

# The effect of Informative Selection on the estimation of parameters related to Spatial Processes

Daniel Bonn  ry \*     Francesco Pantalone<sup>†</sup>     M. Giovanna Ranalli<sup>‡</sup>

## Abstract

This paper extends the concepts of informative selection, population distribution and sample distribution to a spatial process context. These notions were first defined in a context where the output of the random process of interest consists of independent and identically distributed realisations for each individual of a population. It has been showed that informative selection induces stochastic dependence among realisations on selected units. In the context of spatial processes, 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 from a spatial process and how the sample distribution differs from the population distribution.

**Keywords:** analytic inference, point process, sample distribution, population distribution, sample likelihood, unequal probability sample, variogram.

## 1 Introduction

Spatial processes are employed in many fields, such as geology, Earth science, environmental and agricultural surveys, among many others. A large variety of data is surveyed in space, such as rainfall (Ord and Rees, 1979), atmospheric (Thiebaux and Pedder, 1987), forestry (Samra et al., 1989) and soil data (Burgess and Webster, 1980), in order to perform statistical inference on parameters of interest. In these applications, a general approach is to postulate a spatial process for the variable  $Y$  under investigation, that is a stochastic process is assumed to have generated the population values, also called *superpopulation* model. When using sample data, we need to pay attention to the relation between the spatial process and

---

\*Epidemiology and Modelling Group, Department of Plant Sciences, University of Cambridge, UK

<sup>†</sup>Department of Economics, University of Perugia, Italy

<sup>‡</sup>Department of Political Science, University of Perugia, Italy

the selection mechanism. Indeed, the distribution of the observed data can be different from the distribution assumed for the population by means of the spatial process  $P^Y$ . 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 bias 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).

In this paper, we study the effect of informative selection when a spatial process is employed. 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. We start from the notion of population and sample distribution defined by Krieger and Pfeffermann (1992), and we extend those to a spatial process context. Indeed, Krieger and Pfeffermann (1992) consider realizations of a random process of interest as independent and identically distributed, whereas we address the dependence between the realizations, which very likely characterizes a spatial process. As in Pfeffermann et al. (1998), we start from the distribution of the observed values given they were selected, or *sample pdf*. While Pfeffermann et al. (1998) consider the observations as if they were independently distributed according to the sample pdf, our work considers the dependence between the observed units in order to account for the spatial component of the population. Also, while Pfeffermann et al. (1998) assume that the target population  $U$  is finite, our work can be applied to both finite population and continuous populations.

Throughout the paper we assume to have a spatial Gaussian process and a point process that represents the selection mechanism. Given these two processes, we then define the probability density function of the sample and a subsample, and the “sample” and “population” distribution of the signal. These concepts are needed in order to introduce a *density ratio* that we use for investigating the bias possibly introduced by the selection mechanism, or *selection bias*. In doing that, the spatial structure of the population is taken into account.

These general definitions allow us to define the sample counterpart of different characteristics of the population distribution. Indeed, we focus on the *variogram*, which analyzes

the degree of spatial dependence of spatial 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. We introduce the definitions of *population variogram* and *sample variogram*, and we investigate the behaviour 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 briefly reviewed and a simulation study is carried out in order to investigate the behaviour of the naive estimator. In Section 5 we provide the formula for the predictive distribution of the signal conditioned to observed values. Finally, Section 6 provides conclusion and future research.

## 2 Statistical Framework

This section introduces a statistical framework that describes the relationship between the design and sample on one side, and the variable of interest on the other sides. The following concepts borrow to the general statistical framework by Bonnery and Sedransk (2019). The definitions and most properties we provide apply in this framework. We also provide a case study example throughout the paper to illustrate the general definitions and properties. Some results presented here only apply to the specific statistical framework used for the example.

### 2.1 General notation

In this paper, all random variables are defined on a probability space  $(\Omega, \mathcal{A}, P)$ . 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 associates  $g(x)$  to  $x$ . 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 the application  $\mathbf{0}$  with empty domain. The application  $\mathbf{0}$  can be interpreted as the empty sample that matches no draw to the population of interest. 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  $\eta$  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  $\bar{E}$ :  $\bar{\eta} = \eta^{\otimes \emptyset} + \sum_{n \in \mathbb{N}, n \geq 0} \eta^{\otimes \{1, \dots, n\}}$ .

We apply the following rule: Random variables are capital Roman letters, whereas Random variable realisations are bold corresponding lowercase roman letter. By convention, any sum over an empty set is 0 and any product over an empty set is 1. The notation  $f_{V|W}$  denotes the density of  $V$  conditional on  $W$  with respect to a dominating measure on the domain of  $V$ .

## 2.2 Spatial process

We consider a space  $U$ , that is a compact 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:  $Y[S] : \Omega \rightarrow (\{1,2\} \rightarrow \mathcal{Y}), \omega \mapsto (Y(\omega))(S(\omega)) : \ell \mapsto (Y(\omega))(S(\omega))(\ell)$ . Examples and illustrations will be given for the particular case where  $U = [0,1]^2$ . Let  $\eta$  denote a sigma-finite measure on the set  $\mathcal{Y}$  and let  $\nu$  denote a probability measure on  $U$ . The  $\nu$ -averaged 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  $\nu$ -averaged 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 :  $\forall \mathbf{x} \in U^{\{1,2\}}$ ,

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

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

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

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

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

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 to 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. The covariance structure of the signal is then fully characterized by  $C(0)$  and  $G(h), h \neq 0$ . For the simulations and illustrations in this paper, a Gaussian covariogram of the form

$$C(h) = \theta_1 \exp(-\|h\|^2/\theta_2^2) \quad (7)$$

will be used, where the deviation parameter  $\theta_1$ , and the scale parameter  $\theta_2$  are real positive numbers. See Chilès and Delfiner (1999, p. 80) for alternative models.

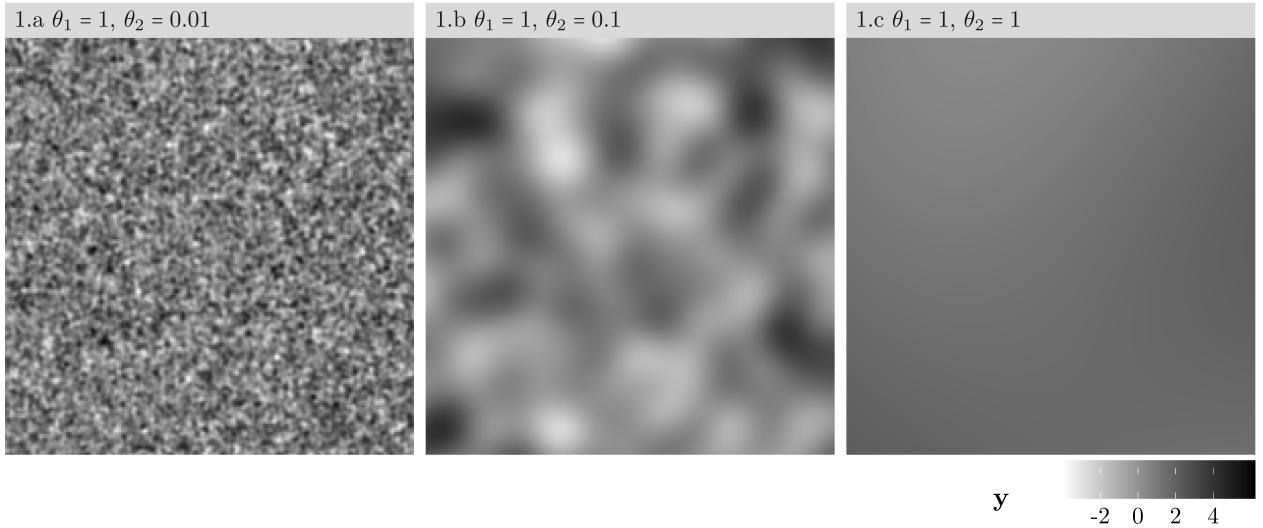
**Definition 2.2** (Isotropic Gaussian Random Process). A *Gaussian random process*  $Y : \Omega \rightarrow (U \rightarrow \mathcal{Y})$  is such that its distribution can be derived from the distributions of  $Y[\mathbf{x}]$ , where  $\mathbf{x} \in \bar{U}$ . For  $\mathbf{x}, \mathbf{x}' \in \bar{U}$ , denote the expected value of the signal by  $\mu : \bar{U} \rightarrow \mathcal{Y}, \mathbf{x} \mapsto E[Y[\mathbf{x}]]$ , and the covariance of the random vectors  $Y[\mathbf{x}], Y[\mathbf{x}']$  by  $\Sigma_{Y;\mathbf{x},\mathbf{x}'} = \text{Cov}[Y[\mathbf{x}], Y[\mathbf{x}']]$ . The distribution of  $Y$  is then fully characterized by  $\mu$  and  $\Sigma$ : for  $n \in \mathbb{N}$ ,  $\mathbf{x} \in U^{\{1,\dots,n\}}$ ,  $Y[\mathbf{x}]$  has the following density with respect to  $\eta^{\otimes\{1,\dots,n\}}$ :

$$f_{Y[\mathbf{x}]}(\mathbf{y}) = \left(2\pi^{n/2}|\Sigma_{Y;\mathbf{x},\mathbf{x}}|^{\frac{1}{2}}\right)^{-1} \exp\left(-\frac{1}{2}(\mathbf{y} - \mu(\mathbf{x}))\Sigma_{Y;\mathbf{x},\mathbf{x}}^{-1}(\mathbf{y} - \mu(\mathbf{x}))^T\right). \quad (8)$$

The distribution of an *isotropic Gaussian process* is such that  $\forall \mathbf{x}, \mathbf{x}' \in U, \Sigma_{Y;\mathbf{x},\mathbf{x}'} = C(\|\mathbf{x} - \mathbf{x}'\|)$  and is fully characterized by  $\mu$  and  $C$ .

**Example 2.1** (Isotropic Gaussian process  $Y$  with Gaussian Covariogram). *We simulate three independent replications of an isotropic Gaussian process with  $Y : \Omega \rightarrow (\mathbf{U} = [0, 1]^2 \rightarrow \mathbb{R})$ , with  $\forall \mathbf{x} \in \mathbf{U}$ ,  $\mu[\mathbf{x}] = 0$ , with a Gaussian Covariogram (7) with deviation parameter  $\theta_1 = 1$ , and scale parameter  $\theta_2 = \{0.01, 0.1, 1\}$ , respectively. Figure 1 represents independent realizations of  $Y$ .*

Figure 1: Heat maps of realisations of the random process  $Y : \Omega \rightarrow (\mathbf{U} = [0, 1]^2 \rightarrow \mathbb{R})$  for different values of  $\theta_2$ . For each realisation  $\mathbf{y}$  of the random process  $Y$ , for each  $\mathbf{x}$  in  $\mathbf{U}$ , 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 design variables and fixed designs

By definition, a fixed design  $\mathbf{d}$  is a probability distribution on  $\bar{\mathbf{U}}$ . 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  $\mathbf{x} \in \mathbf{U}^{\{1, \dots, n\}}$ ,  $P(N = 0)$  if  $\mathbf{x} = \mathbf{0}$ . A fixed design variable is a function  $\mathbf{z} : \mathbf{U} \mapsto \mathcal{Z}$ . A fixed design  $\mathbf{d}$  is usually defined as a function of a fixed design variable and characterized by its density with respect to  $\bar{\nu}$ . For example, the Probability Proportional to Size  $\mathbf{z}$  With Replacement and size  $n$  design, with  $\mathcal{Z} = [0, +\infty)$ , for  $\mathbf{x} \in \bar{\mathbf{U}}$ , is characterized by:

$$f_S(\mathbf{x}) = (dP^S/d\bar{\nu})(\mathbf{x}) = \begin{cases} \left( \int_U z(\mathbf{x}') d\nu(\mathbf{x}') \right)^{-n} \left( \prod_{\ell=1}^n z(\mathbf{x}(\ell)) \right) & \text{if } \mathbf{x} \in \mathbf{U}^{\{1, \dots, n\}}, \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

We remind that with our notation,  $(\mathbf{z}.\nu)(U) = \int_U \mathbf{z}(\mathbf{x})d\nu(\mathbf{x})$ . The point process  $S$  characterized by Equation (9) is a binomial point process of  $\mathbf{n}$  points in  $U$  with intensity  $U \rightarrow \mathbb{R}, \mathbf{x} \rightarrow ((\mathbf{z}.\nu)(U))^{-1} \mathbf{z}(\mathbf{x})$ , which we abbreviate by  $S \sim \text{bpp}(\mathbf{z}, \mathbf{n})$ . Simple random sampling with replacement is a binomial point process with a constant intensity, and we refer to as  $S \sim \text{bpp}(1, \mathbf{n})$ . 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  $\nu$ -measurable subset  $A$  of  $U$ ,

$$S \sim \text{Ppp}(\mathbf{z}) \Leftrightarrow \text{cardinality}(S^{-1}[A]) \sim \text{Poisson}((\mathbf{z}.\nu)(A)), \quad (10)$$

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  $\mathbf{x} \in \bar{U}$ , by:

$$f_S(\mathbf{x}) = (dP^S/d\bar{\nu})(\mathbf{x}) = (\text{size}(\mathbf{x})!)^{-1} \exp(-(\mathbf{z}.\nu)(U)) \prod_{\ell \in \text{domain}(\mathbf{x})} \mathbf{z}(\mathbf{x}(\ell)). \quad (11)$$

For a point process  $S$  with values in the measured space  $(U, \nu)$ , 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  $\lambda_S$  as the density of  $\Lambda_S$  with respect to  $\nu$ :  $\lambda_S = d\Lambda_S/d\nu$ .

### 2.3.2 Random design variables and random designs

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 design variable  $Z$  and study variable  $Y$ .

The design  $D$  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  $S$  conditionally to the design *is* the design, e.g:

$$P^D - a.s.(\mathbf{d}), \quad P^{S|D=\mathbf{d}} = \mathbf{d}. \quad (12)$$

We assume that the distribution of the sample conditionally on the design, the signal and the design variable is the design:

$$P^{D,Y,Z} - a.s.(\mathbf{d}, \mathbf{y}, \mathbf{z}), \quad P^{S|D=\mathbf{d}, Y=\mathbf{y}, Z=\mathbf{z}} = \mathbf{d}. \quad (13)$$

This assumption is an independence of the sample  $S$  on the design variable and the signal conditionally on the design, which does not imply independence of the sample and the signal.

For simplicity, we only consider exchangeable index sample designs, in the sense that  $\forall n \in \mathbb{N}$ , for all permutation  $\tau$  of  $\{1, \dots, n\}$ ,

$$P^Z - a.s.(\mathbf{z}), \quad P^{S|N=\mathbf{n}, Z=\mathbf{z}} = P^{S[\tau]|N=\mathbf{n}, Z=\mathbf{z}}. \quad (14)$$

which ensures that

$$P^{S|N=\mathbf{n}} = P^{S[\tau]|N=\mathbf{n}}. \quad (15)$$

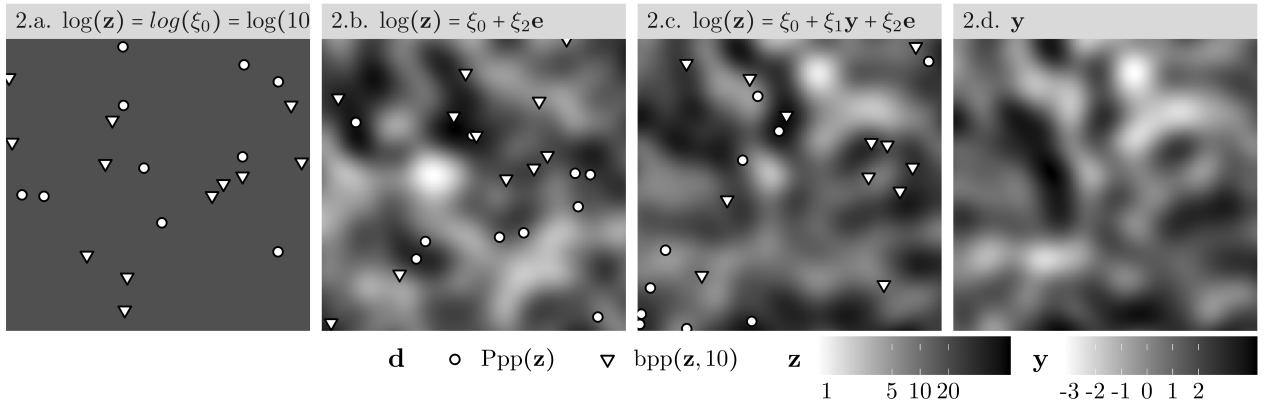
**Example 2.2** (Example 2.1 continued: distribution of  $Z$  conditionally on  $Y$ ). *We assume that  $Z$  satisfies:*

$$Z = \exp(\xi_0 + \xi_1 Y + \xi_2 \varepsilon),$$

where  $\xi_0, \xi_1, \xi_2$  are real positive numbers,  $\varepsilon : \Omega \rightarrow (\mathcal{U} \rightarrow \mathbb{R})$  is an isotropic Gaussian process with mean  $\mu = 0$  and Gaussian covariogram  $C(h) = \exp(-\|h\|^2/\xi_{scale}^2)$  with scale parameter  $\xi_{scale} = 0.1$ . The variable  $\mathbf{y} : \mathcal{U} \rightarrow \mathbb{R}$  (resp  $e : \mathcal{U} \rightarrow \mathbb{R}$ ) is generated once by sampling from  $Y$  (resp  $\varepsilon$ ). The variable  $\mathbf{z} = \exp(\xi_0 + \xi_1 \mathbf{y} + \xi_2 \mathbf{e})$  is computed for 3 different values of the vector  $(\xi_0, \xi_1, \xi_2)$ , chosen so that the expected value of the sampling intensity is the same for each simulation, more specifically so that  $E[\log(Z)] = \exp(\xi_0 + ((\theta_1 \times \xi_1)^2 + \xi_2^2)/2) = 10$ . Two samples are drawn. The first sample, conditionally on  $Z = \mathbf{z}$ , follows  $P^{S|Z=\mathbf{z}} = \text{bpp}(\mathbf{z}, 10)$ . The unconditional distribution of  $S$  is denoted  $P^S = \text{bpp}(Z, 10)$  in this case. The second sample, conditionally on  $Z = \mathbf{z}$ , follows  $P^{S|Z=\mathbf{z}} = \text{Ppp}(\mathbf{z})$ . The unconditional distribution of  $S$  is denoted  $P^S = \text{Ppp}(Z)$ . The variable  $\mathbf{z}$  generated with different parameters is mapped in Figure 2.a, 2.b, 2.c, and the variable  $\mathbf{y}$  is mapped in figure 2.d. In particular, Figures 2.a, 2.b, 2.c. show how sampled units tend to concentrate where the sampling intensity  $\mathbf{z}$  is the highest. By comparing Figure 2.b and Figure 2.d, as  $\xi_1$  is 0, the high sampling intensity in Figure 2.b and sampled points does not necessarily correspond to high values of the signal  $\mathbf{y}$  in Figure 2.d. In the opposite, the high sampling intensity in Figure 2.c and sampled points often corresponds to high values of the signal  $\mathbf{y}$  in Figure 2.d. Figure 3 represents the values of  $\mathbf{y}(\mathbf{x})$  vs  $\mathbf{z}(\mathbf{x})$  for values of  $\mathbf{x}$  on a regular grid of  $\mathcal{U}$ . The values of  $\mathbf{z}$  in Figure 3.a (resp 3.b,

3.c) correspond to the values of Figure 2.a (resp Figure 2.b, Figure 2.c). Figure 3 illustrates that high sampling intensity correspond to high values of  $\mathbf{y}$  when  $\xi_2 > 0$ , as in Figure 3.c.

Figure 2: Heat maps of the design variable  $\mathbf{z}$  and plot of realisations of  $S$  for three different design variables and designs. Sub-figures correspond to the heatmap of a design variable  $\mathbf{z}$  (Sub-figures 2.a, 2.b, 2.c) to the heatmap of  $\mathbf{y}$  (Sub-figure 2.d) and the plot of the samples (circle and triangle dots) drawn according to the two different designs. The values of  $(\xi_0, \xi_1, \xi_2)$  for each sub-figure are: 2.a:  $(\log(10), 0, 0)$ , 2.b:  $(\log(10) - 0.125, 0, 0.5)$ , 2.c:  $(\log(10) - 0.125, 0.4, 0.3)$ .



## 2.4 Probability density function of the sample

The conditional distributions of the sample and of subsamples can be derived from the conditional distribution of the design:

$$f_{S|W}(\mathbf{x} | \mathbf{w}) = \int \left( dP^{S|D,W} / d\bar{\nu} \right) (\mathbf{x} | \mathbf{d}, \mathbf{w}) dP^{D|W}(\mathbf{d} | \mathbf{w}) \quad (16)$$

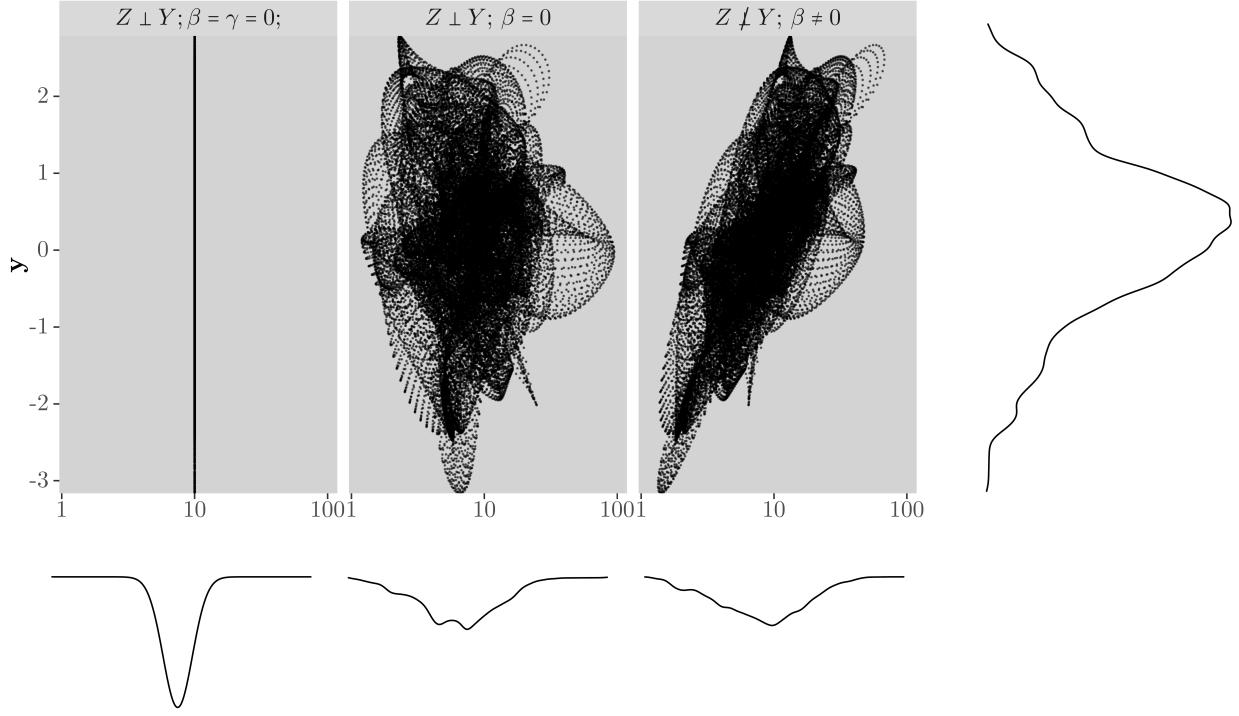
In particular when  $W$  is a function of  $Y$  and  $Z$ , assumption (13) implies that the probability to select the sample  $\mathbf{x}$  conditionally on  $W = \mathbf{w}$  is the probability to select the sample  $\mathbf{x}$  given the design, averaged on all possible designs conditionally on  $W = \mathbf{w}$ :

$$f_{S|W}(\mathbf{x} | \mathbf{w}) = \int (d\mathbf{d} / d\bar{\nu})(\mathbf{x}) dP^{D|W}(\mathbf{d} | \mathbf{w}). \quad (17)$$

**Example 2.3** ( Example 2.2 continued: Sample density for  $S \sim \text{Ppp}(Z)$  and  $S \sim \text{bpp}(Z, n)$ ). For  $\mathbf{x} \in U^{\{1, \dots, n\}}$ , when  $S \sim \text{Ppp}(Z)$ ,

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

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  $(\xi_0, \xi_1, \xi_2)$  for each Sub-figure are: 3.a:  $(\log(10), 0, 0)$ , 3.b:  $(\log(10) - 0.125, 0, 0.5)$ , 3.c:  $(\log(10) - 0.125, 0.4, 0.3)$ . 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).



When  $S \sim \text{bpp}(Z, \mathbf{n})$ ,

$$f_{S|W}(\mathbf{x} | \mathbf{w}) = \int ((\mathbf{z} \cdot \nu)(\mathbf{U}))^{-\mathbf{n}} \left( \prod_{\ell=1}^n \mathbf{z}(\mathbf{x}(\ell)) \right) dP^{Z|W}(\mathbf{z} | \mathbf{w}). \quad (19)$$

## 2.5 Probability density function of a subsample

Theoretical developments in Section 3 require to express the density function of a random size or fixed size subsample of a random or fixed size sample.

For a fixed set  $K$ , the density  $f_{S_K|W}(\mathbf{x} | \mathbf{w})$  can be derived from  $f_{S|W}$  via:

$$f_{S_K|W}(\mathbf{x} | \mathbf{w}) = \int f_{S|W}(\mathbf{x}' | \mathbf{w}) \mathbb{1}(\mathbf{x}'_K == \mathbf{x}) d\nu(\mathbf{x}'), \quad (20)$$

where  $\mathbb{1}(\mathbf{x}'_K == \mathbf{x}) = 0$  when  $\mathbf{x}'_K != \mathbf{x}$ , 1 otherwise. More generally, for a random set  $K$  the

density  $f_{S|W}(\mathbf{x} \mid \mathbf{w})$  can be derived from  $f_{S_K|W}$  via:

$$f_{S_K|W}(\mathbf{x} \mid \mathbf{w}) = \int \left( \int f_{S|W}(\mathbf{x}' \mid \mathbf{w}) \mathbb{1}(\mathbf{x}'_K == \mathbf{x}) d\bar{\nu}(\mathbf{x}') \right) dP^{K|W=\mathbf{w}}. \quad (21)$$

*Remark 1.* For two fixed size sets  $K_1, K_2$  fixed such that  $K_1 \subset K_2$ , the equation

$$f_{S_{K_1}}(\mathbf{x} \mid \mathbf{w}) = \int f_{S_{K_2}}(\mathbf{x}' \mid \mathbf{w}) \mathbb{1}(\mathbf{x}'_{K_1} = \mathbf{x}) d\nu^{\otimes K_2}(\mathbf{x}') \quad (22)$$

is not necessarily true. For example, when  $S \sim \text{Ppp}(Z)$ ,  $K_1 = \{1\}$ ,  $K_2 = \{1, 2\}$ , for  $\mathbf{x} \in U^{\{1\}}$ ,  $f_{S_{\{1\}}}(\mathbf{x}) = f_S(\mathbf{x}) + \int f_{S_{K_2}}(\mathbf{x}' \mid \mathbf{w}) \mathbb{1}(\mathbf{x}'_{K_1} = \mathbf{x}) d\nu^{\otimes K_2}(\mathbf{x}')$ . A sufficient condition for the equation (22) would be that  $N$  is not random.

Let  $K$  be a finite subset of  $\mathbb{N} \setminus \{0\}$ . Let  $K'$  be an element of  $\{\{\ell \in K : \ell \leq \mathbf{n}\} : \mathbf{n} \in \mathbb{N}\}$ , and let  $\mathbf{x} \in U^{K'}$ . Let  $W$  be a function of  $Z, Y$ . In the following we derive the density of  $S_K$  conditionally on  $W = \mathbf{w}$ . For  $\mathbf{x} = \mathbf{0}$ ,

$$f_{S_K|W}(\mathbf{0} \mid \mathbf{w}) = \begin{cases} 1 & \text{if } K = \emptyset, \\ P(N < \min(K) \mid W = \mathbf{w}) & \text{otherwise.} \end{cases} \quad (23)$$

For  $\mathbf{x} \neq \mathbf{0}$ ,

$$f_{S_K|W}(\mathbf{x} \mid \mathbf{w}) = \int f_{S_K|Z}(\mathbf{x} \mid \mathbf{z}) dP^{Z|W}(\mathbf{z} \mid \mathbf{w}). \quad (24)$$

**Example 2.4** (Example 2.3 continued: Probability density function of a subsample for  $S \sim \text{Ppp}(Z)$  and  $S \sim \text{bpp}(Z, \mathbf{n})$ ). *Let  $K$  be a non random finite subset of  $\mathbb{N} \setminus \{0\}$*

*When  $S \sim \text{Ppp}(Z)$ ,*

$$f_{S_K|Z}(\mathbf{x} \mid \mathbf{z}) = \left( \sum_{\substack{\mathbf{n} \in \mathbb{N}: \\ \{\ell \in K : \ell \leq \mathbf{n}\} = K'}} P(N = \mathbf{n} \mid Z = \mathbf{z}) \right) \prod_{\ell \in K'} \frac{\mathbf{z}(\mathbf{x}(\ell))}{(\mathbf{z} \cdot \nu)(U)}. \quad (25)$$

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

$$f_{S_K|Z}(\mathbf{x} \mid \mathbf{z}) = \begin{cases} \prod_{\ell \in K'} ((\mathbf{z} \cdot \nu)(U))^{-1} (\mathbf{z}(\mathbf{x}(\ell))) & \text{if } K' = \{1, \dots, \mathbf{n}\} \cap K \\ 0 & \text{otherwise.} \end{cases} \quad (26)$$

Note that the distribution of the sample unconditionally on the design or the design variable can be obtained via:  $f_S(\mathbf{x}) = \int f_{S|Z}(\mathbf{x}) dP^Z$ . Contrary to the particular case of population index exchangeability and sample index exchangeability, the sample index exchangeability (15) condition alone does not ensure that conditionally on the sample size,  $N = \mathbf{n}$ , the distribution of the sample  $P^{S|N=\mathbf{n}}$  is the uniform distribution on  $U^{\{1, \dots, \mathbf{n}\}}$ . For example,  $f_{S_{\{1\}}}$  can be uniform, but  $f_{S_{\{1, 2\}}}$  not.

### 3 “Sample” vs “Population” distribution of the signal

The observation consists of the realisations of the random variables  $S$  and  $Y[S]$ . 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  $Z$ . We resort to the Bayes formula to express the density of  $(Y[S], S)$  in  $(\mathbf{y}, \mathbf{x})$ , as the product of the density of the sample  $S$  in  $\mathbf{x}$  conditionally on  $Y[\mathbf{x}] = \mathbf{y}$  by the density of the signal  $Y[\mathbf{x}]$  in  $\mathbf{y}$ .

#### 3.1 Density ratio

##### 3.1.1 Definition and properties

We start with the following definition.

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

$$P^{(S_K, Y[S_K])} - \text{a.s}(\mathbf{x}, \mathbf{y}), \quad f_{Y[\mathbf{x}]|S_K=\mathbf{x}}(\mathbf{y}) = f_{Y[\mathbf{x}]}(\mathbf{y}) \rho_K(\mathbf{x} | \mathbf{y}) \quad (27)$$

**Property 3.1.**

$$P^{(S_K, Y[S_K])} - a.s(\mathbf{x}, \mathbf{y}), \quad \rho_K(\mathbf{x} | \mathbf{y}) f_{S_K}(\mathbf{x}) = f_{S_K|Y[\mathbf{x}]}(\mathbf{x} | \mathbf{y}) \quad (28)$$

*Proof.* From the Bayes formula:

$$f_{Y[\mathbf{x}]|S_K=\mathbf{x}}(\mathbf{y}) = (f_{S_K}(\mathbf{x}))^{-1} f_{Y[\mathbf{x}], S_K}(\mathbf{y}, \mathbf{x}) \quad (29)$$

$$= (f_{S_K}(\mathbf{x}))^{-1} f_{S_K|Y[\mathbf{x}]}(\mathbf{x} | \mathbf{y}) f_{Y[\mathbf{x}]}(\mathbf{y}) \quad (30)$$

So, combining Equations (30) and (27):

$$\rho_K(\mathbf{x} | \mathbf{y}) = (f_{S_K}(\mathbf{x}))^{-1} f_{S_K|Y[\mathbf{x}]}(\mathbf{x} | \mathbf{y}). \quad (31)$$

□

Property 3.1 provides an interpretation of  $\rho_K(\mathbf{y} | \mathbf{x})$  as the probability that the draws indexed by  $K$  correspond to the sample  $\mathbf{x}$  conditionally on the values of  $Y$  in  $\mathbf{x}$  divided by the same probability unconditionally on the values of  $Y$ . In this sense, the density ratio  $\rho_K(\mathbf{y} | \mathbf{x})$  is a relative probability of selection.

**Property 3.2.** *The density ratio  $\rho$  can be derived from the distribution of  $Z$  via:*

$$\rho_K(\mathbf{x} \mid \mathbf{y}) = \left( \int f_{S_K|Z}(\mathbf{x} \mid \mathbf{z}) dP^Z \right)^{-1} \int f_{S_K|Z}(\mathbf{x} \mid \mathbf{z}) dP^{Z|Y[\mathbf{x}]=\mathbf{y}}(\mathbf{z}) \quad (32)$$

**Property 3.3.** *If  $Z$  and  $Y$  are independent, then  $\rho_K(\mathbf{x} \mid \mathbf{y}) = 1$ .*

*Proof.*  $Z \perp Y \Rightarrow P^Z = P^{Z|Y[\mathbf{x}]=\mathbf{y}}$ . Replacing  $P^{Z|Y[\mathbf{x}]=\mathbf{y}}$  by  $P^Z$  in the numerator of Equation (32) yields  $\rho_K(\mathbf{x} \mid \mathbf{y}) = 1$ .  $\square$

### 3.1.2 Computation of $\rho$ on a Case study

In this subsection, we use the framework of Example 2.4 to illustrate the properties and provide graphical representations of the density ratio. We remind that in this framework,  $S \sim \text{Ppp}(Z)$  or  $S \sim \text{bpp}(Z, \mathbf{n})$ . Conditionally on  $Y[\mathbf{x}] = \mathbf{y}$ ,  $Z$  is a log normal random process with distribution characterized by  $E[\log(Z[\mathbf{x}']) \mid Y[\mathbf{x}] = \mathbf{y}] = \xi_0 + \xi_1(\mu + \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1}(\mathbf{y} - \mu))$  and  $\text{Var}[\log(Z[\mathbf{x}']) \mid Y[\mathbf{x}] = \mathbf{y}] = \xi_2^2 \Sigma_{\varepsilon; \mathbf{x}', \mathbf{x}'} + \xi_1^2 (\Sigma_{Y; \mathbf{x}', \mathbf{x}'} - \Sigma_{Y; \mathbf{x}, \mathbf{x}} \Sigma_{Y; \mathbf{x}, \mathbf{x}}^{-1} \Sigma_{Y; \mathbf{x}', \mathbf{x}'})$ .

*Proof.* See Appendix A.1  $\square$

For  $\mathbf{x} = \mathbf{0}$  and  $\mathbf{y} = \mathbf{0}$ ,

$$\begin{aligned} & f_{S_{\{1\}}|Y[\mathbf{0}]=\mathbf{0}}(\mathbf{0}) \\ &= f_{S_{\{1\}}}(\mathbf{0}) \\ &= \begin{cases} \int \exp(-(\mathbf{z} \cdot \nu)(U)) dP^Z(\mathbf{z}) & \text{if } S \sim \text{Ppp}(Z) \\ 0 & \text{if } S \sim \text{bpp}(Z, \mathbf{n}), \end{cases} \end{aligned}$$

so  $\rho_{\{1\}}(\mathbf{0} \mid \mathbf{0}) = f_{S_{\{1\}}|Y[\mathbf{0}]=\mathbf{0}}(\mathbf{0})/f_{S_{\{1\}}}(\mathbf{0}) = 1$ . More generally, for any random set  $K$ ,  $\rho_K(\mathbf{0} \mid \mathbf{0}) = 1$ . When  $S \sim \text{Ppp}(Z)$ , for  $\mathbf{x} \in U^{\{1\}}$ , as a consequence of Equations (24) and (25),

$$\begin{aligned} & \rho_{\{1\}}(\mathbf{x} \mid \mathbf{y}) \\ &= \frac{\int (\sum_{\mathbf{n} \geq 1} (\mathbf{n}!)^{-1} (\mathbf{z} \cdot \nu)(U)^{\mathbf{n}}) \times \exp(-(\mathbf{z} \cdot \nu)(U)) ((\mathbf{z} \cdot \nu)(U))^{-1} \mathbf{z}(\mathbf{x}(1)) dP^{Z|Y[\mathbf{x}]}(\mathbf{z} \mid \mathbf{y})}{\int (\sum_{\mathbf{n} \geq 1} (\mathbf{n}!)^{-1} (\mathbf{z} \cdot \nu)(U)^{\mathbf{n}}) \times \exp(-(\mathbf{z} \cdot \nu)(U)) ((\mathbf{z} \cdot \nu)(U))^{-1} \mathbf{z}(\mathbf{x}(1)) dP^Z(\mathbf{z})} \\ &= \frac{\int (1 - \exp(-(\mathbf{z} \cdot \nu)(U))) ((\mathbf{z} \cdot \nu)(U))^{-1} \mathbf{z}(\mathbf{x}(1)) dP^{Z|Y[\mathbf{x}]}(\mathbf{z} \mid \mathbf{y})}{\int (1 - \exp(-(\mathbf{z} \cdot \nu)(U))) ((\mathbf{z} \cdot \nu)(U))^{-1} \mathbf{z}(\mathbf{x}(1)) dP^Z(\mathbf{z})} \\ &= \frac{\int (1 - \exp^{-\int \exp(\xi_0 + \xi_1 Y[\mathbf{x}'] + \xi_2 \varepsilon[\mathbf{x}']) d\nu(\mathbf{x}')} ) \left( \int \exp^{-\xi_1(Y[\mathbf{x}'] - \mathbf{y}) - \xi_2(\varepsilon[\mathbf{x}'] - \varepsilon[\mathbf{x}])} d\nu(\mathbf{x}') \right)^{-1} dP^{Y, \varepsilon|Y[\mathbf{x}]=\mathbf{y}}}{\int (1 - \exp^{-\int \exp(\xi_0 + \xi_1 Y[\mathbf{x}'] + \xi_2 \varepsilon[\mathbf{x}']) d\nu(\mathbf{x}')} ) \left( \int \exp^{-\xi_1(Y[\mathbf{x}'] - Y[\mathbf{x}(1)]) - \xi_2(\varepsilon[\mathbf{x}'] - \varepsilon[\mathbf{x}(1)])} d\nu(\mathbf{x}') \right)^{-1} dP^{Y, \varepsilon}}$$

and for  $\mathbf{x} \in U^{\{1,2\}}$ ,

$$\begin{aligned} & \rho_{\{1,2\}}(\mathbf{x} \mid \mathbf{y}) \\ &= \frac{\int \frac{(1-(1+(\int \exp(\xi_0+\xi_1 Y[\mathbf{x}']+\xi_2 \varepsilon[\mathbf{x}']) d\nu(\mathbf{x}'))))}{\exp(\int \exp(\xi_0+\xi_1 Y[\mathbf{x}']+\xi_2 \varepsilon[\mathbf{x}']) d\nu(\mathbf{x}'))} \frac{\exp^{-\xi_1(\sum_{\ell=1}^2 y^{(\ell)})-\xi_2(\sum_{\ell=1}^2 \varepsilon[\mathbf{x}^{(\ell)}])}}{\left(\int \exp(-\xi_1(Y[\mathbf{x}'])-\xi_2(\varepsilon[\mathbf{x}'])) d\nu(\mathbf{x}')\right)^2} dP_{Y,\varepsilon|Y[\mathbf{x}]=\mathbf{y}}} \\ &= \frac{\int \frac{(1-(1+(\int \exp(\xi_0+\xi_1 Y[\mathbf{x}']+\xi_2 \varepsilon[\mathbf{x}']) d\nu(\mathbf{x}'))))}{\exp(\int \exp(\xi_0+\xi_1 Y[\mathbf{x}']+\xi_2 \varepsilon[\mathbf{x}']) d\nu(\mathbf{x}'))} \frac{\exp^{-\xi_1(\sum_{\ell=1}^2 Y(\mathbf{x}^{(\ell)}))-\xi_2(\sum_{\ell=1}^2 \varepsilon[\mathbf{x}^{(\ell)}])}}{\left(\int \exp(-\xi_1(Y[\mathbf{x}'])-\xi_2(\varepsilon[\mathbf{x}'])) d\nu(\mathbf{x}')\right)^2} dP_{Y,\varepsilon} \end{aligned}$$

and for  $\mathbf{x} \in U^{\{1,\dots,n\}}$ , as a consequence of Equations (24) and (26)

$$\begin{aligned} & \rho_{\{1,\dots,n\}}(\mathbf{x} \mid \mathbf{y}) \\ &= \frac{\int \frac{(\int \exp(\xi_0+\xi_1 Y[\mathbf{x}']+\xi_2 \varepsilon[\mathbf{x}']) d\nu(\mathbf{x}'))^n}{n! \exp(\int \xi_0+\xi_1 Y[\mathbf{x}']+\xi_2 \varepsilon[\mathbf{x}']) d\nu(\mathbf{x}'))} \frac{\exp^{-\xi_1(\sum_{\ell=1}^n y^{(\ell)})-\xi_2(\sum_{\ell=1}^n \varepsilon[\mathbf{x}^{(\ell)}])}}{\left(\int \exp(-\xi_1(Y[\mathbf{x}'])-\xi_2(\varepsilon[\mathbf{x}'])) d\nu(\mathbf{x}')\right)^n} dP_{Y,\varepsilon|Y[\mathbf{x}]=\mathbf{y}}} \\ &= \frac{\int \frac{(\int \exp(\xi_0+\xi_1 Y[\mathbf{x}']+\xi_2 \varepsilon[\mathbf{x}']) d\nu(\mathbf{x}'))^n}{n! \exp(\int \exp(\xi_0+\xi_1 Y[\mathbf{x}']+\xi_2 \varepsilon[\mathbf{x}']) d\nu(\mathbf{x}'))} \frac{\exp^{-\xi_1(\sum_{\ell=1}^2 Y(\mathbf{x}^{(\ell)}))-\xi_2(\sum_{\ell=1}^2 \varepsilon[\mathbf{x}^{(\ell)}])}}{\left(\int \exp(-\xi_1(Y[\mathbf{x}'])-\xi_2(\varepsilon[\mathbf{x}'])) d\nu(\mathbf{x}')\right)^n} dP_{Y,\varepsilon} \end{aligned}$$

When  $S \sim \text{bpp}(Z, \mathbf{n})$ , for  $\mathbf{x} \in U^{\{1\}}$ :

$$\rho_{\{1\}}(\mathbf{x} \mid \mathbf{y}) = \frac{\int \left( \int \exp^{-\xi_1(Y[\mathbf{x}'])-\xi_2(\varepsilon[\mathbf{x}']-\varepsilon[\mathbf{x}])} d\nu(\mathbf{x}') \right)^{-1} dP_{Y,\varepsilon|Y[\mathbf{x}]=\mathbf{y}}}{\int \left( \int \exp^{-\xi_1(Y[\mathbf{x}'])-Y[\mathbf{x}(1)]}-\xi_2(\varepsilon[\mathbf{x}']-\varepsilon[\mathbf{x}(1)]) d\nu(\mathbf{x}') \right)^{-1} dP_{Y,\varepsilon}}.$$

for  $\mathbf{x} \in U^{\{1,2\}}$ :

$$\rho_{\{1,2\}}(\mathbf{x} \mid \mathbf{y}) = \frac{\int \frac{\exp(-\xi_1(\sum_{\ell=1}^2 y^{(\ell)})-\xi_2(\sum_{\ell=1}^2 \varepsilon[\mathbf{x}^{(\ell)}]))}{\left(\int \exp(-\xi_1 Y[\mathbf{x}']-\xi_2 \varepsilon[\mathbf{x}']) d\nu(\mathbf{x}')\right)^2} dP_{Y,\varepsilon|Y[\mathbf{x}]=\mathbf{y}}}{\int \frac{\exp(-\xi_1(\sum_{\ell=1}^2 Y(\mathbf{x}^{(\ell)}))-\xi_2(\sum_{\ell=1}^2 \varepsilon[\mathbf{x}^{(\ell)}]))}{\left(\int \exp(-\xi_1 Y[\mathbf{x}']-\xi_2 \varepsilon[\mathbf{x}']) d\nu(\mathbf{x}')\right)^2} dP_{Y,\varepsilon}}.$$

and for  $\mathbf{x} \in U^{\{1,\dots,n'\}}$ ,  $0 \leq \mathbf{n}' \leq \mathbf{n}$ :

$$\rho_{\{1,\dots,n'\}}(\mathbf{x} \mid \mathbf{y}) = \frac{\int \frac{\exp(-\xi_1(\sum_{\ell=1}^2 y^{(\ell)})-\xi_2(\sum_{\ell=1}^{n'} \varepsilon[\mathbf{x}^{(\ell)}]))}{\left(\int \exp(-\xi_1 Y[\mathbf{x}']-\xi_2 \varepsilon[\mathbf{x}']) d\nu(\mathbf{x}')\right)^{n'}} dP_{Y,\varepsilon|Y[\mathbf{x}]=\mathbf{y}}}{\int \frac{\exp(-\xi_1(\sum_{\ell=1}^{n'} Y(\mathbf{x}^{(\ell)}))-\xi_2(\sum_{\ell=1}^{n'} \varepsilon[\mathbf{x}^{(\ell)}]))}{\left(\int \exp(-\xi_1 Y[\mathbf{x}']-\xi_2 \varepsilon[\mathbf{x}']) d\nu(\mathbf{x}')\right)^{n'}} dP_{Y,\varepsilon}}.$$

The density ratio is a function of  $\mathbf{x}$ ,  $\mathbf{y}$ ,  $\xi$ , and  $\theta$ .

**Properties 3.4.** *Properties of  $\rho$  common to Poisson and binomial point processes*

- i.  $\xi_1 = 0 \Rightarrow \rho(\mathbf{x}, \mathbf{y}; \xi, \theta) = 1$ .
- ii.  $\rho_K(\mathbf{x} \mid \mathbf{y}; \xi, \theta) - 1 = o_{\xi_2 \rightarrow \infty}(1)$ ,
- iii.  $\rho_K(\mathbf{x} \mid \mathbf{y}; \xi, \theta) - 1 = o_{\theta_{scale} \rightarrow \infty}(1)$ ,

- iv.  $\rho_K(\mathbf{x} \mid \mathbf{y}; \xi, \theta) - 1 = o_{\theta_{scale} \rightarrow \infty}(1)$ ,
- v.  $\rho_K(\mathbf{x}, -\mathbf{y}; (\theta, (\xi_0, -\xi_1, \xi_2, \xi_{scale})) = \rho_K(\mathbf{x} \mid \mathbf{y}; (\theta, (\xi_0, \xi_1, \xi_2, \xi_{scale})$ )
- vi.  $\text{sign}(\xi_1) = 1 \Rightarrow \lim_{\mathbf{y} \rightarrow \infty} \rho_{\{1\}}(\mathbf{x} \mid \mathbf{y}; \xi, \theta) = +\infty$ ,
- vii.  $\text{sign}(\xi_1) = 1 \Rightarrow \lim_{\mathbf{y} \rightarrow -\infty} \rho_{\{1\}}(\mathbf{x} \mid \mathbf{y}; \xi, \theta) = 0$ ,

**Properties 3.5.** *Properties of  $\rho$  when  $S \sim \text{bpp}(Z, \mathbf{n})$*

- i.  $\rho_{\{1\}}(\mathbf{x} \mid \mathbf{y}; \xi, \theta) = \exp(\xi_1(\mathbf{y} - \xi/2) + o_{\theta_{scale} \rightarrow 0}(1))$ .
- ii.  $\mathbf{n}' \leq \mathbf{n} \Rightarrow \rho_{\{1, \dots, \mathbf{n}'\}}(\mathbf{x} \mid \mathbf{y}; \xi, \theta) = \exp\left(\xi_1\left(\sum_{\ell=1}^{\mathbf{n}'} \mathbf{y}(\ell) - \mathbf{n}'\xi_1/2\right)\right) + o_{\theta_{scale} \rightarrow 0}(1)$ .

The density ratio  $\rho(\mathbf{x} \mid \mathbf{y})$  is the product of two integrals as shown in Equation (32). No close form could be obtained for any of these two integrals. However, we can use Monte Carlo approximations to compute  $\rho(\mathbf{x} \mid \mathbf{y})$ . In the case where  $U$  is finite, we can simulate realisations of  $P^Y$  and  $P^\varepsilon$  on the population independently  $J$  times. Let denote by  $Y^{(j)}$  and  $\varepsilon^{(j)}$  the  $j$ -th realisation of each distribution  $Y$  and  $\varepsilon$ , respectively.

For  $j \in 1, \dots, J$ , let define  $Y_c^{(j)}$  by  $Y_c^{(j)}[\mathbf{x}'] = Y^{(j)}[\mathbf{x}'] + \Sigma_{Y; \mathbf{x}', \mathbf{x}} \Sigma_{Y; \mathbf{x}, \mathbf{x}}^{-1}(\mathbf{y} - Y^{(j)}[\mathbf{x}])$ , which ensures that  $P^{Y_c^{(j)}} = P^{Y|Y[\mathbf{x}] = \mathbf{y}}$  by the Matheron rule (Wilson et al., 2020; Doucet, 2010).

Define, for a non random set  $K$ ,  $\emptyset \neq K \subset \{1, \dots, \mathbf{n}\}$ , and  $\mathbf{x} \in U^K$ ,  $\mathbf{y} \in \mathcal{Y}^K$ ,  $m = \text{size}(K)$ :

$$\tilde{\rho}_K(\mathbf{x} \mid \mathbf{y}) = \frac{\sum_{j=1}^J \left( \left( \left( \exp\left(\xi_1 Y_c^{(j)} + \xi_2 \varepsilon^{(j)}\right) \cdot \nu \right) (U) \right)^{-m} \prod_{\ell \in K} \exp(\xi_1 \mathbf{y}(\ell) + \xi_2 \varepsilon[\mathbf{x}(\ell)]) \right)}{\sum_{j=1}^J \left( \left( \left( \exp\left(\xi_1 Y^{(j)} + \xi_2 \varepsilon^{(j)}\right) \cdot \nu \right) (U) \right)^{-m} \prod_{\ell \in K} \exp(\xi_1 Y^{(j)}[\mathbf{x}(\ell)] + \xi_2 \varepsilon[\mathbf{x}(\ell)]) \right)}.$$

In the case of a binomial point process ( $S \sim \text{bpp}(Z, \mathbf{n})$ ), the ratio of the expected values of the numerators and denominators of  $\tilde{\rho}(\mathbf{x} \mid \mathbf{y})$  is exactly  $\rho_K(\mathbf{x} \mid \mathbf{y})$ . The empirical variance of the summands of the numerator and denominator can be used to estimate the variance of  $\tilde{\rho}(\mathbf{x} \mid \mathbf{y})$ .

Following the aforementioned discussion, we employ a Monte Carlo simulation in order to approximate  $\rho$  by means of  $\tilde{\rho}_K(\mathbf{x} \mid \mathbf{y})$ . As first step, we generate  $J = 10,000$  replications for the computation of  $\tilde{\rho}_{\{1\}}(\mathbf{x} \mid \mathbf{y})$ . In Figure 4 we investigate the role of  $\xi_1$ . Toward this end,  $\xi_2$  is kept fixed to 1, and  $\xi_1$  varies while  $\mathbf{y}$  is kept fixed as well. This procedure is done for different levels of  $\mathbf{y}$ . As expected, the value of  $\tilde{\rho}_{\{1\}}(\mathbf{x} \mid \mathbf{y})$  at  $\xi_1 = 0$  is always equal to 1. Indeed, we recall that  $\xi = 0$  implies independence between  $Y$  and  $Z$ . When  $\xi_1$  differs from zero,  $\tilde{\rho}_{\{1\}}(\mathbf{x} \mid \mathbf{y})$  is different from 1, hence reflecting the fact that dependence

between  $Y$  and  $Z$  introduces bias. In particular, we can observe that as  $\xi_1$  approaches infinity,  $\tilde{\rho}_{\{1\}}(\mathbf{x} \mid \mathbf{y})$  approaches zero. Also, for positive values of  $\mathbf{y}$ , the maximum is achieved around  $\xi_1 = \mathbf{y}$ . The same analysis is performed for the behavior of  $\xi_2$ , where  $\xi_1$  is kept fixed to 1, and different values of  $\mathbf{y}$  are fixed while  $\xi_2$  varies. We report the results in Figure 5. When  $\xi_2 = 0$ ,  $\tilde{\rho}_{\{1\}}(\mathbf{x} \mid \mathbf{y})$  is different from zero. This reflects the fact that  $\xi_1$  is fixed to 1. As  $\xi_2$  increases, the effect of  $\xi_1$  decreases and  $\tilde{\rho}_{\{1\}}(\mathbf{x} \mid \mathbf{y})$  tends to 1. We then turn the attention to  $\tilde{\rho}_{\{1,2\}}(\mathbf{x} \mid \mathbf{y})$ .  $J = 4,000$  replications are generated and values of  $\tilde{\rho}_{\{1,2\}}(\mathbf{x} \mid \mathbf{y})$  are computed for four different couples of units, which have an Euclidean distance (between the two points) of 0.018, 0.074, 0.357, and 0.711, respectively. In Figure 6 the contours of the surface generated by the values of  $\tilde{\rho}_{\{1,2\}}(\mathbf{x} \mid \mathbf{y})$  when the two values of  $\mathbf{y}$  vary are plotted.

Figure 4: Representation of  $\tilde{\rho}_{\{1\}}(\mathbf{x} \mid \mathbf{y})$  when  $S \sim \text{bpp}(Z, \mathbf{n})$ ,  $\xi_2 = 1$ , and  $\xi_1$  varies. The value of  $\mathbf{y}$  is reported on top of the corresponding graph.

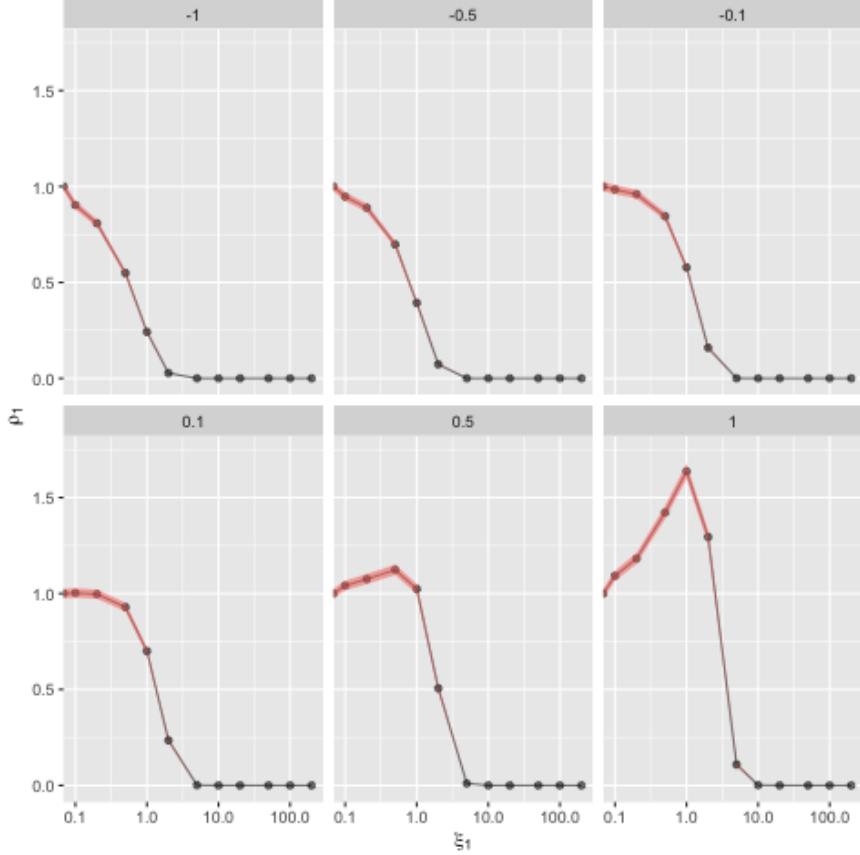


Figure 5: Representation of  $\tilde{\rho}_{\{1\}}(\mathbf{x} \mid \mathbf{y})$  when  $S \sim \text{bpp}(Z, \mathbf{n})$ ,  $\xi_1 = 1$ , and  $\xi_2$  varies. The value of  $\mathbf{y}$  is reported on top of the corresponding graph.

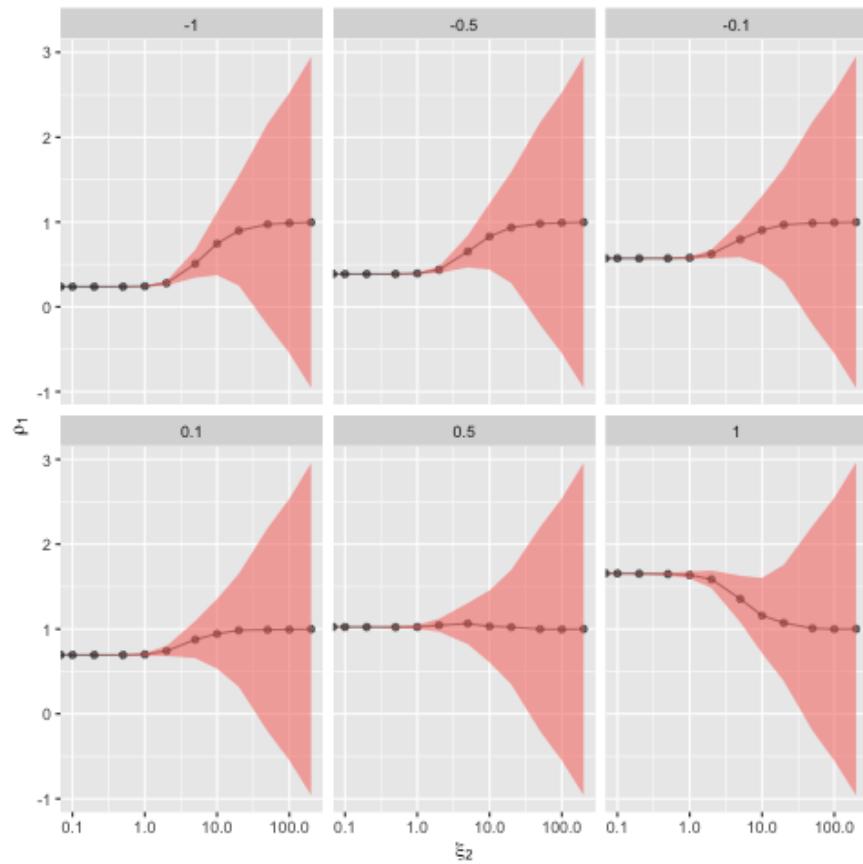
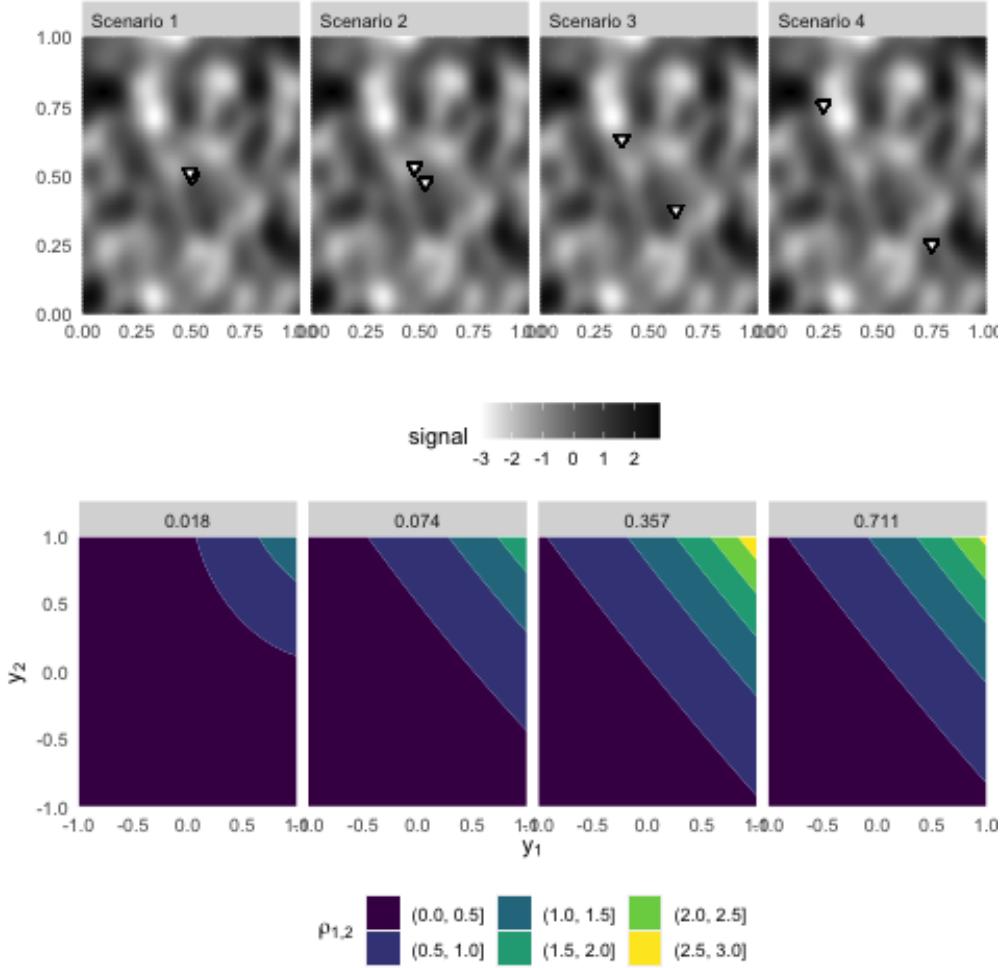


Figure 6: Representation of  $\tilde{\rho}_{\{1,2\}}(\mathbf{x} \mid \mathbf{y})$  when  $S \sim \text{bpp}(Z, \mathbf{n})$ ,  $\xi_1 = 1$ ,  $\xi_2 = 1$ , and for four different couples of points.



### 3.2 Density ratio $\rho$ and distribution of $Y[S]$

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

$$f_{S_K, Y[S_K]}(\mathbf{x} \mid \mathbf{y}) = f_{Y[\mathbf{x}]}(\mathbf{y}) \times \rho_K(\mathbf{x} \mid \mathbf{y}) \times f_{S_K}(\mathbf{x}). \quad (33)$$

$$= f_{Y[\mathbf{x}]}(\mathbf{y}) \times f_{S_K|Y[\mathbf{x}]}(\mathbf{x} \mid \mathbf{y}). \quad (34)$$

In particular,

$$f_{S, Y[S]}(\mathbf{x} \mid \mathbf{y}) = f_{Y[\mathbf{x}]}(\mathbf{y}) \times \rho_{\{1, \dots, N\}}(\mathbf{x} \mid \mathbf{y}) \times f_S(\mathbf{x}). \quad (35)$$

So when  $Z \perp Y$ ,

$$f_{Y[S_K]|S_K}(\mathbf{y} \mid \mathbf{x}) = f_{Y[\mathbf{x}]}(\mathbf{y}). \quad (36)$$

Krieger and Pfeffermann (1992) explore the case where  $K = \{1\}$  and the spatial process satisfies the population independence assumption:

$\forall \mathbf{x} \in U^{\{1, \dots, n\}}$ , such that  $(\mathbf{x}[1], \dots, \mathbf{x}[n])$  are two by two distinct,

$$(Y[\mathbf{x}[1]], \dots, Y[\mathbf{x}[n]]) \text{ are i.i.d variables.} \quad (37)$$

To the “population” distribution, which is the distribution of  $Y$  that satisfies Equation (37) with probability density function  $f_{Y[\mathbf{x}]}$  for  $\mathbf{x} \in U^{\{1\}}$ , Krieger and Pfeffermann (1992) oppose the sample distribution, that is the distribution of a random variable  $Y^*$  that satisfies Equation (37) and with probability density function  $\rho_{\{1\}} f_{Y[\mathbf{x}]}$  for  $\mathbf{x} \in U^{\{1\}}$ .

The random variable  $Y^*$  “does not exist”, in the sense that if the observer could measure  $Y[\mathbf{x}]$  for all values of  $\mathbf{x}$  on  $U$ , or for values of  $\mathbf{x}$  in a non informative sample (e.g. such that  $\rho = 1$ ), the distribution of observations would have probability density  $f_Y[S]$ . With an informative sample, the observations are similar to what one would observe with a simple random sample and if the population was following  $P^{Y^*}$  and not  $P^Y$ . The distribution of  $Y^*$  is called by Krieger and Pfeffermann (1992) the “sample” distribution.

In the case of a spatial process, can we define a sample distribution ?

Sometimes not, due to the property pointed out in Remark 1: the finite dimensional probability density functions  $\rho_{\{1, \dots, n\}}(\mathbf{x} | \mathbf{y}) f_{Y[\mathbf{x}](\mathbf{y})}$  are not necessarily the finite dimensional densities of the same random process distribution. A sufficient condition for the finite dimensional densities to be finite dimensional probability functions of the same process is that the sample size is fixed. But then, there is not necessarily unicity of the random process distribution that have these finite dimensional probability density functions, as they are only defined up to the dimension equal to the sample size.

Generalising Krieger and Pfeffermann (1992), the selection is non informative when  $\forall m \in \mathbb{N} P^{Y[S], S} - \text{a.s.}(\mathbf{y}, \mathbf{x})$ ,

$$\rho_{\{1, \dots, N\}}(\mathbf{x} | \mathbf{y}) = 1.$$

In this case, we consider that the sample distribution corresponds to the population distribution.

In the general case, we then define the sample distribution as a collection of finite dimensional distributions with respective p.d.f.’s  $\rho_{\{1, \dots, n'\}}(\mathbf{x} | \mathbf{y}) f_{Y[\mathbf{x}](\mathbf{y})}$ , with  $n' \in \{\{1, \dots, \max(N(N^{-1}(\mathbb{N})))\}\}$ .

### 3.3 Sample counterpart of distribution characteristics

Even if the definition of the sample distribution is not the distribution of a random process on  $U$ , we propose to define the sample counterpart of different characteristics of the population distribution as for example the sample semivariogram.

In a non spatial framework, Krieger and Pfeffermann (1992) differentiates the characteristics of the distribution of one variable when observed on the sample, and the characteristics of the same variable when observed on the population. The terminology to differentiate consists in adding “populatio” or “sample” to the name of the characteristic. For example, the population probability density function (pdf) is the equivalent  $f_{Y[\mathbf{x}]}$  in our example, and the sample pdf is the density  $f_{Y[S[1]]}$ . Bonnery et al. (2012) and Bonnery and Sedransk (2019) investigate further this concept of population and sample characteristics. In the following, we provide the definition of the sample characteristics of the distribution of  $Y$  that are meaningful in the context of a spatial model.

Some characteristics of the distribution of  $Y$  can be defined as expected values of a transformation of the values of  $Y$  on all samples of  $U$ . For example, we have seen that the distribution of a Gaussian process is characterised by  $E[Y[\mathbf{x}]]$  for  $\mathbf{x} \in U$  and  $\text{Var}[Y[\mathbf{x}]]$  for  $\mathbf{x} \in U^{\{1,2\}}$ . Those characteristics can be obtained by integrals of the form  $\int g(\mathbf{y}[\mathbf{x}])f_{Y[\mathbf{x}]}(\mathbf{y}[\mathbf{x}])d\bar{\eta}$ , which themselves can be seen as expected values of statistics derived from the values of  $Y$  on a simple random sample with replacement. Those characteristics are called “population characteristics”. Their sample counterparts are obtained by substituting  $f_{Y[\mathbf{x}]}$  by  $\rho_{\{1,\dots,\text{size}(\mathbf{x})\}}f_{Y[\mathbf{x}]}$ , and can be interpreted as the expected value of the same statistics derived from the values of  $Y$  on the informative sample  $S$ .

In the case of sampling with fixed size  $\mathbf{n} \geq 2$ , define the sample semivariogram for exchangeable designs as

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

In the case of a random size sample, one can define the sample semivariogram as

$$G^*(h) = \frac{1}{2} \sum_{\mathbf{n} \geq 2} P(N = \mathbf{n}) E \left[ (Y[S[1]] - Y[S[2]])^2 \mid N = \mathbf{n} \text{ and } S[2] - S[1] = h \right].$$

**Property 3.6** (Relationship between  $G$ ,  $G^*$  and  $\rho$ ).

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

*Proof.* Let  $\mathbf{x} \in U^{\{1,2\}}$ , then

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

□

## 4 Estimation

### 4.1 Naive non parametric estimation : the case of the variogram

With the term *naive estimation* we refer to the situation where the estimator does not take into account the mechanism selection. Therefore, in case of informative sampling, the estimator could suffer from selection bias. 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)

$$2\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{(\ell_1, \ell_2) \in N(h)} (Y[S[\ell_1]] - Y[S[\ell_2]])^2, \forall h \in \mathbb{R}^d$$

where  $N(h) = \{(\ell_1, \ell_2) \in \{1, \dots, N\}^2 : S[\ell_1] - S[\ell_2] = h\}$  and  $|N(h)|$  is the number of distinct pairs in  $N(h)$ .

**Property 4.1.**  $E[N(h)\hat{G}(h)] = G(h)$

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

$$2\hat{\gamma}_s(h(l)) = \text{average}\{(Y[S[\ell_1]] - Y[S[\ell_2]])^2 : \|S[\ell_1] - S[\ell_2]\| \in [h \mp \alpha]\}, \quad (38)$$

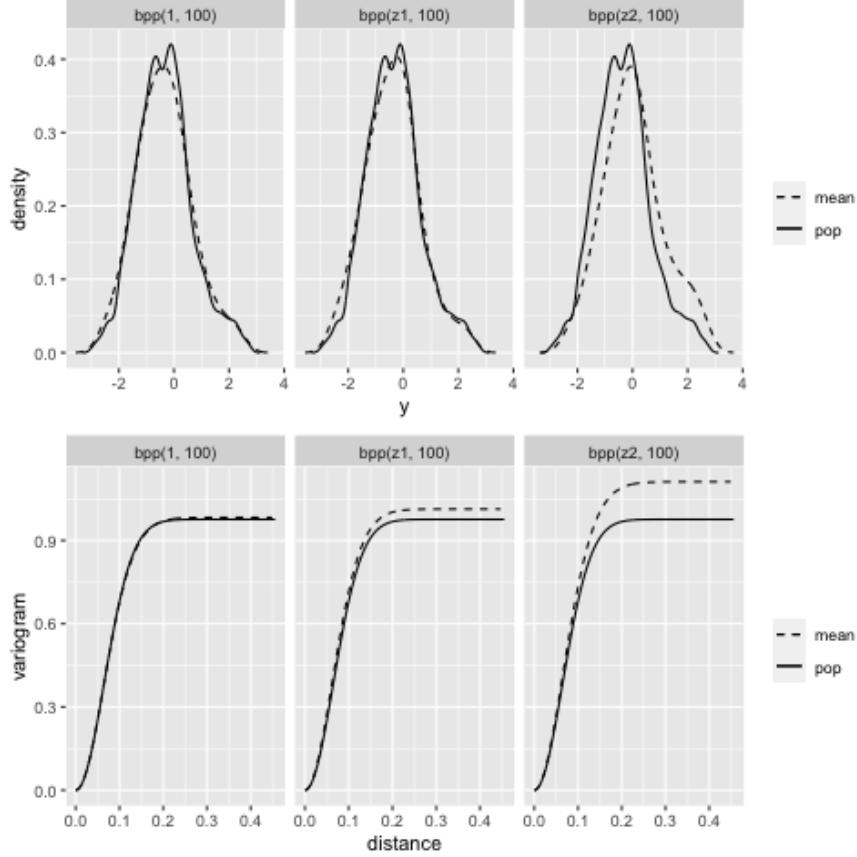
where the tolerance  $\alpha$  is a positive number.

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-fitness criteria, such as maximum likelihood estimator or Least Squares method.

In order to investigate the behaviour of the naive estimation, we carried out a Monte Carlo simulation. We simulated one replication of the same isotropic Gaussian process considered in the Example 1, that is  $Y : \Omega \rightarrow (U = [0, 1]^2 \rightarrow \mathbb{R})$ , with  $\forall \mathbf{x} \in U, \mu[\mathbf{x}] = 0$  and with a Gaussian Covariogram with parameters  $\theta_1 = 5, \theta_2 = 0.1$ . Samples were selected by means of  $bpp(1, 100)$ ,  $bpp(\mathbf{z}_1, 100)$ , and  $bpp(\mathbf{z}_2, 100)$ , where  $\mathbf{z}_1$  is generated with parameters  $(\log(10) - (0.4^2 + 0.3^2), 0, \sqrt{0.4^2 + 0.3^2})$ , and  $\mathbf{z}_2$  is generated with parameters  $(\log(10) - (0.4^2 + 0.3^2), 0.4, 0.3)$ , as in the Example 2. The  $bpp(1, \mathbf{n})$  is by definition a non-informative sampling mechanism, while the  $bpp(\mathbf{z}, \mathbf{n})$  could introduce bias when the variable  $\mathbf{z}$  is correlated with the signal  $\mathbf{y}$ , as explained in the previous Sections. Note that  $\mathbf{z}_1$  was generated independently from  $\mathbf{y}$  while  $\mathbf{z}_2$  was generated dependently from  $\mathbf{y}$ . Indeed,  $\text{cor}(\mathbf{y}, \mathbf{z}_1) \approx -0.08$  and  $\text{cor}(\mathbf{y}, \mathbf{z}_2) \approx 0.75$ , where  $\text{cor}$  indicates the correlation between the two variables. For each sampling mechanism,  $M = 1,000$  samples were selected, and variograms estimated by the method of moments and fitted by the Weighted Least Square criterion. These estimated variograms are compared to the variogram obtained by using all the population values, which we call *population variogram*. In the first row of Figure 7, we compare the averages of the densities obtained by samples selected by means of the three different selection mechanisms. We note that when  $bpp(\mathbf{z}_2, 100)$  is employed, the average density is different from the *population density*. In the second row of Figure 7, we compare the expected value of the variograms obtained by the replications to the population variogram. The simulation shows that when the samples are selected by means of  $bpp(1, 100)$  and  $bpp(\mathbf{z}_1, 100)$ , the Monte Carlo expected value of the sample variogram (indicated by the dotted line) is close to the population variogram (indicated by continuous line), while when the samples are selected by means of  $bpp(\mathbf{z}_2, 100)$ , the Monte Carlo expected value of the sample variogram (indicated by the dotted line) is different from the population variogram (continuous line). Therefore, bias is introduced by the selection mechanism.

Figure 7: Naive estimation of the variogram. In the first row, average densities (dotted lines) are compared with the population density (continuous line). In the second row, expected values of the variogram (dotted lines) are compared with population variogram (continuous line).



## 4.2 Maximum likelihood estimation

Here we briefly present a possible solution that take into account the informativeness of the sample. The proposal is based on the maximum likelihood estimation. In particular, in our setup, the loglikelihood of  $Y[\mathbf{x}]$  is given by

$$\mathcal{L}_{Y[\mathbf{x}]}(\theta; \mathbf{y}) = \log(f_{Y[\mathbf{x}]}(\mathbf{y}; \theta)) \quad (39)$$

$$= \frac{1}{2} \log |\Sigma_{\mathbf{x}, \mathbf{x}'; \theta}| + \frac{1}{2} (Y[S])^T \Sigma_{\mathbf{x}, \mathbf{x}'; \theta}^{-1} Y[S] \quad (40)$$

and the naive estimator maximizes (39). Indeed, this maximization ignores the selection mechanism.

In order to take into account the informativeness of the sample, the *full loglikelihood* should be used, which is composed by three factors: the density ratio,  $\rho_S(\mathbf{x} \mid \mathbf{y})$ , the distribution of  $Y$ ,  $f_{Y[\mathbf{x}]}(\mathbf{y})$ , and the distribution of the sample,  $f_S(\mathbf{x})$ . Therefore the loglikelihood is

$$\mathcal{L}_{S,Y[S],N}(\theta, \xi; \mathbf{x}, \mathbf{y}, \mathbf{n}) = \log(\rho_S(\mathbf{x} \mid \mathbf{y}; \theta, \xi)) + \log(f_{Y[\mathbf{x}]}(\mathbf{y}; \theta)) + \log(f_S(\mathbf{x}; \theta, \xi)) \quad (41)$$

## 5 Prediction

The following property provides the formula for the predictive distribution of  $Y[\mathbf{x}']$  when  $S, Y[S] = \mathbf{x}, \mathbf{y}$  is observed.

**Property 5.1.** *The probability density function of  $Y[\mathbf{x}']$  conditionally on  $(S, Y[S] = (\mathbf{x}, \mathbf{y})$  is given, for  $\mathbf{x}' \in \bar{\mathbf{U}}, \mathbf{y}' \in \mathcal{Y}^{\{1, \dots, \text{size}(\mathbf{x}')\}}$  by:*

$$P^{S,Y[S]} - a.s.(\mathbf{s}, \mathbf{y}), f_{Y[\mathbf{x}']|S,Y[S]}(\mathbf{y}' \mid \mathbf{x}, \mathbf{y}) = \frac{f_{S|Y[\mathbf{x}]=\mathbf{y}, Y[\mathbf{x}']}(\mathbf{x} \mid \mathbf{y}, \mathbf{y}')}{f_{S|Y[\mathbf{x}]}(\mathbf{x} \mid \mathbf{y})} \times f_{Y[\mathbf{x}']|Y[\mathbf{x}]}(\mathbf{y}' \mid \mathbf{y}) \quad (42)$$

$$\frac{f_{S|Y[\mathbf{x}]=\mathbf{y}, Y[\mathbf{x}']}(\mathbf{x} \mid \mathbf{y}, \mathbf{y}')}{f_{S|Y[\mathbf{x}]}(\mathbf{x} \mid \mathbf{y})} = \frac{\int \frac{\exp(-\xi_1(\sum_{\ell=1}^2 \mathbf{y}(\ell) - \xi_2(\sum_{\ell=1}^n \varepsilon[\mathbf{x}(\ell)])))}{(\int \exp(-\xi_1 Y[\mathbf{x}'] - \xi_2 \varepsilon[\mathbf{x}']) d\nu(\mathbf{x}'))^{n'}} dP_{Y,\varepsilon|Y[\mathbf{x}]=\mathbf{y}, Y[\mathbf{x}']=\mathbf{y}'}}{\int \frac{\exp(-\xi_1(\sum_{\ell=1}^n Y[\mathbf{x}(\ell)] - \xi_2(\sum_{\ell=1}^n \varepsilon[\mathbf{x}(\ell)])))}{(\int \exp(-\xi_1 Y[\mathbf{x}'] - \xi_2 \varepsilon[\mathbf{x}']) d\nu(\mathbf{x}'))^{n'}} dP_{Y,\varepsilon|Y[\mathbf{x}]=\mathbf{y}}}.$$

## 6 Conclusions

We have extended the notions of informative selection, population and sample distributions defined by Krieger and Pfeffermann (1992) to a situation where a spatial process is assumed to have generated the whole population. In particular, a Gaussian random field and two point processes that represent the selection mechanism have been considered. An analysis of the naive estimation of the variogram have been performed, and simulation shows that selection bias is introduced when there is dependence between the superpopulation model and the mechanism selection. Finally, we briefly discussed a possible solution in order to correct the naive estimation in presence of informative selection. This solution is based on the use of the maximum likelihood. Future research will be dedicated to a proper development of such proposal.

## Acknowledgements

We developed an R package that allows to reproduce all the simulation results see Bonnery and Pantalone (2021). Francesco Pantalone's work was partially funded by the “*International Graduate Research Fellowships*” at Joint Program in Survey Methodology, University of Maryland, College Park, USA.

## 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  $\text{Var}[(\alpha + \beta Y + \gamma \varepsilon)[\mathbf{x}']] = \gamma^2 \Sigma_{\mathbf{x}', \mathbf{x}'}^\varepsilon + \beta^2 (\Sigma_{\mathbf{x}', \mathbf{x}'} - \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1} \Sigma_{\mathbf{x}, \mathbf{x}'})$ , and consequently, conditionally on  $Y[\mathbf{x}] = \mathbf{y}$ ,  $Z = \exp(\alpha + \beta Y + \gamma \varepsilon)$  is a lognormal spatial process.

*Proof.*

$$\begin{bmatrix} Y[\mathbf{x}'] \\ Y[\mathbf{x}] \end{bmatrix} \sim \text{Normal} \left( \mu, \begin{bmatrix} \Sigma_{\mathbf{x}', \mathbf{x}'} & \Sigma_{\mathbf{x}', \mathbf{x}} \\ \Sigma_{\mathbf{x}, \mathbf{x}'} & \Sigma_{\mathbf{x}, \mathbf{x}} \end{bmatrix} \right) \quad (43)$$

$$\Rightarrow [Y[\mathbf{x}'] | Y[\mathbf{x}] = \mathbf{y}] \sim \text{Normal} (\mu + \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1}(\mathbf{y} - \mu), \Sigma_{\mathbf{x}', \mathbf{x}'} - \Sigma_{\mathbf{x}', \mathbf{x}} \Sigma_{\mathbf{x}, \mathbf{x}}^{-1} \Sigma_{\mathbf{x}, \mathbf{x}'}) \quad (44)$$

The independence of  $\varepsilon$  and  $Y$  implies that:

$$[\varepsilon[\mathbf{x}'] | Y[\mathbf{x}] = \mathbf{y}] \sim \text{Normal} (\mu_\varepsilon, \Sigma_{\mathbf{x}', \mathbf{x}'; \varepsilon}), \quad (45)$$

and that the vector obtained by stacking  $Y[\mathbf{x}]$ ,  $Y[\mathbf{x}']$ , and  $\varepsilon[\mathbf{x}']$  is normal. By combining (45) and (44) 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 Algebra for $\Sigma$

The denominator of  $\rho$  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:

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

## 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.
- Bonney, D. and Sedransk, J. (2019). On the definition of informative vs. ignorable nuisance process. <http://export.arxiv.org/abs/1906.02733>.
- Bonney, D. B. and Pantalone, F. (2021). Spatial informative selection. <https://github.com/DanielBonney/SpatialInformativeSelection>.
- 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 Interscience.
- Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.
- Doucet, A. (2010). A note on efficient conditional simulation of gaussian distributions. *Departments of Computer Science and Statistics, University of British Columbia*, 1020.
- Krieger, A. M. and Pfeffermann, D. (1992). Maximum likelihood estimation from complex sample surveys. *Survey Methodology*, 18(2):225–239.

- 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*.
- Wilson, J. T., Borovitskiy, V., Terenin, A., Mostowsky, P., and Deisenroth, M. P. (2020). Pathwise conditioning of gaussian processes. *arXiv preprint arXiv:2011.04026*.