Algebraic Methods for Gaussian Random Variables

Mathias Drton

Department of Statistics

The University of Chicago

Outline

- **Probability & Statistics**:

  1. Multivariate normal (Gaussian) distribution
  2. Likelihood inference in Gaussian models

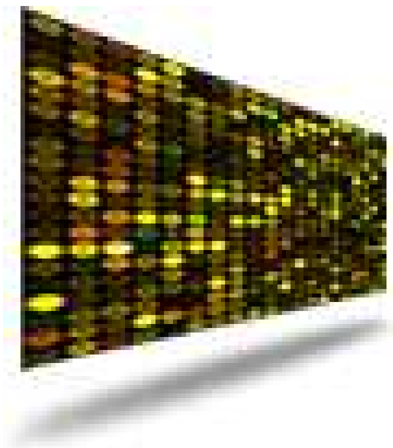
- **Algebra**: Techniques for answering statistical questions

  1. Algebraic maximum likelihood estimation
  2. Model invariants & model fit
  3. Model singularities & parameter identification

“Probability”

The multivariate normal $\equiv$ Gaussian distribution

Data from microarray experiments



YBR243C	1.18	1.18	-0.64	0.16	-0.92	-1.32
YDL093W	0.24	1.10	-0.45	-0.29	-0.33	-0.37
YDL095W	1.22	1.69	-0.54	-0.27	-0.29	-0.66
YER001W	-0.47	2.49	-1.60	-0.15	-2.29	NA

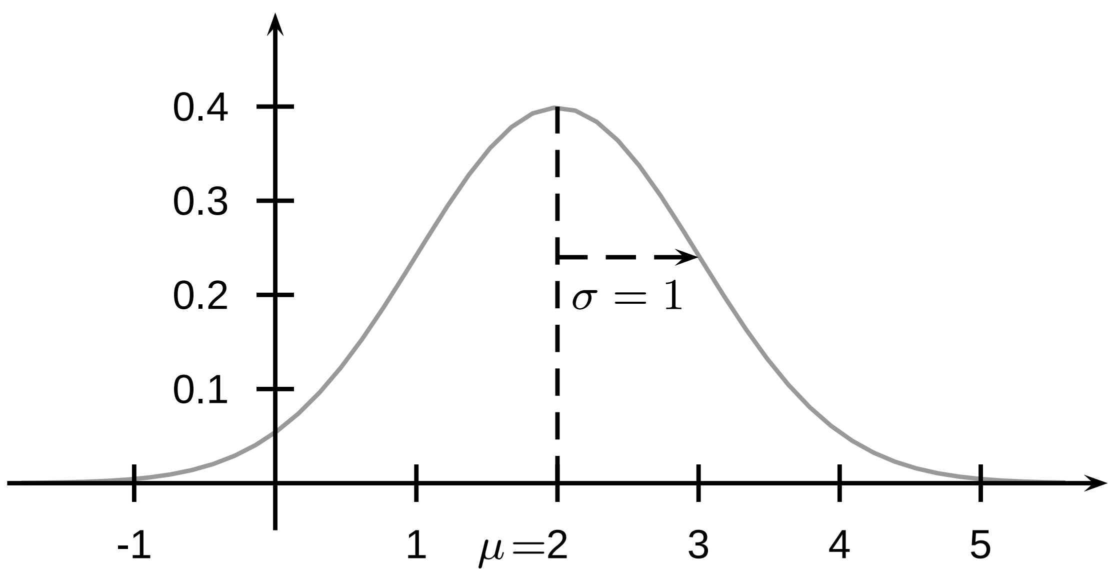
- Gene expression data are **continuous** and generally **correlated**. (physical interactions between genes)
- **Multivariate normal distribution** provides many interesting models for such data.
- Appropriateness of normality assumptions should be checked.

Univariate normal distribution

Normal distribution $\mathcal{N}(\mu, \sigma^2)$ has (Lebesgue) prob. density function

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \qquad x \in \mathbb{R}.$$

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\mathrm{E}[X] = \mu \in \mathbb{R}$ and $\mathrm{Var}[X] = \sigma^2 > 0$.



Multivariate normal distribution

Multivariate normal distribution $\mathcal{N}_p(\mu, \Sigma)$ has pdf on $\mathbb{R}^p$:

$$f_{\mu,\Sigma}(\vec{x}) \;=\; \frac{1}{\sqrt{(2\pi)^p \det(\Sigma)}} \exp\left\{-\frac{1}{2}(\vec{x}-\mu)^t \Sigma^{-1}(\vec{x}-\mu)\right\}.$$

If $\vec{X} \sim \mathcal{N}_p(\mu, \Sigma)$, then $\mathrm{E}[\vec{X}] = \mu = (\mu_1, \ldots, \mu_p)^t \in \mathbb{R}^p$ and

$$\mathrm{Var}[\vec{X}] = \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \ldots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{1p} & \sigma_{2p} & \ldots & \sigma_{pp} \end{pmatrix} \in \mathbb{R}^{p\times p} \text{ is sym. \& pos. def.}$$

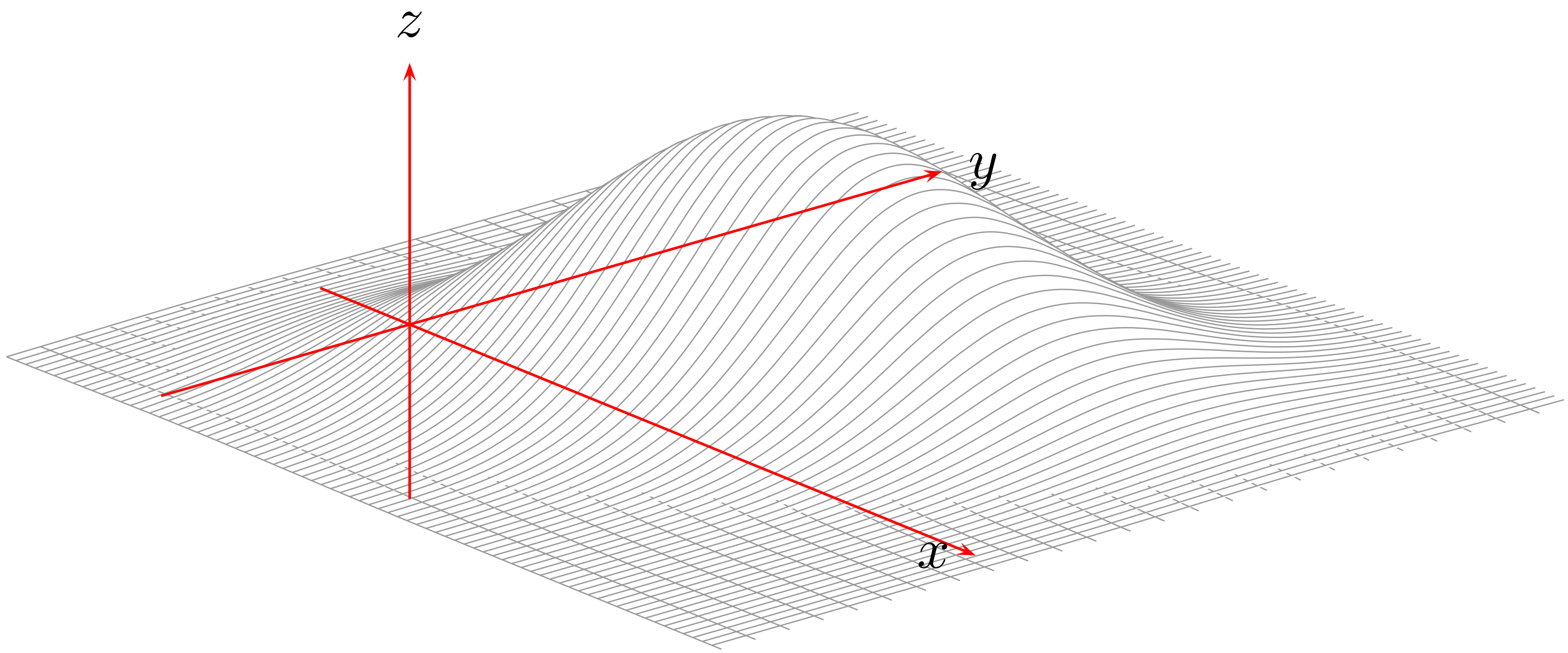
Linear transformations:

$$A\vec{X} + b \sim \mathcal{N}_p(A\mu + b, A\Sigma A^t), \quad A \text{ full rank}$$

A bivariate density

Plot of pdf $f_{\mu,\Sigma}$ of the bivariate normal distribution $\mathcal{N}_2(\mu, \Sigma)$ with

$$\mu = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$



Marginalizing and conditioning

If X is a multivariate normal random vector, $X \sim \mathcal{N}_p(\mu, \Sigma)$, then

- **normal marginal dist.**: For $A \subseteq \{1, \ldots, p\}$,

$$X_A = (X_i \mid i \in A) \sim \mathcal{N}_A(\mu_A, \Sigma_{A\times A}).$$

- **normal conditional dist.**: For $A, C \subseteq \{1, \ldots, p\}$, $A \cap C = \emptyset$:

$$(X_A \mid X_C = x_C) \sim \mathcal{N}_A\Big(\underbrace{\mu_A + B_{A.C}(x_C - \mu_C)}_{\text{linear in } x_C!},\ \underbrace{\Sigma_{AA.C}}_{\text{constant in } x_C!} \Big)$$

where

$$B_{A.C} = \Sigma_{A\times C}\Sigma_{C\times C}^{-1} \qquad \text{(matrix of regression coeff.)}$$

$$\Sigma_{AA.C} = \Sigma_{A\times A} - \Sigma_{A\times C}\Sigma_{C\times C}^{-1}\Sigma_{C\times A} \qquad \text{(cond. covariance matrix)}$$

Marginal independence

If X is a multivariate normal random vector, $X \sim \mathcal{N}_p(\mu, \Sigma)$, then two sub-vectors X_A and X_B are **independent** if

$$(X_A \mid X_B = x_B) \sim X_A \quad \forall x_B \in \mathbb{R}^B.$$

In other words, $(X_A \mid X_B = x_B)$ does not depend on x_B.

Under the assumed multivariate normality,

$$\begin{aligned}
\boldsymbol{X_A \perp\!\!\!\perp X_B} \iff & \; \mathrm{E}[X_A \mid X_B = x_B] = \mathrm{E}[X_A] = \mu_A \quad \forall x_B \in \mathbb{R}^B \\
\iff & \; B_{A.B} = \Sigma_{A\times B}\Sigma_{B\times B}^{-1} = 0 \\
\iff & \; \Sigma_{A\times B} = 0 \\
\iff & \; \boldsymbol{\sigma_{ij} = 0 \quad \forall i \in A,\; j \in B} \\
\iff & \; X_i \perp\!\!\!\perp X_j \quad \forall i \in A,\; j \in B
\end{aligned}$$

Conditional independence

If X is a multivariate normal random vector, $X \sim \mathcal{N}_p(\mu, \Sigma)$, then sub-vectors X_A and X_B are **conditionally independent given** a sub-vector X_C if

$$(X_A \mid X_B = x_B,\, X_C = x_C) \sim (X_A \mid X_C = x_C) \quad \forall x_B,\, x_C.$$

Under the assumed multivariate normality,

$$\begin{aligned} \boldsymbol{X_A \perp\!\!\!\perp X_B \mid X_C} \iff & \ \mathbf{det}(\boldsymbol{\Sigma_{iC \times jC}}) = \mathbf{0} \quad \forall \boldsymbol{i \in A,\ j \in B} \\ \iff & \ X_i \perp\!\!\!\perp X_j \mid X_C \quad \forall i \in A,\ j \in B. \end{aligned}$$

Note that

$$X_i \perp\!\!\!\perp X_j \mid X_{\{1,\ldots,p\} \setminus \{i,j\}} \iff \text{adj}(\Sigma)_{ij} = 0 \iff \Sigma^{-1}_{ij} = 0.$$

“Statistics”

Likelihood inference in Gaussian models

Gaussian models

- Sample of random vectors $X_i = (X_{i1}, \ldots, X_{ip})^t \in \mathbb{R}^p$:

$$X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}_p(\mu, \Sigma).$$

- A **Gaussian statistical model** $\mathbf{M}$ is a family

$$\mathbf{M} \subseteq \left\{ \mathcal{N}_p(\mu, \Sigma) \mid (\mu, \Sigma) \in M \right\}, \qquad M \subseteq \mathbb{R}^p \times \mathbb{R}^{p \times p}_{\text{pos.def.}}$$

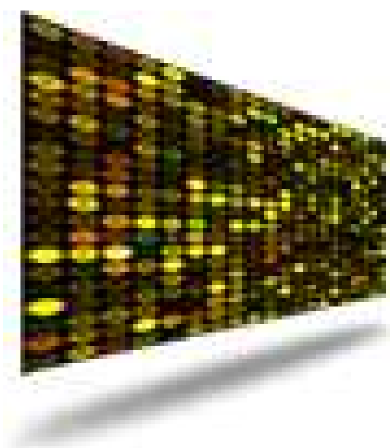
  The set M is the **mean parameter space** of $\mathbf{M}$.

- The **saturated model** is the model $\mathbf{M}$ of all multivariate normal distributions, i.e. $M = \mathbb{R}^p \times \mathbb{R}^{p \times p}_{\text{pd}}$.

Data matrix

- Create matrix $X = (X_{ij}) \in \mathbb{R}^{p \times n}$ by filling columns with random vectors $X_i = (X_{i1}, \ldots, X_{ip})^t \in \mathbb{R}^p$ in the sample.
- Density function for random matrix X is function on $\mathbb{R}^{p \times n}$:

$$f_{\mu,\Sigma}(x) = \frac{1}{\sqrt{(2\pi)^{pn} \det(\Sigma)^n}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^t \Sigma^{-1} (x_i - \mu) \right\}.$$



YBR243C	1.18	1.18	-0.64	0.16	-0.92
YDL093W	0.24	1.10	-0.45	-0.29	-0.33
YDL095W	1.22	1.69	-0.54	-0.27	-0.29
YER001W	-0.47	2.49	-1.60	-0.15	-2.29

Sufficient statistics

- **Sample mean vector**

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \in \mathbb{R}^p$$

- **Sample covariance matrix** (pos.def. with prob. 1 if $n \geq p+1$)

$$S = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^t \in \mathbb{R}^{p\times p}$$

- Sufficient statistics:

$$f_{\mu,\Sigma}(x) = \frac{1}{\sqrt{(2\pi)^{pn}\det(\Sigma)^n}} \times \exp\left\{-\frac{n}{2}\operatorname{tr}(\Sigma^{-1}S) - \frac{n}{2}(\bar{x}-\mu)^t\Sigma^{-1}(\bar{x}-\mu)\right\}$$

Maximum likelihood estimation

- **Log-likelihood function** $\ell : M \longrightarrow \mathbb{R}$ of model $\mathbf{M}$ for data $x \in \mathbb{R}^{p\times n}$:

$$\ell(\mu,\Sigma) \mapsto \log f_{\mu,\Sigma}(x) \propto \frac{n}{2}\log(\det(\Sigma^{-1})) - \frac{n}{2}\operatorname{tr}(\Sigma^{-1}S) - \frac{n}{2}(\bar{x}-\mu)^t\Sigma^{-1}(\bar{x}-\mu).$$

- Maximum likelihood estimates (**MLE**)

$$(\hat{\mu},\hat{\Sigma}) = \arg\max\{\ell(\mu,\Sigma) \mid (\mu,\Sigma) \in M\}$$

- Under regularity conditions

$$(\hat{\mu},\hat{\Sigma}) \longrightarrow_d \mathcal{N}\left(0, \left(\mathrm{E}_{\mu,\Sigma}\left[-D_2\ell(\mu,\Sigma)\right]\right)^{-1}\right)$$

- If $\mathbf{M}$ is saturated model and $S > 0$ then MLE $(\hat{\mu},\hat{\Sigma}) = (\bar{x}, S)$ is unique local = global maximum of ℓ.

“Algebra”

Algebraic models and computational techniques

Algebraic Gaussian models

- A Gaussian model $\mathbf{M}$ is a subfamily

$$\mathbf{M} \subseteq \{\mathcal{N}_p(\mu, \Sigma) \mid (\mu, \Sigma) \in M\}, \qquad M \subseteq \mathbb{R}^p \times \mathbb{R}^{p\times p}_{\text{pos.def.}}$$

- In this talk, an **algebraic Gaussian model** $\mathbf{M}$ is a model with mean parameter space given by a **polynomial parameterization**,

$$M = \mathbf{f}(\Theta), \quad \mathbf{f} \text{ polynomial}, \quad \Theta \subseteq \mathbb{R}^d,$$

or by a **rational parameterization**,

$$M = \mathbf{f}(\Theta) = \frac{\mathbf{h}}{\mathbf{g}}(\Theta), \quad \mathbf{h}, \mathbf{g} \text{ polynomials}, \quad \Theta \subseteq \mathbb{R}^d,$$

Algebra Part 1: Algebraic ML estimation

- Likelihood equations of algebraic Gaussian model $\mathbf{M}$

$$\frac{\partial}{\partial \theta} \ell(\mathbf{f}(\theta)) = 0$$

are rational equations.

- If not already polynomial then these equations can be converted into polynomial equations.

- Computational algebra software permits to solve these equations numerically for given data:

$\Longrightarrow$ we obtain **all solutions** to the likelihood equations.

- Care must be taken in situations where the global maximum may occur on the boundary of the parameter space.

Example: Behrens-Fisher (Sugiura, Gupta, 1987)

- Test equality of means of two indep. populations with arbitrary variances
- Algebraic Gaussian model w. linear par. space $M = \mathbf{f}\big(\mathbb{R} \times (0,\infty)^2\big)$,

$$\mathbf{f}(\nu, \sigma_1, \sigma_2) = \left[\begin{pmatrix} \nu \\ \nu \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \right]$$

- Log-likelihood function simplifies to

$$\ell(\nu, \sigma_1, \sigma_2) \propto -\log(\sigma_1) - \log(\sigma_2) - \left(\frac{s_{11} + (\bar{x}_1 - \nu)^2}{2\sigma_1^2} + \frac{s_{22} + (\bar{x}_2 - \nu)^2}{2\sigma_2^2} \right)$$

- Sufficient statistics $\bar{x} = (\bar{x}_1, \bar{x}_2)$ and diagonal entries (s_{11}, s_{22}) of $S = (s_{ij}) \in \mathbb{R}^{2\times 2}$. Note: $s_{ii} > 0$ with prob. 1 if $n \geq 2$.

Note: linearity in mean pars. (μ, Σ) not immediately benefi cial. Linearity of natural pars. $(\Sigma^{-1}\mu, \Sigma^{-1})$ yields concave log-likelihood function.

Behrens-Fisher (cont.)

- Likelihood equations for generic $(\bar{x}, S)$: $(\times \sigma_1^2 \sigma_2^2$ to clear denom)

$$\begin{aligned} \frac{\partial}{\partial \nu}\ell\big(\mathbf{f}(\nu,\sigma_1,\sigma_2)\big) &= 0 \\ \frac{\partial}{\partial \sigma_i}\ell\big(\mathbf{f}(\nu,\sigma_1,\sigma_2)\big) &= 0 \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} (\bar{x}_1 - \nu)\sigma_2^2 + (\bar{x}_2 - \nu)\sigma_1^2 &= 0, \\ \sigma_i^2 = s_{ii} + (\bar{x}_i - \nu)^2, \quad i &= 1,2. \end{aligned}$$

- Lexicographic Gröbner basis $\sigma_1 \succ \sigma_2 \succ \nu$:

$$\begin{aligned} 2\nu^3 - 3(\bar{x}_1 + \bar{x}_2)\nu^2 + (\bar{x}_1^2 + 4\bar{x}_1\bar{x}_2 + \bar{x}_2^2 + s_{11} + s_{22})\nu \\ - (\bar{x}_1^2\bar{x}_2 + \bar{x}_1\bar{x}_2^2 + \bar{x}_1 s_{22} + \bar{x}_2 s_{11}); \end{aligned}$$

$$\sigma_i^2 - s_{ii} - (\bar{x}_i - \nu)^2, \quad i = 1,2.$$

- Discriminant D (degree 6) tells whether 1 or 3 sols. to lik. equations

- If data in model, $\bar{x}_1 = \bar{x}_2$, then $D = -8(s_{11} + s_{22})^3 < 0$

$$\Longrightarrow \quad \#\{\text{solutions to lik. equations}\} \xrightarrow{a.s.} 1$$

Example: Covariance model (Drton, Richardson, 2004)

- Algebraic Gaussian model with linear par. space M given by

$$\mathbf{f}(\sigma) = \left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & 0 & \sigma_{13} & 0 \\ 0 & \sigma_{22} & 0 & \sigma_{24} \\ \sigma_{13} & 0 & \sigma_{33} & \sigma_{34} \\ 0 & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{pmatrix} \right], \quad \sigma \in \boldsymbol{\Theta} \subset \mathbb{R}^7$$

- Log-likelihood function simplifies to

$$\ell(\sigma) \propto -\log[\det(\mathbf{f}_\Sigma(\sigma))] - \mathrm{tr}[\mathbf{f}_\Sigma(\sigma)^{-1}\bar{S}], \qquad \bar{S} = \frac{1}{n} X X^t \in \mathbb{R}^{4\times 4}.$$

- Sufficient statistics $\bar{S} > 0$ with probability one if $n \geq 4$.

Covariance model (cont.)

- Rational likelihood equations $\quad (ij \neq 12, 14, 23)$

$$\frac{\partial \ell(\mathbf{f}_\Sigma(\sigma))}{\partial \sigma_{ij}} = 0 \iff (\Sigma^{-1})_{ij} = (\Sigma^{-1}\bar{S}\Sigma^{-1})_{ij}$$

$$\iff [\text{adj}(\Sigma)]_{ij} = \frac{[\text{adj}(\Sigma) \cdot \bar{S} \cdot \text{adj}(\Sigma)]_{ij}}{\det(\Sigma)}$$

- Clear denominator $\det(\Sigma)$ to obtain polynomial equations defining an ideal $I \subset \mathbb{R}[\sigma_{ij}]$

$$[\text{adj}(\Sigma)]_{ij} \det(\Sigma) = [\text{adj}(\Sigma) \cdot S \cdot \text{adj}(\Sigma)]_{ij}$$

- Remove noninvertible matrices from $V(I)$ by **saturation** $I : \det(\Sigma)^\infty$.

$\Longrightarrow$
a) 5 sols. to likelihood equations (all 5 may be real & feasible)
b) $\#\{$solutions to lik. equations$\} \xrightarrow{a.s.} 1$

Algebra Part 2: Model invariants

- $M = \mathbf{f}(\Theta)$ defined parametrically
- What is the implicit description? (check whether $(\mu, \Sigma) \in M$)
- **Model invariants** are the equality relations among components of mean parameters

  $$\mu = (\mu_h \mid h = 1, \ldots, p)$$

  and

  $$\Sigma = (\sigma_{ij})_{1 \le i,j \le p}.$$

  In other words,

  $$f \in \mathbb{R}[\mu_h, \sigma_{ij} \mid h, i \le j] \ : \ f(\mu, \Sigma) = 0 \quad \forall \ (\mu, \Sigma) \in M$$

- Note: there may be inequalities among components of (μ, Σ).

Invariants as test statistics

- Statistical test of goodness-of-fit of algebraic Gaussian model $\mathbf{M}$:

$$H: \ (\mu, \Sigma) \in M \quad \text{versus} \quad K: \ (\mu, \Sigma) \notin M$$

- If polynomial f is model invariant for model $\mathbf{M}$:

$$f(\mu, \Sigma) \neq 0 \quad \Longrightarrow \quad (\mu, \Sigma) \notin M$$

- Test H vs K by testing (weaker) substitute hypothesis

$$H_f: \ f(\mu, \Sigma) = 0 \quad \text{versus} \quad K_f: \ f(\mu, \Sigma) \neq 0$$

Rejection of $H_f \quad \Longrightarrow \quad$ rejection of H

Invariants as test statistics (cont.)

- To test $H_f: \; f(\mu,\Sigma)=0$ we can use the test statistic $t_f := f(\bar{x},S)$

  1. Compute variance $\text{Var}_{\mu,\Sigma}[f(\bar{x},S)]$

     (moments of normal and Wishart distribution)

  2. At "regular point":

$$t_f = \frac{f(\bar{x},S)^2}{\text{Var}_{\bar{x},S}[f(\bar{x},S)]} \longrightarrow_d \chi^2_1$$

- Several invariants:

$$H_{f_1,\dots,f_k} : \vec{f}(\mu,\Sigma) = \big(f_1(\mu,\Sigma),\dots,f_k(\mu,\Sigma)\big)^t = 0$$

At "regular point":

$$t_{\vec{f}} = \vec{f}(\mu,\Sigma)^t \left(\text{Var}_{\bar{x},S}[\vec{f}(\bar{x},S)]\right)^- \vec{f}(\mu,\Sigma) \longrightarrow_d \chi^2_{\text{rank}(\nabla \vec{f}(\mu,\Sigma))}$$

Ideal of invariants

- Polynomial ring

$$R = \mathbb{R}[\mu_h, \sigma_{ij} \mid h = 1, \dots, p;\ 1 \leq i \leq j \leq p]$$
$$= \mathbb{R}[\mu_1, \dots, \mu_p, \sigma_{11}, \dots, \sigma_{1p}, \sigma_{22}, \dots, \sigma_{pp}]$$

- Set of model invariants

$$I_M = \{f \in R \mid f(\mu, \Sigma) = 0 \quad \forall (\mu, \Sigma) \in M\}$$

forms an **ideal** in the polynomial ring R.

- Hilbert's basis theorem

$$\exists f_1, \dots, f_k \in R : \quad I_M = \langle f_1, \dots, f_k \rangle$$

- Computational algebra provides the technique of **implicitization** to compute ideal-generators $\{f_1, \dots, f_k\}$.

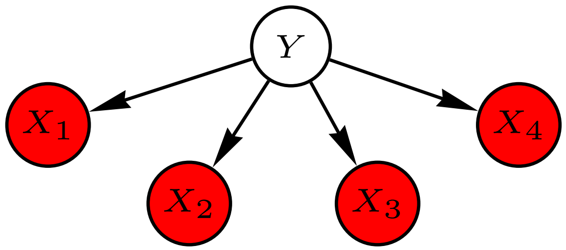
Example: Factor analysis

- From n "experiments", obtain expression levels of $p = 4$ genes, and compute sample covariance matrix:

$$S = \begin{pmatrix} 1.00 & 0.72 & 0.63 & 0.54 \\ 0.72 & 1.00 & 0.56 & 0.48 \\ 0.63 & 0.56 & 1.00 & 0.42 \\ 0.54 & 0.48 & 0.42 & 1.00 \end{pmatrix} = \begin{pmatrix} 0.19 & 0 & 0 & 0 \\ 0 & 0.36 & 0 & 0 \\ 0 & 0 & 0.51 & 0 \\ 0 & 0 & 0 & 0.64 \end{pmatrix} + \begin{pmatrix} 0.9 \\ 0.8 \\ 0.7 \\ 0.6 \end{pmatrix} \begin{pmatrix} 0.9 \\ 0.8 \\ 0.7 \\ 0.6 \end{pmatrix}^t$$

- Unknown (hidden) factor Y explains correlations (Xie & Bentler, 2003):

$$X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3 \perp\!\!\!\perp X_4 \mid Y$$



One-factor model

- Factor analysis model ($m = 1$ factor):

$$\{\mathcal{N}_p(\mu, \Sigma) \mid \mu \in \mathbb{R}^p,\ \Sigma \in F_p\}, \quad M = \mathbb{R}^p \times F_p,$$

  with covariance matrix parameter space

$$F_p = \{\Delta + \underbrace{\Lambda\Lambda^t}_{\text{rank } \leq 1} \mid \Delta > 0 \text{ diag.},\ \Lambda \in \mathbb{R}^p\}$$

- Model is algebraic: $F_p = \mathbf{f}\big((0,\infty)^p \times \mathbb{R}^p\big)$ where

$$\mathbf{f}_{ij}(\Delta, \Lambda) = \begin{cases} \delta_{ii} + \lambda_i^2 & \text{if } i = j, \\ \lambda_i \lambda_j & \text{if } i \neq j \end{cases}$$

- Model dimension: $\dim(F_p) = \min\left\{2p, \binom{p+1}{2}\right\}$

Tetrads

- Decomposition: $\Sigma = \Delta + \Lambda\Lambda^t =$ diag $+$ (rank $\leq 1) \in F_p$
- Off-diagonal 2×2-minors = **tetrads** vanish if $p \geq 4$ variables observed,

  e.g. $\det(\Sigma_{12\times 34}) = \sigma_{13}\sigma_{24} - \sigma_{14}\sigma_{23} = 0 \quad \forall\, \Sigma \in F_p$

- Statistical use of tetrads: e.g. Software TETRAD (Spirtes et al. 2000)
- **Theorem (De Loera/Sturmfels/Thomas, 1995):** *If $p \leq 3$, then $I_{F_p} = \{0\}$. If $p \geq 4$, then the set of $2\binom{p}{4}$ tetrads*

  $$\mathcal{T}_p = \{\underline{\sigma_{ij}\sigma_{k\ell}} - \sigma_{ik}\sigma_{j\ell}, \underline{\sigma_{i\ell}\sigma_{jk}} - \sigma_{ik}\sigma_{j\ell} \mid 1 \leq i < j < k < \ell \leq p\}$$

  is the reduced Gröbner basis of I_{F_p} wrto a certain monomial order.

- Example: $I_{F_4} = \langle \underline{\sigma_{12}\sigma_{34}} - \sigma_{13}\sigma_{24},\ \underline{\sigma_{14}\sigma_{23}} - \sigma_{13}\sigma_{24} \rangle$

More than one factor

- Factor analysis model with $m \geq 2$ factors:

$$\Sigma \in F_{p,m} = \{\Delta + \Lambda\Lambda^t \,:\, \Sigma > 0 \text{ diag.},\ \Lambda \in \mathbb{R}^{p\times m}\}.$$

- $F_{p,m} = \mathbf{f}\big((0,\infty)^p \times \mathbb{R}^{p\times m}\big)$ where

$$\mathbf{f}_{ij}(\Delta,\Lambda) = \begin{cases} \delta_{ii} + \sum_{k=1}^m \lambda_{ik}^2 & \text{if } i = j \\ \sum_{k=1}^m \lambda_{ik}\lambda_{jk} & \text{if } i \neq j \end{cases},$$

- Model dimension: $\dim(F_{p,m}) = \min\left\{p(m+1) - \binom{m}{2}, \binom{p+1}{2}\right\}$

- See Drton, Sturmfels, Sullivant (2005) for details.

Algebra Part 3: Singularities & identifiability

$$f_{\mu,\Sigma}(x) = \frac{1}{\sqrt{(2\pi)^p \det(\Sigma)}} \exp\left\{-\tfrac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right\}$$

$$= c(\mu,\Sigma) \exp\left\{\mu^t \Sigma^{-1} x - \tfrac{1}{2}\operatorname{tr}\left(\Sigma^{-1} x x^t\right)\right\}$$

$\Longrightarrow$ saturated model of all multivariate normal distributions

forms **exponential family**:

sufficient statistics: $x \in \mathbb{R}^p$, $xx^t \in \mathbb{R}^{p\times p}$

natural parameters: $\Sigma^{-1}\mu \in \mathbb{R}^p$, $\Sigma^{-1} \in \mathbb{R}^{p\times p}$

- Gaussian model $\mathbf{M}$ is **curved exponential family** if parameter space M = smooth manifold.
- **Singularities** in M $\Longrightarrow$ Model $\mathbf{M}$ is **not** a CEF.

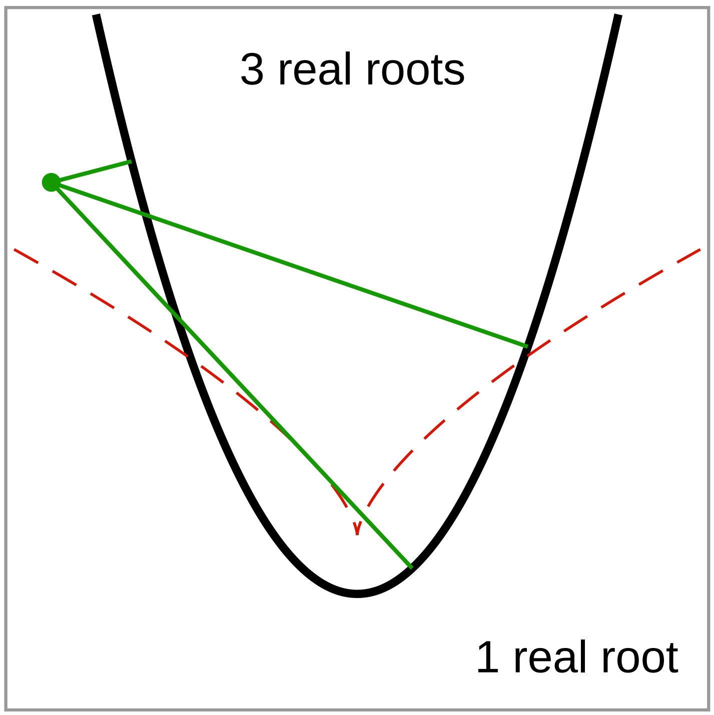
  (usual asymptotic distribution theory for MLE may fail at singularities)

Toy example 1: Parabola ($\Sigma = I_2$ known)

Mean vector on algebraic curve C; covariance matrix = identity matrix:

$$\{\mathcal{N}_2(\mu, \Sigma) \mid \mu \in C,\ \Sigma = I_2\}, \quad M = C \times \{I_2\}$$

Parabola: $C = \{\mu \mid \mu_2 = \mu_1^2\}$; Parameterization $\mathbf{f}(\theta) = (\theta, \theta^2)$

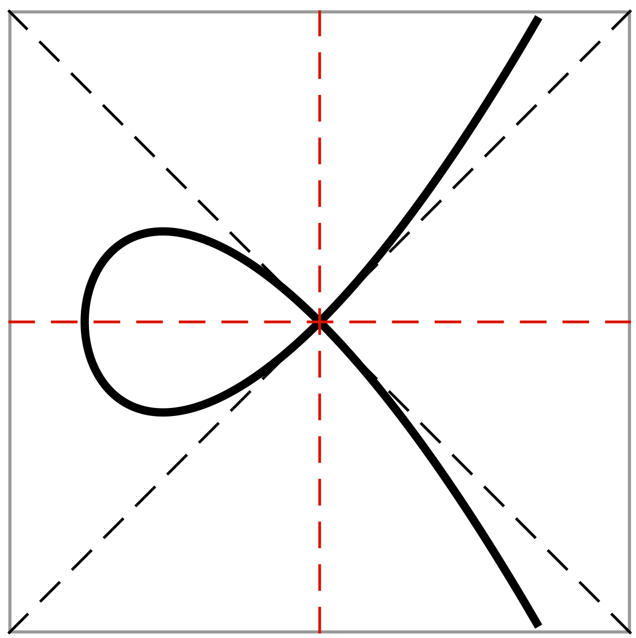


Toy example 2: "Folium of Descartes" ($\Sigma = I_2$)

Mean vector on algebraic curve C; covariance matrix = identity matrix:

$$\{\mathcal{N}_2(\mu, \Sigma) \mid \mu \in C,\ \Sigma = I_2\}, \quad M = C \times \{I_2\}$$

Curve $C = \{\mu \mid \mu_2^2 = \mu_1^3 + \mu_1^2\}$; Parametr. $\mathbf{f}(\theta) = [\theta^2 - 1, \theta(\theta^2 - 1)]$

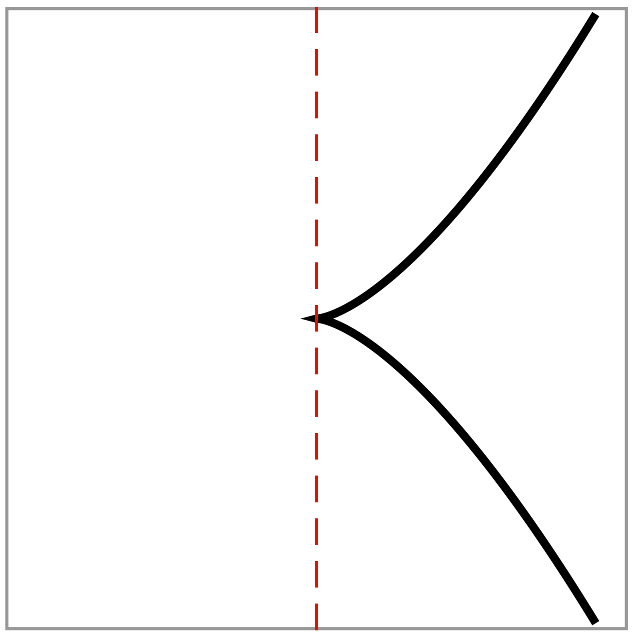


Toy example 3: Neil's parabola ($\Sigma = I_2$)

Mean vector on algebraic curve C; covariance matrix = identity matrix:

$$\{\mathcal{N}_2(\mu, \Sigma) \mid \mu \in C,\ \Sigma = I_2\}, \quad M = C \times \{I_2\}$$

Curve $C = \{\mu \mid \mu_2^2 = \mu_1^3\}$; Parameterization $\mathbf{f}(\theta) = (\theta^2, \theta^3)$



Note: If Σ unknown then MLE not obtained by minimizing a distance.

Singularities

- Consider algebraic Gaussian model $\mathbf{M}$ with ideal of model invariants

$$I_M = \langle f_1, \ldots, f_k \rangle \subset \mathbb{R}[\mu_h, \sigma_{ij} \mid h = 1, \ldots, p;\ 1 \leq i \leq j \leq p]$$

- Jacobi-matrix

$$\nabla f(\mu, \Sigma) = [\nabla f_1(\mu, \Sigma), \ldots, \nabla f_k(\mu, \Sigma)] \in \mathbb{R}^{[p+\binom{p}{2}]\times k}$$

- The singular locus of the **variety** $V(I_M) \supset M$ is the set

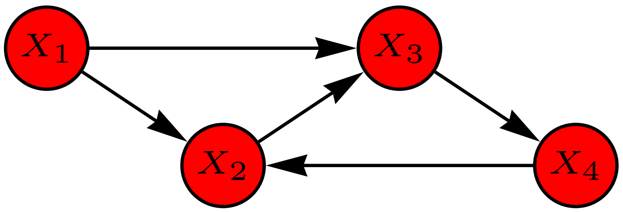
$$V_{\text{sing}}(I_M) = \{(\mu, \Sigma) \in V(I_M) \mid \text{rank}[\nabla f(\mu, \Sigma)] < \text{codim}(M)\}$$

- $V(I_M) \setminus V_{\text{sing}}(I_M)$ is smooth manifold

$$\overset{i.g.}{\not\Longrightarrow} \quad M \setminus V_{\text{sing}}(I_M) \text{ smooth manifold (CEF).}$$

- `Singular` provides useful routines for computing singular locus.

Example: Cyclic graphical model



- Regression equations

$$X_1 = \varepsilon_1, \qquad X_3 = \beta_{31}X_1 + \beta_{32}X_2 + \varepsilon_3,$$
$$X_2 = \beta_{21}X_1 + \beta_{42}X_4 + \varepsilon_2, \qquad X_4 = \beta_{43}X_3 + \varepsilon_4.$$

- Algebraic Gaussian model with par. space $M = \mathbb{R}^4 \times \mathbf{f}(\Theta)$, where

$$\mathbf{f}(\omega,\beta) = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ -\beta_{21} & 1 & 0 & -\beta_{24} \\ -\beta_{31} & -\beta_{32} & 1 & 0 \\ 0 & 0 & -\beta_{43} & 1 \end{pmatrix}}_{B}{}^{-1} \underbrace{\begin{pmatrix} \omega_{11} & 0 & 0 & 0 \\ 0 & \omega_{22} & 0 & 0 \\ 0 & 0 & \omega_{33} & 0 \\ 0 & 0 & 0 & \omega_{44} \end{pmatrix}}_{\Omega} B^{-t}$$

and $\quad \Theta = (0,\infty)^4 \times \{\beta \in \mathbb{R}^5 \mid \beta_{32}\beta_{24}\beta_{43} \neq 1\}$

Singularities of cyclic graphical model

- Model is embedded in hypersurface of degree 6 given by

$$f = \sigma_{13}\sigma_{14}^3\sigma_{23}^2 - 2\sigma_{13}^2\sigma_{14}^2\sigma_{23}\sigma_{24} + \sigma_{13}^3\sigma_{14}\sigma_{24}^2 - \sigma_{12}\sigma_{14}^3\sigma_{23}\sigma_{33}$$
$$\cdots + \sigma_{11}\sigma_{13}\sigma_{14}\sigma_{22}\sigma_{33}\sigma_{44} + \sigma_{12}^2\sigma_{13}^2\sigma_{34}\sigma_{44} - \sigma_{11}\sigma_{13}^2\sigma_{22}\sigma_{34}\sigma_{44}.$$

(19 terms)

- Singular locus:

$$V_{\text{sing}}(I_M) \cap M = \{(\mu, \Sigma) \in M \mid \sigma_{13} = \sigma_{14} = 0\}$$

- Singularities with $\sigma_{13} = \sigma_{14} = 0$ arise if $\beta_{31} + \beta_{21}\beta_{32} = 0$.

- One can choose $(\omega_0, \beta_0) \in \mathbf{f}^{-1}\big(V_{\text{sing}}(I_M) \cap M\big)$ such that by studying tangent vectors at $\mathbf{f}(\omega_0, \beta_0) \Longrightarrow$ model is not CEF

Parameter identifiability

- Algebraic Gaussian model $\mathbf{M}$ with parameter space $M = \mathbf{f}(\Theta)$
- **Identifiability** of $\mathbf{M}$ $:\Longleftrightarrow$ $\mathbf{f} : \Theta \to M$ injective
- **Global identifiability** of $\mathbf{M}$ at $\theta_0 \in \Theta$ with $(\mu_0, \Sigma_0) = \mathbf{f}(\theta_0)$:

$$\mathbf{f}^{-1}(\mu_0, \Sigma_0) = \{\theta_0\}$$

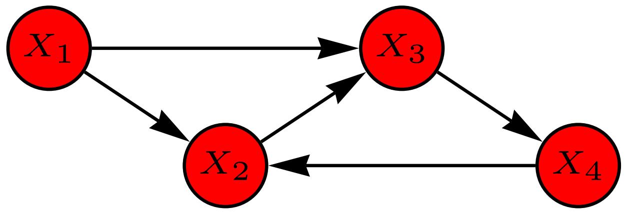
- **Local identifiability** of $\mathbf{M}$ at $\theta_0 \in \Theta$ with $(\mu_0, \Sigma_0) = \mathbf{f}(\theta_0)$:

$$\exists \varepsilon > 0 \quad \text{s.t.} \quad \mathbf{f}^{-1}(\mu_0, \Sigma_0) \cap B_\varepsilon(\theta_0) = \{\theta_0\}$$

- Can check identifiability by solving the system of polynomial/rational equations:

$$(\mu_0, \Sigma_0) = \mathbf{f}(\theta)$$

Example: Cyclic graphical model (cont.)



- Model $\subset$ hypersurface (degree 6)

  with singular locus:

$$V_{\text{sing}}(I_M) \cap M = \{(\mu, \Sigma) \in M \mid \sigma_{13} = \sigma_{14} = 0\}$$

$$= \mathbf{f}\Big(\{(\omega, \beta) \in \Theta \mid \beta_{31} + \beta_{21}\beta_{32} = 0\}\Big)$$

- One can choose $(\omega_0, \beta_0) \in \mathbf{f}^{-1}\big(V_{\text{sing}}(I_M) \cap M\big)$ such that

$$\exists (\omega_1, \beta_1) \neq (\omega_0, \beta_0) : \mathbf{f}(\omega_0, \beta_0) = \mathbf{f}(\omega, \beta)$$

$\Longrightarrow$ $\exists$ "singular" point (ω_0, β_0) at which model is not globally identified.

A nonrepresentative sample of literature

- N. Sugiura & A.K. Gupta (1987). Maximum Likelihood Estimates for Behrens-Fisher Problem. *J. Japan Statist. Soc.* 17(1): 55–60.
- MD & T.S. Richardson (2004). Multimodality of the Likelihood in the Bivariate Seemingly Unrelated Regressions Model. *Biometrika* 91(2): 383–392.
- MD (2005). Computing All Roots of the Likelihood Equations of Seemingly Unrelated Regressions. *J. Symbolic Comput.*, to appear. `arXiv:math.ST/0508437`.
- MD, B. Sturmfels & S. Sullivant (2005). Algebraic Factor Analysis: Tetrads, Pentads and Beyond. `arXiv:math.ST/0509390`.
- MD, M. Eichler, T.S. Richardson (2006). Identification and Likelihood Inference for Recursive Linear Models with Correlated Errors.