widehat{G}(\alpha) \subseteq G)) \leq \alpha.$$

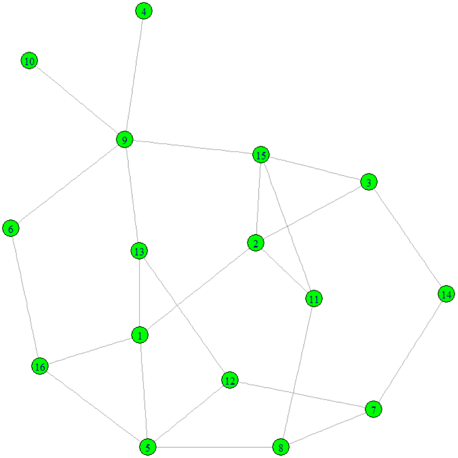
- Guarantees only hold asymptotically. For finite $n$, may have low power, may not control rate of false edge inclusion, or both.

# Experiment: Undirected case

- Vary $n$ and set $p = 16$ with fixed sparse population graph.
- $n \in 2^{5:10}$.
- Monotinically transform marginals to create *non-paranormal data* (a Gaussian copula) to test robustness.
- Show false positive and false negative rates.
- '1' is gaussian data. '2' is non-paranormal data.

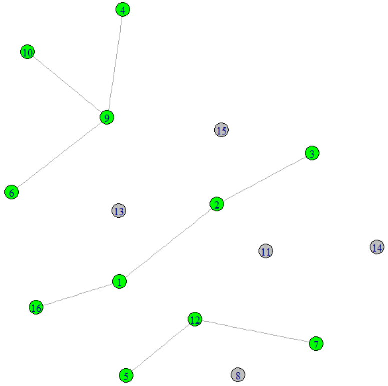
# True (Population) Graph

- $p = 16$ in this case.

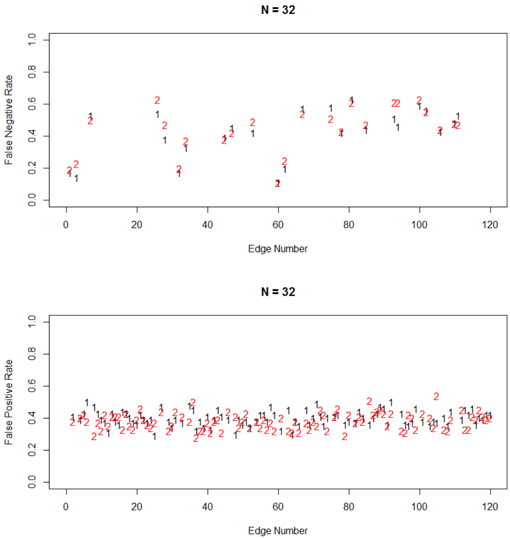


# Estimated Graph: $n = 128$

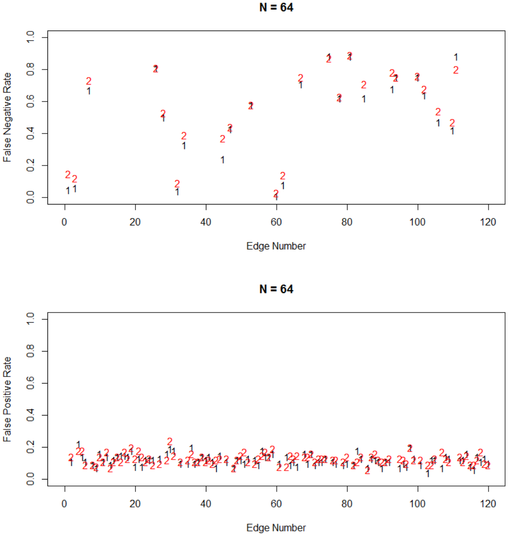
- Very conservative estimate for $\alpha = 0.05$.



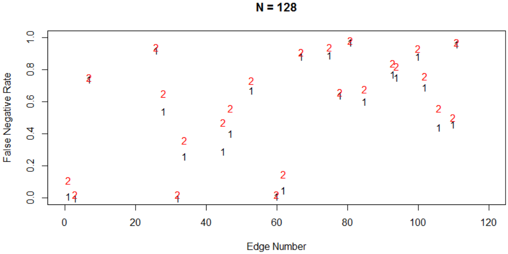
# Error Rates: $n = 32$

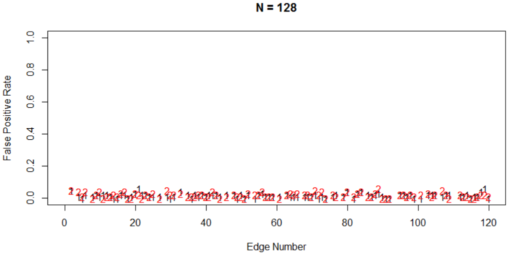


# Error Rates: $n = 64$

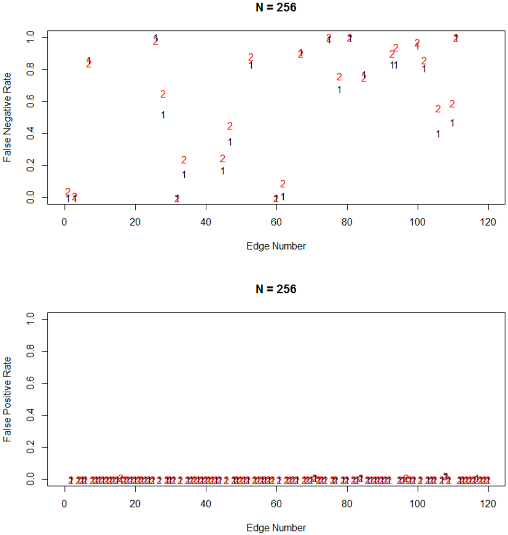


# Error Rates: $n = 128$

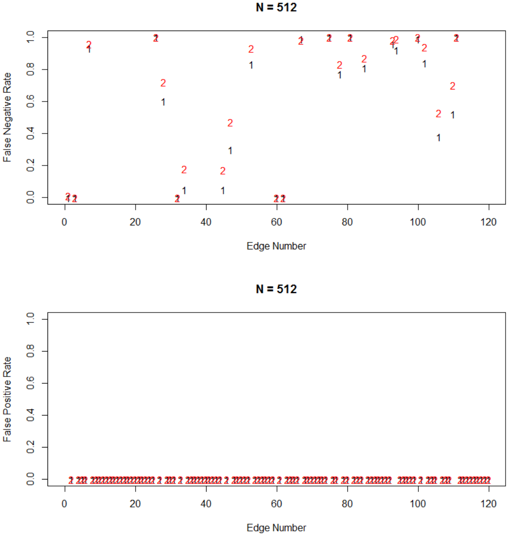


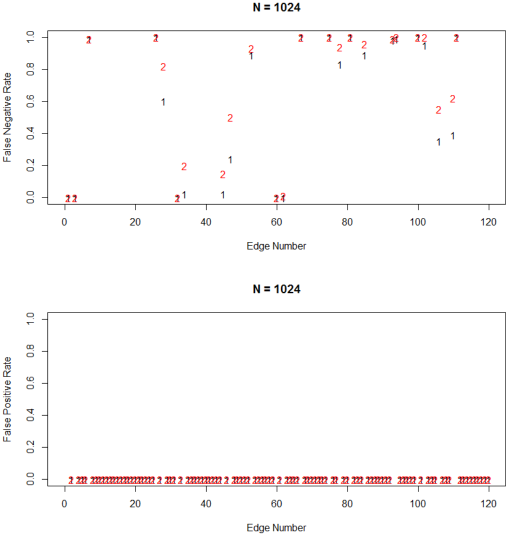
# Error Rates: $n = 256$



# Error Rates: $n = 512$



# Error Rates: $n = 1024$



# Comments and Conclusions

- Still have to do experiments for bidirected, directed, and chain graph cases.
- Method seems low power in terms of edge selection.
- Method does seem robust to non-gaussian data with same conditional indepenence structure.
- Method requires a priori knowledge of the well-numbering for directed and chain graphs – a *very* strong assumption.
- Method is simple in terms of derivation and implementation.