Statistics and Probability Letters 110 (2016) 191–197

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Asymptotic near-efficiency of the "Gibbs-energy and empirical-variance" estimating functions for fitting Matérn models — I: Densely sampled processes

Didier A. Girard

CNRS and Grenoble-Alpes University, Lab. LJK, F-38000, Grenoble, France

ARTICLE INFO

Article history:
Received 31 October 2015
Received in revised form 11 December 2015
Accepted 17 December 2015
Available online 24 December 2015

Keywords:
ARMA
Gaussian process
Maximum likelihood
Estimating functions
Matérn autocorrelation

ABSTRACT

Consider one realization of a continuous-time Gaussian process Z which belongs to the Matérn family with known regularity index $\nu > 0$. For estimating the autocorrelation-range and the variance of Z from n observations on a fine grid, we propose two simple estimating functions based on the "candidate Gibbs energy" (GE) and the empirical variance (EV). Here a candidate GE designates the quadratic form $\mathbf{z}^T R^{-1}\mathbf{z}/n$ where $\mathbf{z}$ is the vector of observations and R is the autocorrelation matrix for $\mathbf{z}$ associated with a candidate range. We show that the ratio of the large-n mean squared error of the resulting GE–EV estimate of the range-parameter to the one of its maximum likelihood estimate, and the analog ratio for the variance-parameter, both converge, when the grid-step tends to 0, toward a constant, only function of ν, surprisingly close to 1 provided ν is not too large. This latter condition on ν has not to be imposed to obtain the convergence to 1 of the analog ratio for the microergodic-parameter. Possible extensions of this approach, which could be rather easily implemented, are briefly discussed.

1. Introduction

We consider time-series of length n obtained by observing, at n equispaced times, a continuous-time process Z which is Gaussian, has mean zero and an autocorrelation function which belongs to the Matérn family with regularity index $\nu > 0$. This family is commonly used, for instance in geostatistics or for turbulence models; see e.g. Stein (1999), Guttorp and Gneiting (2006), Gaetan and Guyon (2010). Recall that $\nu = 1/2$ corresponds to the well known exponential autocorrelation (in other words, Z is a stationary Ornstein–Uhlenbeck (OU) process). The definition of Matérn processes on $\mathbb{R}$ can be easily formulated in terms of the Fourier transform of their autocorrelation function, namely their spectral density over $(-\infty, +\infty)$:

$$f^*_{\nu,b,\theta}(\omega) = b\, g^*_{\nu,\theta}(\omega), \quad \text{with } g^*_{\nu,\theta}(\omega) := \frac{C_\nu\, \theta^{2\nu}}{(\theta^2+\omega^2)^{\nu+\frac{1}{2}}} \quad \text{where } C_\nu = \frac{\Gamma\left(\nu+\frac{1}{2}\right)}{\sqrt{\pi}\,\Gamma(\nu)}. \tag{1.1}$$

In this paper the constant C_ν (where $\Gamma(\cdot)$ is the classical gamma function, see e.g. Weisstein, 1999) is chosen so that $\int_{-\infty}^{\infty} g^*_{\nu,\theta}(\omega)d\omega = 1$. Thus b is the variance of $Z(t)$ and θ is the so-called "inverse-range parameter" (in fact, it is $\nu^{1/2}/\theta$ which can be interpreted as an effective range or "correlation length" independently of ν, cf. Stein, 1999, Section 2.10); we will often drop the term "inverse".

E-mail address: didier.girard@imag.fr.

http://dx.doi.org/10.1016/j.spl.2015.12.021
0167-7152/

We are mainly concerned here with dense grid for the observation "locations" (or "times") in the sense that the distance $\delta > 0$ between two successive locations is small relatively to $1/\theta$. The considered processes being mean square continuous, this means that two successive observations are strongly correlated. Stein (1999, Chapter 3) shows that a standard (i.e. fixed $\delta > 0$) large-n asymptotic analysis followed by a less standard small-δ analysis yields useful insights and good approximations for various real (finite-size) problems. We shall also use such an increasing domain framework with an "infill component".

In this article, we assume that ν is known (see Wu et al., 2013, for a recent work on the estimation of ν) and we consider the problem of estimating the variance and the range parameters (denoted respectively by $b_0 > 0$ and $\theta_0 > 0$) from a perfect sampling of Z. Notice that the case of measurement errors, called "nugget effect" in geostatistics, is developed in a companion paper (Girard, 2015c). More precisely, given a known $\delta > 0$, one observes

$$\mathbf{z} := (Z(\delta), Z(2\delta), \ldots, Z(n\delta))^T \sim N(0, b_0 R_{\theta_0}), \tag{1.2}$$

where R_θ is the Toeplitz matrix of coefficients $[R_\theta]_{j,k} = K_{\nu,\theta}(\delta|j-k|),\ j,k = 1,\ldots,n$, with $K_{\nu,\theta}(t) = \int_{-\infty}^{\infty} g^*_{\nu,\theta}(\omega) e^{i\omega t} d\omega$. Expressions for these autocorrelation functions $K_{\nu,\theta}(\cdot)$ are now classical in terms of exponential or Bessel functions, depending on ν, see e.g. Stein (1999, Section 2.5).

As is well known, implementing the maximum likelihood (ML) method is simplified after noticing that, if one constraints θ to a fixed, >0, value, then the maximizer b is simply (e.g. Zhang, 2004):

$$\hat{b}_{\mathrm{ML}}(\theta) = (1/n)\mathbf{z}^T {R_\theta}^{-1}\mathbf{z}. \tag{1.3}$$

Thus, substituting to b this expression in the likelihood, one only has to maximize a one-parameter criterion (the so-called concentrated likelihood) of θ. Furthermore failures of the classical two-parameters Fisher-scoring algorithm may occur when one does not use this profiling of the likelihood, especially in our Matérn context (Zhang, 2004). However even the numerical one-dimensional maximization of this concentrated likelihood may often require a global grid search and thus requires computing a (sometimes very) large number of $n \times n$ matrix determinants in addition to solving the associated $n \times n$ linear systems. Numerical aspects will not be further discussed here.

Zhang and Zimmerman (2007) proposed to use the classical weighted least square method (not statistically fully efficient but whose implementation is often much easier than maximum likelihood since it does not require computing $n \times n$ determinant-terms) to estimate the range parameter θ, next, to chose $\hat{b}_{\mathrm{ML}}(\theta)$ as estimate for b_0 (thus only one linear $n \times n$ system has to be solved). The idea underlying this method is that, in the infill asymptotic framework (i.e. $\delta = 1/n$ and n large), even if θ is fixed at a wrong value θ_1, the product $\hat{b}_{\mathrm{ML}}(\theta_1)\theta_1^{2\nu}$ still remains an *efficient* estimator of $c_0 := b_0\theta_0^{2\nu}$, which is called the "microergodic parameter" of the Matérn model (1.1) (see Du et al., 2009; Wang and Loh, 2011; Kaufman and Shaby, 2013).

The method we study here (proposed in the first arXiv version of Girard, 2015c), firstly reverses the above roles of variance and range, in that it is based on a very simple nonparametric estimate for the variance, namely the empirical variance:

$$\hat{b}_{\mathrm{EV}} := n^{-1}\mathbf{z}^T\mathbf{z}. \tag{1.4}$$

Secondly the maximization of the likelihood (w.r.t. θ) is replaced by the estimating equation method:

$$\text{solve in } \theta \text{ the equation } n^{-1}\mathbf{z}^T {R_\theta}^{-1}\mathbf{z} = b, \tag{1.5}$$

with b fixed at $\hat{b}_{\mathrm{EV}}$. The quantity $(1/n)\mathbf{z}^T {R_\theta}^{-1}\mathbf{z}$ for a candidate θ may be called "Generalized global energy after candidate de-correlation" (or candidate "Gibbs energy", GE, in short) of the discretely sampled process. So our proposal is called the GE–EV estimation method and we denote by $\hat{\theta}_{\mathrm{GEV}}$ this range parameter estimate.

Let us give two heuristic justifications. First, since it is quite plausible that the idea underlying Zhang and Zimmerman (2007)'s proposal remains true for a random θ_1 (Kaufman and Shaby, 2013 make this argument rigorous), then, instead of "fixing" θ at θ_1, one may as well *adjust* θ so that $\hat{b}_{\mathrm{ML}}(\theta)$ coincides with a given value b_1 for the variance (i.e. solve (1.5) with b fixed at b_1); and, denoting $\hat{\theta}_1$ the so-obtained θ, the product $\hat{b}_{\mathrm{ML}}(\theta)\theta^{2\nu} = b_1\hat{\theta}_1^{2\nu}$ will plausibly be an efficient estimator of $b_0\theta_0^{2\nu}$. Second, for the case $\nu = 1/2$ and $b_0\theta_0 = c_0$ known, it has been shown (see Kessler, 1997) that $c_0/(n^{-1}\sum Z(\delta j)^2)$ is a successful estimator of θ_0, and thus $n^{-1}\sum Z(\delta j)^2$ is a successful estimator of b_0. The second point is an admittedly weak justification for selecting $\hat{b}_{\mathrm{EV}}$, beyond the case $\nu = 1/2$. The theoretical results we give in this article, will provide in our context a quite strong, and rather unexpected, justification of this GE–EV approach for ν not too large (which is very often the case in applications, see e.g. Stein, 1999; Gaetan and Guyon, 2010).

Our objective in this article is to provide first insights into the capability of the GE–EV method with the hope that the theoretical justification obtained here can be extended to more computationally complex settings. Indeed, this approach is not limited to observations on a one dimensional lattice, and is potentially not limited to regular grids (a weighted version, with Riemann-sum type coefficients, of the empirical variance should then be used instead). Successful experiments with GE–EV (and its extension CGEM–EV to the case with measurement errors, see Section 5) and its Riemann-sum version, for various simulated two-dimensional Matérn random fields, are described in Girard (2010). See also the two Mathematica Demos which allow one, via the Internet, to easily assess GE–EV for the case $\nu = 3/2$ (Girard, 2015a) and the case $\nu = 1$ (Girard, 2015b).