On the choice of correlation function and cross-validation for Gaussian processes

MUCM technical report, MS 3.3c.2

AP, JO

December 12, 2009

Contents

Chapter 1

On cross-validation for the analysis of computer experiments

In applications of the Gaussian processes, the predetermined parametric form of correlation function is typically used. The aim of this chapter is to show that cross-validation is an appropriate way to choose a correlation function among a given set of functions. Moreover, it is natural to simultaneously estimate the parameters of the Gaussian process via cross-validation. The efficiency of such estimation is shown by a number of examples for the analysis of computer experiments.

1.1 Introduction

Gaussian processes are widely used for prediction in machine learning (Rasmussen, Williams, 2006), geostatistics (Matheron, 1973, Ripley, 1981, Cressie, 1993), Bayesian analysis of computer experiments (Sacks et al., 1989, Kennedy, O'Hagan, 2001), and global optimization (Jones, 2001, Zhigljavsky, Zilinskas, 2008).

In these fields, measurements of a model under study are available at n different points and a meta-model, which predicts the output of the model for untried points, is of interest. This concern is motivated by the considerably time consuming evaluations of the model for a particular input, while the evaluation of the meta-model should be fast and the meta-model should serve as a replacement of the model. If an output of the model is deterministic, then the meta-model should interpolate the values of a given dataset. The Gaussian process framework supplies the constructing of a meta-model for an arbitrary black-box type model. This meta-model is called a nonparametric regression, and provides an elegant interpolation without specifying a basis of regression functions.

In the present chapter, the research is concentrated around an application of Gaussian processes for the analysis of computer experiments. During the last decade, the Bayesian analysis of deterministic computer models has been actively developing, see

Kennedy, O'Hagan, (2001), Paulo (2005), Bayarri et al. (2007), Bastos, O'Hagan (2009), Conti et al. (2009). In these references, the choice of the correlation function of the Gaussian process is a matter of belief. If there are no beliefs on the form of a correlation function, one is faced with the problem of its choice. It will be shown that the accuracy of a meta-model is sensitive to this choice, which can be done by the cross-validation (CV) technique (see e.g. Stone, 1974, Hastie et al., 2001).

In the Bayesian analysis of computer experiments, a so-called emulator is of particular importance. The emulator is a stochastic presentation of the meta-model and includes information on the uncertainty of prediction. The knowledge of the emulator allows one to compute the credible interval for the output of model at an untried point. The problem of constructing a good emulator is more difficult than the problem of constructing a meta-model. Note that for any specified values of parameters of the Gaussian process, the meta-model in the form of the conditional mean interpolates the values of a given dataset. However, the range of credible interval strongly depends on the parameters of Gaussian process. Thus, the determination of the parameters values is very important for probabilistic judgments about the output of model.

Different methods are proposed for the estimation of parameters of Gaussian processes, see Santner et al. (2003). Since the CV criterion is used for the choice of correlation function, it is natural to estimate parameters of Gaussian process via cross-validation (Stone, 1974, 1977, Cressie, 1993, p. 104).

In the present chapter, a specific way of cross-validation estimation is being used. First, the parameter of the mean of the Gaussian process is estimated by ordinary least squares and, after that, the parameter of the correlation function is estimated by the cross-validation technique. This two-stage estimation avoids making a very narrow credible interval such that the output of model lies outside the credible interval.

The present chapter is organized as follows. A procedure of CV estimation and criteria of goodness-of-fit for the emulator are introduced in Section 2. A broad numerical study is presented in Section 3.

1.2 Statement of the problem

In the analysis of computer experiments the output of complex mathematical model can be obtained at n input conditions and one is interested in predicting the output of model for untried inputs. According to the Gaussian process framework, we assume that the vector of output values of the model is a realization of a Gaussian process, that is, the vector has a multivariate normal distribution. As a result, a prediction is given by a posterior Gaussian process with the hyper-parameter to be estimated from the data.

There are several ways to carry out the estimation. One way is to use parameter estimators, which are commonly used in statistics, under the premeditated assumption

that the data set is a realization of a Gaussian process. For example, one can use a maximum likelihood estimator (Ripley, 1981), a restricted maximum likelihood estimator (Harville, 1974) a Bayesian estimator (Kennedy, O'Hagan, 2001) or a variogram estimator (Cressie, 1993). Another way of estimation is to use an indirect procedure, for example, the cross-validation (Hastie et al., 2001).

Let $y_1, \ldots, y_n$ be output values of the model $\eta(x)$ at points $x_1, \ldots, x_n$. Let a Gaussian process have a linear mean $m(x) = \beta^T h(x) = \beta_0 + \sum_{k=1}^{d} \beta_k x_{k'}$, a variance σ^2 and a correlation function $r(x, \tilde{x})$. One may consider the Gaussian correlation function

$$r(x, \tilde{x}|\psi) = \prod_{k=1}^{d} \exp\Big(- (x_{k'} - \tilde{x}_{k'})^2/\psi_k \Big),$$

the rational quadratic correlation function

$$r(x, \tilde{x}|\psi) = \prod_{k=1}^{d} \frac{1}{1 + (x_{k'} - \tilde{x}_{k'})^2/\psi_k}$$

and the Matern correlation function

$$r_\nu(x, \tilde{x}|\psi) = \prod_{k=1}^{d} \frac{(|x_{k'} - \tilde{x}_{k'}|/\psi_k)^\nu \mathcal{K}_\nu(|x_{k'} - \tilde{x}_{k'}|/\psi_k)}{2^{\nu-1}\Gamma(\nu)}$$

where $\mathcal{K}_\nu(d)$ is the modified Bessel function of the second kind and order ν, see Koehler, Owen (1996), Paulo (2005), Rasmussen, Williams, (2006). The Matern correlation function for $\nu - 1/2 \in \mathbb{N}$ has a simple algebraic form, for example, for $\nu = 7/2$ we have

$$r_{7/2}(x, \tilde{x}|\psi) = \prod_{k=1}^{d} \frac{1}{15}(s_k^3 + 6s_k^2 + 15s_k + 15)e^{-s_k}\Big|_{s_k = |x_{k'} - \tilde{x}_{k'}|/\psi_k}.$$

Gaussian, rational and Matern correlation functions are presented in Figure 1.1. We can see totally different shapes of correlation functions.

Let parameters of a Gaussian process be estimating by the Bayesian approach which was motivated in (Kennedy, O'Hagan, 2001) in the context of computer experiments. The Bayesian estimator is obtained by the integration out with respect to non-informative priors and the mean of the posterior Gaussian process for the given dataset is given by

$$m_p(x|\hat{\theta}, y) = \hat{\beta}^T h(x) + t(x)R^{-1}(y - H\hat{\beta}) \tag{1.1}$$

where

$$H = \big(h(x_1), \ldots, h(x_n)\big)^T,$$
$$R = \Big(r(x_i, x_j|\hat{\psi})\Big)_{i,j=1}^{n},$$

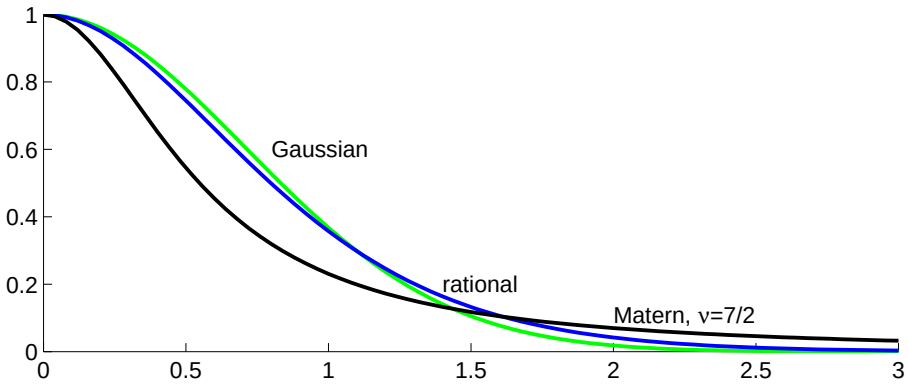

Figure 1.1: Gaussian, rational and Matern correlation functions.

$$t(x) = \left(r(x, x_1|\hat{\psi}), \ldots, r(x, x_n|\hat{\psi})\right)^T.$$

Note that the emulator is the posterior Gaussian process with the estimated hyper-parameter $\theta = (\beta, \sigma, \psi)$ and the meta-model is the its mean $m_p(x|\hat{\theta}, y)$. The credible interval is proportional to the variance of the posterior Gaussian process for which the covariance function is defined by

$$V_p(x, \tilde{x}) = \hat{\sigma}^2 \left(R(x, \tilde{x}|\hat{\psi}) - t^T(x)R^{-1}t(\tilde{x}) + s^T(x)(H^T R^{-1}H)^{-1}s(\tilde{x})\right)$$

where $s(x) = h(x) - H^T R^{-1}t(x)$.

In further numerical studies, the hyper-parameter of Gaussian process is estimated by a specific algorithm used in Bastos, O'Hagan (2009). Specifically, the parameter of the mean and the variance are found by integrating out the likelihood with respect to the prior density $p(\beta, \sigma^2) \propto \sigma^{-2}$ and the estimate of ψ is found by maximization of the posterior density.

1.2.1 Measures of goodness-of-fit for the emulator

In the analysis of computer experiments there are two objectives: the precise prediction and the convincing uncertainty.

The determination of the good prediction means finding a value of the hyper-parameter of the Gaussian process such that the meta-model output is as close as possible to the model output. This closeness can be measured by the root of mean square error

$$\text{RISE} = \left(\int_\Omega \left(m_p(x|\hat{\theta}, y) - \eta(x)\right)^2 \mathrm{d}x\right)^{1/2}$$

where the set Ω is typically the full design space (Sacks et al., 1989, Santner et al., 2003). Thus, we say that the meta-model with minimal RISE is optimal. Unfortunately, the RISE is impossible to compute in practice. Consequently, we need an optimality

criterion which is similar to RISE and depends only on the model output for tried inputs.

Note that the value of RISE has a similar functional form as the cross-validation error (1.2) (see e.g. Stone, 1974) if design points fill the design space uniformly. This similarity will be shown numerically by a number of examples in Section 3.

The meta-model has an appropriate uncertainty if the model output is located within the credible interval. For this characterization we introduce the ratio of volume of domain where this condition is hold

$$\text{RBCI} = \frac{\text{Volume}(x \in \Omega \,:\, \eta(x) \in 0.95\text{-credible interval})}{\text{Volume}(\Omega)}\,.$$

Also we define the averaged length of credible interval

$$\text{ALCI} = \int_{\Omega} V_p(x,x)\,\mathrm{d}x \Big/ \text{Volume}(\Omega)$$

where $V_p(x,x)$ is the variance of prediction at a point x.

1.2.2 Estimation of parameters via cross-validation

According to the cross-validation procedure (see e.g. Santner et al. 2003, p. 68), estimators $\hat{\beta}$ and $\hat{\psi}$ minimize

$$D(\beta,\psi) = \sum_{i=1}^{n} \left(m_p\left(x_i|\beta,\psi,y^{(i)}\right) - y_i \right)^2 \tag{1.2}$$

over $\beta \in \mathbb{R}$, $\psi \in (0,\infty)$ where $y^{(i)} = (y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_n)$. The function $m_p(x|\beta,\psi,y^{(i)})$ is the meta-model constructed with the ith point removed from the dataset. Note that the criterion (1.2) is an analog of the sum of squared deleted residuals in the regression analysis. The estimator of parameter σ^2 is given by $\hat{\sigma}^2 = (Y - H\hat{\beta})^T R^{-1}(Y - H\hat{\beta})/n$ where $R = R(\hat{\psi})$.

In the present chapter, it is proposed to modify the criterion (1.2) by splitting the estimation procedure into two stages: first, estimating β, second, estimating ψ. Namely, to estimate the parameter β by

$$\hat{\beta} = (H^T H)^{-1} H^T Y$$

and, after that, estimate the parameter ψ by minimizing

$$\mathcal{D}(\psi) = D(\hat{\beta},\psi)$$

over $\psi \in (0,\infty)$. If the minimum of $\mathcal{D}(\psi)$ is attained for $\psi = 0$ or for $\psi = \infty$, then we say that the Gaussian process with the given correlation function is not suitable for the data fitting.

Note that the utilizing of ordinary least squares leads to a larger value $\min \mathcal{D}(\psi)$ comparing with $\min D(\beta, \psi)$ (that is observed in tables in Section 3). However, in this way we try to avoid a very narrow credible interval such that the model output is outside the credible interval. In general, the ordinary least squares may be less efficient that other estimators but sometimes it yields better results, see tables in Section 3.

For a comparison with RISE we define the cross-validation error

$$\text{JRSE} = \sqrt{\mathcal{D}(\psi)}.$$

In the notation it is stressed that the Jack-knife procedure as a particular case of cross-validation is used. In the present context it is important to exclude only one point from the data.

Computational complexity of cross-validation

The computation of the cross-validation criterion (1.2) can be simplified by using a classic result for the inversion of block matrices. Let the inverse to the correlation matrix

$$R = \big(r(x_i, x_j)\big)_{i,j=1}^{n}$$

be computed for the whole dataset. Let $R_{(i)}$ be a correlation matrix for the dataset with removed ith point. Note that $R_{(i)}$ is obtained from R by removing ith column and ith row. Then the matrix $R_{(i)}^{-1}$ can be expressed via components of R^{-1} in a simple algebraic form. Indeed, divide the matrix R^{-1} into blocks

$$R^{-1} = \begin{pmatrix} Q & b \\ b^T & \kappa \end{pmatrix}$$

where κ is an (i,i)-element of matrix R^{-1} and b is i-th column of R^{-1} excluding i-th row. Then the matrix $R_{(i)}^{-1}$ can be computed by

$$R_{(i)}^{-1} = Q - \frac{bb^T}{\kappa}.$$

1.2.3 The choice of correlation function

Using the cross-validation criterion we obtain a simple rule for the choice of correlation function. Assume that we have two Gaussian processes with different correlation functions $r_1(x, \tilde{x})$ and $r_2(x, \tilde{x})$. Then we say that the Gaussian process with $r_1(x, \tilde{x})$ is more preferable than the Gaussian process with $r_2(x, \tilde{x})$ if $\mathcal{D}_{r_1} < \mathcal{D}_{r_2}$. In section 3, it is shown that the Gaussian process with $r_1(x, \tilde{x})$ has approximately a smaller value by the RISE criterion.

1.3 Numerical comparison of the emulators

1.3.1 One-dimensional case

Let us start considering examples with a one-dimensional case, $p = 1$, $x \in \mathbb{R}$, which allows a complete visualization of the emulator. Let the mean of the Gaussian process have a linear form, i.e. $h(x) = (1, x)^T$.

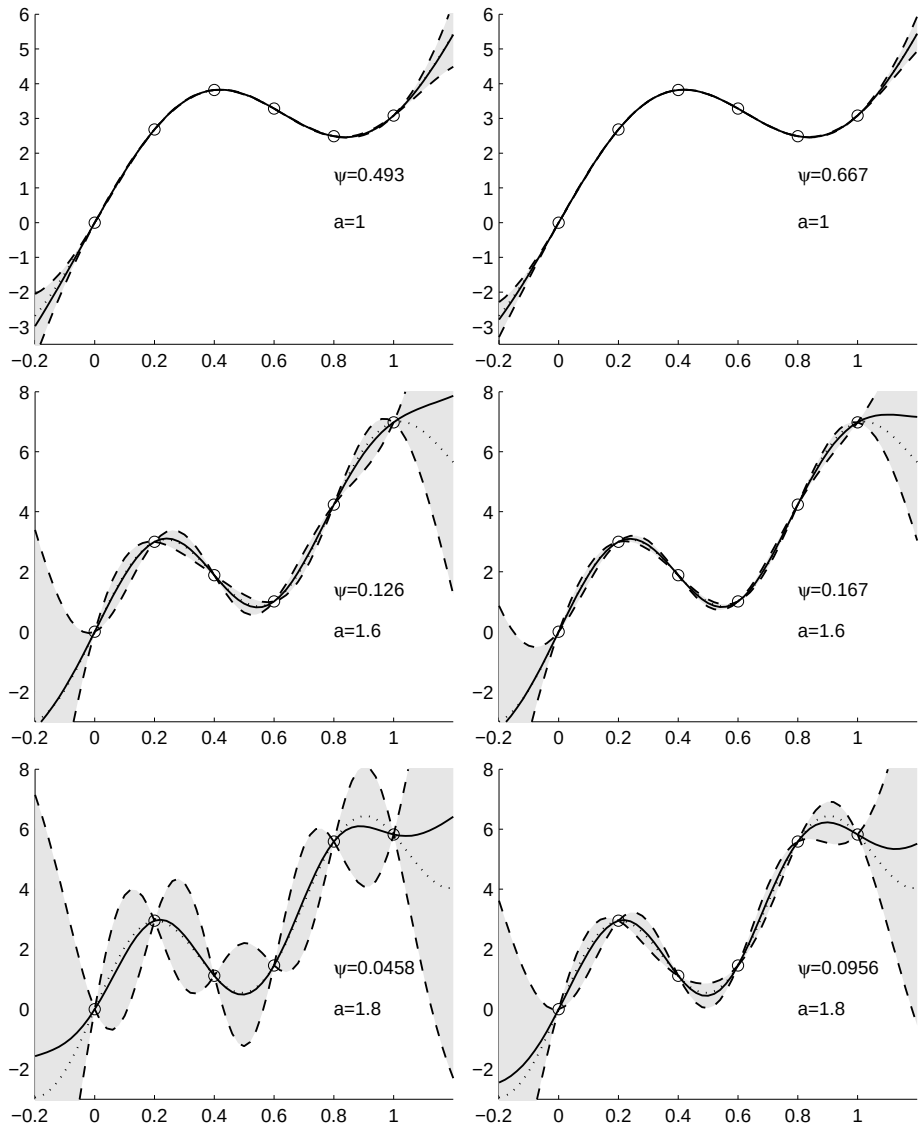

Figure 1.2: Emulators with the Gaussian correlation function using the Bayesian estimation (at left) and the cross-validation estimation (at right) for the model $\eta(x) = 5x + 2\sin(5ax)$ with $a = 1$ (top), $a = 1.6$ (middle) and $a = 1.8$ (bottom), and points $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$. The model output is given by the dotted line.

Emulators for three different models are depicted in Figure 1.2. The mean and credible interval of the emulator are given by solid and dashed lines, the model output

Table 1.1: *Values of MSE and RISE for the Bayesian and cross-validation estimation for the model* $\eta(x) = 5x + 2\sin(5ax)$ *and points* $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$.

	Bayesian			CV		
a	1	1.6	1.8	1	1.6	1.8
	Gaussian corr. function					
JRSE	0.1651	0.672	1.31	0.0509	0.619	1.23
RISE	0.0033	0.067	0.16	0.0013	0.045	0.11
	rational corr. function					
JRSE	0.174	1.003	1.41	0.098	0.962	1.76
RISE	0.006	0.092	0.39	0.003	0.063	0.19
	Matern corr. function, $\nu = 7/2$					
JRSE	0.344	0.982	1.36	0.248	0.981	—
RISE	0.012	0.095	0.24	0.010	0.088	—

is given by the dotted line. Values of JRSE and RISE are presented in Table 1.1. We see that RISE using the CV estimation is smaller than RISE using the Bayesian estimation and this almost holds for the JRSE criterion. Also we observe that the meta-model with the Gaussian correlation function is more accurate than the meta-model with the rational or Matern correlation functions. The empty cell in Table 1.1 for the Matern correlation function and $a = 1.8$ means that the CV estimator does not exist. We can see in Figure 1.2 that the credible interval using the CV estimation is more short than the credible interval using the Bayesian estimation, especially for $a = 1.8$, and the output of model belongs to the credible interval in all cases.

1.3.2 Two-dimensional case

Let us consider a two-dimensional test model which has a typical form and is determined by

$$\eta(x) = \left(1 - e^{-0.5/x_{2'}}\right) \frac{2300x_{1'}^3 + 1900x_{1'}^2 + 2092x_{1'} + 60}{100x_{1'}^3 + 500x_{1'}^2 + 4x_{1'} + 20} \tag{1.3}$$

where $x = (x_{1'}, x_{2'}) \in [0, 1]^2$. In particular, the model (1.3) has been studied by Bastos, O'Hagan, (2009). Let observations of the model be collected at points of 30-point maximin Latin hypercube design. For the sake of visualization, in Figure 1.3 the emulator is depicted with respect to one variable when another variable is fixed along with the model output. The mean of the Gaussian process was taken to be of a linear form, $h(x) = (1, x_{1'}, x_{2'})^T$.

Figure 1.3 shows that the meta-model using the CV estimation is more reliable than using the Bayesian estimation since the model output lies outside the credible interval. Values of RISE for sets $\Omega = [0, 1]^2$ and $\Omega = [0.05, 0.95]^2$ and JRSE are given

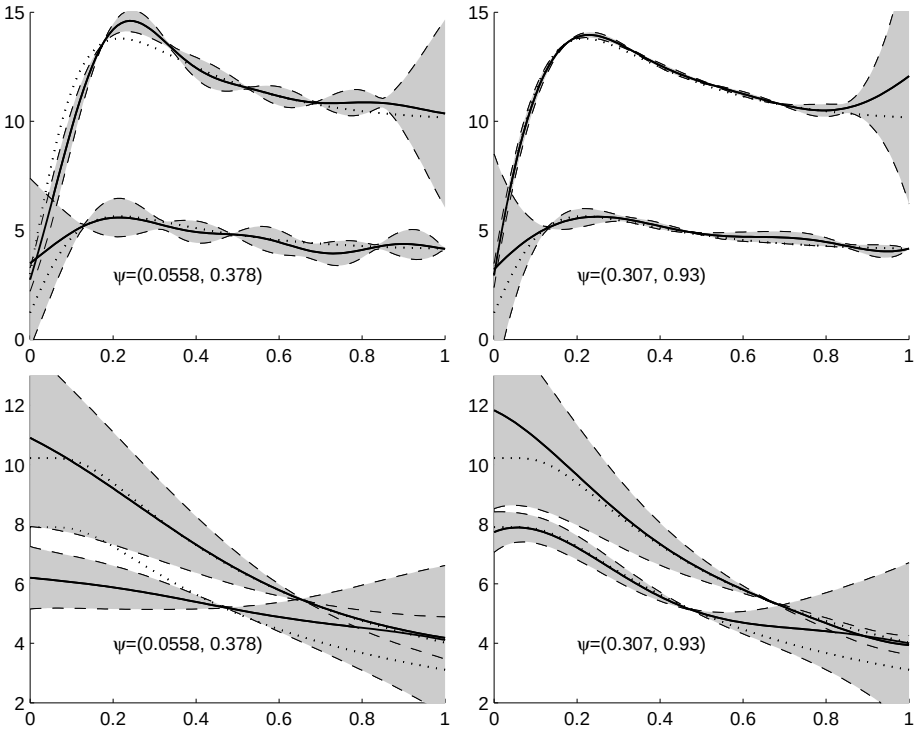

Figure 1.3: Emulators for the model (1.3) w.r.t. $x_{1'}$ with $x_{2'} = 0.05, 0.95$ (top), w.r.t. $x_{2'}$ with $x_{1'} = 0.05, 0.95$ (bottom) using the Bayesian estimation (at left) and the cross-validation estimation (at right). The model output is given by the dotted line.

Table 1.2: *Characteristics of emulators using the Bayesian and cross-validation estimation for the model* (1.3) *and the 30-point maximin LHD.*

corr. fun.	Gaussian		rational		Matern, $\nu = 7/2$	
	Bayesian	CV	Bayesian	CV	Bayesian	CV
JRSE	0.299	0.143	0.169	0.093	0.082	0.073
$\text{RISE}_{[0,1]^2}$	0.410	0.247	0.110	0.138	0.141	0.174
$\text{RISE}_{[0.05,0.95]^2}$	0.307	0.116	0.086	0.068	0.067	0.070
RBCI	0.93	0.91	0.96	0.93	0.99	0.97
ALCI	0.24	0.10	0.11	0.09	0.10	0.09

in Table 1.2. We see that values of JRSE using the CV estimation are smaller than using the Bayesian estimation. Also, the length of the credible interval is a little smaller but the ratio of domain where the model output is outside the credible interval is a little larger using the CV estimation than using the Bayesian estimation.

As a second two-dimensional example, consider the model

$$\eta(x) = 3x_{1'} + 3x_{2'} + e^{-25(x_{1'}-0.7)^2} + 1/(1 + 25(x_{2'} - 0.7)^2) \tag{1.4}$$

which has another curvature.

Table 1.3: *Characteristics of emulators using the Bayesian and cross-validation estimation for the model* (1.4) *and the 30-point maximin LHD.*

corr. fun.	Gaussian		rational		Matern, $\nu = 7/2$	
	Bayesian	CV	Bayesian	CV	Bayesian	CV
JRSE	0.108	0.126	0.072	0.070	0.078	0.074
$\text{RISE}_{[0,1]^2}$	0.064	0.056	0.043	0.036	0.051	0.045
$\text{RISE}_{[0.05,0.95]^2}$	0.042	0.038	0.040	0.034	0.048	0.042
RBCI	0.93	0.98	0.97	1.0	0.96	0.99
ALCI	0.031	0.035	0.027	0.032	0.029	0.032

Values of JRSE and RISE are given in Table 1.3 and they are close enough to each other. For the model (1.4) in contrast to the model (1.3) we observe that the length of the credible interval is a little smaller but the ratio of domain where the model output is outside the credible interval is a little larger using the Bayesian estimation than using the CV estimation.

It is interesting to note that RISE is essentially larger near the boundary of hypercube for the emulator with the Gaussian correlation function. The accuracy of emulators with the rational and Matern correlation functions is the same in all subdomains of the design space. The emulator with the rational correlation function has happened to be the most accurate for both models.

1.3.3 Borehole model

In this subsection, we examine the borehole model which was investigated in Worley (1987), Morris et al. (1993), Ahn, Owen (2001), Fang, Lin (2003). The borehole model determine the flow rate through the borehole that is drilled from the ground surface through two aquifers. The flow rate is given by

$$\eta_B(z) = \frac{2\pi(H_u - H_l)}{\ln(r/r_w)\left(1 + \frac{2LT_u}{\ln(r/r_w)r_w^2 K_w} + \frac{T_u}{T_l}\right)}$$

where $z = (r_w, r, T_u, T_l, H_u, H_l, L, K_w)$. The model has eight inputs. The input $r_w \in [0.05, 0.15]$ is the radius of the borehole, $r \in [100, 50000]$ is the radius of influence, $T_u \in [63070, 115600]$ is the transmissivity of the upper aquifer, $T_l \in [63.1, 116]$ is the transmissivity of the lower aquifer, $H_u \in [990, 1110]$ is the potentiometric head of the upper aquifer, $H_l \in [700, 820]$ is the potentiometric head of the lower aquifer, $L \in [1120, 1680]$ is the length of the borehole $K_w \in [9855, 12045]$ is the hydraulic conductivity of the borehole. For numerical calculations, consider a scaled model $\eta(x)$ defined on the unit hypercube by $\eta(x) = \eta_B(\overline{R}x + \underline{S})$, $x \in [0,1]^8$, $\underline{S}$ is a vector with lower bounds and $\overline{R}$ is a diagonal matrix with input ranges. Values of JRSE and RISE for the borehole meta-model are presented in Table 1.4. We observe that the emulator

using the CV estimation yields smaller RISE than the emulator using the Bayesian estimation, although in some cases, it yields a larger value of JRSE. This happens since the parameter of the mean of the Gaussian process is estimated by the ordinary least squares.

Table 1.4: *Characteristics of emulators using the Bayesian and cross-validation estimation for the Borehole model and the* 60- *and* 80-*point maximin LHDs.*

	60-point LHD		80-point LHD	
	Bayesian	CV	Bayesian	CV
	Gaussian corr. function			
JRSE	3.08	3.46	2.01	1.38
$\text{RISE}_{[0,1]^8}$	1.84	1.41	1.27	1.29
$\text{RISE}_{[0.05,0.95]^8}$	1.66	1.15	1.15	0.96
RBCI	0.99	0.99	0.99	0.99
ALCI	2.14	2.24	1.37	1.63
	rational corr. function			
JRSE	3.58	3.74	2.34	1.64
$\text{RISE}_{[0,1]^8}$	1.94	1.78	1.64	1.46
$\text{RISE}_{[0.05,0.95]^8}$	1.69	1.47	1.29	1.21
RBCI	0.99	0.99	0.99	0.99
ALCI	2.75	2.40	1.69	1.85
	Matern corr. function, $\nu = 7/2$			
JRSE	1.82	1.43	0.92	0.94
$\text{RISE}_{[0,1]^8}$	0.94	0.82	0.42	0.44
$\text{RISE}_{[0.05,0.95]^8}$	0.76	0.63	0.32	0.33
RBCI	0.97	0.98	0.99	0.99
ALCI	0.72	0.65	0.44	0.41

We observe that the meta-model with the Matern correlation function is essentially more accurate than for the Gaussian and rational correlation functions. Also, the CV estimation gives a more accurate meta-model.

1.3.4 Neddermeyer's model

In this subsection, we examine designs for the Neddermeyer's model which was studied in Neddermeyer (1943) and Higdon et al. (2008). The model describes the internal radius of steel cylinder during a high explosive (HE) charge implosion. The internal radius as a function of time is given by a solution of the differential equation

$$y' = \left[\frac{2}{R_1^2 f(y)^2}\left(\frac{v_0^2}{2} - \frac{s}{2\rho}g(y)\right)\right]^{-1/2}$$

where

$$f(y) = \frac{y^2}{(1-\lambda^2)} \ln \frac{y^2 + 1 - \lambda^2}{y^2},$$

$$g(y) = \frac{1}{1-\lambda^2}(y^2 + 1 - \lambda^2)\ln(y^2 + 1 - \lambda^2) - \lambda^2 \ln(\lambda^2),$$

$R_1 = 1.5$ is the initial outer radius of steel cylinder being imploded, $\lambda = 2/3$ is the ratio of outer cylinder radius to the inner radius, $\rho = 7.5$ is the specific density (relative to water), $v_0 = 0.3 \cdot 10^5$ is the initial velocity imparted on outer radius of cylinder from the HE.

Let us consider the internal radius after an implosion. Then the model has a form $\eta_N(z) = y(T)$ with three-dimensional input $z = (s, m, u_0)$, where $T = 5 \cdot 10^5$ is a time moment after an implosion, the input $s \in [1.85 \cdot 10^{10}, 2.25 \cdot 10^{10}]$ is the yield stress of steel, the input $m \in [0.29, 0.35]$ is the mass ratio between the HE and cylinder, the input $u_0 \in [1.4 \cdot 10^{10}, 1.7 \cdot 10^{10}]$ is the detonation energy per gram of exploded gas from the HE. Similarly to the subsection 3.3, we determine the model $\eta(x) = \eta_N(\overline{R}x + \underline{S})$ defined on the unit hypercube $[0, 1]^3$.

Table 1.5 shows that RISE for the emulator using the CV estimation is smaller than for the emulator using the Bayesian estimation. This relation holds for the JRSE criterion also. However, emulators have short credible intervals, consequently, the model output fewer belongs the credible interval. Note that emulators for all three correlation functions yield a similar accuracy.

1.4 Conclusions

A broad numerical study shows that the accuracy of the emulator strongly depends on the choice of a correlation function. In particular, the difference by the RISE and ALCI criteria can reach 2 times larger or even higher. The JRSE criterion based on the cross-validation technique does not guarantee the best possible choice of a correlation function. However, it can extricate the choice rather efficiently.

Note that the cross-validation technique has a restriction. This method is not efficient in the case where a data size is rather small and a complicated shape of a model response is poorly described by sparse observations.

Table 1.5: *Characteristics of emulators using the Bayesian and cross-validation estimation for the Neddermeyer's model and the* 30- *and* 50-*point maximin LHDs.*

	30-point LHD		50-point LHD	
	Bayesian	CV	Bayesian	CV
	Gaussian corr. function			
JRSE	1.41	1.21	1.80	1.32
$\text{RISE}_{[0,1]^3}$	1.78	1.43	1.19	0.94
$\text{RISE}_{[0.05,0.95]^3}$	0.98	0.83	0.76	0.73
RBCI	0.74	0.52	0.89	0.80
ALCI	0.28	0.17	0.45	0.40
	rational corr. function			
JRSE	0.97	0.93	1.39	1.02
$\text{RISE}_{[0,1]^3}$	1.57	1.45	1.41	1.27
$\text{RISE}_{[0.05,0.95]^3}$	0.90	0.87	0.82	0.86
RBCI	0.75	0.64	0.94	0.77
ALCI	0.27	0.22	0.49	0.37
	Matern corr. function, $\nu = 7/2$			
JRSE	0.92	0.69	1.36	1.01
$\text{RISE}_{[0,1]^3}$	1.45	1.26	1.37	1.20
$\text{RISE}_{[0.05,0.95]^3}$	0.86	0.85	0.81	0.79
RBCI	0.76	0.98	0.97	0.92
ALCI	0.26	1.21	0.55	0.54

Bibliography

[1] Ahn J., Owen A. (2001). Quasi-regression. Journal of Complexity, 17, 588–607.

[2] Bayarri M.J., Berger J.O., Cafeo J., Garcia-Donato G., Liu F., Palomo J., Parthasarathy R.J., Paulo R., Sacks J., Walsh D. (2007) Computer model validation with functional output. Ann. Statist. 35, no. 5, 1874–1906.

[3] Bastos L., O'Hagan A. (2009) Diagnostics for Gaussian process emulators. accepted to Technometrics.

[4] Conti S., Gosling J.P., Oakley J.E., O'Hagan A. (2009) Gaussian process emulation of dynamic computer codes. Biometrika 96, 663–676.

[5] Cressie N. A. C. (1993) Statistics for Spatial Data, John Wiley & Sons, New York.

[6] Jones D.R. (2001) A taxonomy of global optimization methods based on response surfaces. J. Global Optim. 21, no. 4, 345–383.

[7] Harville D.A. (1974) Bayesian inference for variance components using only error contrasts. Biometrika 61, 383–385.

[8] Hastie T., Tibshirani R., Friedman J. (2001) The elements of statistical learning. Data mining, inference, and prediction. Springer Series in Statistics. Springer-Verlag, New York.

[9] Higdon D., Gattiker J., Williams B., Rightley M. (2008). Computer model calibration using high-dimensional output. J. Amer. Stat. Ass., 103, 570–583.

[10] Fang K.-T., Lin D.K.J. (2003). Uniform experimental designs and their application in industry. In Handbook of Statistics, Ed. by R. Khattree and C.R. Rao, 131-170.

[11] Kennedy M.C., O'Hagan A. (2001) Bayesian calibration of computer models. J. R. Stat. Soc. Ser. B 63, no. 3, 425–464.

[12] Koehler J.R., Owen A.B. (1996) Computer experiments. In Handbook of Statistics, 261–308.

[13] Matheron G. (1973) The intrinsic random functions and their applications. Advances in Appl. Probability 5, 439–468.

[14] Morris M.D., Mitchell T.J. (1995). Exploratory designs for computer experiments J. Stat. Plan. Inf., 43, 381–402.

[15] Neddermeyer S. (1943). Collapse of hollow steel cylinders. Los Alamos Report, LA-18, Los Alamos National Laboratory.

[16] Paulo R. (2005) Default priors for Gaussian processes. Ann. Statist. 33, no. 2, 556–582.

[17] Rasmussen C.E., Williams C.K.I. (2006) Gaussian processes for machine learning. MIT Press, Cambridge.

[18] Ripley B.D. (1981) Spatial statistics. John Wiley & Sons, Inc., New York.

[19] Sacks J., Welch W.J., Mitchell T.J., Wynn H.P. (1989) Design and analysis of computer experiments. Statist. Sci. 4, no. 4, 409–435.

[20] Santner T.J., Williams B.J., Notz W. (2003) The Design and Analysis of Computer Experiments. Springer-Verlag, New York.

[21] Stone M. (1974) Cross-validatory choice and assessment of statistical predictions. J. Roy. Statist. Soc. Ser. B 36, 111–147.

[22] Stone M. (1977) An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. J. Roy. Statist. Soc. Ser. B 39, no. 1, 44–47.

[23] Worley B.A. (1987). Deterministic uncertainty analysis. OSTI ID: 6104504.

[24] Zhigljavsky A., Zilinskas A. (2008) Stochastic global optimization. Springer, New York.

Chapter 2

Fixed-domain asymptotics of maximum likelihood estimators for observations of deterministic models

A case when data is collected as precise observations of deterministic model is considered. Such data happens in the analysis of computer experiments and machine learning. It is shown that the maximum likelihood estimator of the correlation (roughness) parameter of a Gaussian process with the Matern correlation function tends to infinity (i.e. the correlation between any two points tends to 1) as the size of data increases.

2.1 Introduction

Fixed-domain asymptotics of maximum likelihood (ML) estimators for the realization of stochastic processes have been investigated in a number of papers. Du, Zhang, Mandrekar (2009) and Kaufman, Schervish, Nychka (2008) studied tapered ML estimators for the Matern correlation function. Ying (1993) and Chen, Simpson, Ying (2000) established asymptotic distributions for the exponential correlation function. In particular, Ying proved that the estimators are strongly consistent and asymptotically normal under mild conditions. Loh (2005) proved a weak consistency of estimators for the Matern correlation function with $\nu = 1.5$ under mild conditions. Note that the result of Loh is based on the explicit inversion of the correlation matrix and other calculations using mathematical software of symbolic computations. Zhang (2004) shown that not all parameters in the covariance are consistently estimable and only a function of parameters allows consistent estimation.

In the above references, the authors considered the data drawn from the Gaus-

sian process model. However, in some practical situations, the data is collecting as precise observations of a deterministic model. This happens in computer experiments (Kennedy, O'Hagan, 2001, Santner, Williams, Notz, 2003, Paulo, 2005) and machine learning (Rasmussen, Williams, 2006). For example, deterministic mathematical models are used for developing engineering structures, where building physical prototypes is too costly, and to determine the performance of integrated circuits or total joint replacements in prosthesis devices.

In this work the asymptotic properties of ML estimators is studied in the case of analysis of deterministic models. Such consideration does not correspond to the statistical theory but gives an insight on related questions. It is shown on a number of examples that, for observations of a deterministic model, the maximum likelihood estimator of the correlation parameter tends to the infinity as the number of points is increasing. It means that a deterministic model is approximated by a Gaussian process with large correlations between any points.

2.2 The likelihood for Gaussian process with exponential correlation function

Let $y_i = \eta(x_i)$ be the output of the deterministic model $\eta(x)$ at the point $x_i \in [0,1]$, $i = 1, \ldots, n$. Note that for a deterministic model the replication of observation at any point gives the same output. Without loss of generality, let $x_1 < \ldots < x_n$.

Lemma. *For the model $\eta(x) = x - 1/2$, $x_i = (i-1)/(n-1)$, and the Gaussian process with the constant mean β, the variance σ^2 and the correlation function $r(x, \tilde{x}) = e^{-|x-\tilde{x}|/\psi}$, the maximum likelihood estimator of correlation parameter ψ tends to infinity as the data size n increases.*

It is well-known (Rasmussen, Williams, 2006) that the likelihood for the Gaussian process have a form

$$p(y|\beta, \sigma, \psi) = \frac{|R|^{-1/2}}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}(y-H\beta)^T R^{-1}(y-H\beta)}$$

where $y = (y_1, \ldots, y_n)^T$ is a vector of output values, $R = (r(x_i, x_j|\psi))_{i,j=1}^n$ is the correlation matrix, $H = (h(x_1), \ldots, h(x_n))$, and $h(x) \equiv 1$.

The maximum likelihood estimators of β and σ have the following explicit forms

$$\hat{\beta} = (H^T R^{-1} H)^{-1} H R^{-1} y$$

and

$$\hat{\sigma}^2 = \frac{1}{n}(y - H\hat{\beta})^T R^{-1}(y - H\hat{\beta}).$$

The ML estimator of ψ can be found only numerically in the following way

$$\hat{\psi} = \arg \max_{\psi \in (0,\infty)} p(y|\hat{\beta}, \hat{\sigma}, \psi).$$

After substituting and simplifying, we obtain that the estimator $\hat{\psi}$ maximizes

$$L(\psi) = \ln \left[|R|^{-1/2} \right] - \frac{n}{2} \ln \left[(y - H\hat{\beta})^T R^{-1} (y - H\hat{\beta}) \right].$$

For the exponential correlation function, the inverse of matrix R admits the explicit representation

$$R^{-1} = \frac{1}{1-\lambda^2} \begin{bmatrix} 1 & -\lambda & 0 & \cdots & 0 & 0 \\ -\lambda & 1+\lambda^2 & -\lambda & \ddots & 0 & 0 \\ 0 & -\lambda & 1+\lambda^2 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1+\lambda^2 & -\lambda \\ 0 & 0 & 0 & \cdots & -\lambda & 1 \end{bmatrix}$$

where $\lambda = e^{-\frac{1}{(n-1)\psi}}$. The straightforward calculation yields that

$$yR^{-1}y = \frac{y_1^2 + y_n^2}{1-\lambda^2} + \sum_{i=2}^{n-1} y_i^2 \frac{1+\lambda^2}{1-\lambda^2} - 2 \sum_{i=1}^{n-1} y_i y_{i+1} \frac{\lambda}{1-\lambda^2}$$

and

$$|R|^{-1/2} = \frac{1}{(1-\lambda^2)^{(n-1)/2}}.$$

For the model $\eta(x) = x - 1/2$, we obtain that $\hat{\beta} = 0$ and

$$yR^{-1}y = \frac{1}{2} \frac{1}{1-\lambda^2} + \frac{n^2-5n+6}{12(n-1)} \cdot \frac{1+\lambda^2}{1-\lambda^2} - \frac{n^2-2n-3}{6(n-1)} \cdot \frac{\lambda}{1-\lambda^2}.$$

The estimator $\hat{\psi}$ can be found explicitly in Maple and is not presented since it is a very large expression. Asymptotically we have

$$\hat{\psi} = \hat{\psi}(n) = \frac{n}{2} - \frac{7}{6} - \frac{7}{18n} - \frac{17}{54n^2} + O\left(\frac{1}{n^3}\right).$$

This means that the estimator $\hat{\psi}$ increases almost linearly as the size n of data increases.

2.3 General case

Numerical calculations show non-intuitive asymptotics for arbitrary deterministic models and correlation functions. The maximum likelihood estimators of parameters of

Gaussian process for different correlation functions are presented in Table 1 for the model $\eta(x) = \sqrt{1+x}$, in Table 2 for the model $\eta(x) = e^{-3x}$ and in Table 3 for the model $\eta(x) = \sin(2x)$. These models have different curvatures.

We see that for three models the ML estimator of correlation parameter using the Matern correlation function increases as the data size increases. Thus, it may be conjectured that, for the Gaussian process with the Matern correlation function, this asymptotic holds for any deterministic model given by an analytic function. This asymptotic can be explained by micro oscillations of trajectories of stochastic process or numerous alternations of sign of a high-order derivative. Note that micro oscillations are decreasing if the correlation between points is increasing.

The case of Gaussian and rational correlation functions is significantly more complicated. The estimator of correlation parameter is decreasing for the model $\eta(x) = \sqrt{1+x}$ and is increasing for models $\eta(x) = e^{-3x}$ and $\eta(x) = \sin(2x)$. Moreover, the estimator of mean and variance is significantly increasing for models $\eta(x) = e^{-3x}$ and $\eta(x) = \sin(2x)$. Also, for Gaussian and rational correlation functions the condition number of the correlation matrix $R = R(\hat{\psi})$ is huge for all three models, see also Ababou, Bagtzoglou, Wood (1994). This phenomena is undesirable in numerical computations.

For the correlation functions $\rho(t) = e^{-|t|}$ and $\rho(t) = \exp(-|t|^{1.9})$ the condition number of the correlation matrix is rather small and the estimators slightly depend on the data size n. The estimator of mean β belongs to the interval of model response for any n that does not hold true for other correlation functions.

Asymptotics for the estimators using the Matern correlation function with $\nu = 1.5$ and $\nu = 2.5$ is rather surprising. In particular, the estimator of variance using the Matern correlation function with $\nu \geq 1.5$ is quite different from one using Gaussian and rational correlation functions.

It is interesting to note different asymptotics of estimators for two fairly close correlation functions $\rho(t) = \exp(-t^2)$ and $\rho(t) = \exp(-|t|^{1.9})$. This may be explained that that the correlation function $\rho(t) = e^{-|t|^\gamma}$ is defined for $\gamma \in (0, 2]$, and a trajectory of Gaussian process is infinitely mean square differentiable with $\gamma = 2$ and is not mean square differentiable with $\gamma < 2$. The function $\rho(t) = e^{-|t|^\gamma}$ is called a stable correlation function that have been motivated by Davis (1973) in applications to geology. Results given in Tables 1-3 show that the estimators of parameters using the correlation function $\rho(t) = e^{-|t|^{1.9}}$ are rather "stable" in the sense that the estimators depend slightly on n. Thus, the stable correlation function with $\gamma \approx 1.9$ has very attractive features and is appealing for the application to the analysis of deterministic models. The studying of related issues is deferred to future research.

Table 2.1: *Maximum likelihood estimators for Gaussian processes with mean β, variance σ^2 and different correlation functions $r(t) = \rho(t/\psi)$ for the data from the model $\eta(x) = \sqrt{1+x}$ observed at n points $x_i = (i-1)/(n-1)$, and the order of the condition number of correlation matrix, $s = \log_{10} \mathrm{cond}(R)$.*

n	6	8	10	14	18
exponential, $\rho(t) = e^{-\lvert t\rvert}$					
$\hat{\psi}$	1.80	2.78	3.78	5.76	7.74
$\hat{\sigma}^2$	0.034	0.036	0.037	0.039	0.040
$\hat{\beta}$	1.210	1.209	1.208	1.208	1.208
s	1.96	2.45	2.8	3.3	3.66
Matern $\nu = 1.5$, $\rho(t) = e^{-\lvert t\rvert}(1 + \lvert t\rvert)$					
$\hat{\psi}$	24.6	39.4	54.8	86.9	118.5
$\hat{\sigma}^2$	30.8	75.4	143.4	356.5	659.3
$\hat{\beta}$	0.44	-0.09	-0.64	-1.80	-2.95
s	8.1	9.3	10.2	11.4	12.3
Matern $\nu = 2.5$, $\rho(t) = e^{-\lvert t\rvert}(1 + \lvert t\rvert + t^2/3)$					
$\hat{\psi}$	5.10	6.66	8.09	10.75	13.31
$\hat{\sigma}^2$	9.07	19.99	36.86	93.92	196.4
$\hat{\beta}$	-0.33	-1.43	-2.69	-5.63	-9.22
s	9.5	11.1	12.2	13.8	15.0
stable, $\rho(t) = \exp(-\lvert t\rvert^{1.9})$					
$\hat{\psi}$	2.56	2.88	3.11	3.47	3.74
$\hat{\sigma}^2$	0.11	0.10	0.09	0.08	0.07
$\hat{\beta}$	1.14	1.13	1.13	1.12	1.11
s	4.2	4.7	5.1	5.6	6.0
Gaussian, $\rho(t) = \exp(-t^2)$					
$\hat{\psi}$	2.39	1.89	1.57	1.19	0.99
$\hat{\sigma}^2$	1.31	1.49	1.63	1.97	3.50
$\hat{\beta}$	0.82	0.71	0.64	0.53	0.30
s	10.5	14.5	18.2	25.5	32.8
rational, $\rho(t) = 1/(1 + t^2)$					
$\hat{\psi}$	4.34	3.84	3.47	3.02	2.76
$\hat{\sigma}^2$	2.23	2.76	3.34	5.30	9.76
$\hat{\beta}$	0.66	0.45	0.27	-0.14	-0.70
s	10.7	14.6	18.2	25.2	32.1

Table 2.2: *Maximum likelihood estimators for Gaussian processes with mean β, variance σ^2 and different correlation functions $r(t) = \rho(t/\psi)$ for the data from the model $\eta(x) = \exp(-3x)$ observed at n points $x_i = (i-1)/(n-1)$, and the order of the condition number of correlation matrix, $s = \log_{10} \mathrm{cond}(R)$.*

n	6	8	10	14	18
exponential, $\rho(t) = e^{-\lvert t\rvert}$					
$\hat{\psi}$	0.996	1.599	2.195	3.404	4.612
$\hat{\sigma}^2$	0.166	0.179	0.188	0.198	0.204
$\hat{\beta}$	0.459	0.476	0.487	0.498	0.505
s	1.65	2.18	2.54	3.06	3.43
Matern $\nu = 1.5$, $\rho(t) = e^{-\lvert t\rvert}(1 + \lvert t\rvert)$					
$\hat{\psi}$	0.93	1.51	2.11	3.30	4.49
$\hat{\sigma}^2$	0.92	2.26	4.32	10.54	19.57
$\hat{\beta}$	0.86	1.26	1.68	2.51	3.36
s	3.8	5.0	5.9	7.1	8.0
Matern $\nu = 2.5$, $\rho(t) = e^{-\lvert t\rvert}(1 + \lvert t\rvert + t^2/3)$					
$\hat{\psi}$	0.65	1.05	1.46	2.29	3.12
$\hat{\sigma}^2$	2.13	9.81	30.98	161.5	524.4
$\hat{\beta}$	1.28	2.58	4.43	9.79	17.36
s	5.0	7.1	8.5	10.5	11.8
stable, $\rho(t) = \exp(-\lvert t\rvert^{1.9})$					
$\hat{\psi}$	1.04	1.13	1.17	1.20	1.22
$\hat{\sigma}^2$	0.49	0.462	0.41	0.322	0.261
$\hat{\beta}$	0.77	0.838	0.871	0.903	0.917
s	3.4	3.9	4.3	4.7	5.1
Gaussian, $\rho(t) = \exp(-t^2)$					
$\hat{\psi}$	0.72	0.90	1.03	1.24	1.41
$\hat{\sigma}^2$	0.935	7.348	60.8	4646	390726
$\hat{\beta}$	0.84	2.04	5.37	41.78	354.29
s	5.3	9.9	14.9	26.0	38.0
rational, $\rho(t) = 1/(1 + t^2)$					
$\hat{\psi}$	1.25	1.80	2.32	3.32	4.30
$\hat{\sigma}^2$	1.84	19.23	230.1	40347	8313987
$\hat{\beta}$	1.09	3.137	10.12	123.1	1676.2
s	5.5	10.1	15.2	26.3	38.5

Table 2.3: *Maximum likelihood estimators for Gaussian processes with mean β, variance σ^2 and different correlation functions $r(t) = \rho(t/\psi)$ for the data from the model $\eta(x) = \sin(2x)$ observed at n points $x_i = (i-1)/(n-1)$, and the order of the condition number of correlation matrix, $s = \log_{10} \mathrm{cond}(R)$.*

n	6	8	10	14	18
exponential, $\rho(t) = e^{-\lvert t\rvert}$					
$\hat{\psi}$	0.99	1.51	2.02	3.05	4.07
$\hat{\sigma}^2$	0.176	0.183	0.187	0.193	0.196
$\hat{\beta}$	0.536	0.517	0.504	0.490	0.482
s	1.7	2.2	2.5	3.0	3.4
Matern $\nu = 1.5$, $\rho(t) = e^{-\lvert t\rvert}(1 + \lvert t\rvert)$					
$\hat{\psi}$	0.92	1.38	1.81	2.63	3.43
$\hat{\sigma}^2$	0.74	1.39	2.17	4.13	6.65
$\hat{\beta}$	0.08	-0.26	-0.57	-1.16	-1.73
s	3.8	4.9	5.7	6.8	7.6
Matern $\nu = 2.5$, $\rho(t) = e^{-\lvert t\rvert}(1 + \lvert t\rvert + t^2/3)$					
$\hat{\psi}$	0.90	1.36	1.77	2.52	3.24
$\hat{\sigma}^2$	2.77	8.76	19.67	62.26	146.90
$\hat{\beta}$	-0.86	-2.21	-3.65	-6.79	-10.41
s	5.7	7.6	8.9	10.7	11.9
stable, $\rho(t) = \exp(-\lvert t\rvert^{1.9})$					
$\hat{\psi}$	0.950	0.981	0.993	1.003	1.006
$\hat{\sigma}^2$	0.266	0.215	0.177	0.130	0.102
$\hat{\beta}$	0.244	0.214	0.201	0.190	0.185
s	3.3	3.8	4.1	4.6	4.9
Gaussian, $\rho(t) = \exp(-t^2)$					
$\hat{\psi}$	1.63	1.90	2.14	2.55	2.95
$\hat{\sigma}^2$	1.85	4.62	9.94	59.919	500.6
$\hat{\beta}$	-0.52	0.53	0.42	0.24	-3.41
s	8.8	14.5	20.6	34.1	48.9
rational, $\rho(t) = 1/(1 + t^2)$					
$\hat{\psi}$	2.81	3.52	4.77	6.66	8.57
$\hat{\sigma}^2$	8.20	28.87	247.3	7907	283833
$\hat{\beta}$	-1.21	0.62	3.57	-10.21	14.23
s	8.8	14.0	20.7	34.1	48.7

Bibliography

[1] Ababou R., Bagtzoglou A.C., Wood E.F. (1994) On the condition number of covariance matrices in kriging, estimation, and simulation of random fields. Mathematical Geology 26, 99–133.

[2] Chen H.-S., Simpson D.G., Ying Z. (2000) Fixed-domain asymptotics for a stochastic process model with measurement error. Statist. Sinica 10 141-156.

[3] Du J., Zhang H., Mandrekar V.S. (2009) Fixed-domain asymptotic properties of tapered maximum likelihood estimators. Annals of Stat. 37, 3330-3361.

[4] Davis J. (1973) Statistics and Data Analysis in Geology, Wiley, New York.

[5] Kaufman C., Schervish M., Nychka D. (2008). Covariance tapering for likelihood-based estimation in large spatial datasets. J. Amer. Statist. Assoc. 103, 1545-1555.

[6] Loh W.-L. (2005). Fixed-domain asymptotics for a subclass of Matern-type Gaussian random fields. Ann. Statist. 33, 2344-2394.

[4] Kennedy M.C., O'Hagan A. Bayesian calibration of computer models. J. R. Stat. Soc. Ser. B 63 (2001), no. 3, 425–464.

[8] Paulo R. (2005) Default priors for Gaussian processes. Ann. Statist. 33, no. 2, 556–582.

[13] Rasmussen C.E., Williams C.K.I. (2006) Gaussian processes for machine learning. MIT Press, Cambridge.

[16] Santner T.J., Williams B.J., Notz W. (2003) The Design and Analysis of Computer Experiments. Springer-Verlag, New York.

[11] Ying Z. (1993) Maximum likelihood estimation of parameters under a spatial sampling scheme. Ann. Statist. 21, 1567-1590.

[12] Zhang H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. J. Amer. Statist. Assoc. 99 250-261.

Chapter 3

The role of the nugget term in the Gaussian process method

The maximum likelihood estimate of the correlation parameter of a Gaussian process with and without of a nugget term is studied in the case of the analysis of deterministic models.

3.1 Introduction

The Gaussian process method is an elegant way to analyze the results of experiments in many areas of science including machine learning (Rasmussen, Williams 2006), spatial statistics (Matheron 1973, Ripley 1981, Cressie 1993, Muller 2007), and the Bayesian analysis of computer experiments (Sacks et al. 1989, Kennedy, O'Hagan 2001, Santner et al. 2003). Each area has its own specific ways of employing and interpreting the Gaussian processes. The purpose of this chapter is not to give a full overview, that can be found in the above references, but to discuss some issues on the nugget term for the analysis of computer experiments.

The conception of the nugget term was first introduced in geostatistics by Matheron (1962). Roughly speaking, the variogram and covariance often show a discontinuity at the origin, termed the nugget effect. The nugget effect is considered as a random noise and may represent a measurement error or short scale variabilities. The nugget term is a well explored object in spatial statistics (Pitard 1993).

Another area of the application of Gaussian processes is the Bayesian approach developed for the analysis of computer experiments. In this approach, a so-called emulator is introduced for making probabilistic judgments on the true output of the given computer model, which is called a simulator. A Gaussian process is used for a full probabilistic specification of the emulator. Thus, the emulator is utilized to measure uncertainty of different kinds, see (Kennedy, O'Hagan 2001).

Formally, there is no nugget term in the Gaussian process method for the analysis of

deterministic models, but the nugget term can be introduced artificially, for example, for the regularization of the inversion of a covariance matrix, see (Neal 1997) for details. Gramacy and Lee (2009) reported on the usefulness of the nugget term in their research of supercomputer experiments.

The presence of the nugget term in the Gaussian process method is natural for the analysis of stochastic and simulation models. The nugget effect may represent a measurement error or an effect of random values used inside computer models (Kleijnen 2008, Kleijnen, Beers 2005).

The influence of the nugget term for optimal designs of experiments for a number of cases have been studied in (Zhu, Stein 2005, Stehlik et al. 2008).

The present chapter focuses on the Gaussian process method applied for the analysis of deterministic models. It is shown that the nugget term has a great impact on the likelihood and the estimate of correlation parameter.

3.2 The likelihood for a Gaussian process without the nugget term

In this section, it is shown that the likelihood of a Gaussian process has an unexpected behaviour in the analysis of non-stochastic models. More precisely, for a deterministic model of observations, the maximum likelihood estimate of the correlation parameter may tend to the infinity as the number of points increases. It means that a deterministic model is approximated by a Gaussian process with the correlation function $r(x) \approx 1$ for any x.

Indeed, let $y_i = \eta(x_i)$ be the output of the model $\eta(x)$ at the point $x_i \in [0,1]$, $i = 1, \ldots, n$. Note that for a deterministic model, the replication of an observation at some point gives the same output. Without loss of generality, let $x_1 < \ldots < x_n$. The likelihood for a Gaussian process with constant mean β, variance σ^2 and correlation function $r(x, \tilde{x}) = e^{-|x-\tilde{x}|/\psi}$ have the form

$$p(y|\beta, \sigma, \psi) = \frac{|R|^{-1/2}}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}(y-H\beta)^T R^{-1}(y-H\beta)}$$

where $y = (y_1, \ldots, y_n)^T$ is the vector of output values, $R = (r(x_i, x_j|\psi))_{i,j=1}^n$ is the correlation matrix, $H = (h(x_1), \ldots, h(x_n))$, and $h(x) \equiv 1$.

The maximum likelihood (ML) estimates of β and σ have the following explicit forms

$$\hat{\beta}_{ML} = (H^T R^{-1} H)^{-1} H R^{-1} y$$

and

$$\hat{\sigma}^2_{ML} = \frac{1}{n}(y - H\hat{\beta}_{ML})^T R^{-1}(y - H\hat{\beta}_{ML}).$$

The ML estimate of ψ can be found only numerically in the following way

$$\hat{\psi}_{ML} = \arg \max_{\psi \in (0,\infty)} p(y|\hat{\beta}_{ML}, \hat{\sigma}_{ML}, \psi).$$

After substituting and simplifying, we obtain that the estimate $\hat{\psi}_{ML}$ maximizes

$$L(\psi) = \ln \left[|R|^{-1/2} \right] - \frac{n}{2} \ln \left[(y - H\hat{\beta}_{ML})^T R^{-1} (y - H\hat{\beta}_{ML}) \right].$$

For the exponential correlation function, the inverse of matrix R admits the explicit representation $R^{-1} = V^T V$ where the matrix V is defined by

$$V = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -\frac{\mu_2}{\sqrt{1-\mu_2^2}} & \frac{1}{\sqrt{1-\mu_2^2}} & 0 & \cdots & 0 & 0 \\ 0 & -\frac{\mu_3}{\sqrt{1-\mu_3^2}} & \frac{1}{\sqrt{1-\mu_3^2}} & \cdots & 0 & 0 \\ & & & & & \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -\frac{\mu_n}{\sqrt{1-\mu_n^2}} & \frac{1}{\sqrt{1-\mu_n^2}} \end{bmatrix},$$

$\mu_i = e^{-(x_i - x_{i-1})/\psi}$. For n equidistant points $x_i = (i-1)/(n-1)$, $i = 1, \ldots, n$, straightforward calculation shows that

$$yR^{-1}y = \frac{y_1^2 + y_n^2}{1-\lambda^2} + \sum_{i=2}^{n-1} y_i^2 \frac{1+\lambda^2}{1-\lambda^2} - 2\sum_{i=1}^{n-1} y_i y_{i+1} \frac{\lambda}{1-\lambda^2}$$

where $\lambda = e^{-\frac{1}{(n-1)\psi}}$, and

$$|R|^{-1/2} = \frac{1}{(1-\lambda^2)^{(n-1)/2}}.$$

For the model $\eta(x) = x - 1/2$, we obtain that $\hat{\beta}_{ML} = 0$ and

$$yR^{-1}y = \frac{1}{2}\frac{1}{1-\lambda^2} + \frac{n^2 - 5n + 6}{12(n-1)} \cdot \frac{1+\lambda^2}{1-\lambda^2} - \frac{n^2 - 2n - 3}{6(n-1)} \cdot \frac{\lambda}{1-\lambda^2}.$$

The estimate $\hat{\psi}_{ML}$ can be found explicitly in Maple and is not presented since it is a very large expression. Applying the power series expansion, we have

$$e^{-\frac{1}{(n-1)\hat{\psi}_{ML}}} = 1 - \frac{2}{n^2} - \frac{20}{3n^2} + O\left(\frac{1}{n^3}\right) \text{ and } \hat{\psi}_{ML} = \frac{n}{2} - \frac{7}{6} - \frac{7}{18n} - \frac{17}{54n^2} + O\left(\frac{1}{n^3}\right).$$

The dependence of $\hat{\psi}_{ML}$ on n is given in Figure 3.1 for the model $\eta(x) = x - 1/2$ at the left part and for the model $\eta(x) = \sin(2\pi x)$ at the right part. We observe that the estimate $\hat{\psi}_{ML}$ increases almost linearly as n increases for both models.

The maximum likelihood estimate of ψ for the Gaussian correlation function $r(x, \tilde{x}) = e^{-(x-\tilde{x})^2/\psi}$ is given in Figure 3.2. For the model $\eta(x) = x - 1/2$ we have that $\hat{\psi}_{ML} = \infty$ for any n. Note that for the model $\eta(x) = \sin(2\pi x)$, the condition number of the correlation matrix $R(\hat{\psi}_{ML})$ is of order 10^7, 10^{14}, 10^{22}, 10^{30}, and 10^{38} for $n = 8$, 11, 14, 17,

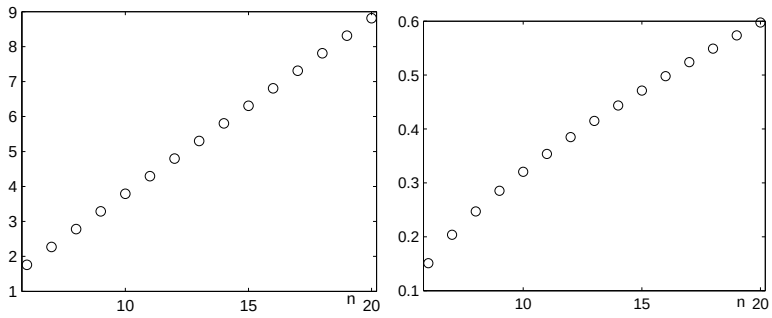

Figure 3.1: The maximum likelihood estimate of ψ for the Gaussian process with the exponential correlation function and n equidistant points on the interval $[0, 1]$ for the model $\eta(x) = x - 1/2$ (left part) and for the model $\eta(x) = \sin(2\pi x)$ (right part) for $n = 6, \ldots, 20$.

and 20, respectively. These calculations were done in Maple with 45 digits precision. However, the computer representation of floating numbers typically has only 17 digits. Thus, it is impossible to find the maximum likelihood estimate for large n using the ordinary floating representation in a computer. In particular, Ababou et al. (1994) have shown that the condition number grows linearly for the exponential correlation function and grows exponentially for the Gaussian correlation function.

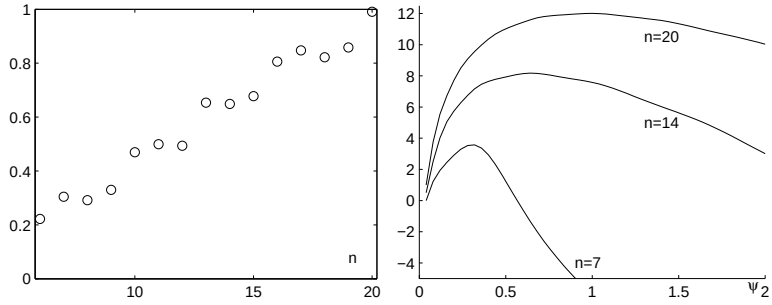

Figure 3.2: At left: The maximum likelihood estimate of ψ for the Gaussian process with the Gaussian correlation function and n equidistant points on the interval $[0, 1]$ for the model $\eta(x) = \sin(2\pi x)$ for $n = 6, \ldots, 20$. At right: The likelihood function of ψ for $n = 7, 14, 20$.

In more general situations for other correlation functions and other models, the dependence of the maximum likelihood estimate and the restricted maximum likelihood estimate of ψ on n remains typically the same and can be verified numerically (see Chapter 2).

Thus, roughly speaking, the estimate of parameters of a Gaussian process is associated with the given data set and is not associated with the deterministic model. This estimation is not simple and is not well-defined. It is easy to observe that if one divides an input space into several regions, one may get quite different estimates of parameters

for different regions. However, if one is looking for one Gaussian process over the full space, one has difficulty in finding the single estimate.

3.3 The likelihood for a Gaussian process with the nugget term

In this section, the likelihood with the presence of the nugget term is investigated. For this case, the correlation matrix R in formulae from Section 2 should be replaced to the correlation matrix

$$R_\nu = \big((1-\nu)r(x_i - x_j) + \nu\delta_{i,j}\big)_{i,j}$$

where ν is the nugget term.

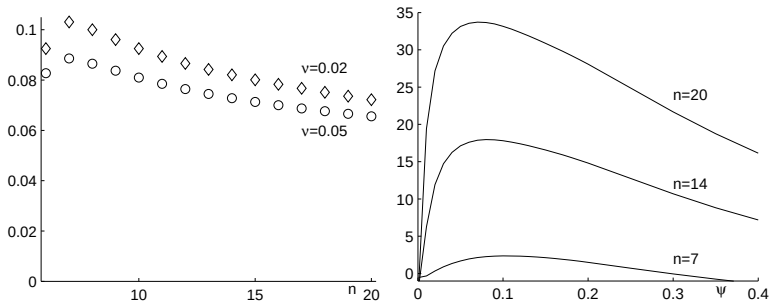

Figure 3.3: At left: The maximum likelihood estimate of ψ for the Gaussian process with the Gaussian correlation function and the nugget term $\nu = 0.02, 0.05$ for measurements of the model $\eta(x) = \sin(2\pi x)$ at n equidistant points on the interval $[0, 1]$, $n = 6, \ldots, 20$. At right: The likelihood function of ψ for the nugget term $\nu = 0.02$ and for $n = 7, 14, 20$.

The likelihood function and the maximum likelihood estimate for fixed values of the nugget term are presented in Figure 3.3. One can observe that the nugget term essentially changes the maximum likelihood estimate of ψ (and also σ). The estimate $\hat{\psi}_{ML}$ does not increase to infinity as n increases, since the Gaussian process is fitted to a band around the deterministic function. It should also be noted that the condition number of the correlation matrix R_α is of order 10^2 and is increasing very slowly as n is increasing. Moreover, the estimate $\hat{\psi}_{ML}$ is smaller with the presence of the nugget term that also reduces the condition number of the correlation matrix. Ababou et al. (1994) have shown that the condition number of the correlation matrix for the Gaussian process models increases to a finite limit with the presence of the nugget term.

Note one undesired effect of the nugget term. The likelihood may have the second mode for large values of the correlation parameter, see Figure 3.4. The second mode strongly depends on a value of the nugget term and can be considered as a false mode.

For some data, the likelihood function at the second mode may have a larger value than at the first mode.

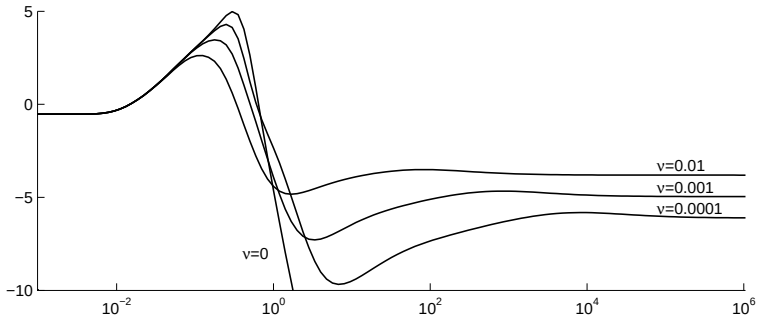

Figure 3.4: The likelihood function of ψ for the Gaussian process with the Gaussian correlation function and the nugget term $\nu = 0, 0.01, 0.001, 0.0001$ for measurements of the model $\eta(x) = \sin(2\pi x)$ at 7 equidistant points on the interval $[0, 1]$.

Note that in the presence of the nugget term, the meta-model

$$m_\nu(x) = H\beta + t^T(x)R_\nu^{-1}(y - H\beta)$$

where $t(x) = (r(x, x_1), \ldots, r(x, x_n))^T$ does not possess the interpolation property. Nevertheless, the deviations $\varepsilon_i = y_i - m_\nu(x_i)$ are very small. One may construct a meta-model, that interpolates the dataset $\{(x_i, \varepsilon_i)\}_{i=1}^n$, by a method given in Cressie (1993, Sect. 5.9). It is not necessary for the deviations ε_i to use the Kriging approach without the nugget term. One may use the inverse distance weighted interpolation (Cressie 1993, p. 371; Lu, Wong 2008) and define the meta-model in the following form

$$m(x) = m_\nu(x) + \frac{\sum\limits_{i=1}^{n} \varepsilon_i ||x - x_i||_2^{-2}}{\sum\limits_{i=1}^{n} ||x - x_i||_2^{-2}}\,.$$

3.3.1 MLE for stationary processes

Let us perform a small simulation study. Assume that the results of experiments satisfy

$$y(x_i) = \beta + \sigma^2\varepsilon^{(1)}(x_i) + \tau^2\varepsilon^{(2)}(x_i)$$

where $x_1, \ldots, x_n$ are points of measurements, $\varepsilon^{(1)}(x)$ denotes a stationary Gaussian process with correlation function $r(x) = e^{-x^2/\psi}$ and $\varepsilon^{(2)}(x)$ is white noise. Let $\mathbf{E}\varepsilon^{(j)}(x) = 0$, $\mathbf{D}\varepsilon^{(j)}(x) = 1$, processes $\varepsilon^{(1)}(x)$ and $\varepsilon^{(2)}(x)$ be independent. The values $\beta + \sigma^2\varepsilon^{(1)}(x_i)$ may be conceived as true values of a physical process. The values $\tau^2\varepsilon^{(2)}(x_i)$ may be interpreted as a measurement error or a rough rounding of measured values. Let us compute the maximum likelihood estimators for 1000 realizations obtained for $n = 8$,

$x_i = (i-1)/7$, $i = 1, \ldots, 8$, $\beta = 2$, $\psi = 1.5$, $\sigma = 1$, $\tau = 0$ or $\tau = 0.01$. Results of the maximum likelihood estimation for different values of the nugget term are presented in Table 3.1.

Table 3.1: The mean of maximum likelihood estimators of parameters using different values of the nugget term. Standard deviations are given in brackets.

	$\tau = 0$			$\tau = 0.01$		
ν	0	0.01	0.02	0	0.01	0.02
$\hat{\beta}_{ML}$	2.03(0.68)	2.01(0.85)	2.02(0.86)	2.02(0.92)	2.04(0.85)	2.04(0.86)
$\hat{\sigma}_{ML}$	0.83(0.40)	0.29(0.17)	0.27(0.16)	0.33(0.23)	0.30(0.17)	0.28(0.16)
$\hat{\psi}_{ML}$	1.44(0.37)	0.54(0.25)	0.47(0.20)	0.14(0.06)	0.58(0.29)	0.49(0.23)

One can observe that the maximum likelihood estimators with a nonzero nugget term does not depend on small perturbations $\{\tau^2\varepsilon^{(2)}(x_i)\}_i$ of the data $\{\beta + \sigma^2\varepsilon^{(1)}(x_i)\}_i$. In contrast, for $\nu = 0$, the maximum likelihood estimators of σ and ψ are significantly changed due to adding small perturbations. In all cases, the accuracy of $\hat{\beta}_{ML}$ is approximately the same. Thus, as can be seen, the nugget term yields a regularization effect on the maximum likelihood estimators.

3.4 Conclusions

In the analysis of deterministic models the presence of a nugget term has a significant impact on the likelihood of a Gaussian process. The maximum likelihood estimate of the correlation parameter with a nonzero nugget term is more reliable and the condition number of the correlation matrix is moderate. Even if a deterministic model does not have any internal computational errors or other perturbations, the artificial introduction of the nugget term can be recommended.

Bibliography

[1] Ababou, R., Bagtzoglou, A.C., Wood, E.F.: On the condition number of covariance matrices in kriging, estimation, and simulation of random fields. Mathematical Geology 26, 99–133 (1994)

[2] Cressie, N.A.C.: Statistics for Spatial Data. John Wiley & Sons, New York (1993)

[3] Gramacy, R.B., Lee, H.K.H.: Adaptive Design and Analysis of Supercomputer Experiments. Technometrics **51**, 130–145 (2009)

[4] Kennedy, M.C., O'Hagan A.: Bayesian calibration of computer models. J. R. Stat. Soc. Ser. B **63**, 425–464 (2001)

[5] Kleijnen, J.P.C.: Design and Analysis of Simulation Experiments. Springer, New York (2008)

[6] Kleijnen, J.P.C., van Beers W.C.M.: Robustness of Kriging when interpolating in random simulation with heterogeneous variances – some experiments. European J. Oper. Res. **165**, 826–834 (2005)

[7] Lu, G.Y., Wong, D.W.: An adaptive inverse-distance weighting spatial interpolation technique. Computers & Geosciences 34, 1044–1055 (2008)

[8] Matheron, G.: Traite de geostatistique appliquee: Memoires du Bureau de Recherche Geologiques et Minieres 24, Paris (1962)

[9] Matheron, G.: The intrinsic random functions and their applications. Advances in Appl. Probability **5**, 439–468 (1973)

[10] Muller, W.G.: Collecting Spatial Data (3rd edn). Physica-Verlag, Heidelberg (2007)

[11] Neal, R.: Monte Carlo implementation of Gaussian process models for Bayesian regression and classification. Tech. Rep. CRGTR972, Dept. of Computer Science, University of Toronto (1997)

[12] Pitard, F.F.: Exploration of the Nugget Effect. Geostatistics for the Next Century. Kluwer Academic Pub. Doordrecht, The Netherlands (1993)

[13] Rasmussen, C.E., Williams, C.K.I.: Gaussian processes for machine learning. MIT Press, Cambridge (2006)

[14] Ripley, B.D.: Spatial statistics. John Wiley & Sons, Inc., New York (1981)

[15] Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P.: Design and analysis of computer experiments. Statist. Sci. **4**, 409–435 (1989)

[16] Santner, T.J., Williams, B.J., Notz, W.: The Design and Analysis of Computer Experiments. Springer-Verlag, New York (2003)

[17] Stehlik, M., Rodriguez-Diaz, J.M., Muller, W.G., Lopez-Fidalgo, J.: Optimal allocation of bioassays in the case of parameterized covariance functions: an application in lung's retention of radioactive particles. Test **17**, 56-68. (2008)

[18] Zhu, Z., Stein, M.L.: Spatial sampling design for parameter estimation of the covariance function. J. Stat. Plan. Inf. **134**, 583–603 (2005)

Chapter 4

Generalized Latin hypercube design for computer experiments

Space filling designs, which satisfy a uniformity property, are widely used in computer experiments. In the present chapter, the performance of non-uniform experimental designs, which locate more points in a neighborhood of the boundary of the design space, is investigated. These designs are obtained by a quantile transformation of the one-dimensional projections of commonly used space filling designs. This transformation is motivated by logarithmic potential theory, which yields the arc-sine measure as an equilibrium distribution. The methodology is illustrated for maximin Latin hypercube designs by several examples. In particular, it is demonstrated that the new designs yield a smaller integrated mean square error for prediction.

4.1 Introduction

In modern scientific studies, mathematical computer models are widely used to describe complex processes. These models are a replacement for real (physical, chemical, biological, etc.) experiments which are too time consuming or too costly. Typical examples of computer models include the flow of air over an airplane wing, behavior of a metal structure under stress, combustion of gasses in a flame, VLSI-circuit and engine design, plan ecology, thermal-energy storage, health medical models and many other processes. Moreover, mathematical models may describe phenomena which can not be reproduced, for example, weather conditions or climate change. Computer experiments usually involve a large number of input variables and are typically extremely time consuming. One run of a computer model may require several hours or longer. Therefore, the choice of an appropriate set of input conditions is of particular importance for making inferences on the basis of a computer model. In the literature, this rapidly growing field is called design for computer experiments [see Sacks, Welch, Mitchell and Wynn (1989), Santner, Williams and Notz (2003), Fang, Li, Sudjianto (2006) or

Kleijnen (2008) among many others].

For computer experiments, space-filling designs are very popular. The class of space-filling designs includes a wide variety of designs such as Latin hypercube designs (LHD), designs based on random or pseudo-random sequences, sphere packing designs and others. Most of these designs are uniformly space-filling in the sense that the design points are uniformly scattered on the domain.

The purpose of the present note is to give some motivation for practitioners to choose non-uniform space-filling designs in computer experiments. In the new design, it is optimized an average measure of inter-point distances of each marginal projection of the design. In particular, we consider Latin hypercube designs with marginals minimizing a logarithmic potential. This optimization problem yields to the arc-sine distribution and we propose to transform the marginals of maximin Latin hypercube designs by the corresponding quantile function. The transformed designs are called generalized maximin Latin hypercube designs. These designs locate more points in a neighborhood of the boundary of the design space. We show their efficiency for prediction with respect to the mean squared criterion by several numerical examples.

The remaining part of the chapter is organized as follows. In Section 2 we review some notations from the design of computer experiments and give heuristic arguments for introducing generalized maximin Latin hypercube designs, which are obtained by transforming the one-dimensional projections of a maximin Latin hypercube design by the quantile function of a Beta-distribution. In particular, we consider the arc-sine distribution and show the performance of the corresponding generalized Latin hypercube designs, which are compared with commonly used designs, by several numerical examples in Section 3. Finally, some conclusions are given in Section 4.

4.2 Efficient designs for computer experiments

In computer experiments, one of the primary objectives is to construct a meta-model for the estimation of the response at untried inputs. Correspondingly, appropriate optimality criteria for choosing a design in computer experiments should be based on the error of prediction. A very popular optimality criterion is the minimization of the mean square error over the design space. This criterion is adopted to lead to a design which should fill the entire design space uniformly. The examples of space-filling designs include Latin hypercube designs, sphere packing designs, distance based designs, uniform designs, minimal discrepancy designs, designs based on random or pseudo-random sequences, see Santner et al. (2003), Fang et al. (2006). A "good" design should yield a dense set in the projection on each coordinate and, at the same time, should form a dense set in the entire design space. By construction, Latin hypercube designs have projections that are spread out evenly over the values of each input variable. Unfortunately, these designs may have a poor filling of the entire hypercube (in

the extreme case - if the design space is a square - all points are evenly spread out of the diagonal of this square); see e.g. Liefvendahl, Stocki (2006), Pepelyshev (2009) for the analysis of the random LHDs. Several criteria of optimality are introduced in order to choose a "good" design in a class of Latin hypercube designs. Most criteria are based on a measure which quantifies how spread out a set of points is. A typical example is the maximin criterion, which determines the Latin hypercube design maximizing the minimal distance

$$\Psi_p(L) = \min_{\substack{i,j=1,\dots,n \\ i \neq j}} ||x_i - x_j||_p = \min_{\substack{i,j=1,\dots,n \\ i \neq j}} \left(\sum_{s=1}^{d} |x_{s,i} - x_{s,j}|^p \right)^{1/p}, \tag{4.1}$$

where d denotes the dimension and $x_i = (x_{1,i}, \dots, x_{d,i})^T$ is ith point of the design $L = \{x_1, \dots, x_n\}$. Usually this criterion is used with $p = 2$ for computational reasons or the isotropic argument. The maximin objective for the choice of points have been shown to be useful first in Niederreiter (1983). A Latin hypercube design, which maximizes $\Psi_p(L)$ in the class of all LHD, is called a maximin Latin hypercube design (MLHD).

An alternative criterion was introduced by Audze, Eglais (1977) and is based on the analogy of minimizing forces between charged particles. These authors proposed the minimization of the function

$$\Psi_{AE}(L) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{1}{||x_i - x_j||_2^2}. \tag{4.2}$$

Other criteria of uniformity are star L_2-discrepancy, centered L_2-discrepancy, wrap-around L_2-discrepancy which are motivated by quasi-Monte-Carlo methods and the Koksma-Hlawka inequality [see Hickernell (1998), Fang et al. (2000, 2002)].

Uniform designs are particularly important for inference about the mean $\int G(x)dx$ of some known arbitrary integrable function G over the design space. In this case, the sample mean of G, computed from a uniform design, minimizes a certain bound for the absolute error from the true mean [see, for example, Hickernell (1998)]. However, in most computer experiments, one is typically not interested in estimating the mean of some function G. In many cases, the output values of a computer model are used for the construction of a meta-model for prediction. In this meta-model, the parameters have to be estimated. Note that this objective is not taken into account by the choice of uniform designs.

In order to address the problem of prediction and fitting, we propose to choose designs points which are more densely located near the boundary of the design space. In the following discussion, we define a transformation for the projections of a Latin hypercube design. Our proposal is motivated by the following observations.

First, for a uniform design, if a ball with center in the middle of the design space contains k points, then a ball with the same radius and a center near the boundary of

the design space contains fewer points. This effect is increasing with the dimension of the design space. Thus, there is substantially less information for accurate prediction near the boundary of the design space. This statement is confirmed numerically in Tables 4.1–4.3.

Second, in the Gaussian process method in the analysis of computer experiments, the uncertainty of the prediction at a point x is essentially of the form

$$V(x) = 1 - t^T(x)R^{-1}t(x),$$

where $t(x) = (r(x, x_1), \ldots, r(x, x_n))^T$ is the vector of correlations between the point x and the design points $x_1, \ldots, x_n$ and $r(\cdot)$ is the correlation function of Gaussian process. For a one-dimensional input variable, the typical form of the function $V(x)$ with Gaussian correlation function $r(\cdot)$ for the design points $0, 0.1, \ldots, 1$ is given in Figure 4.1. We observe that the uncertainty is larger near the boundary of the design space. Therefore, to reduce the uncertainty (for example, in a minimax approach), the design points should be located more densely near the boundary.

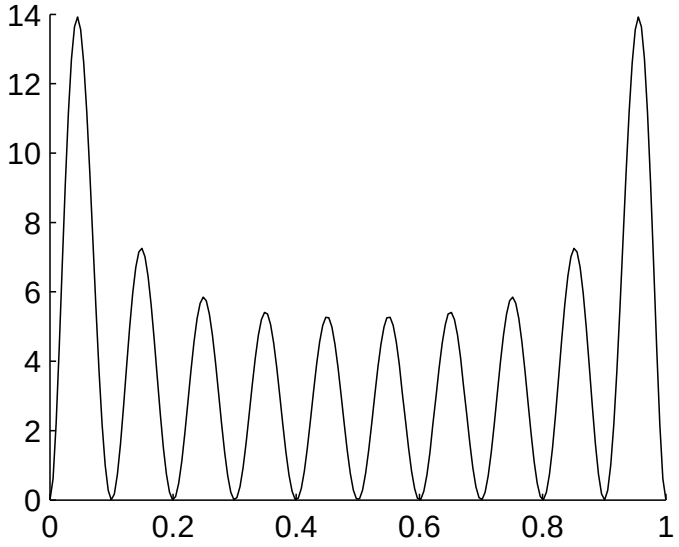

Figure 4.1: *The typical form of the uncertainty in the Gaussian process method using equidistant design points at the interval* $[0, 1]$.

Third, it is known in theory of experimental designs, that the arc-sine distribution is the limit distribution of the D-optimal design for the large degree polynomial regression [see Pukelsheim (2006), Ch. 9.5]. Moreover, Wiens and Zhou (1997) showed that V-optimal designs for linear models, which are robust against small departures from the assumed regression, converge weakly to the arc-sine distribution.

4.2.1 Algorithm of computing a generalized LHD

First, we recall an procedure for the construction of Latin hypercube design, which was introduced in McKay et al. (1979). This procedure generates n points in dimension d in the following manner.

(1) Generate n uniform equidistant points $\bar{x}_{s,1}, \ldots, \bar{x}_{s,n}$ in the range of each input variable, $s = 1, \ldots, d$. Without loss of generality, we assume that the range of each input is the interval $[0, 1]$ and $\bar{x}_{s,j} \in \mathcal{R} = \{0, 1/(n-1), 2/(n-1), \ldots, 1\}$.

(2) Generate a matrix $(p_{i,j})$ of size $d \times n$ such that each row is a random permutation of the numbers $1, \ldots, n$ and these permutations are independent.

(3) Compute $(x_{1,j}, \ldots, x_{d,j}) = (\bar{x}_{1,p_{1,j}}, \ldots, \bar{x}_{d,p_{d,j}})$, the jth point of the Latin hypercube design. Note that the columns of the matrix $(p_{i,j})$ determine the design points.

For constructing a non-uniform design, we propose the following algorithm.

(A) Determine a maximin Latin hypercube design by one of the algorithms which can be found in Liefvendahl and Stocki (2006), Jin et al. (2005), Fang and Qin (2003) among many others.

(B) Define a generalized (non-uniform) design supported at the points $z_1, \ldots, z_n$, $z_i = (z_{1,i}, \ldots, z_{d,i})$, by the transformation of the points $x_{j,i}$ using the quantile function of the Beta density

$$p_{\frac{1+a}{2}}(t) = \frac{1}{B(\frac{1+a}{2}, \frac{1+a}{2})} t^{\frac{a-1}{2}} (1-t)^{\frac{a-1}{2}} \; ; \quad a \in [0, 1].$$

Namely, define $z_{j,i}$ as a solution of the equation

$$x_{j,i} = \int_0^{z_{j,i}} p_{\frac{1+a}{2}}(t) \, \mathrm{d}t$$

$i = 1, \ldots, n$, $j = 1, \ldots, d$, $\{x_{j,1}, \ldots, x_{j,n}\}$ is the one-dimensional projection of a maximin Latin hypercube design.

The resulting design will be called the generalized maximin Latin hypercube design throughout this chapter. By the previous discussion, we expect that a good choice for the distribution in step (B) of the algorithm is to use one of the densities $\{p_{\frac{1+a}{2}}\}_{a \in [0,1]}$. The parameter a can be considered as a tuning parameter, which specifies the importance of the boundary. The larger the parameter $a \in [0, 1]$, the more mass is put at the boundary of the interval $[0, 1]$. Note that the extreme case $a = 1$ yields the uniform distribution, while the other extreme case $a = 0$ is corresponding to the arc-sine

distribution. In the following discussion, we will concentrate on the arc-sine density $p_{\frac{1}{2}}(x)$. For this density, we obtain the simple explicit transformation

$$z_{j,i} = \Big(\cos\Big(x_{j,i}\ \pi\Big) + 1\Big)\Big/2,\ i = 1, \ldots, n.$$

In particular, for the arc-sine transformation, we have

$$\int_0^{1/4} p_{\frac{1}{2}}(t)dt = \int_{1/2-1/4}^{1/2+1/4} p_{\frac{1}{2}}(t)dt = 1/3,$$

which indicates that – in contrast to the uniform distribution – the arc-sine distribution puts more mass in the neighbourhood of the boundary of the design space. In the following section, we investigate the performance of the nonuniform designs by several numerical examples.

4.3 Examples

In this section, we compare two designs: the maximin Latin hypercube design (MLHD) and the generalized maximin Latin hypercube design (GMLHD) obtained from the MLHD by applying the arc-sine transformation. We downloaded maximin Latin hypercube designs from the website http://www.spacefillingdesigns.nl/. The following illustrative examples of models are not hard time consuming computer models. However, they have a typical form of the response and allow to compute the accuracy of the meta-model for different designs.

We use the Bayesian framework, in which the meta-model is the mean of a posterior Gaussian process [see Sacks et al. (1989), Kennedy and O'Hagan (2001)]. More precisely, we consider a Gaussian process with constant or linear mean and the Gaussian correlation function

$$r(x) = e^{-\sum_{s=1}^{d} x_s^2/\psi_s}$$

where $\psi = (\psi_1, \ldots, \psi_d)$ is a vector of correlation parameters. In the following examples, we estimate the hyper-parameter of the Gaussian process for each dataset by the algorithm given in Kennedy and O'Hagan (2001) and Bastos and O'Hagan (2009). Specifically, the parameters of the mean and variance are determined by the integration of the likelihood with respect to the noninformative prior and the estimate of ψ is determined by the maximization of the posterior density.

To study the performance of the meta-model for different designs, we use the mean square error criterion defined by

$$\mathrm{MSE}_\Omega(L) = \int_\Omega \big(\eta(x) - \hat{\eta}(x)\big)^2 w(x)\,\mathrm{d}x, \tag{4.3}$$

where $\hat{\eta}(x)$ is the meta-model, constructed upon the outputs of the computer model $\eta(x)$ at points of the design L, and $w(x)$ is a weight function. For the sake of brevity, we assume that $w(x) \equiv 1$ and calculate the mean square error by Monte-Carlo sampling,

$$\mathrm{MSE}_{\Omega}(L) = \frac{1}{N} \sum_{k=1}^{N} \big(\eta(x_k) - \hat{\eta}(x_k)\big)^2$$

with $N = 10000$, where the points x_k are randomly chosen from Ω. We consider three choices of the domain Ω to study the accuracy of prediction, namely, the full design space $[0,1]^d$, the inscribed hyperball $S(0.5)$ and the smaller hyperball $S(0.4)$, where $S(r) = S_d(r) = \{x : ||x - 0.5 \cdot \mathbf{1}_d||_2 \le r\}$.

4.3.1 Comparison using test functions

The one-dimensional case and, even, the two-dimensional case are not of interest in studying computer experiments. We start the comparison of the MLHD and GMLHD for three-dimensional test functions and consider two models with different curvatures. Let a model having asymptotes be given by

$$\eta(x) = 100(e^{-2/x_1^{1.75}} + e^{-2/x_2^{1.5}} + e^{-2/x_3^{1.25}}). \tag{4.4}$$

Let a model, that is highly curved, be given by

$$\eta(x) = 4(x_1 - 2 + 8x_2 - 8x_2^2)^2 + (3 - 4x_2)^2 + 16\sqrt{x_3 + 1}(2x_3 - 1)^2. \tag{4.5}$$

Comparison of meta-models based on Gaussian processes with constant and linear mean

For the model (4.4), the square root of the mean square errors for the different meta-models are presented in Table 4.1. We observe that, for different domains Ω, the RMSE for the meta-model based on the Gaussian process with constant mean is slightly larger than the RMSE for the meta-model based on the Gaussian process with linear mean.

Recall that the volume of the ball $S_d(r)$ is $\mathrm{Vol}S_d(r) = \pi^{d/2} r^d / \Gamma(d/2 + 1)$. In particular, we have $\mathrm{Vol}S_3(0.5) \approx 0.52$ and $\mathrm{Vol}S_3(0.4) \approx 0.27$. We see in Table 4.1 that, for the MLHD, the RMSE over the 52% central domain $S_3(0.5)$ of 3D-hypercube is smaller than the RMSE over the entire design space. Moreover, the RMSE over the 27% central domain $S_3(0.4)$ is even more smaller. This confirm the heuristic arguments presented in Section 2. In contrast, for the GMLHD, the mean square error is the same for all domains Ω.

Also, we see that $\mathrm{RMSE}_{[0,1]^3}$ for the GMLHD is smaller than for the MLHD, and the difference increases as n increases.

For the model (4.5), the mean square errors for the different meta-models are presented in Table 4.2. Results for the model (4.5) are very similar to results for the model

Table 4.1: *Square root of the mean square error for n-point MLHDs and GMLHDs and the model* (4.4) *which have asymptotes.*

	n	30	40	50	70	90	150
	Gaussian process with constant mean						
MLHD	$\text{RMSE}_{[0,1]^3}$	0.363	0.287	0.312	0.168	0.1681	0.0482
	$\text{RMSE}_{S(0.5)}$	0.267	0.177	0.141	0.054	0.0459	0.0217
	$\text{RMSE}_{S(0.4)}$	0.216	0.163	0.107	0.041	0.0262	0.0114
GMLHD	$\text{RMSE}_{[0,1]^3}$	0.333	0.268	0.232	0.152	0.0604	0.0338
	$\text{RMSE}_{S(0.5)}$	0.330	0.270	0.227	0.119	0.0571	0.0333
	$\text{RMSE}_{S(0.4)}$	0.314	0.266	0.239	0.122	0.0558	0.0339
	Gaussian process with linear mean						
MLHD	$\text{RMSE}_{[0,1]^3}$	0.344	0.263	0.281	0.166	0.1491	0.0512
	$\text{RMSE}_{S(0.5)}$	0.251	0.164	0.135	0.055	0.0435	0.0196
	$\text{RMSE}_{S(0.4)}$	0.205	0.149	0.102	0.044	0.0264	0.0102
GMLHD	$\text{RMSE}_{[0,1]^3}$	0.321	0.243	0.228	0.141	0.0577	0.0303
	$\text{RMSE}_{S(0.5)}$	0.319	0.245	0.233	0.147	0.0575	0.0296
	$\text{RMSE}_{S(0.4)}$	0.305	0.238	0.241	0.151	0.0571	0.0310

(4.4). We observe that, for different domains Ω, the RMSE for the meta-model based on the Gaussian process with constant mean is slightly larger than the RMSE for the meta-model based on the Gaussian process with linear mean. Thus, in further studies, we present results for the meta-model based on the Gaussian process with linear mean.

Again we see that, for the MLHD, the RMSE is smaller in the middle of hypercube. In contrast, for the GMLHD, the RMSE is the same for all domains Ω.

Also, we see that $\text{RMSE}_{[0,1]^3}$ for the GMLHD is smaller than for the MLHD, and the difference increases as n increases.

Table 4.2: *Square root of the mean square error for n-point MLHDs and GMLHDs and the model* (4.5) *which is highly curved.*

	n	30	40	50	70	90	150
	Gaussian process with constant mean						
MLHD	$\text{RMSE}_{[0,1]^3}$	0.947	0.254	0.166	0.103	0.0659	0.0225
	$\text{RMSE}_{S(0.5)}$	0.386	0.083	0.043	0.036	0.0144	0.0087
	$\text{RMSE}_{S(0.4)}$	0.354	0.077	0.036	0.033	0.0139	0.0077
GMLHD	$\text{RMSE}_{[0,1]^3}$	0.815	0.235	0.078	0.062	0.0516	0.0136
	$\text{RMSE}_{S(0.5)}$	0.750	0.244	0.073	0.068	0.0580	0.0141
	$\text{RMSE}_{S(0.4)}$	0.665	0.240	0.080	0.071	0.0595	0.0137
	Gaussian process with linear mean						
MLHD	$\text{RMSE}_{[0,1]^3}$	0.944	0.152	0.128	0.093	0.0595	0.0182
	$\text{RMSE}_{S(0.5)}$	0.403	0.061	0.036	0.033	0.0128	0.0069
	$\text{RMSE}_{S(0.4)}$	0.388	0.060	0.029	0.030	0.0122	0.0058
GMLHD	$\text{RMSE}_{[0,1]^3}$	0.688	0.146	0.064	0.058	0.0237	0.0107
	$\text{RMSE}_{S(0.5)}$	0.648	0.151	0.060	0.063	0.0255	0.0111
	$\text{RMSE}_{S(0.4)}$	0.575	0.150	0.065	0.065	0.0256	0.0113

Comparison for models with stochastic term

Let us consider experiments with a stochastic term. Assume that the output is modelled as the sum of the deterministic model $\eta(x)$ and random perturbations. Assume that the perturbations are independent identically distributed as $\mathrm{N}(0, \sigma^2)$ with $\sigma = 2$. For the construction of the meta-model, we use the Bayesian framework for a Gaussian process with linear mean, Gaussian correlation function and nugget term $\nu = 0.01$. The performance of the meta-model is presented in Table 4.3.

Note that, similar to models without the stochastic term, for the MLHD, the RMSE is smaller in the middle of hypercube. In contrast, for the GMLHD, the RMSE is the same for all domains Ω.

Also, we see that $\mathrm{RMSE}_{[0,1]^3}$ for the GMLHD is smaller than for the MLHD.

Table 4.3: *Square root of the mean square error for n-point MLHDs and GMLHDs and the models* (4.4) *and* (4.5) *with the stochastic term.*

	n	30	40	50	70	90	150
	model (4.4) with stochastic term $\varepsilon \sim \mathrm{N}(0, 2^2)$						
MLHD	$\mathrm{RMSE}_{[0,1]^3}$	2.155	2.089	2.107	2.057	1.935	0.689
	$\mathrm{RMSE}_{S(0.5)}$	1.472	1.660	1.731	1.777	1.678	0.473
	$\mathrm{RMSE}_{S(0.4)}$	1.417	1.540	1.658	1.536	1.531	0.447
GMLHD	$\mathrm{RMSE}_{[0,1]^3}$	1.604	1.754	1.492	1.610	1.790	0.612
	$\mathrm{RMSE}_{S(0.5)}$	1.491	1.736	1.480	1.423	1.776	0.609
	$\mathrm{RMSE}_{S(0.4)}$	1.327	1.747	1.449	1.304	1.741	0.568
	model (4.5) with stochastic term $\varepsilon \sim \mathrm{N}(0, 2^2)$						
MLHD	$\mathrm{RMSE}_{[0,1]^3}$	2.258	1.936	1.988	1.310	1.275	0.968
	$\mathrm{RMSE}_{S(0.5)}$	1.562	1.436	1.395	1.063	0.910	0.689
	$\mathrm{RMSE}_{S(0.4)}$	1.600	1.473	1.361	0.956	0.814	0.592
GMLHD	$\mathrm{RMSE}_{[0,1]^3}$	2.085	1.818	1.173	1.066	0.930	0.937
	$\mathrm{RMSE}_{S(0.5)}$	2.004	1.820	1.135	0.989	0.791	0.782
	$\mathrm{RMSE}_{S(0.4)}$	1.753	1.795	1.153	0.910	0.680	0.612

Comparison using a 8D test function

Let us consider 8-dimensional model

$$\begin{aligned} \eta(x) \;=\; & 4(x_1 - 2 + 8x_2 - 8x_2^2)^2 + (3 - 4x_2)^2 \\ & +16\sqrt{x_3 + 1}(2x_3 - 1)^2 + \sum_{k=4}^{8} k \ln(1 + \sum_{i=3}^{k} x_i) \end{aligned} \tag{4.6}$$

which is highly curved in some variables and showing asymptotes by another variables.

The results of the simulation study are depicted in Table 4.4, where we show the square root of the mean square error for $\Omega = [0,1]^8$. Note that the consideration of the inscribed ball may be misleading since $\mathrm{Vol}S_8(0.5) \approx 0.016$, meanwhile the volume of the whole design space is 1. In general, we have $\mathrm{Vol}S_d(r) \approx r^d(2\pi e/d)^{d/2}$ for a large dimension d. It means that the volume of the inscribed ball is negligibly small in cases with many variables.

From Table 4.4 we see that the mean square error obtained by the MLHD is noticeable larger than the mean square error obtained by the GMLHD. For the model (4.6) without stochastic term, the improvement with respect to $\mathrm{MSE}_{[0,1]^8}$ varies between 50% ($n = 30$) to 20% ($n = 200$). For the model (4.6) with stochastic term, the improvement with respect to $\mathrm{MSE}_{[0,1]^8}$ is about 35%. Thus, for the model (4.6), the GMLHD is substantially more efficient than the MLHD.

Table 4.4: *Square root of the mean square error for n-point MLHDs and GMLHDs and the model* (4.6).

	n	80	90	140	200
	model (4.6) without stochastic term				
MLHD	$\mathrm{RMSE}_{[0,1]^8}$	3.08	1.74	1.07	0.62
GMLHD	$\mathrm{RMSE}_{[0,1]^8}$	1.47	1.08	0.67	0.52
	model (4.6) with stochastic term $\varepsilon \sim \mathrm{N}(0, 4^2)$				
MLHD	$\mathrm{RMSE}_{[0,1]^8}$	5.07	3.50	3.56	3.15
GMLHD	$\mathrm{RMSE}_{[0,1]^8}$	2.88	2.59	2.07	2.02

4.3.2 Borehole model

In this subsection, we examine the performance of designs for a borehole model which was investigated in Worley (1987), Morris and Mitchell (1995), Ahn and Owen (2001), Fang and Lin (2003). The borehole model describes the flow rate through the borehole that is drilled from the ground surface through two aquifers. The flow rate is given by

$$\eta(x) = \frac{2\pi(H_u - H_l)}{\ln(r/r_w)\left(1 + \frac{2LT_u}{\ln(r/r_w)r_w^2 K_w} + \frac{T_u}{T_l}\right)},$$

where the dimension of the vector of input variables is 8, and

$$x = (r_w, r, T_u, T_l, H_u, H_l, L, K_w).$$

The input $r_w \in [0.05, 0.15]$ is the radius of the borehole, $r \in [100, 50000]$ is the radius of the influence, $T_u \in [63070, 115600]$ is the transmissivity of the upper aquifer, $T_l \in [63.1, 116]$ is the transmissivity of the lower aquifer, $H_u \in [990, 1110]$ is the potentiometric head of the upper aquifer, $H_l \in [700, 820]$ is the potentiometric head of the

lower aquifer, $L \in [1120, 1680]$ is the length of the borehole and $K_w \in [9855, 12045]$ is the hydraulic conductivity of the borehole.

The computed square root of the mean square error is depicted in Table 4.5. We see that, for the model without stochastic term, if the mean square error over the full design space is considered, the generalized maximin Latin hypercube designs show a better performance than the initial designs in most cases, in particular, if (compared to the dimension) the number of designs points is relatively large. If the sample sizes are small compared to the dimension (such as $d = 8$ and $n = 80$), the maximin Latin hypercube designs show a slightly better performance. However, in such cases an comparison of two designs with small number of points is adequate only within a narrow predetermined class of computer models. Consequently, numerical results on small-point designs for a particular model can not be transferred to other models.

For the model with stochastic term, the GMLHD provide the smaller value of $\text{MSE}_{[0,1]^8}$ on about 25% with respect to the MLHD. Thus, this is clearly show a superiority of the GMLHD over the MLHD.

Table 4.5: *Square root of the mean square error for n-point MLHD and GMLHD for the borehole model.*

	n	80	90	140	200
	the borehole model without stochastic term				
MLHD	$\text{MSE}_{[0,1]^8}$	0.61	0.53	0.51	0.37
GMLHD	$\text{MSE}_{[0,1]^8}$	0.65	0.54	0.48	0.32
	the borehole model with stochastic term $\varepsilon \sim \text{N}(0, 4^2)$				
MLHD	$\text{MSE}_{[0,1]^8}$	3.63	2.99	2.89	2.52
GMLHD	$\text{MSE}_{[0,1]^8}$	3.02	2.68	2.07	1.89

4.3.3 Mechanical model

In this subsection, we examine designs for a non-linear mechanical model which was studied in Crisfield (1991, p. 2) and Liefvendahl and Stocki (2006). The model consists from the long bar and the spring as given in Figure 1. In the mechanical system, the angle θ is assumed to be small.

We consider the vertical displacement w which is given by a solution of the equation

$$S = \frac{EA}{l^3}(z^2 w + 1.5 z w^2 + 0.5 w^3) + K_s w$$

with the smallest absolute value, where the input $E \in [4.8 \cdot 10^5, 5.2 \cdot 10^5]$ is the Young modulus of the bar material, $A \in [95, 105]$ is the cross-sectional area of the bar, $l \in [2650, 2750]$ is the length of the unloaded bar, $z \in [24.5, 25.5]$ is the vertical coordinate

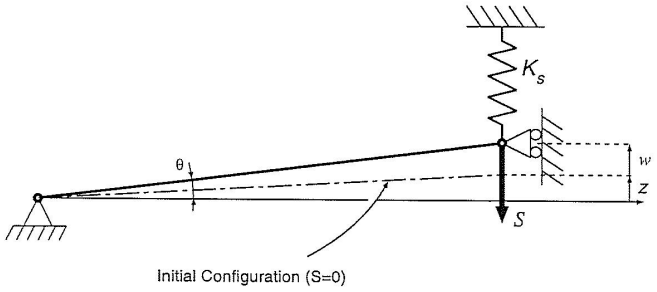

Figure 4.2: *The mechanical system. The force S has negative sign.*

of the right node of the bar when it is unloaded, $K_s \in [0.88, 0.92]$ is the stiffness of the spring, $S \in [-23, -22]$ is the vertical force applied to the right node of the bar. Thus, the model $\eta(x) = w$ has six inputs and $x = (E, A, l, z, K_s, S)$. The computed mean square error for the MLHD and GMLHD is displayed in Table 4.6. We see that the performance of the designs for the mechanical model is similar to the performance for the Borehole model. For the model without stochastic term, the MLHD with a smaller number of design points performs slightly better than the GMLHD, while the opposite behaviour is observed if n is larger than 80. For all considered values of n, the differences between the two designs are rather small.

However, for the model without stochastic term, the GMLHD has substantially a better performance, about 20%, than the MLHD.

Table 4.6: *Square root of the mean square error for n-point MLHD and GMLHD for the mechanical model.*

	n	60	70	80	90	140	200
	the mechanical model without stochastic term						
MLHD	$\text{RMSE}_{[0,1]^6}$	0.55	0.53	0.50	0.48	0.47	0.45
GMLHD	$\text{RMSE}_{[0,1]^6}$	0.60	0.57	0.52	0.47	0.43	0.41
	the mechanical model with stochastic term $\varepsilon \sim \text{N}(0, 2^2)$						
MLHD	$\text{RMSE}_{[0,1]^6}$	1.91	1.87	1.79	1.73	1.45	1.31
GMLHD	$\text{RMSE}_{[0,1]^6}$	1.58	1.62	1.57	1.48	1.27	1.08

4.3.4 Neddermeyer's model

In this subsection, we examine the performance of the designs for a model which was studied by Neddermeyer (1943) and Higdon et al. (2008). The model describes the internal radius of steel cylinder during a high explosive (HE) charge implosion. The

internal radius as a function of time is given by a solution of the differential equation

$$y' = \left[\frac{2}{R_1^2 f(y)^2} \left(\frac{v_0^2}{2} - \frac{s}{2\rho} g(y) \right) \right]^{-1/2},$$

where

$$f(y) = \frac{y^2}{(1-\lambda^2)} \ln \frac{y^2 + 1 - \lambda^2}{y^2},$$

$$g(y) = \frac{1}{1-\lambda^2} (y^2 + 1 - \lambda^2) \ln(y^2 + 1 - \lambda^2) - \lambda^2 \ln(\lambda^2),$$

$R_1 = 1.5$ is the initial outer radius of steel cylinder being imploded, $\lambda = 2/3$ is the ratio of outer cylinder radius to the inner radius, $\rho = 7.5$ is the specific density (relative to water), $v_0 = 0.3 \cdot 10^5$ is the initial velocity imparted on the outer radius of the cylinder from the HE.

We consider the internal radius after an implosion and define the model in the form $\eta(x) = y(T, x)$ with three input variables $x = (s, m, u_0)$. Here $T = 5 \cdot 10^5$ is a time moment after an implosion, the input variable s varies in the interval $[1.85 \cdot 10^{10}, 2.25 \cdot 10^{10}]$ and represents the yield stress of steel, while $m \in [0.29, 0.35]$ is the mass ratio between the HE and cylinder and $u_0 \in [1.4 \cdot 10^{10}, 1.7 \cdot 10^{10}]$ is the detonation energy per gram of exploded gas from the HE. The computed square root of mean square error for the MLHD and GMLHD is shown in Table 4.7. For this model, we observe a convincing superiority of the GMLHD. For the model without stochastic term, the GMLHD yields between 18% ($n = 30$) and 39% ($n = 150$) smaller values for the $\mathrm{MSE}_{[0,1]^3}$ than the MLHD. For the model with stochastic term, the improvement is about 20%.

Table 4.7: *Square root of the mean square error for n-point MLHD and GMLHD for the Neddermeyer's model.*

	n	30	40	50	70	90	150
	the Neddermeyer's model without stochastic term						
MLHD	$\mathrm{RMSE}_{[0,1]^3}$	1.34	1.13	1.08	1.07	0.86	0.90
GMLHD	$\mathrm{RMSE}_{[0,1]^3}$	1.10	0.99	0.86	0.82	0.68	0.55
	the Neddermeyer's model with stochastic term $\varepsilon \sim \mathrm{N}(0, 4^2)$						
MLHD	$\mathrm{RMSE}_{[0,1]^3}$	5.09	4.48	2.80	2.76	2.06	1.88
GMLHD	$\mathrm{RMSE}_{[0,1]^3}$	4.14	3.62	2.26	2.16	1.77	1.64

4.4 Conclusions

In this chapter, we have proposed new designs for computer experiments. The computing of new designs is based on a quantile transformation of the projections of a commonly used space-filling design for computer experiments (such as maximin Latin hypercube, minimum distance, pseudo-random sequences, minimum discrepancy, etc.). For the quantile transformation, we use the arc-sine distribution and, as a consequence, the new designs allocate more input conditions in the neighborhood of the boundary of the design space.

Using (logarithmic) potential theory, some explanation is given why the new designs (obtained by the quantiles of the arc-sine distribution) should perform better than the original uniform space-filling designs. The transformed designs are called generalized maximin Latin hypercube designs (GMLHD). The performance of MLHDs and GMLHDs is illustrated by four examples, which are commonly considered in the literature on computer experiments. The results of comparison of these designs can be summarized as follows.

- In all examples, for models with stochastic term, the square root of the mean square error, obtained by the generalized maximin Latin hypercube designs, is substantially smaller (about 20%-35%) than the square root of the mean square error, obtained by the maximin Latin hypercube designs.

- For models without stochastic term, the generalized maximin Latin hypercube designs show a better performance than the initial designs in most cases, in particular, if (compared to the dimension) the number of designs points is relatively large. If the sample sizes are small compared to the dimension (such as $d = 8$ and $n = 80$), the maximin Latin hypercube designs show a slightly better performance.

- For the MLHD, the mean square error at the middle of the design space is considerably smaller than the mean square error over the full design space. For the GMLHD, the mean square error is almost constant among different domains of the design space.

We believe that the GMLHD is a compromise between designs for parameter estimation/prediction and designs for exploration of the design space. Also, the GMLHD is a compromise between optimal designs for regression experiments and the naive uniform space-filling designs. We shown that GMLHD has the evident superiority in the analysis of computer models with stochastic term. Also, we believe that the GMLHD is a good design at the initial stage of sequential experimentation in the analysis of computer models, that is confirmed in our preliminary investigations.

Bibliography

[1] Al-Mharmah H., Calvin J. M. (1996). Optimal random non-adaptive algorithm for global optimization of brownian motion. J. Global Optim. 8, 81-90.

[2] Audze P., Eglais V. (1977). New approach for planning out of experiments. Problems of Dynamics and Strengths 35, 104–107.

[3] Ahn J., Owen A. (2001). Quasi-regression. Journal of Complexity 17, 588–607.

[4] Bastos L., O'Hagan A. (2009). Diagnostics for gaussian process emulators. Technometrics 51, 425–438.

[5] Crisfield M.A. (1991). Non-linear finite element analysis of solids and structures.

[6] Dette H., Pepelyshev A., Zhigljavsky A. (2009). A new approach to optimal designs for models with correlated observations. Preprint, Ruhr-Universität Bochum. http://www.ruhr-uni-bochum.de/imperia/md/content/mathematik3/publications/pzcorr10.pdf.

[7] Dette, H. and Studden, W. J. (1995). Some new asymptotic properties for the zeros of Jacobi, Laguerre and Hermite polynomials. Constructive Approximation 11, 227–238.

[8] Fang K.-T., Lin D.K.J. (2003). Uniform experimental designs and their application in industry. In Handbook of Statistics, Ed. by R. Khattree and C.R. Rao, 131–170.

[9] Fang K.-T., Qin H. (2003). A note on construction of nearly uniform designs with large number of runs. Statist. Probab. Lett. 61, no. 2, 215–224.

[10] Fang K.-T., Lin D.K.J., Winker P., Zhang Y. (2000). Uniform design: theory and application. Technometrics 42, no. 3, 237–248.

[11] Fang K.-T. Li R., Sudjianto A. (2006). Design and modeling for computer experiments. Chapman & Hall/CRC.

[12] Fang K.-T., Ma C.-X., Winker P. (2002). Centered L_2-discrepancy of random sampling and Latin hypercube design, and construction of uniform designs. Math. Comp. 71, no. 237, 275–296.

[13] Hickernell F.J. (1998). A generalized discrepancy and quadrature error bound. Math. Comp. 67, no. 221, 299–322.

[14] Higdon D., Gattiker J., Williams B., Rightley M. (2008). Computer model calibration using high-dimensional output. J. Amer. Stat. Ass. 103, 570–583.

[15] Jin R., Chen, W., Sudjianto A. (2005). An efficient algorithm for constructing optimal design of computer experiments. J. Statist. Plann. Inf. 134, 268–287.

[16] Kennedy M.C., O'Hagan A. (2001). Bayesian calibration of computer models. J. R. Stat. Soc. Ser. B 63, no. 3, 425–464.

[17] Kleijnen J.P.C. (2008). Design and Analysis of Simulation Experiments. Springer, New York.

[18] Liefvendahl M., Stocki R. (2006). A study on algorithms for optimization of Latin hypercubes. J. Statist. Plann. Inference 136, 3231–3247.

[19] McKay M. D., Beckman R. J., Conover W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21, no. 2, 239–245.

[20] Morris M.D., Mitchell T.J. (1995). Exploratory designs for computer experiments J. Stat. Plan. Inf. 43, 381–402.

[21] Neddermeyer S. (1943). Collapse of hollow steel cylinders. Los Alamos Report, LA-18, Los Alamos National Laboratory.

[22] Niederreiter H. (1983). A quasi-Monte Carlo method for the approximate computation of the extreme values of a function. Studies in pure mathematics, 523–529, Birkhauser, Basel.

[23] Pepelyshev A. (2009). Improvement of random LHD for high dimensions. Proceedings of the 6th St. Petersburg Workshop on Simulation, VVM, St. Petersburg, 1091–1096.

[24] Pukelsheim F. (2006) Optimal Design of Experiments. Society for Industrial and Applied Mathematics (SIAM), Philadelphia.

[25] Saff E.B., Totik V. (1997). Logarithmic potentials with external fields. Springer-Verlag, Berlin.

[26] Sacks J., Welch W.J., Mitchell T.J., Wynn H.P. (1989). Design and analysis of computer experiments. With comments and a rejoinder by the authors. Statist. Sci. 4, no. 4, 409–435.

[27] Santner T.J., Williams B.J., Notz W. (2003). The Design and Analysis of Computer Experiments.

[28] Szegö, G. (1975). Orthogonal Polynomials. Amer. Math. Soc. Colloq. Publ., Vol. 23, Amer. Math. Soc., NY.

[29] Wiens D.P., Zhou J. (1997). Robust Designs Based on the Infinitesimal Approach. Journal of the American Statistical Association 92, 1503–1511.

[30] Worley B.A. (1987). Deterministic uncertainty analysis. OSTI ID: 6104504.

Chapter 5

Use of Mahalanobis distance and pivoted Cholesky decomposition diagnostics

In this chapter, we consider the use of Mahalanobis distance and pivoted Cholesky decomposition diagnostics from Bastos and O'Hagan (2009) to test validity of the emulators.

Recall that one need a validation set of output values of the model. Denote this set by $\{x_i^*, y_i^*\}_{i=1,\ldots,k}$. Recall that the Mahalonobis distance is defined by

$$\text{MD} = (y^* - \mathbf{E}\eta(x^*|y))^T V_\eta^{-1}(x^*|y)(y^* - \mathbf{E}\eta(x^*|y))$$

where $y^* = (y_1, \ldots, y_k)^T$ and $x^* = (x_1, \ldots, x_k)^T$. The vector of pivoted Cholesky errors is defined by

$$E_{PC} = (\varepsilon_1^{PC}, \ldots, \varepsilon_k^{PC}) = G^{-1}(y^* - \mathbf{E}\eta(x^*|y))$$

where $G = PR^T$, $P^T V_\eta(x^*|y)P = R^T R$, P is a permutation matrix corresponded to the pivoted Cholesky decomposition, and R is the unique upper triangular matrix.

Define

$$\text{RGPCE} = \frac{1}{k}\#\Big\{i \in \{1, \ldots, k\} \,:\, |\varepsilon_i^{PC}| < 2\Big\}.$$

and

$$\text{SD} = \text{std.dev.}(E_{PC})$$

In further numerical studies, we use the validation set at points of the 10-, 12- and 15-point maximin Latin hypercube designs.

5.1 Test function 1

Consider the model

$$\eta(x) = \left(1 - e^{-0.5/x_2}\right) \frac{2300x_1^3 + 1900x_1^2 + 2092x_1 + 60}{100x_1^3 + 500x_1^2 + 4x_1 + 20}, \tag{5.1}$$

where the input variable $x = (x_1, x_2)$ varies in the square $[0, 1]^2$.

The performance of the emulator for the model (5.1) is presented in Table 5.1.

Table 5.1: *The characteristics of GP emulators with Gaussian correlation function and linear mean for the n-point maximin LHD and the model* (5.1).

n	MD	SD	RGPCE	RMSE	RBCI
	10-point LHD for validation				
30	13.7	1.22	0.9	0.39	0.95
40	84.9	2.90	0.6	0.16	0.93
50	74.9	2.09	0.6	0.06	0.84
60	38.4	1.63	0.7	0.15	0.97
70	2772.8	17.27	0.5	0.07	0.37
80	1607.0	19.45	0.3	0.05	0.67
	12-point LHD for validation				
30	76.96	2.62	0.8	0.39	0.95
40	154.05	3.74	0.5	0.16	0.93
50	50.00	2.04	0.7	0.06	0.84
60	52.82	2.19	0.8	0.15	0.97
70	2431.45	42.46	0.6	0.07	0.37
80	467.76	6.63	0.5	0.05	0.67
	15-point LHD for validation				
30	75.25	2.30	0.7	0.39	0.95
40	99.98	2.67	0.5	0.16	0.93
50	1016.34	7.63	0.2	0.06	0.84
60	167.67	3.37	0.7	0.15	0.97
70	11291.17	24.89	0.6	0.07	0.37
80	801.32	9.63	0.3	0.05	0.67

5.2 Test function 2

Consider the model

$$\eta(x) = 100(e^{-2/x_1^{1.75}} + e^{-2/x_2^{1.5}} + e^{-2/x_3^{1.25}}), \tag{5.2}$$

where the input variable $x = (x_1, x_2, x_3)$ varies in the square $[0, 1]^3$.

The performance of the emulator for the model (5.2) is presented in Table 5.2.

Table 5.2: *The characteristics of GP emulators with Gaussian correlation function and linear mean for the n-point maximin LHD and the model* (5.2).

n	MD	SD	RGPCE	RMSE	RBCI
	10-point LHD for validation				
30	195.7	4.65	0.4	0.34	0.74
40	52.7	2.39	0.7	0.24	0.90
50	40.7	2.12	0.7	0.28	0.81
70	30.5	1.75	0.7	0.13	0.88
90	24.7	1.60	0.8	0.16	0.99
150	73.9	2.74	0.7	0.05	0.84
	12-point LHD for validation				
30	139.97	3.09	0.4	0.34	0.74
40	108.74	3.10	0.4	0.24	0.90
50	49.56	1.97	0.7	0.28	0.81
70	99.81	2.91	0.5	0.13	0.88
90	7.82	0.83	1.0	0.16	0.99
150	107.50	3.04	0.7	0.05	0.84
	15-point LHD for validation				
30	267.21	4.36	0.3	0.34	0.74
40	76.39	2.33	0.5	0.24	0.90
50	87.20	2.48	0.6	0.28	0.81
70	104.68	2.66	0.5	0.13	0.88
90	30.08	0.98	0.9	0.16	0.99
150	94.60	2.56	0.5	0.05	0.84

5.3 Test function 3

Consider the model

$$\eta(x) = 4(x_1 - 2 + 8x_2 - 8x_2^2)^2 + (3 - 4x_2)^2 + 16\sqrt{x_3 + 1}(2x_3 - 1)^2. \tag{5.3}$$

where the input variable $x = (x_1, x_2, x_3)$ varies in the square $[0, 1]^3$.

The performance of the emulator for the model (5.3) is presented in Table 5.3.

Table 5.3: *The characteristics of GP emulators with Gaussian correlation function and linear mean for the n-point maximin LHD and the model* (5.3).

n	MD	SD	RGPCE	RMSE	RBCI
	10-point LHD for validation				
30	24.8	1.63	0.9	1.03	0.91
40	9.9	1.54	0.8	0.10	0.80
50	240.6	0.12	1.0	0.09	0.92
	12-point LHD for validation				
30	38.48	1.08	0.8	1.03	0.91
40	13.69	0.70	1.0	0.10	0.80
50	0.70	0.27	1.0	0.09	0.92
	15-point LHD for validation				
30	56.59	1.84	0.7	1.03	0.91
40	42.96	0.40	1.0	0.10	0.80
50	19.13	0.14	1.0	0.09	0.92

5.4 Conclusions

Upon a broad numerical study, we obtain the following conclusions.

- There is a similarity of MD, SD and RGPCE for different sizes of validation set.
- It is better (in average) to check the emulator validity for small sizes of validation set.
- Mahalonobis distance typically provide approximate information about validity of emulator.
- Pivoted Cholesky decomposition yields better diagnostic than one through Mahalonobis distance.

Chapter 6

Comparison with the 'Durham approach'

In this chapter, we do a comparison with the 'Durham approach' of using a more complex mean function in combination with a single correlation function parameter, estimated using variogram methods, on emulator prediction and validity.

6.0.1 Bayesian approach

Recall the conception of the emulator $\eta(x)$. The emulator is represented by Gaussian process with the mean $m_e(\cdot)$ and covariance function $V_e(\cdot,\cdot)$, i.e. *Gaussian process emulator* is

$$\eta(x)|\beta,\sigma^2,\psi \sim \mathrm{GP}(m_e(\cdot), V_e(\cdot,\cdot)) \tag{6.1}$$

where $x = (x_{1'},\ldots,x_{d'}) \in \mathbb{R}^d$ is a set of inputs, the mean function have a form

$$m_e(x) = \beta^T h(x)$$

where $h(x)$ is a vector of known function, $\beta \in \mathbb{R}^q$ is a vector of unknown parameters and the output is one-dimensional. We assume that the covariance function $V_e(\cdot,\cdot)$ is given by

$$V_e(x,\tilde{x}) = \sigma^2 R(x,\tilde{x}|\psi)$$

where $R(x,\tilde{x}) = R(x,\tilde{x}|\psi)$ is a known correlation function and σ and ψ are unknown parameters. A typical choice of $h(x)$ is

$$h(x) \;=\; (1, x_{1'},\ldots,x_{d'})^T.$$

A typical choice of $R_s(x,\tilde{x})$ is

$$R(x,\tilde{x}|\psi) = e^{-\sum\limits_{i=1}^{d} \frac{(x_{i'}-\tilde{x}_{i'})^r}{\psi_i}} \tag{6.2}$$

with $r = 2$ where $\psi = (\psi_1, \ldots, \psi_p)^T$ is called as *smoothing parameters* or *squared correlation length parameters*. A process with the correlation function (6.2) is infinitely mean square differentiable for $r = 2$ and is not mean square differentiable for $0 < r < 2$.

Let $y = (y_1, \ldots, y_n)$ be a vector of simulator outputs at points $x_1, \ldots, x_n$, $y_i = \eta(x_i) = \eta(x_i, \beta, \sigma^2, \psi)$. The pairs (x_i, y_i) are called *the training dataset* or *the calibration data* since we use them to estimate unknown parameters β, σ^2, ψ.

A predictor is given by *the posterior emulator* which is associated with statistical presentation of the simulator. Namely, the posterior emulator of (6.1) is given by (conditional) Gaussian process

$$\eta(x)|\beta, \sigma^2, \psi, y \sim \mathrm{GP}(m_p(x), V_p(x, \tilde{x})). \tag{6.3}$$

For computing the posterior emulator we should specify a prior distribution for parameters β, σ^2, ψ. We assume that densities are given by weak priors

$$p_a(\beta, \sigma^2) \propto \sigma^{-2}, \quad p_a(\psi) \propto 1.$$

Note that calculations are considerably simplified for these priors.

The algorithm for computing Gaussian posterior emulator is the following.

1. First, define the posterior density for ψ

$$p_p(\psi|y) \propto |R|^{-1/2} |H^T R^{-1} H|^{-1/2} \left(\hat{\sigma}^2\right)^{-(n-q)/2} \tag{6.4}$$

   where

$$\hat{\sigma}^2 = \frac{1}{n-q-2} y^T \left(R^{-1} - R^{-1} H (H^T R^{-1} H)^{-1} H^T R^{-1}\right) y, \tag{6.5}$$

   $H = (h(x_1), \ldots, h(x_n))^T$ is the design matrix and $R = (R(x_i, x_j|\psi))_{i,j=1}^n$ is the covariance matrix.

2. Second, calculate the maximum likelihood estimator of ψ, i.e. find $\hat{\psi}$ such that maximize (6.4) and let $R = R(\hat{\psi})$.

3. Calculate the posterior mean of β by

$$\hat{\beta} = (H^T R^{-1} H)^{-1} H^T R^{-1} y. \tag{6.6}$$

4. Third, compute the posterior mean of σ^2 by (6.5).

5. Then, define the mean of posterior emulator by

$$m_p(x) = \hat{\beta}^T h(x) + t(x) R^{-1} (y - H\beta) \tag{6.7}$$

   where $t(x) = (R(x, x_1|\hat{\psi}), \ldots, R(x, x_n|\hat{\psi}))^T$.

6. After that, define the covariance function of posterior emulator by

$$V_p(x,\tilde{x}) = \hat{\sigma}^2 \left(R(x,\tilde{x}|\hat{\psi}) - t^T(x)R^{-1}t(\tilde{x}) + s^T(x)(H^TR^{-1}H)^{-1}s(\tilde{x}) \right) \quad (6.8)$$

where $s(x) = h(x) - H^TR^{-1}t(x)$.

For a graphical representation of posterior emulator, we plot the mean and its 95%-credible interval which is calculated by

$$(m_p(x) - \gamma\sqrt{V_p(x,x)}, m_p(x) + \gamma\sqrt{V_p(x,x)}) \quad (6.9)$$

where γ is a 0.975-quintile of Student distribution with $(n-q)$ d.f. Note that $V_p(x_i,x_i) = 0$, $i = 1,\ldots,n$.

6.0.2 Kriging approach

Firstly, we note that the Kriging and 'Durham' approaches have similar computational procedures, but they have different motivations. In this subsection, we use the motivation of the Kriging approach. In this approach, the model is given by

$$\eta(x) = m(x) + Z(x)$$

where $Z(x)$ is a stationary process with zero mean and known covariance function. The mean $m(x) = \beta^T f(x)$ is assumed to be constructed by stepwise technique and $Z(x)$ is a isotropic process with gaussian covariance function

$$V(x,\tilde{x}|\psi_0) = \sigma^2 e^{-\frac{1}{\psi_0}\sum\limits_{i=1}^{d}(x_{i'}-\tilde{x}_{i'})^2}$$

with unknown parameter ψ_0. Let the parameter β be estimated by ordinary least squares,

$$\hat{\beta} = (F^TF)^{-1}F^TY$$

where $Y = (y_1,\ldots,y_n)^T$ and $F = (f(x_1),\ldots,f(x_n))^T$. Let the correlation parameter be estimated by variogram method for the residuals $Z_i = y_i - \hat{\beta}^T f(x_i)$, see formula (2.4.12) and (2.6.12) in Cressie (1993); namely, the estimator of variogram is given by

$$2\hat{\gamma}(h) = \frac{\left(\frac{1}{|N(h)|}\sum_{N(h)}|Z_i - Z_j|^{1/2}\right)^4}{0.457 + 0.494/|N(h)|}$$

where $|N(h)|$ means the number of elements in the set $N(h) = \{(i,j) : ||x_i - x_j|| = h,\ i,j = 1,\ldots,n\}$, and the parameter ψ_0 is estimated by minimizing the weighted least squares

$$\sum_{j=1}^{K}|N(h_j)|\left(\frac{\hat{\gamma}(h_j)}{\gamma(h_j,\psi_0)} - 1\right)^2.$$

Finally, due to linear Bayesian approach, the posterior emulator is given by

$$m_p(x) = \hat{\beta}^T f(x) + t^T(x) V^{-1}(y - F\hat{\beta})$$

where $V = (V(x_i, x_j|\hat{\psi}))_{i,j=1}^n$ is the covariance matrix and

$$t(x) = (V(x, x_1|\hat{\psi}_0), \ldots, V(x, x_n|\hat{\psi}_0))^T.$$

The covariance function of posterior emulator is given by

$$V_p(x, \tilde{x}) = V(x, \tilde{x}|\hat{\psi}_0) - t^T(x) V^{-1} t(\tilde{x}).$$

6.1 One-dimensional example

Let the first dataset be given by

x_i	0	0.25	0.5	0.75	1
y_i	0.28	0.37	0.5	0.19	0.07

(6.10)

and the second dataset be given by

x_i	0	0.25	0.5	0.75	1
y_i	0	0.03	0.30	0.50	0.59

(6.11)

Let $h(x) = (1, x)^T$ in Bayesian approach and $f(x) = (1, x, x^2)^T$ in Kriging approach.

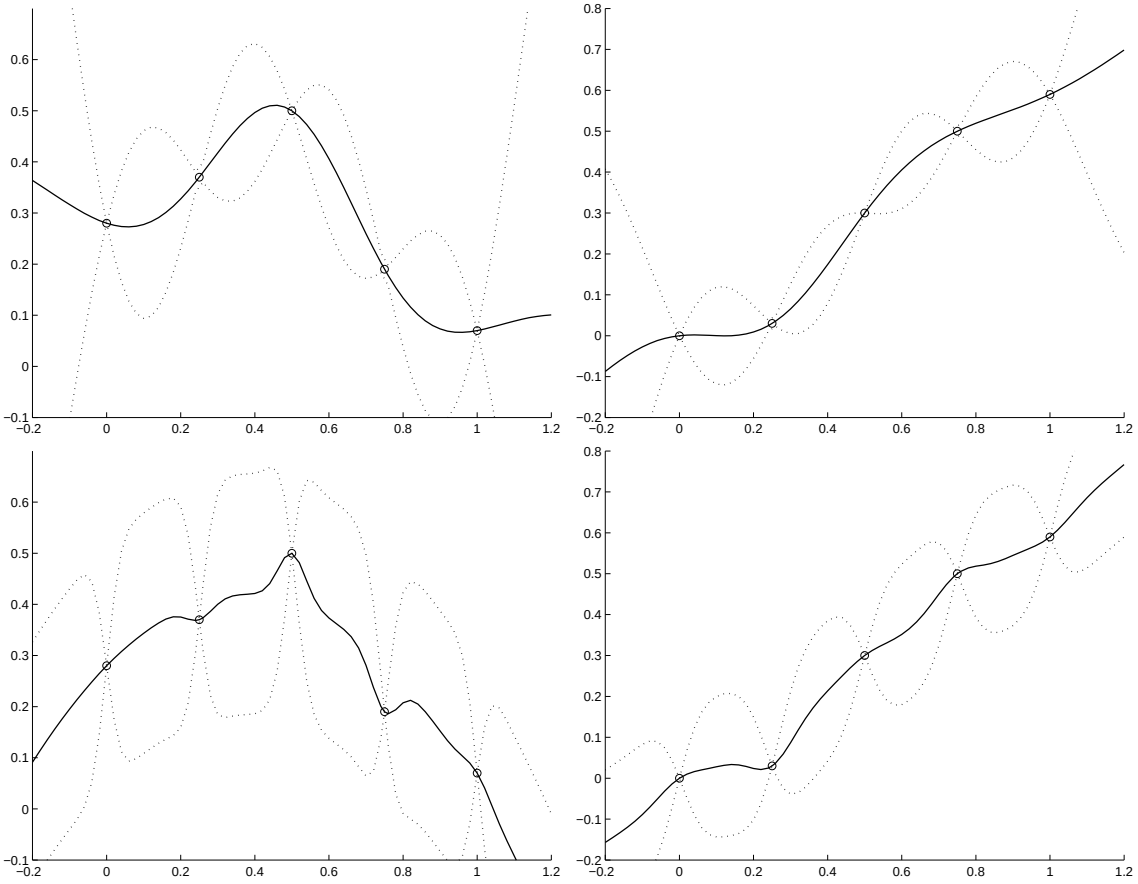

Figure 6.1: Prediction and uncertainty for dataset (6.10) (left column), dataset (6.11) (right column) for Bayesian approach (top row), Kriging approach (bottom row).

Let the third dataset be given by

x_i	0	0.25	0.35	0.5	0.6	0.75	1
y_i	0.28	0.37	0.43	0.5	0.41	0.19	0.07

(6.12)

and the forth dataset be given by

x_i	0	0.25	0.35	0.5	0.6	0.75	1
y_i	0	0.03	0.2	0.30	0.4	0.50	0.59

(6.13)

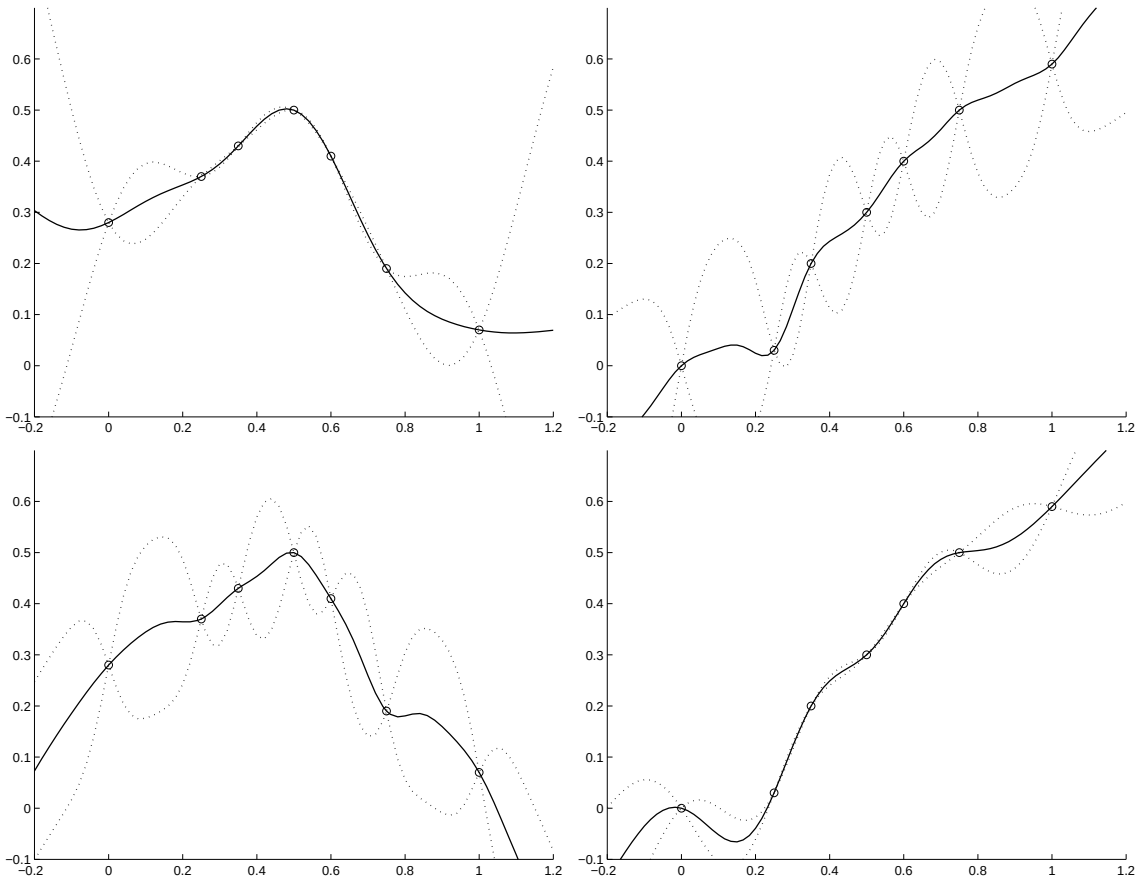

Figure 6.2: Prediction and uncertainty for dataset (6.12) (left column), dataset (6.13) (right column) for Bayesian approach (top row), Kriging approach (bottom row).

Let the firth dataset be given by

x_i	0	0.1	0.25	0.5	0.75	0.9	1
y_i	0.28	0.33	0.37	0.5	0.19	0.1	0.07

(6.14)

and the sixth dataset be given by

x_i	0	0.1	0.25	0.5	0.75	0.9	1
y_i	0	0.02	0.03	0.30	0.50	0.59	0.59

(6.15)

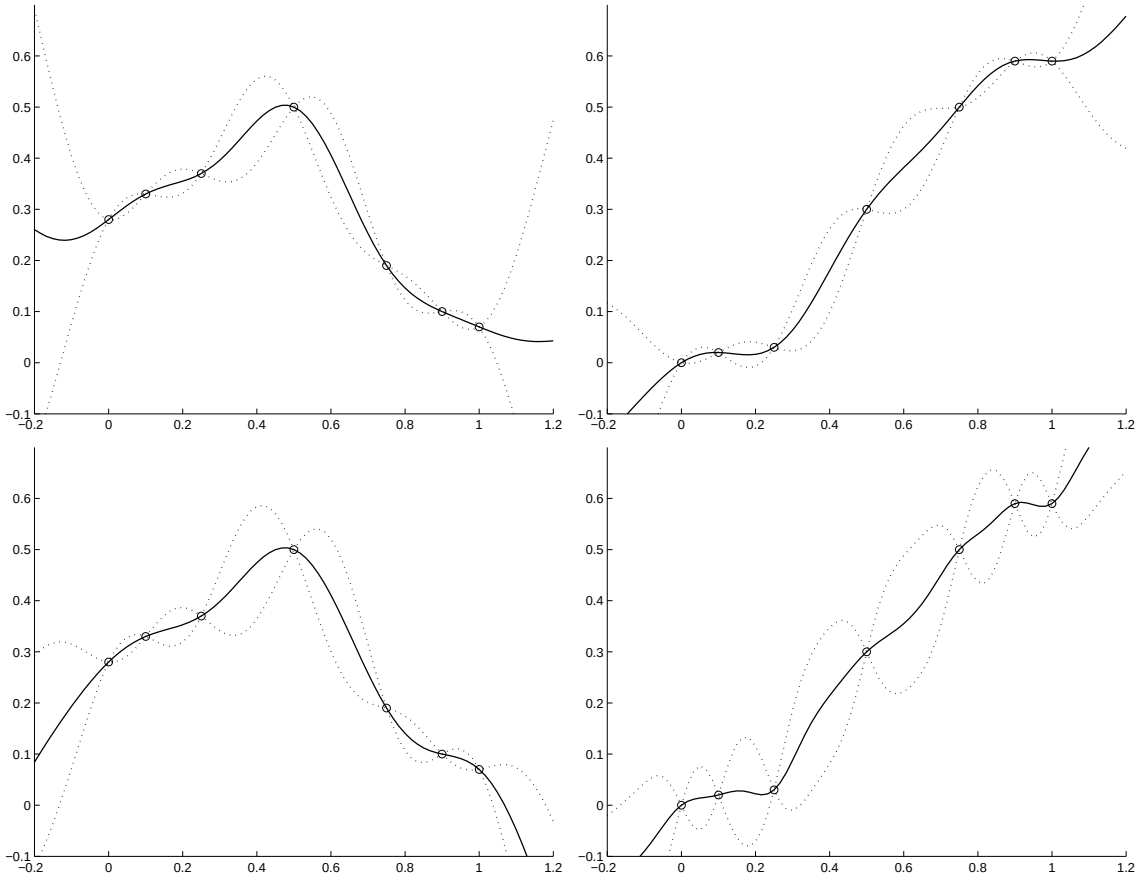

Figure 6.3: Prediction and uncertainty for dataset (6.14) (left column), dataset (6.15) (right column) for Bayesian approach (top row), Kriging approach (bottom row).

Let the seventh dataset be given by

x_i	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
y_i	0.0	0.26	0.33	0.43	0.47	0.5	0.30	0.20	0.10	0.05	0

(6.16)

and the eighth dataset be given by

x_i	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
y_i	0.0	0.16	0.23	0.33	0.37	0.4	0.45	0.47	0.50	0.55	0.50

(6.17)

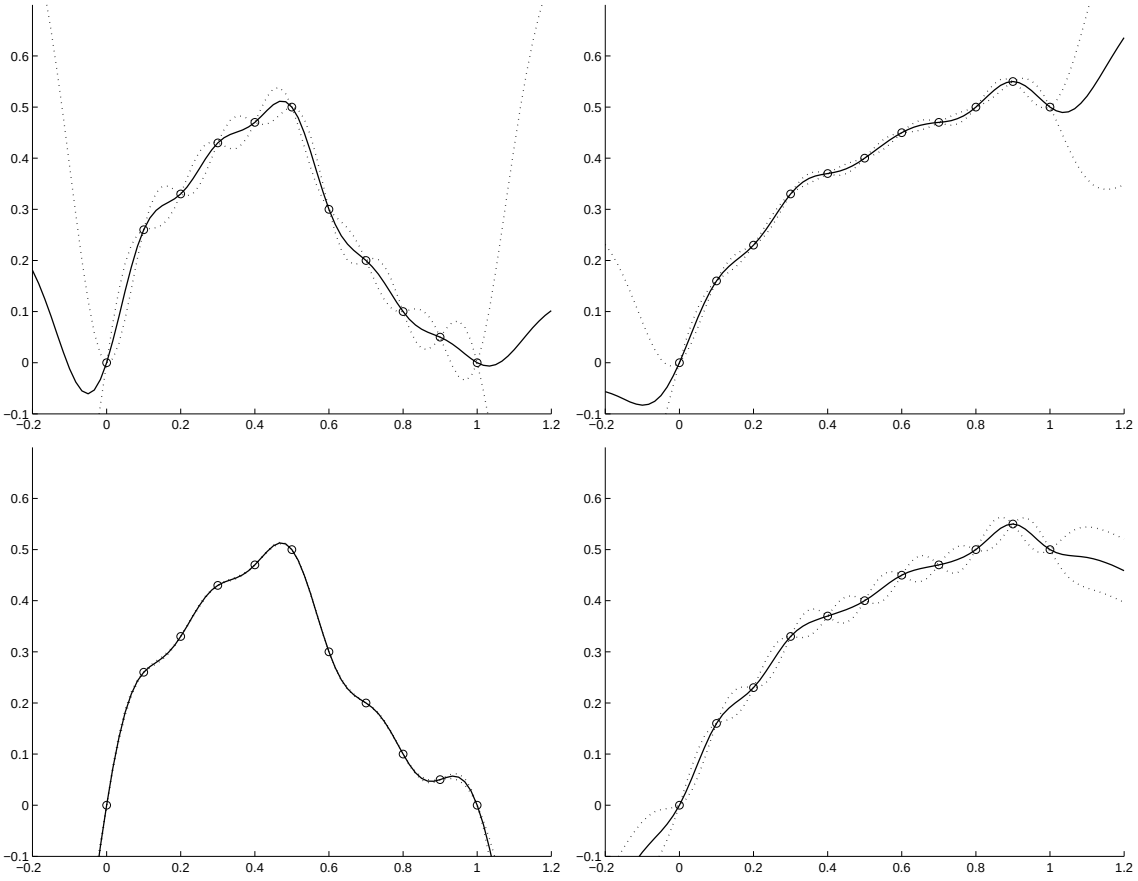

Figure 6.4: Prediction and uncertainty for dataset (6.16) (left column), dataset (6.17) (right column) for Bayesian approach (top row), Kriging approach (bottom row).

6.2 Two-dimensional example

Let us consider a two-dimensional toy model which have a typical form in computer experiments. Let the model is determined by

$$\eta(x) = \left(1 - e^{-0.5/x_{2'}}\right) \frac{2300x_{1'}^3 + 1900x_{1'}^2 + 2092x_{1'} + 60}{100x_{1'}^3 + 500x_{1'}^2 + 4x_{1'} + 20} \tag{6.18}$$

where $x = (x_{1'}, x_{2'}) \in [0,1]^2$. In particular, the model (6.18) was studied in Bastos, O'Hagan, (2008). Let experiment be performed at points of 30-point maximin Latin hypercube design. For the sake of visualization, in Figure 6.5 we plot a predictor with respect to (w.r.t.) one variable when other variable is fixed at some values along with true model (dashed line). Let $h(x) = (1, x_{1'}, x_{2'})^T$ and

$$f(x) = (1, x_{1'}, x_{2'}, x_{1'}^2, x_{1'}x_{2'}, x_{2'}^2, x_{1'}^3, x_{1'}^2x_{2'}, x_{1'}x_{2'}^2, x_{2'}^3)^T. \tag{6.19}$$

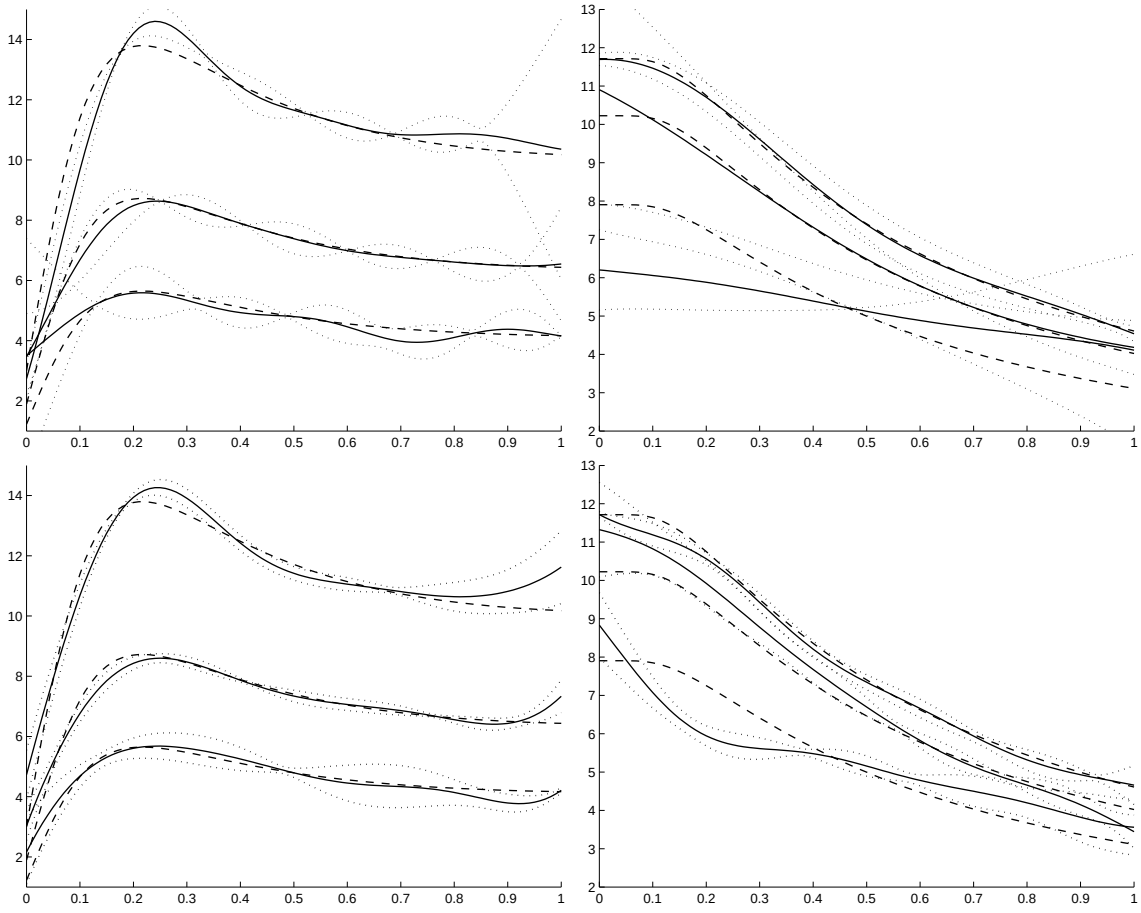

Figure 6.5: Prediction and uncertainty for model (6.18) w.r.t. $x_{1'}$ with $x_{2'} = 0.05, 0.5, 0.95$ (left column), w.r.t. $x_{2'}$ with $x_{1'} = 0.05, 0.5, 0.95$ (right column) for Bayesian approach (top row), Kriging approach (bottom row).

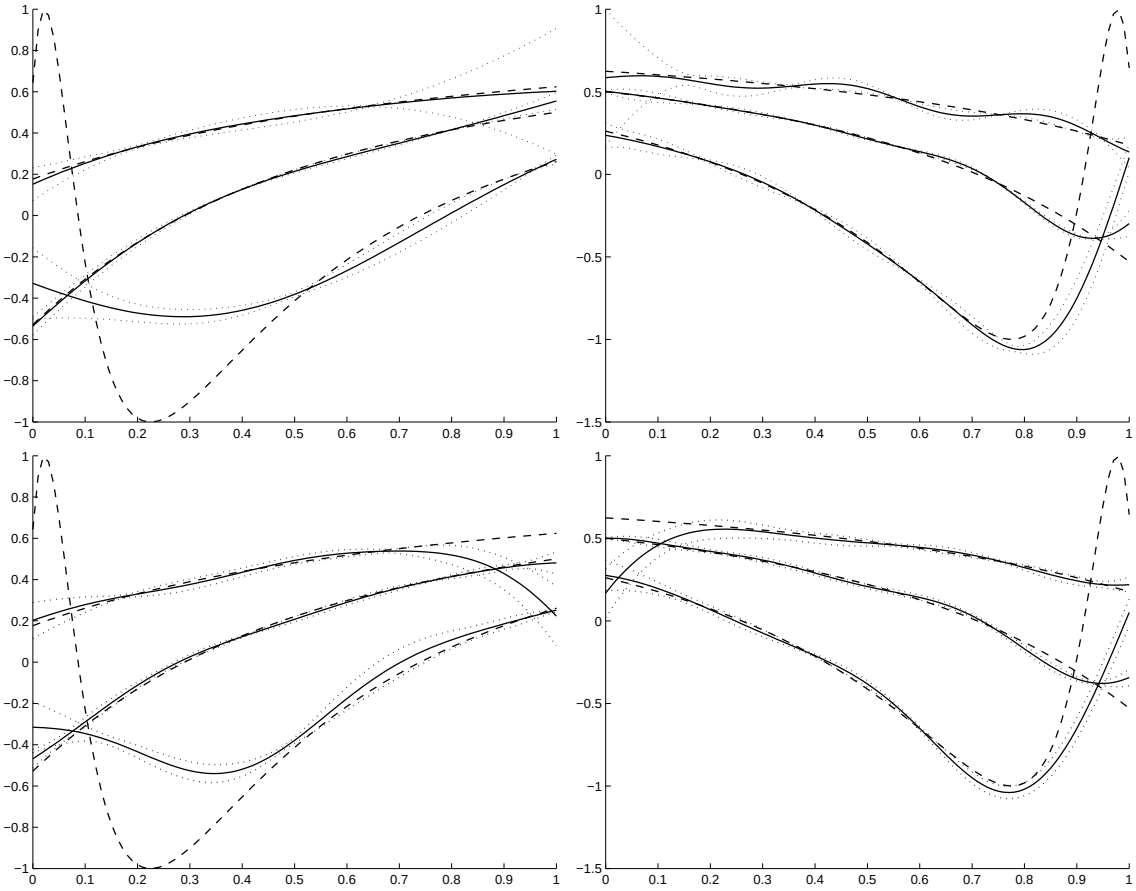

Figure 6.6: Prediction and uncertainty for model $y = \cos(1/\ln(x_{1'} - x_{2'} + 2.1))$ w.r.t. $x_{1'}$ with $x_{2'} = 0.05, 0.5, 0.95$ (left column), w.r.t. $x_{2'}$ with $x_{1'} = 0.05, 0.5, 0.95$ (right column) for Bayesian approach (top row), Kriging approach (bottom row).

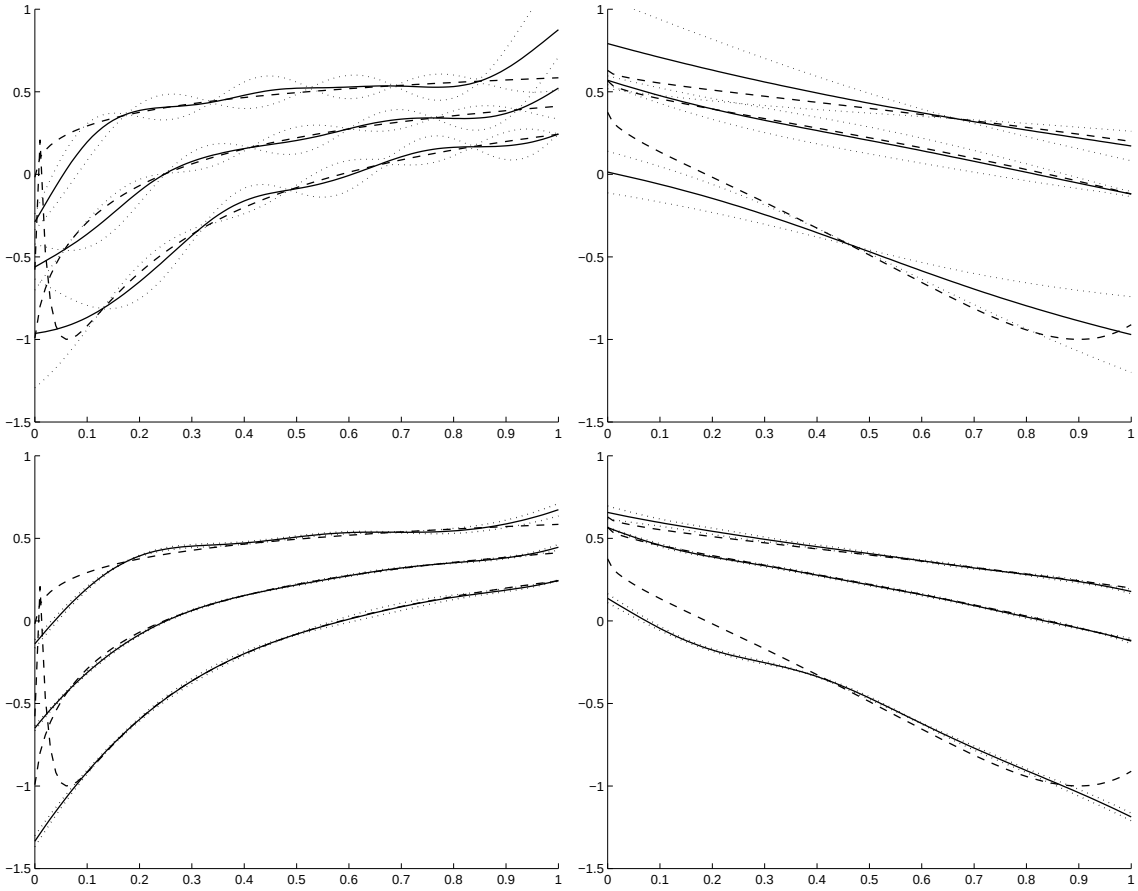

Figure 6.7: Prediction and uncertainty for model $y = \cos(1/\ln(x_{1'}^{0.5} - x_{2'}^{0.5} + 2.1))$ w.r.t. $x_{1'}$ with $x_{2'} = 0.05, 0.5, 0.95$ (left column), w.r.t. $x_{2'}$ with $x_{1'} = 0.05, 0.5, 0.95$ (right column) for Bayesian approach (top row), Kriging approach (bottom row).

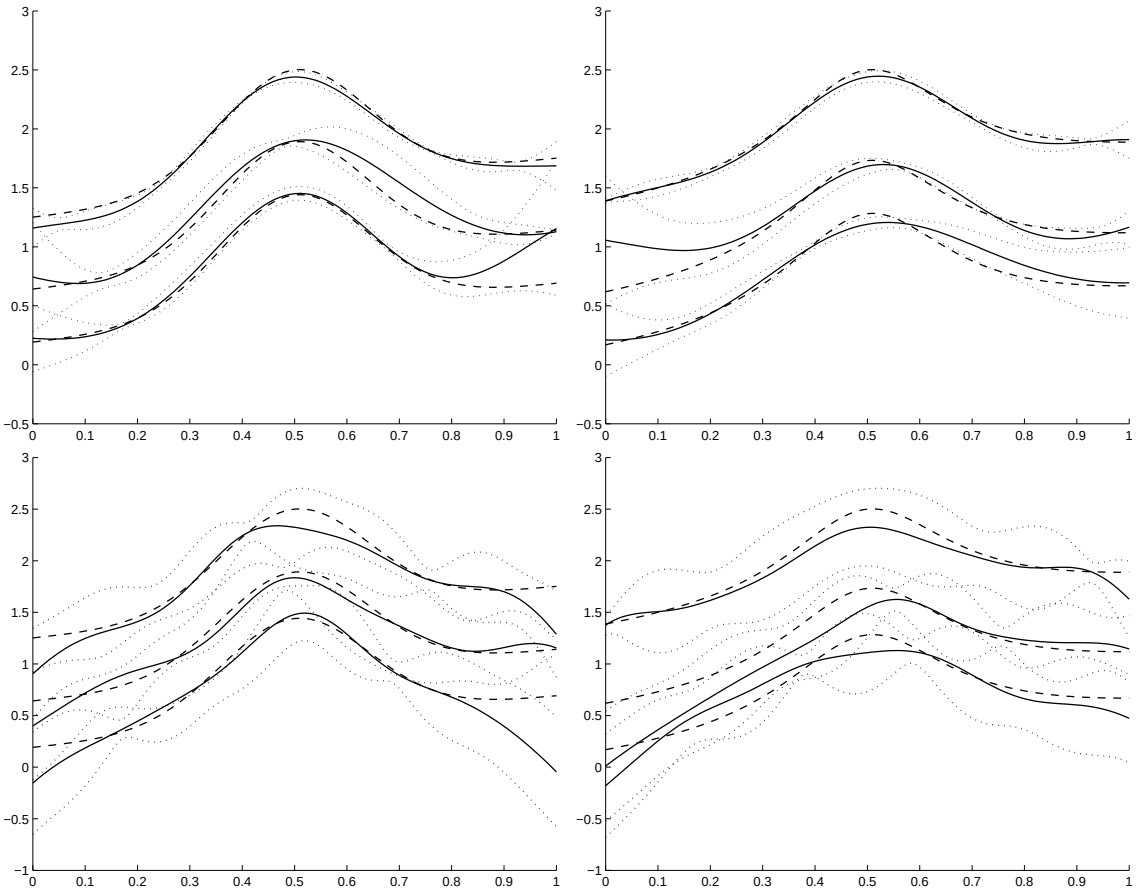

Figure 6.8: Prediction and uncertainty for model $y = x_{1'}/2 + x_{2'}/2 + e^{-25(x_{1'}-0.5)^2} + 1/(1+25(x_{2'}-0.5)^2)$ w.r.t. $x_{1'}$ with $x_{2'} = 0.05, 0.5, 0.95$ (left column), w.r.t. $x_{2'}$ with $x_{1'} = 0.05, 0.5, 0.95$ (right column) for Bayesian approach (top row), Kriging approach (bottom row).

6.3 Conclusions

For one-dimensional case, the true output of model belongs to the confidence interval of the emulators in both approaches. However, the size of the confidence interval strongly depends on a data. For some data, the size of the confidence interval of the emulator constructed by the Kriging approach is smaller than the size of the confidence interval of the emulator constructed by the Bayesian approach. For another data, the relation is the opposite.

For two-dimensional case, the true output of model does not belongs to the confidence interval of the emulators in both approaches. Again, the size of the confidence interval strongly depends on a data. The Bayesian approach is slightly better than the Kriging approach.

Chapter 7

Investigation of the effect of including additional terms in the prior mean function

In this chapter, we investigate the effect of including additional terms in the prior mean function on both correlation parameter estimation and emulator prediction and validity. We present tables with the estimators of the correlation parameter and the variance for data collected by n-point maximin LHD. We also present the RMSE criterion and the RBCI criterion to show the performance of the emulator prediction and the emulator validity, respectively. We consider the mean of the Gaussian process in the following forms: the constant meant

$$m(x, \beta) = \beta_0,$$

the linear mean

$$m(x, \beta) = \beta_0 + \sum_{i=1}^{d} \beta_i x_i$$

the pure quadratic mean

$$m(x, \beta) = \beta_0 + \sum_{i=1}^{d} \beta_i x_i + \sum_{i=1}^{d} \beta_{ii} x_i^2,$$

and the full quadratic mean

$$m(x, \beta) = \beta_0 + \sum_{i=1}^{d} \beta_i x_i + \sum_{1 \le i \le j \le d} \beta_{ij} x_i x_j.$$

7.1 Test function 1

Consider the model

$$\eta(x) = \left(1 - e^{-0.5/x_2}\right) \frac{2300x_1^3 + 1900x_1^2 + 2092x_1 + 60}{100x_1^3 + 500x_1^2 + 4x_1 + 20}, \tag{7.1}$$

where the input variable $x = (x_1, x_2)$ varies in the square $[0, 1]^2$.

Table 7.1: *The estimators of parameters and characteristics of GP emulators with Gaussian correlation function and different mean for the n-point maximin LHD and the model* (7.1).

n	$\hat{\psi}_1$	$\hat{\psi}_2$	$\hat{\sigma}^2$	RMSE	RBCI
	constant mean				
30	0.064	0.400	13.40	0.47	0.93
40	0.075	0.270	16.40	0.28	0.93
50	0.118	0.286	59.03	0.05	0.87
60	0.064	0.187	15.01	0.20	0.96
70	0.041	0.165	8.26	0.20	0.98
80	0.095	0.146	24.57	0.05	0.66
	linear mean				
30	0.056	0.378	10.35	0.41	0.94
40	0.150	0.368	92.80	0.14	0.94
50	0.123	0.287	54.62	0.06	0.84
60	0.070	0.188	17.45	0.16	0.96
70	0.233	0.155	493.59	0.07	0.36
80	0.095	0.141	20.30	0.05	0.65
	pure quadratic mean				
30	0.062	0.344	8.98	0.33	0.95
40	0.214	0.474	252.32	0.19	0.82
50	0.121	0.279	41.39	0.07	0.83
60	0.072	0.181	13.76	0.11	0.96
70	0.217	0.143	212.46	0.08	0.38
80	0.088	0.133	11.84	0.04	0.69
	full quadratic mean				
30	0.072	0.343	10.91	0.30	0.95
40	0.233	0.508	370.09	0.19	0.81
50	0.122	0.284	43.59	0.07	0.83
60	0.075	0.182	14.80	0.10	0.96
70	0.221	0.144	228.30	0.07	0.36
80	0.089	0.134	12.34	0.04	0.69

7.2 Test function 2

Consider the model

$$\eta(x) = 100(e^{-2/x_1^{1.75}} + e^{-2/x_2^{1.5}} + e^{-2/x_3^{1.25}}), \tag{7.2}$$

where the input variable $x = (x_1, x_2, x_3)$ varies in the square $[0, 1]^3$.

Table 7.2: *The estimators of parameters and characteristics of GP emulators with Gaussian correlation function and different mean for the n-point maximin LHD and the model* (7.2).

n	$\hat{\psi}_1$	$\hat{\psi}_2$	$\hat{\psi}_3$	$\hat{\sigma}^2$	RMSE	RBCI
	constant mean					
30	1.467	1.786	3.526	1521.65	0.36	0.71
40	0.931	1.202	1.080	346.92	0.29	0.85
50	0.778	0.731	1.373	336.11	0.31	0.78
70	0.561	0.598	0.810	125.66	0.13	0.84
90	0.262	0.498	0.478	69.55	0.17	0.98
150	0.310	0.429	0.409	79.88	0.05	0.80
	linear mean					
30	1.228	1.449	2.750	598.68	0.34	0.73
40	0.738	0.932	0.821	126.96	0.24	0.90
50	0.541	0.594	0.853	72.41	0.28	0.84
70	0.426	0.456	0.610	35.16	0.15	0.88
90	0.235	0.377	0.377	20.21	0.15	0.99
150	0.264	0.351	0.341	21.65	0.05	0.83
	pure quadratic mean					
30	0.033	0.149	20.000	2.04	0.61	0.90
40	0.323	0.407	0.440	9.35	0.34	0.96
50	0.347	0.405	0.545	13.20	0.30	0.88
70	0.308	0.319	0.389	7.52	0.08	0.99
90	0.179	0.260	0.273	3.42	0.05	1.00
150	0.189	0.259	0.261	3.19	0.03	0.89
	full quadratic mean					
30	1.385	1.395	3.733	709.74	0.39	0.75
40	0.378	0.503	0.510	15.57	0.37	0.93
50	0.381	0.435	0.643	18.59	0.38	0.78
70	0.325	0.338	0.414	9.49	0.08	0.99
90	0.183	0.266	0.280	3.82	0.06	1.00
150	0.192	0.262	0.264	3.44	0.03	0.89

7.3 Test function 3

Consider the model

$$\eta(x) = 4(x_1 - 2 + 8x_2 - 8x_2^2)^2 + (3 - 4x_2)^2 + 16\sqrt{x_3 + 1}(2x_3 - 1)^2. \qquad (7.3)$$

where the input variable $x = (x_1, x_2, x_3)$ varies in the square $[0, 1]^3$.

Table 7.3: *The estimators of parameters and characteristics of GP emulators with Gaussian correlation function and different mean for the n-point maximin LHD and the model* (7.3).

n	$\hat{\psi}_1$	$\hat{\psi}_2$	$\hat{\psi}_3$	$\hat{\sigma}^2$	RMSE	RBCI
	constant mean					
30	3.46	0.95	2.51	3368.0	0.94	0.94
40	5.69	1.39	6.23	26610.3	0.23	0.97
50	8.44	1.44	13.28	170390.4	0.18	0.25
	linear mean					
30	5.37	1.25	4.48	11142.4	0.94	0.93
40	10.13	2.64	18.79	306270.6	0.10	0.81
50	10.49	1.97	20.00	427386.4	0.08	0.92
	pure quadratic mean					
30	1.00	0.26	0.75	82.1	1.67	0.94
40	12.33	3.20	20.00	351660.5	0.08	0.99
50	13.62	2.71	20.00	559405.5	0.04	0.00
	full quadratic mean					
30	7.82	1.70	8.06	45941.2	0.70	1.00
40	12.91	3.31	20.00	427523.7	0.08	1.00
50	14.93	3.32	20.00	905580.7	0.04	0.91

7.4 Conclusions

Upon a broad numerical study, we obtain the following conclusions.

- The emulator validity is better (in average) for the GP emulator with full quadratic mean.
- The emulator prediction is better (in average) for the GP emulator with full quadratic mean.
- The correlation parameter (in average) slightly depends on the form of mean term.