Gaussian Graphical Models

Jami Jackson Mulgrave
April 7th, 2017

What is a Graph?

- A graph is a mathematical object defined as a pair $G = (V, E)$, where V is a set of vertices or nodes and E is a set of edges.
- Each edge is associated with a pair of nodes, its endpoints. Edges may be directed, undirected, or bidirected.
- Nodes are represented by circles or points, and edges by lines, arrows, or bidirected arrows.
- We use the notation $\alpha - \beta$ to denote undirected edges between α and β.

Undirected Graphs

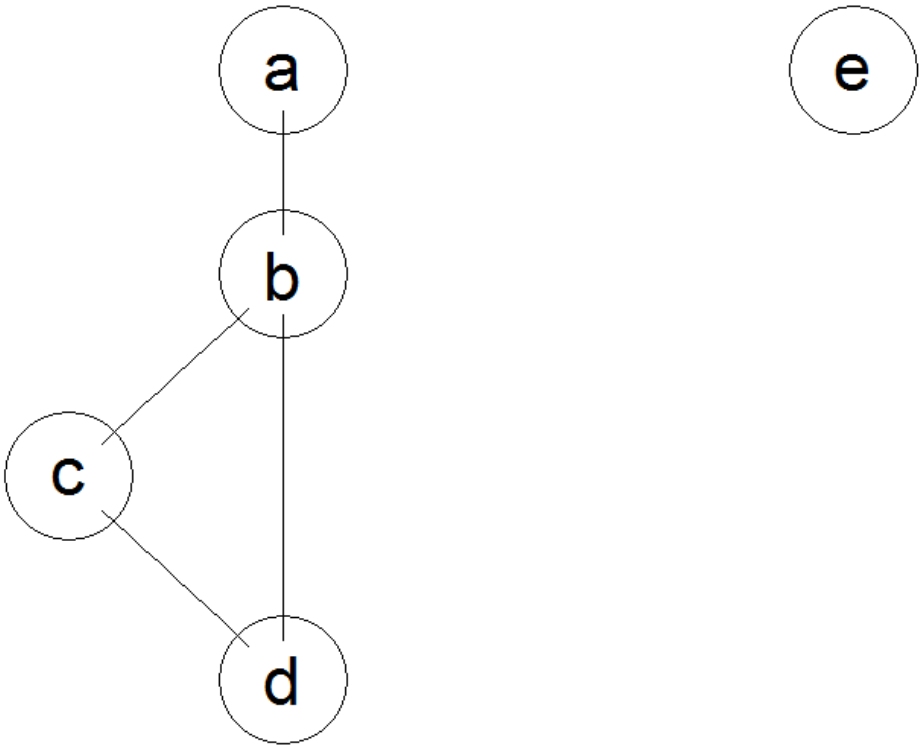
- Packages **gRbase** and **Rgraphviz** are very useful for graphical modelling
- An undirected graph may be created using the ug() function.

```
library(gRbase)
library(RBGL)


ug0 <- ug(c("a","b"),c("b","c","d"),"e")
```

Undirected Graphs

```
plot(ugo)
```



Complete Graphs

- A subset $A \subseteq V$ is complete if all vertex pairs in A are connected by an edge.
- A graph $G = (V, E)$ is complete if the vertex set V is complete.
- A clique is a maximal complete subset, that is to say, a complete subset that is not contained in a larger complete subset.
- The set of cliques of a graph G is denoted by $C(G)$.

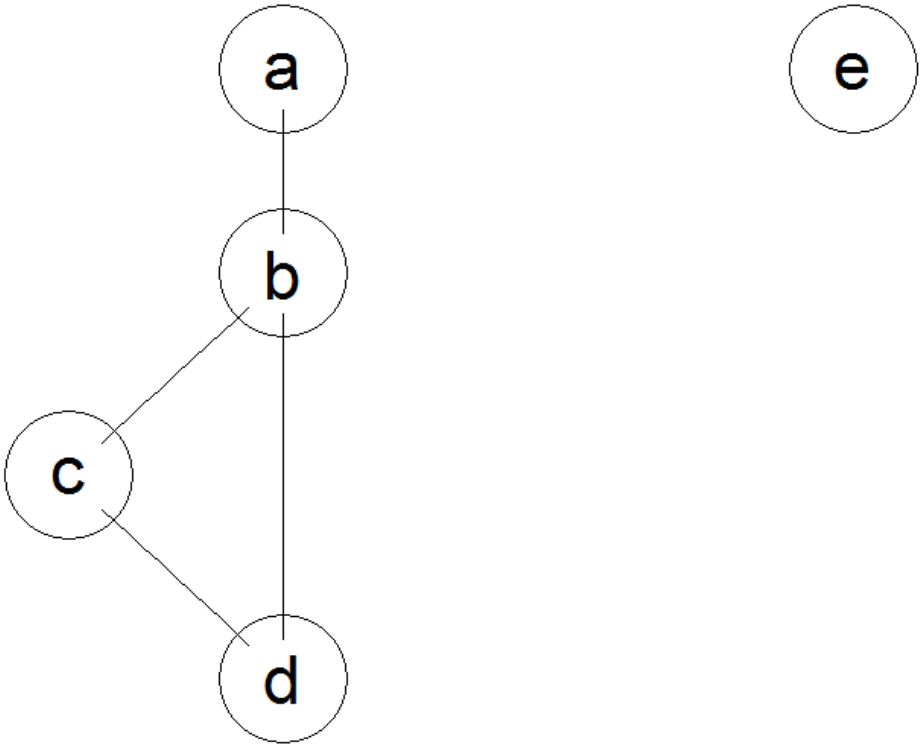
s our Undirected Graph Complete?

```
is co plete(ug0)
```

```
      L
```

```
is co plete(ug0, c("b","c","d"))
```

```
     R
```


a
e
b
c
d

What are the ma imal cli ues?

```
a  li ue(ug0)
```

```
 a  li ues
 a  li ues
   "b" "c" "d"

 a  li ues
   "b" "a"

 a  li ues
   "e"
```

eparation

- A *path* (of length n) between α and β in an undirected graph is a set of vertices $\alpha = \alpha_0, \alpha_1, \ldots, \alpha_n = \beta$ where $\alpha_{i-1} - \alpha_i$ for $i = 1, \ldots, n$.
- If a path $\alpha = \alpha_0, \alpha_1, \ldots, \alpha_n = \beta$ has $\alpha = \beta$ then the path is said to be a cycle of length n.
- A subset $D \subset V$ in an undirected graph is said to separate $A \subset V$ from $B \subset V$ if every path between a vertex in A and a vertex in B contains a vertex from D.

```
separates("a", "d", c("b", "c"), ug0)
```

```
R
```

This shows that $\{b, c\}$ separates $\{a\}$ and $\{d\}$.

Conditional ndependence and Graphs

Suppose that we have a collection of random variables $(X_v)_{v \in V}$ with a joint density. Let A, B and C be subsets of V and let $X_A = (X_v)_{v \in A}$ and similarly for X_B and X_C.

Then the statement that X_A and X_B are conditionally independent given X_C, written $A \coprod B | C$, means that for each possible value of x_C of X_C, X_A and X_B are independent in the conditional distribution given $X_C = x_c$.

So if we write $f()$ for a generic density or probability mass function, then one characterization of $A \coprod B | C$ is that

$$f(x_A, x_B | x_C) = f(x_A | x_C) f(x_B | x_C).$$

actori ation Criterion

An equivalent characterization that the joint density of (X_A, X_B, X_C) factorizes as

$$f(x_A, x_B, x_C) = g(x_A, x_C)h(x_B, x_C)$$

that is, as a product of two functions $g()$ and $h()$, where $g()$ does not depend on x_B and $h()$ does not depend on x_A.

ensit actori in ccordin to a Graph

Parametric models for $(X_v)_{v \in V}$ may be thought of as specifying a set of joint densities.

These may admit factorisations that give rise to conditional independence relations between the variables.

Some models give rise to patterns of conditional independences that can be represented as an undirected graph.

ensit actori in ccordin to a Graph

More specifically, let $G = (V, E)$ be an undirected graph with cliques $C_1, \ldots, C_k$. Consider a joint density $f()$ of the variables in V . If this admits a factorization of the form

$$f(x_V) = \prod_{i=1}^{k} g_i(x_{C_i})$$

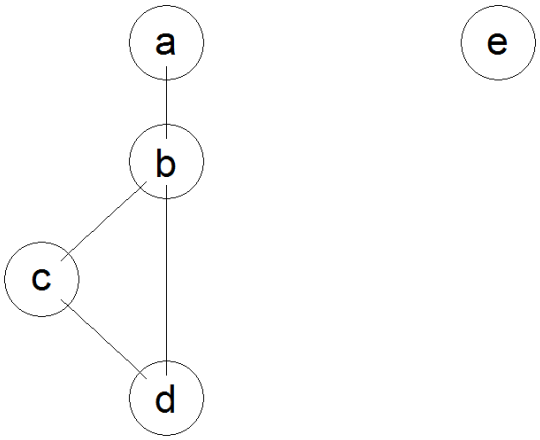
for some functions $g_1() \ldots g_k()$ where $g_j()$ depends on x only through x_{C_j} then we say that $f()$ factorizes according to G .

he Glo al Mar o ropert

If all the densities under a model factorize according to G , then G encodes the conditional independence structure of the model, through the following result,

The global Markov property : whenever sets A and B are separated by a set C in the graph, then $A \perp\!\!\!\perp B|C$ under the model.

a is Conditionall ndependent o d Gi en




```
separates("a", "d", "b", ug0)
```

```
      R
```

Gaussian Graphical Models

Carcass Data

The carcass data from gRbase contains measurements of the thickness of meat and fat layers at different locations on the back of a slaughter pig together with the lean meat percentage on each of 344 carcasses.

These data have been used to predict the lean meat percentage on the basis of the thickness measurements.

Carcass ata

Here Fat12 is the thickness of the fat layer between rib number 12 and 13.

LeanMeat is the percentage of meat in the carcass calculated as the weight of meat divided by the total weight of the carcass.

```
library(gRbase)
data(carcass)
 ead(carcass)
```

```
   at   Meat     at   Meat     at   Meat
Lea Meat

                                          0

       0        0              0
         0
```

Gaussian Graphical Models

Gaussian graphical models provide a framework for modelling how these variables are mutually related.

Consider a random vector $y = (y_1, \ldots, y_d)$ following a *multivariate normal* $N_d(\mu, \Sigma)$ distribution.

Gaussian Graphical Models

The key quantity in Gaussian graphical models is the inverse of the covariance matrix $K = \Sigma^{-1}$ known as the *concentration matrix*:

$$K = \begin{pmatrix} k_{1,1} & k_{1,2} & \cdots & k_{1,d} \\ k_{2,1} & k_{2,2} & \cdots & k_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ k_{d,1} & a_{d,2} & \cdots & k_{d,d} \end{pmatrix}$$

Gaussian Graphical Models GGM

The partial correlation between y_u and y_v given all other variables can be derived from K as $\rho_{uv}|V_{\setminus u,v} = -k_{uv}/\sqrt{k_{uu}k_{vv}}$.

Thus $k_{uv} = 0$ if and only if y_u and y_v are conditionally independent given all other variables.

In contrast to the concentrations, the partial correlations are invariant under a change of scale and origin.

Carcass ata

Returning to the carcass data, the concentration matrix can be estimated as

```
  carc <- co   t (carcass,  et od "ML") co
  carc <-sol e(  carc)
rou d( 00   carc)
```

```
          at   Meat    at   Meat    at
Meat   Lea Meat
 at                    - 0     -    -
         0
Meat                    -      -     -
-         -
 at         - 0     -                 -
-
Meat         -      -                  -
-          0
 at         -       -    -       -

Meat               -     -      -
          -
Lea Meat     0     -             0
-
```

Carcass ata

Concentrations depend on the scale on which the variables are measured.

Since lean meat percentage is measured on a different scale than the meat and fat measurements, the partial correlation matrix seems more appropriate for measuring dependence.

Carcass ata

The cov2pcor() function provides such a measure:

```
    carc <- co  pcor(  carc)
rou d( 00    carc)
```

```
          at   Meat    at   Meat    at
Meat   Lea Meat
 at          00    -              0
-         -
Meat        -        00

 at                      00    -
        -
Meat          0           -        00

 at                                 00
-        -
Meat        -                        -
 00
Lea Meat    -             -            -
         00
```

Gaussian Graphical Models

Two of the partial correlations relating to LeanMeat are very small which suggests that Fat13 is conditionally independent of Meat12 given the remaining variables and this also holds for LeanMeat and Meat12.

In particular this implies that Meat12 can be left out without loss of accuracy in the linear prediction of LeanMeat from the meat and fat measurements.

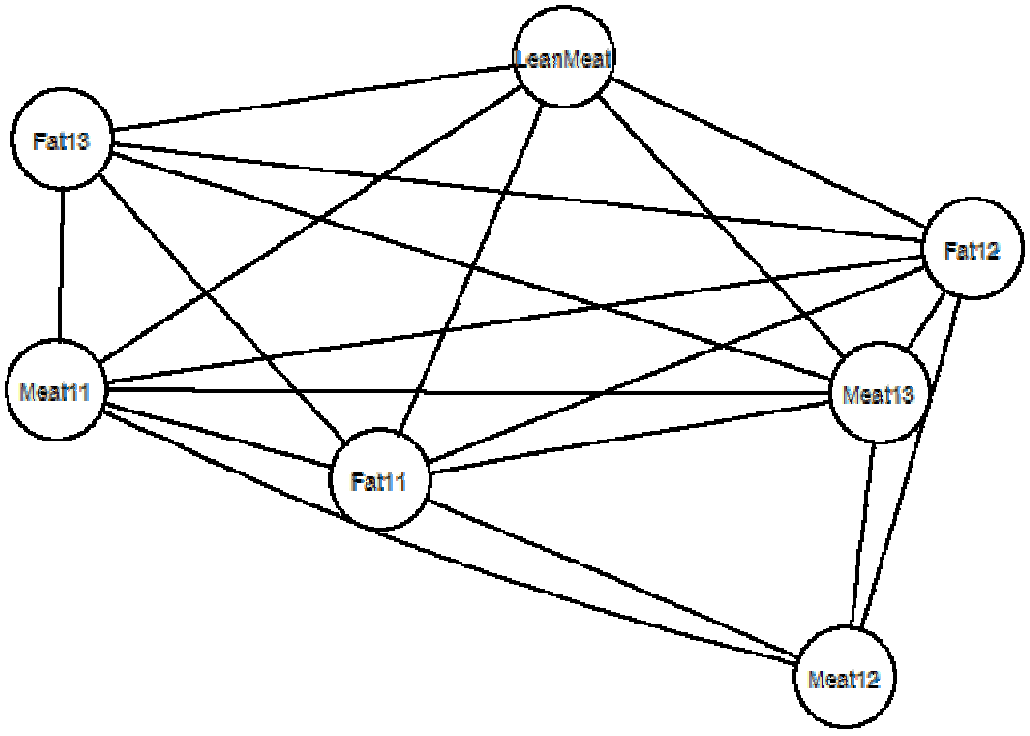
Gaussian Graphical Models

A stepwise backward model selection procedure using AIC from the saturated model yields the following graphical model.

This has two edges removed, corresponding to the conditional independences observed above.

```
library(gRi )
sat carc <- c od(    , data carcass)
aic carc<- step ise(sat carc)
library(Rgrap  i )
```

```
plot(as(aic carc,"grap   L"),"  dp")
```



GGM usin C

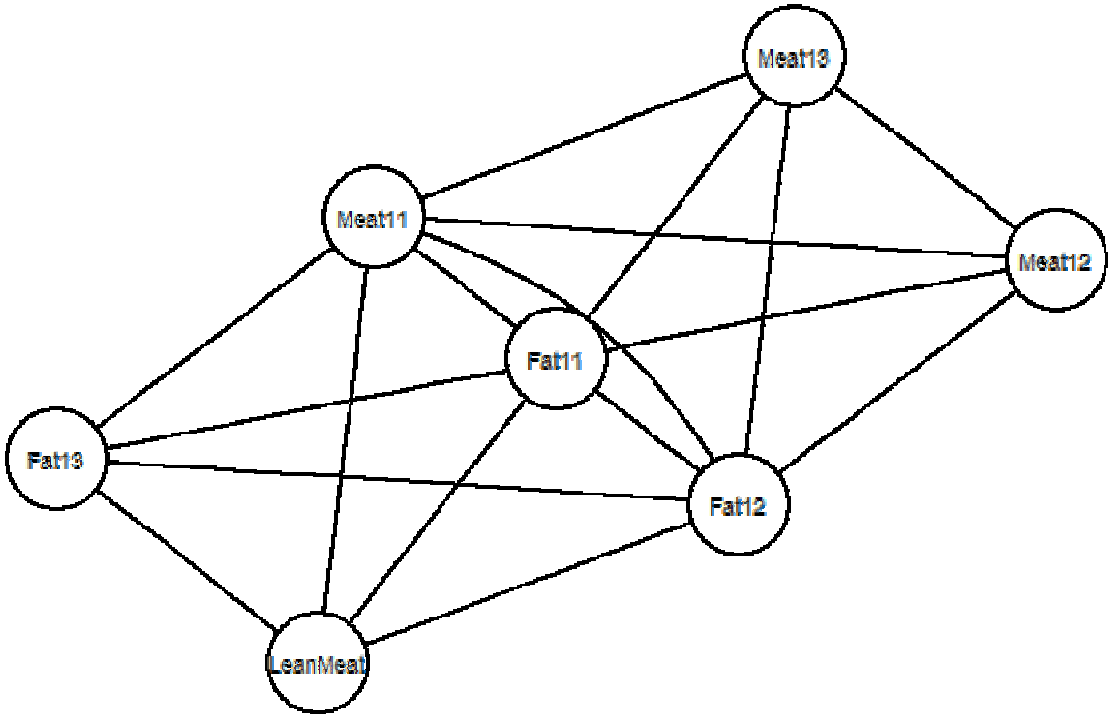
Using BIC, yielding a higher penalty for complexity, we get a simpler graph, where also edges between Fat13 and Meat13 as well as LeanMeat and Meat13 are removed:

```
bic carc<-step ise(sat carc,  log( ro
(carcass)))
bic carc
```

```
Model    cModel  it     ariables
 - logL                  0   di          aic
      0
 ide ia ce                   id          bic

 de ia ce                    d
```

```
plot(as(bic carc,"grap   L"),"  dp")
```



GGM usin C

It specifies a model with only two cliques.

The conditional independence relations can be summarised as (LeanMeat,Fat13) $\perp\!\!\!\perp$ (Meat12,Meat13)|(Fat11,Fat12,Meat11).

In this model neither Meat12 nor Meat13 contribute directly to the prediction of LeanMeat.

od at ata

These data in the gRbase package were collected to illustrate issues in making multiple regression models for prediction of percentage of body fat for an individual based on simple measurements of weight, circumferences of body parts, etc.

```
data(Body at)
 ead(Body at)
```

```
   e sity Body at  ge  eig t  eig t  ec
  est  bdo e     ip   ig    ee
     0 0
                      0
     0
           0
     0                    00            0

     0         0
 0             0       0
     0  0
        00 0  0
     0 0       0          0                0
 0             0         0    0
     le Biceps  orear   rist
              0
           0
      0

      0
                  0
```

od at ata

The measurement of body density is made by an elaborate underwater weighing procedure and the estimated percentage of body fat is then calculated from the body density by a linear expression in the reciprocal of the latter.

Outliers (obvious errors) are removed in the data. Density is removed because it's related functionally to BodyFat. In addition, Age and Weight are transformed by the square root to gain roughly linear relationships among the variables.

```
Body at <- Body at -c
(  ,  ,  ,  ,  ,  ,  ,  ,  ,  , 0 ),
Body at  ge <- s rt(Body at  ge)
Body at  eig t <- s rt(Body at  eig t)
gRbody at <- Body at ,
```

od at ata

After these changes, we get the following partial correlation matrix

```
  body<-co   t(gRbody at,  et od  "ML") co
   body<-co  pcor(  body)
rou d( 00    body)
```

```
        Body at  ge  eig t  eig t  ec    est
 bdo e   ip   ig    ee    le
Body at      00                -   -      -
   -           0    -
 ge             00    -       -     0
    -    -          -
 eig t          -       00
                0
 eig t       -   -            00  -
-       -   -     -          -
 ec         -     0           -     00    -
  -           -    -
  est        -                -     -     00
   - 0   -     -     -
 bdo e                        -
 00        -          -
 ip         -    -           -    -     - 0
    00               -
  ig            -              -
-        -         00
  ee           0          0          -     -
               00
   le         -  -             -    -
-      -    -                 00
Biceps                       -
      -    -      0   -    -
 orear          -               -
```

```
      -   - 0                -
 rist       -                   -          - 0
         -     -
        Biceps  orear   rist
Body at                  -
 ge                -
 eig t
 eig t     -         -     -
 ec
  est                    - 0
 bdo e     -        -
 ip         -       - 0
  ig         0             -
  ee         -              -
   le       -         -
Biceps      00
 orear              00
 rist                    00
```

od at ata

There's a high partial correlation between BodyFat and Abdomen.

If we again fit a model by stepwise selection using BIC we get

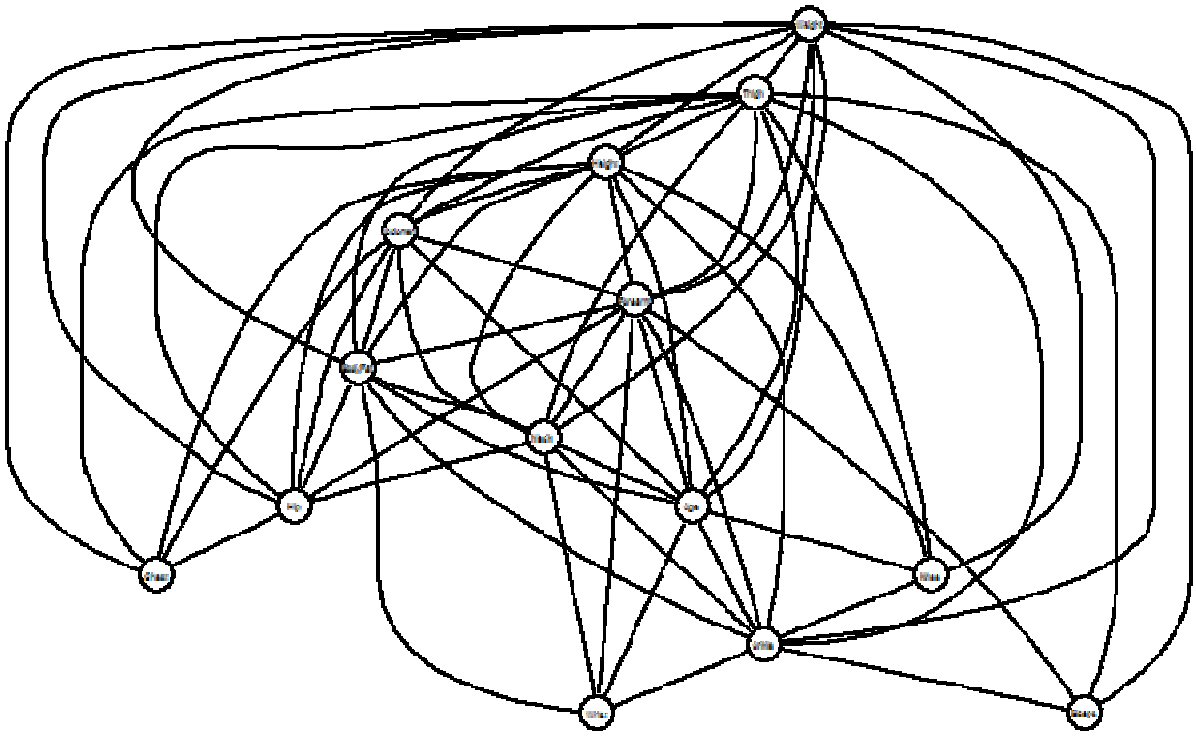
```
sat body <- c od(    , data gRbody at)
bic body<-step ise(sat body,  log( ro
(gRbody at)))
bic body
```

```
Model    cModel  it      ariables
 - logL                     di          aic

 ide ia ce                  id          bic

 de ia ce                   d         0
```

```
plot(bic body," eato")
```



GGM or od at ata

It has all variables except Chest, Knee, and Biceps as direct predictors for BodyFat.

Note that the model is rather dense.

Indeed 61 out of 91 possible edges are present in the model:

```
grap   degree(as(bic body,"grap   L"))
```

```
  eig t      ig     eig t  bdo e   orear
Body at      ec        ip      ge

 0        0                 0
    est       ee       le  Biceps    rist
                         0
```

Undirected Gaussian Graphical Models UGGM

An undirected Gaussian graphical model (hereafter abbreviated UGGM) is represented by an undirected graph $G = (\Gamma, E)$ where the vertices $\Gamma = \{1, \ldots, d\}$ represent the set of variables and E is a set of undirected edges.

When a random vector y follows a Gaussian distribution $N_d(\mu, \Sigma)$, the graph G represents the model where $K = \Sigma^{-1}$ is a positive definite matrix with $k_{uv} = 0$ whenever there is no edge between vertices u and v in G .

This graph is called the dependence graph of the model because it holds for all u, v that if u and v have no edge between them, then $u \coprod v | \Gamma_{\setminus u,v}$.

UGGM

Recall that a probability distribution factorizes according to an undirected graph $G = (V, E)$ if it admits a factorisation of the form

$$f(x)V) = \prod_{i=1,\ldots,k} g_i(x_{C_i})$$

where $C_1 \ldots C_k$ are the cliques of G , and $g_1() \ldots g_k()$ are arbitrary functions.

We can factorize the multivariate normal density and show that if we have, for a set of vertices A and B separated by a set C in the dependence graph, that $k_{uv} = 0$ for $u \in A$ and $v \in B$, then for Gaussian graphical models, the global Markov property holds: $A \coprod B|C$.

stimation, i elihood, and Model ittin

Consider a sample $y_1, \ldots, y_n$ of n observations of $y \sim N_d(\mu, \Sigma)$.

Let W denote the matrix of sums of squares and products, $W = \sum_{\nu=1}^{n} (y^\nu - \bar{y})(y^\nu - \bar{y})^T$, and let $S = W/n$ denote the empirical covariance matrix.

The log-likelihood function based on the sample is

$$\log L(K, \mu) = \frac{n}{2} \log \det(K) - \frac{n}{2} \mathrm{tr}(KS) - \frac{n}{2} (\bar{y} - \mu)^T K (\bar{y} - \mu)$$

stimation, i elihood, and Model ittin

For fixed K this is clearly maximized for $\hat{\mu} = \bar{x}$ which renders the last term equal to zero. The profile likelihood for K thus becomes

$$\log L(K, \hat{\mu}) = \frac{n}{2} \log \det(K) - \frac{n}{2} \operatorname{tr}(KS)$$

Since $\operatorname{tr}(KS) = \sum_u \sum_v s_{uv} k_{uv}$ it follows that only elements s_{uv} of S for which the corresponding elements k_{uv} of K are non-zero will contribute to the likelihood.

The MLE for Σ under the saturated model with no conditional independence restrictions satisfies $\hat{\Sigma} = S$, where S is the sample covariance matrix, so in that case we have $\hat{K} = S^{-1}$, provided S is not singular.

pothesis estin

Because $\hat{\Sigma}$ and S differ exactly on those entries for which $k_{uv} = 0$ it holds $\mathrm{tr}(\hat{K}S) = \mathrm{tr}(\hat{K}\Sigma) = d\,.$

Thus the maximized value of the log likelihood is $\hat{l} = n \log\det(\hat{K})/2 - nd/2\,.$

Thus the *deviance* of a model M is

$$D = \mathrm{dev} = 2(\hat{l}_{sat} - \hat{l}) = n \log\{\det(S^{-1})/\det(\hat{K})\} = -n \log\det(S\hat{K})$$

pothesis estin

The likelihood ratio test statistic for testing M_1 under M_0 where $M_1 \subseteq M_0$ is the difference in deviance (or ideviance) between the two models:

$$lrt = 2(\hat{l}_0 - \hat{l}_1) = n \log(\det(\hat{K}_0)/\det(\hat{K}_1))$$

The likelihood ratio test statistic can be used for testing M_1 under M_0 : large values of lrt suggest that the null hypothesis M_1 is false.

Under the hypothesis that M_1 holds, lrt has an approximate χ^2_f distribution where f is the difference in the number of parameters of the two models, which is the same as the difference in the number of edges.

pothesis estin

```
co pare odels <- u ctio ( , )
lrt <- iti o de - iti o de
 d di <- iti o di e sio -
  iti o di e sio
  a es(d di ) <- LL
list( lrt lrt, d d di )

co pare odels(aic carc,bic carc)
```

```
 lrt


 d
```

indicating that the simpler model does not quite fit.

pothesis estin

The ciTest_mvn() function can be used for testing a single conditional independence hypothesis, or when put in the terminology of graphs, for testing whether a single edge can be deleted from the saturated model (the model for which the dependence graph is complete).

Default is to use the likelihood ratio test for testing against the alternative hypothesis, the saturated model, using the deviance.

pothesis estin

For example to test for LeanMeat ⟦ Meat13|rest we can use:

```
ci est    (list(co    carc,   obs  ro
(carcass)), set
 Lea Meat Meat   Meat   Meat    at    at    at  )
```

```
 esti g Lea Meat     Meat     Meat   Meat
 at    at    at
 tatistic (   )          d     p- alue
0    0  et od
```

so ciTest_mvn() interprets set by testing conditional independence of the two first variables given the remaining.

Concentration and e ression

There is a close connection between the concentration matrix K and multiple linear regression. Let $A \subset \Gamma$ and $B = \Gamma \setminus A$. This defines a partitioning of y, μ, Σ and K according to A and B as

$$y = \begin{pmatrix} y_A \\ y_B \end{pmatrix}, \ \mu = \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{pmatrix}, \ K = \begin{pmatrix} K_{AA} & K_{AB} \\ K_{BA} & K_{BB} \end{pmatrix}$$

Concentration and e ression

Then

$$y_A|y_B \sim N(\mu_A|B, \Sigma_A|B)$$

where

$$\mu_A|B = \mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(y_B - \mu_B)$$
$$\Sigma_A|B = \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{AB}$$

Standard results on the inverse of partitioned matrices gives that the quantities can be expressed in terms of K as

$$\Sigma_{AB}\Sigma_{BB}^{-1} = -K_{AA}^{-1}K_{AB}$$
$$\Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{AB} = K_{AA}^{-1}$$

Concentration and e ression

Consider a multiple regression of y_1 on $y_2, \ldots, y_d$ as explanatory variables,

$$y_1 = a_1 + \beta_{13} y_2 + \ldots + \beta_{1d} y_d + \epsilon_1 \text{ where } \epsilon_1 \sim N(0, \sigma_1^2)$$

Then

$$\sigma_1^2 = 1/k_{11}$$

while the regression coefficients $\beta_{12}, \ldots, \beta_{1d}$ can be derived from K as

$$(\beta_{12}, \ldots, \beta_{1d}) = -(k_{12}, \ldots, k_{1d})/k_{11}$$

Concentration and e ression

Returning to K for the carcass data we find that the regression coefficients for predicting LeanMeat are

```
-  carc  ,-     carc  ,
```

```
        at          Meat           at
Meat           at          Meat
-0             0           -0
0 0          -0       0     0 0
```

while the residual variance of the lean meat percentage is

```
    carc  ,
```

```

```

Model election

We can use stepwise selection using AIC and BIC for model selection, but another method is the glasso algorithm, implemented in the glasso package.

This gives a fast technique to find the Gaussian graphical model that maximizes a log-likelihood for K which is penalized by the L_1 -norm.

Model election

An L_1 -penalized log-likelihood is equivalent to

$$L_{pen}(K, \hat{\mu}) = \log\det(K) - \mathrm{tr}(KS) - \rho||K||_1$$

where ρ is a non-negative penalty parameter.

This penalized log-likelihood is convex in K and can be optimized by convex programming methods.

The smaller the value of ρ , the denser the graph that results.

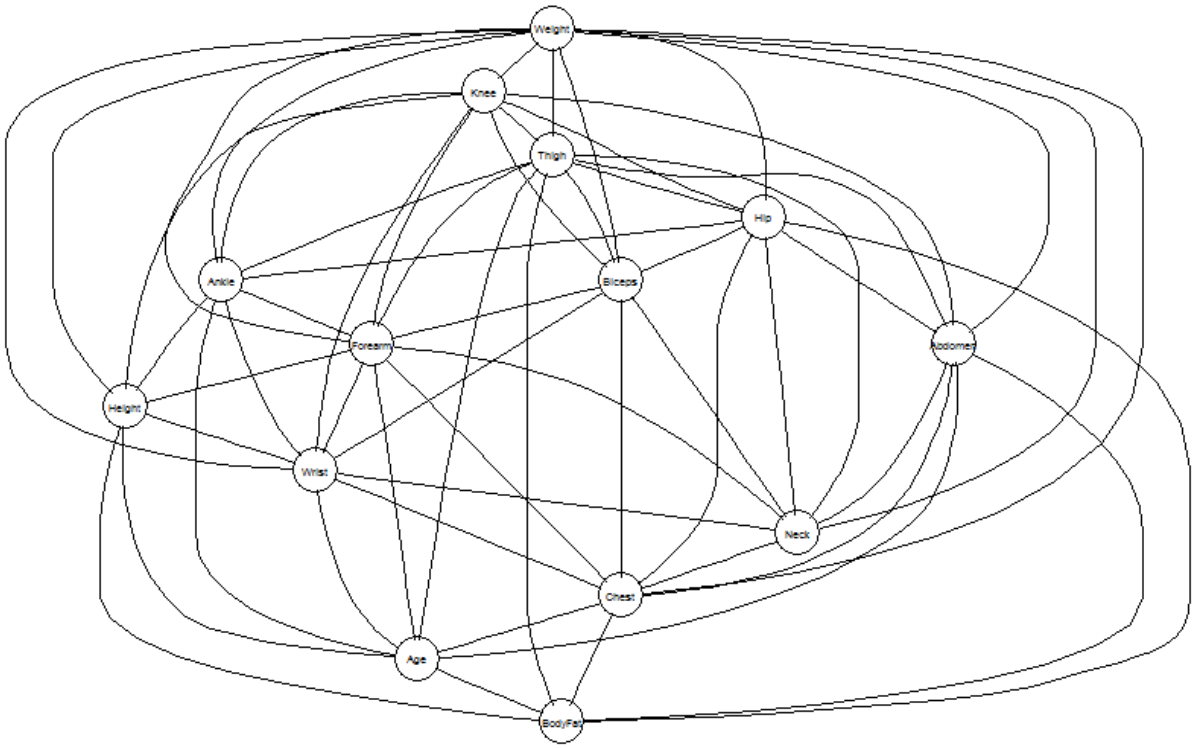
Model election

Here the the glasso algorithm implemented on the gRbodyfat data:

```
  body<-co  cor(  body)
library(glasso)
 res lasso<-glasso(  body,r o 0  )
  M <- res lasso  i    0
 diag( M) <-
 g lasso <- as( M, "grap   L")
  odes(g lasso)<- a es(gRbody at)
 glasso body<-c od(g lasso,data gRbody at)
```

o od at is onl connected to e, ei ht, Chest, domen, ip, and hi h

```
plot(glasso body, " eato")
```



ittle it out m esearch

What if the data are not normal? Can we still build a graphical model?

Nonparanormal Graphical Model

- A nonparametric extension of the Gaussian graphical model
- Random variables $X = (X_1, \ldots, X_p)$ are replaced with the transformed random variables through some unknown smooth monotone transformations $f_1, \ldots, f_p$:

$$Z = f(X) = (f_1(X_1), \ldots, f_p(X_p)) \sim \text{Normal}(\mu, \Sigma)$$

onparanormal Graphical Model

- The transformation function f_j is estimated using

$$\tilde{f}_j(x) \equiv \hat{\mu}_j + \hat{\sigma}_j \tilde{h}_j(x)$$

where $\tilde{h}_j(x) = \Phi^{-1}(\tilde{F}_j(x))$ and $\hat{\mu}_j$ and $\hat{\sigma}_j$ are the sample mean and standard deviation of the observed variables and $\tilde{F}_j(x)$ is a truncated marginal empirical distribution function.

- They estimate sparse Ω using $\tilde{f}(X_1), \ldots, \tilde{f}(X_n)$ by applying the graphical lasso.

onparanormal Graphical Model

Overall idea: Plug in the observed data into these estimated transformed functions to get your transformed variables and then you have a classical sparsity problem to estimate the precision matrix and learn the graphical structure.

a esian onparanormal Graphical Model

- My research involves estimating these transformation functions using a B-splines random series prior, where the coefficients of the B-splines are given a normal prior, ordered to induce monotonicity.

- Then with the transformed variables, I use a spike and slab prior to estimate a sparse precision matrix and construct the graph.

- My research is looking at other transformation functions as well as other ways to estimate the sparse precision matrix.

Conclusion

- Gaussian graphical models are mathematical objects that can visually depict how variables are mutually related to each other.
- They can be useful in variable selection problems, particularly for prediction.
- There are many different ways to select a graph, including AIC, BIC, and glasso. This is an area of active research.