Significance Tests for Gaussian Graphical Models Based on Shrunken Densities

Bernal Arzola, Victor; Guryev, Victor; Bischoff, Rainer; Horvatovich, Peter; Grzegorczyk, Marco

Published in:
proceedings of the 33rd Inter- national Workshop on Statistical Modelling (IWSM), University of Bristol, UK, 16-20 July 2018

Significance Tests for Gaussian Graphical Models Based on Shrunken Densities

Victor Bernal [12], Victor Guryev[3], Rainer Bischoff[2], Peter Horvatovich[2], Marco Grzegorczyk [1]

[1] Johann Bernoulli Institute, University of Groningen, Groningen, NL.
[2] Department of Pharmacy, Analytical Biochemistry, University of Groningen, Groningen, NL.
[3] Universitair Medisch Centrum Groningen (UMCG), ERIBA, University of Groningen, Groningen, NL.

E-mail for correspondence: v.a.bernal.arzola@rug.nl

Abstract: Gaussian Graphical Models (GGMs) are important probabilistic graphical models in Statistics. Inferring a GGM's structure from data implies computing the inverse of the covariance matrix (i.e. the precision matrix). When the number of variables p is larger than the sample size n, the (sample) covariance estimator is not invertible and therefore another estimator is required. Covariance estimators based on shrinkage are more stable (and invertible), however, classical hypothesis testing for the "shrunk" coefficients is an open challenge. In this paper we present an exact null-density that naturally includes the shrinkage, and allows an accurate parametric significance test that is accurate and computationally efficient.

Keywords: Gaussian Graphical Models; Shrinkage; Genetic Networks, "small n, large " problem.

1 Introduction

Gaussian Graphical Models (GGMs) are important network models in Statistics. A GGM is represented as a network where each variable is a node, and an edge is present between a pair of nodes if their respective partial correlation is (statistically) significant. Partial correlations measure linear dependences between a pair of variables adjusted for all other nodes. Inferring the matrix of pair-wise partial correlations (i.e. the GGM's structure) demands the estimation of the covariance matrix $\hat{\mathbf{C}}$ and its inverse,

This paper was published as a part of the proceedings of the 33rd International Workshop on Statistical Modelling (IWSM), University of Bristol, UK, 16-20 July 2018.

therefore the importance that the covariance estimator is invertible, and well-conditioned (i.e. numerical errors are not magnified). The sample covariance estimator $\hat{\mathbf{C}}_{sm}$ with p variables and n samples is not invertible if $n \ll p$, thus another estimator must be employed. This is a common scenario in systems biology (e.g large number of genes with few measurements), and is usually refered to as the "small n, large p" problem, symbolically $n \ll p$.

Covariance estimators based on shrinkage produce a more stable estimator by using a (convex) linear combination of $\hat{\mathbf{C}}_{sm}$ with a target estimator $\mathbf{T}$ (e.g. a diagonal matrix). The result is a well-conditioned estimator, and its inverse can be used to compute the "shrunk" partial correlations. A significance test have been developed by Schäfer, J. and Strimmer, K. (2005) but it does not take the shinkage intensity into account. This is an open and important challenge as the reconstruction is a multiple testing problem (testing $\frac{p(p-1)}{2}$ edges), thus an slight bias would translate into an error repeated systematically during the inference. In this work we aim to obtain an exact density that includes the shrinkage effects. Our empirical results in Section 3 demonstrate that this leads to a substantial improvement over the earlier approach.

2 Shrinkage based Gaussian Graphical Models

Partial correlations are a measure of linear dependence between two variables adjusting the effects coming from all other variables. GGMs are undirected graphical models represented by a matrix of partial correlations. The matrix entry ρ_{ij} in a GGM represents the partial correlation between the variables i and j and can be computed from the inverse $\mathbf{C}^{-1}$ of the covariance matrix $\mathbf{C}$,

$$\rho_{ij} = -\frac{\mathbf{C}^{-1}_{ij}}{\sqrt[2]{\mathbf{C}^{-1}_{ii}}\sqrt[2]{\mathbf{C}^{-1}_{jj}}} \tag{1}$$

where $\mathbf{C}$ needs to be estimated from the data. However, when $n \leq p$ the sample covariance estimator $\hat{\mathbf{C}}_{sm}$ is ill-conditioned and cannot be used. Instead, the shrinkage based estimator $\hat{\mathbf{C}}^{[\lambda]}$ is a linear combination of $\hat{\mathbf{C}}_{sm}$ with a target matrix $\mathbf{T}$ in the form $\hat{\mathbf{C}}^{[\lambda]} = \lambda\mathbf{T} + (1-\lambda)\hat{\mathbf{C}}_{sm}$, where $\lambda \in [0, 1]$. The resulting estimator is well-conditioned, and is implemented in the widely used R package *GeneNet* (see. Schäfer, J. and Strimmer, K. (2005)) where λ is chosen following an optimization criteria. Moreover, significance is tested with the density of the standard partial correlation $f(\rho, k)$.

In the same way the correlation matrix $\mathbf{R}$ (i.e. the standarized $\hat{\mathbf{C}}_{sm}$) can be combined with (or shrunk towards) the identity matrix $\mathbf{I}$. In this case

the diagonal elements of $\mathbf{R}$ (i.e. the variances) remains equal to 1, and the off-diagonal r_{ij} (i.e. the pair-wise correlations coefficients) are scaled by a factor of $(1-\lambda)$. Therefore, the probability density function (pdf) of the "shrunk" correlation $r^{[\lambda]}$ can be found via the transformation $r^{[\lambda]} = (1-\lambda)r$,

$$f(r^{[\lambda]}, k) = \frac{[(1-\lambda)^2 - (r^{[\lambda]})^2]^{(\frac{k-3}{2})}}{Beta(\frac{1}{2}, \frac{k-1}{2})(1-\lambda)(1-\lambda)^{\frac{k-3}{2}}} \tag{2}$$

where k denotes the degrees of freedom. We now use a classical result from Fisher (1924) to obtain the probability density of the "shrunk" partial correlation $f(r^{[\lambda]}, k)$. Here it was proved that ρ and r have the same density differing only in the value of k. The main idea is to study the problem in *subject space* where each random variable is represented with a vector, and probabilistic relationships can be interpreted geometrically (see Wickens, T. D. (2014)). For the purpose of illustration, lets consider three random variables X, Y, and Z with expectation equal to zero (i.e. $E[X] = E[Y] = E[Z] = 0$). Given n data points for each variable, the corresponding random vectors $\vec{x}$, $\vec{y}$, and $\vec{z}$ are in an space of dimension n. The correlation between X and Y can be writen as

$$r = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{j=1}^{n} y_j^2}} = cos(\angle \vec{x}, \vec{y}) \tag{3}$$

where the last equality comes from the product $\vec{x} \cdot \vec{y} = ||\vec{x}||||\vec{y}||cos(\angle \vec{x}, \vec{y})$ under the Euclidean norm. The rationale behind the proof is that r is related to the angle between the vectors (see Eq 3), and that this angle is invariant under rotations of the coordinate axes. Therefore, the rotation can be peformed in such a way that one of the axis coincides with $\vec{z}$, and conditioning on the random variable Z is equivalent to decreasing k by one. This procedure can be continued by rotating again, and conditioning over a new variable. The same idea can be used for $r^{[\lambda]}$ (as it is a scaled correlation), and$f(\rho^{[\lambda]}, k)$ is the probability density of $\rho^{[\lambda]}$.

To test the null hypothesis H_0 : (the "shrunk" partial correlation is zero) with $f(\rho^{[\lambda]}, k)$ we propose the following approach: Suppose the data D consists of p variables and sample size n.

1. For D find the optimal shrinkage λ_{opt}, and estimate $\rho_{ij}^{[\lambda_{opt}]}$ (Schäfer, J. and Strimmer, K. (2005)).
2. Estimate k under H_0:
   (a) Simulate data of length n from H_0 (i.e. the precision matrix is the p x p identity), and using λ_{opt} (from step 1) infer the null-hypothetic coefficients $\rho_{0_{ij}}^{[\lambda_{opt}]}$.
   (b) Find $\hat{k}$ by maximizing the likelihood of the $\rho_{0_{ij}}^{[\lambda_{opt}]}$ with Eq 2.

3. Test the coefficients $\rho_{ij}^{[\lambda_{opt}]}$ from step 1 with $f(\rho_{ij}^{[\lambda_{opt}]}, \hat{k})$.

We will refer to this approach as "Shrunk MLE" in the following section.

3 Results

In this section we provide empirical evidence that the proposed "Shrunk MLE" approach is significantly superior to *GeneNet* 1.2.13. We cross-compare the methods on synthetic, and real gene expression data by testing the null hypothesis H_0 : (the "shrunk" partial correlation is zero). The Positive Predictive Values (PPVs) are compared on (i) syntethic data were the true structure is known, and (ii) on real data were we use MC (a computationally expensive approach) to generate a reliable goldstandard network. To simulate GGMs with a fixed percentage of edges δ we used *GeneNet* (for the algorithm see Schäfer, J. and Strimmer, K. (2005)). Figure 1 shows the $PPV = \frac{TP}{(TP+FP)}$ for different samples sizes n.

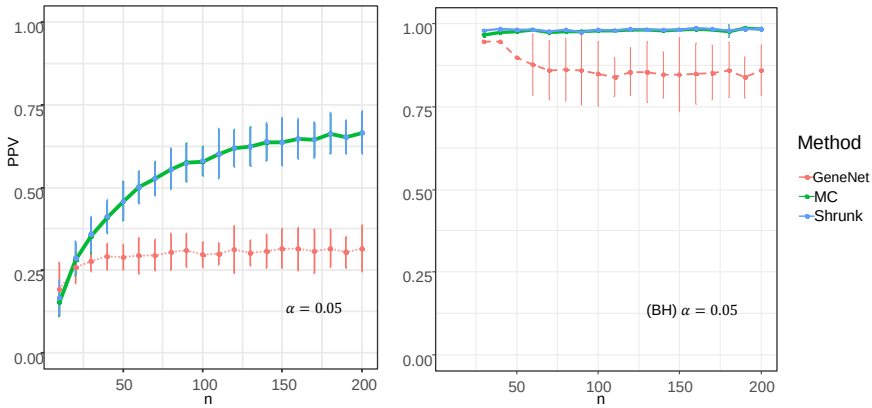

FIGURE 1. **Positive predictive value.** *On the left*: GGM simulation for $p = 100$, and n ranging from 10 to 200 in steps of size 10. The simulated GGM structure has 148 correlations (i.e. $\delta = 0.03$). The Positive predictive values (PPV) are computed using p-values at $\alpha = 0.05$. Dots (and bars) represent the average PPV ($\pm$ 2 standard errors) over 25 repeated simulations, and MC was performed 15 times. Three methods are displayed:*GeneNet* (in dashed red), MC (green), and Shrunk MLE (thick blue). Note that the green and blue curves are superposed.*On the left*: The PPV are computed using Benjamini Hochberg adjusted p-values at $\alpha = 0.05$.

The results show a close agreement between the PPV obtained by MC, and with "Shrunk MLE". On the other hand, *GeneNet* has a lower PPV suggesting that it learns too many False Positives (FPs) Figure 2. We also

analyzed Escherichia Coli microarray data from Schmidt-Heck, W. et al. (2004), consisting of stress temporal response of 102 genes in 9 time points after IPTG (induction of the recombinant protein SOD). Figure 2 shows a Venn diagram for the edges found by each method, here we can observe that "Shrunk MLE" learns nearly the same edges as MC.

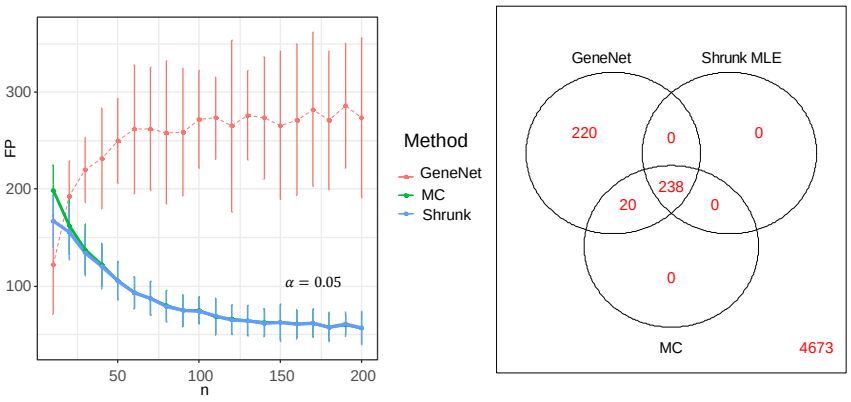

FIGURE 2. **False positives and Empirical results.** *On the left*: GGM simulation for $p = 100$, and n ranging from 10 to 200 in steps of size 10. The simulated GGM structure has 148 correlations (i.e. $\delta = 0.03$). The False Positives (FPs) are computed using p-values at $\alpha = 0.05$. Dots (and bars) represent the average FPs ($\pm$ 2 standard errors) over 25 repeated simulations, and MC was performed 15 times. Three methods are displayed: *GeneNet* (in dashed red), MC (green), and Shrunk MLE (thick blue). Note that the green and blue curves are superposed. *On the right*: Venn diagram for the inferred edges in *E. coli*. Taking MC as a gold standard *GeneNet*'s sensitivity is 258/258=1, with a low PPV of 258/478 $\approx$ 0.54. Shrunk MLE has a slightly decreased sensitivity of 238/258 $\approx$ 0.92, but yields a perfect PPV of 1.

A GO enrichment analysis (http://geneontology.org/) with False Discovery Rate (FDR$<$ 0.05) shows that the connected genes belong significantly to stress response (FDR$=$ $2.02E^{-}02$), in contrast with *GeneNet* (FDR=$7.73E^{-}02$). This suggests a dillution of the GO's significance due higher rate of FPs. The strongest connections were lacA–lacZ, lacY– lacZ, and lacA–lacY related to the lac operon (known to be triggered by IPTG).

4 Conclusions

Gaussian Graphical Models (GGMs) are an important tool for network learning. Reconstructing the network demands the estimation of the covariance matrix, which is ill-conditioned if the sample size is smaller than the number of variables. Covariance estimators based in shrinkage make the

covariance matrix invertible, however, for an accurate (parametric) significanc tests the shrinkage value needs to be included, otherwise the inference will have a systematic error (e.g. biased p-values). In this paper a new shrunk density was introduced, and to our knowledge is the only test that includes the regularization effects. In the "small n, large p" scenario the new density allows an accurate inference for any shrinkage value.

References

Fisher, R. (1924). The Distribution of the Partial Correlation Coefficient. *Metron*, **3**, 329 – 332.

Schäfer, J. and Strimmer, K. (2005). A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Funtional Genomics. *Statistical Applications in Genetics and Molecular Biology*, **4**, 1 – 30.

Schmidt-Heck, W. et al. (2004). Reverse Engineering of the Stress Response during Expression of a Recombinant Protein. *Proceedings of the EUNITE symposium*, Aachen, 10 – 12.

Wickens, T. D. (2014). *The Geometry of Multivariate Statistics.* Psychology Press.