Journal of Multivariate Analysis 160 (2017) 42–67

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Cross-validation estimation of covariance parameters under fixed-domain asymptotics

François Bachoc *, Agnès Lagnoux, Thi Mong Ngoc Nguyen

Institut de mathématiques de Toulouse, Université Toulouse 3, 31062 Toulouse Cédex 9, France

ARTICLE INFO

Article history:
Received 7 October 2016
Available online 16 June 2017

AMS 2010 subject classifications:
62M40
62F12

Keywords:
Asymptotic normality
Cross validation
Fixed-domain asymptotics
Kriging
Spatial sampling
Strong consistency

ABSTRACT

We consider a one-dimensional Gaussian process having exponential covariance function. Under fixed-domain asymptotics, we prove the strong consistency and asymptotic normality of a cross validation estimator of the microergodic covariance parameter. In this setting, Ying (1991) proved the same asymptotic properties for the maximum likelihood estimator. Our proof includes several original or more involved components, compared to that of Ying. Also, while the asymptotic variance of maximum likelihood does not depend on the triangular array of observation points under consideration, that of cross validation does, and is shown to be lower and upper bounded. The lower bound coincides with the asymptotic variance of maximum likelihood. We provide examples of triangular arrays of observation points achieving the lower and upper bounds. We illustrate our asymptotic results with simulations, and provide extensions to the case of an unknown mean function. To our knowledge, this work constitutes the first fixed-domain asymptotic analysis of cross validation.

1. Introduction

Kriging [28,35] consists in inferring the values of a Gaussian random field given observations at a finite set of observation points. It has become a popular method for a large range of applications, such as geostatistics [25], numerical code approximation [8,29,30] and calibration [9,27] or global optimization [20].

Before Kriging can be applied, a covariance function must be chosen. The most common practice is to estimate statistically the covariance function from a set of observations of the Gaussian process and to plug the estimate in the Kriging equations [35, Chap. 6.8]. Usually, it is assumed that the covariance function belongs to a given parametric family; see [1] for a review of classical families. In this case, the estimation boils down to estimating the corresponding covariance parameters. For covariance parameter estimation, the method of maximum likelihood (ML) is the most frequently studied and used, while cross validation (CV) is an alternative technique [5,36,43]. CV has been shown to have attractive properties, compared to ML, when the parametric family of covariance functions is misspecified [5,7].

There is a fair amount of literature on the asymptotic properties of ML. In this regard, the two main frameworks are increasing-domain and fixed-domain asymptotics [35, p. 62]. Under increasing-domain asymptotics, the average density of observation points is bounded, so that the infinite sequence of observation points is unbounded. Under fixed-domain asymptotics, this sequence is dense in a bounded domain.

Consider first increasing-domain asymptotics. Generally speaking, for all (identifiable) covariance parameters, the ML estimator is consistent and asymptotically normal under some mild regularity conditions. The asymptotic covariance matrix

* Corresponding author.
E-mail addresses: francois.bachoc@math.univ-toulouse.fr (F. Bachoc), lagnoux@univ-tlse2.fr (A. Lagnoux), thi_mong_ngoc.nguyen@math.univ-toulouse.fr (T.M.N. Nguyen).

http://dx.doi.org/10.1016/j.jmva.2017.06.003

is equal to the inverse of the (asymptotic) Fisher information matrix. This result was first shown in [24], and then extended in different directions in [6,12,13,16,31].

The situation is significantly different under fixed-domain asymptotics. Indeed, two types of covariance parameters can be distinguished: microergodic and non-microergodic parameters [18,35]. A covariance parameter is microergodic if, for two different values of it, the two corresponding Gaussian measures are orthogonal; see [18,35]. It is non-microergodic if, even for two different values of it, the two corresponding Gaussian measures are equivalent. Non-microergodic parameters cannot be estimated consistently, but have an asymptotically negligible impact on prediction [32–34,42]. However, it is at least possible to consistently estimate microergodic covariance parameters, and misspecifying them can have a strong negative impact on prediction.

Under fixed-domain asymptotics, there exist results indicating which covariance parameters are microergodic, and providing the asymptotic properties of the corresponding ML estimator. Most of these available results are specific to particular covariance models. In dimension $d = 1$ when the covariance model is exponential, only a reparameterized quantity obtained from the variance and scale parameters is microergodic. It is shown in [40] that the ML estimator of this microergodic parameter is strongly consistent and asymptotically normal. These results are extended in [11], by taking into account measurement errors, and in [10], by taking into account both measurement errors and an unknown mean function. When $d > 1$ and for a separable exponential covariance function, all the covariance parameters are microergodic, and the asymptotic normality of the ML estimator is proved in [41]. Other results in this case are also given in [2,37]. Consistency of ML is shown as well in [23] for the scale parameters of the Gaussian covariance function and in [22] for all the covariance parameters of the separable Matérn 3/2 covariance function. Finally, for the entire isotropic Matérn class of covariance functions, all parameters are microergodic for $d > 4$ [3], and only reparameterized parameters obtained from the scale and variance are microergodic for $d \leqslant 3$; see [42]. In [21], the asymptotic normality of the ML estimators for these microergodic parameters is proved, from previous results in [14] and [39]. Finally we remark that, beyond ML, quadratic variation-based estimators have also been extensively studied, under fixed-domain asymptotics; see for instance [19].

In contrast to ML, CV has received less attention from a theoretical point of view. Under increasing-domain asymptotics, the consistency and asymptotic normality of a CV estimator are proved in [6]. Also, under increasing-domain asymptotics, it is shown in [7] that this CV estimator asymptotically minimizes the integrated square prediction error. To the best of our knowledge, no fixed-domain asymptotic analysis of CV exists in the literature.

In this paper, we provide a first fixed-domain asymptotic analysis of the CV estimator minimizing the CV logarithmic score, see Eq. (5.11) in [28] and [43]. We focus on the case of the one-dimensional exponential covariance function, which was historically the first covariance function for which the asymptotic properties of ML were derived [40]. This covariance function is particularly amenable to theoretical analysis, as its Markovian property yields an explicit (matrix-free) expression of the likelihood function. It turns out that the CV logarithmic score can also be expressed in a matrix-free form, which enables us to prove the strong consistency and asymptotic normality of the corresponding CV estimator. We follow the same general proof architecture as in [40] for ML, but our proof, and the nature of our results, contain several new elements.

In terms of proofs, the random CV logarithmic score, and its derivatives, have more complicated expressions than for ML. This is because the CV logarithm score is based on the conditional distributions of the observations, from both their nearest left and right neighbors, while the likelihood function is solely based on the nearest left neighbors; see Lemma 1 and Lemma 1 in [40] for details. As a consequence, the computations are more involved, and some other tools than in [40] are needed. In particular, many of our asymptotic approximations rely on Taylor expansions of functions of several variables, where each variable is an interpoint distance going to zero; see the proofs for details. In contrast, only Taylor approximations with one variable are needed in [40]. In addition, we use central limit theorems for dependent random variables, while only independent variables need to be considered in [40].

The nature of our asymptotic normality result also differs from that in [40]. In this reference, the asymptotic variance does not depend on the triangular array of observation points. On the contrary, in our case, different triangular arrays of observation points can yield different asymptotic variances. We exhibit a lower and an upper bound for these asymptotic variances, and provide examples of triangular arrays reaching them. The lower bound is in fact equal to the asymptotic variance of ML in [40]. Interestingly, the triangular array given by equispaced observation points attains neither the lower nor the upper bound. It is also pointed out in [6] that equispaced observation points need not provide the smallest asymptotic variance for covariance parameter estimation.

Finally, the fact that the asymptotic variance is larger for CV than for ML is a standard finding in the well-specified case considered here, where the covariance function of the Gaussian process does belong to the parametric family of covariance functions under consideration. In contrast, as mentioned above, CV has attractive properties compared to ML when this well-specified case does not hold [5,7].

The rest of the paper is organized as follows. In Section 2, we present in more detail the setting and the CV estimator under consideration. In Section 3, we give our strong consistency result for this estimator. In Section 4, we provide the asymptotic normality result, together with the analysis of the asymptotic variance. In Section 5, we present numerical experiments, illustrating our theoretical findings. In Section 6, we extend the results of Sections 3 and 4 to the case of an unknown mean function. In Section 7, we give a few concluding remarks. All the proofs are postponed to Section 8.