Prediction for computer experiments with both quantitative and qualitative factors

Min Li, Min-Qian Liu, Xiao-Lei Wang, Yong-Dao Zhou *

School of Statistics and Data Science, LPMC & KLMDASR, Nankai University, Tianjin 300071, China

A R T I C L E   I N F O

Article history:
Received 8 December 2019
Received in revised form 9 June 2020
Accepted 14 June 2020
Available online 20 June 2020

Keywords:
Cross-correlation
Multivariate Gaussian process
Prediction accuracy
Sliced Latin hypercube design

A B S T R A C T

Computer experiments with both quantitative and qualitative factors are commonly encountered in practice. Several literatures found that if the cross-correlation between an auxiliary response and the target response (i.e., the response to be predicted) is small, the information of such an auxiliary response may reduce the prediction accuracy of the target response. In this work, we use the prediction accuracy improvement probability to prove the possibility of this case in theory and develop a selection procedure to choose the useful auxiliary responses.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

In recent decades, there has been a growing interest in the use of computer experiments in science, engineering, and business. The underlying computer codes are often used to model many physical processes that are difficult or even impossible to study by conventional physical experiments. Gaussian process models, also called Kriging models, are popularly used as metamodels for computer codes, see Fang et al. (2006) and Santner et al. (2003) for more details.

Most relevant researches have paid attention to computer codes where all the input variables are quantitative (Sacks et al., 1989; Currin et al., 1991; Santner et al., 2003). In many applications, however, computer codes may contain both qualitative and quantitative factors (Long and Bartel, 2006; Rawlinson et al., 2006; Qian et al., 2008). A simple approach to model this kind of computer experiments is to separately construct an individual Gaussian process (iGP) model for each level-combination of the qualitative variables. This method is sufficient when there is no correlation across different output responses. If there exist cross-correlations between responses of different level-combinations of the qualitative variables, then the iGP may lose the efficiency in prediction accuracy. Some modeling techniques were proposed for computer models with both qualitative and quantitative factors. Qian et al. (2008) proposed an unrestricted correlation structure for the qualitative factors via the semidefinite programming technique such that the correlation structure of the qualitative factors is positive definite. Han et al. (2009) proposed a hierarchical Bayesian model. Zhou et al. (2011) applied the hypersphere decomposition to model the cross-correlation matrix and proposed a general framework for building multivariate Gaussian process (MGP) models. Li and Zhou (2016) used the nonseparable covariance for the cross-correlation structure and proposed a pairwise modeling approach to alleviate the computational burden raised in parameter estimation. Deng et al. (2017) proposed an additive Gaussian process model which employs a more flexible covariance structure that is capable of accommodating the complex interaction effects between qualitative factors and quantitative factors.

* Corresponding author.
E-mail address: ydzhou@nankai.edu.cn (Y.-D. Zhou).

https://doi.org/10.1016/j.spl.2020.108858
0167-7152/© 2020 Elsevier B.V. All rights reserved.
When the goal is to predict the response at a specific level-combination of the qualitative factors, many authors utilized the information of responses at all the other level-combinations of the qualitative factors to improve the prediction performance. Such a method is called the integrated modeling approach. However, Li and Zhou (2016) noted that when cross-correlations are not strong enough, the prediction accuracy of the integrated modeling approach is close to iGP in average performance. Huang et al. (2016) also pointed out that, by using the integrated modeling approach, the prediction accuracy might decrease if the target response (i.e., the response to be predicted) and other auxiliary responses are not similar. However, they did not give a theoretical proof for this phenomenon.

In this paper, we give a theoretical inference for the relationship between the integrated modeling approach and the individual method. We use the integrated squared prediction error (ISPE) to measure the prediction accuracy. Then we obtain the distribution of the variable which represents the improvement of the ISPE of the integrated Gaussian process (iGP) predictors compared to that of the iGP predictors. And the probability of the variable being greater than zero reflects the prediction accuracy improvement probability. We prove that when the cross-correlations between the target response and other responses converge to zero, the prediction accuracy improvement probability converges to 1/2. It means that if the cross-correlations are not strong enough, the full data iGP model may not give a more accurate prediction. Moreover, a useful selection procedure is also proposed to choose the useful auxiliary responses to improve the prediction accuracy.

The remainder of this paper is organized as follows. We describe the MGP model and the cross-correlation function in Section 2. In Section 3, we provide some properties of the ISPE of iGP predictors and prove that iGP model may perform worse than iGP model in theory. Section 4 shows the details of the selection procedure and provides several examples.

2. MGP model and cross-correlation function

Consider a computer model with an input vector $\mathbf{w} = (\mathbf{x}^T, \mathbf{z}^T)^T$, where $\mathbf{x} = (x_1, \ldots, x_d)^T$ denotes the $d$ quantitative factors, $\mathbf{z} = (z_1, \ldots, z_q)^T$ consists of all the $q$ qualitative factors, and $z_j$ has $s_j$ levels, $j = 1, \ldots, q$. Let the domain of $\mathbf{x}$ be $\chi = [0, 1]^d$, and $c_1, \ldots, c_m$ denote the $m = \prod_{j=1}^q s_j$ level-combinations of the factors in $\mathbf{z}$. We call each level-combination a category. For each $\mathbf{x} \in \chi$, let $\mathbf{y}(\mathbf{x}) = (y_1(\mathbf{x}), \ldots, y_m(\mathbf{x}))^T$ be the output of the computer model at the $c_1, \ldots, c_m$ categories and have the following model,

$$\mathbf{y}(\mathbf{x}) = \begin{bmatrix} y_1(\mathbf{x}) \\ \vdots \\ y_m(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1(\mathbf{x}) \beta_1 \\ \vdots \\ \mathbf{f}_m(\mathbf{x}) \beta_m \end{bmatrix} + \begin{bmatrix} \epsilon_1(\mathbf{x}) \\ \vdots \\ \epsilon_m(\mathbf{x}) \end{bmatrix},$$

where $\mathbf{f}_k(\mathbf{x}) = (f_{k1}(\mathbf{x}), \ldots, f_{kp_k}(\mathbf{x}))$ is a $1 \times p_k$ vector of the known regression functions and $\mathbf{\beta}_k$ is an unknown $p_k \times 1$ vector of the regression parameters, $k = 1, \ldots, m$. Suppose $\mathbf{e}(\mathbf{x}) = (\epsilon_1(\mathbf{x}), \ldots, \epsilon_m(\mathbf{x}))^T$ is a stationary MGP with a zero mean and covariance function $\text{cov}(\mathbf{e}(\mathbf{x}), \mathbf{e}(\mathbf{x}')) = \sigma^2 \mathbf{K}_0(\mathbf{x}, \mathbf{x}')$, where $\mathbf{K}_0(\mathbf{x}, \mathbf{x}')$ is the correlation between $\mathbf{x}$ at the $c_i$th category and $\mathbf{x}'$ at the $c_j$th category, $\sigma^2$ is the variance of the $c_i$th category, and $\sigma^2$ is the variance of the $c_j$th category respectively. Each marginal process $\epsilon_i(\mathbf{x})$ is a stationary univariate Gaussian process with a zero mean and covariance function $\sigma^2_i \mathbf{K}_i(\mathbf{x}, \mathbf{x}')$. Qian et al. (2008) proposed a general framework for building MGP models and Zhou et al. (2011) further refined it. In their framework, the covariance functions are separable ones where each level-combination has a common marginal covariance function, i.e., $\mathbf{f}_k = \mathbf{f}$, $\mathbf{\beta}_k = \mathbf{\beta}$ and $\sigma^2_i = \sigma^2$ for all $k$. This assumption simplifies the covariance structure and significantly reduces the number of model parameters. The nonseparable covariance functions which allow different covariance functions at different output levels by kernel convolutions have been discussed in Ver Hoef and Barry (1998), Majumdar and Gelfand (2007), Melkumyan and Ramos (2011). A more flexible cross-correlation structure was proposed in Fricker et al. (2013) and adopted by Li and Zhou (2016). The nonseparable method has a high computational cost because the number of parameters to be estimated is large.

Suppose $K_0(\mathbf{x}, \mathbf{x}')$ has the form

$$K_0(\mathbf{x}, \mathbf{x}') = \tau_{ij} R_0(\mathbf{x}_i, \mathbf{x}_j'),$$

where $\tau_{ij} \in [-1, 1]$ is the cross-correlation factor between the $c_i$th and the $c_j$th categories with $\tau_{ii} = 1$. The matrix $\mathbf{T} = (\tau_{ij})_{m \times m}$ is a positive definite matrix with unit diagonal elements. We choose a Gaussian correlation function for the quantitative factors. In the separable covariance function, we adopt the correlation function proposed by Qian et al. (2008)

$$R_0(\mathbf{x}, \mathbf{x}') = \prod_{k=1}^d \exp\{-\theta_k |x_k - x_k'|^2\}. \tag{2}$$

In the nonseparable covariance function, we adopt the correlation function proposed by Fricker et al. (2013), $\theta_i = \text{diag}(\theta_{i1}, \ldots, \theta_{id})$.

$$R_0(\mathbf{x}, \mathbf{x}') = \frac{\exp(-((\mathbf{x} - \mathbf{x}')^T(\theta_1^{-1}/2 + \theta_d^{-1}/2)^{-1}(\mathbf{x} - \mathbf{x}'))}{|((\theta_1/2 + \theta_d/2)(\theta_1^{-1}/2 + \theta_d^{-1}/2)|^{1/4}}. \tag{3}$$
3. Main result

In this section, we propose the prediction accuracy improvement probability of the IGP model over the iGP model to theoretically explain the occurrence that the integrated modeling approach may perform worse than the individual method.

For simplicity, assume that there is no global trend in the output for each category which can be addressed by normalizing the response data for each response surface at each category respectively. For $m \geq 2$, $y(x) = (y_1(x), \ldots, y_m(x))^T$ follows an $m$-variate MGP model with a zero mean, unit variance and the cross-correlation function $K_θ(x, x')$ in (1), denote the parameters in (2) or (3) as $θ$. Sliced Latin hypercube design (SLHD) proposed by Qian (2012) has been used for experiments with quantitative and qualitative factors. An SLHD can be partitioned into $m$ slices, each slice is a Latin hypercube design containing $n$ runs, and can be used under one category. The design we employ to get observations is an SLHD $D = (D_1^T, \ldots, D_m^T)^T$ with $D_i$ being one slice. Denote the output for each category by $y_i = y_i(D_i), i = 1, \ldots, m$, then

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \sim N(0_{mn}, \Sigma), \quad \Sigma = \begin{bmatrix} R_{11}(D_1, D_1) & \cdots & \tau_{1m}R_{1m}(D_1, D_m) \\ \vdots & \ddots & \vdots \\ \tau_{1m}R_{m1}(D_m, D_1) & \cdots & R_{mm}(D_m, D_m) \end{bmatrix},$$

where $R_{ij}(D_i, D_j)$ denotes the $n \times n$ correlation matrix between $y_i(D_i)$ at the $c_i$th category and $y_j(D_j)$ at the $c_j$th category, $i, j = 1, \ldots, m$. We simplify the notation by replacing $R_{ij}(D_i, D_j)$ with $R_{ij}$ for $i, j = 1, \ldots, m$, as long as it will not cause any ambiguity.

Suppose that our objective is to predict the output of the $c_1$th category at any untried point $x \in \chi$. $τ_1 = (r_{12}, \ldots, r_{1m})^T$ consists of the cross-correlations between the $c_1$th and other categories. Two predictions by the iGP and IGP methods and the relationship between them will be given below.

If we use observations at the $c_1$th category only, the predicting equation obtained by the iGP model is

$$\tilde{y}_1(x) = r_{11}(x)R_{11}^{-1}y_1,$$  \hspace{1cm} (4)

where $r_{11}(x)$ is an $n \times 1$ correlation vector between $y_1$ and $y_1(x)$. We can also utilize all the observations to get an IGP predicting equation

$$\hat{y}_1(x) = r^T(x)R^{-1}y,$$  \hspace{1cm} (5)

where $r(x) = [r_{11}(x), r_{12}(x), \ldots, r_{1m}(x)]^T$, and $r_{ij}(x)$ is an $n \times 1$ correlation vector between $y_j$ and $y_1(x)$. Noted that Eqs. (4) and (5) are identical when $τ_1$ is zero. That is, the iGP and IGP modeling approaches are identical when the $c_1$th and other categories are uncorrelated.

To measure the prediction performance of the iGP predictor $\tilde{y}_1(x)$ in Eq. (4) and the IGP predictor $\hat{y}_1(x)$ in Eq. (5), we consider the following criterion,

$$\Delta = \int_{\chi} [y_1(x) - \tilde{y}_1(x)]^2dx - \int_{\chi} [y_1(x) - \hat{y}_1(x)]^2dx$$

$$= 2\int_{\chi} [y_1(x) - \tilde{y}_1(x)]\delta(x)dx + \int_{\chi} \delta^2(x)dx,$$

where $\delta(x) = \tilde{y}_1(x) - \hat{y}_1(x)$ is bounded on $\chi$ if $τ_1 \neq 0_{m-1}$. We call $\int_{\chi} [y_1(x) - \tilde{y}_1(x)]^2dx$ and $\int_{\chi} [y_1(x) - \hat{y}_1(x)]^2dx$ the ISPEs of the predictors $\tilde{y}_1(x)$ and $\hat{y}_1(x)$ respectively. Then the measure $\Delta$ represents the improvement of the ISPE of the IGP over the iGP method, and the distribution of $\Delta$ and the probability of $\Delta > 0$ are of interest.

Note that

$$(y_1(x) - \tilde{y}_1(x))|y \overset{\Delta}{\sim} z(x) \sim GP\left(0, C_{11}(\cdot, \cdot)\right),$$  \hspace{1cm} (6)

where $C_{11}(\cdot, \cdot) = \text{cov}(y_1(x) - \tilde{y}_1(x), y_1(x') - \tilde{y}_1(x'))|y = R_{11}(x, x') - r^T(x)R^{-1}r(x')$ is continuous and bounded on $\chi \times \chi$.

Then we can choose a measurable version of the process $z(x)$ such that

$$\Delta \sim N(\mu_\Delta, \sigma_\Delta^2),$$

where $\mu_\Delta = \int_{\chi} \delta^2(x)dx$, and $\sigma_\Delta^2 = 4\int_{\chi} \int_{\chi} \delta(x)C_{11}(x, x')\delta(x')dxdx'$. See Marcus and Shepp (1972) for more details. The probability that the IGP predictor $\hat{y}_1(x)$ performs better than the iGP predictor $\tilde{y}_1(x)$ on ISPE is

$$p = P(\Delta > 0) = \Phi\left(\frac{\mu_\Delta}{\sigma_\Delta}\right),$$

where $\Phi$ is the standard normal cumulative distribution function. We call $p$ the prediction accuracy improvement probability of the IGP. Then, we have the following result.

**Theorem 1.** For a fixed design $D = (D_1^T, \ldots, D_m^T)^T$ and the parameter $θ$ in the MGP model, when the cross-correlations between the $c_1$th and other categories $τ_1$ tend to $0_{m-1}$, the prediction accuracy improvement probability $p$ converges to $1/2$ in probability.
Although the proof of Theorem 1 does not need the design to be an SLHD, we still suggest employing SLHDs to get observations in practice because of the benefit of using SLHDs in computer experiments with both quantitative and qualitative inputs.

Theorem 1 shows that when the cross-correlations between the target response and the auxiliary responses converge to zero, the prediction accuracy improvement probability converges to 1/2. That is, the probabilities of the ISPE of the iGP model being smaller and larger than that of the iGP model are both approximately equal to 1/2. It means that if the cross-correlations are not strong enough, the iGP model may not give a more accurate prediction. Since the iGP predictor is easier to calculate, we often use the iGP model to predict a target response. If using the improper IGP model to predict, the prediction performance may be worse than that of the iGP model. Therefore, Theorem 1 theoretically justifies the phenomenon noticed by Li and Zhou (2016) and Huang et al. (2016).

After some computer simulations, we believe that the prediction accuracy improvement probability will be greater than 1/2 under some stronger conditions such as the cross-correlations between the response to be predicted and other auxiliary responses converge to one. We will concentrate on this in the future.

4. Selection procedure

The prediction accuracy of the IGP model depends on the cross-correlations between the target response and the auxiliary responses, thus one needs to select the helpful auxiliary responses for improving the prediction accuracy. The idea is simple, we have already proved that the ISPE has a strong connection with the cross-correlation parameters. When the cross-correlations between the target response and the auxiliary responses are not strong enough, the IGP prediction accuracy may not be better than that of the iGP model, then we can use the iGP to predict the target response. When there are some auxiliary responses that have strong cross-correlations with the target response, we prefer to use their information to improve the prediction accuracy of the target response. Suppose the target response is under the $c_1$th category, then the selection procedure is shown in Algorithm 1. Step 3 suggests choosing these auxiliary responses that have strong cross-correlations with the target response for improving the prediction accuracy of the target response. Ignoring the auxiliary responses of the qualitative variable that are weakly correlated helps reduce the number of model parameters that need to be estimated and reduces the computational burden.

**Algorithm 1.** Selection procedure of the auxiliary responses.

- **Step 1.** Set a lower bound or threshold for the cross-correlation parameter, $\tau_0$;
- **Step 2.** Build the MGP model using the data of all the $m$ categories, obtain the cross-correlation matrix $T = (\tau_{ij})_{m \times m}$ and $\tau_1 = (\tau_{12}, \ldots, \tau_{1m})^T$;
- **Step 3.** Find all $j$, such that $|\tau_{ij}| > \tau_0$, define $S$ as the set of all such $j$;
- **Step 4.** Use the data of the $c_1$th and the categories in $S$ to build the new MGP model, then get the prediction at any untried point under the $c_1$th category.

We provide two examples to give the suggestions for the value of $\tau_0$, and show the efficiency of the selection procedure. The method for building the MGP models and estimating the parameters in this section is the same as that in Zhou et al. (2011).

**Example 1.** Generate the data from an MGP model with one quantitative factor defined on $[0, 1]$ and one qualitative factor of four levels. We use the separable covariance function and let the mean be zero, $\sigma = 2$, $\theta = 1/3$, $\tau_1 = (0.60, 0.60, -0.60)$. The cross-correlation matrix is

$$
T = \begin{bmatrix}
1 & 0.60 & 0.60 & -0.60 \\
0.60 & 1 & 0.35 & -0.20 \\
0.60 & 0.35 & 1 & 0.14 \\
-0.60 & -0.20 & 0.14 & 1
\end{bmatrix}.
$$

Use the data to build the iGP and IGP models respectively, then predict the responses of the first category. The training points $D$ are generated by a one-dimensional SLHD in Qian (2012), with 4 slices and $n = 20$ runs in each slice. Test points $u_u$ are the $N = 100$ evenly spaced points at the first category: $[0, 1/99, 2/99, \ldots, 1]$, $u = 1, \ldots, N$. Their corresponding responses $y_0$ and the observed responses $y = (y_1^1, y_1^2, y_1^3, y_1^4)^T$ at the points in $D$, are generated from the MGP model with the parameters $\sigma$, $\theta$ and $T$ given above.

The value of $\tau_1$ may affect the prediction accuracy significantly. In Example 1.2 of Li and Zhou (2016), $\tau_1 = (-0.70, -0.60, 0.80)$, the prediction accuracy of IGP is close to iGP, which is expected as an integrated modeling approach will not gain much advantage if the cross-correlation is not strong enough. Then, in this example, let $|\tau_1| = 0.61_{m-1}$ in
the selection procedure. We repeat the selection procedure 200 times, and use \( RMSE = \sqrt{N^{-1} \sum_{w=1}^{N} (y(w) - \hat{y}(w))^2} \) to assess the prediction accuracy. The boxplots of the RMSEs for the iGP and IGP models are displayed in Fig. 1. From Fig. 1, it is found that the prediction accuracy of IGP is close to iGP, which is reasonable since the cross-correlations may be still not strong enough, as shown by Example 1.2 of Li and Zhou (2016). We have also carried out the simulation when \(|\tau_1| = 0.71_{m-1}\), the prediction performances of IGP and iGP are close to each other.

Example 2. Consider an experiment with five quantitative factors \((x_1, \ldots, x_5) \in [0, 1]^5\), and one qualitative factor \(z\) with three levels. The true models are as follows:

\[
\begin{align*}
  y(x_1, \ldots, x_5, 1) &= a_1 \exp(x_1 x_2 x_3 x_4 x_5) + a_2(x_1 - a_3)^2, \\
  y(x_1, \ldots, x_5, 2) &= b_1 \exp(x_1 x_2 x_3 x_4 x_5), \quad \text{and} \\
  y(x_1, \ldots, x_5, 3) &= c_1 \cos(c_2 \pi x_1 x_2 x_3 x_4 x_5).
\end{align*}
\]

For each model, the coefficients are drawn from independent Gaussian distributions with a standard deviation \(10^{-2}\). The mean values of \((a_1, a_2, a_3), (b_1), (c_1, c_2)\) are set to be \((4, -1, 1), 5, (6, 10)\), respectively. Clearly, Surfaces I and II are similar and they are positively correlated, but Surface III is very different from them. This example is from the supplementary material of Huang et al. (2016). Test points are taken on a grid of \(4^5 = 1024\) equispaced points on \([0, 1]^5\) for the first response. The training points \(D\) are generated by a five dimensional SLHD in Qian (2012), with 3 slices and \(n = 25\) runs in each slice. According to the discussions of Example 1, we set \(\tau_0 = 0.6\) and repeat 200 times to compare the results by the iGP, IGP, the selection methods of Huang et al. (2016) and our selection procedure. The last two methods are denoted by HLLY and LLWZ, respectively. Fig. 2 displays the boxplots of the RMSEs of the four methods. From Fig. 2, we find that the IGP model is worse than the iGP model, since the cross-correlations between the first response and the others are not all strong enough. The prediction performance of our selection procedure is better than that of the IGP and iGP models, and the method of Huang et al. (2016), which is comparable with the iGP model.
Appendix

Derivation of Eq. (6). It is easy to obtain that $y_1(x) - \tilde{y}_1(x) | y$ follows the Gaussian distribution, $E((y_1(x) - \tilde{y}_1(x))|y) = 0$.

$$
C_{11}(x, x') = \text{cov}(y_1(x) - \tilde{y}_1(x), y_1(x') - \tilde{y}_1(x')) = E((y_1(x) - \tilde{y}_1(x))(y_1(x') - \tilde{y}_1(x'))|y) - E(y_1(x) - \tilde{y}_1(x))E(y_1(x') - \tilde{y}_1(x'))|y) = \text{cov}(y_1(x), y_1(x')) - \text{cov}(y_1(x), r'(x')\mathbf{R}^{-1}y) + \text{cov}(r'(x)\mathbf{R}^{-1}y, r'(x')\mathbf{R}^{-1}y) = R_{11}(x, x') - r'(x)\mathbf{R}^{-1}r(x) - r'(x)\mathbf{R}^{-1}r(x') + r'(x)\mathbf{R}^{-1}r(x') = R_{11}(x, x') - r'(x)\mathbf{R}^{-1}r(x')
$$

Proof of Theorem 1. To get an explicit expression of $\mu_\Delta/\sigma_\Delta$, we partition the covariance matrix $\mathbf{R}$ into the following blocks,

$$
\mathbf{R} = 
\begin{bmatrix}
\mathbf{R}_{11} & \mathbf{R}_{12} \\
\mathbf{R}_{21} & \mathbf{R}_{22} \\
\mathbf{R}_{m1} & \mathbf{R}_{m2} & \cdots & \mathbf{R}_{mm}
\end{bmatrix}
= 
\begin{bmatrix}
\mathbf{R}_{11} & \mathbf{V}_{12} \\
\mathbf{V}_{21} & \mathbf{V}_{22}
\end{bmatrix}^{-1},
$$

and then get a corresponding expression of $\delta(x)$.

Denote $\mathbf{V}_{(22)} = [\mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{R}_{11}^{-1}\mathbf{V}_{12}]^{-1}$, then

$$
\mathbf{R}^{-1} = 
\begin{bmatrix}
\mathbf{R}_{11}^{-1} & + & \mathbf{R}_{11}^{-1}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21} & + & -\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{22} \\
-\mathbf{V}_{22}^{-1}\mathbf{V}_{21} & + & \mathbf{V}_{22}^{-1} & + & -\mathbf{V}_{22}^{-1}\mathbf{V}_{22} \\
\end{bmatrix}.
$$

Denote $y_{(-1)} = (y_2^T, \ldots, y_m^T)^T$, $g(x) = \mathbf{r}_1^T(x) - \mathbf{r}_1^T(\mathbf{x})\mathbf{R}_{11}^{-1}\mathbf{R}_{12}, \ldots, \mathbf{r}_m^T(x) - \mathbf{r}_1^T(\mathbf{x})\mathbf{R}_{11}^{-1}\mathbf{R}_{1m}$, $\chi_{(-1)} = [y_{(-1)} - \mathbf{V}_{21}\mathbf{R}_{11}^{-1}y_1]$, $\mathbf{r}(x) = [\mathbf{r}_1^T(x), \mathbf{r}_2^T(x), \ldots, \mathbf{r}_m^T(x)]^T$, $\Gamma = \text{diag}(\tau_{12}, \ldots, \tau_{1m}) \otimes \mathbf{I}_n$, where $\mathbf{I}_n$ is the identity matrix of order $n$. Then after a simple mathematical derivation, $\delta(x) = \tilde{y}_1(x) - \tilde{y}_1(x) = \mathbf{r}(x)\mathbf{R}^{-1}\mathbf{y} - \mathbf{r}_1^T(\mathbf{x})\mathbf{R}_{11}^{-1}\mathbf{y}_1 = g^T(x) \Gamma \mathbf{V}_{(22)} \chi_{(-1)}$. Denote $\mathbf{G} = \int_x g(x)g^T(x)dx$, $\mathbf{H} = 4\int_x \int_x g(x)C_{11}(x, x')g^T(x')dx'dx$. Then,

$$
\mu_\Delta = \mathbf{r}_{(-1)}^T \Gamma \mathbf{V}_{(22)} \chi_{(-1)} = \mathbf{r}_{(-1)}^T \Gamma \mathbf{G} \Gamma \mathbf{V}_{(22)} \chi_{(-1)}^{-1},
$$

$$
\sigma_{\Delta}^2 = 4 \chi_{(-1)}^T \mathbf{V}_{(22)} \Gamma \int_x \int_x g(x)C_{11}(x, x')g^T(x')dx'dx \chi_{(-1)} = \chi_{(-1)}^T \mathbf{V}_{(22)} \Gamma \mathbf{H} \Gamma \mathbf{V}_{(22)} \chi_{(-1)}^{-1}.
$$
and
\[ \eta = \frac{\mu_\Delta}{\sigma_\Delta} = \frac{\xi^{T}_{(1|1)} V_{(22)} \Gamma G \Gamma V_{(22)} \xi_{(1|1)}}{\sqrt{\xi^{T}_{(1|1)} V_{(22)} \Gamma H \Gamma V_{(22)} \xi_{(1|1)}}} \]

is a ratio of the quadratic forms in the normal random variables \( \xi_{(1|1)} \). From Bao and Kan (2013), \( \eta \) has finite first and second order moments when \( \tau_1 \neq 0_{m-1} \) and with probability 1

\[ 0 < \frac{\xi^{T}_{(1|1)} V_{(22)} \Gamma G \Gamma V_{(22)} \xi_{(1|1)}}{\sqrt{\xi^{T}_{(1|1)} V_{(22)} \Gamma H \Gamma V_{(22)} \xi_{(1|1)}}} \leq \lambda_{\text{max}}(H^{-\frac{1}{2}} G H^{-\frac{1}{2}}), \]

where \( \lambda_{\text{max}}(M) \) is the largest eigenvalue of matrix \( M \).

As \( \tau_1 \to 0_{m-1} \), \( \Gamma \to 0_{(m-1)\times(m-1)} \),

\[ V_{(22)} \to V_0^{(22)} = \begin{bmatrix} R_{22} & \cdots & \tau_{2m} R_{2m} \\ \vdots & \ddots & \vdots \\ \tau_{2m} R_{m2} & \cdots & R_{mm} \end{bmatrix}^{-1}, \]

and \( H \to H_0 = 4 \int_0^1 \int_0^1 g(x)g(y) \left( x \cdot x' \right) dx dy \), where \( H_{11} = R_{11} - \tau_{1j} R_{1j} R_{j1} \). The hypersphere decomposition method in Zhou et al. (2011) guarantees that the \( T = (\tau_{ij}) \) is a positive definite matrix with unit diagonal elements, \( \tau_{ij} \) is the limit of \( \tau_{ij} \) as \( \tau_{ij} \to 0 \), \( 2 \leq i < j \leq m \).

It is easy to obtain that \( E(\xi_{(1|1)}) = 0 \), \( \text{cov}(\xi_{(1|1)}) = V_{(22)}^{-1} \). Then according to the expectation of the quadratic form of a random vector,

\[ E \left[ \xi^{T}_{(1|1)} V_{(22)} \Gamma G \Gamma V_{(22)} \xi_{(1|1)} \right] = E(\xi_{(1|1)} V_{(22)} \Gamma G \Gamma V_{(22)} E(\xi_{(1|1)} + \text{trace}(V_{(22)} \Gamma G \Gamma V_{(22)} \text{cov}(\xi_{(1|1)}))) \right] = \text{trace}(V_{(22)} \Gamma G \Gamma). \]

Then \( \lambda_{\text{max}}(H^{-\frac{1}{2}} G H^{-\frac{1}{2}}) \to \lambda_{\text{max}}(H_0^{-\frac{1}{2}} G H_0^{-\frac{1}{2}}) \) and

\[ E\eta^2 = E \left[ \frac{\xi^{T}_{(1|1)} V_{(22)} \Gamma G \Gamma V_{(22)} \xi_{(1|1)}}{\sqrt{\xi^{T}_{(1|1)} V_{(22)} \Gamma H \Gamma V_{(22)} \xi_{(1|1)}}} \right] \leq \lambda_{\text{max}}(H^{-\frac{1}{2}} G H^{-\frac{1}{2}}) E \left[ \xi^{T}_{(1|1)} V_{(22)} \Gamma G \Gamma V_{(22)} \xi_{(1|1)} \right] \]

\[ = \lambda_{\text{max}}(H^{-\frac{1}{2}} G H^{-\frac{1}{2}}) \text{trace}(V_{(22)} \Gamma G \Gamma) \]

\[ \to \lambda_{\text{max}}(H_0^{-\frac{1}{2}} G H_0^{-\frac{1}{2}}) \cdot 0 = 0, \]

\[ E\eta^2 \to 0, \]

i.e. \( \frac{\mu_\Delta}{\sigma_\Delta} \xrightarrow{L^2} 0 \) (\( \tau_1 \to 0_{m-1} \)), where \( \xrightarrow{L^2} \) means converging in second moment. Then from the continuous mapping theorem we have \( p = \Phi \left( \frac{\mu_\Delta}{\sigma_\Delta} \right) \xrightarrow{L^2} \Phi(0) = \frac{1}{2} \). This implies \( p \xrightarrow{P} \frac{1}{2} \). \( \square \)

References


