Uniform Inference in High-Dimensional Gaussian Graphical Models

Jannis Kück
jannis.kueck@uni-hamburg.de

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Setting

Let
$$X = (X_1, \ldots, X_p)^T \sim \mathcal{N}(\mu_X, \Sigma_X)$$
and assume that
$$X_j = \beta^{(j)} X_{-j} + \varepsilon^{(j)}.$$
It holds
$$\beta_k^{(j)} = 0 \Leftrightarrow (\Sigma_X^{-1})_{j,k} = 0 \Leftrightarrow X_j \perp X_k | X_{-\{j,k\}}.$$
The conditional independence structure of the distribution can be represented by a graph $G = (V, E)$, where
$$(j,k) \in E \Leftrightarrow \beta_k^{(j)} \neq 0.$$

Hypothesis

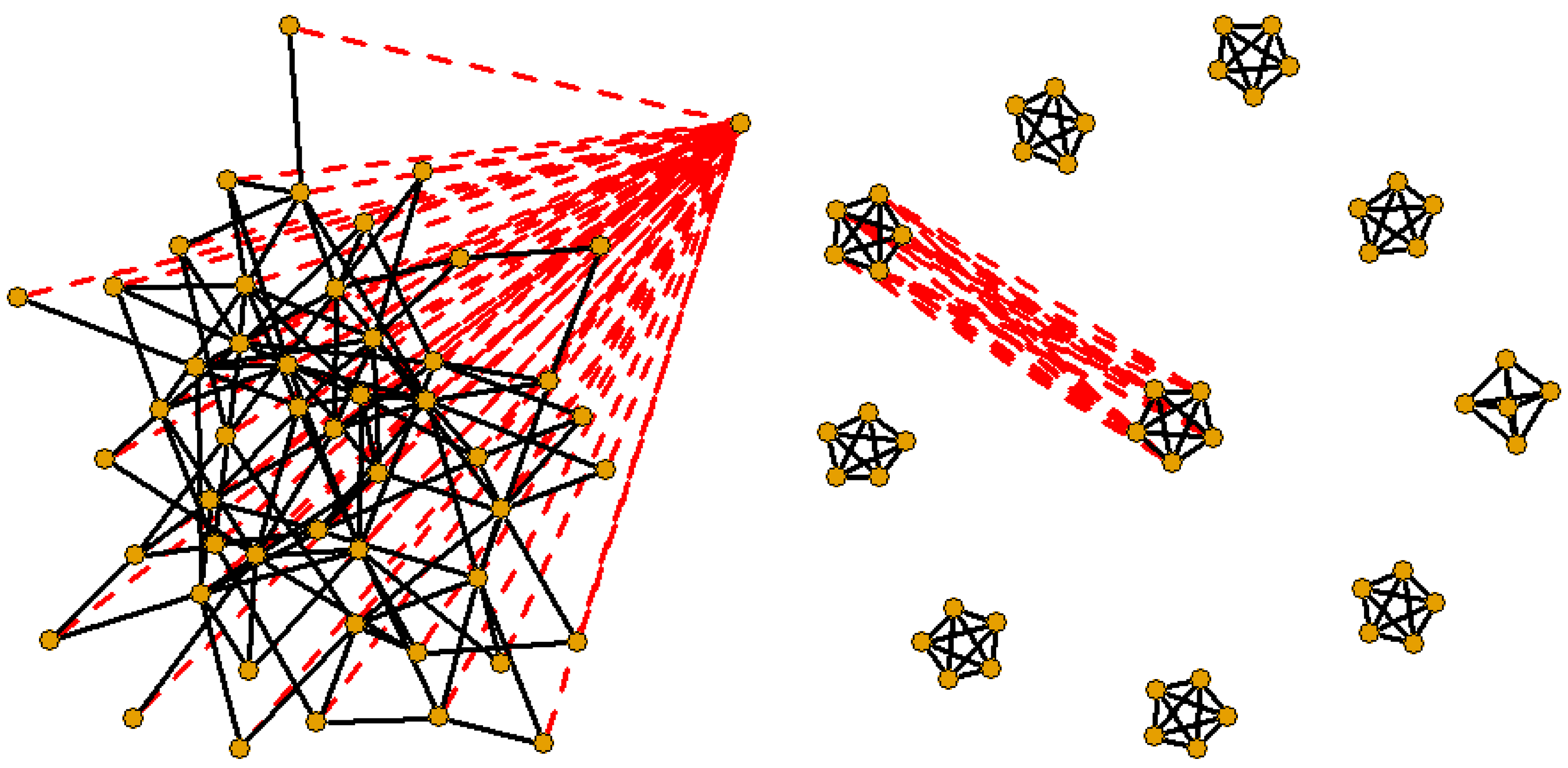
Let $\mathcal{M} = \{m_1, \ldots, m_{d_n}\}$ be a set of possible edges.
We want to test
$$H_0 : \mathcal{M} \cap E = \emptyset,$$
where the cardinality d_n of $\mathcal{M}$ is allowed to grow with sample size.

Assumptions

Sparsity:
For all $m_r = (j,k) \in \mathcal{M}$ with $j \neq k$ we have the following sparse representations
$$X_j = \beta^{(j)} X_{-j} + \varepsilon^{(j)} = \theta_{m_r} X_k + \beta^{(m_r)} X_{-m_r} + \varepsilon^{(m_r)}$$
with $\|\beta^{(m_r)}\|_0 \leq s$.
The sparsity assumption can be extended to approximate sparsity.
Restricted parameter space:
For all $(j,k) = m_r \in \mathcal{M}$ it holds
$$\|\beta^{(m_r)}\|_2 \leq C$$
and
$$\sup_{r=1,\ldots,d} \sup_{\theta \in \Theta_{m_r}} |\theta| \leq C$$
Additionally Θ_{m_r} contains a ball of radius $n^{-1/2} \log^{1/2}(d) \log(n)$ centered at θ_{m_r}.
Restricted correlations:
For all $(j,k) = m_r \in \mathcal{M}$ it holds
$$\inf_{\|\xi\|_2 = 1} \mathbb{E}\left[(\xi X)^2\right] \geq c \text{ and } \sup_{\|\xi\|_2=1} \mathbb{E}\left[(\xi X)^2\right] \leq C$$

Growth conditions:
Let $a_n := \max(p, n, d, e)$. There exists a positive number $\tilde{q} > 0$ such that the following growth condition is fulfilled:
$$n^{\frac{1}{\tilde{q}}} \frac{s \log^4(a_n)}{n} \left(\log^2(\log(a_n)) \vee s\right) = o(1)$$
and
$$\log(d) = o(n^{1/9} \wedge n^{\frac{2}{\tilde{q}}}).$$

Example



Simulations

	p	d	n	Coverage lasso	Coverage post-lasso	Coverage sqrt-lasso
Random graph	20	19	100	0.940	0.950	0.950
	50	49	100	0.965	0.975	0.950
	100	99	100	0.980	0.975	0.965
Cluster graph	20	25	100	0.980	0.985	0.975
	50	25	100	0.970	0.955	0.945
	100	25	100	0.985	0.980	0.985

All results are based on 200 independent simulations for $\alpha = 0.05$.

Estimation

The setting above fits in the general Z-estimation problem of the form
$$\mathbb{E}\left[\psi_{m_r}(X, \theta_{m_r}, \eta_{m_r})\right] = 0$$
with score functions
$$\psi_{m_r}(X, \theta, \eta) := \left(X_j - \theta X_k - \eta^{(1)} X_{-m_r}\right)\left(X_k - \eta^{(2)} X_{-m_r}\right)$$
for $m_r = (j_r, k_r) \equiv (j,k)$, $\eta = (\eta^{(1)}, \eta^{(2)})$ and $r = 1, \ldots, d_n$.
At first we estimate the nuisance parameter $\eta_{m_r} = \left(\eta_{m_r}^{(1)}, \eta_{m_r}^{(2)}\right)$ by running a lasso or post-lasso regression of X_j on X_{-j} to compute $(\tilde{\theta}_{m_r}, \hat{\eta}_{m_r}^{(1)})$ and a lasso or post-lasso regression of X_k on X_{-m_r} to compute $\hat{\eta}_{m_r}^{(2)}$ for each $(j,k) = m_r \in \mathcal{M}$.
The estimator $\hat{\theta}_0$ of the target parameter
$$\theta_0 = (\theta_{m_1}, \ldots, \theta_{m_{d_n}})^T$$
is defined as the solution of
$$\sup_{r=1,\ldots,d_n} \left\{ \left| \mathbb{E}_n\left[\psi_{m_r}(X, \hat{\theta}_{m_r}, \hat{\eta}_{m_r})\right] \right| - \inf_{\theta \in \Theta_{m_r}} \left| \mathbb{E}_n\left[\psi_{m_r}(X, \theta, \hat{\eta}_{m_r})\right] \right| \right\} \leq \epsilon_n,$$
where $\epsilon_n = o\left(\delta_n n^{-1/2}\right)$ is the numerical tolerance and $(\delta_n)_{n \geq 1}$ a sequence of positive constants converging to zero.

Main Result

Under the assumptions with probability $1 - o(1)$ uniformly in $P \in \mathcal{P}_n$ the estimator $\hat{\theta}$ obeys
$$P\left(\hat{\theta}_{m_r} - \frac{c_\alpha \hat{\sigma}_{m_r}}{\sqrt{n}} \leq \theta_{m_r} \leq \hat{\theta}_{m_r} + \frac{c_\alpha \hat{\sigma}_{m_r}}{\sqrt{n}}, r = 1, \ldots, d\right) \to 1 - \alpha.$$
Here c_α is an appropriate critical value, which is estimated using a multiplier bootstrap method.

Some References

[1] Alexandre Belloni, Victor Chernozhukov, Denis Chetverikov, and Ying Wei. Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework. arXiv preprint arXiv:1512.07619, 2015.

[2] Steffen L Lauritzen. Graphical models, volume 17. Clarendon Press, 1996.

[3] Sven Klaassen, Jannis Kueck, and Martin Spindler. Uniform inference in high-dimensional gaussian graphical models. Working Paper.