Saddlepoint approximations for the distribution of some robust estimators of the variogram

Alfonso García-Pérez

Versión final publicada en ***Metrika*, 2020, vol. 83: 69-91.**

https://link.springer.com/article/10.1007/s00184-019-00725-6

Departamento de Estadística, I.O. y C.N., Universidad Nacional de Educación a Distancia (UNED),

Paseo Senda del Rey 9, 28040 Madrid, Spain; agar-per@ccia.uned.es

Abstract: In this paper, we obtain a saddlepoint approximation for the small sample distribution of several variogram estimators such as the classical Matheron's estimator, some M-estimators like the robust Huber's variogram estimator, and also the α-trimmed variogram estimator. The tail probability approximation obtained is very accurate even for small sample sizes. In the approximations we consider that the observations follow a distribution close to the normal, specifically, a scale contaminated normal model. To obtain the approximations we transform the original observations into a new ones, which can be considered independent if a linearized variogram can be accepted as model for them. To check this, a goodness of fit test for a variogram model is defined in the last part of the paper.

Keywords: Robustness; Spatial data; saddlepoint approximations

MSC: 62F35; 62E17; 62H11

1 Introduction and notation

Variogram estimation is a very important issue in Geostatistics where a random variable Z is observed at some known fixed locations $s \in D$, being D a fixed subset of R^d , $d \geq 1$.

Thus, it is also very important to obtain the distribution of the variogram estimator because it can be used to analyze the properties of the estimator and to check the fit of a particular variogram model, which can be used, for instance, in kriging interpolation.

In Matheron (1962), a classical estimator $2\hat{\gamma}^M(\mathbf{h})$ of the variogram was introduced, an estimator which is widely used in applications of spatial statistics. In Cressie and

Hawkins (1980) or Genton (1998), some robust versions were defined but their distributions, not even some approximations of them, were obtained.

In fact, there is no useful approximation for the distribution of $2\hat{\gamma}_M(\mathbf{h})$ when the sample size is small. There are only some approximations if the sample size is large or if the n observations $Z(\mathbf{s}_i) = Z_i$ are i.i.d. Gaussian random variables; in this case, Ali (1987) obtained a small-sample beta approximation. Nevertheless, to suppose that the observations follow a normal model is unrealistic because, in practice, they are contaminated by occasional outliers. For this reason, we shall assume a model close to the normal and, specifically, a scale contaminated normal model,

$$F = (1-\epsilon)\, N(\mu, \sigma) + \epsilon\, N(\mu, g\sigma)$$

with $\epsilon \in (0, 1)$ (usually small) and $g > 1$, for the marginal distributions.

We shall also assume that the variable Z, from which we obtain the sequence of observations $Z(\mathbf{s_i})$, verifies the intrinsic stationarity property, i.e., (Cressie 1993, p. 40) that the differences have zero mean

$$E[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})] = 0, \quad \forall \mathbf{s}, \mathbf{s}+\mathbf{h} \in D$$

and the variance depends only on lag $\mathbf{h}$,

$$V(Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})) = 2\,\gamma(\mathbf{h}), \quad \forall \mathbf{s}, \mathbf{s}+\mathbf{h} \in D.$$

The function $2\,\gamma(\mathbf{h})$ is called the *variogram* (Matheron 1962),

$$2\,\gamma(\mathbf{h}) = V(Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})) = E[(Z(\mathbf{s}) - Z(\mathbf{s}+\mathbf{h}))^2].$$

The estimation of the variogram is made for several fixed lags $\mathbf{h}$, thereby obtaining the estimations $2\hat{\gamma}(\mathbf{h})$. Next, a modeling exercise on the pairs $(||\mathbf{h}||, \hat{\gamma}(\mathbf{h}))$ is carried out, fitting a suitable function to these pairs of data: the *semivariogram model*, where $||.||$ denotes the Euclidean norm.

The knowledge of the distribution of $2\hat{\gamma}(\mathbf{h})$ can be used to analyze its properties and check this fit.

Although the total number of observations is usually large, since the estimation of the variogram $2\,\gamma(\mathbf{h})$ is made for each $\mathbf{h}$, the sample size n used in each estimation, for fixed $\mathbf{h}$, could be small, and the saddlepoint approximations, suitable.

The rest of this paper is organized as follows. In Sect. 2 a transformation of the original variables is done to obtain the required independence of the observations and also the distribution of these new variables. In Sect. 3 the estimation problem is established emphasizing on the M-estimators.

Von Mises approximations and saddlepoint approximations are considered and applied to the estimators of the variogram in Sects. 4 and 5, respectively.

In Sect. 6 these approximations are applied to Matheron's estimator and in Sect. 7 to the α-trimmed variogram estimator. The approximation for the Huber's variogram estimator is obtained in Sect. 8. The independence condition is satisfied by a linearized version of the variogram which is defined in Sect. 9.

After we have obtained the approximations, we use them in Sect. 10 to define a confidence zone and a goodness of fit test for a variogram model.

If the linearized version of the variogram can be accepted, we could also accept the independence of the transformed variables.

Finally, some concluding remarks are given in Sect. 11.

2 Preliminary transformation

If we define the variable $Y_{\mathbf{s}}$ as

$$Y_{\mathbf{s}}(\mathbf{h}) = (Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s}))^2$$

the problem considered in the paper is an estimation problem for the expectation of the variable $Y_{\mathbf{s}}(\mathbf{h})$, because it is $E[Y_{\mathbf{s}}(\mathbf{h})] = 2\,\gamma(\mathbf{h})$. In this way, we follow the via of location estimation, suggested by Cressie and Hawkins (1980), instead of the via of variance estimation used in Genton (1998).

This transformation, usually shortened as

$$Y_s = (Z_{s+h} - Z_s)^2$$

is a key element in the paper because, although variables Z_s are not independent, Cressie and Hawkins (1980, pp. 119–120) argue, "*for problems of practical interest, the interdependence between* Y_t *and* Y_s *will be negligible except for a negligibly small proportion of the* Y_t, Y_s *pairs. We therefore feel justified in treating the* $\{Y_t\}$ as *if they were an independent random sample.*"

Delving into this matter, because the locations $\mathbf{s}_i$ are fixed in advance (for instance they could be sample stations), they can be considered as equally spaced on a transect, as for instance in Fig. 2.1 of Cressie (1993, p. 32). Then, we can match two contiguous Z_i (for which the dependence is supposed to be the strongest), so that it is $t+h=s$. The correlation between $\sqrt{Y_s} = Z_{s+h} - Z_s$ and $\sqrt{Y_t} = Z_{t+h} - Z_t$, will be 0 if and only if (Cressie and Hawkins 1980, p. 119)

$$\rho(\mathbf{h}) - \frac{1}{2}\,[\rho(2\mathbf{h}) + 1] = 0$$

or equivalently, if

$$2\,C(\mathbf{h}) = C(2\mathbf{h}) + C(\mathbf{0}) \tag{1}$$

being $\rho(\mathbf{h}) = C(\mathbf{h})/C(\mathbf{0})$ the *autocorrelation function* and $C(\mathbf{h}) = cov(Z(\mathbf{s}), Z(\mathbf{s}+\mathbf{h}))$ the *covariogram*.

Due to the definition of the variogram, we have that $2\,\gamma(\mathbf{h}) = 2C(\mathbf{0}) - 2C(\mathbf{h})$. Hence, it will be $C(\mathbf{h}) = C(\mathbf{0}) - \gamma(\mathbf{h})$ and $C(2\mathbf{h}) = C(\mathbf{0}) - \gamma(2\mathbf{h})$. Thus, we shall obtain equality (1), and so independence between the Y_s, if the semivariogram model is such that

$$\gamma(2\mathbf{h}) = 2\gamma(\mathbf{h}) \tag{2}$$

condition that is obtained if, for $||\mathbf{h}||$ before the range of the variogram, a linear semivariogram model can be accepted for the pairs $(||\mathbf{h}||, \hat{\gamma}(\mathbf{h}))$. This requirement in the semivariogram model will be checked in the last part of the paper.

Hence, in what follows, we shall consider the $\{Y_s\}$ as if they were a sample of independent variables, and the estimators will be based on them.

2.1 Distribution of variables $\{Y_s\}$

Although the main role is played by the normal marginal distributions of the Z_i, to complete the mathematical framework, let us say that these marginal distributions are obtained from the multivariate contaminated normal distribution with joint probability density function (pdf)

$$f_M(\mathbf{z}) = f_M(z_1, \dots, z_n) = (1-\epsilon) f_N(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) + \epsilon f_N(\mathbf{z}; \boldsymbol{\mu}, g^2 \boldsymbol{\Sigma})$$

where $\epsilon \in (0,1)$, $g > 1$, and where $f_N(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the pdf of a n-variate normal random vector having the multivariate normal distribution with mean vector $\boldsymbol{\mu} = (\mu, \dots, \mu)'$ and covariance matrix $\boldsymbol{\Sigma}$, a matrix where all the elements of its diagonal are equal to σ^2.

From this joint distribution, the marginal distributions will be the scale contaminated normal model,

$$\begin{aligned} f(z_i) &= (1-\epsilon) \int \cdots \int f_N(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) dz_1 \dots dz_{i-1} dz_{i+1} \dots dz_n \\ &\quad + \epsilon \int \cdots \int f_N(\mathbf{z}; \boldsymbol{\mu}, g^2 \boldsymbol{\Sigma}) dz_1 \dots dz_{i-1} dz_{i+1} \dots dz_n \\ &= (1-\epsilon)\, N(\mu, \sigma) + \epsilon\, N(\mu, g\sigma) \end{aligned}$$

$i = 1, \dots, n$, because of the properties a multivariate normal distribution (the marginal distributions are univariate normal) and the meaning of the elements of its covariance matrix.

Under no contamination, it is clear that, if observations Z_s follow a normal distribution,

$$Z_s \equiv N(\mu, \sigma)$$

where $X \equiv H$ stands for "X is distributed as H", then it is $(Z_{s+h} - Z_s) \equiv N\left(0, \sqrt{2\gamma(\mathbf{h})}\right)$ and

$$Y_s \equiv 2\,\gamma(\mathbf{h})\,\chi_1^2.$$

Under contamination, if

$$Z_s \equiv F = (1-\epsilon) N(\mu, \sigma) + \epsilon N(\mu, g\sigma)$$

because the joint distribution of the vector $\mathbf{Z} = (Z_1, \dots, Z_n)$ is the multivariate contaminated normal distribution, the distribution of $Y_s = (Z_{s+h} - Z_s)^2$ is

$$\begin{aligned} P\{Y_s \leq y\} = (1-\epsilon) \int\int_A & f_1(z_s, z_{s+h}; \boldsymbol{\mu_2}, \boldsymbol{\Sigma_2})\, dz_s dz_{s+h} \\ +\epsilon \int\int_A & f_2(z_s, z_{s+h}; \boldsymbol{\mu_2}, g^2 \boldsymbol{\Sigma_2})\, dz_s dz_{s+h} \end{aligned}$$

being A the set

$$A = \{(z_s, z_{s+h}) : (z_{s+h} - z_s)^2 \leq y\},$$

f_1 a bivariate normal distribution with mean vector $\boldsymbol{\mu_2} = (\mu, \mu)'$ and covariance matrix $\boldsymbol{\Sigma_2}$, a matrix in which the two elements in the diagonal are equal to σ^2, and f_2 a bivariate normal distribution with mean vector $\boldsymbol{\mu_2}$ and covariance matrix $g^2 \boldsymbol{\Sigma_2}$.

Hence, under f_1, the variable $\sqrt{Y_s} = Z_{s+h} - Z_s$ is the difference between two univariate normal distributions (and so, with normal distribution, where no correlation means independence) with variance $V(\sqrt{Y_s}) = 2\gamma(\mathbf{h})$ because of the intrinsic stationarity property of the Z_i. Similarly, under f_2, the variable $\sqrt{Y_s} = Z_{s+h} - Z_s$ is the difference of two univariate normal distributions with variance $V(\sqrt{Y_s}) = g^2 2\gamma(\mathbf{h})$.

Then, standardizing, the distribution of Y_s will be

$$\begin{aligned} P\{Y_s \leq y\} &= (1-\epsilon) P_{f_1} \left\{ (Z_{s+h} - Z_s)^2 \leq y \right\} \\ &\quad + \epsilon P_{f_2} \left\{ (Z_{s+h} - Z_s)^2 \leq y \right\} \\ &= (1-\epsilon) P_{f_1} \left\{ \left(\frac{Z_{s+h} - Z_s}{\sqrt{2\gamma(\mathbf{h})}} \right)^2 \leq \frac{y}{2\gamma(\mathbf{h})} \right\} \\ &\quad + \epsilon P_{f_2} \left\{ \left(\frac{Z_{s+h} - Z_s}{\sqrt{g^2 2\gamma(\mathbf{h})}} \right)^2 \leq \frac{y}{g^2 2\gamma(\mathbf{h})} \right\} \\ &= (1-\epsilon) P \left\{ \chi_1^2 \leq \frac{y}{2\gamma(\mathbf{h})} \right\} + \epsilon P \left\{ \chi_1^2 \leq \frac{y}{g^2 2\gamma(\mathbf{h})} \right\} \\ &= (1-\epsilon) G_{\chi_1^2} \left(\frac{y}{2\gamma(\mathbf{h})} \right) + \epsilon G_{\chi_1^2} \left(\frac{y}{g^2 2\gamma(\mathbf{h})} \right) \end{aligned}$$

being $G_{\chi_1^2}$ the cumulative distribution function of a χ_1^2. But this is the cumulative distribution function of a contaminated chi-square model

$$(1-\epsilon)\, 2\gamma(\mathbf{h})\, \chi_1^2 + \epsilon\, g^2\, 2\gamma(\mathbf{h})\, \chi_1^2$$

because of the definition of this model and the total probability theorem. Hence, we can say that

$$Y_s \equiv (1-\epsilon)\, 2\gamma(\mathbf{h})\, \chi_1^2 + \epsilon\, g^2\, 2\gamma(\mathbf{h})\, \chi_1^2$$

Therefore, to say that Z_s follows a distribution model close to the normal, can be formalized in the context of the paper saying that Y_s has a distribution close to the $2\,\gamma(\mathbf{h})\,\chi^2_1$, for instance, $(1-\epsilon)\,2\,\gamma(\mathbf{h})\,\chi^2_1 + \epsilon\, g^2\, 2\,\gamma(\mathbf{h})\,\chi^2_1$.

3 Estimation problem

In this paper we obtain accurate saddlepoint approximations for the distributions of the estimators defined in Cressie and Hawkins (1980), via location estimation instead of the via scale estimation considered in Genton (1998). All these estimators will be based on the, supposedly independent, $\{Y_i\}$.

These approximations can be used to study the properties of the estimators, to test a model for the variogram and also to check the supposed independence of the $\{Y_i\}$.

Some of the variogram estimators T_n will be M-estimators with score function $\psi : \mathcal{Y} \times \Theta \longrightarrow I\!R$, i.e., the solution of the equation

$$\sum_{i=1}^{n} \psi(Y_i, T_n) = 0 \tag{3}$$

assuming that $\psi(y,\theta)$ is monotonic decreasing in θ for all y. In fact, T_n will be an estimator for a location problem, being $\psi(y,\theta)$ of the form $\psi(y-\theta)$, with $\psi(u)$ monotonic increasing in u (Daniels 1983).

We shall also obtain an accurate approximation for the case in which T_n is the α-trimmed mean of the Y_i.

4 von Mises approximation for the distribution of T_n

The first order von Mises expansion of a functional T at a model distribution F is (Withers 1983 or Serfling 1980)

$$T(F) = T(G) + \int \mathrm{IF}\,(x; T, G)\; dF(x) + Rem$$

where IF is the Hampel's Influence Function (Hampel 1974). The remainder term is

$$Rem = \frac{1}{2} \int\int T^{(2)}_{G_F}(x_1, x_2)\, d[F(x_1) - G(x_1)]\, d[F(x_2) - G(x_2)]$$

being $T^{(2)}_{G_F}$ the *second derivative* of the functional T at the mixture distribution $G_F = (1-\lambda)\,G + \lambda\,F$, for some $\lambda \in [0,1]$ and $x, x_1, x_2 \in \mathbb{R}$.

If the distributions F and G are close enough, the remainder term Rem will be small, and the *von Mises approximation*

$$T(F) \simeq T(G) + \int \mathrm{IF}\,(x; T, G)\; dF(x) \tag{4}$$

will be an accurate approximation for the functional T at the model distribution F. Distribution G is chosen so that we know the value of the leading term $T(G)$ and plays an important role; it is called *pivotal distribution*.

In the particular case that we consider the Tail Probability functional

$$T(F) = P_{X_i \equiv F}\{T_n > t\}$$

we obtain the *von Mises approximation for the distribution of* T_n

$$P_F\{T_n > t\} \simeq P_G\{T_n > t\} + \int \mathrm{TAIF}\,(x; t; T_n, G)\; dF(x) \tag{5}$$

where $\mathrm{TAIF}\,(x; t; T_n, G)$ is the Tail Area Influence Function (Field and Ronchetti 1985), defined as

$$\mathrm{TAIF}\,(x; t; T_n, G) = \left. \frac{\partial}{\partial \epsilon} P_{G_{\epsilon,x}}\{T_n > t\} \right|_{\epsilon=0}$$

for all $x \in \mathbb{R}$ where the right hand side exists. This influence function measures the influence on the tail probability of contamination in the underlying model because the TAIF is just the Hampel's influence function of the tail probability functional, i.e. it is calculated changing the underlying model G by a contaminated model $(1-\epsilon)G + \epsilon\delta_x$, before computing the first derivative at $\epsilon = 0$, being δ_x the distribution which assigns mass 1 at x.

If $F = (1-\epsilon)G + \epsilon H$ the von Mises approximation (5) is

$$P_F\{T_n > t\} \simeq P_G\{T_n > t\} + \epsilon \int \mathrm{TAIF}\,(x; t; T_n, G)\; dH(x) \tag{6}$$

because $\int \mathrm{TAIF}\,(x; t; T_n, G)\; dG(x) = 0$.

This approximation will be very useful if we choose, as pivotal distribution G, a model for which the tail probability $P_G\{T_n > t\}$ is easy to compute.

5 Saddlepoint approximation for the distribution of an *M*-estimator of the variogram

Let T_n be an M-estimator. Next we are going to obtain a saddlepoint approximation for the $\mathrm{TAIF}\,(x; t; T_n, G)$ to be included in Eq. (5). The resulting expression will be called *VOM+SAD approximation*.

We shall assume $Y_i \equiv G = 2\,\gamma(\mathbf{h})\,\chi_1^2$ as pivotal distribution because of the comments at the end of Sect. 2, since it is $Z_i \equiv N(\mu, \sigma)$.

To obtain a saddlepoint approximation for the $\mathrm{TAIF}\,(x; t; T_n, G)$, we can use Lugannani and Rice (1980) formula (Jensen 1995, p. 77; or better Huber and Ronchetti 2009, p. 314), to obtain the approximation given in Daniels (1983) for M-estimators T_n:

$$P_{Y_i \equiv G}\left\{T_n(Y_1,\ldots,Y_n) > t\right\} = 1 - \Phi(s) + \phi(s)\left[\frac{1}{r} - \frac{1}{s} + O(n^{-3/2})\right]$$

where s and r are the functionals

$$s = \sqrt{-2nK(z_0,t)}$$
$$r = z_0\sqrt{nK''(z_0,t)}$$

being $K(\lambda,t)$ the function

$$K(\lambda,t) = \log\int_{-\infty}^{\infty} e^{\lambda\psi(y,t)}\,dG(y)$$

$K''(\lambda,t)$ the second partial derivative of $K(\lambda,t)$ with respect to the first variable; z_0 the saddlepoint, i.e., the solution of the *saddlepoint equation*

$$K'(z_0,t) = \int_{-\infty}^{\infty} e^{z_0\psi(y,t)}\,\psi(y,t)\,dG(y) = 0$$

and Φ, ϕ the cumulative and density functions, respectively, of the standard normal distribution.

Hence, if $r_1 = z_0\sqrt{K''(z_0,t)}$, following the same computations than in García-Pérez (García-Pérez 2003, pp. 403–404) it will be

$$\text{TAIF}\,(x;t;T_n,G) = \frac{\phi(s)}{r_1}\,n^{1/2}\left(\frac{e^{z_0\psi(x,t)}}{\int e^{z_0\psi(y,t)}dG(y)} - 1\right) + O(n^{-1/2}).$$

Hence, from (5), the *VOM+SAD approximation* will be

$$P_F\{T_n > t\} \simeq P_G\{T_n > t\} + \int \frac{\phi(s)}{r_1}\,n^{1/2}\left(\frac{e^{z_0\psi(x,t)}}{\int e^{z_0\psi(y,t)}dG(y)} - 1\right)dF(x). \quad (7)$$

If $Z_i \equiv (1-\epsilon)N(\mu,\sigma)+\epsilon N(\mu,g\sigma)$, i.e., $F = (1-\epsilon)\,2\,\gamma(\mathbf{h})\,\chi_1^2+\epsilon\,g^2\,2\,\gamma(\mathbf{h})\,\chi_1^2$, from expression (6), the VOM+SAD approximation will be

$$P_F\{T_n > t\} \simeq P_G\{T_n > t\} + \epsilon\,\frac{\phi(s)}{r_1}\,\sqrt{n}\left(\frac{\int e^{z_0\psi(x,t)}dH(x)}{\int e^{z_0\psi(y,t)}dG(y)} - 1\right) \quad (8)$$

being $G = 2\gamma(\mathbf{h})\chi_1^2$ and $H = g^2 2\gamma(\mathbf{h})\chi_1^2$, and noting that a distribution $b\,\chi_1^2$ is a gamma distribution with parameters $(1/2,\,1/(2b))$.

Sometimes, the distributions F and G are not close enough. In these cases we can improve the approximation considering an iterative procedure as in García-Pérez (2011, 2012, 2016).

Remark 1 One of the keys of using the von Mises expansion is to choose, as pivotal distribution G, a manageable model (a normal, or a χ^2 in this case) for which we can easily compute the tail probability $P_G\{T_n > t\}$. In situations where we can not calculate this tail probability exactly, we can use the previous Lugannani and Rice formula to compute a saddlepoint approximation for the leading term, being the VOM+SAD approximation

$$P_F\{T_n > t\} \simeq 1 - \Phi(s) + \phi(s)\left[\frac{1}{r} - \frac{1}{s}\right] + \int \frac{\phi(s)}{r_1} n^{1/2} \left(\frac{e^{z_0\psi(x,t)}}{\int e^{z_0\psi(y,t)} dG(y)} - 1 \right) dF(x).$$

Remark 2 Alternatively, we could think of using directly a saddlepoint approximation of the tail probability under distribution F and not to use the von Mises expansion first. Namely, assuming that $Y_i \equiv F$, we could use directly the previous saddlepoint approximation given in Daniels (1983),

$$P_{Y_i \equiv F}\{T_n(Y_1, \ldots, Y_n) > t\} \simeq 1 - \Phi(s) + \phi(s)\left[\frac{1}{r} - \frac{1}{s}\right] \tag{9}$$

where all the elements are the same than in the previous section but now, with respect to model F.

Nevertheless, approximation (9) is not useful unless F is so manageable that it allows to compute the elements in the approximation.

For instance, in the simple case that $Z_i \equiv (1-\epsilon)N(\mu, \sigma) + \epsilon N(\mu, g\sigma)$, the distribution F in approximation (9) is $F = (1-\epsilon)\, 2\,\gamma(\mathbf{h})\, \chi_1^2 + \epsilon\, g^2\, 2\,\gamma(\mathbf{h})\, \chi_1^2$ and, if we look for a closed form approach, the fact that the model is a mixture of distributions, makes the saddlepoint equation unviable and, therefore, the saddlepoint approach. For this reason, we shall use the VOM+SAD approximation given by Eqs. (7), or (8) in the particular case of a contaminated Gaussian.

6 Approximation for the distribution of Matheron's estimator

Matheron's estimator is not robust (Cressie and Hawkins 1980 or Genton 1998). However, because its use is so widespread, we obtain in this section an VOM+SAD approximation to its distribution under contamination, studying so its (lack of) robustness.

If $N(\mathbf{h}) = \{(\mathbf{s}_i, \mathbf{s}_j) : \mathbf{s}_i - \mathbf{s}_j = \mathbf{h}\}$ and the sample size is $n = N_h$, the cardinality of $N(\mathbf{h})$, Matheron's estimator is the sample mean (hence, not robust) of the $Y_i = (Z_{i+h} - Z_i)^2$,

$$2\hat{\gamma}_M(\mathbf{h}) = \frac{1}{N_h}\sum_{i=1}^{N_h} Y_i = \overline{y}.$$

Thus, in the unrealistic case of no contamination, i.e., if $Z_i \equiv N(\mu, \sigma)$ and so, $Y_i \equiv 2\,\gamma(\mathbf{h})\,\chi_1^2$, the exact distribution of $2\hat{\gamma}_M(\mathbf{h})$ is the tail of a χ^2 distribution with N_h degrees of freedom,

$$P\{2\hat{\gamma}_M(\mathbf{h}) > t\} = P\left\{\chi^2_{N_h} > \frac{t\,N_h}{2\gamma(\mathbf{h})}\right\}.$$

Hence, using $G = 2\,\gamma(\mathbf{h})\,\chi_1^2$ as pivotal distribution, the von Mises approximation (5) will be

$$P_F\{2\hat{\gamma}_M(\mathbf{h}) > t\} \simeq P\left\{\chi^2_{N_h} > \frac{t\,N_h}{2\gamma(\mathbf{h})}\right\} + \int \mathrm{TAIF}\,(x; t; T_n, G)\; dF(x)$$

and therefore, the von Mises plus saddlepoint (VOM+SAD) approximation (7) for the distribution of Matheron's estimator under a model F will be

$$\begin{aligned}P_F\{2\hat{\gamma}_M(\mathbf{h}) > t\} &\simeq P\left\{\chi^2_{N_h} > \frac{t\,N_h}{2\gamma(\mathbf{h})}\right\}\\ &+ \int \frac{\phi(s)}{r_1}\,N_h^{1/2}\left(\frac{e^{z_0\psi(x,t)}}{\int e^{z_0\psi(y,t)}dG(y)} - 1\right) dF(x).\end{aligned}$$

In the particular case that $F = (1-\epsilon)\,2\,\gamma(\mathbf{h})\,\chi_1^2 + \epsilon\,g^2\,2\,\gamma(\mathbf{h})\,\chi_1^2$, the (VOM+SAD) approximation will be, from (8)

$$P_F\{2\hat{\gamma}_M(\mathbf{h}) > t\} \simeq P\left\{\chi^2_{N_h} > \frac{t\,N_h}{2\gamma(\mathbf{h})}\right\} + \epsilon\,\frac{\phi(s)}{r_1}\,\sqrt{N_h}\left(\frac{\int e^{z_0\psi(x,t)}dH(x)}{\int e^{z_0\psi(y,t)}dG(y)} - 1\right) \tag{10}$$

where G is a gamma distribution with parameters $(1/2,\,1/(4\gamma(\mathbf{h}))$, and H is a gamma distribution with parameters $(1/2,\,1/(4g^2\gamma(\mathbf{h}))$.

Working on approximation (10), z_0 is the solution of

$$K'(z_0, t) = \int_{-\infty}^{\infty} e^{z_0\psi(y,t)}\,\psi(y,t)\,dG(y) = 0$$

where $\psi(y,t) = y - t$. From this equation we obtain the saddlepoint

$$z_0 = \frac{1}{4\gamma(\mathbf{h})} - \frac{1}{2t}.$$

The rest of the elements of (10) are

$$\begin{aligned}K(z_0, t) &= -\frac{t}{4\gamma(\mathbf{h})} + \frac{1}{2} + \frac{1}{2}\log\left(\frac{t}{2\gamma(\mathbf{h})}\right)\\ K''(z_0, t) &= 2\,t^2\end{aligned}$$

Table 1 Tail Probabilities for several values of t and sample size $N_h = 3$

t	Exact	Approximation
2.5	0.12193	0.1241979
3.0	0.07362	0.0750320
3.5	0.04361	0.0449431
4.0	0.02595	0.0267487
4.5	0.01510	0.0158439
5.0	0.00910	0.0093526

$$\begin{aligned} s &= \sqrt{N_h}\sqrt{\frac{t}{2\gamma(\mathbf{h})} - 1 - \log\left(\frac{t}{2\gamma(\mathbf{h})}\right)} \\ r &= \sqrt{N_h}\frac{1}{\sqrt{2}}\left(\frac{t}{2\gamma(\mathbf{h})} - 1\right) \\ \frac{\phi(s)}{r_1} &= \frac{\exp\left\{-\frac{N_h}{2}(t/(2\gamma(\mathbf{h})) - 1 - \log(t/(2\gamma(\mathbf{h}))))\right\}}{\sqrt{\pi}(t/(2\gamma(\mathbf{h})) - 1)} \\ \frac{\int e^{z_0\psi(x,t)} dH(x)}{\int e^{z_0\psi(y,t)} dG(y)} &= \frac{\sqrt{2\gamma(\mathbf{h})}}{\sqrt{t - tg^2 + 2g^2\gamma(\mathbf{h})}} \end{aligned}$$

Then, VOM+SAD approximation (10) is

$$\begin{aligned} P_F\{2\hat{\gamma}_M(\mathbf{h}) > t\} \simeq P&\left\{\chi^2_{N_h} > \frac{t\,N_h}{2\gamma(\mathbf{h})}\right\} \\ &+\epsilon\sqrt{N_h}\,\frac{2\gamma(\mathbf{h})}{\sqrt{\pi}(t - 2\gamma(\mathbf{h}))}\,\exp\left\{-\frac{N_h}{2}\left(\frac{t}{2\gamma(\mathbf{h})} - 1 - \log\frac{t}{2\gamma(\mathbf{h})}\right)\right\} \\ &\cdot\left(\frac{\sqrt{2\gamma(\mathbf{h})}}{\sqrt{t - tg^2 + 2g^2\gamma(\mathbf{h})}} - 1\right) \end{aligned} \qquad (11)$$

Let us observe that, if $\epsilon = 0$ or $g = 1$, the sum of the right hand side of this approximation is zero.

We see the quality of approximation (11) in Table 1 for several values of t, considering a sample size as small as $N_h = 3$, $g = 1.1$ (10% of contamination in scale), $2\gamma(h) = 1.3$, $\epsilon = 0.01$ and where the *exact* values are obtained with a simulation of 100,000 samples.

For other larger values of ϵ or g (i.e., for more distant distributions) we shall obtain accurate approximations with the iterative procedure previously mentioned.

In Fig. 1 we plot this approximation as the dotted line, being the solid line the *exact* distribution.

After we have seen that the VOM+SAD approximation is accurate, we can used it to represent the tail probability of Matheron's estimator with no contamination ($\epsilon = 0$) and with several degrees of contamination: $\epsilon = 0.01$, $\epsilon = 0.05$, $\epsilon = 0.1$ and $\epsilon = 0.2$. They are plotted in Fig. 2. We can see in this figure that, as we increase the

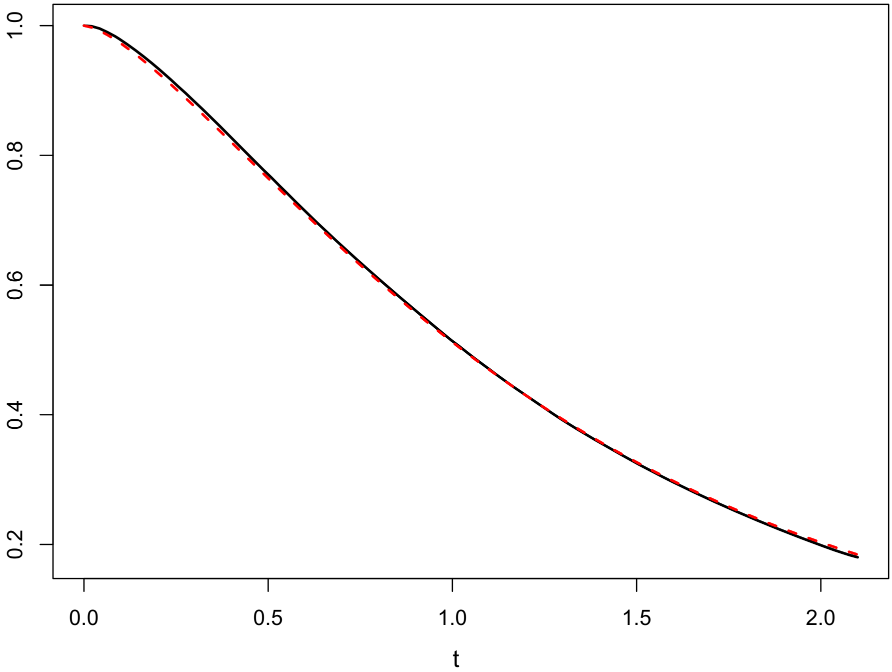

Fig. 1 *Exact* and approximate tail probabilities for Matheron's estimator with $N_h = 3$

contamination percentage, i.e., as we increase the value of ϵ, the p values and critical values are greatly affected.

All the computations and R functions used in the paper are on the website https://www2.uned.es/pea-metodos-estadisticos-aplicados/variogram.htm as Supplementary Material.

As we mentioned in Remark 2, the direct saddlepoint approximation is unviable even in the simple case of being $F = (1-\epsilon)\,2\,\gamma(\mathbf{h})\,\chi_1^2 + \epsilon\, g^2\, 2\,\gamma(\mathbf{h})\,\chi_1^2$, because it would be

$$K(\lambda, t) = \log\left[(1-\epsilon)\frac{e^{-\lambda t}}{\sqrt{1-\lambda 4\gamma(\mathbf{h})}} + \epsilon\frac{e^{-\lambda t}}{\sqrt{1-\lambda 4 g^2\gamma(\mathbf{h})}}\right]$$

and this mixture does not allow to obtain a closed form expression for the saddlepoint z_0, and so, for the sadlepoint approximation.

7 Approximation for the distribution of the α-trimmed variogram estimator

In addition to M-estimates, another usual robust estimator is the sample α-trimmed variogram, defined again for the transformed N_h variables Y_i as follows:

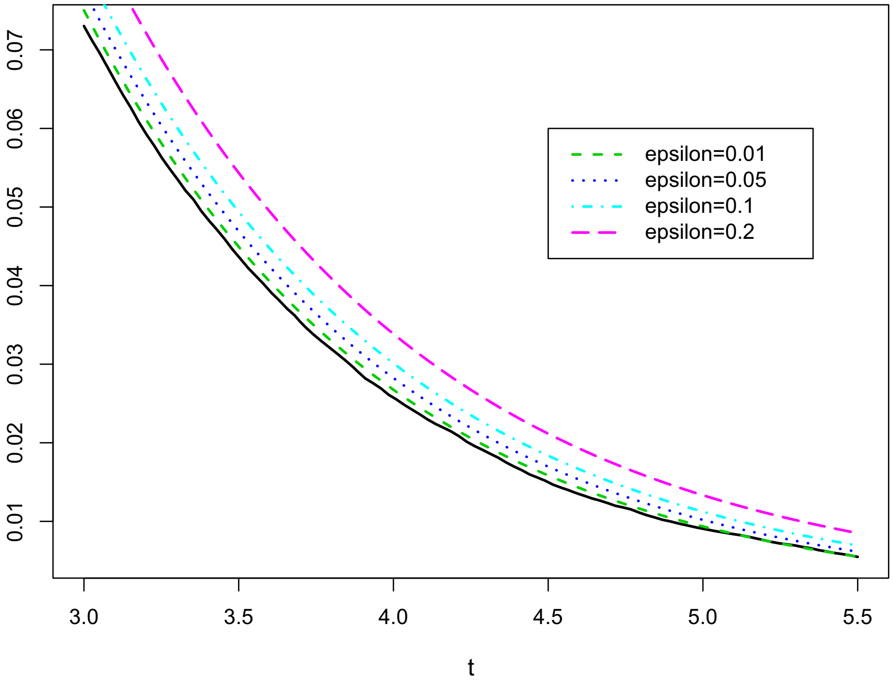

Fig. 2 *Exact* and approximate tail probabilities of Matheron's estimator with contamination $\epsilon = 0.01$, $\epsilon = 0.05$, $\epsilon = 0.1$ and $\epsilon = 0.2$ and sample size $N_h = 3$

If we trim the $100 \cdot \alpha\%$ of the smallest and the $100 \cdot \alpha\%$ of the largest ordered data $Y_{(i)}$, the (symmetrically) sample α*-trimmed variogram* is defined as

$$2\hat{\gamma}_\alpha(\mathbf{h}) = \frac{1}{N_h - 2r}\left(Y_{(r+1)} + \cdots + Y_{(N_h - r)}\right) = \overline{Y}_\alpha$$

where $r = [N_h\alpha]$ if [.] stands for the integer part.

An important remark is that the sample α-trimmed variogram estimates the population α-trimmed variogram, which will usually be different from the population variogram $2\,\gamma(\mathbf{h})$, the population mean of the Y_i, because these observations follow, in the normal case, the highly skewed χ^2 distribution.

Nevertheless because $2\hat{\gamma}_\alpha(\mathbf{h})$ will be more robust than Matheron's estimator, we recommend its use.

In García-Pérez (2016) an accurate VOM+SAD approximation is obtained for the sample α-trimmed mean. From Corollary 1 therein, we can approximate the small sample distribution of the sample α-trimmed variogram $2\hat{\gamma}_\alpha(\mathbf{h})$ when the observations Y_i come from F, with k iterations (k large), by the distribution of $2\hat{\gamma}_M(\mathbf{h})$ when the observations come from F as

$$P_F\left\{2\hat{\gamma}_\alpha(\mathbf{h}) > t\right\} \simeq (1 + N_h\, c_1)^{k+1}\,(1 + N_h\, c_2)^{k+1}\, P_F\left\{2\hat{\gamma}_M(\mathbf{h}) > t\right\}$$

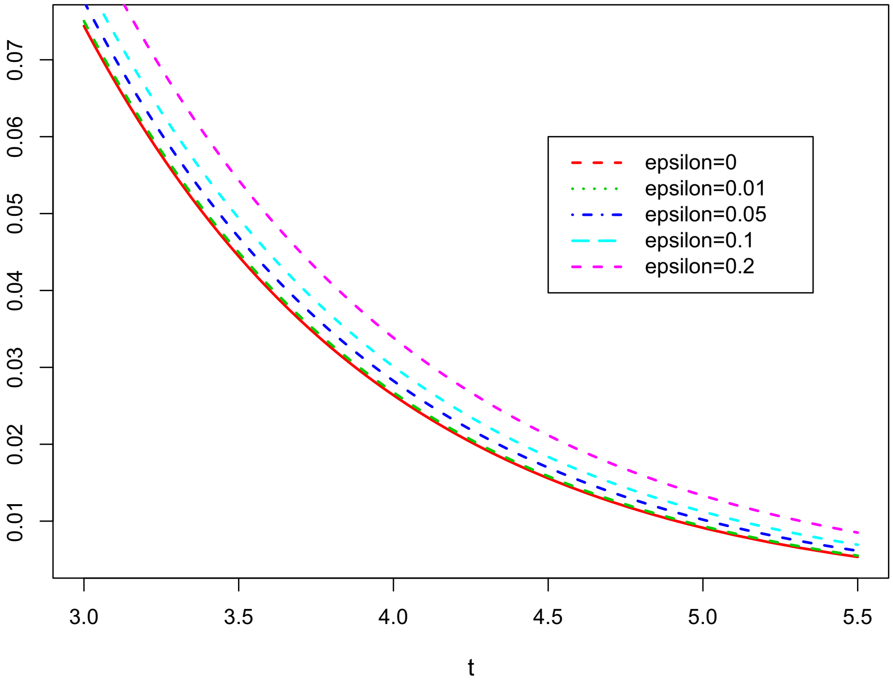

Fig. 3 Tail probabilities of the 0.1-trimmed variogram $2\hat{\gamma}_\alpha(\mathbf{h})$ estimator with no contamination, $\epsilon = 0$, and contamination $\epsilon = 0.01$, $\epsilon = 0.05$, $\epsilon = 0.1$ and $\epsilon = 0.2$

where $c_1 = \left[(1-2\alpha)^{1/(k+1)} - 1\right]$ and $c_2 = \left[1/(1-2\alpha)^{1/(k+1)} - 1\right]$.

Hence, because of the previous approximation (11) for the small sample distribution of Matheron's estimator, the VOM+SAD approximation for the tail probability of the sample α-trimmed variogram $2\hat{\gamma}_\alpha(\mathbf{h})$, under a scale contaminated normal model, is

$$\begin{aligned} P_F\{2\hat{\gamma}_\alpha(\mathbf{h}) > t\} \simeq (1+N_h\, c_1)^{k+1}\,(1+N_h\, c_2)^{k+1} \Bigg[P\left\{\chi^2_{N_h} > \frac{t\,N_h}{2\gamma(h)}\right\} \\ +\epsilon\sqrt{N_h}\,\frac{2\gamma(\mathbf{h})}{\sqrt{\pi}\,(t-2\gamma(\mathbf{h}))}\,\exp\left\{-\frac{N_h}{2}\left(\frac{t}{2\gamma(\mathbf{h})} - 1 - \log\frac{t}{2\gamma(\mathbf{h})}\right)\right\} \\ \cdot\left(\frac{\sqrt{2\gamma(\mathbf{h})}}{\sqrt{t - tg^2 + 2g^2\gamma(\mathbf{h})}} - 1\right)\Bigg] \end{aligned} \qquad (12)$$

In Fig. 3 we plot the tail probability of the 0.1-trimmed variogram estimator $2\hat{\gamma}_\alpha(\mathbf{h})$ with no contamination ($\epsilon = 0$), and with several degrees of contamination: $\epsilon = 0.01$, $\epsilon = 0.05$, $\epsilon = 0.1$ and $\epsilon = 0.2$, being a sample size as small as $N_h = 3$.

We observe from this figure that, as we increase the contamination percentage, i.e., as we increase the value of ϵ, the p-values and critical values are affected but less than with Matheron's estimator, if we compare the differences between the curves in this figure and in Fig. 2.

8 Approximation for the distribution of the Huber's variogram estimator

In the case that the ψ function, $\psi(x,t) = \psi(x-t)$, used to obtain the M-estimator in Eq. (3), is the Huber's function $\psi_b(u) = \min\{b, \max\{u, -b\}\}$, the M-estimator obtained is called the *Huber's variogram estimator* $2\hat{\gamma}_H(\mathbf{h})$ and an approximation for its distribution can be obtained from Eqs. (7), or (8) in the particular case of a contaminated Gaussian distribution.

But the idea of these approximations is to use, as pivotal distribution G, a model for which the leading term $P_G\left\{2\hat{\gamma}_H(\mathbf{h}) > t\right\}$ is easily computable. Nevertheless, this is not the case with a scale contaminated normal model where $G = 2\gamma(\mathbf{h})\chi_1^2$. For this reason, in this case, as it was pointed out in Remark 1, we shall use the Lugannani and Rice formula for this leading term, being the VOM+SAD approximation for the distribution of the Huber's variogram estimator

$$\begin{aligned} P_{Y_i \equiv F}\left\{2\hat{\gamma}_H(\mathbf{h}) > t\right\} &\simeq 1 - \Phi(s) + \phi(s)\left[\frac{1}{r} - \frac{1}{s}\right] \\ &+\epsilon\,\frac{\phi(s)}{r_1}\,\sqrt{n}\left(\frac{\int e^{z_0\psi_b(x-t)}dH(x)}{\int e^{z_0\psi_b(y-t)}dG(y)} - 1\right) \end{aligned} \tag{13}$$

where the saddlepoint z_0 is such that

$$\int e^{z_0\psi_b(y-t)}\,\psi_b(y-t)\,dG(y) = 0$$

being $G = 2\gamma(\mathbf{h})\chi_1^2$ and $H = g^2 2\gamma(\mathbf{h})\chi_1^2$, and where all the functionals r, r_1 and s are computed with respect to model G.

This approximation may seem complicated but it is easy to compute using R software (R Development Core Team 2018).

9 Linearized version of the variogram model

Representing by ra the range of a variogram, we saw in Sect. 2 that if we accept a variogram model such that

$$\gamma(2\mathbf{h}) = 2\gamma(\mathbf{h})$$

$||\mathbf{h}|| < ra$, the observations Y_i, used in the variogram estimation for this $\mathbf{h}$, can be considered independent.

This condition is satisfied by a straight line. Hence, if for a particular $\mathbf{h}$, a *linear semivariogram* can be accepted, the independence of the Y_i used in its estimation, can also be accepted.

For instance, let us observe that the spherical model, the most used one, is such that

$$\gamma(\mathbf{h}) = c||\mathbf{h}|| + O(ra^{-3})$$

for a constant c. Hence, the previous condition is practically satisfied, if the range is large.

Using the classical Matheron's estimator $\hat{\gamma}_M(\mathbf{h})$, next we formally define a *Linearized Semivariogram Model* $\mathcal{M}_L(\mathbf{h})$. If this can be accepted, the independence condition will be satisfied.

We first divide the variogram estimations into two subsets: before the range and after the range. If we represent the parameters of the variogram as nugget nu, sill si(=partial sill+nugget) and range ra, the definition of Linearized Semivariogram Model is based on the regression line through the origin $y = slo\,||\mathbf{h}||$ for the pairs of data $(||\mathbf{h}||, \hat{\gamma}_M(\mathbf{h}) - nu)$, where $||\mathbf{h}|| < ra$.

If

$$y = nu + slo\,||\mathbf{h}||$$

is this linearized part and we intersect it with the horizontal line equal to si, the intersection point is $((si - nu)/slo, si)$.

Finally, in order that the linearized semivariogram model $\mathcal{M}_L(\mathbf{h})$ is continuos and that it represents in a better way a semivariogram model $\mathcal{M}(\mathbf{h})$, $\mathcal{M}_L(\mathbf{h})$ is defined as

$$\mathcal{M}_L(\mathbf{h}) = \begin{cases} nu + slo\,||\mathbf{h}||, & \text{if } ||\mathbf{h}|| \leq (si - nu)/slo \\ si, & \text{if } ||\mathbf{h}|| > (si - nu)/slo. \end{cases}$$

Let us observe that $\mathcal{M}(\mathbf{h})$ (and $\mathcal{M}_L(\mathbf{h})$) are models for the semivariogram $\gamma(\mathbf{h})$, not for the variogram $2\gamma(\mathbf{h})$.

9.1 Robust linearized version of the variogram model

If we use a robust estimate $\hat{\gamma}_R(\mathbf{h})$ of the variogram instead of the Matheron's estimator, the robust linearized version of the variogram model is very similar to the classical one. We divide the variogram again into two parts: before the range ra and after the range, being the range the same obtained in the classical estimation.

Before the range we compute the regression line through the origin, $y = b\,||\mathbf{h}||$ for the pairs of data $(||\mathbf{h}||, \hat{\gamma}_R(\mathbf{h}) - nu)$, where $||\mathbf{h}|| < ra$, being the nugget the same obtained in the classical estimation.

After the range, we define it as the constant line equal to the sample mean $\overline{x}_R$, let us say, of the values $\hat{\gamma}_R(\mathbf{h})$, varying $||\mathbf{h}||$.

The robust linearized version of the variogram model is then defined as

$$\mathcal{RM}_L(\mathbf{h}) = \begin{cases} nu + b\,||\mathbf{h}||, & \text{if } ||\mathbf{h}|| \leq (\overline{x}_R - nu)/b \\ \overline{x}_R, & \text{if } ||\mathbf{h}|| > (\overline{x}_R - nu)/b. \end{cases}$$

10 Inferences with the distribution of the variogram estimator

With the approximations for the distribution of the variogram estimators $\hat{\gamma}(\mathbf{h})$ previously obtained, we can check, for each $\mathbf{h}$, if the value $\gamma(\mathbf{h})$ given by a particular variogram model $\mathcal{M}(\mathbf{h})$ (spherical, exponential, etc.) can be accepted.

More specifically, for each $\mathbf{h}$, assuming it is true the null hypothesis $H_0 : \gamma(\mathbf{h}) = \gamma(\mathbf{h})_0$ given by a variogram model F_0, we can compute a $1-\beta$ confidence interval for $\gamma(\mathbf{h})$, based on $\hat{\gamma}(\mathbf{h})$

$$P_{F_0}\{b_1(\mathbf{h}) < \hat{\gamma}(\mathbf{h}) < b_2(\mathbf{h})\} = 1-\beta$$

being $P_{F_0}\{\hat{\gamma}(\mathbf{h}) < b_1(\mathbf{h})\} = \beta/2$ and $P_{F_0}\{\hat{\gamma}(\mathbf{h}) > b_2(\mathbf{h})\} = \beta/2$. We accept the value $\gamma(\mathbf{h})_0$, given by the semivariogram model, if it is inside this confidence interval.

We could choose β in order to be a significance level for multiple tests, varying $\mathbf{h}$, but the global test, discussed below, is more appropriate for this task.

Anyway, these pairs of points $(b_1(\mathbf{h}), b_2(\mathbf{h}))$, varying $\mathbf{h}$, establish a confidence zone (not a formal confidence region), within which the values of the semivariogram estimator should be. If this is the case, we shall accept the semivariogram model, although the test defined below in Sect. 10.1 is more appropriate.

If the linearized semivariogram model $\mathcal{M}_L(\mathbf{h})$ is inside the confidence zone, we should accept it as model and hence, condition (2), accepting so the independence of the transformed variables Y_i used in the estimator $2\hat{\gamma}_M(\mathbf{h})$.

Besides this, the fact that $\hat{\gamma}(\mathbf{h})$ is outside of a confidence interval associated with a value $\mathbf{h}$, is an indication of the possible presence of spatial outliers at a distance $||\mathbf{h}||$.

10.1 Global test for the variogram model

The confidence zone established in the last section can not be used as a formal test. On the other hand, because of the visual appeal of plotting, we can think in fitting a linear or non-linear regression model to the pairs $(||\mathbf{h}||, \hat{\gamma}(\mathbf{h}))$ and to check a particular variogram model with the usual regression tests. Nevertheless these ones are based on the normality of the observations or in a large sample size.

If the underlying model F is not normal and the sample size not very large, these tests can not be used. Another possibility is to use robust techniques, such as the robust coefficient of determination (Renaud and Victoria-Fesser 2010), the robust Akaike AIC criterion (Ronchetti 1997), the robust Mallows's C_p (Ronchetti and Staudte 1994) or the robust cross-validation (Ronchetti et al. 1997) because these methods do not need the normality assumption.

Nevertheless here we are going to use the previous saddlepoint approximation to test the null hypotheses of a particular variogram model $\mathcal{M}_0(\mathbf{h})$.

Although other more robust test statistics can be used, by analogy with the statistic of the Kolmogorov–Smirnov goodness of fit test, we base our test on

$$S_n = \sup_{\mathbf{h}} \left\| 2\hat{\gamma}(\mathbf{h}) - 2\gamma(\mathbf{h}) \right\| = \max_{1\le ||\mathbf{h}|| \le K} \left\| 2\hat{\gamma}(\mathbf{h}) - 2\gamma(\mathbf{h}) \right\|$$

assuming there are K lags.

If we define the random variable $W(\mathbf{h}) = 2\hat{\gamma}(\mathbf{h}) - 2\gamma(\mathbf{h})$, its cumulative distribution function will be

$$H_{W(\mathbf{h})}(w) = 1 - P\{2\hat{\gamma}(\mathbf{h}) > w + 2\gamma(\mathbf{h})\}$$

where the tail probability can be computed with the previous VOM+SAD approximation.

Now, if $U(\mathbf{h}) = \|W(\mathbf{h})\|$, its cumulative distribution function will be

$$H_{U(\mathbf{h})}(u) = H_{W(\mathbf{h})}(u) - H_{W(\mathbf{h})}(-u).$$

Finally, the cumulative distribution function of S_n will be

$$\begin{aligned} F_{S_n}(v) &= \prod_{||\mathbf{h}||=1}^{K} H_{U(\mathbf{h})}(v) = \prod_{||\mathbf{h}||=1}^{K} \left[H_{W(\mathbf{h})}(v) - H_{W(\mathbf{h})}(-v)\right] \\ &= \prod_{||\mathbf{h}||=1}^{K} \left[P_{2\gamma(h)}\{2\hat{\gamma}(\mathbf{h}) > -v + 2\gamma(\mathbf{h})\} - P_{2\gamma(h)}\{2\hat{\gamma}(\mathbf{h}) > v + 2\gamma(\mathbf{h})\}\right] \end{aligned}$$

because, under the null hypothesis, we assume the independence between two increments of the same size $\mathbf{h}$: $Z_{t+h} - Z_t$ and $Z_{s+h} - Z_s$.

Hence, because we have locations equally spaced on a transect, we shall have also independence between two multiples of this size $Z_{t+h_1} - Z_t$ and $Z_{t+h_2} - Z_t$, i.e., between two $U(\mathbf{h_1})$ and $U(\mathbf{h_2})$.

Example 1 Let us consider Cadmium data from `meuse` data set of `sp` library. The usual variogram model $\mathcal{M}(\mathbf{h})$ admitted is Spherical with $nu = 0.5478482$, $si = 1.3397976 + 0.5478482 = 1.887646$ and $ra = 1149.439$ (see Supplementary Material).

We use the VOM+SAD approximation for the tail probability of the Matheron's estimator $2\hat{\gamma}_M(\mathbf{h})$ given in (11) considering a scale contaminated normal as underlying model, with $\epsilon = 0.01$ and $g = 1.1$.

To obtain the linearized spherical semivariogram model we observe that the fitted regression line is

$$y = nu + slo\,||\mathbf{h}|| = 0.547848 + 0.001392\,||\mathbf{h}||$$

and the interception point $((si - nu)/slo, si) = (962.4983, 1.887646)$. Thus, the linearized spherical semivariogram model $\mathcal{M}_L(\mathbf{h})$ will be

$$\mathcal{M}_L(\mathbf{h}) = \begin{cases} 0.547848 + 0.001392\,||\mathbf{h}||, & \text{if } ||\mathbf{h}|| \leq 962.4983 \\ 1.887646, & \text{if } ||\mathbf{h}|| > 962.4983 \end{cases}$$

For each $\mathbf{h}$ we compute the 95%-confidence limits $(b_1(\mathbf{h}), b_2(\mathbf{h}))$ with approximation (11) that are represented as dots in the center and right side of Fig. 4.

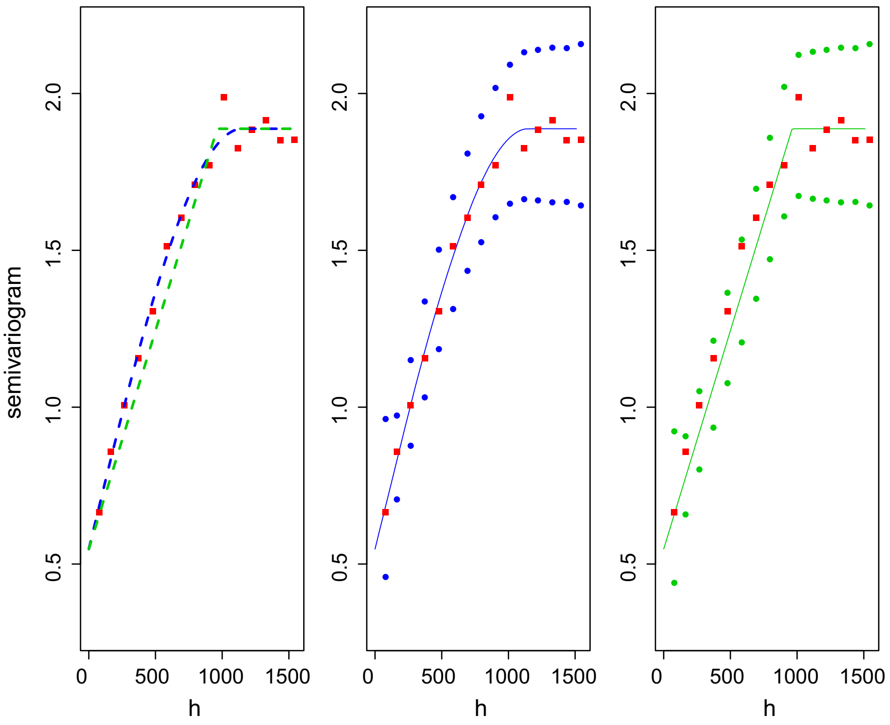

Fig. 4 Matheron's estimates $\hat{\gamma}_M(\mathbf{h})$, Semivariogram model $\mathcal{M}(\mathbf{h})$ and Linearized semivariogram model $\mathcal{M}_L(\mathbf{h})$ (left). Confidence zone created by the distribution of the Matheron's estimator and $\mathcal{M}(\mathbf{h})$ (middle). $\mathcal{M}_L(\mathbf{h})$ Confidence zone created by distribution of the Matheron's estimator and $\mathcal{M}_L(\mathbf{h})$ (right)

In the left-hand plot of this figure are shown the values of the Matheron's estimator $\hat{\gamma}_M(\mathbf{h})$ (red squares), the semivariogram model $\mathcal{M}(\mathbf{h})$ and the linearized semivariogram model $\mathcal{M}_L(\mathbf{h})$, both very close.

The middle plot shows the semivariogram model $\mathcal{M}(\mathbf{h})$ and the confidence zone, established by the 0.95%-confidence limits, represented by blue dots, limits obtained with the distribution of the Matheron's estimator, assuming that this model $\mathcal{M}(\mathbf{h})$ is true. We observe that the values of this estimator (red squares) are inside of the 0.95%-confidence intervals (blue dots), for all the **h** considered, i.e., inside the confidence zone; hence, the spherical model $\mathcal{M}(\mathbf{h})$ (blue curve) fits well.

Finally, the right-hand plot shows the linearized semivariogram model $\mathcal{M}_L(\mathbf{h})$ (green curve) and the 0.95%-confidence limits, represented by green dots, obtained with the distribution of the Matheron's estimator assuming that the linearized model $\mathcal{M}_L(\mathbf{h})$ is true. Since the values of this estimator (red squares) are inside of the 0.95%-confidence intervals (green dots), the linearized semivariogram model fits well.

In Table 2, these 0.95% confidence limits obtained with the model $\mathcal{M}(\mathbf{h})$ and the linearized model $\mathcal{M}_L(\mathbf{h})$ are given, plus the values obtained by the Matheron's estimator. We observe that these last ones are always inside the confidence limits, suggesting to accept both models and so, the independence of the transformed variables Y_i.

Table 2 0.95% confidence intervals for several values of $\mathbf{h}$

$\mathbf{h}$	0.95% conf. lim. with $\mathcal{M}(\mathbf{h})$	0.95% conf. lim. with $\mathcal{M}_L(\mathbf{h})$	$\hat{\gamma}_M(\mathbf{h})$
79.29244	[0.45875, 0.96232]	[0.44000, 0.92300]	0.6650872
163.97367	[0.70596, 0.97329]	[0.65806, 0.90725]	0.8584648
267.36483	[0.87699, 1.15010]	[0.80134, 1.05089]	1.0064382
372.73542	[1.03125, 1.33686]	[0.93484, 1.21188]	1.1567136
478.47670	[1.18495, 1.50212]	[1.07656, 1.36472]	1.3064732
585.34058	[1.31289, 1.66948]	[1.20650, 1.53420]	1.5135658
693.14526	[1.43469, 1.80844]	[1.34560, 1.69615]	1.6040086
796.18365	[1.52592, 1.92738]	[1.47160, 1.85876]	1.7096998
903.14650	[1.60544, 2.01765]	[1.60816, 2.02108]	1.7706890
1011.29177	[1.64862, 2.09173]	[1.67333, 2.12306]	1.9875659
1117.86235	[1.66310, 2.13141]	[1.66441, 2.13310]	1.8259154
1221.32810	[1.65917, 2.13905]	[1.65917, 2.13905]	1.8852099
1329.16407	[1.65305, 2.14600]	[1.65305, 2.14600]	1.9145967
1437.25620	[1.65430, 2.14457]	[1.65430, 2.14457]	1.8505336
1543.20248	[1.64300, 2.15748]	[1.64300, 2.15747]	1.8523791

Finally, the formal global test for checking the linearized semivariogram model gives a value for the test statistic of $S_n = 0.3018476$ and for the p-value

$$P\{S_n > 0.3018476\} = 1 - F_{S_n}(0.3018476) = 0.9011587$$

accepting the linearized semivariogram model and so, the independence of the transformed observations Y_i used in $\hat{\gamma}_M(\mathbf{h})$.

Example 2 Let us consider now log Calcium data (mg/l), one of the eight variables observed in the groundwater data analysis around the city of Madrakah, a town located in the Wadi Usfan region in western Saudi Arabia, (Marko et al. 2014).

We represent in Fig. 5 the values of the Matheron's variogram estimator, 0.05-trimmed variogram estimator and Huber's variogram estimator (with tuning constant $b = 1$) using theses data, with a lag constant $h = 0.002$ (see Supplementary Material). We observe an outlier at lag $h = 0.006$.

The linearized versions of the variogram models are added to the previous plot in Fig. 6.

The test statistics of the global test, for checking the three linearized semivariogram models, are $S_n = 0.1221990$ (classical); $S_n = 0.165044411$ (0.05-trimmed), and $S_n = 0.0711932514$ (Huber).

Besides this, the p-values, computed with the VOM+SAD approximations (11), (12) and (13) respectively, and considering, as underlying model, a scale contaminated normal with $\epsilon = 0.01$ and $g = 1.1$ are: 0.208566, 0.0273871 and 0.9999, accepting the linearized semivariogram model and so, the independence of the transformed observations Y_i.

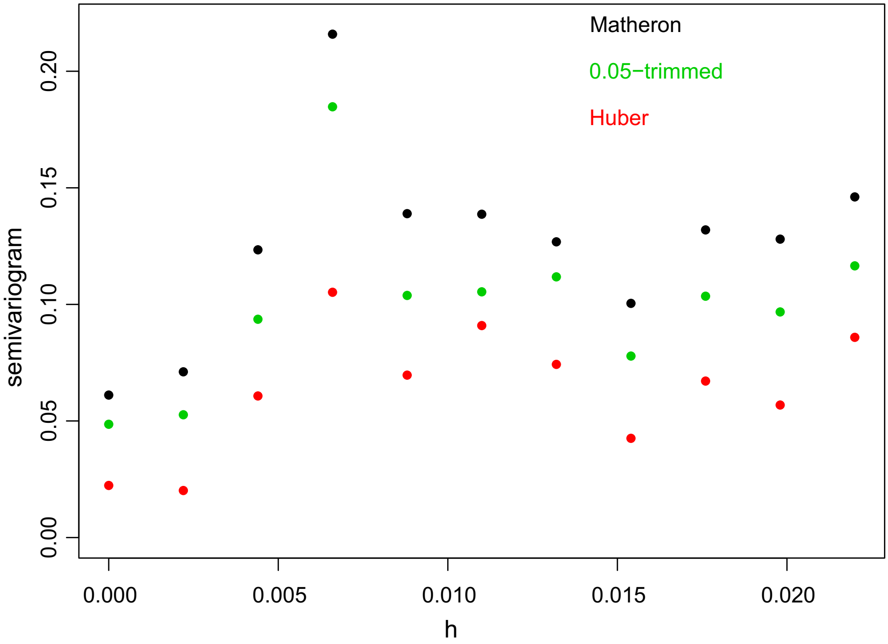

Fig. 5 Matheron's estimates $\hat{\gamma}_M(\mathbf{h})$, 0.05-trimmed estimates $\hat{\gamma}_\alpha(\mathbf{h})$, Huber's estimates $\hat{\gamma}_H(\mathbf{h})$

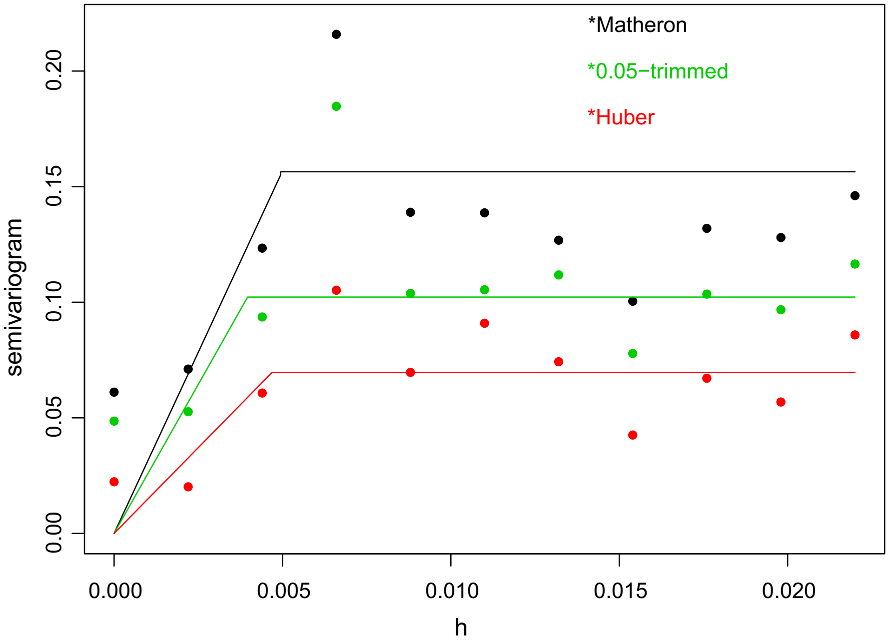

Fig. 6 Estimates and Linearized versions of the variogram model

11 Conclusions

Small sample distributions for classical and robust variogram estimators are obtained, approximations that are based on the von Mises expansion plus a saddlepoint approximation. It is proven that these are very accurate, even for small sample sizes.

The approximations are computed, considering a scale contaminated normal model for the spatial observations Z_i, approximations from which we obtain closed form expressions that allow an easy robust interpretation of the variogram estimators.

With these approximations it is possible to test a particular variogram model, which can be used in kriging interpolation.

Finally, it is also possible to test the independence of the transformed variables $Y_i = (Z_{i+h} - Z_i)^2$, opening so the possibility of using, in Spatial Statistics, the usual statistical methods for an independent sample of Y_i.

Acknowledgements The author is very grateful to the referee and the Editor in Chief for their careful reading and remarks which have improved the paper. This work is partially supported by Grant MTM 2015-67057-P from Ministerio de Economía, Industria y Competitividad (Spain).

References

Ali MM (1987) Durbin–Watson and generalized Durbin–Watson tests for autocorrelations and randomness. J Bus Econ Stat 5(2):195–203. https://doi.org/10.2307/1391900

Cressie NAC (1993) Statistics for spatial data. Wiley, New York

Cressie N, Hawkins DM (1980) Robust estimation of the variogram:I. Math Geol 12(2):115–125. https://doi.org/10.1007/BF01035243

Daniels HE (1983) Saddlepoint approximations for estimating equations. Biometrika 70(1):89–96. https://doi.org/10.1093/biomet/70.1.89

Field CA, Ronchetti E (1985) A tail area influence function and its application to testing. Seq Anal 4(1–2):19–41. https://doi.org/10.1080/07474948508836070

García-Pérez A (2003) Von Mises approximation of the critical value of a test. Test 12(2):385–411. https://doi.org/10.1007/BF02595721

García-Pérez A (2011) Another look at the tail area influence function. Metrika 73(1):77–92. https://doi.org/10.1007/s00184-009-0266-z

García-Pérez A (2012) A linear approximation to the power function of a test. Metrika 75(7):855–875. https://doi.org/10.1007/s00184-011-0356-6

García-Pérez A (2016) A von Mises approximation to the small sample distribution of the trimmed mean. Metrika 79(4):369–388. https://doi.org/10.1007/s00184-015-0559-3

Genton MG (1998) Highly robust variogram estimation. Math Geol 30(2):213–221. https://doi.org/10.1023/A:1021728614555

Hampel FR (1974) The influence curve and its role in robust estimation. J Am Stat Assoc 69:383–393. https://doi.org/10.1080/01621459.1974.10482962

Huber PJ, Ronchetti EM (2009) Robust statistics, 2nd edn. Wiley, New York

Jensen JL (1995) Saddlepoint approximations. Clarendon Press, Oxford

Lugannani R, Rice S (1980) Saddle point approximation for the distribution of the sum of independent random variables. Adv Appl Probab 12:475–490. https://doi.org/10.2307/1426607

Marko K, Al-Amri NS, Elfeki AMM (2014) Geostatistical analysis using GIS for mapping groundwater quality: case study in the recharge area of Wadi Usfan, western Saudi Arabia. Arab J Geosci 7:5239–5252. https://doi.org/10.1007/s12517-013-1156-2

Matheron G (1962) Traité de géostatistique appliquée, Tome I: Mémoires du Bureau de Recherches Géologiques et Minières, 14th edn. Technip, Paris

R Development Core Team (2018) A language and environment for statistical computing. R foundation for statistical computing, Viena, Austria. http://www.R-project.org. Accessed 2018

Renaud O, Victoria-Fesser MP (2010) A robust coefficient of determination for regression. J Stat Plan Inference 140:1852–1862. https://doi.org/10.1016/j.jspi.2010.01.008

Ronchetti E (1997) Robusteness aspects of model choice. Stat Sin 7:327–338

Ronchetti E, Field C, Blanchard W (1997) Robust linear model selection by cross-validation. J Am Stat Assoc 92:1017–1023. https://doi.org/10.2307/2965566

Ronchetti E, Staudte RG (1994) A robust version of Mallows's C_p. J Am Stat Assoc 89:550–559. https://doi.org/10.2307/2290858

Serfling RJ (1980) Approximation theorems of mathematical statistics. Wiley, New York

Withers CS (1983) Expansions for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparametric confidence intervals. Ann Stat 11(2):577–587. https://doi.org/10.1214/aos/1176346163