SMTDA 2016

Bayes and Krige: generalities

Giovanni Pistone [1] Grazia Vicario [2]

[1]de Castro Statistics - Collegio Carlo Alberto

[2]DISMA Politecnico di Torino

June 3, 2016

Abstract

[plain]
Bayes and Krige: generalities

Giovanni Pistone
de Castro Statistics, Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri, Italy
`giovanni.pistone@carloalberto.org`
`www.giannidiorestino.it`

Grazia Vicario
Department of Mathematical Sciences, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
`grazia.vicario@polito.it`
`calvino.polito.it/~vicario/`

In this paper we focus on the use of Kriging models for predicting at untried points from the response values at the tried locations. The underlying Gaussian model requires the modelistration of the covariance structure. In a previous paper we have discussed to this aim generalities about the use of variograms to parameterize Gaussian models. In fact, Geostatisticians, pioneers and practitioners of Kriging, strongly support the variogram considering it more informative of the correlation structure. In particular computations for the case of jointly Gaussian $Y_1, \ldots, Y_n$ with constant variance $\sigma^2 = \mathrm{Var}\,(Y_i)$, $i = 1, \ldots, n$ are performed. In such a case, the model can be parameterized by the common variance σ^2 and the variogram matrix Γ that carries $n(n-1)/2$ degrees of freedom and is conditionally negative definite. The set of variogram matrices is related with a convex set called elliptope (ellipsoid+polytope). The discussion of the domain for the variogram matrices is instrumental when viewing the problem in Bayesian terms. Opposite to the conventional geostatistical Kriging approach that commonly ignores the effect of the uncertainty in the covariance structure on subsequent predictions, a Bayesian approach will provide a general methodology for taking into account the uncertainty about parameters on subsequent predictions. Hence a-priori on the variogram matrices is demanded. We plan to discuss a number of simple typical problems with parameterized subset of variogram matrices and small dimension.

Outline

- Variogram of a normal vector, constant variance
- Kriging and variogram
- Characterisation of variogram matrices
- Inverting the variogram matrix
- Set of correlation matrices
- Parametrization of correlation matrices
- Decomposition of the state vector according to the variogram

Variogram of a normal vector with constant variance

- We consider a Gaussian n-vector, $n \geq 2$, with mean $\boldsymbol{\mu} = \mu \mathbf{1}$ and variance matrix $\Sigma = [\sigma_{ij}]_{ij=1}^{n}$ with constant diagonal $\sigma_{ii} = \sigma^2$, $i = 1, \dots, n$.
- The assumption on the mean and the diagonal terms is intended to be a weak *stationarity assumption*.
- Hence, $Y \sim \mathrm{N}_n(\mu \mathbf{1}, \sigma^2 R)$, where μ is a general mean value and $R = [\rho_{ij}]_{i,j=1}^{n}$ is a correlation matrix.
- The *variogram* of Y is the $n \times n$ matrix $\Gamma = [\gamma_{ij}]_{i,j=1}^{n}$

$$2\gamma_{ij} = \mathrm{Var}\left(Y_i - Y_j\right) = (\boldsymbol{e}_i - \boldsymbol{e}_j)' \sigma^2 R (\boldsymbol{e}_i - \boldsymbol{e}_j) = \\ \sigma^2 \left(\rho_{ii} + \rho_{jj} - 2\rho_{ij}\right) = 2\sigma^2(1 - \rho_{ij}).$$

- In matrix form

$$\Gamma = \sigma^2(\mathbf{1}\mathbf{1}' - R).$$

Kriging model and variogram

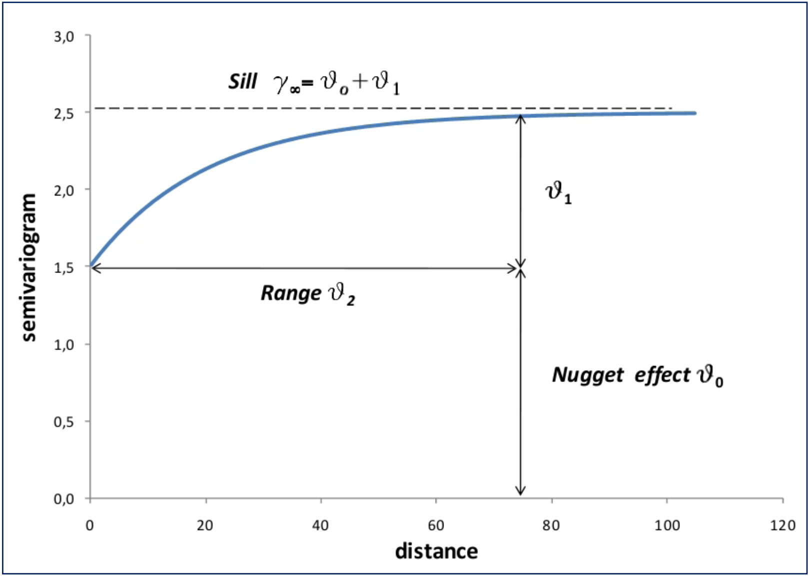
- In Geostatistics applications each component Y_i is associated to a location x_i, $i = 1, \ldots, n$ in a given region X and the covariances are assumed to be a function of the distance between locations: $\Sigma_{i,j} = C(d(x_i, x_j))$.
- In this case the diagonal is constant,

$$\Sigma_{i,i} = C(d(x_i, x_i)) = C(0) = \sigma^2$$

  Moreover it is commonly assumed that the mean is constant, $\mu\mathbf{1}$.
- *First* Krige's modeling idea is to assume positive correlation and assume the variogram to be an *increasing function* γ of the distance, so that to model a variability that increases with the distance and is bounded by the general variance:

$$0 < \frac{1}{2} \operatorname{Var}(Y_i - Y_j) = \Gamma_{ij} = \gamma(d(x_1, x_j)) = \sigma^2(1 - R(d(x_1, x_j)) < 2\sigma^2$$

Semi-variogram function



Krige's prediction

- The parameters in the Krige's model are μ, σ^2, R and are usually estimated over a suitable parametric model. We do not discuss here the modelling aspect, but we adopt a general non-parametric attitude, where μ is real number, σ^2 is a positive real number, R is a positive definite matrix with unit diagonal.
- *Second* Krige's idea is to predict the value Y_{x_0} an untried location x_0 with the conditional expectation based on some estimate of the parameters: if $I = \{1, \dots, n\}$,

$$\widehat{Y_0 - \mu} = \Sigma_{0,I} \Sigma_{I,I}^{-1} (Y_I - \mu \mathbf{1}_I), \quad \text{with} \quad \Sigma = \begin{bmatrix} \Sigma_{I,I} & \Sigma_{I,0} \\ \Sigma_{n,I} & \sigma^2 \end{bmatrix}$$

- Note that the set of data that give the same prediction is an affine plane in $\mathbb{R}^n$.
- **We are going to discuss how to express the prediction formula for $\widehat{Y_0 - \mu}$ as a function of the variogram Γ.**

First properties of variogram matrix Γ

$$\boxed{\Gamma = \sigma^2(\mathbf{1}\mathbf{1}' - R) = \sigma^2\mathbf{1}\mathbf{1}' - \Sigma}$$

- $\frac{1}{n}\mathbf{1}\mathbf{1}'$ is the orthogonal projector on the space of constant vectors $\mathrm{Span}\,(\mathbf{1})$.
- $\Gamma = 0$ if, and only if, $R = \mathbf{1}\mathbf{1}'$.
- $\boldsymbol{x} = \boldsymbol{w} + \alpha\mathbf{1}$ with $\boldsymbol{w}'\mathbf{1} = 0$

$$\begin{aligned} \boldsymbol{x}'\Gamma\boldsymbol{x} &= \boldsymbol{w}'\Gamma\boldsymbol{w} + 2\alpha\boldsymbol{w}'\Gamma\mathbf{1} + \alpha^2\mathbf{1}'\Gamma\mathbf{1} \\ &= -\sigma^2\boldsymbol{w}'R\boldsymbol{w} - 2\alpha\sigma^2\boldsymbol{w}'R\mathbf{1} + \sigma^2(n^2 - \alpha^2\mathbf{1}'R\mathbf{1}). \end{aligned}$$

- Γ is *conditionally negative definite* (i.e. when $\alpha = 0$)
- $\mathbf{1}'\Gamma\mathbf{1} = \sigma^2(n^2 - \mathbf{1}'R\mathbf{1})$.

Characterisation

Is a matrix Γ a variogram?

A nonzero matrix Γ is a variogram of some covariance matrix of the form $\Sigma = \sigma^2 R$, with $\sigma^2 > 0$ and R a correlation matrix, if, and only if, the three following condition hold:

1. Γ is symmetric, and has zero diagonal;
2. Γ is conditionally negative definite;
3. $\sup\{\boldsymbol{x}'\Gamma\boldsymbol{x} | \boldsymbol{x}'\mathbf{1} = 1\} \leq \sigma^2$.

- Note the lower bound for σ^2.
- Γ carries $n(n-1)/2$ degrees of freedom.

Inverse Variogram matrix Γ^{-1}

1. If $\Sigma = \sigma^2 R \in \mathbb{S}_=$ is invertible, then $\sigma^{-2} - \mathbf{1}'\Sigma^{-1}\mathbf{1} \neq 0$ and $\Gamma = \sigma^2(\mathbf{1}\mathbf{1}' - R)$ is invertible, with
$$\Gamma^{-1} = -\Sigma^{-1} - (\sigma^{-2} - \mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1}\Sigma^{-1}\mathbf{1}\mathbf{1}'\Sigma^{-1} \ . \tag{1}$$
2. If $\Sigma = \sigma^2 R \in \mathbb{S}_+$ is invertible, then $\Gamma = \sigma^2(\mathbf{1}\mathbf{1}' - R)$ is invertible. Moreover $1 \neq \sigma^2\mathbf{1}'\Gamma^{-1}\mathbf{1}$, and
$$\Sigma^{-1} = -\Gamma^{-1} - (\sigma^{-2} - \mathbf{1}'\Gamma^{-1}\mathbf{1})^{-1}\Gamma^{-1}\mathbf{1}\mathbf{1}'\Gamma^{-1} \tag{2}$$

- The proof uses the Sherman-Morrison formula and properties of the matrix R.

Elliptope

- In a non parametric approach we want to know the shape of the bounded set of correlation matrices. This convex set is called *elliptope*.
- All principal minors of R are nonnegative, in particular with 3 locations

$$\det(R) = \det\left(\begin{bmatrix} 1 & x & y \\ x & 1 & z \\ y & z & 1 \end{bmatrix}\right) = 1 - x^2 - y^2 - z^2 + 2xyz \geq 0$$

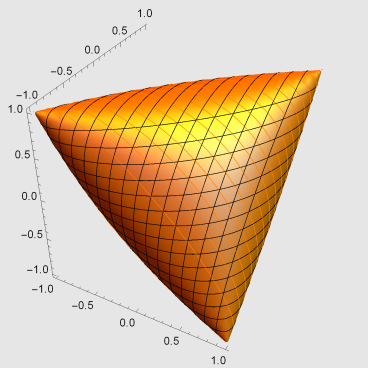
- all horizontal section $z = \text{const}$ are ellipses

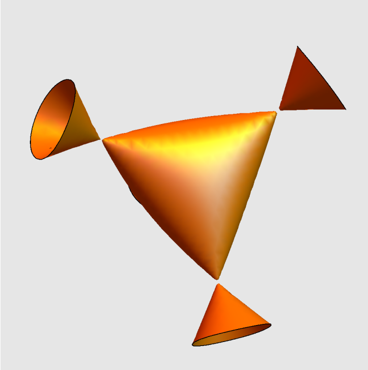
$$1 - x^2 - y^2 + 2cxy \geq c^2$$

Same for other sections.

- The volume is computable: uniform apriori. Simulation is feasible.
- $R = A'A$ where the columns of A are unit vectors: an other possible apriori. Simulation is feasible.

The 3-elliptope





Cholesky decomposition of a correlation matrix

- For a correlation matrix $R = T'T$ with

$$T = \begin{bmatrix} \boldsymbol{t}_1' \\ \boldsymbol{t}_2' \\ \boldsymbol{t}_3' \end{bmatrix} = \begin{bmatrix} \sqrt{1 - t_{12}^2 - t_{13}^2} & t_{12} & t_{13} \\ 0 & \sqrt{1 - t_{23}^2} & t_{23} \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{t}_i \in 0^{i-1} \times S_{n-1+1}^{+}$$

-

$$R = \begin{bmatrix} 1 & \sqrt{1 - t_{12}^2 - t_{13}^2}\, t_{12} & \sqrt{1 - t_{12}^2 - t_{13}^2}\, t_{13} \\ \sqrt{1 - t_{12}^2 - t_{13}^2}\, t_{12} & 1 & t_{12} t_{13} + \sqrt{1 - t_{23}^2}\, t_{23} \\ \sqrt{1 - t_{12}^2 - t_{13}^2}\, t_{13} & t_{12} t_{13} + \sqrt{1 - t_{23}^2}\, t_{23} & 1 \end{bmatrix}$$

$$\det(R) = (1 - t_{12}^2 - t_{13}^2)(1 - t_{23}^2)$$

Projecting on Span $(\mathbf{1})^{\perp}$

- The approach with parameters σ^2, Γ does not appear promising in term of ease of computation.
- We now change our point of view to consider the same problem from a different angle. In fact, we can associate the variogram with the state space description of the Gaussian vector.

1. The matrix Γ is a variogram matrix if, and only if, the matrix

$$\Sigma_0 = -\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)' \Gamma \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}'\right) \tag{3}$$

   is symmetric, positive definite and with constant diagonal. In such a case, the variogram of Σ_0 is Γ.
2. If $Y_0 \sim \mathrm{N}_n(0, \Sigma_0)$, then its variogram is Γ and it is supported by Span $(\mathbf{1})^{\perp}$.

Decomposition of the state vector

- Let $Y \sim \mathsf{N}_n(\boldsymbol{\mu}, \Sigma)$, $\sigma = \sigma^2 R \in \mathbb{S}_+$ with variogram $\Gamma = \sigma^2(\mathbf{1}\mathbf{1}' - R)$.
- Let $Y_0 = \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}'\right) Y$ be the projection of Y onto $\mathrm{Span}\,(\mathbf{1})^{\perp}$ so that we can write $Y = Y_0 + \overline{Y}$, where each component of $\overline{Y}$ is the empirical mean $\frac{1}{n}\mathbf{1}'Y$.

1. The distribution of Y_0 is Gaussian and depends on the mean and the variogram only.
2. The distribution of $\frac{1}{n}\mathbf{1}'Y$, conditionally to Y_0 is Gaussian.

Some references

- Monique Laurent and Svatopluk Poljak. On the facial structure of the set of correlation matrices.

  SIAM J. Matrix Anal. Appl., 17(3):530–547, 1996

- Giovanni Pistone and Grazia Vicario. A note on (semi)variogram.

  In press. arXiv:1503.07686v2, april 2015

- Francesco Rapisarda, Damiano Brigo, and Fabio Mercurio. Parameterizing correlations: a geometric interpretation.

  IMA J. Manag. Math., 18(1):55–73, 2007

- Riccardo Rebonato. *Volatility and Correlation.*

  John Wiley & Sons Ltd, Chichester, England, 2nd edition, 2004.

  The Perfect Hedger and the Fox

- Peter J. Rousseeuw and Geert Molenberghs. The shape of correlation matrices.

  Amer. Statist., 48(4):276–279, 1994

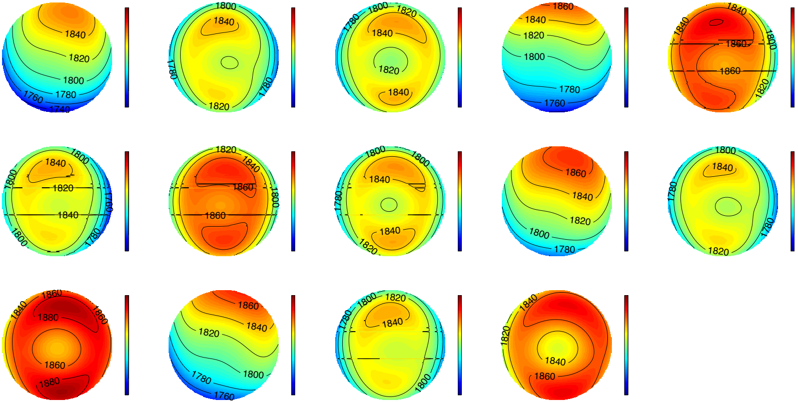
- Suela Ruffa, Grazia Vicario, and Giovanni Pistone. Analysis of the covariance structure in manufactured parts.

  Communications in Statistics - Theory and Methods, 44(21):4540–4551

We thank Guillaume Kon Kam King (CCA, Moncalieri) and Luigi Malagò (CCA, Moncalieri) for suggesting relevant references.

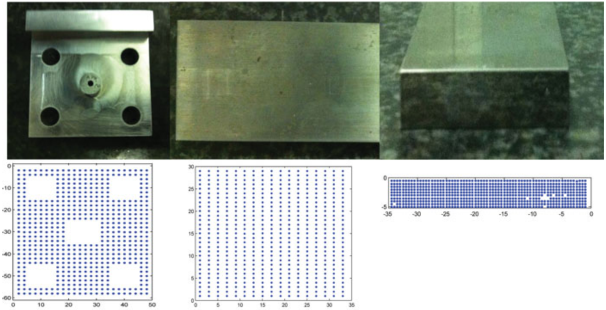
EXTRAS

Example: Wafer diffusion



Giovanni Pistone and Grazia Vicario. Kriging prediction from a circular grid: application to wafer diffusion. *Appl. Stochastic Models Bus. Ind.*, 29(4):350–361, July 2013

Example: CMM measurements



(a) (b) (c)

Figure 1. The three considered surfaces: (a) milled, (b) grinded, and (c) lapped, and the corresponding sampled points on uniform rectangular grids. In (c), some points are missing because they were detected by default as outliers by the CMM software.

Suela Ruffa, Grazia Vicario, and Giovanni Pistone. Analysis of the covariance structure in manufactured parts. *Communications in Statistics - Theory and Methods*, 44(21):4540–4551

$(\sigma^2, \Gamma) \leftrightarrow (\sigma^2, R) \leftrightarrow \Sigma$

1. The mapping from $\Sigma \in \mathbb{S}_=$ to the couple $(\sigma^2, \Gamma) \in \mathbb{R}_> \times \mathbb{V}$ factors as

$$\mathbb{S}_= \ni \Sigma \mapsto (\frac{1}{n}\operatorname{Tr}(\Sigma), \left(\frac{1}{n}\operatorname{Tr}(\Sigma)\right)^{-2}\Sigma) = (\sigma^2, R) \in]0, \infty[\times\mathcal{R}$$

and

$$\begin{aligned}]0, \infty[\times\mathcal{R} \ni (\sigma^2, R) \mapsto (\sigma^2, \sigma^2(\mathbf{1}\mathbf{1}' - R)) = \\ (\sigma^2, \Gamma) \in \left\{(\sigma^2, \Gamma) \middle| \Gamma \in \mathbb{V}, \sup\left\{\boldsymbol{x}'\Gamma\boldsymbol{x} | \boldsymbol{x}'\mathbf{1} = 1\right\} \leq \sigma^2\right\}. \end{aligned}$$

- Inverse is

$$\begin{aligned} \left\{(\sigma^2, \Gamma) \middle| \Gamma \in \mathbb{V}, \sup\left\{\boldsymbol{x}'\Gamma\boldsymbol{x} | \boldsymbol{x}'\mathbf{1} = 1\right\} \leq \sigma^2\right\} \ni (\sigma^2, \Gamma) \mapsto \\ \sigma^2\mathbf{1}\mathbf{1}' - \Gamma = \Sigma \in \mathbb{S}_= \end{aligned}$$

Sherman-Morrison formula

If the matrix A is invertible, then $\mathbf{1}\mathbf{1}' - A$ is invertible if, and only if, $\mathbf{1}'A^{-1}\mathbf{1} \neq 1$. In such a case,

$$\det(\mathbf{1}\mathbf{1}' - A) = (-1)^n \det A(1 - \mathbf{1}'A^{-1}\mathbf{1}) \ ,$$

$$(\mathbf{1}\mathbf{1}' - A)^{-1} = -A^{-1} - (1 - \mathbf{1}'A^{-1}\mathbf{1})^{-1}A^{-1}\mathbf{1}\mathbf{1}'A^{-1}.$$

Proof of SM formula

$$\begin{aligned}\det(\mathbf{1}\mathbf{1}' - A) &= \det(-A) + \sum_{j=1}^{n}\sum_{i=1}^{n}(-A)^{ij} \\ &= (-1)^n \det A - (-1)^{n-1}\sum_{i,j=1}^{n} A^{ij} \\ &= (-1)^n \det A - (-1)^{n-1}\mathbf{1}(\operatorname{adj} A)\mathbf{1}' \ .\end{aligned}$$

If $\det A \neq 0$,

$$\det(\mathbf{1}\mathbf{1}' - A) = (-1)^n(\det A)(1 - \mathbf{1}'A^{-1}\mathbf{1}) \neq 0$$

if, and only if,

$$1 - \mathbf{1}'A^{-1}\mathbf{1} \neq 0 \ .$$

We conclude by checking that

$$(\mathbf{1}\mathbf{1}' - A)(-A^{-1} - \alpha A^{-1}\mathbf{1}\mathbf{1}'A^{-1}) = I$$

if, and only if, $\alpha = (1 - \mathbf{1}'A^{-1}\mathbf{1})^{-1}$.

Properties of the correlation matrix$R$

Let R be a correlation matrix and assume $\det(R) \neq 0$. Let $\lambda_j > 0$, $j = 1, \ldots, n$, be the eigenvalues of R and $\boldsymbol{u}_j$ a set of unit eigenvectors.

1. $\operatorname{Tr} R = \sum_{j=1}^{n} \lambda_j = n$ and $\det(R) = \prod_{j=1}^{n} \lambda_j \leq 1$, with equality if, and only if, $R = I$.
2. $\operatorname{Tr} R^{-1} = \sum_{j=1}^{n} \lambda_j^{-1} \geq n$ and $\det(R)^{-1} \geq 1$ with equality if, and only if, $R = I$.
3. $\mathbf{1}' R^{-1} \mathbf{1} \neq 1$.

Proof

1. $n = \operatorname{Tr} R = \sum_{j=1}^{n} \lambda_j$. From $\det R = \prod_{j=1}^{n} \lambda_j$, as the arithmetic mean is larger than the the geometric mean,
$$1 = \frac{\sum_{j=1}^{n} \lambda_j}{n} \geq \left(\prod_{j=1}^{n} \lambda_j \right)^{\frac{1}{n}} = (\det R)^{\frac{1}{n}} \ ,$$
with equality if, and only if $\lambda_j = 1$ for all $j = 1, \dots, n$, that is, $R = I$.
2. The geometric mean is larger or equal than the harmonic mean,
$$(\det R)^{\frac{1}{n}} = \left(\prod_{j=1}^{n} \lambda_j \right)^{\frac{1}{n}} \geq n \left(\sum_{j=1}^{n} \lambda_j^{-1} \right)^{-1} \ ,$$
with equality if, and only if, $\lambda_j = 1$, $j = 1, \dots, n$. It follows $\frac{1}{n} \sum_{j=1}^{n} \lambda_j^{-1} \geq 1$.
3. As $R^{-1} = \sum_{j=1}^{n} \lambda_j^{-1} \boldsymbol{u}_j \boldsymbol{u}_j'$ and $\sum_{j=1}^{n} (\mathbf{1}' \boldsymbol{u}_j)^2 = \|\mathbf{1}\|^2 = n^2$,
$$1 = \mathbf{1}' R^{-1} \mathbf{1} = \sum_{j=1}^{n} \lambda_j^{-1} (\mathbf{1}' \boldsymbol{u}_j)^2 = n^2 \sum_{j=1}^{n} (\lambda_j)^{-1} \theta_j \ ,$$
where $\theta_j = (\mathbf{1}' \boldsymbol{u})^2 / n^2 \geq 0$ and $\sum_{j=1}^{n} \theta_j = 1$. From the convexity of $\lambda \mapsto \lambda^{-1}$ we obtain
$$1 = n^2 \sum_{j=1}^{n} (\lambda_j)^{-1} \theta_j \geq n^2 \left(\sum_{j=1}^{n} \lambda_j \theta_j \right)^{-1} \ ,$$
hence the contradiction
$$1 \leq \frac{1}{n^2} \sum_{j=1}^{n} \lambda_j \theta_j \leq \frac{1}{n^2} \max \left\{ \lambda_j \middle| j = 1, \dots, n \right\} \leq \frac{1}{n} \ .$$

Likelihood

- $$\det(\Sigma) = \det\left(\sigma^2 \mathbf{1}\mathbf{1}' - \Gamma\right) = \sigma^{2n} \det\left(\mathbf{1}\mathbf{1}' - \sigma^{-2}\Gamma\right) = \sigma^{2n}\left[\det\left(-\sigma^{-2}\Gamma\right) + \mathbf{1}' \operatorname{adj}\left(-\sigma^{-2}\Gamma\right)\mathbf{1}\right] = \det(-\Gamma) - \sigma^2 \mathbf{1}' \operatorname{adj}(-\Gamma)\,\mathbf{1}$$

- $$\boldsymbol{y}'\Sigma^{-1}\boldsymbol{y} = \boldsymbol{y}'\left(-\Gamma^{-1} - (\sigma^{-2} - \mathbf{1}'\Gamma^{-1}\mathbf{1})^{-1}\Gamma^{-1}\mathbf{1}\mathbf{1}'\Gamma^{-1}\right)\boldsymbol{y} = -\boldsymbol{y}'\Gamma^{-1}\boldsymbol{y} - (\sigma^{-2} - \mathbf{1}'\Gamma^{-1}\mathbf{1})^{-1}(\boldsymbol{y}'\Gamma^{-1}\mathbf{1})^2.$$

- $$\log p\left(\boldsymbol{y}\middle|\sigma^2, \Gamma\right) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\left(\det\left(\mathbf{1}\mathbf{1}' - \sigma^{-2}\Gamma\right)\right) - \frac{1}{2\sigma^2}\boldsymbol{y}'(\mathbf{1}\mathbf{1}' - \sigma^{-2}\Gamma)^{-1}\boldsymbol{y} = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\left(\det(-\Gamma) - \sigma^2\mathbf{1}'\operatorname{adj}(-\Gamma)\,\mathbf{1}\right) + \frac{1}{2}\boldsymbol{y}'\Gamma^{-1}\boldsymbol{y} + \frac{1}{2}(\sigma^2 - \mathbf{1}'\Gamma^{-1}\mathbf{1})^{-1}(\boldsymbol{y}'\Gamma^{-1}\mathbf{1})^2\,.$$

- $$d_H(\Gamma \mapsto \log\left(\det\left(\mathbf{1}\mathbf{1}' - \sigma^{-2}\Gamma\right)\right)) = \operatorname{Tr}\left((\sigma^2\mathbf{1}\mathbf{1}' - \Gamma)^{-1}H\right)\;;$$
$$d_H(\Gamma \mapsto \boldsymbol{y}'(\mathbf{1}\mathbf{1}' - \sigma^{-2}\Gamma)^{-1}\boldsymbol{y}) = \sigma^2\operatorname{Tr}\left((\sigma^2\mathbf{1}\mathbf{1}' - \Gamma)^{-1}\boldsymbol{y}\boldsymbol{y}'(\sigma^2\mathbf{1}\mathbf{1}' - \Gamma)^{-1}H\right)\;;$$

- $$-(\sigma^2\mathbf{1}\mathbf{1}' - \Gamma)^{-1} + (\sigma^2\mathbf{1}\mathbf{1}' - \Gamma)^{-1}\boldsymbol{y}\boldsymbol{y}'(\sigma^2\mathbf{1}\mathbf{1}' - \Gamma)^{-1}\quad\text{is diagonal}\,.$$

Cholesky decomposition

- A symmetric matrix A is positive definite if there exists an upper triangular matrix

$$T = \begin{bmatrix} \boldsymbol{t}_1' \\ \boldsymbol{t}_2' \\ \boldsymbol{t}_3' \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ 0 & t_{22} & t_{23} \\ 0 & 0 & t_{33} \end{bmatrix}, \quad t_{ii} \geq 0$$

-

$$A = T'T = [\boldsymbol{t}_i \cdot \boldsymbol{t}_j]_{ij} = \begin{bmatrix} t_{11}^2 & t_{11}t_{12} & t_{11}t_{13} \\ t_{11}t_{12} & t_{12}^2 + t_{22}^2 & t_{12}t_{13} + t_{22}t_{23} \\ t_{11}t_{13} & t_{12}t_{13} + t_{22}t_{23} & t_{13}^2 + t_{23}^2 + t_{33}^2 \end{bmatrix}$$

- $t_{11}t_{22}t_{33} \neq 0 \Leftrightarrow T$ is unique and invertible $\Leftrightarrow A$ is invertible
- A identifiable parametrization for non singular matrices.

Proof

1. If $\Gamma = \sigma^2(\mathbf{1}\mathbf{1}' - R)$ is a variogram matrix, then from Eq. (3) we have

$$\Sigma_0 = \sigma^2 \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}' \right)' R \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}' \right),$$

which is indeed positive definite. Let us compute the diagonal elements of Σ_0.

$$\begin{aligned}(\Sigma_0)_{ii} &= \sigma^2 \boldsymbol{e}_i' \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}' \right)' R \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}' \right) \boldsymbol{e}_i \\ &= \sigma^2 \left(\boldsymbol{e}_i - \frac{1}{n}\mathbf{1} \right)' R \left(\boldsymbol{e}_i - \frac{1}{n}\mathbf{1} \right) \\ &= \sigma^2 \left(\boldsymbol{e}_i' R \boldsymbol{e}_i - \frac{2}{n}\boldsymbol{e}_i R\mathbf{1} + \frac{1}{n^2}\mathbf{1}' R\mathbf{1} \right) \\ &= \sigma^2 \left(\frac{1}{n^2}\mathbf{1}' R\mathbf{1} - 1 \right)\end{aligned}$$

Viceversa, assume Σ_0 is a covariance matrix. As $\boldsymbol{e}_i - \boldsymbol{e}_j \in \text{Span}\,(\mathbf{1})^\perp$, the variogram of Σ_0 has elements

$$\begin{aligned}&(\boldsymbol{e}_i - \boldsymbol{e}_j)'\Sigma_0(\boldsymbol{e}_i - \boldsymbol{e}_j) = \\ &\qquad (\boldsymbol{e}_i - \boldsymbol{e}_j)' \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}' \right)' (-\Gamma) \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}' \right) (\boldsymbol{e}_i - \boldsymbol{e}_j) = \\ &\qquad\qquad -(\boldsymbol{e}_i - \boldsymbol{e}_j)'\Gamma(\boldsymbol{e}_i - \boldsymbol{e}_j) = -\gamma_{ii} - \gamma_{jj} + 2\gamma_{ij} = 2\gamma_{ij}.\end{aligned}$$

2. As $\mathbf{1}'(\boldsymbol{e}_i - \boldsymbol{e}_j) = 0$, then $\mathbf{1}' \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}' \right)' (-\Gamma) \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}' \right) \mathbf{1} = 0$, hence the distribution of Y_0 is supported by the space $\text{Span}\,(\mathbf{1})^\perp$.