Empirical variograms

Giovanni Pistone

Collegio Carlo Alberto

DE CASTRO STATISTICS

Introduction

Consider a response surface on a given real domain D (usually a rectangle). A measurement is available at each testing points $x \in D$. We want to assess the conformity of the shape of the response surface to some standard. For example: "is the surface bended in some direction?" Or: "Is there a waviness of a type associate to a specific technology?" These are possible defects that cannot be specified in a parametric way [5, 6].

A very popular modeling method relies on the assumption that the surface under study is the realization of a random field, for example, a Gaussian random field $(\zeta_x)_{x\in D}$. In such a case, the observed characteristics of the surface will in fact depend on the auto-covariance of the random field.

More specifically, under the intrincic stationarity assumpion $\mathbb{E}(\zeta_x) = m$ and $\mathbb{E}(|\zeta_x - \zeta_y|) = \gamma(d(x,y))$, the properties of the random surface depend on the random field variogram γ [1]. This approach is sometimes called Kriging method or Matheron method and can be used both in a parametric or non-parametric approach. Cf. [2, 3, 4].

G. Vicario has suggested in [7] to use the variogram as a non-parametric method, without assuming any randomness of the surface, but using instead the idea of a sistematic or random sampling of couples of test points on the given domain.

Methodology

The following definition is intended to mimic the definition of the Matheron empirical estimator of the variogram.

Definition *Let X and Y be independent random variables whose common distribution μ is supported by the domain $D \in \mathbb{R}^n$ and let d be a distance on D. Let $F \colon D \to \mathbb{R}$ be a response function. The empirical variogram γ is defined by*

$$\gamma(d(X,Y)) = \frac{1}{2} E\left(|F(X) - F(Y)|^2 \middle| d(X,Y)\right) .$$

If $v \sim T = d(X,Y)$, for an arbitrary Φ it holds

$$\int \Phi(t)\gamma(t)\ \nu(dt) = \int_{D\times D} \Phi(d(x,y))\frac{1}{2}|F(x) - F(y)|^2\ \mu(dx)\mu(dy) .$$

The theoretical computation of the variogram for given F and μ is relevant to test the methodological relevance (if any) of this defintion. Such a computation is simpler when the distance is of the form $d(x,y) = \|x - y\|$, for some norm on $\mathbb{R}^n$ and the μ is a Lebesgue probability measure, $\mu(dx) = dx$. In such a case, the defining equation becomes, with the change of variables $u = x - y$ and $v = x$,

$$\int \Phi(t)\gamma(t)\ \nu(dt) = \int_{D\times D} \Phi(\|u\|)\frac{1}{2}|F(v) - F(u+v)|^2\ dudv .$$

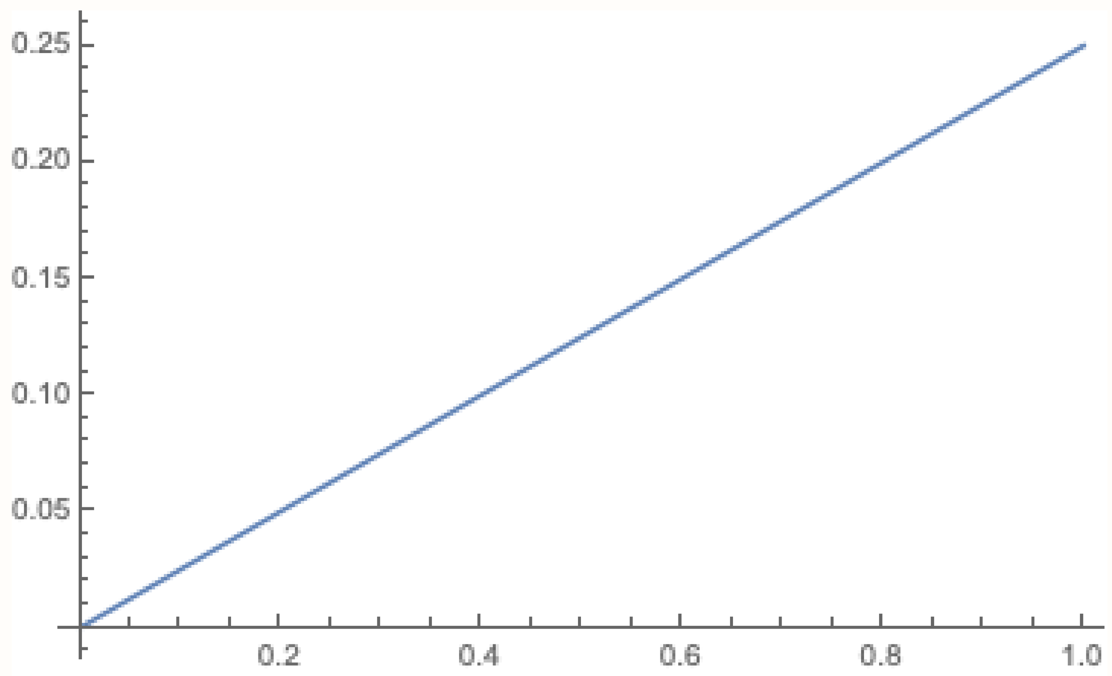
Here are some immediate properties of the empirical variogram in the norm case.

1. If F is constant, then $\gamma = 0$.
2. The variogram depends quadratically on the gradient ∇F. In fact,
$$F(v+u) - F(v) = \int_0^1 \nabla F(v + \theta u)\cdot u\ d\theta .$$
3. If $D =]0,1[$ and X, Y are uniform,
$$\gamma(t) = \frac{1}{2(1-t)}\int_0^{1-t} (F(s) - F(s+t))^2\ ds .$$
4. If F is linear, $F(x) = a \cdot x$, then $|F(v+u) - F(v)|^2 = (a\cdot u)^2$ and the defining equation becomes
$$\int \Phi(t)\gamma(t)\ \nu(dt) = \int_D \Phi(\|u\|)\frac{1}{2}(a\cdot u)^2\ dudv .$$
5. If F is Lipschitz, we can derive an upper bound for the variogram. In fact,
$$|F(v+u) - F(v)|^2 \le \|F\|^2_{\mathrm{Lip}}\|u\|^2 ,$$
implies
$$\gamma(t) \le \frac{1}{2}\|F\|^2_{\mathrm{Lip}}t^2 .$$

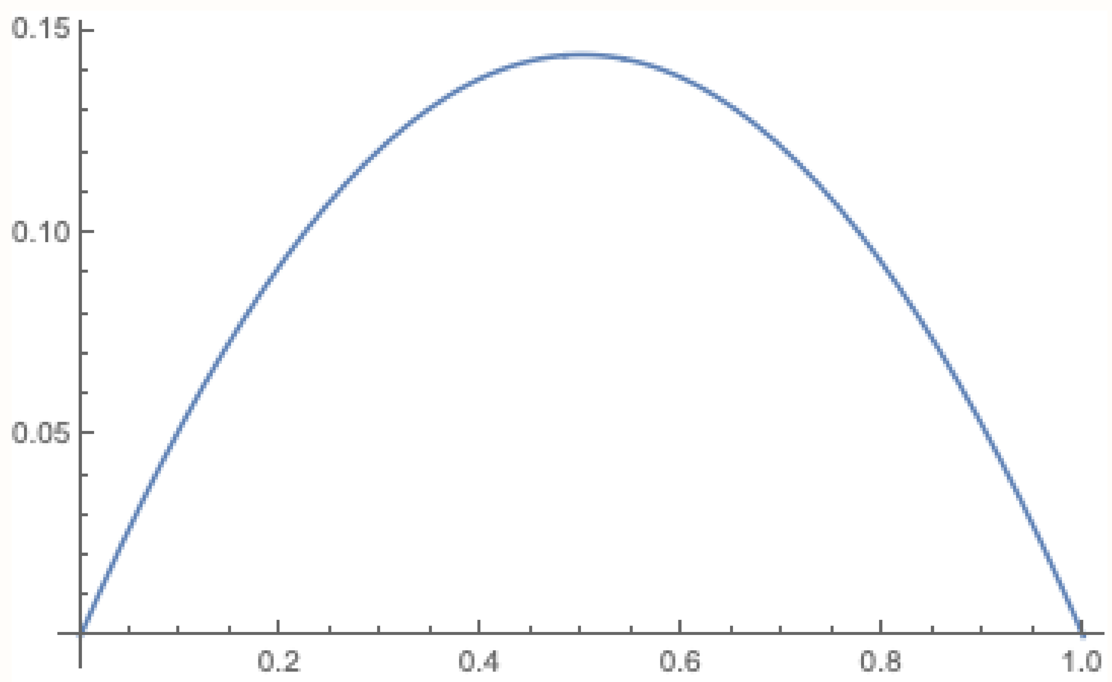
Examples

With $D =]0,1[$ and uniform X, Y, we show below the function F and a graph of $\sqrt{2\gamma}$.

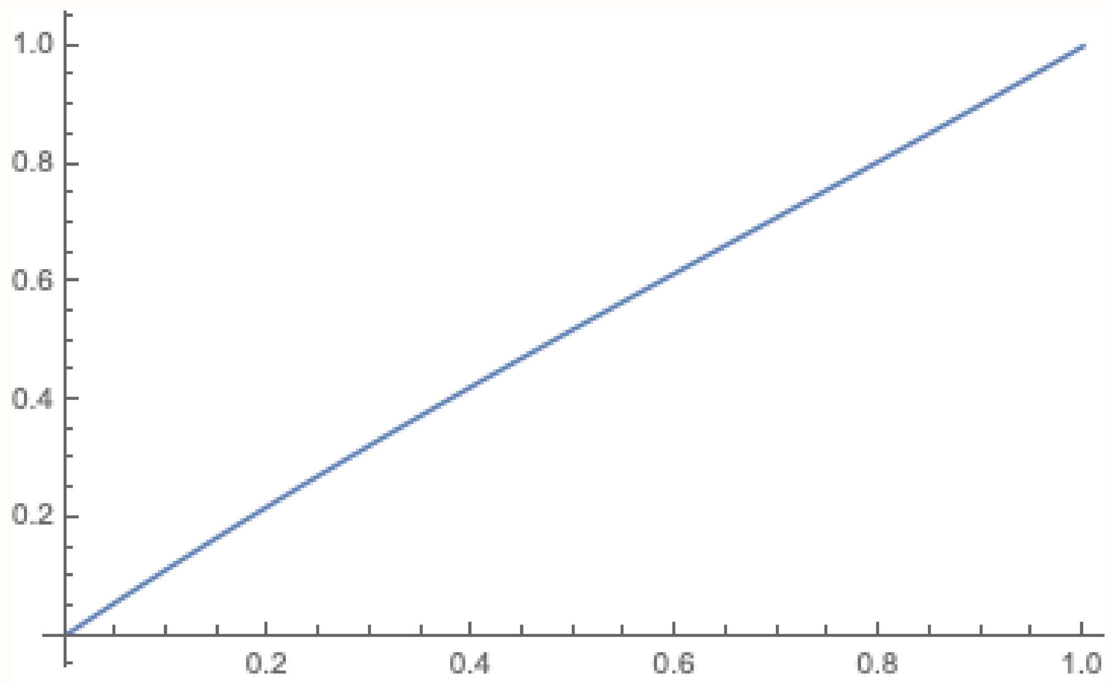
Affine: $F(x) = 1 + \frac{1}{4}x$



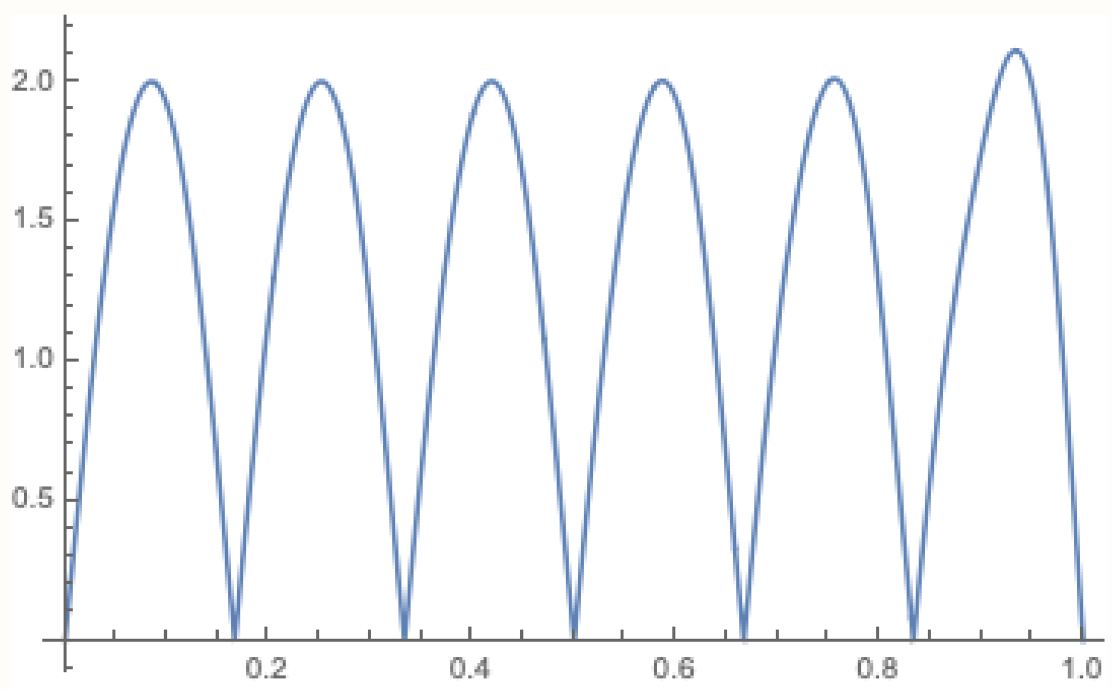
Parabolic bump: $F(x) = x(1-x)$



Parabolic bend: $F(x) = 1 - x^2$



Sine function: $F(x) = \sin(6(2\pi)x)$



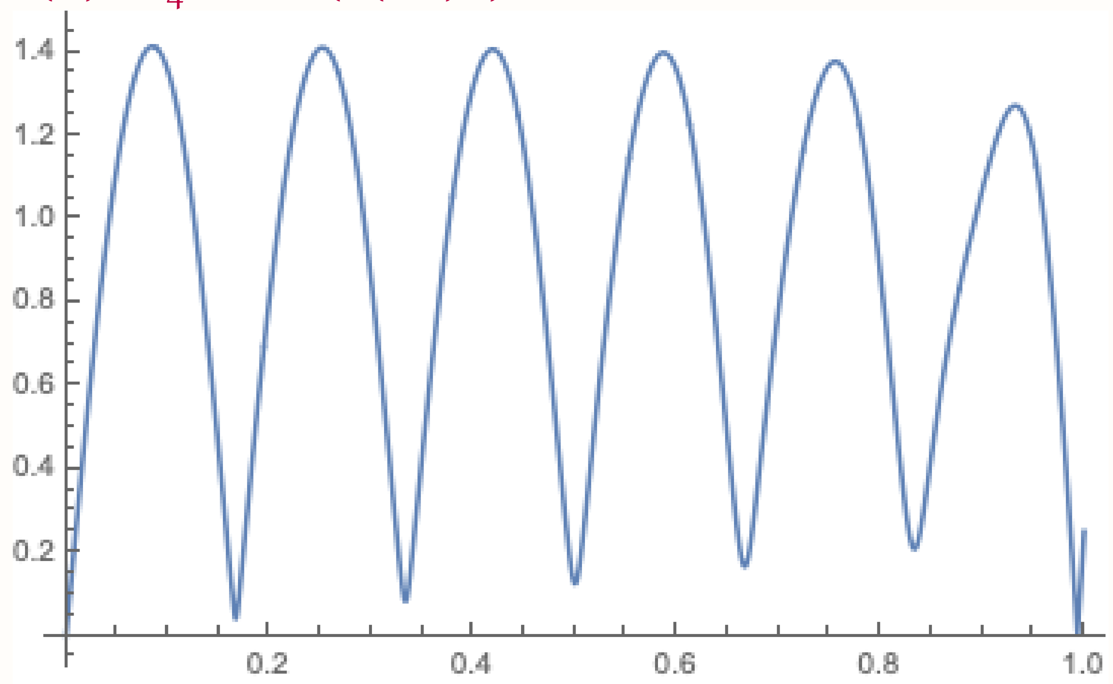
Superposed functions

The variogram of $F_1 + F_2$ appears to be difficult to undestand in terms of the separate variograms because there is an interaction term:

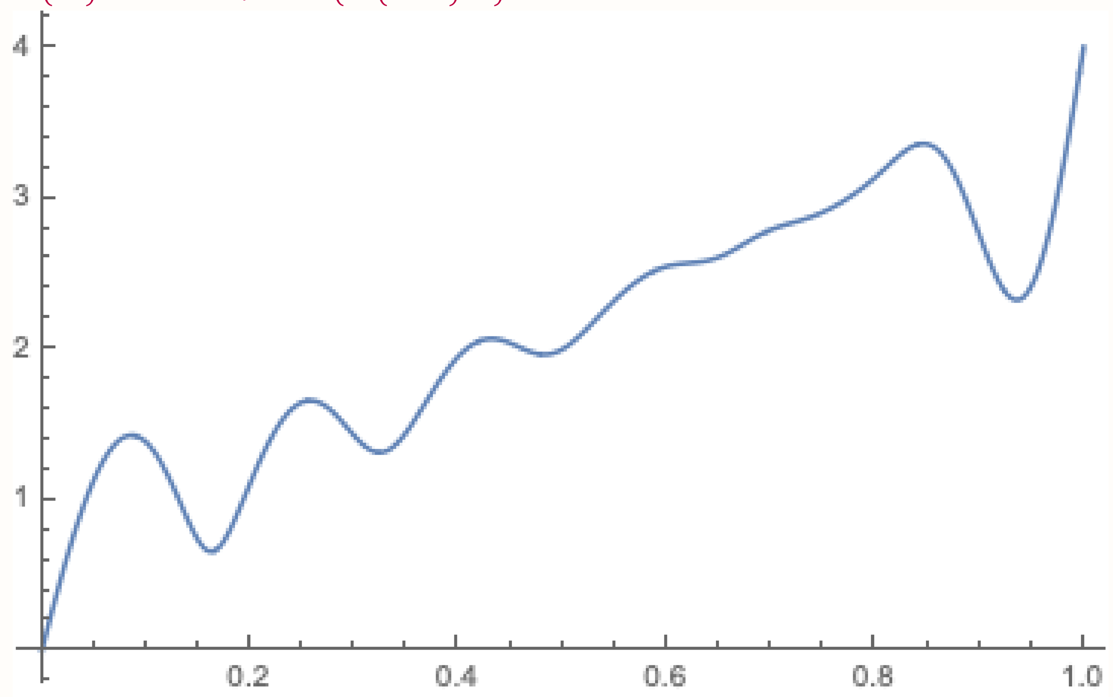
$$\gamma_{1+2} = \gamma_1 + \gamma_2 + \gamma_{12} ,$$

where $\gamma_{1,2}$ is defined by the polarised version of the definition of γ. Here are two examples.

$F(x) = \frac{1}{4}x + \sin(6(2\pi)x)$



$F(x) = 4x + \sin(6(2\pi)x)$



Acknowledgements

The author was supported by de Castro Statistics and by Collegio Carlo Alberto, Piazza Vincenzo Arbarello 8, 10122 Torino. He is a member of INdAM-GNAMPA and of Enbis. A draft version of the full paper will be posted on the author web-site https://www.giannidiorestino.it and will eventually be uploaded to the arXiv. That version will include a sampling version of the empirical variogram with warranty bounds.

REFERENCES

[1] Noel A. C. Cressie, *Statistics for spatial data*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Inc., 1993, Revised reprint of the 1991 edition, A Wiley-Interscience Publication. MR 1239641 (94h:62155)

[2] Georges Matheron, *Traité de géostatistique appliqué*, Mem. Bur. Rech. Geog. Minieres, no. 14, Editions Technip, 1962.

[3] Giovanni Pistone and Grazia Vicario, *Kriging prediction from a circular grid: application to wafer diffusion*, Appl. Stochastic Models Bus. Ind. **29** (2013), no. 4, 350–361.

[4] ———, *A note on semivariogram*, Topics on Methodological and Applied Statistical Inference (W. Racugno T. Di Battista, E. Moreno, ed.), Springer, 2016, pp. 181–190.

[5] Grazia Vicario, *Inference on errors in industrial parts: Kriging and variograms versus geometrical product specifications (GPS) standard*, This Congerence, 2020.

[6] Grazia Vicario, G Craparotta, and Giovanni Pistone, *Meta-models in computer experiments: Kriging vs artificial neural networks*, Quality and Reliability Engineering International **32** (2016), 2055—2065.

[7] Grazia Vicario and Giovanni Pistone, *Simulated variogram-based error inspection of manufactured parts*, Statistical Papers **59** (2018), 1411–1423.