© Board of the Foundation of the Scandinavian Journal of Statistics 2003. Published by Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA Vol 30: 175–192, 2003

The Bootstrap and Kriging Prediction Intervals

SARA SJÖSTEDT-DE LUNA

Umeå University

ALASTAIR YOUNG

University of Cambridge

ABSTRACT. Kriging is a method for spatial prediction that, given observations of a spatial process, gives the optimal linear predictor of the process at a new specified point. The kriging predictor may be used to define a prediction interval for the value of interest. The coverage of the prediction interval will, however, equal the nominal desired coverage only if it is constructed using the correct underlying covariance structure of the process. If this is unknown, it must be estimated from the data. We study the effect on the coverage accuracy of the prediction interval of substituting the true covariance parameters by estimators, and the effect of bootstrap calibration of coverage properties of the resulting 'plugin' interval. We demonstrate that plugin and bootstrap calibrated intervals are asymptotically accurate in some generality and that bootstrap calibration appears to have a significant effect in improving the rate of convergence of coverage error.

Key words: bootstrap, kriging, prediction, resampling

1. Introduction

The purpose of predictive inference is the assessment of the values of a small number, typically one, of as yet unobserved random variables, given a data sample. Data y represent the observed value of a random variable Y, with density $p_Y(y;\theta)$, depending on the unknown parameter θ, and the unobserved Z to be predicted has, conditionally on $Y = y$, the density $q(z \mid y;\theta)$. In this paper we will be concerned specifically with the construction of prediction intervals for a scalar Z. The aim is the construction of an interval $I \equiv I(y)$ which contains Z with some specified probability, say $1 - 2\gamma$. In the more usual unconditional approach we require:

$$P[Z \in I(Y)] = 1 - 2\gamma,$$

with the probability over the joint distribution of (Z, Y). Alternatively, we may adopt a conditional approach, in which the coverage requirement is conditional on the observed y:

$$P[Z \in I(y)] = 1 - 2\gamma,$$

where now the probability is with respect to the conditional distribution of Z, given $Y = y$.

Various approaches to predictive inference have been proposed, including notions of predictive likelihood and fully Bayesian solutions. A brief review is given by Barndorff-Nielsen & Cox (1994, Section 9.4). In some circumstances, the prediction interval may be constructed via a pivotal quantity, a function of (Z, Y) with a known distribution, not depending on θ. More usually, however, an estimative approach is taken, in which an interval, depending on θ, is constructed with the correct coverage properties, and then an estimate of θ is substituted for the unknown true value. The resulting 'plugin' interval then has coverage properties which differ from the nominal required coverage. In order for the prediction interval to achieve the required coverage interpretation asymptotically it may be necessary for prediction limits to be modified to account for the effects of parameter estimation. Beran (1990a, b) suggested the use of bootstrap calibration to control the coverage probability of the prediction interval: see also Hall *et al.* (1999) and Loh (1987, 1988). The idea is to use bootstrapping, simulating from a model estimated from the data y, to estimate how the actual coverage of the prediction interval varies as a function of the nominal desired coverage. The nominal coverage of the plugin interval is then recalibrated to deliver an actual coverage closer to the desired coverage than that obtained from the unmodified interval. Hall *et al.* (1999) provide evidence that bootstrap calibration of estimative approaches to predictive inference may be more effective than predictive likelihood approaches.

A key issue concerns the general effectiveness of bootstrap calibration as a means of accounting for parameter uncertainty in construction of prediction intervals. The work of Beran (1990a, b) and Hall *et al.* (1999) is concerned with independent data settings, or those, such as autoregressive time series and linear models, which can be expressed in terms of independent errors. In this paper, the principal objective is to investigate the effectiveness of bootstrap calibration in handling parameter uncertainty in a specific, dependent data, spatial setting, that of kriging prediction. In this context, a point predictor is constructed for the value of a Gaussian spatial process at a particular spatial location, given observations of the process at a set of sampling locations. Kriging methods are described in detail by Cressie (1993, Chapter 3). This setting is considered by Putter & Young (2001), who establish conditions under which the kriging predictor constructed using estimates of the parameters of the covariance function is asymptotically efficient with respect to the kriging predictor that utilizes the true values of the parameters.

The present paper considers the effect of parameter estimation on the asymptotic coverage accuracy of the plugin prediction interval constructed from the kriging predictor, and considers the effect of bootstrap calibration in reducing the coverage error. In this context, parametric bootstrapping from a fitted Gaussian process is used to estimate the effects of parameter uncertainty on the accuracy of the prediction interval. Informally, it is reasonable to expect that bootstrapping from a fitted model with parameter values 'close' to the true values will provide evidence of the effects of parameter estimation and allow a recalibration of the prediction interval, to yield a reduction in coverage error over that of the crude plugin interval, which makes no attempt to allow for parameter uncertainty. We establish that in some generality asymptotic accuracy is achieved by both the plugin and bootstrap calibrated intervals, and demonstrate further that the bootstrap calibration does indeed have a significant effect in reducing the order of magnitude of the coverage error. We believe that these findings point to the value of bootstrap calibration as a means of modifying an inference to accomodate parameter uncertainty in more general problems of spatial prediction.

Section 2 of the paper describes the problem of kriging prediction, and the procedure of ordinary kriging, which is the primary focus of the rest of the paper. Plugin and bootstrap calibrated unconditional prediction intervals are defined in Section 3, with conditional prediction intervals described in Section 4. Coverage properties and our main theoretical results are described in Section 5. Extension of our results to the technique of universal kriging is detailed in Section 6. Section 7 considers a number of specific covariance models, while a numerical study, which includes illustration of examples of Section 7, as well as more general models, is reported in Section 8. Concluding remarks are given in Section 9 and proofs of the main results are given in the Appendix.

2. Ordinary kriging

Kriging is a method for (spatial) prediction, widely used in mining, hydrology, forestry and other fields. Given a spatial process, observed at sampling locations $x_1, \dots, x_n$, it gives the optimal (minimum Mean Squared Error, MSE) unbiased linear predictor of the process at a given new point, x_p, taking into account the spatial dependence of the observations. In ordinary kriging, it is assumed that the mean of the process $\{Z(x), x \in \mathbb{R}^d\}$ is constant but unknown, and that the covariance structure is known. In particular we consider a stationary Gaussian process $\{Z(x), x \in \mathbb{R}^d\}$ with $E[Z(x)] = \mu$ and covariance function $C_\xi(t) = \mathrm{Cov}[Z(x+t), Z(x)]$, depending on some parameter ξ. Given the observations $\mathbb{Z} = (Z(x_1), \dots, Z(x_n))^{\mathrm{T}}$, we wish to find weights $\alpha = (\alpha_{1n}, \dots, \alpha_{nn})^{\mathrm{T}}$ with $\sum_{i=1}^n \alpha_{in} = 1$ such that the MSE

$$E\left[\sum_{i=1}^{n} \alpha_{in} Z(x_i) - Z(x_p)\right]^2 \tag{1}$$

is minimized. The restriction on the weights ensures that the kriging predictor

$$\mathscr{Z}(x_p) = \sum_{i=1}^{n} \alpha_{in} Z(x_i),$$

is unbiased. The $n \times n$ covariance matrix of $\mathbb{Z}$ is denoted by

$$\Omega = \mathrm{Cov}[\mathbb{Z}, \mathbb{Z}] = \{C_\xi(x_i - x_j)\}_{i,j=1,\dots,n},$$

the n-dimensional covariance vector by

$$\omega = \mathrm{Cov}[Z(x_p), \mathbb{Z}] = \{C_\xi(x_p - x_i)\}_{i=1,\dots,n},$$

and the variance of the process by $\sigma^2 = \mathrm{Var}[Z(x_i)]$. Then,

$$\alpha = \alpha(\xi) = \Omega^{-1}\left(\omega + \mathbf{1}\frac{(1 - \mathbf{1}^{\mathrm{T}}\Omega^{-1}\omega)}{\mathbf{1}^{\mathrm{T}}\Omega^{-1}\mathbf{1}}\right), \tag{2}$$

defines an n-vector which minimizes (1) given $C_\xi(\cdot)$, and satisfies $\mathbf{1}^{\mathrm{T}}\alpha = 1$. The theoretical kriging prediction error in terms of minimum MSE, for the (theoretical) kriging predictor $\mathscr{Z}(x_p)$, using the weights in (2), is

$$v^2 = v^2(\xi) = \mathrm{Var}[\mathscr{Z}(x_p) - Z(x_p)] = \sigma^2 - \omega^{\mathrm{T}}\Omega^{-1}\omega + \frac{(\mathbf{1}^{\mathrm{T}}\Omega^{-1}\omega - 1)^2}{\mathbf{1}^{\mathrm{T}}\Omega^{-1}\mathbf{1}}. \tag{3}$$

Note that (2) and (3) depend uniquely on the covariance structure. In this paper we will consider prediction intervals for $Z(x_p)$ under two types of asymptotics; *infill asymptotics*, where samples are taken from a fixed bounded region and the sampling locations become increasingly dense, and *increasing domain asymptotics*, where the distance between neighbouring sampling locations remains bounded from below and the domain from which sampling takes place necessarily increases.

3. Unconditional prediction intervals

Let $\Phi(z_\gamma) = 1 - \gamma$, where Φ denotes the standard normal distribution function. If the covariance structure is known, it follows that

$$I_\gamma(\xi) = \mathscr{Z}(x_p) \pm z_\gamma v,$$

is an exact $1 - 2\gamma$ prediction interval for $Z(x_p)$. Hence, we have

$$P_\xi[Z(x_p) \in I_\gamma(\xi)] = 1 - 2\gamma.$$

Note that the probability is over the joint distribution of $(Z(x_p), \mathbb{Z})$, and that the interval $I_\gamma(\xi)$ does not depend on μ. The coverage is exactly $1 - 2\gamma$ only if constructed using the correct underlying covariance structure. In practice, this covariance structure will be unknown and is estimated from the data. We will study how the coverage of the prediction interval is affected when the true covariance parameters are substituted by estimates, so called plugin prediction intervals. We will also study the effect of bootstrap calibration of coverage properties of plugin intervals.

3.1. Plugin prediction intervals

When ξ is unknown, the bootstrap substitution principle replaces ξ by an estimator $\hat{\xi}_n$, corresponding to estimating the covariance function $C_\xi(t)$ by $C_{\hat{\xi}_n}(t)$. The *plugin prediction interval* of nominal coverage $1 - 2\gamma$ is $I_\gamma(\hat{\xi}_n)$. Write

$$\pi_n(\gamma, \xi) = P_\xi[Z(x_p) \in I_\gamma(\hat{\xi}_n)],$$

for the true coverage, under ξ, of the prediction interval $I_\gamma(\hat{\xi}_n)$. This is not in general equal to the nominal coverage $1 - 2\gamma$. The probability $\pi_n(\gamma, \xi)$ does not depend on μ if $\hat{\xi}_n$ is a location invariant estimator, since

$$\begin{aligned} P_\xi[Z(x_p) \in I_\gamma(\hat{\xi}_n)|\mu] &= P_\xi[|Z(x_p) - \mu - \hat{\alpha}^{\mathrm{T}}(\mathbb{Z} - \mu\mathbf{1})| \leqslant z_\gamma \hat{v}|\mu] \\ &= P_\xi[|Z(x_p) - \hat{\alpha}^{\mathrm{T}}\mathbb{Z}| \leqslant z_\gamma \hat{v}|\mu = 0]. \end{aligned}$$

Here, and in the sequel a $\hat{}$ denotes when the parameters ξ have been replaced by $\hat{\xi}_n$ in any function. In this paper we will focus on Maximum Likelihood (ML) estimators, which for a Gaussian process makes $\hat{\xi}_n$ location invariant.

3.2. Bootstrap calibrated prediction intervals

Ideally, we want to recalibrate the plugin prediction interval $I_\gamma(\hat{\xi}_n)$ by seeking a $\tilde{\gamma}$ such that

$$\pi_n(\tilde{\gamma}, \xi) = P_\xi[Z(x_p) \in I_{\tilde{\gamma}}(\hat{\xi}_n)] = 1 - 2\gamma.$$

We would hence use the plugin prediction interval of nominal coverage $1 - 2\tilde{\gamma}$. Since $\tilde{\gamma}$ is unknown we construct a bootstrap estimator. Because $\pi_n(\gamma, \xi)$ is monotonically decreasing and continuous in γ (see the proof of Lemma 1), we define $\pi_n^{-1}(\cdot, \xi)$ to be the inverse of $\pi_n(\cdot, \xi)$ with respect to γ. We estimate $\tilde{\gamma}$ by $\pi_n^{-1}(1 - 2\gamma, \hat{\xi}_n)$, where

$$\pi_n(\gamma, \hat{\xi}_n) = P_{\hat{\xi}_n}[Z^*(x_p) \in I_\gamma(\hat{\xi}_n^*)],$$

is a bootstrap estimator of $\pi_n(\gamma, \xi)$. The bootstrap sample $Z^*(x_p), Z^*(x_1), \ldots, Z^*(x_n)$ is a realization of a Gaussian process with covariance parameters $\hat{\xi}_n = \hat{\xi}_n(\mathbb{Z})$ and mean zero. Furthermore, $\hat{\xi}_n^* = \hat{\xi}_n(\mathbb{Z}^*)$ denotes the parameter estimator as constructed from $\mathbb{Z}^* = (Z^*(x_1), \ldots, Z^*(x_n))^{\mathrm{T}}$. The bootstrap calibrated prediction interval is $I_{\pi_n^{-1}(1-2\gamma, \hat{\xi}_n)}(\hat{\xi}_n)$. Monte Carlo simulation can be used to approximate the function $\pi_n(\gamma, \hat{\xi}_n)$ to arbitrary accuracy. In practice, we use a finite number, B, say, of bootstrap samples, to approximate $\pi_n(\gamma, \hat{\xi}_n)$ for a set of values of γ close to the nominal desired coverage, and approximate the bootstrap estimator $\tilde{\gamma}$ by simple quadratic interpolation.

4. Conditional prediction intervals

A conditional prediction interval is constructed in such a way that the coverage probability, computed conditional on the observed values, corresponds to the desired coverage $1-2\gamma$. Therefore, it will in general not be equivalent to an unconditional prediction interval, which is formed to have coverage $1-2\gamma$ without holding the observed values fixed. In this section we will consider conditional prediction intervals for $Z(x_p)$ given $\mathbb{Z}$. The distribution of $Z(x_p)$ given $\mathbb{Z}$ is Gaussian with mean value

$$\eta = E_\xi[Z(x_p)|\mathbb{Z}] = \mu + \omega^{\mathrm{T}}\Omega^{-1}\mathbb{E},$$

where $\mathbb{E} = \mathbb{Z} - \mu\mathbf{1}$, and variance

$$\tau^2 = \mathrm{Var}[Z(x_p)|\mathbb{Z}, \xi] = \sigma^2 - \omega^{\mathrm{T}}\Omega^{-1}\omega.$$

A conditional prediction interval for $Z(x_p)$ given $\mathbb{Z}$ with true coverage $1-2\gamma$ is given by

$$I_\gamma^{\mathrm{C}}(\xi) = \eta \pm z_\gamma\tau.$$

The corresponding conditional plugin prediction interval with nominal coverage $1-2\gamma$ is $I_\gamma^{\mathrm{C}}(\hat{\xi}_n)$. Inserting the ML estimators $\hat{\xi}_n$ and $\hat{\mu} = \mathbf{1}^{\mathrm{T}}\hat{\Omega}^{-1}\mathbb{Z}/\mathbf{1}^{\mathrm{T}}\hat{\Omega}^{-1}\mathbf{1}$ gives us $I_\gamma^{\mathrm{C}}(\hat{\xi}_n) = \hat{\alpha}^{\mathrm{T}}\mathbb{Z} \pm z_\gamma\hat{\tau}$. The true coverage of $I_\gamma^{\mathrm{C}}(\hat{\xi}_n)$ given $\mathbb{Z}$ is

$$P_\xi[Z(x_p) \in I_\gamma^{\mathrm{C}}(\hat{\xi}_n)|\mathbb{Z}] = \Phi(A_n + z_\gamma B_n^{\mathrm{C}}) - \Phi(A_n - z_\gamma B_n^{\mathrm{C}}), \tag{4}$$

where A_n satisfies

$$\tau(\xi)A_n(\xi) = (\hat{\omega}^{\mathrm{T}}\hat{\Omega}^{-1} - \omega^{\mathrm{T}}\Omega^{-1})\mathbb{E} + (1 - \hat{\omega}^{\mathrm{T}}\hat{\Omega}^{-1}\mathbf{1})\hat{\mu}_E, \tag{5}$$

with $\hat{\mu}_E = \hat{\mu} - \mu = \mathbf{1}^{\mathrm{T}}\Omega^{-1}\mathbb{E}/\mathbf{1}^{\mathrm{T}}\hat{\Omega}^{-1}\mathbf{1}$, and

$$B_n^{\mathrm{C}} = B_n^{\mathrm{C}}(\xi) = \tau(\hat{\xi}_n)/\tau(\xi). \tag{6}$$

The coverage of the plugin prediction interval $I_\gamma(\hat{\xi}_n)$, that was proposed in Section 3.1, given $\mathbb{Z}$, is

$$P_\xi[Z(x_p) \in I_\gamma(\hat{\xi}_n)|\mathbb{Z}] = \Phi(A_n + z_\gamma B_n) - \Phi(A_n - z_\gamma B_n), \tag{7}$$

where

$$B_n = B_n(\xi) = v(\hat{\xi}_n)/\tau(\xi). \tag{8}$$

5. Coverage properties

In order to say something about the (asymptotic) coverage of the different plugin prediction intervals we need to know some properties of the parameter estimators. Contiguity is one such property, which can be defined as follows (for more details see e.g. Roussas, 1972). For every n, let $(\mathscr{X}_n, \mathscr{A}_n)$ be a measurable space and let $\{Q_n\}_{n\geqslant 1}$ and $\{Q'_n\}_{n\geqslant 1}$ be two sequences of probability measures on $\{(\mathscr{X}_n, \mathscr{A}_n)\}_{n\geqslant 1}$. Then the sequence $\{Q'_n\}_{n\geqslant 1}$ is said to be *contiguous* with respect to $\{Q_n\}_{n\geqslant 1}$ if for every $D_n \in \mathscr{A}_n$, $Q_n(D_n) \to 0$ implies that $Q'_n(D_n) \to 0$, as $n \to \infty$. Let Ξ denote the parameter space of ξ, and let $M(\xi)$ denote the set of all sequences $\{\xi_n\}_{n\geqslant 1}$ in Ξ such that $\{P_{\xi_n,n}\}_{n\geqslant 1}$ is contiguous with respect to $\{P_{\xi,n}\}_{n\geqslant 1}$. Here $P_{\xi,n}$ denotes the joint distribution of $\mathbb{Z}$ given ξ, and similarly $P_{\xi_n,n}$ denotes the joint distribution of $\mathbb{Z}$ given ξ_n. In both cases we assume that the mean of the process is zero. This implies no restriction in the reasoning below as long as the estimators $\hat{\xi}_n$ are location invariant. We have from (7) that

$$P_{\xi_n}[Z(x_p) \in I_\gamma(\hat{\xi}_n)|\mathbb{Z}] = G(\gamma, A_n(\xi_n), B_n(\xi_n)), \tag{9}$$

where the function G is continuous in all its arguments and $G(\gamma, 0, 1) = 1 - 2\gamma$.

Lemma 1

Assume that

$$A_n(\xi_n) \to 0 \text{ and } B_n(\xi_n) \to 1 \quad \text{in } P_{\xi,n}\text{-probability}, \tag{10}$$

for some $\{\xi_n\} \in M(\xi)$. *Then*

$$P_{\xi_n}[Z(x_p) \in I_\gamma(\hat{\xi}_n)|\mathbb{Z}] \to 1 - 2\gamma \quad \text{in } P_{\xi,n}\text{-probability}, \tag{11}$$

$$\pi_n(\gamma, \xi_n) = P_{\xi_n}[Z(x_p) \in I_\gamma(\hat{\xi}_n)] \to 1 - 2\gamma \quad \text{as } n \to \infty, \tag{12}$$

and

$$\pi_n^{-1}(1 - 2\gamma, \xi_n) \to \gamma \quad \text{as } n \to \infty. \tag{13}$$

The above lemma, with $\xi_n \equiv \xi$, ensures that both the unconditional and the conditional plugin prediction intervals for $Z(x_p)$ asymptotically give the correct coverage $1 - 2\gamma$. A similar property holds for the bootstrap calibrated plugin intervals. Let $M_S(\xi)$ be some subset of $M(\xi)$.

Lemma 2

Assume that (10) holds for all $\{\xi_n\} \in M_S(\xi)$, *and that* $\{\hat{\xi}_n\} \in M_S(\xi)$ *a.s. Then for any* $\{\xi_n\} \in M_S(\xi)$,

$$P_{\xi_n}[Z(x_p) \in I_{\pi_n^{-1}(1-2\gamma,\hat{\xi}_n)}(\hat{\xi}_n)|\mathbb{Z}] \to 1 - 2\gamma \quad \text{in } P_{\xi,n}\text{-probability}, \tag{14}$$

and

$$P_{\xi_n}[Z(x_p) \in I_{\pi_n^{-1}(1-2\gamma,\hat{\xi}_n)}(\hat{\xi}_n)] \to 1 - 2\gamma \quad \text{as } n \to \infty. \tag{15}$$

Lemmas 1 and 2 also hold if $I_\gamma(\hat{\xi}_n)$ and B_n are replaced by $I_\gamma^C(\hat{\xi}_n)$ and B_n^C, respectively.

Remark 1. The condition $\{\hat{\xi}_n\} \in M_S(\xi)$ a.s. in Lemma 2 can be relaxed to the following condition: For each subsequence $\{n_k\}$ there is a subsequence $\{n_{k(l)}\}$ such that $\{\hat{\xi}_{n_{k(l)}}\} \in M_S(\xi)$ a.s., which is equivalent to that $\{P_{\hat{\xi}_n,n}\}$ is contiguous with respect to $\{P_{\xi,n}\}$ a.s. for that subsequence. This is useful if we know how the estimator $\hat{\xi}_n$ behaves in probability (see Section 7).

5.1. Convergence rates

The convergence rate of the bootstrap calibrated prediction intervals turns out to be faster than for the plugin prediction intervals under some additional assumptions. The following lemmas will summarize these properties. First we state two important assumptions.

(A1) Suppose that there are functions $\{a_{ij}(\xi)\}$ such that

$$E_\xi[A_n^2(\xi)] = a_{11}(\xi)/n + o(n^{-1}),$$

$$E_\xi[B_n(\xi) - 1] = a_{12}(\xi)/n + o(n^{-1}),$$

$$E_\xi[(B_n(\xi) - 1)^2] = a_{22}(\xi)/n + o(n^{-1}).$$

(A2) Assume that the sequences $\{nA_n^2(\xi)\}$ and $\{n(B_n(\xi) - 1)^2\}$ are uniformly integrable.

Lemma 3

Suppose that (A1) and (A2) hold. Then, under the assumptions of Lemma 2,

$$\pi_n(\gamma, \xi) = 1 - 2\gamma + b(\gamma, \xi)/n + o(n^{-1}), \tag{16}$$

where

$$b(\gamma, \xi) = z_\gamma \varphi(z_\gamma)[2a_{12}(\xi) - a_{11}(\xi) - z_\gamma^2 a_{22}(\xi)],$$

and $\varphi(\cdot)$ is the density of a standard normal distribution. Moreover,

$$R_n(\gamma, \xi) = \pi_n^{-1}(1 - 2\gamma, \xi) - \gamma - b(\gamma, \xi)/2n = o(n^{-1}). \tag{17}$$

This lemma implies that the (unconditional) plugin prediction interval $I_\gamma(\hat{\xi}_n)$ typically has coverage $1 - 2\gamma + O(n^{-1})$.

Lemma 4

Suppose that $\partial G(\gamma, A_n, B_n)/\partial\gamma$ and $b(\gamma, \xi)$ are bounded. If

$$E_\xi[nR_n(\gamma, \hat{\xi}_n)] \to 0, \quad \text{and} \quad \delta_n = b(\gamma, \hat{\xi}_n) - b(\gamma, \xi) \xrightarrow{P} 0, \quad \text{as } n \to \infty, \tag{18}$$

then, under the assumptions of Lemma 3,

$$\pi_n(\pi_n^{-1}(1 - 2\gamma, \hat{\xi}_n), \xi) = 1 - 2\gamma + o(n^{-1}).$$

Remark 2. Lemmas 3 and 4 hold even if the functions $\{a_{ij}(\xi)\}$ depend on n, when the functions are uniformly bounded with respect to n.

Here, we conclude that the (unconditional) bootstrap calibrated prediction interval $I_{\pi_n^{-1}(1-2\gamma,\hat{\xi}_n)}(\hat{\xi}_n)$ has coverage $1 - 2\gamma + o(n^{-1})$, and thus a faster convergence rate than the plugin prediction interval.

6. Universal kriging

Suppose that the mean of the process $\{Z(x), x \in \mathbb{R}^d\}$ changes with x in a way such that

$$Z(x) = \mu(x) + \epsilon(x),$$

where

$$\mu(x) = \sum_{j=1}^{q} \beta_j f_j(x),$$

and $\epsilon(x)$ is a mean zero Gaussian stationary process with covariance parameters ξ. Here $f_j(x)$, $j = 1, \ldots, q$, are known functions of the coordinates, and $\beta = (\beta_1, \ldots, \beta_q)^{\mathrm{T}}$, are unknown coefficients. Suppose that we have observed the process at n locations, giving us the observed values $\mathbb{Z} = X\beta + \mathbb{E}$, where $X = \{f_j(x_i)\}$ is an $n \times q$ matrix and $\mathbb{E} = (\epsilon(x_1), \dots, \epsilon(x_n))^{\mathrm{T}}$. We want to predict the value of the process at a new location x_p, corresponding to $Z(x_p) = \beta^{\mathrm{T}}\mathbf{f}_p + \epsilon(x_p)$, where $\mathbf{f}_p = (f_1(x_p), \dots, f_q(x_p))^{\mathrm{T}}$. Assuming that the covariance structure is known, the universal kriging predictor corresponds to the best linear unbiased predictor of $Z(x_p)$, given $\mathbb{Z}$, minimizing (1). It is given by $\mathscr{Z}_u(x_p) = \alpha_u^{\mathrm{T}}\mathbb{Z}$, where

$$\alpha_u = \Omega^{-1}[\omega + X(X^{\mathrm{T}}\Omega^{-1}X)^{-1}(\mathbf{f}_p - X^{\mathrm{T}}\Omega^{-1}\omega)].$$

The unbiasedness of $\mathscr{Z}_u(x_p)$ implies that $\alpha_u^{\mathrm{T}}X\beta = \beta^{\mathrm{T}}\mathbf{f}_p$. The MSE is

$$v_u^2 = E[(\mathscr{Z}_u(x_p) - Z(x_p))^2] = \sigma^2 - \omega^{\mathrm{T}}\Omega^{-1}\omega + (\mathbf{f}_p - X^{\mathrm{T}}\Omega^{-1}\omega)^{\mathrm{T}}(X^{\mathrm{T}}\Omega^{-1}X)^{-1}(\mathbf{f}_p - X^{\mathrm{T}}\Omega^{-1}\omega). \tag{19}$$

Hence, an exact $1 - 2\gamma$ (unconditional) prediction interval for $Z(x_p)$ is

$$I_\gamma(\xi) = \alpha_u^{\mathrm{T}}\mathbb{Z} \pm z_\gamma v_u, \tag{20}$$

and the corresponding plugin prediction interval $I_\gamma(\hat{\xi}_n)$, with coverage $\pi_n(\gamma, \xi) = P_\xi[Z(x_p) \in I_\gamma(\hat{\xi}_n)]$. We further have that

$$\zeta = E_\xi[Z(x_p)|\mathbb{Z}] = \beta^{\mathrm{T}}\mathbf{f}_p + \omega^{\mathrm{T}}\Omega^{-1}\mathbb{E},$$

and

$$\mathrm{Var}[Z(x_p)|\mathbb{Z}, \xi] = \tau^2(\xi) = \sigma^2 - \omega^{\mathrm{T}}\Omega^{-1}\omega.$$

From this we can construct an exact $1 - 2\gamma$ prediction interval conditional on $\mathbb{Z}$ as $I_\gamma^{\mathrm{C}}(\xi) = \zeta \pm z_\gamma\tau$, assuming that ξ and β are known. The corresponding conditional plugin prediction interval with nominal coverage $1 - 2\gamma$ is $I_\gamma^{\mathrm{C}}(\hat{\xi}_n)$. Inserting the ML estimators $\hat{\xi}_n$ and $\hat{\beta} = (X^{\mathrm{T}}\hat{\Omega}^{-1}X)^{-1}X^{\mathrm{T}}\hat{\Omega}^{-1}\mathbb{Z}$ gives us

$$I_\gamma^{\mathrm{C}}(\hat{\xi}_n) = \hat{\alpha}_u^{\mathrm{T}}\mathbb{Z} \pm z_\gamma\hat{\tau}.$$

The true coverage of $I_\gamma^{\mathrm{C}}(\hat{\xi}_n)$ given $\mathbb{Z}$ corresponds to (4) with $B_n^{\mathrm{C}}(\xi) = \tau(\hat{\xi}_n)/\tau(\xi)$ and $A_n(\xi)$ satisfying

$$\tau(\xi)A_n(\xi) = (\hat{\omega}^{\mathrm{T}}\hat{\Omega}^{-1} - \omega^{\mathrm{T}}\Omega^{-1})\mathbb{E} + (\mathbf{f}_p^{\mathrm{T}} - \hat{\omega}^{\mathrm{T}}\hat{\Omega}^{-1}X)\hat{\beta}_E, \tag{21}$$

where $\hat{\beta}_E = \hat{\beta} - \beta = (X^{\mathrm{T}}\hat{\Omega}^{-1}X)^{-1}X^{\mathrm{T}}\hat{\Omega}^{-1}\mathbb{E}$. The coverage of the plugin prediction interval of (20), given $\mathbb{Z}$, equals (7), with A_n satisfying (21) and $B_n = v_u(\hat{\xi}_n)/\tau(\xi)$. The ML estimators $\hat{\xi}_n$ are location invariant, which implies that the coverage probabilities of all the plugin prediction intervals are the same whether or not $\beta = 0$. For example, using the fact that $\hat{\alpha}_u^{\mathrm{T}}X\beta = \beta^{\mathrm{T}}\mathbf{f}_p$,

$$\begin{aligned} P_\xi[Z(x_p) \in I_\gamma(\hat{\xi}_n)|\mathbb{Z}, \beta] &= P_\xi[|Z(x_p) - \mathbf{f}_p^{\mathrm{T}}\beta - \hat{\alpha}_u^{\mathrm{T}}(\mathbb{Z} - X\beta)| \leqslant z_\gamma\hat{v}_u|\mathbb{Z}, \beta] \\ &= P_\xi[|Z(x_p) - \hat{\alpha}_u^{\mathrm{T}}\mathbb{Z}| \leqslant z_\gamma\hat{v}_u|\mathbb{Z}, \beta = 0]. \end{aligned}$$

Hence, Lemmas 1–4 hold for universal kriging situations as well, with the corresponding changes of A_n and B_n.

7. Examples

In this section, two examples with exponentially decaying covariance functions are discussed. Both infill and increasing domain asymptotics are considered. We will show that the plugin and the bootstrap calibrated prediction intervals have the correct coverage asymptotically for these examples.

7.1. A one-dimensional example

Consider a stationary Gaussian process, $\{Z(x) \in \mathbb{R}\}$, on the real line with expected value μ and covariance function

$$C_\xi(x_i - x_j) = \mathrm{Cov}[Z(x_i), Z(x_j)] = \sigma^2 \mathrm{e}^{-\theta|x_i - x_j|}, \quad \sigma^2, \theta > 0, \tag{22}$$

with $\xi = (\sigma^2, \theta)^{\mathrm{T}}$. Suppose that the process is observed at locations $x_i = i/a_n, i = 1, \dots, n$. If $a_n = O(1)$ it corresponds to increasing domain asymptotics and if $a_n = O(n)$ it corresponds to infill asymptotics. We will consider these two cases separately, demanding that there are constants $0 < K_1 \leqslant K_2 < \infty$ such that $K_1 \leqslant a_n \leqslant K_2 n$. Let $x_0 = -\infty, x_{n+1} = \infty$, and define $k = k(n)$ to be an integer satisfying $x_k \leqslant x_p \leqslant x_{k+1}$, where x_p is the location where we want to predict the process. This gives us information about whether or not the point of prediction is an interior point, in between the observed locations. The coverage of the prediction intervals depends on several items. From Lemma 1 in Ying (1993) and the notation $c_n = \mathrm{e}^{-\theta(x_p - x_k)}$, $d_n = \mathrm{e}^{-\theta(x_{k+1} - x_p)}$, and $\mathbb{E} = (\epsilon_1, \dots, \epsilon_n)^{\mathrm{T}}$, where $\epsilon_i = Z(x_i) - \mu$, it follows that

$$\mathbf{1}^{\mathrm{T}}\Omega^{-1}\omega = (c_n + d_n)/(1 + \mathrm{e}^{-\theta/a_n}), \tag{23}$$

$$\mathbb{E}^{\mathrm{T}}\Omega^{-1}\omega = \frac{\epsilon_k c_n(1 - d_n^2) + \epsilon_{k+1} d_n (1 - c_n^2)}{1 - \mathrm{e}^{-2\theta/a_n}}, \tag{24}$$

$$\tau^2(\xi) = \frac{\sigma^2(1 - c_n^2)(1 - d_n^2)}{1 - \mathrm{e}^{-2\theta/a_n}}, \tag{25}$$

and

$$\sigma^2\mathbf{1}^{\mathrm{T}}\Omega^{-1}\mathbb{E} = \frac{n(1 - \mathrm{e}^{-\theta/a_n})\bar{\epsilon} + \mathrm{e}^{-\theta/a_n}(\epsilon_1 + \epsilon_n)}{1 + \mathrm{e}^{-\theta/a_n}}, \tag{26}$$

where $\bar{\epsilon} = \sum_{i=1}^n \epsilon_i / n$. Let Ξ_0 be a compact region in $\mathbb{R}_+^2$ (positive orthant) that contains the true parameters $\xi = (\sigma^2, \theta)^{\mathrm{T}}$ as an interior point.

7.1.1. Infill asymptotics

For a mean zero Gaussian process, Putter & Young (2001) showed that if $\{\xi_n = (\sigma_n^2, \theta_n)^{\mathrm{T}}\} \in M_S(\xi)$, where

$$M_S(\xi) = \{\{\xi_n \in \Xi_0\} : \theta_n - \theta = O(1) \quad \text{and} \quad \sigma_n^2\theta_n - \sigma^2\theta = O(n^{-1/2})\},$$

then $\{P_{\xi_n,n}\}$ is contiguous with respect to $\{P_{\xi,n}\}$ under infill asymptotics. Furthermore, Ying (1991) showed that the ML estimators $\hat{\xi}_n \in \Xi_0$ are not consistent estimators of ξ but that $\hat{\sigma}_n^2\hat{\theta}_n \to \sigma^2\theta$ a.s. and $\hat{\sigma}_n^2\hat{\theta}_n - \sigma^2\theta = O_p(n^{-1/2})$. Hence, for each subsequence of $\{\hat{\xi}_n\}$, there is a subsequence that belongs to $M_S(\xi)$ a.s. Assume that x_p is an interior point, which implies that $x_{k+1} - x_k = O(n^{-1})$. By (25) and Taylor expansions, we then have

$$\tau^2(\xi) = 2\sigma^2\theta(x_p - x_k)(x_{k+1} - x_p)a_n[1 + O(n^{-2})] = O(n^{-1}). \tag{27}$$

Hence, $B_n^{\mathrm{C}}(\xi_n) = \tau(\hat{\xi}_n)/\tau(\xi_n) \to 1$ a.s. if $\{\xi_n\} \in M_S(\xi)$. Let us now consider $A_n(\xi_n)$ and let $\omega_n = \omega(\xi_n)$ and $\Omega_n = \Omega(\xi_n)$. From (24) and Taylor expansions, we conclude that

$$(\hat{\omega}^{\mathrm{T}}\hat{\Omega}^{-1} - \omega_n^{\mathrm{T}}\Omega_n^{-1})\mathbb{E} = \epsilon_k o(n^{-1}) + \epsilon_{k+1} o(n^{-1}). \tag{28}$$

Equation (26), with $\mathbb{E} = \mathbf{1}$, implies that $\hat{\sigma}_n^2 \mathbf{1}^{\mathrm{T}}\hat{\Omega}^{-1}\mathbf{1} \geqslant 1$, and further that

$$\hat{\mu}_E = (\epsilon_1 + \epsilon_n)O(1) + \bar{\epsilon}O(1). \tag{29}$$

From (23), we conclude that $1 - \hat{\omega}^{\mathrm{T}}\hat{\Omega}^{-1}\mathbf{1} = O(n^{-2})$. This, together with (27)–(29) implies that

$$A_n(\xi_n) = \epsilon_k o(n^{-1/2}) + \epsilon_{k+1} o(n^{-1/2}) + (\epsilon_1 + \epsilon_n)O(n^{-3/2}) + \bar{\epsilon}O(n^{-3/2}), \tag{30}$$

when $\{\xi_n\} \in M_S(\xi)$. We thus conclude that $A_n(\xi_n) \to 0$ in $P_{\xi,n}$-probability, since the variance of $\bar{\epsilon}$ is bounded. Furthermore, $(1 - \hat{\omega}^{\mathrm{T}}\hat{\Omega}^{-1}\mathbf{1})^2/\mathbf{1}^{\mathrm{T}}\hat{\Omega}^{-1}\mathbf{1} = O(n^{-4})$, which from (3) implies that

$$B_n(\xi_n) = v(\hat{\xi}_n)/\tau(\xi_n) \to 1 \text{ a.s.}$$

All the above ensures that both the conditional and the unconditional prediction intervals, $I_\gamma^{\mathrm{C}}(\hat{\xi}_n)$ and $I_\gamma(\hat{\xi}_n)$, as well as their bootstrap calibrated versions, have the correct coverage asymptotically. A similar reasoning shows that the same holds if x_p is not an interior point.

Remark 3. In order to understand the convergence rates of the plugin and bootstrap calibrated prediction intervals, we need to study the conditions of Lemma 4. For this example, the derivative

$$\delta G(\gamma, A_n, B_n^{\mathrm{C}})/\delta\gamma = -B_n^{\mathrm{C}}[\varphi(A_n + z_\gamma B_n^{\mathrm{C}}) + \varphi(A_n - z_\gamma B_n^{\mathrm{C}})]/\varphi(z_\gamma),$$

is bounded, since B_n^{C} is bounded. Conditions (A1) and (A2) are also verifiable for A_n: From (30) we have that $E_\xi[A_n^2] = o(n^{-1})$, which implies that $a_{11}(\xi) = 0$. Moreover, since $\{\epsilon_i\}$ is Gaussian with finite variance, (30) implies that $\{nA_n^2\}$ is uniformly integrable. Turning our attention to B_n^{C}, we have that

$$(B_n^{\mathrm{C}})^2 - 1 = \frac{\tau^2(\hat{\xi}_n) - \tau^2(\xi)}{\tau^2(\xi)} = \frac{(\hat{\sigma}_n^2\hat{\theta}_n - \sigma^2\theta)}{\sigma^2\theta} + O(n^{-2}).$$

It is possible to verify that the bias and MSE of $\hat{\sigma}_n^2\hat{\theta}_n$ is of order $O(n^{-1})$, and that $\{n((B_n^{\mathrm{C}})^2 - 1)^2\}$ is uniformly integrable (see Sjöstedt-de Luna, 2002). More specifically, $a_{22}(\xi) = 2\sigma^4\theta^2$, and

$$a_{21}(\xi, n) = \frac{n}{2a_n}(E_\xi[2\hat{\sigma}_n^2\hat{\theta}_n^2 - \hat{\theta}_n^2\hat{\mu}_E^2] - \sigma^2\theta^2) + \frac{1}{2a_n}\sum_{i=2}^{n} E_\xi[\hat{\mu}_E\hat{\theta}_n(\epsilon_i + \epsilon_{i-1}) - \hat{\theta}_n^2\epsilon_{i-1}^2],$$

which implies that $b(\gamma, \xi)$ is bounded. However, since $(\hat{\sigma}_n^2, \hat{\theta}_n)$ are not consistent estimators of (σ^2, θ) (cf. Ying, 1991) the rest of the assumptions in Lemma 4 are difficult to verify.

7.1.2. Increasing domain asymptotics

For increasing domain asymptotics the distance between neighbouring observations is bounded from below such that for some $\delta > 0$ it always holds that $|x_i - x_j| \geqslant \delta$ for any two locations. Let

$$M_S(\xi) = \{\{\xi_n \in \Xi_0\} : \theta_n - \theta = O(n^{-1/2}) \quad \text{and} \quad \sigma_n^2\theta_n - \sigma^2\theta = O(n^{-1/2})\}.$$

Putter & Young (2001) showed that $\{P_{\xi_n,n}\}$ is contiguous with respect to $\{P_{\xi,n}\}$ if $\{\xi_n\} \in M_S(\xi)$, for $\mu = 0$. The ML estimators are asymptotically normally distributed, and such that $\hat{\xi}_n - \xi = O_p(n^{-1/2})$ and $\hat{\mu} - \mu = O_p(n^{-1/2})$, when $x_i = i/a_n$, and $a_n = O(1)$. This follows by checking the conditions of Theorem 3 given by Mardia & Marshall (1984). Therefore, $\sqrt{n}(\hat{\sigma}_n^2\hat{\theta}_n - \sigma^2\theta)$ is also asymptotically normally distributed. Hence, for every subsequence of $\{\hat{\xi}_n\}$, there is another subsequence that belongs to $M_S(\xi)$ a.s. Assume that x_p is an interior point such that $x_{k+1} - x_k = O(1)$. It follows from (24) that

$$(\hat{\omega}^{\mathrm{T}}\hat{\Omega}^{-1} - \omega_n^{\mathrm{T}}\Omega_n^{-1})\mathbb{E} = \epsilon_k[h(\hat{\theta}_n, a_n, x_p - x_k, x_{k+1} - x_p) - h(\theta_n, a_n, x_p - x_k, x_{k+1} - x_p)] + \epsilon_{k+1}[h(\hat{\theta}_n, a_n, x_{k+1} - x_p, x_p - x_k) - h(\theta_n, a_n, x_{k+1} - x_p, x_p - x_k)], \quad (31)$$

where $h(\cdot)$ represents the continuous function

$$h(\theta, a, s, t) = \frac{\mathrm{e}^{-\theta s}(1 - \mathrm{e}^{-2\theta t})}{1 - \mathrm{e}^{-2\theta/a}}.$$

Note that $(a_n, x_{k+1} - x_p, x_p - x_k)$ are bounded, and that $(\epsilon_{k+1}, \epsilon_k) = O_p(1)$. This implies that (31) converges in probability to zero if $\{\xi_n\} \in M_S(\xi)$, since a continuous function always is uniformly continuous on a compact set. A similar argument shows that

$$\tau(\hat{\xi}_n) - \tau(\xi_n) \xrightarrow{P} 0 \quad \text{as } n \to \infty,$$

if $\{\xi_n\} \in M_S(\xi)$. Knowing that $\hat{\mu}_E = O_p(n^{-1/2})$, it follows from (23), that

$$(1 - \hat{\omega}^{\mathrm{T}}\hat{\Omega}^{-1}\mathbf{1})\hat{\mu}_E \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

Because $0 < \tau_n(\xi_n) = O(1)$, together with the above, we have that $B_n^{\mathrm{C}}(\xi_n) = \tau_n(\hat{\xi}_n)/\tau_n(\xi_n) \to 1$ and $A_n(\xi_n) \to 0$ in $P_{\xi,n}$-probability for $\{\xi_n\} \in M_S(\xi)$. It follows from (26), with $\mathbb{E} = \mathbf{1}$, that $\hat{\sigma}^2\mathbf{1}^{\mathrm{T}}\hat{\Omega}^{-1}\mathbf{1} \geqslant 1 + n(1 - \mathrm{e}^{-\hat{\theta}\delta})/2$, and hence,

$$(\mathbf{1}^{\mathrm{T}}\hat{\Omega}^{-1}\hat{\omega} - 1)^2/\mathbf{1}^{\mathrm{T}}\hat{\Omega}^{-1}\mathbf{1} \xrightarrow{P} 0 \quad \text{as } n \to \infty. \quad (32)$$

Thus, $B_n(\xi_n) = v(\hat{\xi}_n)/\tau(\xi_n)$ converges to 1 in $P_{\xi,n}$-probability. Therefore, (11)–(15) hold for $I_\gamma^{\mathrm{C}}(\hat{\xi}_n)$ as well as for $I_\gamma(\hat{\xi}_n)$. Also in this case similar reasoning shows that the plugin prediction intervals and the bootstrap calibrated prediction intervals have asymptotically correct coverage if x_p is not an interior point.

7.2. A two-dimensional example

Consider a stationary Gaussian process on $\mathbb{R}^2$, $\{Z(x), x\in\mathbb{R}^2\}$, with mean zero and covariance function

$$C_\xi(t) = \mathrm{Cov}[Z(x+t), Z(x)] = \sigma^2\,\mathrm{e}^{-\lambda|t_1| - \theta|t_2|}, \quad \sigma^2, \theta, \lambda > 0, \quad (33)$$

where $\xi = (\sigma^2, \theta, \lambda)^{\mathrm{T}}, t = (t_1, t_2)^{\mathrm{T}}$, and $t, x \in \mathbb{R}^2$. Suppose that the process is observed on a lattice, at locations $x_{ik} = (u_i, v_k)^{\mathrm{T}}$, $i = 1, \ldots, m$, $k = 1, \ldots, n$. This gives us the nm observation vector $\mathbb{Z} = \mathbb{Z}_{nm} = [\mathbf{Z}_1^{\mathrm{T}}, \ldots, \mathbf{Z}_m^{\mathrm{T}}]^{\mathrm{T}}$, where $\mathbf{Z}_i^{\mathrm{T}} = [Z_{i1}, \ldots, Z_{in}]$ and $Z_{ik} = Z(u_i, v_k)$. Without loss of generality, both $\{u_i\}$ and $\{v_k\}$ are arranged in ascending order. The design need not be nested, i.e. it is not assumed that $\mathbb{Z}_{nm} \subset \mathbb{Z}_{n'm'}$, where $n \leqslant n'$ and $m \leqslant m'$. The coverage of the plugin (bootstrap calibrated) prediction intervals of the ordinary kriging predictor at an interior point $x_{pp} = (u_p, v_p)$ will be studied under infill asymptotics such that

$$\max_i\{\Delta_i = u_i - u_{i-1}\} = O(m^{-1}), \quad \max_k\{\zeta_k = v_k - v_{k-1}\} = O(n^{-1}),$$

and $n/m \to \rho < \infty$, as $n, m \to \infty$. Define $s = s(m)$ and $t = t(n)$ to be integer values satisfying $u_s < u_p < u_{s+1}$, and $v_t < v_p < v_{t+1}$. The covariance matrix of $\mathbb{Z}$ is

$$\Omega = \mathrm{Cov}[\mathbb{Z}, \mathbb{Z}] = \sigma^2 A(\lambda) \otimes B(\theta),$$

where $A(\lambda) = \{\mathrm{e}^{-\lambda|u_i - u_j|}\}_{1 \leqslant i,j \leqslant m}$, $B(\theta) = \{\mathrm{e}^{-\theta|v_k - v_l|}\}_{1 \leqslant k,l \leqslant n}$, and $\otimes$ denotes the Kronecker product. We further have that

$$\omega = \mathrm{Cov}[\mathbb{Z}, Z(x_{pp})] = \sigma^2 a(\lambda) \otimes b(\theta),$$

where $a(\lambda) = \{e^{-\lambda|u_i - u_p|}\}_{1 \leqslant i \leqslant m}$, and $b(\theta) = \{e^{-\theta|v_k - v_p|}\}_{1 \leqslant k \leqslant n}$. Using the notation $a_m = e^{-\lambda(u_p - u_s)}$, $b_m = e^{-\lambda(u_{s+1} - u_p)}$, $c_n = e^{-\theta(v_p - v_t)}$, and $d_n = e^{-\theta(v_{t+1} - v_p)}$, from Lemma 1 in Ying (1993) we have that

$$\mathbb{Z}^{\mathrm{T}} \Omega^{-1} \omega = \frac{\left[a_m(1 - b_m^2)\mathbf{Z}_s^{\mathrm{T}} + b_m(1 - a_m^2)\mathbf{Z}_{s+1}^{\mathrm{T}}\right] B(\theta)^{-1} b(\theta)}{1 - e^{-2\lambda\Delta_{s+1}}}, \tag{34}$$

$$\sigma^2 \mathbb{Z}^{\mathrm{T}} \Omega^{-1} \mathbf{1} = \left(\mathbf{Z}_1^{\mathrm{T}} + \sum_{i=2}^{m} \frac{\mathbf{Z}_i^{\mathrm{T}} - e^{-\lambda\Delta_i} \mathbf{Z}_{i-1}^{\mathrm{T}}}{1 + e^{-\lambda\Delta_i}} \right) B(\theta)^{-1} \mathbf{1}, \tag{35}$$

$$\mathbf{1}^{\mathrm{T}} \Omega^{-1} \omega = \frac{(c_n + d_n)(a_m + b_m)}{(1 + e^{-\theta\zeta_{t+1}})(1 + e^{-\lambda\Delta_{s+1}})}, \tag{36}$$

and

$$\omega^{\mathrm{T}} \Omega^{-1} \omega = \sigma^2 \left(\frac{a_m^2(1 - b_m^2) + b_m^2(1 - a_m^2)}{(1 - e^{-2\lambda\Delta_{s+1}})} \right) \left(\frac{c_n^2(1 - d_n^2) + d_n^2(1 - c_n^2)}{1 - e^{-2\theta\zeta_{t+1}}} \right). \tag{37}$$

Let Ξ_0 be a compact set in $\mathbb{R}_+^3$ that contains the true parameters ξ as an interior point. The ML estimator of ξ over Ξ_0 is location invariant, strongly consistent and asymptotically normally distributed such that $(\hat{\xi}_n - \xi) = O_p(n^{-1/2})$ (see Ying, 1993). Moreover, van der Vaart (1996) showed the Local Asymptotic Normality (LAN) property for this model which implies that $\{P_{\xi_n,n}\}$ is contiguous with respect to $\{P_{\xi,n}\}$ for all $\{\xi_n\} \in M_S(\xi) = \{\{\xi_n \in \Xi_0\} : (\xi_n - \xi) = O(n^{-1/2})\}$ (see Bickel *et al.* 1993, pp. 16–17). Hence, for each subsequence of $\{\hat{\xi}_n\}$, there is another subsequence that belongs to $M_S(\xi)$ a.s. We have by Taylor expansions of (34), noting the close relationship between $\mathbf{Z}_s^{\mathrm{T}} B(\theta)^{-1} b(\theta)$ and (24), that

$$\mathbb{Z}^{\mathrm{T}}[\hat{\Omega}^{-1}\hat{\omega} - \Omega_n^{-1}\omega_n] = Z_{st}O(n^{-1}) + Z_{s,t+1}O(n^{-1}) + Z_{s+1,t}O(n^{-1}) + Z_{s+1,t+1}O(n^{-1}).$$

By inserting $\mathbb{Z} = \mathbf{1}$, in (35) and comparing with (29) we conclude that $\hat{\sigma}_n^2 \mathbf{1}^{\mathrm{T}} \hat{\Omega}^{-1} \mathbf{1} \geqslant 1$. In order to confirm that $\hat{\mu} = O_p(1)$ it is therefore sufficient to show that $\mathbf{Z}_i^{\mathrm{T}} B(\hat{\theta}_n)^{-1} \mathbf{1} = O_p(1)$, which then by similar reasoning implies that $\hat{\sigma}_n^2 \mathbb{Z}^{\mathrm{T}} \hat{\Omega}^{-1} \mathbf{1} = O_p(1)$. By Lemma 1 in Ying (1993) we have

$$\mathbf{Z}_i^{\mathrm{T}} B(\hat{\theta}_n)^{-1} \mathbf{1} = Z_{i1} + \sum_{k=2}^{n} \frac{(Z_{ik} - e^{-\hat{\theta}_n\zeta_k} Z_{i,k-1})}{1 + e^{-\hat{\theta}_n\zeta_k}} = Z_{i1} + \frac{1}{2} \sum_{k=2}^{n} (Z_k - e^{-\theta\zeta_k} Z_{k-1})$$
$$+ \frac{1}{2} \sum_{k=2}^{n} (e^{-\theta\zeta_k} - e^{-\hat{\theta}_n\zeta_k}) Z_{k-1} + \frac{1}{2} \sum_{k=2}^{n} (Z_{ik} - e^{-\hat{\theta}_n\zeta_k} Z_{i,k-1}) O(\zeta_\kappa).$$

Each of the three sums above has bounded variance, and thus $\mathbf{Z}_i^{\mathrm{T}} B(\hat{\theta}_n)^{-1} \mathbf{1} = O_p(1)$. Further, Taylor expansions of (36) and (37) yield, respectively, that $1 - \mathbf{1}^{\mathrm{T}} \hat{\Omega}^{-1} \hat{\omega} = O(n^{-2})$, and

$$\tau^2(\xi) = \frac{\sigma^2 \lambda f(\Delta_{s+1}) + \sigma^2 \theta g(\zeta_{t+1}) + O(n^{-2})}{(1 - \lambda\Delta_{s+1} + O(n^{-2}))(1 - \theta\zeta_{t+1} + O(n^{-2}))} = O(n^{-1}), \tag{38}$$

where

$$f(\Delta_{s+1}) = [\Delta_{s+1}^2 - (u_p - u_s)^2 - (u_{s+1} - u_p)^2]/\Delta_{s+1} = O(n^{-1}),$$

and

$$g(\zeta_{t+1}) = [\zeta_{t+1}^2 - (v_p - v_t)^2 - (v_{t+1} - v_p)^2]/\zeta_{t+1} = O(n^{-1}).$$

All this implies that $A_n(\xi_n) = O(n^{-1/2})$ in $P_{\xi,n}$-probability for all $\{\xi_n\} \in M_S(\xi)$. We further have from (38) and the consistency of $\hat{\xi}_n$, that $B_n^{\mathrm{C}}(\xi_n) \to 1$ in $P_{\xi,n}$-probability if $\xi_n \in M_S(\xi)$.

Also, from the above, (32) divided by $\tau(\xi_n)$ tends to zero a.s. at the rate $O(n^{-7/2})$ which implies that $B_n(\xi_n) \to 1$ in $P_{\xi,n}$-probability. Thus, the correct asymptotic coverage for plugin and bootstrap calibrated intervals for $I_\gamma^C(\hat{\xi}_n)$ as well as $I_\gamma(\hat{\xi}_n)$ is verified.

8. Numerical illustrations

In this section, we provide three numerical illustrations of the coverage properties of plugin and bootstrap calibrated prediction intervals of nominal 90 per cent coverage, for a stationary mean zero Gaussian process. In each case the coverages were estimated by Monte Carlo simulation, for a range of sample sizes n. Parameters were estimated by maximum likelihood, in each case the mean of the process being treated as a nuisance parameter and estimated together with the covariance function parameters. Bootstrap calibration was carried out by estimating coverages at three nominal coverage levels, 0.90, 0.95 and 0.99, with the recalibrated level estimated to yield the nominal desired coverage 0.90 being obtained by simple quadratic interpolation.

Our first example, illustrated graphically in Fig. 1, concerns the one-dimensional example considered in Section 7.1. The covariance function (22) was specified, with parameter values $\sigma^2 = \theta = 1$. The process was observed at sampling locations $x_i = i/a_n$, $i = 1, \ldots, n$, for a range of sample sizes n. An increasing domain asymptotic regime was specified by taking $a_n = 1$, with the process being predicted at 0. Results are shown in Fig. 1(a). The reduction of coverage error with increasing sample size n, as well as the benefits of bootstrap calibration, are evident. An infill asymptotic regime was specified by taking $a_n = n$, with the process now being predicted at $1/\sqrt{2}$. Results for this regime are shown in Fig. 1(b). The plugin interval is seen to yield low coverage error for small sample sizes n, and the benefits of calibration are, contrary to theoretical expectation, not realised for this particular regime. In both the increasing domain and infill cases, coverage properties of the two prediction intervals were simulated from 10,000 Monte Carlo replications, with $B = 500$ bootstrap samples being drawn for each bootstrap calibration.

Our second example concerns a two-dimensional stationary process of mean zero, and with covariance function of the form (33), and true parameter values $\sigma^2 = \lambda = \theta = 1$. The process

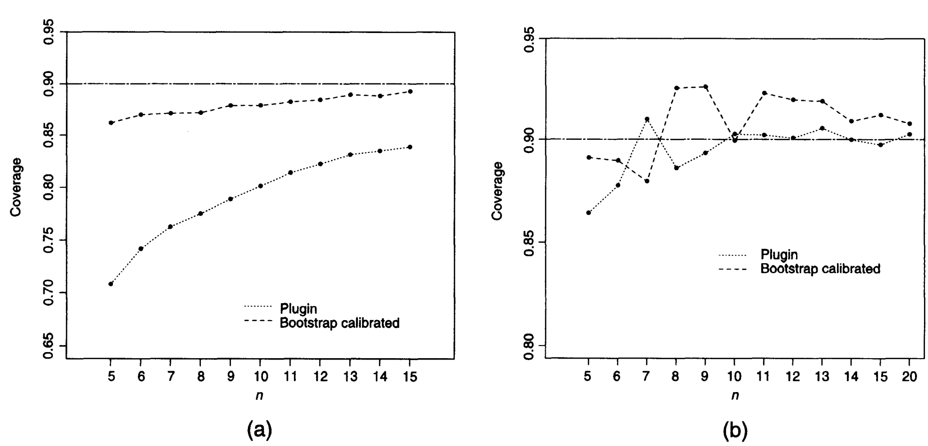

Fig. 1. Empirical coverage, as a function of sample size n, of 90 per cent equal-tailed prediction intervals obtained for the one-dimensional example in Section 8, under: (a) increasing domain asymptotics; and (b) infill asymptotics.

was observed at a regular lattice of sampling locations $x_{ik} = [(i - 1/n - 1), (k - 1/n - 1)]$, $i = 1, \dots, n$; $k = 1, \dots, n$, the effects of infill asymptotics being studied by increasing n, and the process was predicted at $(1/\sqrt{2}, 1/\sqrt{2})$. Coverage figures for this case were approximated from 2000 Monte Carlo replications at each value of n, with $B = 200$ bootstrap samples being drawn for each bootstrap calibration. Results are presented in Fig. 2. Again, the effectiveness of bootstrap calibration in eliminating coverage error is clear.

Our final example concerns a two-dimensional stationary process of mean zero, and with isotropic covariance function

$$C(t) = \mathrm{Cov}[Z(x+t), Z(x)] = \sigma^2 \mathrm{e}^{-\theta \|t\|},$$

where $t = (t_1, t_2)^{\mathrm{T}}$, with $x, t \in \mathbb{R}^2$, and $\|t\|$ denotes the Euclidean length of t. In the simulations we specified $\sigma^2 = \theta = 1$. The process was observed at random sampling locations $x_i = b_n U_i$, $i = 1, \dots, n$, where $U_1, \dots, U_n$ are independent, identically distributed on the unit disc $\{t : \|t\| < 1\}$, and predicted at (0,0). As the sample size n was varied, nested sampling locations were used, so that, for example, the sampling locations for $n = 10$ consisted of the sampling locations for $n = 9$, together with another uniformly random sampling location. An infill asymptotic regime was specified by taking $b_n = 1$, while an increasing domain regime was specified by taking $b_n = n^{1/2}$. Coverage figures were approximated from 2000 Monte Carlo replications, with $B = 200$ bootstrap samples being used for the calibration. Results are presented in Figs 3(a) and (b) for the increasing domain and infill cases, respectively. Again the effectiveness of bootstrap calibration is striking.

9. Final remarks

In this paper, we have studied plugin and bootstrap calibrated prediction intervals for kriging predictors. It is shown that, in some generality, both intervals give the correct coverage asymptotically. We have focused our attention on Gaussian random fields. It would be of

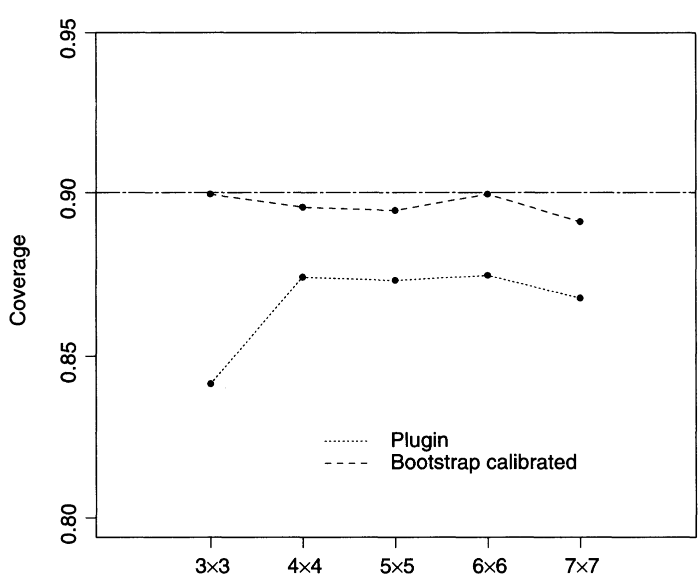

Fig. 2. Empirical coverage, as a function of sample size $n \times n$, of 90 per cent equal-tailed prediction intervals obtained for the two-dimensional lattice example in Section 8, under infill asymptotics.

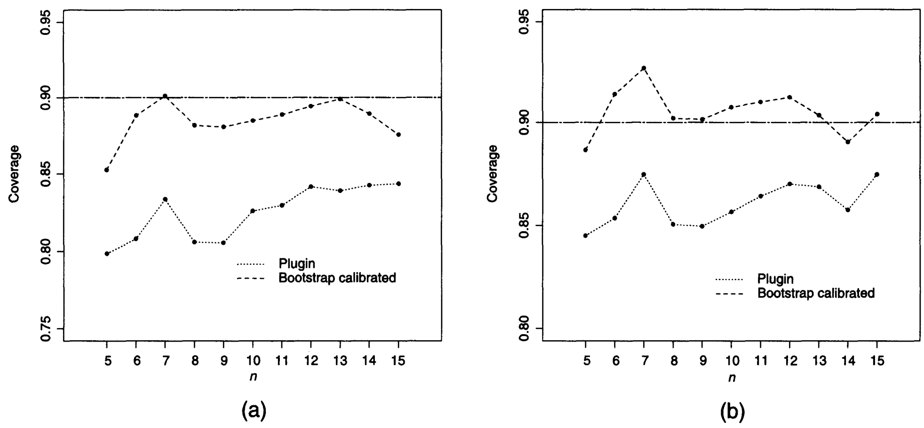

Fig. 3. Empirical coverage, as a function of sample size n, of 90 per cent equal-tailed prediction intervals obtained for the two-dimensional example with isotropic covariance function in Section 8, under: (a) increasing domain asymptotics; and (b) infill asymptotics.

interest to study whether the same type of results remain valid if distributional assumptions are avoided and some form of non-parametric block bootstrap is used instead of the parametric bootstrap in the calibration. However, we doubt that similar infill asymptotics results can be shown for the non-parametric bootstrap: infill sampling implies a long range dependence structure in the data, for which block bootstrap methods are known to perform poorly (cf. Lahiri, 1993).

Acknowledgements

Research was funded by the EU TMR network on Computational and Statistical Aspects of the Analysis of Spatial data (ERB-FMRX-CT96-0095). The authors are grateful to Professor L. Bondesson and the reviewers for helpful comments.

References

Barndorff-Nielsen, O. E. & Cox, D. R. (1994). *Inference and asymptotics.* Chapman & Hall, London.

Beran, R. (1990a). Refining bootstrap simultaneous confidence sets. *J. Amer. Statist. Assoc.* **85**, 417–428.

Beran, R. (1990b). Calibrating prediction regions. *J. Amer. Statist. Assoc.* **85**, 715–723.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. & Wellner, J. A. (1993). *Efficient and adaptive estimation for semiparametric models.* Johns Hopkins University Press, Baltimore, MD.

Cressie, N. A. C. (1993). *Statistics for spatial data.* Wiley, New York.

Hall, P., Peng, L. & Tajvidi, N. (1999). On prediction limits based on predictive likelihood or bootstrap methods. *Biometrika* **86**, 871–880.

Lahiri, S. N. (1993). On the moving block bootstrap under long range dependence. *Statist. Probab. Lett.* **18**, 405–413.

Loh, W.-Y. (1987). Calibrating confidence coefficients. *J. Amer. Statist. Assoc.* **82**, 155–162.

Loh, W.-Y. (1988). Contribution to the discussion of P. Hall's paper. *Ann. Statist.* **16**, 972–976.

Mardia, K. V. & Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135–146.

Putter, H. & Young, G. A. (2001). On the effect of covariance function estimation on the accuracy of kriging predictors. *Bernoulli* **7**, 421–438.

Roussas, G. G. (1972). *Contiguity of probability measures: some applications in statistics.* Cambridge University Press, Cambridge.

Sjöstedt-de Luna, S. (2002). Some properties of a maximum likelihood estimator with data from a Gaussian process. Research Report 2002-1, Department of Mathematical Statistics, Umeå University, Sweden.

van der Vaart, A. (1996). Maximum likelihood estimation under a spatial sampling scheme. *Ann. Statist.* **24**, 2049–2057.

Ying, Z. (1991). Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *J. Multivariate Anal.* **36**, 280–296.

Ying, Z. (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme. *Ann. Statist.* **21**, 1567–1590.

Received June 2001, in final form March 2002

Sara Sjöstedt-de Luna, Department of Mathematical Statistics, Umeå University, SE-901 87 Umeå, Sweden.
E-mail: sara@matstat.umu.se

Appendix

Proof of Lemma 1. The contiguity and (10) imply that $A_n(\xi_n) \to 0$ and $B_n(\xi_n) \to 1$ in $P_{\xi_n,n}$-probability. Since G is continuous in all its arguments it follows that $G(\gamma, A_n, B_n) \to G(\gamma, 0, 1) = 1 - 2\gamma$ in $P_{\xi_n,n}$-probability, that is, (11) holds. By dominated convergence it then follows that (12) holds. Further, note that $\pi_n(\gamma, \xi_n)$ is continuous in γ for each n, motivated by the following reasoning. We have that

$$\pi_n(\gamma, \xi_n) = E_{\xi_n}[G(\gamma, A_n, B_n)] = \int_{\mathbb{R}^n} H(\gamma, \xi_n, \mathbb{E})\mathrm{d}\mathbb{E},$$

where $H(\gamma, \xi_n, \mathbb{E}) = G(\gamma, A_n(\xi_n, \mathbb{E}), B_n(\xi_n, \mathbb{E}))f(\mathbb{E}, \xi_n)$ and $f(\mathbb{E}, \xi_n)$ is the normal density of $\mathbb{E}$ with mean zero and covariance parameters ξ_n. Let $\{\gamma_m\}_{m \geqslant 1}$ be a sequence converging to γ. The continuity of H in γ guarantees that $H(\gamma_m, \xi_n, \mathbb{E}) \to H(\gamma, \xi_n, \mathbb{E})$ as $m \to \infty$ pointwise for all $\mathbb{E} \in \mathbb{R}^n$ and $\xi_n \in \Xi$. Further $0 \leqslant H(\gamma_m, \xi_n, \mathbb{E}) \leqslant f(\mathbb{E}, \xi_n)$ for all m and $f(\mathbb{E}, \xi_n)$ is integrable. Then, by dominated convergence, $\pi_n(\gamma_m, \xi_n) \to \pi_n(\gamma, \xi_n)$ as $m \to \infty$, and hence π_n is continuous in γ. Because $G(\gamma, A_n, B_n)$ is decreasing in γ it follows that π_n is decreasing in γ as well. Since (12) holds for each $\gamma \in (0, 0.5)$ it ensures that for any $\{\xi_n\} \in M(\xi)$, the equation

$$[\pi_n(\pi_n^{-1}(1 - 2\gamma, \xi_n), \xi_n) - 1 + 2\pi_n^{-1}(1 - 2\gamma, \xi_n)]/2 = \pi_n^{-1}(1 - 2\gamma, \xi_n) - \gamma,$$

tends to zero as $n \to \infty$.

Proof of Lemma 2. Suppose that (14) does not hold. Then there is a subsequence $\{n_1(k)\}$ and $\epsilon, \delta > 0$ such that

$$P_\xi[|G(\pi_{n_1(k)}^{-1}(1 - 2\gamma, \hat{\xi}_{n_1(k)}), A_{n_1(k)}(\xi_{n_1(k)}), B_{n_1(k)}(\xi_{n_1(k)})) - 1 + 2\gamma| > \epsilon] > \delta, \tag{39}$$

for some sequence $\{\xi_n\} \in M_S(\xi)$. By going to a further subsequence $\{n_2(k)\} \subset \{n_1(k)\}$ we have from the assumptions that $\{P_{\hat{\xi}_n,n}\}$ is contiguous with respect to $\{P_{\xi,n}\}$ a.s. for that subsequence. This implies that $\pi_{n_2(k)}^{-1}(1 - 2\gamma, \hat{\xi}_{n_2(k)}) \to \gamma$ a.s., as $k \to \infty$. From (10) and knowing that $\{\xi_{n_2(k)}\} \in M_S(\xi)$ it follows that $A_{n_2(k)}(\xi_{n_2(k)}) \to 0$ and $B_{n_2(k)}(\xi_{n_2(k)}) \to 1$ in $P_{\xi,n}$-probability. Therefore, it is possible to find a further subsequence $\{n_3(k)\} \subset \{n_2(k)\}$ such that $A_{n_3(k)}(\xi_{n_3(k)}) \stackrel{\text{a.s.}}{\to} 0$ and $B_{n_3(k)}(\xi_{n_3(k)}) \stackrel{\text{a.s.}}{\to} 1$ as $k \to \infty$. From this and the assumed properties of G it follows that

$$G(\pi_{n_3(k)}^{-1}(1 - 2\gamma, \hat{\xi}_{n_3(k)}), A_{n_3(k)}, B_{n_3(k)}) \to G(\gamma, 0, 1) = 1 - 2\gamma \text{ a.s.},$$

which contradicts (39). Hence (14) holds. By dominated convergence it then follows that (15) holds.

Let $G^{(i,j,k)}(\gamma, a, b)$ denote a partial derivative of $G(\gamma, a, b)$.

Proof of Lemma 3. Taylor expansions of G over (A_n, B_n) around the point $(0, 1)$, where $A_n = A_n(\xi)$ and $B_n = B_n(\xi)$ yield

$$\begin{aligned} P_\xi[Z(x_p) \in I_\gamma(\hat{\xi}_n)|\mathbb{E}] = G(\gamma, A_n, B_n) &= 1 - 2\gamma + 2z_\gamma\varphi(z_\gamma)(B_n - 1) \\ &- z_\gamma\varphi(z_\gamma)A_n^2 - z_\gamma^3\varphi(z_\gamma)(B_n - 1)^2 + A_n^2[G^{(0,2,0)}(\gamma, \bar{A}_n, \bar{B}_n) \\ &- G^{(0,2,0)}(\gamma, 0, 1)]/2 + A_n(B_n - 1)G^{(0,1,1)}(\gamma, \bar{A}_n, \bar{B}_n) \\ &+ (B_n - 1)^2[G^{(0,0,2)}(\gamma, \bar{A}_n, \bar{B}_n) - G^{(0,0,2)}(\gamma, 0, 1)]/2. \end{aligned} \tag{40}$$

Here, $\bar{A}_n$ is between A_n and 0, and $\bar{B}_n$ is between B_n and 1. The second moments in (A1) and the continuity of the derivatives of G entail that

$$nA_n^2[G^{(0,2,0)}(\gamma, \bar{A}_n, \bar{B}_n) - G^{(0,2,0)}(\gamma, 0, 1)] \to 0 \quad \text{in } P_{\xi,n}\text{-probability.} \tag{41}$$

A uniform integrability argument, based on (A2) and the boundedness of $G^{(0,2,0)}$, shows that the convergence (41) also occurs in expectation. Similar arguments give

$$E_\xi[nA_n(B_n - 1)G^{(0,1,1)}(\gamma, \bar{A}_n, \bar{B}_n)] \to 0 \quad \text{as } n \to \infty,$$

and

$$E_\xi[n(B_n - 1)^2[G^{(0,0,2)}(\gamma, \bar{A}_n, \bar{B}_n) - G^{(0,0,2)}(\gamma, 0, 1)]] \to 0 \quad \text{as } n \to \infty.$$

Hence, conclusion (16) follows from the fact just established, by taking expectations over (40). By substituting $\pi_n^{-1}(1 - 2\gamma, \xi)$ for γ into (16) and rearranging terms, we have

$$\begin{aligned} \pi_n^{-1}(1 - 2\gamma, \xi) &= \gamma + b(\gamma, \xi)/2n + [b(\pi_n^{-1}(1 - 2\gamma, \xi), \xi) - b(\gamma, \xi)]/2n + o(n^{-1}) \\ &= \gamma + b(\gamma, \xi)/2n + o(n^{-1}), \end{aligned}$$

since $b(\gamma, \xi)$ is continuous in γ, which implies (17).

Proof of Lemma 4. Let $K_n(\gamma, \xi) = \pi_n^{-1}(1 - 2\gamma, \xi)$ and $K_{n0}(\gamma, \xi) = \gamma + b(\gamma, \xi)/2n$. Then

$$\pi_n(\pi_n^{-1}(1 - 2\gamma, \hat{\xi}_n), \xi) = E_\xi[G(K_n(\gamma, \hat{\xi}_n), A_n, B_n)].$$

By Taylor expansions around $K_{n0}(\gamma, \hat{\xi}_n)$, we have

$$G(K_n(\gamma, \hat{\xi}_n), A_n, B_n) = G(K_{n0}(\gamma, \hat{\xi}_n), A_n, B_n) + R_n(\gamma, \hat{\xi}_n)G^{(1,0,0)}(k_n, A_n, B_n),$$

where k_n is between $K_n(\gamma, \hat{\xi}_n)$ and $K_{n0}(\gamma, \hat{\xi}_n)$. Taking expectations and using (18) and the boundedness of $G^{(1,0,0)}$ we get

$$E_\xi[G(K_n(\gamma, \hat{\xi}_n), A_n, B_n)] = E_\xi[G(K_{n0}(\gamma, \hat{\xi}_n), A_n, B_n)] + o(n^{-1}).$$

We further have that

$$G(K_{n0}(\gamma, \hat{\xi}_n), A_n, B_n) = G(K_{n0}(\gamma, \xi), A_n, B_n) + \delta_n G^{(1,0,0)}(\bar{k}_n, A_n, B_n)/2n,$$

where $\bar{k}_n$ is between $K_{n0}(\gamma, \hat{\xi}_n)$ and $K_{n0}(\gamma, \xi)$. From (18) and the boundedness of $b(\gamma, \xi)$ and $G^{(1,0,0)}$, by dominated convergence, we have that

$$E_\xi[\delta_n G^{(1,0,0)}(\bar{k}_n, A_n, B_n)] \to 0 \quad \text{as } n \to \infty. \tag{42}$$

Taking expectations and using (42) we get

$$E_{\xi}[G(K_{n0}(\gamma, \hat{\xi}_n), A_n, B_n)] = E_{\xi}[G(K_{n0}(\gamma, \xi), A_n, B_n)] + o(n^{-1}).$$

Finally,

$$\begin{aligned} E_{\xi}[G(K_{n0}(\gamma, \xi), A_n, B_n)] &= E_{\xi}[G(K_n(\gamma, \xi), A_n, B_n)] - R_n(\gamma, \xi)E_{\xi}[G^{(1,0,0)}(\tilde{k}_n, A_n, B_n)] \\ &= 1 - 2\gamma + o(n^{-1}), \end{aligned}$$

by (17). Therefore, $\pi_n(\pi_n^{-1}(1 - 2\gamma, \hat{\xi}_n), \xi) = 1 - 2\gamma + o(n^{-1})$.