Références

Contents

1 Cours de Géostatistiques **2**

2 Covariance Function Estimation **4**
2.1 Non Parametric Estimation 4
2.2 Parametric Estimation 5

3 Non Parametric Variogram Estimators **6**

4 Robust Variogram Estimation and Uncertainty **7**

5 Ordinary Kriging: Kriging when the mean is not known and when only the variogram is available **10**

6 Kriging and non parametric Covariance Estimation with several Simulations **11**

7 Regularisation of Covariance Estimation **13**

8 Chi Square **13**

9 Distribution of a linear combination of Chi Square random variables **13**

10 Gaussian Random Variables **14**

11 Other **15**

1 Cours de Géostatistiques

[**Allard, 2012**]:

- définitions et propriétés: stationnarités, fonction de covariance, variogramme
- représentation spectrale: théorème de Bochner
- estimation du variogramme: variogramme expérimental, modèle paramétrique

[**Gaetan and Guyon, 2009**]:

- définitions et propriétés: stationnaire du second ordre, intrinsèque, isotrope, variogramme
- modélisation spatiale par convolution: modèle de variogramme gaussien, exponentiel, sphérique
- modèle linéaire de regression spatiale
- krigrage simple (prédiction à covariance connue)
- simulations: champs gaussien stationnaire, méthodes bandes tournantes
- estimation: analyse variographique, estimation empirique du variogramme:

$$\hat{\gamma}(|h|) = \frac{1}{2|N(h)|} \sum_{(s_i,s_j)\in N(h)} (X_{s_i} - X_{s_j})^2$$

where $N(h) = \{(s_i, s_j), \|h\| - \varepsilon \leq \|s_i - s_j\| \leq \|h\| + \varepsilon\}$

Proposition 1.1. *In the Gaussian case $X \sim \mathcal{N}(0, \Sigma)$: the law of the empirical variogram of a Gaussian process is:*

$$\hat{\gamma}(h) = \sum_{i=1}^{|N(h)|} \lambda_i(h)\chi^2_{i1}$$

for i.i.d. χ^2 with one degree of freedom

[**Kanevski et al., 2009**]:

- analysis, modelling, visualisation of spatial data
- develop machine learning methods to learn from data with the goal of making spatial prediction
- motivations: machine learning methods are adaptive, non linear, robust, efficient, well suited for decision support systems and can be used for classification and regression
- software: Machine Learning Office: multilayer perceptron, regression neural networks, support vector machines, Gaussian mixture models
- learning from geospatial data: new major consideration is the spatial coordinates as inputs:
  - Machine Learning: data are i.i.d. samples drawn from some unknown distribution, statistical estimators depend on choice of loss function
  - Geostatistics: method of kriging, measures of spatial continuity closely related to covariance under stationarity assumptions
  - Machine Learning and Geostatistics can be considered as complementary approaches when working with spatial data, Machine Learning has advantages in high dimensional data
- spatial predictions: in most cases, geostatistics deal with only one realisation of the phenomena under study $\Longrightarrow$ need to formulate hypotheses about the phenomena to make statistical inferences possible: second order stationarity, ergodicity

- variography, kriging

[**Stein, 1999**]:

- kriging method for interpolating data: optimal linear prediction (BLUP) that requires knowing the covariance structure of the random field
- But, usually the covariance is unknown $\Longrightarrow$ estimate the covariance function, using the same data that will be used for the interpolation
- predicting with estimated parameters: microergodicity, plug in method
-  properties of interpolants based on estimated covariance structure are not well understood $\Longrightarrow$ often, effect of uncertainty in the covariance structure is ignored
- behavior of semi variogram near the origin plays a central role $\Longrightarrow$ key mathematical concept is equivalence and orthogonality of Gaussian measures to prove connections between behavior of semi variogram near the origin and the properties of kriging predictors
- in fill asymptotic framework

[**Vazquez, 2015**]:

- stepwise uncertainty reduction
- kriging
- consistency, reproducing kernel Hilbert space theory

[**Le Riche and Durrande, 2019**]:

- random process, Gaussian process
- kriging: statistical (BLUP, constrained optimization problem) and functional (RKHS) point of views
- kriging issues: covariance matrix (Gram) ill conditioned $\Longrightarrow$ solution: regularize it

Allard: Introduction aux Statistiques Spatiales:

- krigeage (BLUP), krigeage espérance conditionnelle dans le cas de champ gaussien
- krigeage de la moyenne

2 Covariance Function Estimation

2.1 Non Parametric Estimation

[**Elogne et al., 2008**]:

- motivation: problem of covariance estimation is crucial in modelling
- goal: estimate the covariance function at any lag point without parametric assumptions, then evaluate the impact of the covariance estimation on the kriging predictions
- assumption: smoothness of the covariance
- problem: non parametric approaches have a difficulty that lies in imposing the positive definiteness property of the estimator
- ⚠ positive definiteness property of estimator is fundamental: it ensures that the prediction variances are non neagative for BLUP
- solution: inspiring from the Parzen estimator, obtain an estimator that fulfills the initial requirement (the proof uses the Bochner spectral theorem)

[**Hall and Patil, 1994**]:

- non parametric estimator of the autocovariance of a stationary random field
- estimator is itself a covariance
- assumptions: smoothness and tail conditions on the surface, but does not require isotropy
- result: reduction of the integrated squared error

[**Putter and Alastair Young, 2001**]:

- motivations: kriging procedure gives an optimal linear predictor of a spatial process at a point, given observations, taking into account the spatial dependence of observations
- problem: kriging predictor is optimal if the weights are calculated from the correct covariance, which is unknown $\Longrightarrow$ estimate it from the data $\Longrightarrow$ accuracy of kriging predictor is effected when the optimal weights are substituted by weights derived from the estimated covariance
- solution: effect of estimation is negligible if the joint Gaussian distributions of the process under the true and the estimated covariances are contiguous
- two approaches:
  - [Diamond and Armstrong, 1984]: approximating covariance function results in a perturbation of the covariance matrix $\Longrightarrow$ effect can be expressed in terms of the condition number of the covariance matrix $\Longrightarrow$ but, these bounds are not sharp: with more observations, bounds become wider
  - true covariance is assumed to be misspecified by a second covariance: How the approximation of the covariance function by a sequence of estimators covariance function affects the accuracy of the kriging predictor?

[**Loukas, 2017**]:

- How many samples are sufficient to guarantee that the eigenvectors of the sample covariance matrix are close to those of the actual covariance matrix?

[**Vershynin, 2010**]:

- What is the optimal sample size that guarantees estimation with a fixed accuracy in the operator norm?
- ⚠ sample covariance matrix obtained from a sample of independent points

2.2 Parametric Estimation

[**Zimmerman and Cressie, 1992**]:

- predict the value of a linear functional of a random field when the parameter of the covariance function is unknown
- EBLUP predictor
- performance depends on the strength of spatial correlation: estimated mean squared predictor error is at its best when the spatial correlation is strong

[**Müller and Zimmerman, 1999**]

[**Zimmerman, 1989**]

3 Non Parametric Variogram Estimators

[**Mingoti and Rosa, 2008**]:

- note on robust and non robust variogram estimators
- presentation of five variogram estimators:
  - Matheron estimator based on the method of moments:

$$\hat{\gamma}(|h|) = \frac{1}{2|N(h)|} \sum_{(s_i,s_j)\in N(h)} (X_{s_i} - X_{s_j})^2 \tag{1}$$

    unbiased estimator, but very affected by the presence of outliers
  - Cressie Hawkins estimator [Cressie, 1980]:

$$\hat{\gamma}_{CH}(|h|) = \frac{1}{2C_h} \left(\frac{1}{|N(h)|} \sum_{(s_i,s_j)\in N(h)} (X_{s_i} - X_{s_j})^{1/2} \right)^4 \tag{2}$$

    where $C_h = 0.457 + \frac{0.494}{|N(h)|} + \frac{0.494}{|N(h)|^2}$ is the correction factor for bias.
    The estimator is robust against outliers and non normality for distributions that are normal in the central region and heavier in the tails. The authors show that the use of fourth root transformation yields stable robust estimated of the variogram.
  - Median estimator
  - Genton estimator [Genton, 1998]: use of the theory of the M estimators of scale to derive robustness properties
  - Haslett estimator: in time series context
- best estimators are:
  - for contaminated data: Genton and Median
  - for non contaminated data: Matheron and Haslett

[**Cressie, 1980**]:

- study of robust estimation of the variogram when the distribution is normal like in the central region and heavier than normal in the tails
- use of fourth root transformation yields stable robust estimates of the variogram

[**Genton, 1998**]:

- motivations: variogram estimation is a crucial stage of spatial prediction, since it determines kriging weights
- problem: classical variogram estimator (Matheron) is not robust against outliers in the data
- solution: estimator of the variogram based on a highly robust estimator of scale
- compare robustness properties of Matheron, Cressie and Hawkins and new approach estimators
- results show that the higly robust variogram estimator improves the estimation significantly

[**Kerry et al., 2008**]:

- residual maximum likelihood estimator variogram (REML)
- require fewer samples than method of moments estimator

4 Robust Variogram Estimation and Uncertainty

[**Marchant and Lark, 2004**]:

- motivations: variogram is central to any geostatistical study, important to be able to quantify variogram uncertainty to ensure that the variogram estimate is sufficiently accurate for kriging, variogram characterizes the structure of spatial correlation of a variable
- problem: precision of an estimated variogram (by the method of moments) is unknown
- results: for a Gaussian variable with known variogram, the expression for the uncertainty of the experimental estimated variogram is accurate
- ergodic variogram: variogram averaged over all realizations of the underlying random process
- non ergodic variogram: exhaustive variogram of a single realization
- test accuracy of theoretical expressions for uncertainty of the methods of moments variogram estimator of a known ergodic variogram
- study of uncertainty associated with experimental variogram estimation to a non ergodic variogram when the ergodic variogram is known

[**Diamond and Armstrong, 1984**]:

- robustness in geostatistics concentrate upon estimation of the experimental variogram
- problem: predictive methods can be very sensitive to small perturbations in data and in variogram model
- quantify notion of robustness: nearness of variogram models (since it is reflected in the sensitivity of their corresponding kriging estimators)
- condition number of kriging matrices has a central role in robustness
- previous studies: attention has largely been centered upon robust estimation of the experimental variogram and tended to ignore the possible sensitivity of further procedures (kriging, for example) that are based on the choice of the variogram model
- δ neighbourhood for a valid variogram γ, $\delta \in]0,1[$, $0 < r < \infty$:

$$N_\delta(\gamma) = \left\{ g \in \mathcal{V}, \left| \frac{g}{\gamma} - 1 \right| < \delta \right\} = \left\{ g \in \mathcal{V}, 1 - \delta < \frac{g}{\gamma} < 1 + \delta \right\}$$

  where $|\cdot|$ is the sup norm

- kriging system:

$$\begin{cases} \sum_{j=1}^{N} \lambda_j \gamma(||s_i - s_j||) + \mu = \bar{\gamma}(s_i, \mathcal{V}) \\ \sum_{j=1}^{N} \lambda_j = 1 \end{cases} \tag{3}$$

  where $\bar{\gamma}(s_i, \mathcal{V}) = V^{-1} \int_V \gamma(||s_i - s||) d_s$

- kriging system in matrix notation:

$$\Gamma \lambda = B_\gamma \tag{4}$$

  where $\Gamma = (\gamma_{i,j})_{i,j} = (\gamma(||s_i - s_j||)_{i,j \leq N}$, $\lambda^T = (\lambda_1, \cdots, \lambda_N, \mu)$ and $B_\gamma^T = (\bar{\gamma}(s_i, \mathcal{V}))_{i \leq N}$

- for any $g \in N_\delta(\gamma)$, we define the corresponding system:

$$G \lambda_g = B_g \tag{5}$$

  where $G = (g_{i,j})_{i,j} = (g(||s_i - s_j||)_{i,j \leq N}$, $\lambda_g^T = (\lambda_1^g, \cdots, \lambda_N^g, \mu^g)$ and $B_g^T = (\bar{g}(s_i, \mathcal{V}))_{i \leq N}$

- $\Delta\lambda = \lambda_g - \lambda$
- $\Delta B = B_\gamma - B_g$
- $\Delta\Gamma = G - \Gamma$
- result:

$$\boxed{\frac{||\Delta\lambda||}{||\lambda||} \leq 2\delta \frac{\kappa(\Gamma)}{1 - \delta\kappa(\Gamma)} \qquad (6)}$$

  for any variogram γ and any $\varepsilon > 0$, there exists a δ neighbourhood which produces a relative error $\frac{||\Delta\lambda||}{||\lambda||}$ in kriging that is no greater than ε (such that $\delta < \min\left\{\frac{1}{\kappa(\Gamma)}, \frac{\varepsilon}{2+\varepsilon}\kappa(\Gamma)\right\}\varepsilon$).

- configuration of sample will perturb Γ and B in the kriging equation (but not the variogram γ)

[**Armstrong, 1984**]:

- motivation: efficiency of geostatistics depends on the quality of the estimate obtained for the variogram, attention to the variability of the experimental variogram for large values of the distance h
- goal: study sensitivity of the kriging weights and of the kriging variance to small changes in the variogram model or in the location of sample points
- steps involved in geostatistical studies: data, experimental variogram, variogram model, kriging
- causes of non robust behaviour: outliers, influence of a trend on the calculation of the variogram, presence of artefacts, unsuitable choice of the distance classes (when samples are on an almost regular grid, this can lead to highly erratic variograms)
- three types of solution: downweighting the high values, robust variograms, change of scale methods (lognormal kriging)
- sensitivity of kriging: to what extent the kriging weights and the kriging variance depend on the choice of the variogram model and on the choice of the location of samples? $\Longrightarrow$ find mathematical way of characterizing the change in the kriging weights and the kriging variance provoked by a small change in the variogram model
- motivation: able to quantify the robustness of kriging to small perturbations $\Longrightarrow$ help identify potentially unstable cases and allow the user to take appropriate action
- results:

$$\frac{||\Delta\lambda||}{||\lambda||} \leq 2\delta \frac{\kappa(\Gamma)}{1 - \delta\kappa(\Gamma)}$$

  and

$$|\sigma_g^2 - \sigma_\gamma^2| \leq \delta\left(\bar{\gamma} + ||\Gamma||\,||\lambda||^2 + 2||\Gamma||\,||\lambda||^2(1+\delta)\frac{\kappa(\Gamma)}{1 - \delta\kappa(\Gamma)}\right)$$

[**Bardossy, 1988**]:

- motivation: in ordinary kriging, the weights attached to the measurements and the estimation variance are computed thanks to the theoretical variogram and the locations of the measurements points (values of measurements used in two steps of the procedure: for the estimation of the variogram and for the estimator value after computing the kriging weights)
- study the robustness of the kriging system with respect to uncertainty of the estimation of the theoretical variogram
- goal: develop measure of the effect of the variogram uncertainty on the kriging weights

- inequalities for possible changes of the kriging estimator and the kriging variance
- inequalities for change in the estimation variance for a modified nugget effect
- changes of kriging weights can predicted partly thanks to the maximal kriging weights
- previous studies: no studies about the effect of measurements points configuration on the robustness of kriging estimator
- previous studies: [Diamond and Armstrong, 1984]: problem: inequality depends mostly on the configuration of locations of measurements points and on the variogram and not much influenced by the location of the point to be estimated, but the estimator might be very different depending on the location of the point to be estimated $\Longrightarrow$ the condition number is not affected by the location of the point to be estimated

5 Ordinary Kriging: Kriging when the mean is not known and when only the variogram is available

[Chiles and Delfiner, 1999]:

- problem: in most practical situations, the mean is not known $\Longrightarrow$ simple kriging cannot be applied
- solution: ordinary kriging, that does not require any knowledge of the mean
- only variogram known
- kriging system to estimate the random function at a location s_0 is:

$$\begin{cases} \gamma_{i,0} = \sum_{j=1}^{N} \lambda_j \gamma_{i,j} - \mu, \quad \forall i \in \{1, \cdots, N\} \\ \sum_{j=1}^{N} \lambda_j = 1 \end{cases} \tag{7}$$

  where $\gamma_{i,j} = \gamma(h_{i,j}) = \gamma(||s_i - s_j||)$

[Gratton, 2002]:

- matrix version:

$$\Gamma\lambda = B \tag{8}$$

$$\Gamma = \begin{pmatrix} \gamma(h_{1,1}) & \cdots & \gamma(h_{1,N}) & 1 \\ \vdots & & \vdots & \vdots \\ \gamma(h_{N,1}) & \cdots & \gamma(h_{N,N}) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} \gamma(h_{1,0}) \\ \vdots \\ \gamma(h_{N,0}) \\ 1 \end{pmatrix} \qquad \lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_N \\ \mu \end{pmatrix}$$

- in order to obtain an unbiased estimator, we introduce the constraint that the weights must add up to 1 (the kriging error becomes an allowable linear combination)

6 Kriging and non parametric Covariance Estimation with several Simulations

[**Qiao et al., 2018**]:

- geophysics: goal is to interpolate the value of a physical quantity of interest at an unsampled spatial location by exploiting the correlation between measurements collected from a number of suitably spaced sensors $\Longrightarrow$ solution is linear function of observed data
- assumptions: second order stationarity: correlation between measurements depends only on difference between locations coordinates
- use two dimensional nested array in kriging $\Longrightarrow$ reduce the number of sensors deployed in the field $\Longrightarrow$ enable efficient interpolation
- under spatial stationarity assumptions, the number of predictable locations is determined by the distinct mutual location differences of the sensors
- predictions beyond the field by constructing a virtual covariance matrix of larger dimension from the covariance matrix of the physical measurements: relate kriging to Toeplitz matrix estimation
- robustness analysis and concentration bounds of sample covariance:

  Theorem 6.1. *We assume that $(X^l)_{l\leq L}$ are zero mean i.i.d. Gaussian random vectors, such that: $X^l \sim \mathcal{N}(0, \Gamma_N)$, where Γ_N is a positive semi definite matrix of full rank. For $u \geq 0$, if the number of simulations verifies*

$$L > \max\left\{ \frac{4(N+u)}{\left(1+\sqrt{1+\frac{4}{\kappa C}}\right)^2}, \left(\frac{Tr(\Gamma_N)\kappa}{\|c_N\|_2}\right)^2 \right\}$$

  then, with probability at least $1 - 2e^{-u} - 2e^{-2c_1\sqrt{L}}$:

$$\frac{\|\hat{\lambda} - \lambda^*\|_2}{\|\lambda^*\|} \leq \kappa\varepsilon \left(1 + \frac{2}{1-\kappa\varepsilon}\right) \tag{9}$$

  where we let

$$\varepsilon = \max\left\{ C\left(\sqrt{\frac{N+u}{L}} + \frac{N+u}{L}\right), \frac{Tr(\Gamma_N)}{\sqrt{L}\|c_N\|_2} \right\}$$

- given a number L of snapshots, total least squares are used to show that a stable estimate of the linear combination coefficients can be obtained
- relative estimation error $\frac{\|\hat{\lambda}-\lambda^*\|_2}{\|\lambda^*\|}$ is a decreasing function of L (tends to 0)

[**Wei, 1989**]:

- results for the proof of the concentration bounds of [Qiao et al., 2018]
- perturbation of consistent least squares problems
- linear system $\Gamma_N\lambda^* = c_N$ and estimated system $\hat{\Gamma}\hat{\lambda} = \hat{c}$, assumption $Rank(\hat{\Gamma}) = Rank(\Gamma_N) = r \leq N$ (the matrices are not necessarly of full rank)

  Theorem 6.2. *We assume that $\exists\, \varepsilon > 0$, such that $\|\Delta\Gamma_N\| \leq \varepsilon\|\Gamma_N\|$, $\|\Delta c_N\|_2 \leq \varepsilon\|c_N\|_2$ and $\kappa\varepsilon < 1$. Then, we have:*

$$\frac{\|\hat{\lambda} - \lambda^*\|_2}{\|\lambda^*\|} \leq \kappa\varepsilon \left(1 + \frac{2}{1-\kappa\varepsilon}\right) \tag{10}$$

- use of lemma:

  Lemma 6.1. *Let $\Gamma_N \in^{N\times N}$ a matrix and its empirical estimation $\hat{\Gamma} \in^{N\times N}$: $\hat{\Gamma} = \Gamma_N + \Delta\Gamma_N$. Under the assumptions that the two matrices are of same rank ($Rank(\hat{\Gamma}) = Rank(\Gamma_N) = r \leq N$) and that $\|\Delta\Gamma_N\|\|\Gamma_N^{-1}\| < 1$, we have*

$$\|\hat{\Gamma}^{-1}\| \leq \frac{\|\Gamma_N^{-1}\|}{1 - \|\Delta\Gamma_N\|\|\Gamma_N^{-1}\|} \tag{11}$$

[Vershynin, 2018]:

- results for the proof of the concentration bounds of [Qiao et al., 2018]
- results for i.i.d. Gaussian random vectors
- covariance matrix can be estimated accurately by the sample covariance matrix if sample size m is proportional to dimension n
- concentration bounds of sample covariance:

  Corollary 6.1. *We assume that $(X^l)_{l\leq L}$ are zero mean i.i.d. Gaussian random vectors, such that: $X^l \sim \mathcal{N}(0, \Gamma_N)$, where Γ_N is a positive semi definite matrix. For $u \geq 0$, with probability at least $1 - 2e^{-u}$:*

$$\|\Delta\Gamma_N\| = \|\hat{\Gamma} - \Gamma_N\| \leq C\left(\sqrt{\frac{N+u}{L}} + \frac{N+u}{L}\right)\|\Gamma_N\| \tag{12}$$

  where $C > 0$.

7 Regularisation of Covariance Estimation

[Bickel and Levina, 2008a]

[Bickel and Levina, 2008b]

[Pourahmadi, 2011]

8 Chi Square

[Laurent and Massart, 2000]:

Lemma 8.1. *We let $(Y_i)_{i\leq k}$ independent and identically distributed centred Gaussian random variables with variance 1. Given $(a_i)_{i\leq k}$ non negative, we define $|a|_\infty = \sup_{i\leq k}\{|a_i|\}$ and $|a|_2^2 = \sum_{i=1}^k a_i^2$.*

Let the random variable $Z = \sum_{i=1}^k a_i(Y_i^2 - 1)$.
Then Z satisfies the two inequalities for $x \geq 0$

$$\mathbb{P}\left(Z \geq 2|a|_2\sqrt{x} + 2|a|_\infty x\right) \leq e^{-x} \tag{13}$$

$$\mathbb{P}\left(Z \leq -2|a|_2\sqrt{x}\right) \leq e^{-x} \tag{14}$$

Corollary 8.1. *We assume $X \sim \chi_k^2$ chi square with k degrees of freedom. Then we have*

$$\mathbb{P}\left(X - k \geq 2\sqrt{kx} + 2x\right) \leq e^{-x} \tag{15}$$

$$\mathbb{P}\left(X - k \leq -2\sqrt{kx}\right) \leq e^{-x} \tag{16}$$

9 Distribution of a linear combination of Chi Square random variables

[Mathai, 1982]:

- derive exact representations of the distribution of a linear combination of Chi Square limitation to small values of number of terms in the sum for computational purposes
- Let $(X_i)_{i\leq n}$ be n independent Gamma random variables $\Gamma(\alpha_i, \beta_i)$ where $\alpha_i, \beta_i > 0, \forall i \in \{1, \cdots, n\}$. Define: $Z = \sum_{i=1}^n X_i$. The density function of Z is

$$f_Z(z) = \left(\prod_{i=1}^n \frac{1}{\beta_i^{\alpha_i}\Gamma(i)}\right)^{-1} z^{\alpha-1} - e^{-z\beta_1}\phi\left(\alpha_2, \cdots, \alpha_n, a, \left(\frac{1}{\beta_1} - \frac{1}{\beta_2}\right)z, \cdots, \left(\frac{1}{\beta_1} - \frac{1}{\beta_n}\right)z\right) \tag{17}$$

  where $\beta_1 = \min_i \beta_i$, $a = \sum_{i=1}^n \alpha_i$ and ϕ is the confluent hypergeometric function (defined in the previous subsection).

[Solomon and Stephens, 1977]: propose various moments based approximations

[Oman and Zacks, 1981]: propose method of negative binomial mixtures, which reduces the amount of computation and produces sufficiently accurate results

[Moschopoulos and Canada, 1984]: calculate the distribution function by inverting the moment generating function, the distribution function is expressed as an infinite Gamma series whose terms can be computed efficiently to a sufficient degree of accuracy

[Moschopoulos, 1985]: Let $(X_i)_{i\leq n}$ be n independent Gamma random variables $\Gamma(\alpha_i, \beta_i)$ where $\alpha_i,\ \beta_i > 0,\ \forall i \in \{1, \cdots, n\}$. Define: $Z = \sum_{i=1}^{n} X_i$. The density function of Z is

$$f_Z(z) = C \sum_{k=0}^{\infty} \delta_k f(x, a+k, \beta_1) \tag{18}$$

where f is the density function of a Gamma variable with parameters $\Gamma(a+k, \beta_1)$ defined in the previous subsection by $f(x, a+k, \beta_1) = \frac{\beta_1^{a+k}}{\Gamma(a+k)} x^{a+k-1} e^{-\beta_1 x},\ \forall x > 0$. We define $\beta_1 = \min_i \beta_i$, $C = \prod_{i=1}^{n} \left(\frac{\beta_i}{\beta_1}\right)^{\alpha_i}$ and $a = \sum_{i=1}^{n} \alpha_i$.

The coefficients can be calculated recursively by: $\delta_k = \frac{1}{k} \sum_{i=1}^{k} i\rho_i \delta_{k-i},\ \forall k > 0$ and $\delta_0 = 1$, where $\rho_i = \sum_{j=1}^{n} \alpha_j \left(1 - \frac{\beta_i}{\beta_1}\right)^k \frac{1}{k}$.

The distribution function of Z is

$$F_Z(z) = C \sum_{k=0}^{\infty} \delta_k G(x, a+k, \beta_1) \tag{19}$$

where G is the distribution function of a Gamma variable with parameters $\Gamma(a+k, \beta_1)$ and the other variables are defined as above.

[Bausch, 2013]:

- previous methods are not computationally efficient (exponential growth in the number of terms) and are limited to small number of terms
- propose computationally efficient algorithm to numerically calculate the linear combination of Chi Square random variables, starting out from an analytic expression for the error bound, technique computationally efficient (polynomial in growth in the number of terms)
- Let $X, Y \sim \chi^2_k$ two i.i.d. Chi Square random variables and define $Z = aX + bY$ where $a, b \in \mathbb{R}_+$. The distribution function of Z is

$$f_Z(z) = \frac{1}{(4ab)^k} \left(\frac{a-b}{8ab}\right)^{(1-k)/2} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma(k)} e^{-\frac{a+b}{4ab}z} x^{(k-1)/2} I_{(k-1)/2}\left(\frac{b-a}{4ab}z\right),\ \forall z > 0 \tag{20}$$

- For two i.i.d. Chi Square random variables with 1 degree of freedom, the distribution function is

$$f_Z(z) = \frac{1}{4ab} e^{-\frac{a+b}{4ab}z} I_0\left(\frac{b-a}{4ab}z\right),\ \forall z > 0 \tag{21}$$

10 Gaussian Random Variables

[Nadarajah and Pogány, 2016]:
product of dependent Gaussian random variables gives a much more complicated function distribution (compared to sum of dependent Gaussian random variables) $\Longrightarrow$ we prefer variogram empirical estimation (compared to covariance empirical estimation).

11 Other

[Lantuéjoul, 1991]

[Bel et al., 2009]

[Wedin, 1973]

[DeVore, 1998]

[Hsu et al., 2011]

[Zajkowski, 2020]

[Cressie, 1993]

References

[Allard, 2012] Allard, D. (2012). Cours de statistiques spatiales : introduction à la géostatistique. *Université de Montpellier*.

[Armstrong, 1984] Armstrong, M. (1984). Improving the estimation and modelling of the variogram. In *Geostatistics for natural resources characterization*, pages 1–19. Springer.

[Bardossy, 1988] Bardossy, A. (1988). Notes on the robustness of the kriging system. *Mathematical geology*, 20(3):189–203.

[Bausch, 2013] Bausch, J. (2013). On the efficient calculation of a linear combination of chi-square random variables with an application in counting string vacua. *Journal of Physics A: Mathematical and Theoretical*, 46(50):505202.

[Bel et al., 2009] Bel, L., Allard, D., Laurent, J., Cheddadi, R., and Bar-Hen, A. (2009). CART algorithm for spatial data: Application to environmental and ecological data. *Computational Statistics and Data Analysis*, 53(8):3082–3093.

[Bickel and Levina, 2008a] Bickel, P. J. and Levina, E. (2008a). Covariance regularization by thresholding. *The Annals of Statistics*, 36(6):2577–2604.

[Bickel and Levina, 2008b] Bickel, P. J. and Levina, E. (2008b). Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36(1):199–227.

[Chiles and Delfiner, 1999] Chiles, J.-P. and Delfiner, P. (1999). *Geostatistics: modeling spatial uncertainty*. Wiley, New York.

[Cressie, 1980] Cressie, N., H. D. (1980). Robust estimation of the variogram. *Mathematical Geology*, 12:115–125.

[Cressie, 1993] Cressie, N. (1993). *Statistics for Spatial Data*, chapter 1, pages 1–26. John Wiley Sons, Ltd.

[DeVore, 1998] DeVore, R. A. (1998). Nonlinear approximation. *Acta Numerica*, 7:51–150.

[Diamond and Armstrong, 1984] Diamond, P. and Armstrong, M. (1984). Robustness of variograms and conditioning of kriging matrices. *Journal of the International Association for Mathematical Geology*, 16(8):809–822.

[Elogne et al., 2008] Elogne, S., Perrin, O., and Thomas-Agnan, C. (2008). Non parametric estimation of smooth stationary covariance functions by interpolation methods. *Statistical Inference for Stochastic Processes*, 11(2):177–205.

[Gaetan and Guyon, 2009] Gaetan, C. and Guyon, X. (2009). *Spatial Statistics and Modeling*. Springer Series in Statistics. Springer New York.

[Genton, 1998] Genton, M. (1998). Highly robust variogram estimation. *Mathematical Geosciences*, 30(2):213–221.

[Gratton, 2002] Gratton, Y. (2002). Le krigeage: La méthode optimale d'interpolation spatiale.

[Hall and Patil, 1994] Hall, P. and Patil, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. *Probab. Th. Rel. Fields*, 99:399–424.

[Hsu et al., 2011] Hsu, D., Kakade, S. M., and Zhang, T. (2011). A tail inequality for quadratic forms of subgaussian random vectors.

[Kanevski et al., 2009] Kanevski, M., Timonin, V., and Pozdnukhov, A. (2009). *Machine Learning for Spatial Environmental Data: Theory, Applications, and Software*. Environmental sciences research report. CRC Press.

[Kerry et al., 2008] Kerry, R., Ingram, B., Goovaerts, P., and Oliver, M. (2008). *How many samples are required to estimate a reliable REML variogram?*

[Lantuéjoul, 1991] Lantuéjoul, C. (1991). Ergodicity and integral range. *Journal of Microscopy*, 161(3):387–403.

[Laurent and Massart, 2000] Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302 – 1338.

[Le Riche and Durrande, 2019] Le Riche, R. and Durrande, N. (2019). An overview of kriging for researchers. Lecture.

[Loukas, 2017] Loukas, A. (2017). How close are the eigenvectors of the sample and actual covariance matrices? volume 70 of *Proceedings of Machine Learning Research*, pages 2228–2237, International Convention Centre, Sydney, Australia. PMLR.

[Marchant and Lark, 2004] Marchant, B. and Lark, R. (2004). Estimating variogram uncertainty. *Mathematical Geology*, 36:867–898.

[Mathai, 1982] Mathai, A. (1982). Storage capacity of a dam with gamma type inputs. *Annals of the Institute of Statistical Mathematics*, 34(3):591–597.

[Mingoti and Rosa, 2008] Mingoti, S. A. and Rosa, G. (2008). A note on robust and non-robust variogram estimators. *Rem: Revista Escola de Minas*, 61:87 – 95.

[Moschopoulos, 1985] Moschopoulos, P. (1985). The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics*, 37:541–544.

[Moschopoulos and Canada, 1984] Moschopoulos, P. and Canada, W. (1984). The distribution function of a linear combination of chi-squares. *Computers Mathematics with Applications*, 10(4):383–386.

[Müller and Zimmerman, 1999] Müller, W. G. and Zimmerman, D. (1999). Optimal designs for variogram estimation. *Environmetrics*, 10:23–37.

[Nadarajah and Pogány, 2016] Nadarajah, S. and Pogány, T. K. (2016). On the distribution of the product of correlated normal random variables. *Comptes Rendus Mathematique*, 354(2):201–204.

[Oman and Zacks, 1981] Oman, S. D. and Zacks, S. (1981). A mixture approximation to the distribution of a weighted sum of chi-squared variables. *Journal of Statistical Computation and Simulation*, 13(3-4):215–224.

[Pourahmadi, 2011] Pourahmadi, M. (2011). Covariance estimation: The glm and regularization perspectives. *Statist. Sci.*, 26(3):369–387.

[Putter and Alastair Young, 2001] Putter, H. and Alastair Young, G. (2001). On the effect of covariance function estimation on the accuracy of kriging predictors. *Bernoulli*, 7(3):421–438.

[Qiao et al., 2018] Qiao, H., Hucumenoglu, M. C., and Pal, P. (2018). Compressive kriging using multidimensional generalized nested sampling. In *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, pages 84–88.

[Solomon and Stephens, 1977] Solomon, H. and Stephens, M. A. (1977). Distribution of a sum of weighted chi-square variables. *Journal of the American Statistical Association*, 72(360):881–885.

[Stein, 1999] Stein, M. L. (1999). *Interpolation of spatial data*. Springer Series in Statistics. Springer-Verlag, New York. Some theory for Kriging.

[Vazquez, 2015] Vazquez, E. (2015). *Sequential search strategies based on kriging*. Habilitation à diriger des recherches, Université Paris-Sud.

[Vershynin, 2010] Vershynin, R. (2010). How close is the sample covariance matrix to the actual covariance matrix?

[Vershynin, 2018] Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

[Wedin, 1973] Wedin, P. (1973). Perturbation theory for pseudo-inverses. *BIT Numerical Mathematics*, 13:217–232.

[Wei, 1989] Wei, M. (1989). The perturbation of consistent least squares problems. *Linear Algebra and its Applications*, 112:231 – 245.

[Zajkowski, 2020] Zajkowski, K. (2020). Bounds on tail probabilities for quadratic forms in dependent sub-gaussian random variables. *Statistics Probability Letters*, 167:108898.

[Zimmerman and Cressie, 1992] Zimmerman, D. and Cressie, N. (1992). Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Annals of the Institute of Statistical Mathematics*, 44:27–43.

[Zimmerman, 1989] Zimmerman, D. L. (1989). Computationally exploitable structure of covariance matrices and generalized convariance matrices in spatial models. *Journal of Statistical Computation and Simulation*, 32(1-2):1–15.