

Informative Selection and Spatial Process

Bonnéry, Pantalone, Ranalli

Contents

1	Introduction	2
2	Statistical Framework	3
2.1	General notations	3
2.2	Spatial process	4
2.3	Sample, design and design variable	6
2.3.1	Fixed designs and design variables	6
2.3.2	Random design variables, sample and random design	7
2.4	Observations	9
3	Sample distribution	9
3.1	Density ratio and weighted density.	9
3.2	Distribution of $Y[S]$	11
3.3	Sample Intensity	12
3.4	Sample Variogram	12
4	Estimation	12
4.1	Naive estimators and their properties in the non informative selection case .	12
4.2	Properties of the sample in the general case (including informative)	14
4.3	Naive estimation	14
5	Prediction	15
5.1	Accounting for informative selection	16
A	Algebra for the example	17
A.1	Distribution of Z conditionally on $Y[\mathbf{x}] = \mathbf{y}$	17
A.2	Proof of Equation (35)	17
A.3	Algebra for the denominator of ρ	18
B	R code	18

C May be needed later	20
C.1 Algebra for $G^*(h)$	20
C.2 Likelihood	21
C.3 Intensity ratios	22
C.4 Sample distribution	23
C.5 Proof of Property C.1	24

Abstract

This paper extends the concept of informative selection, population distribution and sample distribution in a spatial process context. Those notions were defined by M. Krieger and Pfeffermann (1992) in a context where the output of the random process of interest consists of independent and identically distributed realisations for each individuals of a population. ? showed that informative selection was inducing a stochastic dependence among realisations on the selected units. In the context of spatial process, the "population" is a continuous space and realisations for two different elements of the population are not independent. We show how informative selection may induce a different dependence among selected units, how the sample distribution differs from the "population" distribution, and how one can account for this effect in an simulated study when doing statistical inference, including Semi-variogram parametric and semi parametric estimation, as well as prediction on the part of the space for which the random process was not observed.

The references in abstracts should be either written in full or not given.

1 Introduction

Spatial processes are employed in many fields, like geology, Earth science, environmental and agricultural surveys, among many others. A huge variety of data can be employed, such as rainfall data (Ord and Rees, 1979), atmospheric data (Thiebaux and Pedder, 1987), forestry data (Samra et al., 1989) and soils data (Burgess and Webster, 1980) **FP: these references are a little bit old, so not sure if we can leave it.** In these applications, a general approach is to postulate a spatial process for the variable Y under investigation, that is a stochastic process which is assumed to have generated the population values, also called *superpopulation* model. When using sample data, we need to pay attention at the relation between the spatial process and the mechanism selection. Indeed, the distribution of the observed data can be different from the distribution assumed for the population by means of the spatial process. In other words, the density obtained by the sample data may not be the density obtained by reducing the number of terms in P^Y , as if the units were selected completely at random. In this case, we say that the sampling is *informative*. Ignoring an informative sample may lead to biases and erroneous inference, as illustrated for example in Skinner et al. (1989). This could happen when there is dependence between the assumed stochastic model and the sampling mechanism, that is the units are selected dependently to

the variable of interest. Examples of this type of mechanism selection are, among many others, length-biased sampling, endogenous stratification, adaptive sampling, sequential quota sampling and cut-off sampling. For more details, see Section 3 of Bonnery et al. (2012).

When dealing with spatial processes, an important tool is the *variogram*, which analyzes the degree of spatial dependence of such processes and provides useful insights on the phenomenon under investigation. For instance, the variogram is the key component for the well-known *kriging* method (Matheron, 1962), and it plays a crucial role on prediction since it is used to compute the kriging weights.

In this paper, we study the effect of informative selection when the variogram is of interest. We consider informative selection as a situation where the sample responses, given that they were selected, are not i.i.d. from the superpopulation model. As in Pfeffermann et al. (1998), we start from the distribution of the observed values given they were selected, or *sample pdf*. The major difference in our approach is that, while Pfeffermann et al. (1998) consider the observations as if they were independently distributed according to the sample pdf, our work consider the dependence between the observed units. We provide a theoretical background for the definitions of *population variogram* and *sample variogram*, and some properties of the *naive* estimator, i.e. the estimator that does not take into account the informativeness of the selection mechanism.

The paper is organized as follow. In Section 2 we define the statistical framework used throughout the paper. In particular, general notions and the concepts of spatial process, sample, design and design variables are introduced. In Section 3 the sample distribution and the population distribution are defined, and in Section 4 estimation of the variogram is analyzed and properties of the naive estimator are presented. Finally, Section ?? provides conclusion and future research.

2 Statistical Framework

This section illustrates the concepts and notations defined in ?.

2.1 General notations

In this paper, all random variables are defined on a probability space (Ω, \mathcal{A}, P) . The expected value and variance/covariance operators are defined with respect to the probability measure P . For two sets E and F , $(E \rightarrow F)$ or F^E designate the set of the functions from E to F . The notation " $g : E \rightarrow F, x \mapsto g(x)$ " means: let g be a mapping from set E to set F that to x associates $g(x)$. For $f : E \rightarrow (F \rightarrow G)$, $g : E \rightarrow F$, $h : E \rightarrow (H \rightarrow F)$, the notation $f[g]$ designates the function : $f[g] : E \rightarrow G$, such that $f[g](x) = (f(x))(g(x))$, and the notation $f[h]$ designates the function : $f[h] : E \rightarrow (H \rightarrow G)$, such that $f[h](x) = (f(x) \circ h(x))$.

For a set E , \bar{E} designates the set $\{\mathbf{0}\} \cup \bigcup_{n \in \mathbb{N}, n \geq 1} E^{\{1, \dots, n\}}$, where $\mathbf{0} = E^\emptyset$ corresponds to the set containing an application with empty domain. Let denote by size the application that maps an application to the cardinality of its domain: $\bar{E} \rightarrow \mathbb{N}$, $\mathbf{e} \mapsto n$ if $\mathbf{x} \in E^{\{1, \dots, n\}}$, 0 if $\mathbf{x} = \mathbf{0}$. For an application $\mathbf{x} \in \bar{E}$, a set K , \mathbf{x}_K is the application: $\mathbf{x}_K : K \cap \text{domain}(\mathbf{x}) \rightarrow$

$E : \ell \mapsto \mathbf{x}(\ell)$, in the case of a random application $S : \Omega \rightarrow \bar{U}$ and a random set $:K \rightarrow (\mathcal{P}(\mathbb{N})$ ($\mathcal{P}(\mathbb{N})$ is the set of all subsets of \mathbb{N}), then S_K is the random application: $\Omega \rightarrow \bar{U}, \omega \mapsto S_K(\omega) : (K(\omega) \cap \text{domain}(S(\omega)) \rightarrow U), \ell \mapsto (S(\omega))(\ell)$. For a measure η of E , a non random finite set K , $\eta^{\otimes K}$ is the measure such that for any collection $(A_\ell)_{\ell \in K}$ of subsets of E : $\eta^{\otimes K}(\bigcap_{\ell \in K} \{\mathbf{x} \in E^K : \mathbf{x}(\ell) \in A_\ell\}) = \prod_{\ell \in K} \eta(A_\ell)$, and $\eta^{\otimes \emptyset}(\{\mathbf{0}\}) = 1$, then define the measure $\bar{\eta}$ on \bar{E} : $\bar{\eta} = \eta^{\otimes \emptyset} + \sum_{n \in \mathbb{N}, n \geq 0} \eta^{\otimes \{1, \dots, n\}}$.

2.2 Spatial process

We consider a space U , that is a compact (non necessarily convex) subset of a finite dimensional real vector space \mathbb{R}^d , with its associated Borel sigma-field, and a random process Y defined on U with value in another finite dimension real vector space \mathcal{Y} , e.g. $Y : \Omega \rightarrow (U \rightarrow \mathcal{Y})$. For example for a random variable $S : \Omega \rightarrow U^{\{1,2\}}$, and a random variable $Y : \Omega \rightarrow (U \rightarrow \mathcal{Y})$, $Y[S]$ is the random variable: $Y[S] : \Omega \rightarrow (\{1,2\} \rightarrow \mathcal{Y})$, $\omega \mapsto (Y(\omega))(S(\omega)) : \ell \mapsto (Y(\omega))(S(\omega))(\ell)$. The set of all functions All definitions will be given under the general statistical framework described above. Examples and illustrations will be given for the particular case where $U = [0, 1]^2$. Let η (resp. ν) denote a sigma-finite measure on the set \mathcal{Y} (resp. U). The notation $f_{V|W}$ denotes the density of V conditional on W with respect to a dominating measure on the domain of V . The average theoretical semivariogram is defined as the function:

$$G : \mathbb{R}^d \rightarrow [0, +\infty), h \mapsto \frac{1}{2} \int_{U^{\{1,2\}}} \text{Var}[Y[\mathbf{x}(2)] - Y[\mathbf{x}(1)]] d(\nu^{\otimes \{1,2\}})^{X|X[2]-X[1]=h}(\mathbf{x}), \quad (1)$$

where (X) is the identity of $U^{\{1,2\}}$, and the average theoretical covariogram is the function $\nu^{X_2-X_1} - a.s(h)$ -defined:

$$C : \mathbb{R}^d \rightarrow \mathbb{R}, h \mapsto \int_{U^{\{1,2\}}} \text{Cov}[Y[\mathbf{x}(1)], Y[\mathbf{x}(2)]] d(\nu^{\otimes \{1,2\}})^{X|X[2]-X[1]=h}(\mathbf{x}). \quad (2)$$

The covariogram and semivariogram satisfy the relationship: $\forall h \in \mathbb{R}^d, G(h) = C(0) - C(h)$.

Definition 2.1 (Intrinsic stationarity and Second order stationarity, (Cressie, 2015, p. 53)). A process is intrinsic stationary when the following conditions are satisfied : $\forall \mathbf{x} \in U^{\{1,2\}}$,

$$\mathbb{E}[Y[\mathbf{x}(2)] - Y[\mathbf{x}(1)]] = 0 \quad (3)$$

$$\frac{1}{2} \text{Var}[Y[\mathbf{x}(2)] - Y[\mathbf{x}(1)]] = G(\mathbf{x}(2) - \mathbf{x}(1)) \quad (4)$$

A process is second order stationary when the following are satisfied:

$$\exists \mu \in \mathbb{R}, \forall \mathbf{x} \in U, \mathbb{E}[Y[\mathbf{x}]] = \mu \quad (5)$$

$$\forall \mathbf{x} \in U^{\{1,2\}}, \text{Cov}[Y[\mathbf{x}(1)], Y[\mathbf{x}(2)]] = C(\mathbf{x}(1) - \mathbf{x}(2)) \quad (6)$$

In the case of a first order stationary process, the variance operator $\text{Var}[\cdot]$ can equivalently be replaced by the square expected value operator $E[(\cdot)^2]$ in equation (1). In the case of a second order stationary process, equations (2) and (1) correspond to the definition of the theoretical covariogram and semivariogram as found in Cressie (2015, p. 53 and p. 58). The random process is isotropic if in addition on being second order stationary, the covariogram function $h \mapsto C(h)$ only depends on h via $h \mapsto \|h\|$.

Common model assumptions on the process Y consist in assuming second order stationarity and isotropy. Covariance structure of the signal is then fully characterized by $C(0)$ and $G(h), h \neq 0$. A Gaussian covariogram is a function of the form: $h \mapsto C(h) = c_0 + c_1(1 - \exp(-\|h\|^2/(2c_2^2)))$ where $c_0, c_1, c_2 \in [0, +\infty)$ (see (Chilès and Delfiner, 1999, p. 80) for more models).

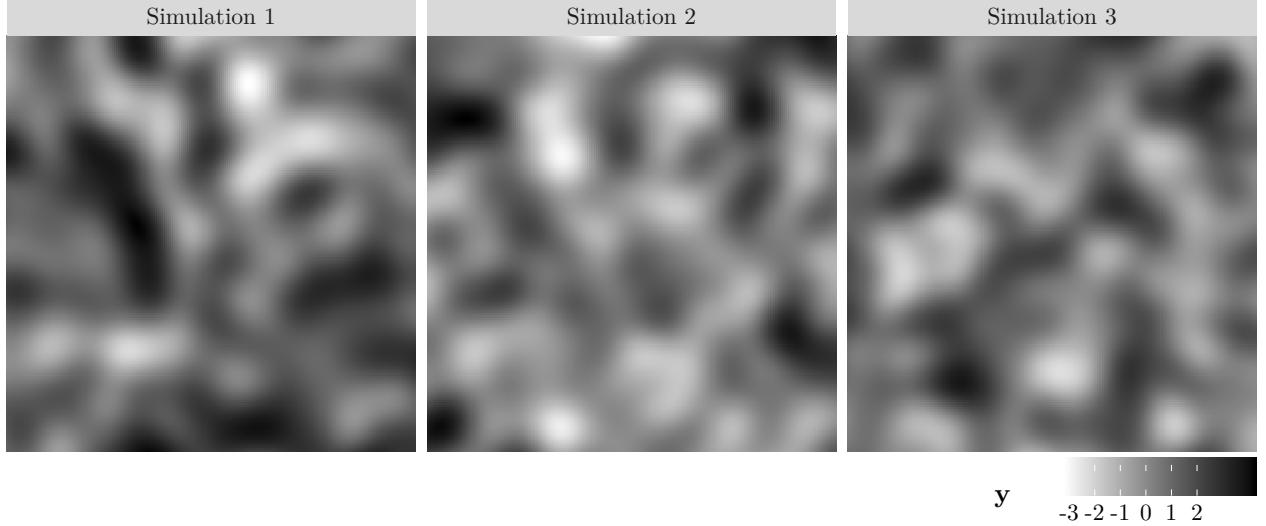
Example with simulations: isotropic Gaussian process Y .

In the case where $Y : \Omega \rightarrow (\mathbb{U} \rightarrow \mathcal{Y})$ is a Gaussian random process, its distribution can be derived from the distributions of $Y[\mathbf{x}]$, where $\mathbf{x} \in \bar{\mathbb{U}}$. For $\mathbf{x}, \mathbf{x}' \in \bar{\mathbb{U}}$, denote the expected value of the signal by $\mu : \bar{\mathbb{U}} \rightarrow \bar{\mathcal{Y}}, \mathbf{x} \mapsto E[Y[\mathbf{x}]]$, and the covariance of the the random vectors $Y[\mathbf{x}], Y[\mathbf{x}']$ by $\Sigma_{\mathbf{x}, \mathbf{x}'} = \text{Cov}[Y[\mathbf{x}], Y[\mathbf{x}']]$. The distribution of Y is then fully characterized by μ and Σ : for $n \in \mathbb{N}$, $\mathbf{x} \in \mathbb{U}^{\{1, \dots, n\}}$, $Y[\mathbf{x}]$ has the following density with respect to $\eta^{\otimes \{1, \dots, n\}}$:

$$f_{Y[\mathbf{x}]}(\mathbf{y}) = \left(2\pi^{n/2}|\Sigma_{\mathbf{x}, \mathbf{x}}|^{\frac{1}{2}}\right)^{-1} \exp\left(-\frac{1}{2}(\mathbf{y} - \mu(\mathbf{x}))\Sigma_{\mathbf{x}, \mathbf{x}}^{-1}(\mathbf{y} - \mu(\mathbf{x}))^T\right). \quad (7)$$

In the case of an isotropic Gaussian process, the distribution of Y is fully characterized by μ and G . We simulate three independant replications of an isotropic Gaussian process with $Y : \Omega \rightarrow (\mathbb{U} = [0, 1]^2 \rightarrow \mathbb{R})$, with $\forall \mathbf{x} \in \mathbb{U}, \mu[\mathbf{x}] = 0$ and with a Gaussian Covariogram with parameters $c_0 = 0, c_1 = 5, c_2 = 10$. Figure 1 represents the independent realizations of Y .

Figure 1: Heat maps of realisations of the random process $Y : \Omega \rightarrow (U = [0, 1]^2 \rightarrow \mathbb{R})$



For each realisation \mathbf{y} of the random process Y , for each \mathbf{x} in U , the value of $\mathbf{y}[\mathbf{x}]$ is color coded with a grayscale gradient.

2.3 Sample, design and design variable

2.3.1 Fixed designs and design variables

By definition a design \mathbf{d} is a probability distribution on \bar{U} . A sample S drawn from \mathbf{d} is a random variable or point process of distribution \mathbf{d} , e.g. a random variable S such that $P^S = \mathbf{d}$. Define the size $N = \text{size} \circ S$ of the sample S . The sample density with respect to $\bar{\nu}$, is defined by: $(dP^S)/(d\bar{\nu})(\mathbf{x}) = P(N = n) \times (dP^{S|N=n})/(d\nu^{\otimes \{1, \dots, n\}})(\mathbf{x})$, if $\mathbf{x} \in U^{\{1, \dots, n\}}$, $P(N = 0)$ if $\mathbf{x} = \mathbf{0}$. A fixed design variable is a function $\mathbf{z} : U \mapsto \mathcal{Z}$. A fixed size design \mathbf{d} is usually defined as a function of a fixed design variable, and characterised by its density with respect to $\bar{\nu}$. For example, the Probability Proportional to Size \mathbf{z} With Replacement and size n (PPSWR(\mathbf{z}, n)) design, with $\mathcal{Z} = [0, +\infty)$ is characterized by:

$$(dP^S/d\bar{\nu})(\mathbf{x}) = \left(\int_U (\mathbf{z}[\mathbf{x}']) d\nu(\mathbf{x}') \right)^{-n} \left(\prod_{\ell=1}^n (\mathbf{z}[\mathbf{x}(\ell)]) \right) \text{ if } \mathbf{x} \in U^{\{1, \dots, n\}}, 0 \text{ otherwise.} \quad (8)$$

The point process S characterized by Equation (8) is a binomial point process of \mathbf{n} points in U with intensity $U \rightarrow \mathbb{R}, \mathbf{x} \rightarrow ((\mathbf{z} \cdot \nu)(U))^{-1} \mathbf{z}[\mathbf{x}]$, which we abbreviate by $S \sim \text{bpp}(\mathbf{z}, 10)$. Simple random sampling with replacement is a binomial point process with a constant intensity. For simplicity, we only consider exchangeable sample designs, in the sense that $\forall \mathbf{n} \in \mathbb{N}$, for all permutation τ of $\{1, \dots, n\}$, $P^{S|N=\mathbf{n}} = P^{S[\tau]|N=\mathbf{n}}$. For example, given a measurable real function $\mathbf{z} : U \rightarrow \mathbb{R}$, a spatial Poisson Process of intensity \mathbf{z} is a point process $S : \Omega \rightarrow \bigcup_{\mathbf{n} \in \mathbb{N}} U^{\mathbf{n}}$, such that for all ν -measurable subset A of U ,

$$S \sim \text{Ppp}(\mathbf{z}) \Leftrightarrow \text{cardinality}(S^{-1}[A]) \sim \text{Poisson}((\mathbf{z} \cdot \nu)(A)), \quad (9)$$

where $S^{-1}[A]$ is the random variable with domain the finite subsets of U defined by:
 $\omega \mapsto \{\ell \in \{1, \dots, N(\omega)\}; (S(\omega))(\ell) \in A\}$ if $N(\omega) > 0$, \emptyset otherwise.

The density of such process with respect to $\bar{\nu}$ is defined, for $\mathbf{n} \in \mathbb{N}$, $\mathbf{x} \in \bar{U}$, by:

$$f_S(\mathbf{x}) = (\text{size}(\mathbf{x})!)^{-1} \exp(-(\mathbf{z} \cdot \nu)(U)) \prod_{\ell \in \text{domain}(\mathbf{x})} \mathbf{z}[\mathbf{x}(\ell)]. \quad (10)$$

2.3.2 Random design variables, sample and random design

In practice, the design parameter \mathbf{z} is modeled as the output of a random process $Z : \Omega \rightarrow (U \rightarrow \mathcal{Z})$ that we will refer to as the design variable. The selection process, when controlled, is in practice a function of an auxiliary variable, called design variable, that is a process defined on the same space U . When the selection process is not chosen by the experimenter it can also be modelled as a function of such a process, that can be observed, partially observed or latent. In practice, it may not be reasonable to assume independence of the the design and study variables Z and Y .

The design is by definition a random variable with domain the set of probability distributions on \bar{U} . The sample is a random variable S with domain \bar{U} such that the distribution of S conditionnaly to the design *is* the design, e.g:

$$P^D - a.s.(\mathbf{d}), \quad P^{S|D=\mathbf{d}} = \mathbf{d}.$$

For a point process S with values in the measured space (U, ν) , for a random variable W , define $\Lambda_{S|W=w}$ as the intensity measure of S conditionally to $W = w$ with respect to the measure on U , where for each measurable subset A of U , $\Lambda_S(A) = E[\text{cardinality}(S^{-1}[A]) | W = w]$, and λ_S as the density of Λ_S with respect to ν : $\lambda_S = d\Lambda_S/d\nu$.

Example with simulations (continued): distribution of Z conditionally on Y

We assume that Z satisfies:

$$Z = \exp(\alpha + \beta Y + \gamma \varepsilon),$$

where α, β, γ are real positive numbers, $\varepsilon : \Omega \rightarrow (U \rightarrow \mathbb{R})$ is a Gaussian process with mean: $\mu = 0$ and Gaussian covariogram of parameters $c_0 = 0, c_1 = 5, c_2 = 10$.

The variable $\mathbf{y} : U \rightarrow \mathbb{R}$ (resp $e : U \rightarrow \mathbb{R}$) is generated once by sampling from Y (resp ε). The variable $\mathbf{z} = \exp(\alpha + \beta \mathbf{y} + \gamma \mathbf{e})$ is computed for 3 different values of the vector (α, β, γ) : $(\log(10), 0, 0)$, $(\log(10) - (0.5^2 + 0.3^2), 0, \sqrt{0.5^2 + 0.3^2})$, and $(\log(10) - (0.5^2 + 0.3^2), 0.5, 0.3)$. Two samples are drawn, one following $\mathbf{d} = \text{bpp}(\mathbf{z}, 10)$ and the other following $\mathbf{d} = \text{Ppp}(\mathbf{z})$. The variables \mathbf{z} and \mathbf{y} and the two samples are mapped in figures 2 and 3. Sampled units tend to concentrate where the sampling intensity is the highest.

Figure 2: Heat maps of the design variable \mathbf{z} and plot of realisations of S for three different design variables and designs.

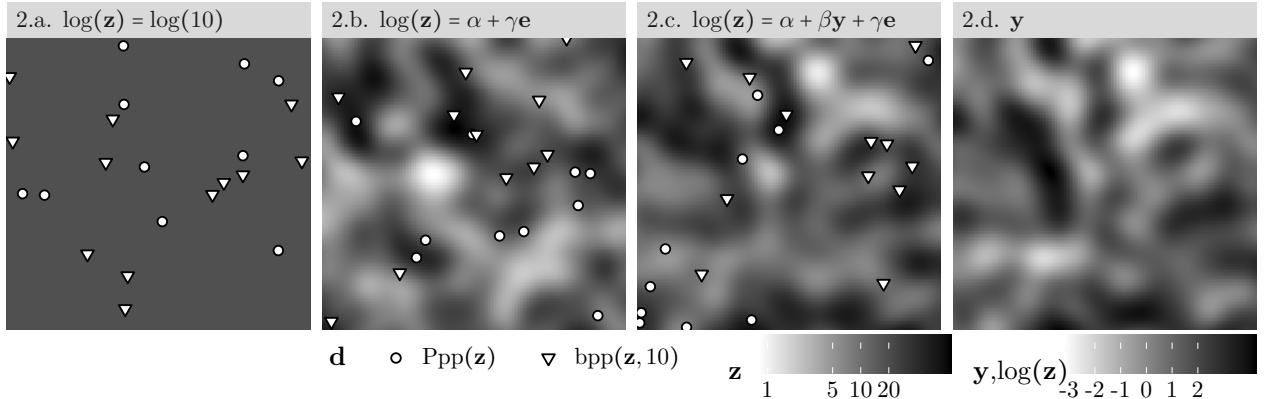
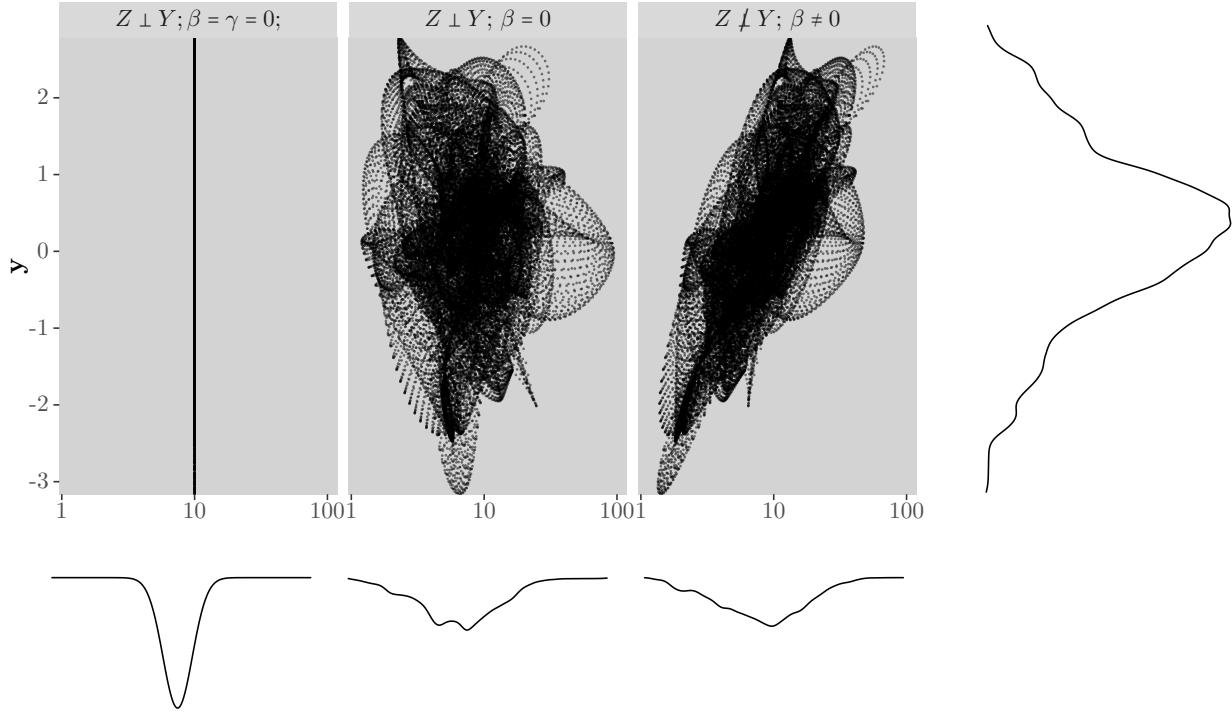


Figure 3: Joint and marginal densities of \mathbf{z} and \mathbf{y}



Each sub-figure contains the scatter plot of $\{(\mathbf{z}[\mathbf{x}], \mathbf{y}[\mathbf{x}]) : \mathbf{x} \in \text{Grid}\}$, where Grid is a regularly spaced grid of \mathbf{U} . The values of (α, β, γ) for each Sub-figure are: 3.1: $(\log(10), 0, 0)$, 3.2: $(\log(10) - (0.5^2 + 0.3^2), 0, \sqrt{0.5^2 + 0.3^2})$, 3.3: $(\log(10) - (0.5^2 + 0.3^2), 0.5, 0.3)$. The vertical axis corresponds to \mathbf{y} , the horizontal to \mathbf{z} . The marginal plots correspond to the density of \mathbf{z} (right margin) and to the densities of \mathbf{y} (bottom margins).

2.4 Observations

The observation consists of the realisations of the random variables S and $Y[S]$.

3 Sample distribution

3.1 Density ratio and weighted density.

In this section, we derive the distribution of the observed values of the signal on the sample, (e.g. the distribution of $Y[S]$) from the distribution of the design variable conditionally Z to the signal Y and the function that links the design to the design variable, or equivalently the distribution of the sample S conditionally to the Design variable Z . To this end, we proceed step by step and resort to the Bayes formula.

Definition 3.1. For a random set K , define $\rho_K(\cdot | \cdot)$ as any function that satisfies:

$$P^{(S_K, Y[S_K])} - \text{a.s}(\mathbf{x}, \mathbf{y}), \quad f_{Y[\mathbf{x}]|S_K=\mathbf{x}}(\mathbf{y}) = f_{Y[\mathbf{x}]}(\mathbf{y}) \rho_K(\mathbf{x} | \mathbf{y}) \quad (11)$$

Property 3.1.

$$P^{(S_K, Y[S_K])} - a.s(\mathbf{x}, \mathbf{y}), \quad \rho_K(\mathbf{x} | \mathbf{y}) = \frac{f_{S_K|Y[\mathbf{x}]}(\mathbf{x} | \mathbf{y})}{f_{S_K}(\mathbf{x})} \quad (12)$$

Proof. From the Bayes formula:

$$f_{Y[\mathbf{x}]|S_K=\mathbf{x}}(\mathbf{y}) = (f_{S_K}(\mathbf{x}))^{-1} f_{Y[\mathbf{x}], S_K}(\mathbf{y}, \mathbf{x}) \quad (13)$$

$$= (f_{S_K}(\mathbf{x}))^{-1} f_{S_K|Y[\mathbf{x}]}(\mathbf{x} | \mathbf{y}) f_{Y[\mathbf{x}]}(\mathbf{y}) \quad (14)$$

So, combining Equations (14) and (11):

$$\rho_K(\mathbf{x} | \mathbf{y}) = (f_{S_K}(\mathbf{x}))^{-1} f_{S_K|Y[\mathbf{x}]}(\mathbf{x} | \mathbf{y}) \quad (15)$$

□

The density ratio ρ can be derived from the distribution of Z :

$$\rho_K(\mathbf{x} | \mathbf{y}) = \left(\int f_{S_K|Z}(\mathbf{x} | \mathbf{z}) dP^Z \right)^{-1} \int f_{S_K|Z}(\mathbf{x} | \mathbf{z}) dP^{Z|Y[\mathbf{x}]=\mathbf{y}}(\mathbf{z}) \quad (16)$$

However, there is not necessarily a close form for the integration of $f_{S_K|Z}$ over \mathbf{z} , as shown in the two examples below.

Example (continued): Sample density ratios for $S \sim \text{Ppp}(Z)$ and $S \sim \text{bpp}(Z, n)$

Conditionally on $Y[\mathbf{x}] = \mathbf{y}$, Z is a log normal random process with distribution characterized by $E[\log(Z[\mathbf{x}']) | Y[\mathbf{x}] = \mathbf{y}] = \alpha + \beta(\mu + \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1} (\mathbf{y} - \mu))$ and $\text{Var}[\log(Z[\mathbf{x}']) | Y[\mathbf{x}] = \mathbf{y}] = \gamma^2 \Sigma_{\mathbf{x}'} + \beta^2 (\Sigma_{\mathbf{x}', \mathbf{x}'} - \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1} \Sigma_{\mathbf{x}, \mathbf{x}'})$.

Proof. See Appendix A.1

□

For $\mathbf{x} \in U^{\{1, \dots, n\}}$, when $S \sim \text{Ppp}(Z)$,

$$f_{S|W}(\mathbf{x} | \mathbf{w}) = \int \exp(-(\mathbf{z} \cdot \boldsymbol{\nu})(U)) (n!)^{-1} \left(\prod_{\ell=1}^n \mathbf{z}[\mathbf{x}(\ell)] \right) dP^{Z|W}(\mathbf{z} | \mathbf{w}), \quad (17)$$

and when $S \sim \text{bpp}(Z, n)$,

$$f_{S|W}(\mathbf{x} | \mathbf{w}) = \int (-(\mathbf{z} \cdot \boldsymbol{\nu})(U))^{-n} \left(\prod_{\ell=1}^n \mathbf{z}[\mathbf{x}(\ell)] \right) dP^{Z|W}(\mathbf{z} | \mathbf{w}). \quad (18)$$

A close form for (18) or (17) cannot be achieved when $W = \mathbf{w}$ is replaced by $Y[\mathbf{x}] = \mathbf{y}$ (for the numerator of ρ) or $1 = 1$ (for the denominator). And numerical approximation is

numerically intensive. However, we propose to use a very crude approximation: $\rho_K(\mathbf{x} \mid \mathbf{y}) \approx \tilde{\rho}_K(\mathbf{x} \mid \mathbf{y})$, with $\tilde{\rho}$ that is $P - a.s(S_K, Y[S_K])(\mathbf{x}, \mathbf{y})$ - a.s. defined:

$$\tilde{\rho}_K(\mathbf{x} \mid \mathbf{y}) = \frac{\mathbb{E} \left[\exp \left(\sum_{\ell \in \mathbf{k}} \beta Y[\mathbf{x}(\ell)] + \gamma \varepsilon[\mathbf{x}(\ell)] \right) \mid Y[\mathbf{x}] = \mathbf{y} \right]}{\mathbb{E} \left[\exp \left(\sum_{\ell \in \mathbf{k}} \beta Y[\mathbf{x}(\ell)] + \gamma \varepsilon[\mathbf{x}(\ell)] \right) \right]} \times \frac{P(K \cap \{1, \dots, N\}) = \mathbf{k} \mid Y[\mathbf{x}] = \mathbf{y}}{P(L \cap \{1, \dots, N\} = \mathbf{k})} \quad (19)$$

with $\mathbf{k} = \text{domain}(\mathbf{x})$.

A cruder approximation consists in neglecting the second factor in (19):

$$\tilde{\tilde{\rho}}_K(\mathbf{x} \mid \mathbf{y}) = \frac{\mathbb{E} \left[\exp \left(\sum_{\ell \in \mathbf{k}} \beta Y[\mathbf{x}(\ell)] + \gamma \varepsilon[\mathbf{x}(\ell)] \right) \mid Y[\mathbf{x}] = \mathbf{y} \right]}{\mathbb{E} \left[\exp \left(\sum_{\ell \in \mathbf{k}} \beta Y[\mathbf{x}(\ell)] + \gamma \varepsilon[\mathbf{x}(\ell)] \right) \right]}.$$

We obtain a close form for $\tilde{\tilde{\rho}}$:

$$P^{(S, Y[S])} - a.s.(\mathbf{x}, \mathbf{y}), \tilde{\tilde{\rho}}_{\{1, \dots, N\}}(\mathbf{x} \mid \mathbf{y}) = \exp \left(\mathbf{1}^T \beta (\mathbf{y} - \mu \mathbf{1}) - \frac{1}{2} \mathbf{1}^T (\beta^2 \Sigma_{\mathbf{x}, \mathbf{x}}) \mathbf{1} \right) \quad (20)$$

Proof. See Appendix A.2 □

This approximation is not satisfying: it does not depend on γ . We need to work on that by investigating the distribution of the integral of a lognormal point process. Cressie has some approximations that we could use.

3.2 Distribution of $Y[S]$

The density of $Y[S]$ with respect to $\bar{\eta}$ is defined by

$$f_{S, Y[S]}(\mathbf{x}, \mathbf{y}) = \rho_{\{1, \dots, N\}}(\mathbf{x} \mid \mathbf{y}) \times f_{Y[\mathbf{x}]}(\mathbf{y}) \times f_S(\mathbf{x}). \quad (21)$$

The "population distribution" is the distribution of Y , from which can be derived the distribution of $Y[\mathbf{x}]$ for any $n \in \mathbb{N}$, and any $\mathbf{x} \in U^n$. The distribution of Y is defined by all its finite dimensional distributions. The "sample distribution" is the distribution of an imaginary process Y^* such that its finite-dimensional distributions $f_{Y^*}[\mathbf{x}]$, for any $\mathbf{x} \in \bar{U}$ are given by

$$f_{Y^*}[\mathbf{x}] = \rho_{\{1, \dots, N\}}(\mathbf{x} \mid \mathbf{y}) \times f_{Y[\mathbf{x}]}.$$

Generalising M. Krieger and Pfeffermann (1992), the selection is non informative when $P^{Y[S], S} - a.s.(\mathbf{y}, \mathbf{x})$, $\rho(\mathbf{x}, \mathbf{y}) = 1$. In this case, we consider that the sample distribution corresponds to the population distribution.

This general definition of a sample process by opposition to the population process allows to define the sample counterpart of different characteristics of the population distribution as for example the intensity and the covariogram.

Although we gave a theoretical definition of ρ , as in practice a close form may be difficult to obtain, one can also define the approximated finite dimensional densities of Y^*

$$\tilde{f}_{Y^*[\mathbf{x}]}(\mathbf{y}) = \tilde{f}_{Y[S]|S}(\mathbf{y} \mid \mathbf{x}) = \tilde{\rho}_{\{1, \dots, N\}}(\mathbf{x} \mid \mathbf{y}) f_{Y[\mathbf{x}]}(\mathbf{y}).$$

3.3 Sample Intensity

3.4 Sample Variogram

Define the sample semi variogram for exchangeable designs as

$$G^*(h) = \frac{1}{2} E \left[(Y[S[1]] - Y[S[2]])^2 \mid S[2] - S[1] = h \right].$$

Property 3.2 (Relationship between G , G^* and ρ).

$$G(h)^* = \frac{1}{2} \int_{U^{\{1,2\}}} \left[\int_{\mathcal{Y}^{\{1,2\}}} (\mathbf{y}(2) - \mathbf{y}(1))^2 f_{Y[\mathbf{x}]}(\mathbf{y}) \rho_{\{1,2\}}(\mathbf{x}, \mathbf{y}) d\eta^{\otimes 2}(\mathbf{y}) \right] d(\nu^{\otimes \{1,2\}})^{X|X[2]-X[1]=h}(\mathbf{x})$$

Proof.

$$\begin{aligned} & E \left[(Y[\mathbf{x}(2)] - Y[\mathbf{x}(1)])^2 \mid S_{\{1,2\}} = \mathbf{x} \right] \\ &= \int_{\mathcal{Y}^2} (\mathbf{y}(2) - \mathbf{y}(1))^2 f_{Y[\mathbf{x}]|S_{\{1,2\}}}(\mathbf{y} \mid \mathbf{x}) d\eta^{\otimes \{1,2\}}(\mathbf{y}) \\ &= \int_{\mathcal{Y}^2} (\mathbf{y}(2) - \mathbf{y}(1))^2 \rho_{\{1,2\}}(\mathbf{x}, \mathbf{y}) f_{Y[\mathbf{x}]}(\mathbf{y}) d\eta^{\otimes \{1,2\}}(\mathbf{y}) \end{aligned}$$

□

4 Estimation

4.1 Naive estimators and their properties in the non informative selection case

A common approach to achieve a valid variogram estimator is composed by *estimation* and *fitting* of the variogram. In the former, an estimate of the variogram is obtained, while the latter phase is necessary since the estimators used in the first phase are usually not conditionally negative-definite.

Under the assumption of constant-mean, an estimator based on the method of moments is (Matheron, 1962)

$$2\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} (Y(\mathbf{x}_i) - Y(\mathbf{x}_j))^2, \forall h \in \mathbb{R}^d \quad (22)$$

where $N(h) = \{(\mathbf{x}_i, \mathbf{x}_j) : \mathbf{x}_i - \mathbf{x}_j = h; i, j = 1, \dots, n\}$ and $|N(h)|$ is the number of distinct pairs in $N(h)$.

When data are irregularly spaced in \mathbb{R}^d , we can use

$$2\hat{\gamma}_s(h(l)) = ave \{(Y(\mathbf{x}_i) - Y(\mathbf{x}_j))^2 : (\mathbf{x}_i, \mathbf{x}_j) \in N(h); h \in T(h(l))\} \quad (23)$$

where the region $T(h(l))$ is a specified tolerance region in \mathbb{R}^d around $h(l)$ $l = 1, \dots, K$
Check variable l .

A robust version of (22) is given by (Cressie and Hawkins, 1980)

$$2\hat{\gamma}_r(h) = \left\{ \frac{1}{|N(h)|} \sum_{N(h)} |Y(\mathbf{x}_i) - Y(\mathbf{x}_j)|^{1/2} \right\}^4 / (0.457 + 0.494/|N(h)|) \quad (24)$$

and corresponding smoothed version

$$2\hat{\gamma}_{rs} = [med\{|Y(\mathbf{x}_i) - Y(\mathbf{x}_j)|^{1/2} : (\mathbf{x}_i, \mathbf{x}_j) \in N(h); h \in T(h(l))\}]^4 / B(h) \quad (25)$$

where med is the median of the sequence and $B(h)$ is a correction-term for bias (asymptotically $B(h) = 0.457$).

A non-parametric approach to variogram estimation can be achieved by the use of kernel estimator

$$2\hat{\gamma}(h) = \frac{\sum_i \sum_j w_{ij}(h) (Y(\mathbf{x}_i) - Y(\mathbf{x}_j))^2}{\sum_i \sum_j w_{ij}(h)} \quad (26)$$

where $w_{ij} = K\left(\frac{h - \|\mathbf{x}_i - \mathbf{x}_j\|}{g}\right)$, K is a symmetric, zero-mean (bounded) and g is a positive number called bandwidth.

FP: I do not know if the following makes sense. To sum up, we can write a general estimator form for the *naive estimated variogram*

$$\hat{G}(h) = \sum_{ij \in S} \beta_{ij}(h) (Y(\mathbf{x}_i) - Y(\mathbf{x}_j))(Y(\mathbf{x}_i) - Y(\mathbf{x}_j))^T \quad (27)$$

where the term $\beta_{i,j}$ is a mapping that depends on the particular estimator used (and sample selected).

Once the estimated variogram is obtained (or *empirical*), a model is fitted to it in order to achieve a valid variogram. At this stage, we are searching for a valid variogram “closest” to the empirical one, and typically we look into a subset of valid variograms $P = \{2\gamma : 2\gamma(\cdot) = 2\gamma(\cdot; \theta); \theta \in \Theta\}$. The best element of P can be searched through several good-of-fit criteria. Maximum Likelihood estimator relies heavily on Gaussian assumption, while Least Squares method requires few assumption about $Y(\mathbf{x})$.

In particular, with ML we assume that the data Y are multivariate Gaussian $(\mathbf{X}\boldsymbol{\beta}, \Sigma_{\mathbf{x}, \mathbf{x}'; \theta})$. The negative loglikelihood is **FP: notation needs to be checked**.

$$L(\boldsymbol{\beta}, \theta) = (n/2) \log(2\pi) + (1/2) \log |\Sigma_{\mathbf{x}, \mathbf{x}'; \theta}| + (1/2) (Y - \mathbf{X}\boldsymbol{\beta})^T \Sigma_{\mathbf{x}, \mathbf{x}'; \theta}^{-1} (Y - \mathbf{X}\boldsymbol{\beta}) \quad (28)$$

with estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\theta}$ satisfying $L(\hat{\boldsymbol{\beta}}, \hat{\theta}) = \inf \{L(\boldsymbol{\beta}, \theta) : \boldsymbol{\beta} \in \mathbb{R}^q, \theta \in \Theta\}$.

The LS minimizing problem can be written as

$$\min \{(2\tilde{\gamma} - 2\gamma(\theta))^T W^{-1} (\tilde{\gamma} - \gamma(\theta))\} \quad (29)$$

where $2\tilde{\gamma}$ is the empirical variogram, $2\gamma(\theta)$ is a valid model variogram with the exact form known except the parameter θ , and W is a weight matrix. If W is an identity matrix, Ordinary Least Square criterion is employed; in case of $W = V$, with V variance-covariance matrix, we have Generalized Least Square criterion, and if V is diagonal, Weighted Least Square criterion is achieved.

Exact finite-sample distribution theory for estimators and corresponding variance estimators are available only in special circumstances (e.g. Jensen 1988). Therefore, simulations or approximation theory need to be employed in order to deal with intractable distribution theory. Zimmerman and Zimmerman (1991) presented a Monte Carlo comparison of different estimators, when two intrinsically stationary isotropic Gaussian random processes in \mathbb{R}^2 and various sampling intensities are taken into account. With regard to the MLE, different Authors agree that such approach can suffer from bias. Warnes and Ripley (1987) illustrate potential problems by simulations of some simple spatial process, and Mardia and Watkins (1989) indicate that these problems are caused by likelihoods not twice differentiable in θ . Moreover, a general conclusion that emerges from a set of several works based on simulations (Haining 1978c, Mardia and Meshall 1984, Swallow and Monahan 1984, Haining et al 1989, Zimmerman and Zimmerman 1991) is that the bias of MLE tends to be negative when the spatial dependence is positive, especially for small sample size.

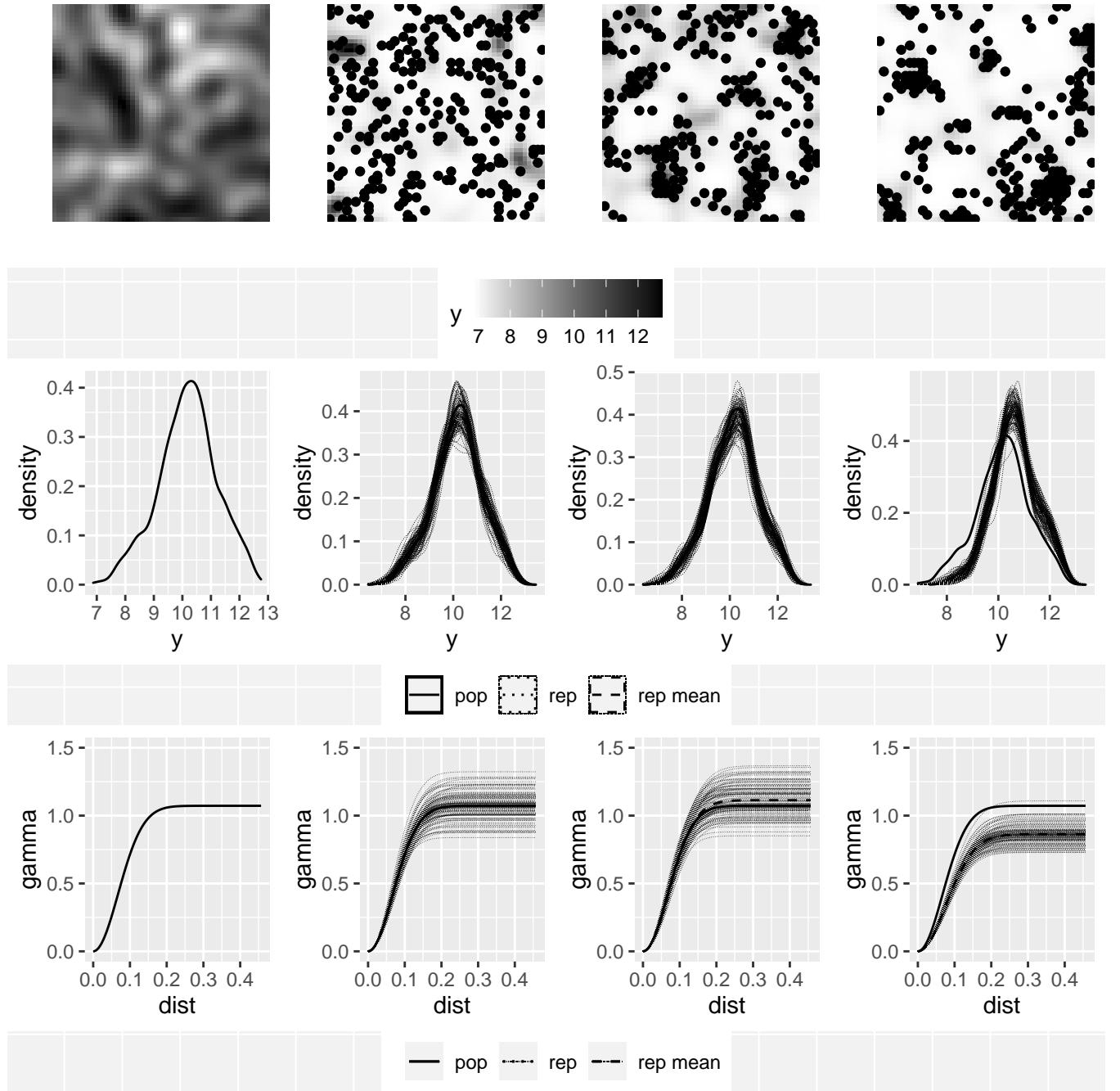
For approximation theory, Cressie talks about two type of asymptotic theories pp . *Infill asymptotics* refers to the the possibility to increase the sample size to infinity while the finite domain D is kept fixed, whereas *increase-domain asymptotics* refers to the situation where we take more units by means of increase the domain D .

4.2 Properties of the sample in the general case (including informative)

4.3 Naive estimation

Naive estimation of the population covariogram: $E[\text{naive Covariogram}] \neq \text{pop cov}$, $E[\text{naive Covariogram}]$ sample theoretical cov, with illustrative figures. I had to move it there. We cannot have this figure without having defined the estimates of the covariogram

Figure 4: Naive estimation of the semivariogram



5 Prediction

The variogram is a useful tool in order to quantify the spatial dependence in the data/process. Therefore, use of it in the inferential process has been investigated. We focus on the *kriging*

(Matheron, 1962), which is a minimum mean squared error method of spatial prediction that takes full advantage of the variogram. In fact, the kriging estimator incorporates the covariance structure of the Y into the weights. In particular, the weights are based on the covariance among sample points and the covariance between sample points and predicted points. Note that this does not happen in distance-based method, where the weights depend only on the location of the points.

In the follow just some notes about ordinary kriging, it needs to be completed/reviewed
Ordinary kriging assumes following model

$$Y(\mathbf{x}) = \mu + \delta(\mathbf{x}) \quad (30)$$

where $\mathbf{x} \in \mathbb{R}^d$, $\mu \in \mathbb{R}$, and μ unknown, and following predictor

$$P(Y, X) = \sum_{i=1}^n \lambda_i Y(\mathbf{x}) \quad (31)$$

The weights λ_i s are computed such that the estimator is unbiased and the variance is minimized. [Needs more here](#)

5.1 Accounting for informative selection

A Algebra for the example

A.1 Distribution of Z conditionally on $Y[\mathbf{x}] = \mathbf{y}$

Under the condition of the paper example, conditionally on $Y[\mathbf{x}] = \mathbf{y}$, the process $(\alpha + \beta Y + \gamma \varepsilon)$ is a Gaussian Process characterised by $E[(\alpha + \beta Y + \gamma \varepsilon)[\mathbf{x}']] = \alpha + \beta(\mu + \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1} (\mathbf{y} - \mu))$ and $\text{Var}[(\alpha + \beta Y + \gamma \varepsilon)[\mathbf{x}']] = \gamma^2 \Sigma_{\mathbf{x}', \mathbf{x}'} + \beta^2 (\Sigma_{\mathbf{x}', \mathbf{x}'} - \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1} \Sigma_{\mathbf{x}, \mathbf{x}'})$, and consequently, conditionally on $Y[\mathbf{x}] = \mathbf{y}$, $Z = \exp(\alpha + \beta Y + \gamma \varepsilon)$ is a lognormal spatial process.

Proof.

$$\begin{bmatrix} Y[\mathbf{x}'] \\ Y[\mathbf{x}] \end{bmatrix} \sim \text{Normal} \left(\mu, \begin{bmatrix} \Sigma_{\mathbf{x}', \mathbf{x}'} & \Sigma_{\mathbf{x}', \mathbf{x}} \\ \Sigma_{\mathbf{x}, \mathbf{x}'} & \Sigma_{\mathbf{x}, \mathbf{x}} \end{bmatrix} \right) \quad (32)$$

$$\Rightarrow [Y[\mathbf{x}'] | Y[\mathbf{x}] = \mathbf{y}] \sim \text{Normal} \left(\mu + \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1} (\mathbf{y} - \mu), \Sigma_{\mathbf{x}', \mathbf{x}'} - \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1} \Sigma_{\mathbf{x}, \mathbf{x}'} \right) \quad (33)$$

The independence of ε and Y implies that:

$$[\varepsilon[\mathbf{x}'] | Y[\mathbf{x}] = \mathbf{y}] \sim \text{Normal} (\mu_\varepsilon, \Sigma_{\mathbf{x}', \mathbf{x}'; \varepsilon}), \quad (34)$$

and that the vector obtained by stacking $Y[\mathbf{x}]$, $Y[\mathbf{x}']$, and $\varepsilon[\mathbf{x}']$ is normal. By combining (34) and (33) we obtain that:

$$\begin{bmatrix} [Y[\mathbf{x}']] \\ [\varepsilon[\mathbf{x}']] \end{bmatrix} \Big| Y[\mathbf{x}] = \mathbf{y} \sim \text{Normal} \left(\begin{bmatrix} \mu + \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1} (\mathbf{y} - \mu) \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{x}', \mathbf{x}'} - \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1} \Sigma_{\mathbf{x}, \mathbf{x}'} & 0 \\ 0 & \Sigma_{\mathbf{x}', \mathbf{x}'; \varepsilon} \end{bmatrix} \right).$$

We obtain the moments of $Z[\mathbf{x}'] = \gamma \varepsilon[\mathbf{x}'] + \alpha + \beta Y[\mathbf{x}']$ conditionally on $Y[\mathbf{x}] = \mathbf{y}$:

$$E[\alpha + \beta Y[\mathbf{x}'] + \gamma \varepsilon[\mathbf{x}'] | Y[\mathbf{x}] = \mathbf{y}] = \alpha + \beta(\mu + \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1} (\mathbf{y} - \mu)),$$

and

$$\text{Var}[\alpha + \beta Y[\mathbf{x}'] + \gamma \varepsilon[\mathbf{x}'] | Y[\mathbf{x}] = \mathbf{y}] = \gamma^2 \Sigma_{\mathbf{x}', \mathbf{x}'; \varepsilon} + \beta^2 (\Sigma_{\mathbf{x}', \mathbf{x}'} - \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1} \Sigma_{\mathbf{x}, \mathbf{x}'}).$$

□

A.2 Proof of Equation (35)

The first moment of a lognormal distribution of parameters m and s^2 is $\exp(m + s^2/2)$, so

$$\begin{aligned} E \left[\exp \left(\sum_{\ell \in \mathbf{k}'} \beta Y[\mathbf{x}'(\ell)] + \gamma \varepsilon[\mathbf{x}'(\ell)] \right) \mid Y[\mathbf{x}] = \mathbf{y} \right] \\ = \exp \left(\mathbb{1}^T \beta (\mu \mathbb{1} + \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1} (\mathbf{y} - \mu \mathbb{1})) + \frac{1}{2} \mathbb{1}^T (\gamma^2 \Sigma_{\mathbf{x}'} + \beta^2 (\Sigma_{\mathbf{x}', \mathbf{x}'} - \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1} \Sigma_{\mathbf{x}, \mathbf{x}'})) \mathbb{1} \right) \end{aligned}$$

which simplifies when $\mathbf{x}' = \mathbf{x}$ to

$$E \left[\exp \left(\sum_{\ell \in k} \beta Y[\mathbf{x}(\ell)] + \gamma \varepsilon[\mathbf{x}(\ell)] \right) \mid Y[\mathbf{x}] = \mathbf{y} \right] = \exp \left(\mathbf{1}^T \beta \mathbf{y} + \frac{1}{2} \mathbf{1}^T (\gamma^2 \Sigma_{\mathbf{x}, \mathbf{x}; \varepsilon}) \mathbf{1} \right)$$

and when $\mathbf{x} = \mathbf{0}$:

$$E \left[\exp \left(\sum_{\ell \in k'} \beta Y[\mathbf{x}'(\ell)] + \gamma \varepsilon[\mathbf{x}'(\ell)] \right) \right] = \exp \left(\mathbf{1}^T \beta \mu \mathbf{1} + \frac{1}{2} \mathbf{1}^T (\gamma^2 \Sigma_{\mathbf{x}', \mathbf{x}', \varepsilon} + \beta^2 (\Sigma_{\mathbf{x}', \mathbf{x}'})) \mathbf{1} \right)$$

So:

$$\tilde{\rho}_{\{1, \dots, N\}}(\mathbf{x} \mid \mathbf{y}) = \exp \left(\mathbf{1}^T \beta (\mathbf{y} - \mu \mathbf{1}) + \frac{1}{2} \mathbf{1}^T (-\beta^2 (\Sigma_{\mathbf{x}, \mathbf{x}})) \mathbf{1} \right) \quad (35)$$

A.3 Algebra for the denominator of ρ

The denominator of ρ requires the computation of $\int_U \Sigma_{\mathbf{x}', \mathbf{x}} d\nu$, which is itself a function of $\int_U C(\mathbf{x}' - \mathbf{x}_j) d\nu(\mathbf{x}')$

For $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$, such that $\|\mathbf{x}_1 - \mathbf{x}_2\| = h$, we have the following:

$$\begin{aligned} \Sigma_{\mathbf{x}, \mathbf{x}} &= \begin{bmatrix} C(0) & C(h) \\ C(h) & C(0) \end{bmatrix} \\ \Sigma_{\mathbf{x}, \mathbf{x}}^{-1} &= (C(0)^2 - C(h)^2)^{-1} \begin{bmatrix} C(0) & -C(h) \\ -C(h) & C(0) \end{bmatrix} \\ \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1} &= (C(0)^2 - C(h)^2)^{-1} [C(\mathbf{x}' - \mathbf{x}_1) \quad C(\mathbf{x}' - \mathbf{x}_2)] \begin{bmatrix} C(0) & -C(h) \\ -C(h) & C(0) \end{bmatrix} \\ &= \frac{[C(0)C(\mathbf{x}' - \mathbf{x}_1) - C(h)C(\mathbf{x}' - \mathbf{x}_2) \quad C(0)C(\mathbf{x}' - \mathbf{x}_2) - C(h)C(\mathbf{x}' - \mathbf{x}_1)]}{C(0)^2 - C(h)^2} \\ \Sigma_{\mathbf{x}', \mathbf{x}} &= [C(\mathbf{x}' - \mathbf{x}_1) \quad C(\mathbf{x}' - \mathbf{x}_2)] \end{aligned}$$

B R code

We generate a spatial Gaussian random process by means of the package `RandomFields`.

References

Bonnéry, D., Breidt, F. J., Coquet, F., et al. (2012). Uniform convergence of the empirical cumulative distribution function under informative selection from a finite population. *Bernoulli*, 18(4):1361–1385.

- Burgess, T. and Webster, R. (1980). Optimal interpolation and isarithmic mapping of soil properties: I the semi-variogram and punctual kriging. *Journal of soil science*, 31(2):315–331.
- Chilès, J. and Delfiner, A. (1999). *Geostatistics: Modelling spatial uncertainty*: Wiley Inter-science.
- Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.
- Cressie, N. and Hawkins, D. M. (1980). Robust estimation of the variogram: I. *Journal of the International Association for Mathematical Geology*, 12(2):115–125.
- M. Krieger, A. and Pfeffermann, D. (1992). Maximum likelihood estimation from complex sample surveys. *Survey Methodology*, 18(2):225–239.
- Mardia, K. and Watkins, A. (1989). On multimodality of the likelihood in the spatial linear model. *Biometrika*, 76(2):289–295.
- Matheron, G. (1962). *Traité de géostatistique appliquée. 1 (1962)*, volume 1. Editions Technip.
- Ord, J. and Rees, M. (1979). Spatial processes: Recent developments with applications to hydrology.
- Pfeffermann, D., Krieger, A. M., and Rinott, Y. (1998). Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica*, pages 1087–1114.
- Samra, J., Gill, H., and Bhatia, V. (1989). Spatial stochastic modeling of growth and forest resource evaluation. *Forest Science*, 35(3):663–676.
- Skinner, C. J., Holt, D., and Smith, T. F. (1989). *Analysis of complex surveys*. John Wiley & Sons.
- Thiebaux, H. and Pedder, M. (1987). Spatial objective analysis with applications in atmospheric science((book)). London and Orlando, FL, Academic Press, 1987, 308.
- Warnes, J. and Ripley, B. (1987). Problems with likelihood estimation of covariance functions of spatial gaussian processes. *Biometrika*, 74(3):640–642.
- Zimmerman, D. L. and Zimmerman, M. B. (1991). A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors. *Technometrics*, 33(1):77–91.
https://warwick.ac.uk/fac/sci/statistics/staff/academic-research/nichols/research/spatbayes/johnson_spatialpointproc.pdf
Kaar Matheron. Pfeffermann

C May be needed later

C.1 Algebra for $G^*(h)$

How to approximate

$$\bar{G}^*(h) \quad (36)$$

$$= \int_{U^2} g^*(\mathbf{x}_1, \mathbf{x}_2) d(\nu^{\otimes 2})^{(X_1, X_2) | X_2 - X_1 = h}(\mathbf{x}_1, \mathbf{x}_2) \quad (37)$$

$$= \int_{U^2} \left(\int_{\mathcal{Y}^2} \frac{(\mathbf{y}_2 - \mathbf{y}_1)^2}{2} f_{Y[\mathbf{x}]}(\mathbf{y}_1, \mathbf{y}_2) \rho_{\{1,2\}}(\mathbf{x}, \mathbf{y}) d\eta^{\otimes 2}(\mathbf{y}) \right) d(\nu^{\otimes 2})^{(X_1, X_2) | X_2 - X_1 = h}(\mathbf{x}_1, \mathbf{x}_2) \quad (38)$$

$$= \int_{U^2} \left(\int_{\mathcal{Y}^2} \frac{(\mathbf{y}_2 - \mathbf{y}_1)^2}{2} \frac{\exp\left(\frac{-(\mathbf{y} - \mu \mathbb{1})^T A(h)^{-2}(\mathbf{y} - \mu \mathbb{1})}{2}\right)}{2\pi \det(A(h))} \rho_{\{1,2\}}(\mathbf{x}, \mathbf{y}) d\eta^{\otimes 2}(\mathbf{y}) \right) d(\nu^{\otimes 2})^{X | X_2 - X_1 = h}(\mathbf{x}) \quad (39)$$

$$= G(h) \int_{\mathcal{Y}^2} (\mathbf{y}'_2 - \mathbf{y}'_1)^2 \frac{e^{-\frac{(\mathbf{y}')_1^2 + (\mathbf{y}')_2^2}{2}}}{2\pi} \left(\int_U \rho_{\{1,2\}}((\mathbf{x}', \mathbf{x}' + h e_\alpha), \mu + A(h)\mathbf{y}) \gamma(\mathbf{x}', h) d\nu(\mathbf{x}') \right) d\eta^{\otimes 2}(\mathbf{y}) \quad (40)$$

Variable change : $(\mathbf{y}'_1, \mathbf{y}'_2) = (A_h^{-1}(\mathbf{y}_1 - \mu), A_h^{-1}(\mathbf{y}_2 - \mu))$;

$\mathbf{y} = A(h)\mathbf{y}' + \mu d\eta(\mathbf{y}) = \det(A(h))d\eta(\mathbf{y}')$; $\frac{1}{2}(\mathbf{y}_1 - \mathbf{y}_2)^2 = G(h)(\mathbf{y}'_1 - \mathbf{y}'_2)^2$

Variable change : $(\mathbf{x}'_1, \alpha) = (\mathbf{x}_1, \cos(\text{angle}((0, 1), \mathbf{x}_2 - \mathbf{x}_1)))$

Setup a grid $\tilde{U} \subset U$ (triangular tiling with h).

Draw $R = 1000$, \mathbf{y} from a $\text{Normal}((0, 0), Id_2)$. and transform to get a $\text{Normal}((0, 0), \Sigma_{\mathbf{x}, \mathbf{x}})$ by multiplying the vector by the matrix $\Sigma_{\mathbf{x}, \mathbf{x}}^{1/2}$ which is defined as $\Sigma_{\mathbf{x}, \mathbf{x}}$ is symmetric positive

$$A(h) = \begin{bmatrix} C(0) & C(h) \\ C(h) & C(0) \end{bmatrix}^{1/2}$$

$$A(h) = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} C(0) + C(h) & 0 \\ 0 & C(0) - C(h) \end{bmatrix}^{\frac{1}{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$A(h) = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} (C(0) + C(h))^{\frac{1}{2}} & 0 \\ 0 & (C(0) - C(h))^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} (C(0) + C(h))^{\frac{1}{2}} & (C(0) + C(h))^{\frac{1}{2}} \\ (C(0) - C(h))^{\frac{1}{2}} & -(C(0) - C(h))^{\frac{1}{2}} \end{bmatrix}$$

$$= \frac{1}{2} \begin{bmatrix} (C(0) + C(h))^{\frac{1}{2}} + (C(0) + C(h))^{\frac{1}{2}} & (C(0) + C(h))^{\frac{1}{2}} - (C(0) - C(h))^{\frac{1}{2}} \\ (C(0) + C(h))^{\frac{1}{2}} - (C(0) + C(h))^{\frac{1}{2}} & (C(0) + C(h))^{\frac{1}{2}} + (C(0) - C(h))^{\frac{1}{2}} \end{bmatrix}$$

$$\begin{aligned}
A(h)\mathbf{y} &= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} (C(0) + C(h))^{\frac{1}{2}} & 0 \\ 0 & (C(0) - C(h))^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \mathbf{y} \\
&= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} (C(0) + C(h))^{\frac{1}{2}} & 0 \\ 0 & (C(0) - C(h))^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 + \mathbf{y}_2 \\ \mathbf{y}_1 - \mathbf{y}_2 \end{bmatrix} \\
&= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} (C(0) + C(h))^{\frac{1}{2}}(\mathbf{y}_1 + \mathbf{y}_2) \\ (C(0) - C(h))^{\frac{1}{2}}(\mathbf{y}_1 - \mathbf{y}_2) \end{bmatrix} \\
&= \frac{1}{2} \left[(C(0) + C(h))^{\frac{1}{2}}(\mathbf{y}_1 + \mathbf{y}_2) + (C(0) - C(h))^{\frac{1}{2}}(\mathbf{y}_1 - \mathbf{y}_2) \right] \\
&\quad - \left[(C(0) + C(h))^{\frac{1}{2}}(\mathbf{y}_1 + \mathbf{y}_2) - (C(0) - C(h))^{\frac{1}{2}}(\mathbf{y}_1 - \mathbf{y}_2) \right]
\end{aligned}$$

$$\begin{aligned}
(A(h)\mathbf{y})_1 \times (A(h)\mathbf{y})_2 &= \frac{1}{4} \left(((C(0) + C(h))(\mathbf{y}_1 + \mathbf{y}_2)^2) \right. \\
&\quad \left. - ((C(0) - C(h))(\mathbf{y}_1 - \mathbf{y}_2)^2) \right) \\
&= \frac{1}{4} \left(((C(0) + C(h))(\mathbf{y}_1^2 + \mathbf{y}_1\mathbf{y}_2 + \mathbf{y}_2^2)) \right. \\
&\quad \left. - ((C(0) - C(h))(\mathbf{y}_1^2 - \mathbf{y}_1\mathbf{y}_2 + \mathbf{y}_2^2)) \right) \\
&= \frac{1}{2} (C(h)(\mathbf{y}_1^2 + \mathbf{y}_2^2) + 2C(0)\mathbf{y}_1\mathbf{y}_2)
\end{aligned}$$

$$\begin{aligned}
(\mu + A(h)\mathbf{y})_1 \times (\mu + A(h)\mathbf{y})_2 &= \mu^2 + 2\mu((A(h)\mathbf{y})_1 + (A(h)\mathbf{y})_2) + \frac{1}{2} (C(h)(\mathbf{y}_1^2 + \mathbf{y}_2^2) + 2C(0)\mathbf{y}_1\mathbf{y}_2) \\
&= \mu^2 + 2\mu((C(0) + C(h))^{\frac{1}{2}}(\mathbf{y}_1 + \mathbf{y}_2)) + \frac{1}{2} (C(h)(\mathbf{y}_1^2 + \mathbf{y}_2^2) + 2C(0)\mathbf{y}_1\mathbf{y}_2)
\end{aligned}$$

$$(A(h) \times \mathbf{y}) + \mu \sim \text{Normal}((\mu, \mu), \Sigma_{\mathbf{x}, \mathbf{x}})$$

$$\begin{aligned}
B(h) &= A(h) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} A(h) \\
&= (C(0) - C(h)) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}
\end{aligned}$$

$$\mathbf{y}^T B(h) \mathbf{y} = (C(0) - C(h))(\mathbf{y}_1 - \mathbf{y}_2)^2$$

C.2 Likelihood

Let $\mathbf{n} \in \mathbb{N}$, $\mathbf{x} \in U^n$, $\mathbf{y} \in \mathcal{Y}^n$ then the likelihood is defined as:

$$\mathcal{L}_{S,Y[S]}(\theta, \xi; \mathbf{x}, \mathbf{y}) = f_{Y[\mathbf{x}]}(\mathbf{y}; \theta) \times f_{S|Y[\mathbf{x}]|N}(\mathbf{x} \mid \mathbf{y}; \theta, \xi)$$

C.3 Intensity ratios

In the litterature, ρ can be interpreted as the density ratio, in this section we show that ρ can also be interpreted as an intensity ratio. Let W be a random variable, define the intensity ratio of the point process S conditionally on $W = w$ as :

$$\rho_{S|W}(\cdot | w) : U \rightarrow, \mathbf{x} \mapsto \rho_{S|W}(\mathbf{x} | w) = \frac{\lambda_{S|W}(\mathbf{x} | w)}{\lambda_S(\mathbf{x})} \quad (41)$$

Property C.1 (Intensity ratios). *Let $\mathbf{x} \in U$, let $\mathbf{y} \in \mathcal{Y}$, then*

$$\lambda_{(S,Y[S])}(\mathbf{x}, \mathbf{y}) = f_{Y[\mathbf{x}]}(\mathbf{y}) \times \rho_{S|Y[\mathbf{x}]}(\mathbf{x} | \mathbf{y}) \times \lambda_S(\mathbf{x}) \quad (42)$$

The interest of this formulation is to distinguish between a non informative and an informative selection. In the statistical frameworks used in this paper, selection is informative when $\rho_{S|Y[S]} \neq 1$. Ignoring the selection mechanism consists in ignoring the term in $\rho_{S|Y[S]}$ in the expression of the intensity. When the selection mechanism D is independent of Y , then necessarily $\rho_{S|Y[S]} = 1$. Independence of D and Y is a sufficient condition for non informativeness.

For $\delta \subset \mathbb{N}$, and a point process S on U , define S^δ as the point process of K -uples of S as the sample process in U^δ :

$$S^\delta = \{S_{\{\ell_1, \dots, \ell_\delta\}} : \{\ell_1, \dots, \ell_\delta\} \subset \{1, \dots, N\} \text{ and } \ell_1 < \dots < \ell_\delta\}.$$

Note that S^δ is the empty set when $\delta > N$.

Corollary C.2. *Let $\delta \in \mathbb{N}$, let $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_\delta) \in U^\delta$, let $\mathbf{y} = (\mathbf{y}(1), \dots, \mathbf{y}_\delta) \in \mathcal{Y}^\delta$, then*

$$\lambda_{(S,Y[S])^\delta}(\mathbf{x}, \mathbf{y}) = f_{Y[\mathbf{x}]}(\mathbf{y}) \times \rho_{S^\delta|Y[\mathbf{x}]}(\mathbf{x} | \mathbf{y}) \times \lambda_{S^\delta}(\mathbf{x}) \quad (43)$$

Proof. We apply Property C.1 to the point process $(S, Y[S])^\delta$. \square

To characterize informative selection, ? does not use an intensity ratio but a density ratio. The use of density ratios is only suitable for the case of fixed size sampling under the exchangeable condition. In this subsection, we show that the intensity ratio equals the density ratio in the case of fixed size sampling under the exchangeable condition. The use of intensity ratio is another way to characterize informative selection that coincides with the use of density ratio when suitable.

Property C.3 (Relationship between density and intensity ratios for fixed size sampling). *Let S be a fixed size exchangeable sample. Let $\delta \in \{1, \dots, n\}$, and let $K \subset \{1, \dots, n\}$ such that $\text{cardinality}(K) = \delta$. Then for $\mathbf{x} \in U^\delta$, $\mathbf{y} \in \mathcal{Y}^\delta$,*

$$\rho_{S^\delta|Y[\mathbf{x}]}(\mathbf{x} | \mathbf{y}) = \frac{f_{S_K|Y[\mathbf{x}]}(\mathbf{x} | \mathbf{y})}{f_{S_K}(\mathbf{x})} \quad (44)$$

and

$$f_{Y[\mathbf{x}]|S_K}(\mathbf{y} | \mathbf{x}) = \rho_{S^\delta|Y[\mathbf{x}]}(\mathbf{x} | \mathbf{y}) \times f_{Y[\mathbf{x}]}(\mathbf{y}) \quad (45)$$

Proof. See Appendix ??.

□

Property C.4 (Intensity and density of fixed size exchangeable samples). *For fixed size sampling of size n , under the exchangeability of the sample index assumption, the following relationship holds:*

$$\forall \delta \in \{1, \dots, n\}, \forall \mathbf{x} \in U^\delta, \forall \mathbf{y} \in Y^\delta, \forall L \subset \{1, \dots, n\} \text{ such that } \text{cardinality}(L) = \delta,$$

$$\lambda_{S^\delta|Y[\mathbf{x}]}(\mathbf{x} | \mathbf{y}) = B_{n,\delta} \times f_{S_K|Y[\mathbf{x}]}(\mathbf{x} | \mathbf{y}).$$

Where $B_{n,\delta} = \delta! \text{cardinality}\{K \mid K \subset \{1, \dots, n\}, \text{cardinality}(K) = \delta\}$

Examples (continued)

For $\delta \in \mathbb{N}$, $\mathbf{x} \in U^\delta$, $\mathbf{y} \in Y^\delta$ then when $D = \text{Ppp}[Z]$,

$$\rho_{S^\delta|Y[\mathbf{x}]} = . \quad (46)$$

Proof. See Appendix ??

□

And when $D = \text{bpp}[Z, n]$

$$\rho_{S^\delta|Y[\mathbf{x}]} = . \quad (47)$$

Proof. See Appendix

□

Add plots

C.4 Sample distribution

Let $\mathbf{n} \in \mathbb{N}$, $\mathbf{x} \in U^\mathbf{n}$, $\mathbf{y} \in Y^\mathbf{n}$

$$f_{S,Y[S]}(\mathbf{x}, \mathbf{y}) = f_{Y[\mathbf{x}]}(\mathbf{y}; \theta) \times f_{S|Y[S]}(\mathbf{x} | \mathbf{y})$$

As mentioned, the codomain of the sample S is \bar{U} . For $\mathbf{n} \in \mathbb{N}$, $\mathbf{x} \in U^{\{1, \dots, \mathbf{n}\}}$,

$$f_S(\mathbf{x}) = P(n = \mathbf{n}) \times \frac{dP^{S|N=\mathbf{n}}}{d\nu^{\otimes \mathbf{n}}}(\mathbf{x})$$

and for $\mathbf{y} \in Y^\mathbf{n}$,

$$f_{S,Y[S]}(\mathbf{x}, \mathbf{y}) = P(n = \mathbf{n}) \times \frac{dP^{S,Y[S]|n=\mathbf{n}}}{d(\nu^{\otimes n} \otimes \eta^{\otimes n})}(\mathbf{x}, \mathbf{y})$$

$$f_{S|Y[S]}(\mathbf{x} | \mathbf{y}) = P(n = \mathbf{n} \mid Y[\mathbf{x}] = \mathbf{y}) \times \frac{dP^{S|Y[\mathbf{x}]=\mathbf{x}, n=\mathbf{n}}}{d\nu^{\otimes n}}(\mathbf{x} | \mathbf{y})$$

so

$$f_{S|Y[S]}(\mathbf{x} | \mathbf{y}) = P(n = \mathbf{n} \mid Y[\mathbf{x}] = \mathbf{x}) \times \frac{dP^{S|Y[\mathbf{x}]=\mathbf{x}, n=\mathbf{n}}}{d\nu^{\otimes n}}(\mathbf{x} | \mathbf{y})$$

$$f_S(\mathbf{x}) = \int \frac{dD}{d\nu}(\mathbf{x}) dP^D$$

Examples (continued)

For $\delta \in \mathbb{N}$, $\mathbf{x} \in U^\delta$, $\mathbf{y} \in \mathcal{Y}^\delta$ then when $D = \text{Ppp}[Z]$,

$$\begin{aligned} f_S(\mathbf{x}) &= \int \frac{dD}{d\nu}(\mathbf{x}) dP^D \\ &= \int \exp(-(Z.\nu)(U)) \frac{((Z.\nu)(U))^{\mathbf{n}}}{\mathbf{n}!} \frac{\prod_{\ell=1}^{\mathbf{n}} Z[\ell]}{((Z.\nu)(U))^{\mathbf{n}}} dP^Z \\ &= \int \exp(-(Z.\nu)(U)) \frac{\prod_{\ell=1}^{\mathbf{n}} Z[\ell]}{\mathbf{n}!} dP^Z \\ &= \int \exp(-(Z.\nu)(U)) \frac{\prod_{\ell=1}^{\mathbf{n}} Z[\ell]}{\mathbf{n}!} dP^Z \end{aligned}$$

Proof. See Appendix ?? □

And when $D = \text{bpp}[Z, n]$

$$\rho_{S^\delta | Y[\mathbf{x}]} = . \quad (48)$$

C.5 Proof of Property C.1

Proof. Let A_1, A_2 measurable subsets of U and \mathcal{Y} . For $\mathbf{x} \in U$, define the random variable $\mathbf{n}_x : \Omega \rightarrow \mathbb{N}, \omega \mapsto \text{cardinality}(\{\ell \in \text{domain}(S(\omega)) \mid ((S(\omega))(\ell)) = \mathbf{x}\})$. When U and \mathcal{Y} are discrete, and η and ν are the counting measures, then

$$\Lambda_{(S, Y[S])}(A_1 \times A_2) = \sum_{(\mathbf{x}, \mathbf{y}) \in A_1 \times A_2} \sum_{\mathbf{n}^* \in \mathbb{N}} \mathbf{n}^* P(\{\mathbf{n}_x = \mathbf{n}^*\} \cap \{Y[\mathbf{x}] = \mathbf{y}\})$$

So,

$$\begin{aligned} \lambda_{S, Y[S]}(\mathbf{x}, \mathbf{y}) &= \sum_{\mathbf{n}^* \in \mathbb{N}} \mathbf{n}^* P(\{\mathbf{n}_x = \mathbf{n}^*\} \cap \{Y[\mathbf{x}] = \mathbf{y}\}) \\ &= \sum_{\mathbf{n}^* \in \mathbb{N}} \mathbf{n}^* P(\{\mathbf{n}_x = \mathbf{n}^*\} \mid \{Y[\mathbf{x}] = \mathbf{y}\}) P(\{Y[\mathbf{x}] = \mathbf{y}\}) \\ &= \lambda_{S, Y[S] | Y[\mathbf{x}]}(\mathbf{x}, \mathbf{y}) \times P(\{Y[\mathbf{x}] = \mathbf{y}\}) \end{aligned}$$

When transposing the same reasoning to the general case, we obtain that:

$$\Lambda_{(S, Y[S])}(A_1 \times A_2) = \int_{A_1 \times A_2} [\lambda_{S | Y[\mathbf{x}]}(\mathbf{x} \mid \mathbf{y}) \times f_{Y[\mathbf{x}]}(\mathbf{y})] d(\nu \otimes \eta)(\mathbf{x}, \mathbf{y})$$

□