CONTRIBUTIONS TO CROSSOVER DESIGNS AND QUANTILE ANALYSIS FOR COMPUTER EXPERIMENTS

by

LIWEI WANG

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Statistics

Written under the direction of

Ying Hung

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

OCTOBER, 2019

ABSTRACT OF THE DISSERTATION

Contributions to crossover designs and quantile analysis for computer experiments

By LIWEI WANG

Dissertation Director:

Ying Hung

This dissertation develops methodologies for optimal crossover designs and quantile modeling in computer experiments. It consists of two parts. The fist part of the dissertation is regarding finding optimal crossover designs for qualitative treatment effect. The second part of the dissertation is to investigate quantile analysis in computer experiments.

Crossover designs for quantitative variables are common in practice but the theoretical developments are overlooked in the design literature. Motivated by an experimental design problem in cell biology. new classes of optimal crossover designs are introduced under different model assumptions on carryover effects with quantitative variables. Theoretical properties of optimal designs are derived which are different from their counterparts with qualitative variables. To efficiently construct optimal designs, systematic procedures are proposed based on a collection of swap operations. The

proposed optimal designs are demonstrated by simulation studies and an application in cell adhesion experiments.

The characteristic of computer experiments are that they are deterministic, time-consuming and involve large number of variables. Due to these features of computer experiments, Gaussian process model is a widely used interpolator as an emulator in computer experiments to model mean structure of response. It is also of interest to model different quantiles. Motivated by finding an analogy to Gaussian process models for quantile analysis in computer experiments, a new Bayesian quantile analysis model for computer experiments is introduced. New model developed could capture the non-linearity and smoothness of underlying quantile functions. Quantile predictions in the new model for observed inputs agree on true quantiles asymptotically. With some constraints, asymptotic consistency of coefficients estimation is derived. There is no issue of quantile curves crossing each other in proposed model for quantile prediction of observed inputs.

Acknowledgements

First of all, I would like to express my deepest gratitude to my advisor Professor Ying Hung. Her constant encouragement and endless support helped me get through many difficulties during my research. She not only taught me many valuable research experience but also a lot of life experience, which will benefit me tremendously in the long run. Without her help, I wouldn't have completed my PhD study. I cannot be more fortunate to have Professor Ying Hung as my advisors and to be influenced by her attitude and experience toward life and work as well as teaching kids.

I am also grateful for the support and help I received from statistics department at Rutgers, especially when I had my baby, Professor John Kolassa and Professor Regina Liu offered a lot of support both financially and mentally. I also learned a lot from Professor Minge Xie and Professor Zhiqiang Tan's class. I also want to thank Arlene Gray, Lisa Curtin and Marcy Collins for attending all my need. The time I spent at statistics department is so memorable that I will cherish forever.

I also want to thank Professor Jerry Cheng, Professor Steve Buyske and Professor Tirthankar DasGupta for taking their valuable time being my committee members and for making them so available in scheduling my defense.

Last but not least, my gratitude goes to my families and friends. I thank my husband Changlei Li for his strong support and valuable insight on everything. I also want to thank my son Alexander Li for allowing me to work during our regular playtime and

being a sweet boy that I will love forever. My parents Anqiang Wang and Hongmin Yu are always there to help no matter what I need. Their unconditional love has passed on to me and is the best gift I received ever. I also want to thank Long Feng, Yi Fan, Guang Yang, Yibo Zhao and Wenxiu Dong for being my friends. I also want to thank Wenting Wang and Cuijuan Wang for being my sweet sisters.

Dedication

To everyone I love.

Table of Contents

List of Tables

List of Figures

Chapter 1

Optimal Crossover Designs for Quantitative Variables

1.1 Introduction

Crossover designs, also called repeated measurement designs, refer to designs where each subject in the experiment receives each treatment in succession (Stufken, 1996; Bose and Dey, 2009; Padgett, 2014; Jones and Kenward, 2015; Lui, 2016). Crossover designs have been widely applied to different scientific studies and, based on different models, optimal crossover designs are intensively studied in the literature (Hedayat and Afsarinejad, 1975, 1978; Cheng and Wu, 1980; Sen and Mukerjee, 1987; Stufken, 1991; Afsarinejad and Hedayat, 2002; Kunert and Stufken, 2002).

However, to the best of our knowledge, the existing works are limited to the modeling assumption where the treatment effects are qualitative. Quantitative treatment effects are of interest in many applications and direct applications of the existing designs by treating different levels of a quantitative factor as distinct categories can significantly reduce the design efficiency. Therefore, the focus of this paper is to construct new classes of optimal designs for quantitative variables.

This research is motivated by the experimental design problem in a study of cell adhesion, which plays a key role in tumor metastasis in cancer study (Huang et al., 2010; Zarnitsyna et al., 2007). In the experiment, two cells are put together for a predetermined duration, then retracted away. After a predetermined waiting time,

the two cells are put together again for the next contact. Such procedure is performed repeatedly and adhesion properties, such as bond lifetime, are measured as the outcome of the repeated tests. Such repeated experiments are conducted for multiple pairs of cells. There are two quantitative control variables in the experiment, the contact time and waiting time, which control the duration that molecules are put into contact and the time between retraction and the next contact. Due to a significant amount of cell-cell variability, the idea of crossover designs is desirable in practice because the effects from different contact and waiting time are no longer confounded with cell effects. As a result, the systematic bias in estimating the treatment effects could be reduced substantially.

Constructing an optimal crossover design is challenging due to the carryover effects. Based on different model assumptions on the carryover effects with quantitative variables, new classes of optimal crossover designs are introduced. Optimal design properties are derived which are different from their counterparts for qualitative variables. Instead of requiring balanced properties which is common for optimal designs with qualitatively variables, it is shown that the design optimality for quantitative variables is achieved by the orthogonality between the current design and the design in the previous period for each subject. Systematic swapping procedures are proposed to construct optimal designs.

1.2 Crossover designs with one quantitative variable

1.2.1 Two models for crossover designs

The crossover designs are constructed based on two models. The first model studies a direct treatment effect and a constant carryover effect from the previous treatment, and

the second model further distinguish the carryover effect into self and mixed carryover effects. For both models, assume that there is one quantitative variable and it has t levels. Without loss of generality, we can assume our experiment region is $[0, 1]$. The response for the ith subject at the jth period, is denoted by $y_{i,j}$, where $1 \leq i \leq n$, $1 \leq j \leq p$.

Model 1: Assume that α_i is the effect from subject i, β_j is the effect in period j, τ is the direct treatment effect, and γ is the carryover effect. We have

$$y_{i,j} = \alpha_i + \beta_j + \tau d_{i,j} + \gamma d_{i,j-1} + e_{i,j}, \tag{1.2.1}$$

where $d_{i,j}$ is the design of the quantitative variable for subject i at period j, $d_{i,0}$ is the design for the initial period, and $e_{i,j}$ are i.i.d random variables with mean 0 and variance σ^2. Without loss of generality, the design is assumed to be centered around zero, i.e., $\sum_{j=1}^{p} d_{i,j} = 0$ for all i. This model can be rewritten in a matrix form as follows

$$\boldsymbol{y} = \boldsymbol{U}\boldsymbol{\alpha} + \boldsymbol{P}\boldsymbol{\beta} + \boldsymbol{t_d}\tau + \boldsymbol{p_d}\gamma + \boldsymbol{e}, \tag{1.2.2}$$

where $\boldsymbol{U} = \boldsymbol{I_n} \otimes \boldsymbol{1_p}$ and $\boldsymbol{P} = \boldsymbol{1_n} \otimes \boldsymbol{I_p}$ are the design matrix for the subject and period effect, $\boldsymbol{t_d^T} = (d_{1,1}, \cdots, d_{1,p}, \cdots, d_{n,1}, \cdots, d_{n,p})$ is the design vector for the direct treatment effect and $\boldsymbol{p_d^T} = (d_{1,0}, \cdots, d_{1,p-1}, \cdots, d_{n,0}, \cdots, d_{n,p-1})$ is the design vector for the carryover effect.

Model 1 is different from the model considered by Hedayat and Afsarinejad (1978) where the treatment is qualitative. For a qualitative variable with t categories, t direct treatment effects $\tau_{d_{i,j}}$ and t carryover effects $\gamma_{d_{i,j-1}}$ are assumed, and the model can be written as

$$y_{i,j} = \alpha_i + \beta_j + \tau_{d_{i,j}} + \gamma_{d_{i,j-1}} + e_{i,j}.$$

In contrast, Model 1 has only one treatment effect and one carryover effect for the quantitative variable regardless of the level of the variable.

In many applications, the carryover effect is different according to the treatment in the previous period. If the treatment in the jth period is the same as the previous one, i.e., $d_{i,j-1} = d_{i,j}$, then the carryover effect is called self-carryover (Kunert and Stufken, 2002). On the other hand, if $d_{i,j-1} \neq d_{i,j}$, then the carryover effect is called mixed-carryover. To distinguish these two types of carryover effects, we consider Model 2 as follows.

Model 2: Assume that ϕ is the self-carryover effect and γ is the mixed-carryover effect. We have

$$y_{i,j} = \begin{cases} \alpha_i + \beta_j + \tau d_{i,j} + \phi d_{i,j-1} + e_{i,j} & d_{i,j-1} = d_{i,j}, \\ \alpha_i + \beta_j + \tau d_{i,j} + \gamma d_{i,j-1} + e_{i,j} & d_{i,j-1} \neq d_{i,j}, \end{cases} \tag{1.2.3}$$

where $d_{i,j}$, $d_{i,0}$, α_i, τ, and $e_{i,j}$ are the same as in Model 1.

The matrix form of above model can be written as

$$\boldsymbol{y} = \boldsymbol{U\alpha} + \boldsymbol{P\beta} + \boldsymbol{t_d}\tau + \boldsymbol{m_d}\gamma + \boldsymbol{s_d}\phi + \boldsymbol{e} \tag{1.2.4}$$

where $\boldsymbol{m_d^T} = (d_{1,0}\mathbb{1}_{d_{1,0}\neq d_{1,1}}, \cdots, d_{n,p-1}\mathbb{1}_{d_{n,p-1}\neq d_{n,p}})$ is design vector for mixed carryover effect and $\boldsymbol{s_d^T} = (d_{1,0}\mathbb{1}_{d_{1,0}= d_{1,1}}, \cdots, d_{n,p-1}\mathbb{1}_{d_{n,p-1}= d_{n,p}})$ is the design vector for self-carryover effects, where $\mathbb{1}$ is the indicator function. Similar to Model 1, this model is also different from its counterpart for qualitative variables discussed in Kunert and Stufken (2002).

1.2.2 Optimal crossover designs

Based on the two models, optimal crossover design criteria are introduced and properties of the optimal designs are discussed. We focus on the estimation of the treatment

effect. Since there is only one control variable, the information matrix for estimating the treatment effect is a scalar. As a result, a design $d^* \in \Omega(t, n, p)$ for which maximizes the information matrix is A-optimal, D-optimal, E-optimal and T-optimal (Chernoff, 1953; Kiefer, 1959; Ehrenfeld, 1955), where $\Omega(t, n, p)$ is the design space for experiments with t treatment levels, n subjects, and p periods. Designs that are A, D, E, and T-optimal are known as universally optimal designs.

The information matrix for the direct treatment effect τ under Model 1 can be written as (Kunert and Stufken, 2002)

$$C_d = \boldsymbol{t}_{\boldsymbol{d}}^T \omega^{\perp}([\boldsymbol{P}, \boldsymbol{U}, \boldsymbol{p}_{\boldsymbol{d}}])\boldsymbol{t}_{\boldsymbol{d}},$$

where $\omega^{\perp}(A) = I - A(A^TA)^-A^T$ is the projection on the space of all vectors that are orthogonal to A^T and $(A^TA)^-$ is the generalized inverse of A^TA. An upper bound of C_d is derived based on Proposition 2.3 in Kunert (1983) and the sufficient condition to attain this bound is given in the following theorem .

Theorem 1.2.1. *(i) We have* $C_d \leq \boldsymbol{t}_{\boldsymbol{d}}^T \omega^{\perp}([\boldsymbol{U}, \boldsymbol{p}_{\boldsymbol{d}}])\boldsymbol{t}_{\boldsymbol{d}}$ *and the equality holds if and only if*

$$\boldsymbol{t}_{\boldsymbol{d}}^T \omega^{\perp}([\boldsymbol{U}, \boldsymbol{p}_{\boldsymbol{d}}])\boldsymbol{P} = 0. \tag{1.2.5}$$

(ii) The sufficient condition that satisfies equation (1.2.5) is:

$$\sum_{i=1}^{n} d_{i,0} = \sum_{i=1}^{n} d_{i,1} = \sum_{i=1}^{n} d_{i,2} = \cdots = \sum_{i=1}^{n} d_{i,p}. \tag{1.2.6}$$

Based on Theorem 1.2.1, the upper bound of information matrix can be reached by conditions given in (1.2.6). Therefore, given designs satisfying (1.2.6), universally optimal designs can be obtained by maximizing the upper bound of the information matrix. Based on this idea, properties of the universally optimal designs are discussed

in the next Theorem. We focus on the universally optimal designs within the class of *uniform designs* which can be defined as follows.

Definition 1.2.1. *A design $d \in \Omega(t, n, p)$ is defined as uniform if every treatment appears in every period exactly n/t times; and, for each subjects, every treatment appears exactly p/t times.*

Definition 1.2.1 is a direct extension of Hedayat and Afsarinejad (1978) to the cases where $p = \lambda t$. Based on the definition, for a design d to be uniform, $d \in \Omega(t, \mu t, \lambda t)$. λ, μ are positive integers. In Theorem 1.2.2, properties of the universally optimal designs under Model 1 are specified within the class of uniform designs.

Theorem 1.2.2. *An uniform design $d^* \in \Omega(t, \mu t, \lambda t)$ that satisfies $d_{i,0} = d_{i,p}$, $\forall$ i, is universally optimal under Model 1 if*

$$d^* = \arg\min_{d} \mid \sum_{i=1}^{n} \sum_{j=1}^{p} d_{i,j} d_{i,j-1} \mid . \tag{1.2.7}$$

The objective function, $\mid \sum_{i=1}^{n} \sum_{j=1}^{p} d_{i,j} d_{i,j-1} \mid$, in (1.2.7) reaches the minimum value zero if, for each subject, the design is orthogonal to the design in the previous period. This result is different from its counterpart for qualitative variable in Cheng and Wu (1980) where the optimal designs in $\Omega(t, \mu t, \lambda t)$ need to be uniform and strongly balanced. In particular, the strong balance property implies that every treatment is immediately preceded by every treatment (including itself) equally often. This balance property places a key role in optimal designs for qualitative variables, but not for quantitative variables in general. Based on the construction method proposed in the next section, the optimal designs for quantitative variables are not necessarily balanced. Instead of the balanced property, the design optimality for qualitative variables

is achieved by controlling the orthogonality between the current and the previous treatment for each subject.

For Model 2, the information matrix for the direct treatment effect can be written as (Kunert and Stufken, 2002)

$$C'_d = \boldsymbol{t_d}^T \omega^{\perp}([\boldsymbol{P}, \boldsymbol{U}, \boldsymbol{m_d}, \boldsymbol{s_d}])\boldsymbol{t_d}.$$

An upper bound for C'_d and the sufficient condition is given below.

Corollary 1.2.1. *(i)* $C'_d \leq \boldsymbol{t_d}^T \omega^{\perp}([\boldsymbol{U}, \boldsymbol{m_d}, \boldsymbol{s_d}])\boldsymbol{t_d}$ *and the equality holds if and only if*

$$\boldsymbol{t_d}^T \omega^{\perp}([\boldsymbol{U}, \boldsymbol{m_d}, \boldsymbol{s_d}])\boldsymbol{P} = 0 \tag{1.2.8}$$

(ii) The sufficient conditions for (1.2.8) are the condition in (1.2.6) and

$$\boldsymbol{s_d} = 0. \tag{1.2.9}$$

Note that, this additional condition, (1.2.9), implies no identical treatments in any consecutive settings of the design.

Corollary 1.2.2. *For any universally optimal design under Model 1, if the additional constraint (1.2.9) is satisfied, then it is universally optimal under Model 2.*

Similar to Model 1, the universally optimal designs under Model 2 is also different from its corresponding counterpart developed for qualitative variables in Kunert and Stufken (2002), where the designs need to be totally balanced to be optimal. The totally balanced property implies that every treatment is immediately preceded by every other treatment equal often. This implies the requirement of $\boldsymbol{s_d} = 0$.

1.2.3 Construction of optimal designs

In this section, we introduce construction rules for optimal designs discussed in Section 1.2.2. In particular, we focus on the class of uniform designs in $\Omega(t, c\lambda t, \lambda t)$, c is also positive integer. Let $\boldsymbol{S}_m$ denotes all the possible permutation of treatments with p period that contains each level of the quantitative variable exactly m times, where $m = p/t$. Let $d_i = [d_{i,1}, \cdots, d_{i,p}]$ be the design for the ith subject and $d_{i,0} = d_{i,p}$. For uniform designs, we need $d_i \in \boldsymbol{S}_m$ for all i.

Based on the results in Theorem 1.2.2, universally optimal designs d^* under Model 1 can be constructed by the following two steps when $c = 1$.

Step 1 Without loss of generality, find the optimal treatment assignment for the first subject, $d_1^* = [d_{1,1}^*, \cdots, d_{1,p}^*] \in \boldsymbol{S_m}$, such that

$$\sum_{j=1}^{p} d_{1,j}^* d_{1,j-1}^* = 0 \tag{1.2.10}$$

Step 2 The design for the rest of the subjects can be obtained by sequentially shifting the design d_1^* as follows.

Subjects	Period 1	Period 2	$\cdots$	Period p
1	$d_{1,1}^*$	$d_{1,2}^*$	$\cdots$	$d_{1,p}^*$
2	$d_{1,2}^*$	$d_{1,3}^*$	$\cdots$	$d_{1,1}^*$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
p	$d_{1,p}^*$	$d_{1,1}^*$	$\cdots$	$d_{1,p-1}^*$

Then set the design at period 0 by $d_{i,0}^* = d_{i,p}^*$ for $i = 1, ..., n$. The resulting design d^* is universally optimal under Model 1.

For $c > 1$, the universally optimal designs can be obtained by stacking c times of the optimal designs obtained from the above two steps.

Obtaining d_1^* in (1.2.10) is crucial to the construction of universally optimal designs. Apart from finding the optimal by searching over the feasible region $\boldsymbol{S_m}$, we propose a constructive procedure in Theorem 1.2.3. Using the d_1^* constructed by Theorem 1.2.3, the resulting optimal designs also satisfy $\boldsymbol{s_d} = 0$. Therefore, the resulting design is universally optimal for both models 1 and 2.

For notation simplicity, it is assumed that the t levels of the quantitative variable are equally spaced and denote them by $\{\frac{1-t}{2}, \frac{3-t}{2}, \cdots, \frac{t-1}{2}\}$. We first introduce some notation. Define $d_1^0 = [\frac{1-t}{2}, d_{1,2}^0, \cdots, d_{1,t}^0]$, where, $d_{1,j}^0 = (j \bmod 2)(j - \frac{t+3}{2}) - (j \bmod 2 - 1)(t - j + \frac{3-t}{2})$, when t is odd; and

$$d_{1,j}^0 = \begin{cases} (j \bmod 2)(j - \frac{t+3}{2}) - (j \bmod 2 - 1)(t - j + \frac{3-t}{2}) & j = 2, \cdots, (t/2) + 1, \\ (j \bmod 2)(t - j + \frac{3-t}{2}) - (j \bmod 2 - 1)(j - \frac{t+3}{2}) & j = (t/2) + 2, \cdots, t, \end{cases}$$

when t is even. Define $swap(i, j)$ as an operation that swap the elements i and j, and define $W_r(d_i)$ as a collection of swap operations for the design of subject i, where $r = 0, 1, 345, 7$. Denote $W_r(d_i) = U_r(d_i)$ for $r = 1$ and $W_r(d_i) = (U_r(d_i), V_r(d_i))$ for $r \neq 1$. The operations $U_r(d_i)$ and $V_r(d_i)$ are given below.

(i) $U_r(d_i) = swap(i_r, j_r)$ with $i_r + j_r = 0$ and $i_r = 4(k-1) + \frac{3-t}{2}, 4(k-1) + \frac{5-t}{2}$ for $k = 1, \cdots, \frac{t-r}{8}$.

(ii) $V_3 = swap(0,1)$, $V_4 = swap(\frac{k-1}{2}, \frac{k+1}{2})$ for $k = 2$ and 6, $V_5 = swap(0,2)$, $V_6 = swap(-\frac{5}{2}, \frac{3}{2})$, $V_7 = swap(-2,3)$, and $V_0 = swap(\frac{k-1}{2}, \frac{k+1}{2})$ for $k = 0, 2$, and 6.

Given d_1^0 as the initial design, the optimal design d_1^* can be constructed as follows.

Theorem 1.2.3. *(i) When $p = t > 8$ and $t \bmod 8 = r$, where $r \neq 2, r \neq 6$, the optimal*

treatment setting d_1^* *can be constructed by* $d_1^* = W_r(d_1^0)$ *and*

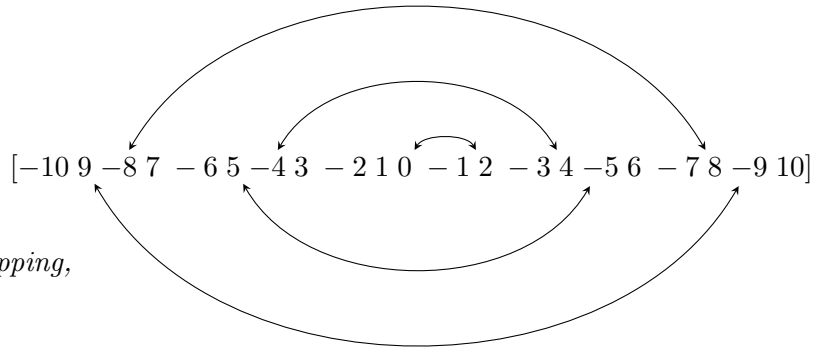
$$\sum_{j=1}^{t} d_{1,j}^* d_{1,j-1}^* = 0$$

(ii) When $p = \lambda t$*, where* $\lambda > 1$*,* $r \neq 2, r \neq 6$*,* d_1^* *can be constructed by repeating* d_1^* *from (i)* λ *times across the* p *periods and* $\sum_{j=1}^{p} d_{1,j}^* d_{1,j-1}^* = 0$*.*

An example of d_1^* is given below for $p = t = 21$.

Example 1.2.1. *Start with*

$d_1^0 = [-10\ 9\ -8\ 7\ -6\ 5\ -4\ 3\ -2\ 1\ 0\ -1\ 2\ -3\ 4\ -5\ 6\ -7\ 8\ -9\ 10]$*. According to Theorem 1.2.3,* $W_5(d_1^0) = (U_5(d_1^0), V_5(d_1^0))$*, where* $U_5 = swap(i_5, j_5)$*,* $i_5 = -9, -8, -5, -4$ *and the corresponding* $j_5 = 9, 8, 5, 4$*,* $V_5 = swap(0, 2)$*. Pairs needed to be swapped are illustrated as follows.*



After swapping,

$$d_1^* = [-10\ -9\ 8\ 7\ -6\ -5\ 4\ 3\ -2\ 1\ 2\ -1\ 0\ -3\ -4\ 5\ 6\ -7\ -8\ 9\ 10].$$

For $t < 9$, the design space is relatively smaller and therefore the optimal designs can be found by exhaustive search.

1.3 Crossover designs with multiple quantitative variables

Assume there are k independent quantitative variables and each of them has t levels. To construct the corresponding crossover designs, we consider the following model.

Model 3: Assume that τ_l is the direct treatment effect from variable l, γ_l is the carryover effect due to the previous design on variable l, where $l = 1, \ldots, k$. We have

$$y_{i,j} = \alpha_i + \beta_j + \sum_{l=1}^{k} \tau_l d_{i,j}^{(l)} + \sum_{l=1}^{k} \gamma_l d_{i,j-1}^{(l)} + e_{i,j}, \tag{1.3.1}$$

where α_i, β_j, $e_{i,j}$ are same as in Model 1, $d_{i,j}^{(l)}$ is the design for variable l of subject i at period j, and $d_{i,0}^{(l)}$ is the corresponding design for the initial period.

This model can be rewritten in a matrix form as follows

$$\boldsymbol{y} = \boldsymbol{U\alpha} + \boldsymbol{P\beta} + \boldsymbol{T_d}\tau + \boldsymbol{Q_d}\gamma + \boldsymbol{e}, \tag{1.3.2}$$

where $\boldsymbol{T_d} = [\boldsymbol{t_{d1}}, \cdots, \boldsymbol{t_{dk}}]$ and $\boldsymbol{Q_d} = [\boldsymbol{p_{d1}}, \cdots, \boldsymbol{p_{dk}}]$ are the design matrix of direct treatment effects and carryover effects for the k variables. $\boldsymbol{t_{dl}^T} = (d_{1,1}^{(l)}, \cdots, d_{n,p}^{(l)})$ and $\boldsymbol{p_{dl}^T} = (d_{1,0}^{(l)}, \cdots, d_{n,p-1}^{(l)})$. The information matrix for the direct treatment effect in Model 3 can be written as

$$C_d'' = \boldsymbol{T_d^T}\omega^{\perp}([\boldsymbol{P}, \boldsymbol{U}, \boldsymbol{Q_d}])\boldsymbol{T_d}.$$

Similar to Model 1, an upper bound of C_d'' can be derived and the sufficient conditions to achieve this upper bound is given below.

Corollary 1.3.1. *(i)* $C_d'' \leq \boldsymbol{T_d^T}\omega^{\perp}([\boldsymbol{U}, \boldsymbol{Q_d}])\boldsymbol{T_d}$ *and the equality holds if and only if*

$$\boldsymbol{T_d^T}\omega^{\perp}([\boldsymbol{U}, \boldsymbol{Q_d}])\boldsymbol{P} = 0. \tag{1.3.3}$$

(ii) The sufficient condition that satisfies equation (1.3.3) is:

$$\sum_{i=1}^{n} d_{i,0}^{(l)} = \sum_{i=1}^{n} d_{i,1}^{(l)} = \sum_{i=1}^{n} d_{i,2}^{(l)} = \cdots = \sum_{i=1}^{n} d_{i,p}^{(l)} \tag{1.3.4}$$

for all $1 \leq l \leq k$.

Denote $\Omega(k, t, n, p)$ as the design space for n subjects, p periods, and k quantitative variables with t levels. A design $d^* \in \Omega(k, t, n, p)$ which maximizes the trace of the

information matrix is T-optimal. Based on Corollary 1.3.1, the upper bound of C_d'' can be reached by the conditions given in (1.3.4). Therefore, T-optimal designs can be obtained by maximizing the trace of the upper bound of C_d'', given the conditions in (1.3.4) is satisfied. However, there is no explicit form for this maximization problem when $k > 2$ due to the matrix inversion in the calculation of $\omega^{\perp}([\boldsymbol{U}, \boldsymbol{Q_d}]) = \omega^{\perp}(\boldsymbol{U}) - \omega^{\perp}(\boldsymbol{U})\boldsymbol{Q_d}\{\boldsymbol{Q_d^T}\omega^{\perp}(\boldsymbol{U})\boldsymbol{Q_d}\}^{-}\boldsymbol{Q_d^T}\omega^{\perp}(\boldsymbol{U})$.

For $k = 2$, a closed form expression for the properties of T-optimal designs can be obtained. Define

$$a_1 = \sum_{i=1}^{n}\sum_{j=1}^{p} d_{i,j}^{(1)} d_{i,j-1}^{(1)},\ a_2 = \sum_{i=1}^{n}\sum_{j=1}^{p} d_{i,j}^{(2)} d_{i,j-1}^{(1)},$$
$$b_1 = \sum_{i=1}^{n}\sum_{j=1}^{p} d_{i,j}^{(1)} d_{i,j-1}^{(2)},\ b_2 = \sum_{i=1}^{n}\sum_{j=1}^{p} d_{i,j}^{(2)} d_{i,j-1}^{(2)},$$
$$x = \sum_{i=1}^{n}\sum_{j=1}^{p} (d_{i,j}^{(1)})^2,\ y = \sum_{i=1}^{n}\sum_{j=1}^{p} (d_{i,j}^{(2)})^2 \text{ and } z = \sum_{i=1}^{n}\sum_{j=1}^{p} d_{i,j}^{(1)} d_{i,j}^{(2)}.$$

In the next Theorem, properties of the T-optimal designs under Model 3 with two variables are specified. Similar to Model 1, we focus on the T-optimal designs within the class of k-variable uniform designs which is defined to be uniform for all the k variables.

Theorem 1.3.1. *A 2-variable uniform design $d^* \in \Omega(2, t, \mu t, \lambda t)$ is the T-optimal crossover design under Model 3 if it satisfies $xy - z^2 \neq 0$, $a_1 = a_2 = b_1 = b_2 = 0$, and $d_{i,0}^{(l)} = d_{i,p}^{(l)}$ for $i = 1, \ldots, n$ and $l = 1, 2$.*

For individual variables, the conditions are the same as those in Theorem 1.2.2 for one variable case, including $a_1 = b_2 = 0$ and $d_{i,0}^{(l)} = d_{i,p}^{(l)}$. The condition, $xy - z^2 \neq 0$, implies that designs for the two variables cannot be identical. The condition $a_2 = b_1 = 0$ implies that the design of any variable is orthogonal to the design of the other variable

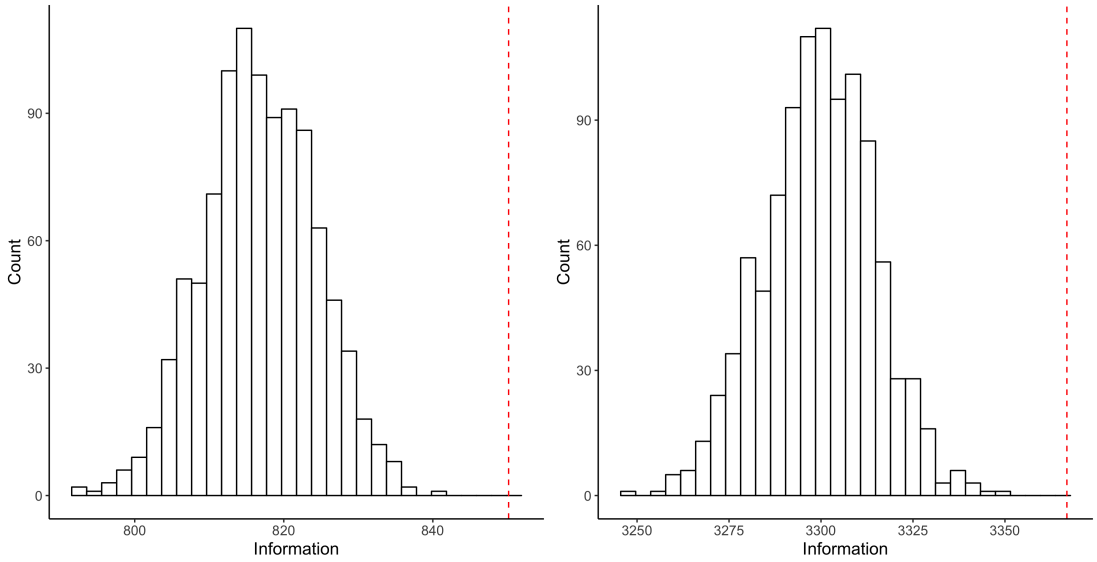
in the previous period.

1.4 Examples and comparisons

1.4.1 One quantitative variable

Based on the construction procedure in Theorem 1.2.3, universally optimal designs can be efficiently obtained for Models 1 and 2. The procedure is particularly useful for designs with large n, p, or t, where extensive search is often infeasible. Here we demonstrate two designs constructed by the proposed procedure with $n = p = t = 100$ and $n = p = t = 200$. For each sample size, the universally optimal design is compared with designs by a naive approach where, for each subject, design settings of the repeated experiments are randomly sampled from the experimental region.

The comparisons are summarized by Figure 1.1. The x-axis in Figure 1.1 stands for the values of the Fisher's information. The histogram demonstrates the distribution of the Fisher's information obtained based on 1000 replicates of the naive approach. The vertical dashed line corresponds to the Fisher information obtained by the proposed optimal design. It is clear that the optimal crossover design contains the largest information in estimating the treatment effect, compared with the randomly generated designs.



t=100 t=200

Figure 1.1: Comparison of the universally optimal crossover designs with randomly generated designs.

1.4.2 Crossover design for cell adhesion experiments with two quantitative variables

In this section, the cell adhesion experiment is revised and the optimal design is constructed for the two quantitative variables, contact time and waiting time. Assume that there are eight levels for both variables and there are eight periods, i.e., $p = t = 8$. The sample size is assumed to be $n = 50t$. Under Model 3, a 2-variable T-optimal crossover design can be constructed by Theorem 1.3.1 and the resulting d_1^* for each variable is given in Table 1.1.

Table 1.1: Optimal crossover design for cell adhesion experiments

Contact	-3.5	-1.5	-0.5	-2.5	3.5	0.5	2.5	1.5
Wait	-3.5	0.5	2.5	-2.5	-1.5	1.5	3.5	-0.5

The T-optimal design is compared with the naive approach as in Section 4.1. The result is shown in Figure 1.2. The histogram illustrates the distribution of the trace of the information matrix obtained based on 1000 replicates of the randomly generated designs. The red vertical dashed line corresponds to the trace of the information obtained by the optimal crossover design. According to Figure 1.2, it appears that the optimal crossover design significantly outperforms the design obtained by the naive approach in estimating the treatment effect. Compared with one variable cases, the estimation efficiency herein is improved by the proposed method with a much significant margin. This is because the condition of $a_2 = b_1 = 0$ can be easily violated in randomly generated designs and therefore the estimation efficiency deteriorates.

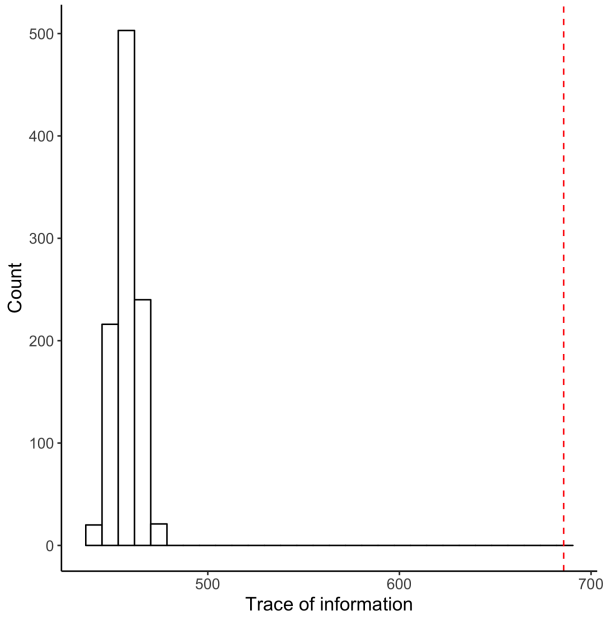


Figure 1.2: Comparison of the 2-variable optimal crossover design with randomly generated designs

1.5 Concluding remarks

Despite intensive studies of crossover designs, most of the results are developed for qualitative variables. Motivated by an experimental design problem in cell biology, new classes of optimal crossover designs are introduced for quantitative variables. Theoretical design properties are derived and compared with their counterparts with qualitative variables. To obtain these optimal designs efficiently, systematic construction procedures are proposed based on a collection of swap operations. Numerical studies demonstrate the estimation efficiency of the proposed optimal crossover designs.

1.6 Technical Proofs

Proof of Theorem 1.2.1. The proof of upper bound and equation (2.2.6) is same as **?**, Proposition 2.3.

The derivation of equation (1.2.6) is as below.

Since $\omega^{\perp}([\boldsymbol{U}, \boldsymbol{p_d}]) = \omega^{\perp}(\boldsymbol{U}) - \omega^{\perp}(\boldsymbol{U})\boldsymbol{p_d}\{\boldsymbol{p_d}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{p_d}\}^{-}\boldsymbol{p_d}^T\omega^{\perp}(\boldsymbol{U})$ (Kunert and Martin, 2000), $\boldsymbol{t_d}^T\omega^{\perp}([\boldsymbol{U}, \boldsymbol{p_d}])\boldsymbol{P}$

$= \boldsymbol{t_d}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{P} - \boldsymbol{t_d}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{p_d}\{\boldsymbol{p_d}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{p_d}\}^{-}\boldsymbol{p_d}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{P} = \boldsymbol{0}$ Sufficient condition satisfying (1.2.5) is: $\boldsymbol{t_d}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{P} = \boldsymbol{0}$ and $\boldsymbol{p_d}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{P} = \boldsymbol{0}$. Since $\omega^{\perp}(\boldsymbol{U}) = \boldsymbol{I} - 1/p\boldsymbol{U}\boldsymbol{U}^T$,

$$\boldsymbol{t_d}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{P} = (\textstyle\sum_{i=1}^{n} d_{i,1}, \sum_{i=1}^{n} d_{i,2}, \cdots, \sum_{i=1}^{n} d_{i,p}) = \boldsymbol{0}$$

$$\boldsymbol{p_d}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{P} = (\textstyle\sum_{i=1}^{n} d_{i,0}, \sum_{i=1}^{n} d_{i,1}, \cdots, \sum_{i=1}^{n} d_{i,p-1}) = \boldsymbol{0}$$

Therefore, $\sum_{i=1}^{n} d_{i,0} = \sum_{i=1}^{n} d_{i,1} = \sum_{i=1}^{n} d_{i,2} = \cdots = \sum_{i=1}^{n} d_{i,p} = 0$. This complete our proof. □

Proof of Theorem 1.2.2. Any designs in $\Omega(t, \mu t, \lambda t)$ that are uniform on periods with $d_{i,0} = d_{i,p}$ for all i will satisfy condition (1.2.6) in Theorem 1.2.1, therefore attaining upper bound of information. Next, we simplify the objective of obtaining optimal designs. By plugging in,

$$\begin{aligned}
&\max_{d} \boldsymbol{t_d}^T\omega^{\perp}([\boldsymbol{U}, \boldsymbol{p_d}])\boldsymbol{t_d} \\
=&\max_{d} \boldsymbol{t_d}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{t_d} - \boldsymbol{t_d}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{p_d}\{\boldsymbol{p_d}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{p_d}\}^{-}\boldsymbol{p_d}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{t_d} \qquad (1.6.1)\\
=&\max_{d} \sum_{i,j} d_{i,j}^2 - \Big(\sum_{i=1}^{n}\sum_{j=1}^{p} d_{i,j}d_{i,j-1}\Big)^2\Big(\sum_{i=1}^{n}\sum_{j=1}^{p} d_{i,j-1}^2\Big)^{-1}
\end{aligned}$$

Since designs are uniform in $\Omega(t, \mu t, \lambda t)$, all t settings will appear exactly np/t times for all n subjects across p periods. So $\sum_{i,j} d_{i,j}^2$ are constant. Also, $d_{i,0} = d_{i,p}$ for all

i indicates $\sum_{i,j} d^2_{i,j-1} = \sum_{i,j} d^2_{i,j}$. Therefore the objective function can be reduced to (1.2.7) in Theorem 1.2.2. This complete our proof. □

Proof of Corollary 1.2.1. Following Kunert and Martin (2000) and **?**,

$$\begin{aligned}
&\boldsymbol{t}_{\boldsymbol{d}}^T\omega^{\perp}([\boldsymbol{U},\boldsymbol{m_d},\boldsymbol{s_d}])\boldsymbol{t_d}\\
&= A_{d11} - A_{d12}A_{d22}^{-}A_{d12}^T - (A_{d13} - A_{d12}A_{d22}^{-}A_{d23})\\
&\times (A_{d33} - A_{d23}^TA_{d22}^{-}A_{d23})^{-}(A_{d13} - A_{d12}A_{d22}^{-}A_{d23})^T,
\end{aligned}$$

where $A_{d11} = \boldsymbol{t}_{\boldsymbol{d}}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{t_d}$, $A_{d12} = \boldsymbol{t}_{\boldsymbol{d}}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{m_d}$, $A_{d13} = \boldsymbol{t}_{\boldsymbol{d}}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{s_d}$,

$A_{d22} = \boldsymbol{m}_{\boldsymbol{d}}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{m_d}$, $A_{d23} = \boldsymbol{m}_{\boldsymbol{d}}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{s_d}$, $A_{d33} = \boldsymbol{s}_{\boldsymbol{d}}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{s_d}$.

With above notation, rewrite condition (1.2.8) as:

$$\begin{aligned}
&\boldsymbol{t}_{\boldsymbol{d}}^T\omega^{\perp}([\boldsymbol{U},\boldsymbol{m_d},\boldsymbol{s_d}])\boldsymbol{P}\\
&= \boldsymbol{t}_{\boldsymbol{d}}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{P} - A_{d12}A_{d22}^{-}\boldsymbol{m}_{\boldsymbol{d}}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{P} - (A_{d13} - A_{d12}A_{d22}^{-}A_{d23})\\
&\times (A_{d33} - A_{d23}^TA_{d22}^{-}A_{d23})^{-}(\boldsymbol{s}_{\boldsymbol{d}}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{P} - A_{d23}^TA_{d22}^{-}\boldsymbol{m}_{\boldsymbol{d}}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{P})\\
&= \boldsymbol{t}_{\boldsymbol{d}}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{P} - x_1\boldsymbol{m}_{\boldsymbol{d}}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{P} - x_2(\boldsymbol{s}_{\boldsymbol{d}}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{P} - x_3\boldsymbol{m}_{\boldsymbol{d}}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{P})\\
&= 0,
\end{aligned}$$

where x_1, x_2, x_3 are some constant since all A_{dij}, $1 \le i \le j \le 3$ are scalars. Therefore, sufficient condition satisfying (1.2.8) is: $\boldsymbol{s_d} = 0$, $\boldsymbol{t}_{\boldsymbol{d}}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{P} = 0$ and $\boldsymbol{m}_{\boldsymbol{d}}^T\omega^{\perp}(\boldsymbol{U})\boldsymbol{P} = 0$. That is,

$$\boldsymbol{s_d} = 0 \text{ and } \sum_{i=1}^{n} d_{i,0} = \sum_{i=1}^{n} d_{i,1} = \sum_{i=1}^{n} d_{i,2} = \cdots = \sum_{i=1}^{n} d_{i,p}$$

This complete our proof. □

Proof of Corollary 1.2.2. Since $\boldsymbol{s_d} = 0$, by Corollary 1.2.1, upper bound of information can be attained. Next, note that $A_{d13} = A_{d23} = A_{d33} = 0$ by $\boldsymbol{s_d} = 0$, $\max_d \boldsymbol{t_d^T}\omega^{\perp}([\boldsymbol{U}, \boldsymbol{m_d}, \boldsymbol{s_d}])\boldsymbol{t_d} = \max_d \left(A_{d11} - A_{d12}A_{d22}^{-}A_{d12}^{T}\right)$. This is same with (2.6.3). □

Proof of construction steps in section 1.2.3. $\min_d \mid \sum_{i=1}^{n}\sum_{j=1}^{p} d_{i,j}d_{i,j-1} \mid \geq 0$. So when $\sum_{j=1}^{p} d_{i,j}^{*}d_{i,j-1}^{*} = 0$ as in Step 1 and by construction in Step 2, the design d^* generated will satisfy $\sum_{i=1}^{n}\sum_{j=1}^{p} d_{i,j}^{*}d_{i,j-1}^{*} = 0$. Therefore it's optimal for $c = 1$.

When $c > 1$, i.e. $n = cp$, multiple of optimal designs for $c = 1$ are still optimal in $\Omega(t, c\lambda t, \lambda t)$ since it still satisfies minimal value is 0. □

Lemma 1.

$$\sum_{j=1}^{t} d_{1,j}^{0}d_{1,j-1}^{0} = \begin{cases} -\frac{1}{12}t^3 + \frac{7}{12}t - \frac{1}{2} & t \bmod 2 = 1 \\ -\frac{1}{12}t^3 + \frac{7}{12}t - 1 & t \bmod 2 = 0 \end{cases} \tag{1.6.2}$$

Proof of Lemma 1. For simplicity, we first add $\frac{t+1}{2}$ to each element of d_1^0.

When $t \bmod 2 = 1$, according to arrangement of d_1^0, after adding $\frac{t+1}{2}$, two adjacent numbers of j is $t - j$ and $t - j + 2$ for $1 < j < t$. Two adjacent numbers of 1 is t and

$t-1$ and two adjacent numbers of t is 1 and 2. So

$$\begin{aligned}
&\sum_{j=1}^{t}(d_{1,j}^0+\frac{t+1}{2})(d_{1,j-1}^0+\frac{t+1}{2})\\
=&\frac{1}{2}\Big((2t-1)+\sum_{j=2}^{t-1}j[(t-j)+(t-j+2)]+t(1+2)\Big)\\
=&\frac{1}{2}\Big(2t\sum_{j=2}^{t-1}j-2\sum_{j=2}^{t-1}j^2+2\sum_{j=2}^{t-1}j+5t-1\Big)\\
=&\frac{1}{2}\Big(t(t+1)(t-2)-2(\frac{2t^3+3t^2+t}{6}-1-t^2)+(t+1)(t-2)+5t-1\Big)\\
=&\frac{1}{6}t^3+\frac{1}{2}t^2+\frac{5}{6}t-\frac{1}{2}
\end{aligned}$$

So $\sum_{j=1}^{t} d_{1,j}^0 d_{1,j-1}^0$

$$\begin{aligned}
=&(\frac{1}{6}t^3+\frac{1}{2}t^2+\frac{5}{6}t-\frac{1}{2})-t(\frac{(t+1)}{2})^2\\
=&-\frac{1}{12}t^3+\frac{7}{12}t-\frac{1}{2}
\end{aligned}$$

When t is even, after adding $\frac{t+1}{2}$ to d_1^0, two adjacent numbers of j is $t-j$ and $t-j+2$ for $1<j<t$ and $j\neq\frac{t}{2},\frac{t}{2}+1$. The two adjacent number of $\frac{t}{2}$ is $\frac{t}{2}+1$ and $\frac{t}{2}+2$. And the two adjacent number of $\frac{t}{2}+1$ is $\frac{t}{2}$ and $\frac{t}{2}-1$. So similarly, $\sum_{j=1}^{t} d_{1,j}^0 d_{1,j-1}^0=\big(\frac{1}{6}t^3+\frac{1}{2}t^2+\frac{5}{6}t-1\big)-t(\frac{(t+1)}{2})^2=-\frac{1}{12}t^3+\frac{7}{12}t-1$ □

Proof of Theorem 1.2.3. (i) $p=t$.

Since the objective is to find $d_1^*=\underset{d_1\in\boldsymbol{S_m}}{\arg\min}\mid\big(f(d_1)-f(d_1^0)\big)+f(d_1^0)\mid$ and second term is fixed given by (1.6.2), d_1^* can be constructed by a collection of swap operations from d_1^0 such that $f(d_1)-f(d_1^0)=-f(d_1^0)$ if exist.

It remains to show $d_1^*=W_r(d_1^0)$. For any element $q\neq-\frac{1}{2},\frac{1}{2}$ in d_1^0, the two adjacent numbers are $1-q$ and $-1-q$. So for $u<v$, $u\neq-1-v,1-v$ and $u,v\neq-\frac{1}{2},\frac{1}{2}$, first $swap(u,v)$ and then $swap(-1-u,1-v)$ on d_1^0 will change $f(d_1^0)$ by $2(v-u)^2-4(v-u-2)$. So after $swap(4(k-1)+\frac{3-t}{2},-4(k-1)-\frac{3-t}{2})$ and $swap(4(k-1)+\frac{5-t}{2},-4(k-1)-\frac{5-t}{2})$,

$f(d_1^0)$ is changed by $2\big(8(k-1)+3-t\big)^2-4(8(k-1)+5-t)$, where $k=1,\cdots,\frac{t-r}{8}$. Denote $d_1^1=U_r(d_1^0)$. So the total change from $f(d_1^0)$ to $f(d_1^1)$ is:

$$
\begin{aligned}
&f(d_1^1)-f(d_1^0)\\
=&\sum_{k=1}^{\frac{t-r}{8}}\{2\big(8(k-1)+3-t\big)^2-4(8(k-1)+5-t)\}\\
=&\sum_{k=1}^{\frac{t-r}{8}}2\big(8k-3+r\big)^2-\sum_{k=1}^{\frac{t-r}{8}}4(8k-5+r)\\
=&\sum_{k=1}^{\frac{t-r}{8}}\big(128k^2+32(r-4)k+2r^2-16r+38\big)\\
=&128\sum_{k=1}^{\frac{t-r}{8}}k^2+32(r-4)\sum_{k=1}^{\frac{t-r}{8}}k+\frac{t-r}{8}(2r^2-16r+38)\\
=&\frac{1}{12}t^3+(\frac{-r}{12}+\frac{r}{4}-\frac{r}{6})t^2+(\frac{r^2}{3}-\frac{2r^2}{4}+\frac{t}{6}-\frac{7}{12})t+(\frac{-r^3}{3}+\frac{r^3}{4}+\frac{7r}{12})\\
=&\frac{1}{12}t^3-\frac{7}{12}t+\frac{7r-r^3}{12}
\end{aligned}
$$

When $r=1$, $f(d_1^1)-f(d_1^0)=\frac{1}{12}t^3-\frac{7}{12}t+\frac{1}{2}$. $f(d_1^1)=0$. So $d_1^*=d_1^1$.

When $r=2$, $f(d_1^1)-f(d_1^0)=\frac{1}{12}t^3-\frac{7}{12}t+\frac{1}{2}$. $f(d_1^1)=-\frac{1}{2}$. No swaps on d_1^1 can increase $f(d_1^1)$ by $\frac{1}{2}$. So $d_1^*=d_1^1$.

When $r=3$, $f(d_1^1)-f(d_1^0)=\frac{1}{12}t^3-\frac{7}{12}t-\frac{1}{2}$. $f(d_1^1)=-1$. $swap(0,1)$ will increase $f(d_1^1)$ by 1. So $d_1^*=V_3(U_3(d_1^0))=W_3(d_1^0)$.

When $r=4$, $f(d_1^1)-f(d_1^0)=\frac{1}{12}t^3-\frac{7}{12}t-3$. $f(d_1^1)=-4$. $V_4=swap(\frac{k-1}{2},\frac{k+1}{2})$ for $k=2$ and 6 will increase $Sp(d_1^1)$ by 4. So $d_1^*=V_4(U_4(d_1^0))=W_4(d_1^0)$.

When $r=5$, $f(d_1^1)-f(d_1^0)=\frac{1}{12}t^3-\frac{7}{12}t-\frac{90}{12}$. $f(d_1^1)=-8$. $swap(0,2)$ will increase $f(d_1^1)$ by 8. So $d_1^*=V_5(U_5(d_1^0))=W_5(d_1^0)$.

When $r=6$, $f(d_1^1)-f(d_1^0)=\frac{1}{12}t^3-\frac{7}{12}t-14.5$. $f(d_1^1)=-15.5$. $swap(-\frac{5}{2},\frac{3}{2})$ will increase $f(d_1^1)$ by 16. So $d_1^*=V_6(U_6(d_1^0))=W_6(d_1^0)$.

When $r=7$, $f(d_1^1)-f(d_1^0)=\frac{1}{12}t^3-\frac{7}{12}t-\frac{1}{2}$. $f(d_1^1)=-25$. $swap(-2,3)$ will increase

$f(d_1^1)$ by 25. So $d_1^* = V_7(U_7(d_1^0)) = W_7(d_1^0)$.

When $r = 0$, $f(d_1^1) - f(d_1^0) = \frac{1}{12}t^3 - \frac{7}{12}t - \frac{1}{2}$. So $f(d_1^1) = -1$. $swap(\frac{k-1}{2}, \frac{k+1}{2})$ for $k = 0, 2,$ and 6 in a sequence will increase $f(d_1^1)$ by 1. So $d_1^* = V_0(U_0(d_1^0)) = W_0(d_1^0)$.

(ii) $p = \lambda t$, where $\lambda > 1$ integers. Denote d_1^* constructed from (i) as d_1^{**}. When $t \bmod 2 = 1$ or $t \bmod 4 = 0$, $| \sum_{j=1}^{t} d_{1,j}^{**} d_{1,j-1}^{**} | = 0$. If d_1^* is constructed by repeating d_1^{**} p/t times aross p periods, then $| \sum_{j=1}^{p} d_{1,j}^{*} d_{1,j-1}^{*} | = | \frac{p}{t} \sum_{j=1}^{t} d_{1,j}^{**} d_{1,j-1}^{**} | = \lambda | \sum_{j=1}^{t} d_{1,j}^{**} d_{1,j-1}^{**} | = 0$ for $t \bmod 2 = 1$ or $t \bmod 4 = 0$.

This completes the proof. □

Proof of Theorem 1.3.1. Before we prove Theorem 1.3.1, we introduce Corollary 1.6.1.

Corollary 1.6.1. *A 2-variable uniform designs $d^* \in \Omega(2, t, \mu t, \lambda t)$ that satisfy $d_{i,0}^{(l)} = d_{i,p}^{(l)}$, for all i and l, is T-optimal if*

$$d^* = \arg\max_d \{x + y - \frac{1}{xy - z^2}\left((a_1^2 + a_2^2)y + (b_1^2 + b_2^2)x - 2(a_1 b_1 + a_2 b_2)z\right)\}, \quad (1.6.3)$$

where $x, y, z, a_1, a_2, b_1, b_2$ are defined in Theorem 1.3.1.

We first prove Corollary 1.6.1.

$$\begin{aligned} d^* &= \arg\max_d \left[\boldsymbol{t}_{\boldsymbol{d1}}^T \omega^{\perp}([\boldsymbol{U}, \boldsymbol{P_d}]) \boldsymbol{t}_{\boldsymbol{d1}} + \boldsymbol{t}_{\boldsymbol{d2}}^T \omega^{\perp}([\boldsymbol{U}, \boldsymbol{P_d}]) \boldsymbol{t}_{\boldsymbol{d2}} \right] \\ &= \arg\max_d \left[\left(\boldsymbol{t}_{\boldsymbol{d1}}^T \omega^{\perp}(\boldsymbol{U}) \boldsymbol{t}_{\boldsymbol{d1}} + \boldsymbol{t}_{\boldsymbol{d2}}^T \omega^{\perp}(\boldsymbol{U}) \boldsymbol{t}_{\boldsymbol{d2}} \right) \right. \\ &\quad - \left(\boldsymbol{t}_{\boldsymbol{d1}}^T \omega^{\perp}(\boldsymbol{U}) \boldsymbol{P_d} \{ \boldsymbol{P_d}^T \omega^{\perp}(\boldsymbol{U}) \boldsymbol{P_d} \}^{-} \boldsymbol{P_d}^T \omega^{\perp}(\boldsymbol{U}) \boldsymbol{t}_{\boldsymbol{d1}} \right. \\ &\quad \left. \left. + \boldsymbol{t}_{\boldsymbol{d2}}^T \omega^{\perp}(\boldsymbol{U}) \boldsymbol{P_d} \{ \boldsymbol{P_d}^T \omega^{\perp}(\boldsymbol{U}) \boldsymbol{P_d} \}^{-} \boldsymbol{P_d}^T \omega^{\perp}(\boldsymbol{U}) \boldsymbol{t}_{\boldsymbol{d2}} \right) \right] \end{aligned}$$

The rest can be proved by plugging in $\boldsymbol{t_{dl}}$, $l = 1, 2$, $\omega^{\perp}(\boldsymbol{U})$ and $\boldsymbol{P_d}$.

Now we prove Theorem 1.3.1. Second term in the parenthesis can be rewritten into: $\boldsymbol{t}_{\boldsymbol{d1}}^T \omega(\omega^{\perp}(\boldsymbol{U}) \boldsymbol{P_d}) \boldsymbol{t}_{\boldsymbol{d1}} + \boldsymbol{t}_{\boldsymbol{d2}}^T \omega(\omega^{\perp}(\boldsymbol{U}) \boldsymbol{P_d}) \boldsymbol{t}_{\boldsymbol{d2}}$. Since $\omega(\omega^{\perp}(\boldsymbol{U}) \boldsymbol{P_d})$ is a projection matrix

and it is positive semi-definite so $\boldsymbol{t_{d1}^T}\omega(\omega^{\perp}(\boldsymbol{U})\boldsymbol{P_d})\boldsymbol{t_{d1}} \geq 0$ and $\boldsymbol{t_{d2}^T}\omega(\omega^{\perp}(\boldsymbol{U})\boldsymbol{P_d})\boldsymbol{t_{d2}} \geq 0$. Therefore, if exists, d^* can be obtained when second term $\frac{1}{xy-z^2}\big((a_1^2+a_2^2)y+(b_1^2+b_2^2)x-2(a_1b_1+a_2b_2)z\big)=0$. That is,

$$(a_1^2+a_2^2)y+(b_1^2+b_2^2)x-2(a_1b_1+a_2b_2)z=0,\ xy-z^2\neq 0 \tag{1.6.4}$$

A sufficient condition satisfying (1.6.4) is $a_1=a_2=b_1=b_2=0$ and $xy-z^2\neq 0$. □

Proof of $d_1^0=\underset{d_1\in\boldsymbol{S_m}}{\arg\min}\sum_{j=1}^{p}d_{1,j}d_{1,j-1}$. We first prove $d_1^0+\frac{t+1}{2}=\underset{d_1\in\boldsymbol{S_m}}{\arg\min}\sum_{j=1}^{p}(d_{1,j}+\frac{t+1}{2})(d_{1,j-1}+\frac{t+1}{2})$.

Denote first row as $a_1,\cdots,a_t$, where $a_1,\cdots,a_t$ is a permutation of $1,2,\cdots,t$. Now the proof remains to show that arrangement of $a_1,\cdots,a_t$ as in the d_1^0 will minimize

$$\sum_{i=2}^{t}a_ia_{i-1}+a_1a_t \tag{1.6.5}$$

We will finish our proof in the following steps.

Denote two adjacent numbers for i as x_i and y_i. $x_i<y_i\neq i$, $1\leq x_i<y_i\leq t$, $i=1,\cdots,t$

Note that we don't distinguish the position (left or right) of x_i and y_i. Two adjacent of last element a_t in the row is a_{t-1} and a_1 and two adjacent of first element a_1 in the row is a_t and a_2.

The original minimization problem is equivalent to below:

$$\min\Big[(a_t+a_2)\times a_1+(a_1+a_3)\times a_2+\cdots+(a_{t-1}+a_1)\times a_t\Big] \tag{1.6.6}$$

or

$$\min\Big[(x_1+y_1)\times 1+(x_2+y_2)\times 2+\cdots+(x_t+y_t)\times t\Big] \tag{1.6.7}$$

That satisfies the following conditions.

(1) $x_i < y_i$, $1 \leq x_i < y_i \leq t$,

(2) Each value of 1 to t is repeated exactly two times

(3) $x_i < y_i \neq i$ for $i = 1, \cdots, t$

(4) x_i, y_i for $i = 1, \cdots, t$ satisfying first two conditions can be rearranged into same layout as in (1.6.6)

Value in objective function (1.6.6) doubles value in original objective function. Conditions in (1.6.7) are to make sure that (1.6.6) and (1.6.7) are equivalent.

$$\begin{aligned}
&(x_1 + y_1) \times 1 + (x_2 + y_2) \times 2 + \cdots + (x_t + y_t) \times t \\
=&x_1 \times 1 + x_2 \times 2 + \cdots + x_t \times t + y_1 \times 1 + y_2 \times 2 + \cdots + y_t \times t \\
=&\sum_1^t x_i \times i + \sum_1^t y_j \times j
\end{aligned}$$

Now we first look at a broader minimization problem that only satisfies condition 1 and condition 2. Assume $\boldsymbol{x} = \{x_1, \cdots, x_t\} \subset \boldsymbol{F} = \{1, 1, 2, 2, \cdots, t, t\}$ and $\boldsymbol{Z} = \{y_1, \cdots, y_t\} = F \setminus S$ such that $x_i < y_i$ for all i. $x_{(1)} \leq x_{(2)} \leq \cdots x_{(t)}$ and $y_{(1)} \leq y_{(2)} \leq \cdots y_{(t)}$ are the corresponding ordered numbers.

Note that after ordering, $x_{(i)} < y_{(i)}$ still satisfies condition (1). We can prove this in the following. For any $x_{(i)}$, the corresponding pair in y_j is denoted as $y_{x_{(i)}}$. Since $x_{(i)} < y_{x_{(i)}}$ for $k < i \leq t$ and $x_{(k)} \leq x_{(i)} < y_{x_{(i)}}$ for all $k \leq i \leq t$. $x_{(k)}$ is smaller than at least $t - k + 1$ y_j. Therefore, $x_{(k)} < y_{(k)}$.

By rearrangement inequality G. H & Littlewood (1952), for this given arrangement

of $x_1, \cdots, x_t$ and $y_1, \cdots, y_t$ under condition 1 and condition 2,

$$\begin{aligned} &\sum_1^t x_i \times i + \sum_1^t y_j \times j \\ \geq &\sum_1^t x_{(t+1-i)} \times i + \sum_1^t y_{(t+1-j)} \times j \end{aligned} \tag{1.6.8}$$

Denote the lower bound above as $LB(x_i, y_j)$. If $\min LB(x_i, y_j)$ attained also satisfies condition (3) and (4), then it must be the optimal solution satisfying all conditions. If not, denote the lower bound of arrangement in $d_1^0 + \frac{t+1}{2}$ as LB^*. If for all $LB(x_i, y_j) < LB^*$, they don't satisfy other two conditions no matter how we permute x_i and y_j under the first two conditions, then LB^* is the our optimal solution satisfying all conditions.

Next we will show that LB^* is the optimal solution that satisfies all conditions.

If we write $d_1^0 + \frac{t+1}{2}$ according to (1.6.8), clearly, $x_{(1)} = x_{(2)} = 1$, $y_{(t)} = y_{(t-1)} = t$, $x_{(k)} = k - 1$ for $k > 2$ and $y_{(k)} = k + 1$ for $k < t - 1$. This arrangement gives LB^*.

Since $x_{(1)}, x_{(2)}, y_{(t-1)}, y_{(t)}$ are fixed, we will first find all $2 \leq x_{(i)} \leq (t-1)$ for $i > 2$ and $2 \leq y_{(j)} \leq (t-1)$ for $j < t - 1$ that corresponding $LB(x_i, y_j) < LB^*$.

$$\begin{aligned} &\sum_1^t x_i \times i + \sum_1^t y_j \times j \\ \geq &\sum_1^t x_{(t+1-i)} \times i + \sum_1^t y_{(t+1-j)} \times j \\ = &t + (t-1) + \sum_1^{t-2} x_{(t+1-i)} \times i + \sum_3^t y_{(t+1-j)} \times j + t + 2t \\ = &t + (t-1) + \sum_3^t x_{(i)} \times (t+1-i) + \sum_1^{t-2} y_{(j)} \times (t+1-j) + t + 2t \end{aligned}$$

Denote the ordered repeated numbers in $x_{(i)}$ (exclude 1) as $s_1, \cdots, s_r$ and ordered repeated numbers in $y_{(j)}$ (exclude t) as $t_1, \cdots, t_r$. (The total number of repeated values in $x_{(i)}$ and $y_{(j)}$ must be same since we have t numbers with each repeating two times. The number of non-repeating values must be same for $x_{(i)}$ and $y_{(j)}$. Therefore making

number of repeated values in $x_{(i)}$ and $y_{(j)}$ same since $x_{(i)}$ and $x_{(i)}$ have same length.) Compared with $x_{(i)} = i-1$ for $i > 2$ and $y_{(j)} = j+1$ for $j < t-1$, a necessary condition for achieving smaller lower bound is there exist k such that $s_k > t_k$.

Assume $s_i < t_i$ for all $i = 1, \cdots, r$. We first look at the case when $r = 1$. $x_{(i)} \times (t + 1 - i) = (i-1) \times (t + 1 - i)$ will be decreased by $((t-1) - s_1) + ((t-1) - (s_1 + 1)) + \cdots + ((t-1) - (t_1 - 1))$.

$y_{(j)} \times (t + 1 - j) = (j + 1) \times (t + 1 - j)$ will be increased by $((t + 3) - (s_1 + 1)) + ((t + 3) - s_1) + ((t + 3) - (s_1 + 1)) + \cdots + ((t + 3) - t_1)$.

Therefore, compared with $x_{(i)} = i - 1$ for $i > 2$ and $y_{(j)} = j + 1$ for $j < t - 1$, the lower bound is increased by $3 \times (t_1 - s_1)$.

When $r > 1$, similarly, the increase in $y_{(j)} \times (t + 1 - j)$ is always greater than the decrease in $x_{(i)} \times (t+1-i)$ making the lower bound increased. Only when $s_k > t_k$, the lower bound is possibly decreased.

Now we prove that for any arrangement generating smaller lower bound, no matter how we permute x_i and y_j satisfying condition 1, they don't satisfy condition 4.

Recall the necessary condition of possible smaller lower bound is there exist some k such that $s_k > t_k$, where s_k is kth repeated number in $x_{(i)}$ excluding 1 and t_k is kth repeated number in $y_{(j)}$ excluding t. Assume that smallest k satisfying $s_k > t_k$ is k^*, $s_{k^*} > t_{k^*}$ and $s_i < t_i$ for all $i < k^*$.

For $y_{(j)} \leq t_{k^*-1}$, there are $k^* - 1$ repeated numbers and jumps (One jump is counted as jump from $m - 1$ to $m + 1$. If skipping two numbers, it is counted as two jumps. If $y_{(1)} = l$, $l > 2$, its counted as $l - 2$ jumps.) Therefore, $y_{(t_{k^*-1}-2)} = y_{(t_{k^*-1}-1)} = t_{k^*-1}$. And $y_{(j)} = j + 1$ for all $t_{k^*-1} - 1 \leq j < t_k^*$. $y_{(t_{k^*}-1)} = y_{(t_{k^*})} = t_{k^*}$.

Similarly, when $t_{k^*} - t_{k^*-1} > 1$ i.e. k^*th repeated number in $y_{(j)}$ is not $t_{k^*-1} + 1$,

then for $x_{(i)} \leq t_{k^*-1}+1$, there are k^*-1 repeated numbers and jumps. Therefore, $x_{(t_{k^*-1}+2)} = t_{k^*-1}+1$. And $x_{(i)} = i-1$ for all $t_{k^*-1}+2 \leq i < t_k^*+1$. $x_{(t_{k^*})} = t_{k^*}-1$ and $x_{(t_{k^*}+1)} = t_{k^*}+1$. Therefore, for $i \leq t_k^*$, all $x_{(i)} < y_{(i)} \leq t_k^*$ and for $i > t_k^*$, $x_{(i)} > t_k^*$ and $y_{(i)} > t_k^*$.

When $t_{k^*} - t_{k^*-1} = 1$, $y_{(t_{k^*-1})} = y_{(t_{k^*-1}+1)} = t_{k^*}$, $x_{(t_{k^*-1}+2)} = x_{(t_{k^*-1}+3)} = t_{k^*-1}+2 = t_{k^*}+1 = s_k^*$ since $y_{(t_{k^*-1}+2)} \neq t_{k^*}+1$. (If $y_{(t_{k^*-1}+2)} = t_{k^*}+1$, $x_{(t_{k^*-1}+2)} < t_{k^*}+1$. And $x_{(t_{k^*-1}+2)} > t_{k^*-1}-1$ from above. So $x_{(t_{k^*-1}+2)} = t_{k^*-1}$ or t_{k^*}, which is not possible.) Therefore, for $i \leq t_{k^*-1}+1 = t_{k^*}$, all $x_{(i)} \leq t_k^*$ and $y_{(i)} \leq t_k^*$ and for $i > t_{k^*-1}+1 = t_k^*$, all $x_{(i)} > t_k^*$ and $y_{(i)} > t_k^*$.

In both cases, partition $\{x_{(i)}, y_{(i)}\}$ pairs into two blocks $i \leq t_k^*$ and $i \geq t_k^*+1$, where first block takes values $\{11, 22, \cdots, t_k^*, t_k^*\}$ and second block takes values $\{t_k^*+1, t_k^*+1, \cdots, t, t\}$.

Any permutation on x_i and y_j must still satisfy $x_k < y_k$. Despite the position of $y_{(1)}, \cdots, y_{(t_k^*)}$ in the new permutation, corresponding x_i values could only be a permutation within the first block satisfying $x_i < y_i$. Therefore, if switching position of the triplets of $\{i, x_i, y_i\}$ so that all $y_{(1)}, \cdots, y_{(t_k^*)}$ are together, $\{1, 1, 2, 2, \cdots, t_k^*, t_k^*\}$ will appear in the same $\{x_i, y_i\}$ block and separated from remaining numbers.

However, when t is odd, according to condition (3), there exist no partition in all $\{x_{a_i}, y_{a_i}\}$ such that $\{1, 1, \cdots, f, f\}$ and $f+1, f+1, \cdots, t, t$ are separated for any $1 \leq f \leq t-1$.

Therefore, even though there exist k such that $s_k > t_k$ will possibly gives a smaller lower bound, all eligible permutations under first two conditions will not satisfy condition 4. The proof is therefore complete when t is odd.

When t is even, according to condition (3), there exist one and only one partition on the indices set for all triplets $\{a_i, x_{a_i}, y_{a_i}\}$, $I_1 = \{1, 3, 5, \cdots, t-1\}$ and $I_2 = \{2, 4, 6, \cdots, t\}$ such that $\{x_{a_i}, y_{a_i}\}$ are partitioned into $\{a_2, a_2, a_4, a_4 \cdots, a_t, a_t\}$ and $\{a_1, a_1, a_3, a_3, \cdots, a_{t-1}, a_{t-1}\}$. Since we are dividing all pairs of $\{x_{a_i}, y_{a_i}\}$ into half, there is only one repeated number $t_1^* = t/2$ such that $t_1^* < s_1^*$. There shall be only one repeated number such that $t_k < s_k$ (excluding 1 and t) since there will be more partitions if not. There is no other repeated numbers satisfying $s_i < t_i$ since this will make the lower bound larger than our optimal solution. This can be seen by recalling above that even with only one $s_1 < t_1$, the lower bound is increased by $3 \times (t_1 - s_1)$ compared with $x_{(i)} = i - 1$ for $i > 2$ and $y_{(j)} = j + 1$ for $j < t - 1$. In the extreme case, $t_1 - s_1 = 1$. The lower bound is still increased by 3. However, with only one repeated number $s_k > t_k$. The lower bound is only decreased by 1. This can be seen as below.

For $s_k > t_k$, we have as above: $y_{(t_k-1)} = y_{(t_k)} = t_k$ and $x_{(t_k)} = t_k - 1$, $x_{(t_k+1)} = t_k + 1$. $x_{(i)} \times (t + 1 - i) = (i - 1) \times (t + 1 - i)$ will be increased by $t + 1 - (t_k + 1)$ while $y_{(j)} \times (t + 1 - j) = (j + 1) \times (t + 1 - j)$ will be decreased by $t + 1 - t_k$. So the total lower bound is decreased only by 1.

Therefore, the optimal solution for t is even case is as below: $y_{(t/2)} = t/2$ and $y_{(j)} = j + 1$ for $j < t \neq t/2$. $x_{(t/2+1)} = t/2 + 1$ and $x_{(i)} = i - 1$ for $1 < i \neq t/2 + 1$. This complete our proof.

□

Chapter 2

Bayesian Quantile Analysis for Computer Experiments

2.1 Introduction

Computer experiments refer to those experiments that are performed in computers using complex mathematical models. Computer experiments are becoming popular because many physical experiments are difficult or impossible to perform. However, computer simulations are typically time-consuming–it is normal for code to run for 12 hours or longer to produce a single response and the number of variables involved is usually large–15 to 20 or more variables, so it is infeasible to perform all the combinations of experiments. Given these two features of computer experiments, it is common to construct statistical models as a surrogate and perform inference and optimization based on it. In addition, computer experiments are deterministic in the sense that a particular input produces the same output if given to the computer experiments on another occasion. Therefore, it is desirable to build an interpolating surrogate model for computer experiments outputs. A Gaussian process (GP) model, also known as Kriging, is a widely used surrogate model for the analysis of computer experiments because of its interpolation property (Santner et al., 2003; Fang et al., 2005).

Despite the popularity of GP modeling in various applications, GP is mainly designed to model the mean structure of computer outputs. In many scientific studies,

it is often of interest to understand the impacts of input variables on different quantiles of the response in computer experiments. However, to the best of our knowledge, research on quantile regression for computer experiments is limited. Therefore a new model which analogue to GP for quantile analysis in computer experiments is called for. In the example of Community Ice Sheet Model (CISM) (Rutt et al., 2009; Price et al., 2011), thickness of ice sheet can be used to understand ice sheet behavior and its impact on climate. By predicting the quantiles of ice sheet thickness, a distribution of ice sheet thickness can be obtained. Predictions on extreme values of icesheet thickness are critical in the study of climate change.

Conventional parametric linear quantile regression (Koenker and Bassett Jr, 1978; Koenker and Hallock, 2001) can not capture the non-linearity of quantile function. Non-parametric methods including spline based or nearest neighbor or Kernal based quantile regression models (Stone, 1977; Koenker et al., 1994; Yu and Jones, 1998; Koenker, 2005; Roger Koenker, 2017) can capture the non-linearity of quantile function but the inference is resampling based and choosing proper band width is also a challenging problem. To capture the non-linearity and smoothness in quantile prediction for computer experiments and to build a parametric model, we propose a Bayesian quantile analysis model with asymmetric Laplace distribution (ALP) as likelihood and a GP prior on coefficients. Asymmetric Laplace distribution on error is used to build a parametric model and GP prior on coefficients is used for building a nonlinear quantile model with interpolation property. In our proposed model, quantile predictions achieve a quantile interpolation property asymptotically and coefficient estimates are shown to be asymptotically consistent with some constraint.

The reminder of this chapter is organized as follows. In Section 2.2, the new Bayesian

quantile model for computer experiments is introduced. The asymptotic properties are given in section 2.3. Simulations and real data example are given in Section 2.4. Final discussions are given in Section 2.5. All technical proofs are given in section 2.6.

2.2 Bayesian quantile model for computer experiments

Suppose we have a set of data $(\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_n, y_n)$ with input $\boldsymbol{x}_i \in R^p$ and output $y_i \in R$. Quantile regression is about modeling conditional quantile functions (Koenker and Bassett Jr, 1978)

$$g = \arg\min_{g \in G} \sum_{i=1}^{n} \rho_\tau(y_i - g(\boldsymbol{x}_i)),$$

where $g(\boldsymbol{x})$ is the quantile function and $\rho_\tau(u) = u(\tau - \mathbb{1}\{u < 0\})$ is the check function with $\mathbb{1}$ denoting the indicator function. In regular linear quantile regression, $g(\boldsymbol{x}_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}$. In non-parametric quantile regression, $g(\boldsymbol{x}_i)$ is represented using splines. Kernal based non-parametric quantile regresssino is to add Kernal weight to the loss function.

In conventional Bayesian linear quantile regression, the distribution of y is assumed to be asymmetric Laplace (ALP) distribution with probability density function $f(y|\sigma_\epsilon^2, \tau) = \frac{\tau(1-\tau)}{\sigma_\epsilon^2} \exp(-\rho_\tau(\frac{y - \boldsymbol{x}^T\boldsymbol{\beta}}{\sigma_\epsilon^2}))$. ALP is used since minimizing check function corresponds to maximizing likelihood of ALP distribution (Yu and Moyeed, 2001). To build a Bayesian parametric model for computer experiments, y is also assumed to be ALP in our model.

O'Hagan (1978) used GP prior on coefficients in mean regression to achieve non-linear curve fitting. GP prior can characterize the dependence of coefficient β on x that can reflect beliefs about the smoothness of the true response function. So to capture the non-linearity and smoothness of quantile function, similarly, the prior of coefficient is assumed to be GP in the Bayesian quantile model for computer experiments.

It is shown in the next section that with GP prior assumption on coefficients and ALP likelihood distribution, asymptotic interpolation property is achieved for quantile prediction. Before that, the new Bayesian quantile analysis models for computer experiments is explained for one covariate i.e. $p = 1$ case below.

2.2.1 One covariate model

In Bayesian quantile analysis for computer experiments, to model conditional quantile of output y with ALP likelihood and GP prior on coefficient, y can be written as

$$y = x\beta(\tau, x) + \epsilon_\tau, \tag{2.2.1}$$

where $\epsilon_\tau \sim \text{ALP}(0, \sigma_\epsilon^2, \tau)$ and $\beta(\tau, x)$ follows a stationary Gaussian process with mean $b(\tau)$ and covariance $\sigma^2(\tau)\phi$ and covariance function is defined by $\text{cov}(\beta(x_k), \beta(x_{k'})) = \sigma^2(\tau)\phi\big((x_k - x_{k'}); \theta(\tau)\big)$. $\beta(\tau, x)$, $b(\tau)$, $\sigma^2(\tau)$ and $\theta(\tau)$ includes τ since for different quantile τ, all parameters are different. σ_ϵ^2 is scale parameter and τ is the skewness parameter in the ALP distribution. According to Kozubowski et al. (2013) and Lum and Gelfand (2012), ϵ_τ can be represented by $\epsilon_\tau = \sigma_\epsilon\sqrt{\frac{2\xi}{\tau(1-\tau)}}Z + \sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}\xi$, where $Z \sim N(0,1)$ and $\xi \sim Gamma(1,1)$. So $\epsilon_\tau|\xi \sim N\big(\sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}\xi, \frac{2\sigma_\epsilon^2\xi}{\tau(1-\tau)}\big)$. Let $q_\tau(x) = x\beta(\tau, x)$,

$$(y - \sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}\xi)|q_\tau(x), \xi \sim N\big(q_\tau(x), \frac{2\sigma_\epsilon^2\xi}{\tau(1-\tau)}\big) \tag{2.2.2}$$

Assume parameters $\sigma^2(\tau)$, $\theta(\tau)$, $b(\tau)$, σ_ϵ^2 are known. Given observed inputs realizations $x_1, \cdots, x_M$. When multiple outputs are observed for same set of $x_1, \cdots, x_M$ and denote the observed outputs as $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n$, where $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{iM})^T$ is output vector corresponding to $x_1, \cdots, x_M$. Posterior mean and variance of $q_\tau(x)$ given $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n$

and ξ is

$$
\begin{aligned}
&E(q_\tau(x)|\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n,\xi)\\
=&h_\tau(x)+\boldsymbol{r}_\tau(x)^T(\frac{1}{n}\frac{2\sigma_\epsilon^2\xi}{\tau(1-\tau)}I+\Sigma_\tau)^{-1}(\frac{1}{n}\sum_{i=1}^n\boldsymbol{y}_i-\sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}\xi 1-\boldsymbol{h_\tau}) \qquad (2.2.3)\\
\to&h_\tau(x)+\boldsymbol{r}_\tau(x)^T{\Sigma_\tau}^{-1}(\frac{1}{n}\sum_{i=1}^n\boldsymbol{y}_i-\sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}\xi 1-\boldsymbol{h}_\tau)
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Var}(q_\tau(x)|\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n,\xi)\\
=&v(x,x)-\boldsymbol{r}(x)^T(\frac{2\sigma_\epsilon^2\xi}{n\tau(1-\tau)}I+\Sigma_\tau)^{-1}\boldsymbol{r}(x), \qquad (2.2.4)
\end{aligned}
$$

where 1 is length M all-ones vector. Given prior information, $h_\tau(x)=xb(\tau)$ is the mean of $q_\tau(x)$, $\boldsymbol{h}_\tau=(h_\tau(x_1),\cdots,h_\tau(x_M))^T$ is the mean of $(q_\tau(x_1),\cdots,q_\tau(x_M))^T$, $v_\tau(x_k,x_{k'})=x_kx_{k'}\sigma^2(\tau)\phi(|x_k-x_{k'}|,\theta(\tau))$ is covariance between $q_\tau(x_k)$ and $q_\tau(x_{k'})$, $\boldsymbol{r}_\tau(x)$ is the $M\times 1$ vector whose k^{th} element is $v_\tau(x,x_k)$, I is an identity matrix and Σ_τ is $M\times M$ covariance matrix whose $(k,k')^{th}$ elements is $v_\tau(x_k,x_{k'})$.

The covariance structure of response y in Bayesian quantile model is dependent on x so y is not a stationary process. This is different from GP model for mean regression in the sense that covariance of responses y are only dependent on distance of different x and variance of response y is constant for all x. The correlation structure of y in above Bayesian quantile model with one covariate, however, is dependent only on distance of x which is same with correlation of response y in GP model for mean regression.

The general case derivation of (2.2.3) and (2.2.4) is given in the next section.

Based on (2.2.3) and (2.2.4), we have the following results.

Theorem 2.2.1. *Posterior mean and variance of $q_\tau(x)$ given $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n$ as $n\to\infty$ is*

$$
\begin{aligned}
&E(q_\tau(x))|\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n)\to h_\tau(x)+\boldsymbol{r}_\tau(x)^T{\Sigma_\tau}^{-1}(\frac{1}{n}\sum_{i=1}^n\boldsymbol{y}_i-\frac{(1-2\tau)\sigma_\epsilon}{\tau(1-\tau)}1-\boldsymbol{h_\tau}) \qquad (2.2.5)\\
&Var(q_\tau(\boldsymbol{x})|\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n)\to(\frac{(1-2\tau)\sigma_\epsilon}{\tau(1-\tau)}\boldsymbol{r}(x)^T\Sigma_\tau^{-1}1)^2+v(x,x)-\boldsymbol{r}(x)^T\Sigma_\tau^{-1}\boldsymbol{r}(x)
\end{aligned}
$$

Posterior mean in (2.2.5) can be used for quantile prediction for new x_{M+1} and when new x_{M+1} coincides with some x_k for $k = 1, \cdots, M$, the quantile prediction is:

$$\begin{aligned}
\widehat{q_\tau(x_k)} &= h_\tau(x_k) + \boldsymbol{r}_\tau(x_k)^T V^{-1}(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i - \sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}1 - \boldsymbol{h}_\tau) \\
&= h_\tau(x_k) + \frac{1}{n}\sum_{i=1}^{n} y_{ik} - \sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)} - h_\tau(x_k) \\
&= \frac{1}{n}\sum_{i=1}^{n} y_{ik} - \sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)} \\
&\xrightarrow{p} E(y_k) - \sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)},
\end{aligned}$$

as $n \to \infty$. Also by (2.2.2), $E(y_k|q_\tau(x_k)) = E[E(y_k|q_\tau(x_k), \xi)|\xi] = E[\sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}\xi + q_\tau(x_k)|\xi] = \sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)} + q_\tau(x_k)$, so $E(y_k) - \sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)} = q_\tau(x_k)$. Therefore, $\widehat{q_\tau(x_k)} \xrightarrow{p} q_\tau(x_k)$.

Definition 2.2.1. *Given observed input-output pairs $(\boldsymbol{x}, y)$. Define $q_\tau(\boldsymbol{x})$, τ^{th} quantile of y given $\boldsymbol{x}$. So for any observed $\boldsymbol{x}$, we have underlying input-quantile pairs $(\boldsymbol{x}, q_\tau(\boldsymbol{x}))$. When quantile predictions agree on $q_\tau(\boldsymbol{x})$ for all observed inputs $\boldsymbol{x}$ and quantile prediction of unobserved inputs are dependent on these known input-quantile pairs, corresponding quantile prediction model is called a quantile interpolator.*

Therefore quantile prediction using proposed model for one covariate case is a quantile interpolator as $n \to \infty$.

In the next subsection, a general model with p covariates is given.

2.2.2 Multiple covariates model

For $p > 1$ covariates, assume the effects of different covariates are additive and assume the effect of covariate x_j depends only on the value of x_j. Using assumptions above, y's τth conditional quantile given inputs $\boldsymbol{x} = (x_1, x_2, \cdots, x_p)$ is modeled in the following.

$$y = x_1\beta_1(\tau, x_1) + \cdots + x_p\beta_p(\tau, x_p) + \epsilon_\tau, \tag{2.2.6}$$

where ϵ_τ is same as one covariate case. To simplify notation, only linear covariate terms are assumed in model (2.2.6). All results hold when covariates are in polynomial terms.

Assume the prior on the coefficients for jth covariate x_j is as below.

$$\beta_j(\tau, x_j) \sim \mathrm{GP}\big(b_j(\tau), \sigma_j^2(\tau)\phi\big), \tag{2.2.7}$$

where $b_j(\tau)$ is mean and covariance function is defined by $\mathrm{cov}(\beta_j(\tau, x_{js}), \beta_j(\tau, x_{jt})) = \sigma_j^2(\tau)\phi\big((x_{js} - x_{jt}); \theta_j(\tau)\big)$, where x_{js} and x_{jt} denotes two levels of jth covariate x_j. $\sigma_j^2(\tau)$ and $\theta_j(\tau)$ are parameters in the covariance matrix. $\beta_j(\tau, x_j)$ are independent of each other for different j and τ. Without loss of generality, assume correlation function is in L_1 norm: $\phi\big((x_{js} - x_{jt}; \theta_j(\tau)\big) = \exp\left(-\theta_j(\tau)\,|x_{js} - x_{jt}|\right)$.

Let $\boldsymbol{\beta}(\tau, \boldsymbol{x})$ denote the p dimensional vector of $(\beta_1(\tau, x_1), \cdots, \beta_p(\tau, x_p))$, the matrix form of model (2.2.6) is:

$$y = \boldsymbol{x}^T \boldsymbol{\beta}(\tau, \boldsymbol{x}) + \epsilon_\tau$$

By prior distribution (2.2.7), the prior distribution of $\boldsymbol{\beta}(\tau, \boldsymbol{x})$ is p dimensional Gaussian distribution with mean $\boldsymbol{b}(\tau) = (b_1(\tau), \cdots, b_p(\tau))^T$ and covariance

$$C(\boldsymbol{x}_k, \boldsymbol{x}_{k'}) = \mathrm{cov}(\boldsymbol{\beta}(\tau, \boldsymbol{x}_k), \boldsymbol{\beta}(\tau, \boldsymbol{x}_{k'})) = \begin{bmatrix} \sigma_1^2\phi(x_{k1} - x_{k'1}; \theta_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_p^2\phi(x_{kp} - x_{k'p}; \theta_p) \end{bmatrix}$$

In the following section, main theory of interpolation property of quantile prediction is given. Bayesian estimation of coefficients and its theoretical property is also provided.

2.3 General Theory

In this section, since all below properties are true for any given τ, for simplicity, we ignore index τ for all notations. b_j, σ_j^2, θ_j are used to denote $b_j(\tau)$, $\sigma_j^2(\tau)$, $\theta_j(\tau)$.

Denote $q_\tau(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta}(\tau, \boldsymbol{x})$, similar to (2.2.2), for $p > 1$,

$$y - \sigma_\epsilon \frac{1-2\tau}{\tau(1-\tau)}\xi | q_\tau(\boldsymbol{x}), \xi \sim N\big(q_\tau(\boldsymbol{x}), \frac{2\sigma_\epsilon^2 \xi}{\tau(1-\tau)}\big) \tag{2.3.1}$$

Assume σ_j^2, θ_j, b_j, σ_ϵ^2 are known. Given observed inputs realizations $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_M$, where $\boldsymbol{x}_k = (x_{k1}, \cdots, x_{kp})^T$ is length p covariates vector corresponding to kth setting and $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n$ are same as above. Similar to (2.2.5), the objective is to find posterior mean and variance of $q_\tau(\boldsymbol{x})$ given $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n$, where $\boldsymbol{x} = (x_1, \cdots, x_p)^T$. The main results are given in Theorem 2.3.1.

Before introducing Theorem 2.3.1, we first introduce some notations.

Based on prior assumption, the mean of $q_\tau(\boldsymbol{x})$ is denoted as $h(\boldsymbol{x}) = \sum_{j=1}^p x_j b_j$. The mean of $\boldsymbol{y}_i$ for $\forall i = 1, \cdots, n$ is denoted as $\boldsymbol{h} = (h(\boldsymbol{x}_1), \cdots, h(\boldsymbol{x}_M))^T$. Covariance between $q_\tau(\boldsymbol{x}_k)$ and $q_\tau(\boldsymbol{x}_{k'})$ is denoted as $v(\boldsymbol{x}_k, \boldsymbol{x}_{k'}) = \boldsymbol{x}_k^T C(\boldsymbol{x}_k, \boldsymbol{x}_{k'})\boldsymbol{x}_{k'}$, $\boldsymbol{t}(\boldsymbol{x}) = (v(\boldsymbol{x}, \boldsymbol{x}_1), \cdots, v(\boldsymbol{x}, \boldsymbol{x}_M))^T$ denotes the covariance vector between $\boldsymbol{x}$ and $\boldsymbol{x}_k$ for $k = 1, \cdots, M$ and V is $M \times M$ covariance matrix of $(q_\tau(\boldsymbol{x}_1), \cdots, q_\tau(\boldsymbol{x}_M))^T$ whose $(k, k')^{th}$ element is $v(\boldsymbol{x}_k, \boldsymbol{x}_{k'})$.

Similar to the one covariate case, given prior information, the mean and covariance structure of response y are dependent on $\boldsymbol{x}$ so y is not a stationary process.

Theorem 2.3.1. *Assume (2.2.6) and (2.2.7) is satisfied and for given τ,*

(i) As $n \to \infty$, quantile prediction and its variance can be estimated using posterior mean and variance of $q_\tau(\boldsymbol{x})|\boldsymbol{y_1}, \cdots, \boldsymbol{y_n}$ below

$$E(q_\tau(\boldsymbol{x})|\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n) \to h(\boldsymbol{x}) + \boldsymbol{t}(\boldsymbol{x})^T V^{-1}\big(\frac{1}{n}\sum_{i=1}^n \boldsymbol{y}_i - \sigma_\epsilon \frac{1-2\tau}{\tau(1-\tau)}1 - \boldsymbol{h}\big) \tag{2.3.2}$$

$$Var(q_\tau(\boldsymbol{x})|\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n) \to (\sigma_\epsilon \frac{1-2\tau}{\tau(1-\tau)}\boldsymbol{t}(\boldsymbol{x})^T V^{-1}1)^2 + v(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{t}(\boldsymbol{x})^T V^{-1}\boldsymbol{t}(\boldsymbol{x}) \tag{2.3.3}$$

(ii) Quantile prediction model via (2.3.2) is a quantile interpolator as $n \to \infty$.

(iii) Quantile prediction variance for observed inputs is $(\frac{\sigma_\epsilon(1-2\tau)}{\tau(1-\tau)})^2$ as $n \to \infty$.

The posterior inference in Theorem 2.3.1 depends on both prior information and information on the data like any Bayesian analysis. Posterior mean in (2.3.2) for quantile prediction of $q_\tau(\boldsymbol{x})$ has similar form with mean GP model for computer experiment and posterior variance in (2.3.3) has additional first term compared with variance estimation in mean GP model.

When $\tau = 0.5$, (2.3.2) indicates that median prediction coincides with mean $E(\boldsymbol{y})$ when $\boldsymbol{x}$ condicide with $\boldsymbol{x}_k$ as $n \to \infty$. This is true for any symmetric likelihood distribution. (2.3.3) indicates that $\text{Var}(q_\tau(\boldsymbol{x})|\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n) \to v(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{t}(\boldsymbol{x})^T V^{-1} \boldsymbol{t}(\boldsymbol{x})$. When $\boldsymbol{x}$ coincide with $\boldsymbol{x}_k$, $\text{Var}(q_\tau(\boldsymbol{x}_k)|\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n) \to 0$ for $k = 1, \cdots, M$, which is same with GP mean regression model.

In the next part, properties of coefficient estimations based on quantile prediction $q_\tau(\boldsymbol{x_1}), \cdots, q_\tau(\boldsymbol{x_M})$ are given. A direct result from Theorem 2.3.1 for one covariate case is introduced below.

Corollary 2.3.1. *Assume (2.2.6) and (2.2.7) is satisfied and* $p = 1$,

denote $\boldsymbol{\beta}(\tau) = (\beta(\tau, x_1), \cdots, \beta(\tau, x_M))^T$

$\widehat{\boldsymbol{\beta}(\tau)} = E(\boldsymbol{\beta}(\tau)|\boldsymbol{y_1}, \cdots, \boldsymbol{y_n}) \xrightarrow{p} \boldsymbol{\beta}(\tau)$ *when* $n \to \infty$.

The proof is complete by noticing $q_\tau(x_k) = x_k \beta(\tau, x_k)$ and $\widehat{q_\tau(x_k)} \xrightarrow{p} q_\tau(x_k)$ as $n \to \infty$ for all $k = 1, \cdots, M$.

When there are $p > 1$ covariates, by model assumption (2.2.6) and (2.2.7), τ^{th} quantile of y_k can be represented using: $q_\tau(\boldsymbol{x_k}) = \sum_{j=1}^{p} \beta_j(\tau, x_{kj}) x_{kj}$ and $y_k|q_\tau(\boldsymbol{x_k}) \sim$

$\text{ALP}(q_\tau(\boldsymbol{x_k}), \sigma_\epsilon^2, \tau)$ are independent for $k = 1, \cdots, M$. So

$$\boldsymbol{q_\tau} := \begin{bmatrix} q_\tau(\boldsymbol{x_1}) \\ \vdots \\ q_\tau(\boldsymbol{x_M}) \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^p \beta_j(\tau, x_{1j}) x_{1j} \\ \vdots \\ \sum_{j=1}^p \beta_j(\tau, x_{Mj}) x_{Mj} \end{bmatrix} \tag{2.3.4}$$

Without loss of generality, assume covariate x_j has l levels for $j = 1, \cdots, p$ in the design, so the number of distinct $\{x_{kj}\}_{k=1,\cdots,M}$ and $\{\beta_j(\tau, x_{kj})\}_{k=1,\cdots,M}$ is l for j^{th} covariate. We only consider the case when $M \geq lp - p$ for coefficients estimation.

Let $\boldsymbol{\beta_j}(\tau) = (\beta_j(\tau, x_{j(1)}), \beta_j(\tau, x_{j(2)}), \cdots, \beta_j(\tau, x_{j(l)}))^T$ denote the unknown coefficient vector for j^{th} covariate, where $x_{j(1)}, \cdots, x_{j(l)}$ denotes the l distinct levels of covariate x_j. Let $\boldsymbol{\beta}(\tau) = (\boldsymbol{\beta_1}(\tau), \cdots, \boldsymbol{\beta_p}(\tau))^T$ denote length lp coefficients vector for all covariates.

By (2.3.4), $\boldsymbol{q_\tau} = A_1 D \boldsymbol{\beta}(\tau)$, where $D = \text{diag}(x_{1(1)}, \cdots, x_{1(l)}, \cdots, x_{p(1)}, \cdots, x_{p(l)})$ and A_1 is $M \times lp$ transformation matrix with 0 and 1 elements and its row sums are all p. Since $\text{rank}(A_1) < lp$, so $\boldsymbol{\beta}(\tau)$ cannot be obtained uniquely from $\boldsymbol{q_\tau}$.

Additional prior assumption is needed to solve identifiability issue for coefficients estimation. Assume

$$1/l \sum_{r=1}^{l} \beta_j(\tau, x_{j(r)}) = b_j(\tau), \tag{2.3.5}$$

where $b_j(\tau)$ is mean of $\beta_j(\tau, x_j)$ known

Theorem 2.3.2. *Assume conditions (2.2.6), (2.2.7) and (2.3.5) are true and for given τ, as $n \to \infty$, when $M \geq lp - p$,*

$$\widehat{\boldsymbol{\beta}(\tau)} = E(\boldsymbol{\beta}(\tau) | \boldsymbol{y_1}, \cdots, \boldsymbol{y_n}) \xrightarrow{p} \boldsymbol{\beta}(\tau),$$

2.4 Simulation and Real data Examples

2.4.1 Simulation with one covariate

In this section, we simulate $y_{ik} \sim N(5\sin(x_k - 1), 1)$, where x_k $k = 1, \cdots, 10$ are equally spaced between 1 and 6, and $i = 1, \cdots, 30$. 25^{th}, 50^{th} and 75^{th} quantile predictions and confidence bands results are given in Figure 2.1. Grey circles are simulated response y. Black curve is underlying quantile function. Red dots give quantile prediction for observed inputs. Blue triangles give quantile prediction for unobserved inputs. Quantile prediction for both observed and unobserved inputs x are close to true quantile in Figure 2.1. The confidence band is smallest for median prediction.

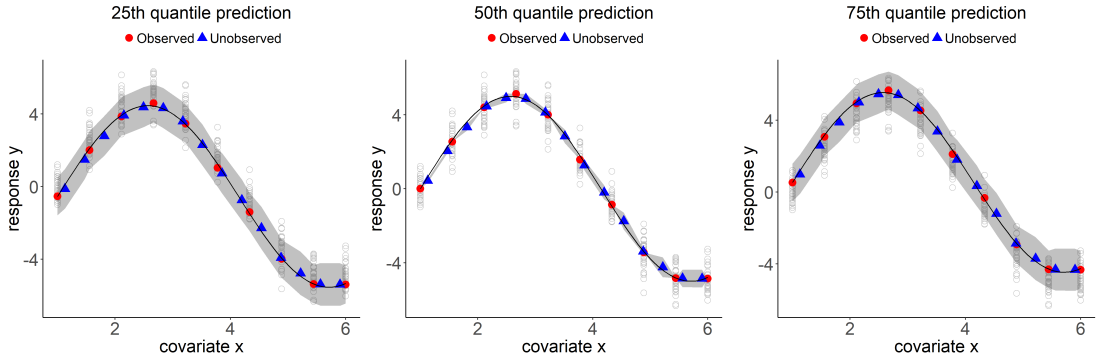


Figure 2.1: Quantile prediction with confidence band for $M = 10$ in one covariate model

When there are only 6 observed inputs in the data i.e. $k = 1, \cdots, 6$. Quantile prediction for observed inputs x are still close to true quantile. The confidence band of quantile prediction for unobserved input is larger than $M = 10$ case.

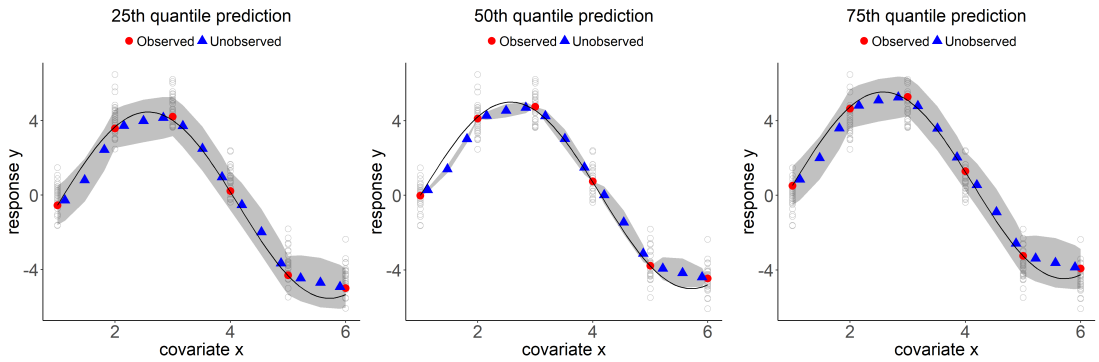


Figure 2.2: Quantile prediction with confidence band for $M = 6$ in one covariate model

The quantile prediction and standard deviation for $M = 10$ and $M = 6$ are summarized in Table 2.1, 2.2 and 2.3. Standard deviation increases with value of x and decreases when number of observed inputs M increases. Standard deviation is smallest when $\tau = 0.5$. This confirm observations in section 2.3.

Table 2.1: Simulation study: 25th Quantile prediction for unobserved inputs in one covariate model

x	True 25^{th} quantile	$M = 10$, $\tau = 0.25$	$M = 6$, $\tau = 0.25$
1.13	0.109	-0.229(0.545)	-0.327(0.551)
1.471	1.728	1.571(0.543)	0.787(0.575)
1.811	3.087	2.874(0.548)	2.466(0.562)
2.152	4.029	3.914(0.538)	3.748(0.55)
2.493	4.445	4.421(0.55)	3.911(0.572)
2.834	4.289	4.363(0.552)	3.964(0.56)
3.174	3.577	3.524(0.542)	3.495(0.563)
3.515	2.392	2.453(0.566)	2.392(0.593)
3.856	0.87	1.092(0.552)	1.01(0.569)
4.196	-0.814	-0.641(0.567)	-0.399(0.585)
4.537	-2.466	-2.34(0.582)	-1.845(0.624)
4.878	-3.897	-4.049(0.538)	-3.502(0.58)
5.219	-4.942	-4.831(0.598)	-4.433(0.615)
5.559	-5.481	-5.342(0.584)	-4.896(0.662)
5.9	-5.452	-5.261(0.585)	-5.382(0.59)

Table 2.2: Simulation study: 50th Quantile prediction for unobserved inputs in one covariate model

x	True 50^{th} quantile	$M = 10$, $\tau = 0.5$	$M = 6$, $\tau = 0.5$
1.13	0.648	0.434(0.05)	0.355(0.054)
1.471	2.268	1.998(0.056)	1.552(0.104)
1.811	3.626	3.396(0.095)	3.256(0.1)
2.152	4.568	4.665(0.059)	4.522(0.109)
2.493	4.985	4.801(0.122)	4.636(0.176)
2.834	4.828	4.672(0.137)	4.594(0.149)
3.174	4.117	4.233(0.094)	4.017(0.17)
3.515	2.932	3.067(0.185)	2.804(0.248)
3.856	1.41	1.571(0.141)	1.278(0.192)
4.196	-0.274	0.022(0.191)	-0.173(0.236)
4.537	-1.927	-1.558(0.23)	-1.565(0.32)
4.878	-3.358	-3.201(0.072)	-3.164(0.226)
5.219	-4.403	-4.221(0.27)	-3.966(0.305)
5.559	-4.942	-4.919(0.237)	-4.247(0.39)
5.9	-4.912	-4.837(0.239)	-4.532(0.25)

Table 2.3: Simulation study: 75th Quantile prediction for unobserved inputs in one covariate model

x	True 75^{th} quantile	$M = 10$, $\tau = 0.75$	$M = 6$, $\tau = 0.75$
1.13	1.188	1.043(0.545)	0.884(0.551)
1.471	2.807	2.662(0.543)	2.035(0.575)
1.811	4.166	3.955(0.548)	3.598(0.562)
2.152	5.108	5.033(0.538)	4.818(0.55)
2.493	5.524	5.336(0.55)	5.132(0.572)
2.834	5.368	5.198(0.552)	5.327(0.56)
3.174	4.656	4.495(0.542)	4.91(0.563)
3.515	3.472	3.331(0.566)	3.747(0.593)
3.856	1.95	1.874(0.552)	2.257(0.569)
4.196	0.266	0.312(0.567)	0.772(0.585)
4.537	-1.387	-1.265(0.582)	-0.726(0.624)
4.878	-2.818	-2.896(0.538)	-2.462(0.58)
5.219	-3.863	-3.658(0.598)	-3.37(0.615)
5.559	-4.402	-4.21(0.584)	-3.746(0.662)
5.9	-4.373	-4.279(0.585)	-4.15(0.59)

2.4.2 Simulation with multiple covariates

In this section, we consider a model with 3 covarites in the model. Assume there are inputs x_1, x_2, x_3 and each covariate has $l = 8$ levels and factorial design is used in the simulation. There are in total $M = 6^3 = 216$ different settings. Coefficients are generated according to assumption (2.2.7) and (2.3.5). Then simulate $y_{ik} \sim \text{ALP}(\sum_{j=1}^{3} x_{kj}\beta_i(\tau, x_{kj}), \sigma_\epsilon = 1, \tau)$, where $k = 1, \cdots, 216$, $i = 1, \cdots, 50$ and $\tau = 0.25, 0.5, 0.75$. Figure 2.3 below gives 25^{th}, 50^{th} and 75^{th} quantile predictions results compared with true quantiles. From Figure 2.3, quantile predictions are very close to true quantiles.

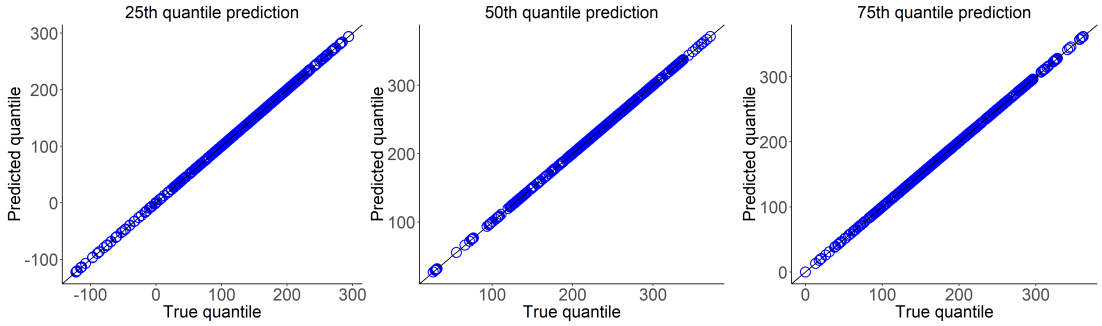


Figure 2.3: Quantile predictions versus true quantiles for multiple covariates model

Figure 2.4 below shows coefficients estimation results compared with true coefficients for $\tau = 0.25$, $\tau = 0.5$ and $\tau = 0.75$. From Figure 2.4, estimated coefficients are close to true coefficients.

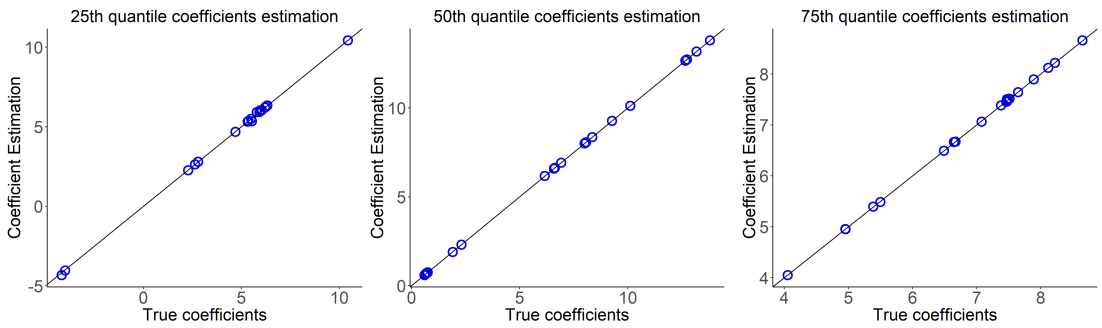


Figure 2.4: Coefficients estimation versus true coefficients for multiple covariates model

2.4.3 Real data example

Icesheet data is simulated based on the community ice sheet model (CISM)(Rutt et al., 2009; Price et al., 2011; Higdon et al., 2013). There are two inputs in the CISM model. x_1 is a constant in the GlenNye flow law, controlling the deformation of the ice sheet and x_2 controls the heat conductivity in the ice sheet. The thickness of the present ice sheet could be measured to inform about the model inputs x_1 and x_2. Icesheet thickness was collected in a 3D field in space and time. We use all thickness measure in the 27×32 grid space at time index 10 as response and used an ensemble of $M = 10$ model runs at different (x_1, x_2) input settings to generate non-linear quantile prediction as well as coefficients estimation. There are in total 20 model runs in Higdon et al. (2013) paper. 10 out of 20 model runs are selected to satisfy the condition that $M > l_1 + l_2 - 2$, where l_i is # of distinct levels of x_i. Among the 10 model runs, $l_1 = 6$ and $l_2 = 5$.

Assume $\theta_1(x_1)$ and $\theta_2(x_2)$ are corresponding coefficients of x_1 and x_2. So the quantile prediction can be written as $q_\tau(\boldsymbol{x}) = \theta_1(x_1)x_1 + \theta_2(x_2)x_2$. After standardization on the input design matrix, estimate for for $\theta_1(x_1)$ and $\theta_2(x_2)$ over different quantile levels is given in Figure 2.5 and Figure 2.6.

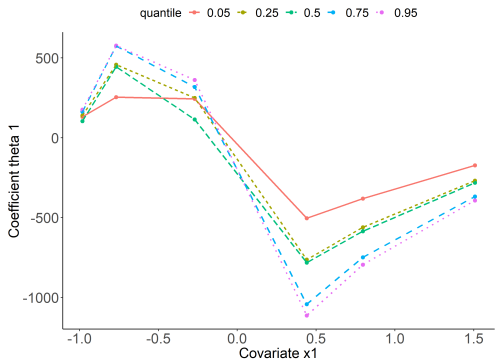


Figure 2.5: θ_1 estimates for different quantiles in ice sheet data

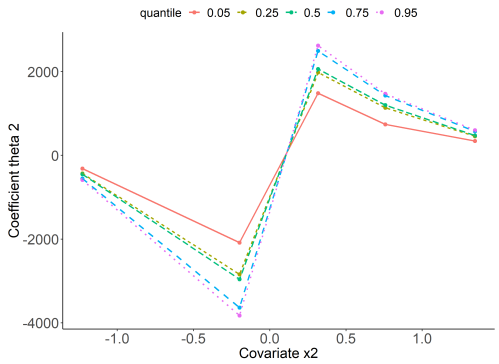


Figure 2.6: θ_2 estimate for different quantiles in ice sheet data

From Figure 2.5 and Figure 2.6, both x_1 and x_2 are more sensitive to higher quantile levels than lower quantile levels. So x_1 and x_2 may be main contributing factor for higher quantile levels. In addition, the sign of coefficient estimates indicates the slope of coefficient of corresponding input. When the sign is negative, coefficient effect is decreasing with the corresponding covariate.

2.5 Discussion

A new Bayesian quantile analysis model for computer experiments is introduced. We show that with Gaussian process prior assumption on the coefficients and asymmetric Laplace distributed error distribution, the quantile predictions for observed inputs agree on true quantiles asymptotically. In addition, with some constraints, asymptotic consistency of coefficients estimation is derived. What's more, there is no issue of quantile curves crossing each other in proposed model for quantile prediction of observed inputs.

2.6 Technical Proofs

Note that Theorem 2.2.1 is a special case of Theorem 2.3.1 so only proof of Theorem 2.3.1 is given. A Lemma that is useful to prove Theorem 2.3.1 is introduced first.

Lemma 2. *Given $nm \times nm$ block matrix $\boldsymbol{C}$ with $n \times n$ blocks. Assume all diagonal blocks are $aI_m + B$ and all other blocks are B, where a is a scalar, B is $m \times m$ matrix and I_m is $m \times m$ identity matrix, then $1^T \boldsymbol{C}^{-1} = (aI_m + nB)^{-1} 1^T$, where 1 is all-ones vector with length n.*

Proof of Lemma 2. Denote $\boldsymbol{I} = \text{diag}(I_m, \cdots, I_m)$ is block diagonal matrix with $n \times n$ blocks and 1 is all-ones vector of size n, so $\boldsymbol{C} = a\boldsymbol{I} + 1B1^T$. By Woodbury matrix identity,

$$\boldsymbol{C}^{-1} = (a\boldsymbol{I} + 1B1^T)^{-1} = a^{-1}\boldsymbol{I} - a^{-1}1\left(B^{-1} + a^{-1}1^T\boldsymbol{I}1\right)^{-1}1^T a^{-1}.$$

So

$$\boldsymbol{C}^{-1} = \begin{bmatrix} \frac{1}{a}I_m - \frac{1}{a^2}E & -\frac{1}{a^2}E & \cdots & -\frac{1}{a^2}E \\ -\frac{1}{a^2}E & \frac{1}{a}I_m - \frac{1}{a^2}E & \cdots & -\frac{1}{a}E \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{a^2}E & -\frac{1}{a^2}E & \cdots & \frac{1}{a}I_m - \frac{1}{a^2}E \end{bmatrix},$$

where $E = (B^{-1} + a^{-1}nI_m)^{-1}$.

Since all column sums are the same, $1^T\boldsymbol{C}^{-1}\boldsymbol{e}_i = \frac{1}{a}I_m - \frac{n}{a^2}E$ for $i = 1, \cdots, n$, where $\boldsymbol{e}_i$ is $n \times 1$ vector with i^{th} entry 1 and 0 elsewhere. It remains to show that

$\frac{1}{a}I_m - \frac{n}{a^2}E = (aI_m + nB)^{-1}$.

$$\begin{aligned}
&\frac{1}{a}I_m - \frac{n}{a^2}E \\
=&\frac{1}{a}I_m - \frac{n}{a^2}(B^{-1} + \frac{n}{a}I_m)^{-1} \\
=&\frac{1}{a}I_m - \frac{n}{a^2}\frac{a}{n}(I_m + \frac{a}{n}B^{-1})^{-1} \\
=&\frac{1}{a}I_m - \frac{1}{a}\big(I_m - (I_m + \frac{n}{a}B)^{-1}\big) \\
=&(aI_m + nB)^{-1}
\end{aligned}$$

The proof is complete. □

Proof of Theorem 2.3.1. By GP prior (2.2.7), distribution of $q_\tau(\boldsymbol{x})$ is normal with mean and variance below.

$$E(q_\tau(\boldsymbol{x})) = h(\boldsymbol{x})$$

$$\text{var}(q_\tau(\boldsymbol{x})) = v(\boldsymbol{x}, \boldsymbol{x})$$

Let $\boldsymbol{y}_i' = \boldsymbol{y}_i - \sigma_\epsilon \frac{1-2\tau}{\tau(1-\tau)}\xi 1$. By (2.3.1), $\boldsymbol{y}_i'|\xi \sim N(\boldsymbol{h}, \frac{2\sigma_\epsilon^2\xi}{\tau(1-\tau)}I)$, where 1 is all-ones vector with length M and I is $M \times M$ identity matrix.

Let $\boldsymbol{\mu} = (\boldsymbol{h}, \cdots, \boldsymbol{h})^T$ and

$$\boldsymbol{\Sigma} = \begin{bmatrix} \frac{2\sigma_\epsilon^2\xi}{\tau(1-\tau)}I & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \frac{2\sigma_\epsilon^2\xi}{\tau(1-\tau)}I \end{bmatrix} + \begin{bmatrix} V & \cdots & V \\ \vdots & \vdots & \vdots \\ V & \cdots & V \end{bmatrix},$$

So $(\boldsymbol{y}_1', \cdots, \boldsymbol{y}_n')^T|\xi \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.Denote $\boldsymbol{y}^* = (\boldsymbol{y_1'}, \cdots, \boldsymbol{y_n'})^T$, $\boldsymbol{g}(\boldsymbol{x}) = (\boldsymbol{t}(\boldsymbol{x})^T, \cdots, \boldsymbol{t}(\boldsymbol{x})^T)^T$ with $\boldsymbol{t}(\boldsymbol{x})^T$ repeating n times, the joint distribution of $\boldsymbol{y}^*$ and $q_\tau(\boldsymbol{x})$ conditioned on ξ is:

$$\begin{bmatrix} \boldsymbol{y}^* \\ q_\tau(\boldsymbol{x}) \end{bmatrix} |\xi \sim N\Big(\begin{bmatrix} \boldsymbol{\mu} \\ h(\boldsymbol{x}) \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{g}(\boldsymbol{x}) \\ \boldsymbol{g}(\boldsymbol{x})^T & v(\boldsymbol{x}, \boldsymbol{x}) \end{bmatrix} \Big)$$

Therefore, by formula of conditional distribution of multivariate Gaussian distribution,

$$q_\tau(\boldsymbol{x})|\boldsymbol{y}^*,\xi \sim N\big(h(\boldsymbol{x}) + g(\boldsymbol{x})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}^* - \boldsymbol{\mu}), v(\boldsymbol{x},\boldsymbol{x}) - g(\boldsymbol{x})^T\boldsymbol{\Sigma}^{-1}g(\boldsymbol{x})\big) \tag{2.6.1}$$

Since $\boldsymbol{\Sigma}$ is block matrix with $n \times n$ blocks and its all diagonal blocks are $\frac{2\sigma_\epsilon^2\xi}{\tau(1-\tau)}I + V$ and V elsewhere, by Lemma 2, $1^T\boldsymbol{\Sigma}^{-1} = (\frac{2\sigma_\epsilon^2\xi}{\tau(1-\tau)}I + nV)^{-1}1^T$. So $\boldsymbol{g}(\boldsymbol{x})^T\boldsymbol{\Sigma}^{-1} = \boldsymbol{t}(\boldsymbol{x})^T1^T\boldsymbol{\Sigma}^{-1} = \boldsymbol{t}(\boldsymbol{x})^T(\frac{2\sigma_\epsilon^2\xi}{\tau(1-\tau)}I+nV)^{-1}1^T$. Mean and variance in (2.6.1) can be rewritten as

$$E(q_\tau(\boldsymbol{x})|\boldsymbol{y}^*,\xi) = h(\boldsymbol{x}) + \boldsymbol{t}(\boldsymbol{x})^T(\frac{2\sigma_\epsilon^2\xi}{\tau(1-\tau)}I + nV)^{-1}\sum_{i=1}^n(\boldsymbol{y}_i^* - \boldsymbol{h})$$

$$\mathrm{Var}(q_\tau(\boldsymbol{x})|\boldsymbol{y}^*,\xi) = v(\boldsymbol{x},\boldsymbol{x}) - \boldsymbol{t}(\boldsymbol{x})^T(\frac{2\sigma_\epsilon^2\xi}{\tau(1-\tau)}I + nV)^{-1}n\boldsymbol{t}(\boldsymbol{x})$$

So

$$E(q_\tau(\boldsymbol{x})|\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n,\xi) = h(\boldsymbol{x}) + \boldsymbol{t}(\boldsymbol{x})^T(\frac{2\sigma_\epsilon^2\xi}{n\tau(1-\tau)}I + V)^{-1}\frac{1}{n}\sum_{i=1}^n(\boldsymbol{y}_i - \sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}\xi 1 - \boldsymbol{h})$$

$$\mathrm{Var}(q_\tau(\boldsymbol{x})|\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n,\xi) = v(\boldsymbol{x},\boldsymbol{x}) - \boldsymbol{t}(\boldsymbol{x})^T(\frac{2\sigma_\epsilon^2\xi}{n\tau(1-\tau)}I + V)^{-1}\boldsymbol{t}(\boldsymbol{x})$$

Therefore, posterior mean and posterior variance of $q_\tau(\boldsymbol{x})|\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n$ is

$$\begin{aligned}
&E(q_\tau(\boldsymbol{x})|\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n)\\
=&E[E(q_\tau(\boldsymbol{x})|\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n,\xi)|\xi]\\
=&E[h(\boldsymbol{x}) + \boldsymbol{t}(\boldsymbol{x})^T(\frac{2\sigma_\epsilon^2\xi}{n\tau(1-\tau)}I + V)^{-1}\frac{1}{n}\sum_{i=1}^n(\boldsymbol{y}_i - \sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}\xi 1 - \boldsymbol{h})|\xi]\\
\to&E[h(\boldsymbol{x}) + \boldsymbol{t}(\boldsymbol{x})^TV^{-1}(\frac{1}{n}\sum_{i=1}^n\boldsymbol{y}_i - \sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}\xi 1 - \boldsymbol{h})|\xi]\\
=&h(\boldsymbol{x}) + \boldsymbol{t}(\boldsymbol{x})^TV^{-1}(\frac{1}{n}\sum_{i=1}^n\boldsymbol{y}_i - \sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}1 - \boldsymbol{h})
\end{aligned} \tag{2.6.2}$$

$$
\begin{aligned}
&\mathrm{Var}(q_\tau(\boldsymbol{x})|\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n)\\
=&\mathrm{Var}[E(q_\tau(\boldsymbol{x})|\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n,\xi)|\xi]+E[\mathrm{Var}(q_\tau(\boldsymbol{x})|\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n,\xi)|\xi]\\
=&\mathrm{Var}[h(\boldsymbol{x})+\boldsymbol{t}(\boldsymbol{x})^T(\frac{2\sigma_\epsilon^2\xi}{n\tau(1-\tau)}I+V)^{-1}\frac{1}{n}\sum_{i=1}^n(\boldsymbol{y}_i-\sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}\xi 1-\boldsymbol{h})|\xi]\\
&+E[v(\boldsymbol{x},\boldsymbol{x})-\boldsymbol{t}(\boldsymbol{x})^T(\frac{2\sigma_\epsilon^2\xi}{n\tau(1-\tau)}I+V)^{-1}\boldsymbol{t}(\boldsymbol{x})|\xi]\\
\to&\mathrm{Var}[\boldsymbol{t}(\boldsymbol{x})^TV^{-1}(\frac{1}{n}\sum_{i=1}^n\boldsymbol{y}_i-\sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}\xi 1-\boldsymbol{h})|\xi]+v(\boldsymbol{x},\boldsymbol{x})-\boldsymbol{t}(\boldsymbol{x})^TV^{-1}\boldsymbol{t}(\boldsymbol{x})\\
=&(\sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}\boldsymbol{t}(\boldsymbol{x})^TV^{-1}1)^2+v(\boldsymbol{x},\boldsymbol{x})-\boldsymbol{t}(\boldsymbol{x})^TV^{-1}\boldsymbol{t}(\boldsymbol{x})
\end{aligned}
$$

(ii) According to (2.6.2), when $\boldsymbol{x}$ coincide with some $\boldsymbol{x}_k$ for $k=1,\cdots,M$, the quantile prediction $q_\tau(\boldsymbol{x}_k)$ with respect to $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n$ is given in the following.

$$
\begin{aligned}
&h(\boldsymbol{x}_k)+\boldsymbol{t}(\boldsymbol{x}_k)^TV^{-1}(\frac{1}{n}\sum_{i=1}^n\boldsymbol{y}_i-\sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}1-\boldsymbol{h})\\
=&h(\boldsymbol{x}_k)+\frac{1}{n}\sum_{i=1}^n y_{ik}-\sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}-h(\boldsymbol{x}_k)\\
=&\frac{1}{n}\sum_{i=1}^n y_{ik}-\sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}\\
\xrightarrow{p}&E(y_k)-\sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)},
\end{aligned}
$$

as $n\to\infty$. Also $q_\tau(\boldsymbol{x}_k)=E(y_k)-\sigma_\epsilon\frac{1-2\tau}{\tau(1-\tau)}$. Therefore quantile prediction with GP prior on coefficients interpolate true quantile asymptotically.

□

Proof of Theorem 2.3.2. Define $\boldsymbol{\beta'}=(\boldsymbol{\beta'_1},\cdots,\boldsymbol{\beta'_p})$,where $\boldsymbol{\beta'_j}=(\beta_j(\tau,x_{j(1)}),\cdots,\beta_j(\tau,x_{j(l-1)}))$ is corresponding to coefficients for jth covariate at all $l-1$ levels. By assumption, $\boldsymbol{\beta'_j}$

is distributed as below.

$$\boldsymbol{\beta'_j} = \begin{bmatrix} \beta_j(\tau, x_{j(1)}) \\ \vdots \\ \beta_j(\tau, x_{j(l-1)}) \end{bmatrix} \Big| (\theta_j, \sigma_j^2) \sim N\big(b_j 1_{l-1}, \Sigma'_j\big), \tag{2.6.3}$$

where $j = 1, \cdots, p$, 1_{l-1} is length $l-1$ all-ones vector and Σ'_j is $(l-1)\times(l-1)$ covariance matrix with same structure as $R(x_{js}, x_{js'})$ defined in (2.2.7).

Note that

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1(\tau, x_{1(1)}) \\ \vdots \\ \beta_1(\tau, x_{1(l)}) \\ \vdots \\ \beta_p(\tau, x_{p(1)}) \\ \vdots \\ \beta_p(\tau, x_{p(l)}) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ lb_1 \\ \vdots \\ 0 \\ \vdots \\ lb_p \end{bmatrix} + \boldsymbol{\beta'} = \boldsymbol{b'} + Q\boldsymbol{\beta'},$$

where $Q = diag(I'_1, \cdots, I'_p)$ is a $(lp)\times(l-1)p$ matrix, $I'_1 = \cdots = I'_p = \begin{bmatrix} I_{l-1} \\ (-1, \cdots, -1) \end{bmatrix}_{l\times(l-1)}$ and I_{l-1} is $(l-1)\times(l-1)$ identity matrix.

Therefore,

$$\begin{aligned} \boldsymbol{q}(\tau) &= A_1 D\boldsymbol{\beta} \\ &= A_1 D\boldsymbol{b'} + A_1 DQ\boldsymbol{\beta'} = A_1 D\boldsymbol{b'} + A_2\boldsymbol{\beta'} \end{aligned}$$

$A_2 = A_1 DQ$ is $M \times p(l-1)$ matrix with rank $p(l-1)$.

Given the above affine relation and using Theorem 2.3.1, we can derive consistency

for $\hat{\boldsymbol{\beta}}'$ as below.

$$\begin{aligned}\hat{\boldsymbol{\beta}}' = E(\boldsymbol{\beta}'|\boldsymbol{y_1}, \cdots, \boldsymbol{y_n}) &= E(H(\boldsymbol{q}(\tau) - A_1 D\boldsymbol{b}')|\boldsymbol{y_1}, \cdots, \boldsymbol{y_n}) \\ &= H(E(\boldsymbol{q}(\tau)|\boldsymbol{y_1}, \cdots, \boldsymbol{y_n}) - AD\boldsymbol{b}') \\ &\xrightarrow{p} H(\boldsymbol{q}(\tau) - A_1 D\boldsymbol{b}') \\ &= \boldsymbol{\beta}'\end{aligned}$$

where $H = (A_2^T A_2)^{-1} A_2^T$ is a full rank $(l-1)p \times (l-1)p$ matrix. Since $\boldsymbol{\beta} = \boldsymbol{b}' + Q\boldsymbol{\beta}'$, clearly $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$ as $n \to \infty$ □

Bibliography

Afsarinejad, K. and Hedayat, A. (2002). Repeated measurements designs for a model with self and simple mixed carryover effects. *Journal of Statistical Planning and Inference*, 106(1):449 – 459.

Bose, M. and Dey, A. (2009). *Optimal Crossover Designs*. World Scientific.

Cheng, C.-S. and Wu, C.-F. (1980). Balanced repeated measurements designs. *Ann. Statist.*, 8(6):1272–1283.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.*, 24(4):586–602.

Ehrenfeld, S. (1955). On the efficiency of experimental designs. *Ann. Math. Statist.*, 26(2):247–255.

Fang, K.-T., Li, R., and Sudjianto, A. (2005). *Design and modeling for computer experiments*. Chapman and Hall/CRC.

G. H & Littlewood, J. E & Polya, G. (1952). *Inequalities (2nd ed)*. Cambridge University Press, West Nyack.

Hedayat, A. and Afsarinejad, K. (1975). Repeated measurements designs, I. *In A survey of statistical design and linear models (Proc. Internat. Sympos., Colorado State Univ., Ft. Collins, Colo., 1973)*, pages 229–242.

Hedayat, A. and Afsarinejad, K. (1978). Repeated measurements designs, II. *Ann. Statist.*, 6(3):619–628.

Higdon, D., Gattiker, J., Lawrence, E., Jackson, C., Tobis, M., Pratola, M., Habib, S., Heitmann, K., and Price, S. (2013). Computer model calibration using the ensemble kalman filter. *Technometrics*, 55(4):488–500.

Huang, J., Zarnitsyna, V. I., Liu, B., Edwards, L. J., Jiang, N., Evavold, B. D., and Zhu, C. (2010). The kinetics of two-dimensional tcr and pmhc interactions determine t-cell responsiveness. *Nature*, 464(7290):932–936.

Jones, B. and Kenward, M. G. (2015). *Design and analysis of cross-over trials*. New York: Chapman and Hall/CRC.

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society. Series B (Methodological)*, 21(2):272–319.

Koenker, R. (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of economic perspectives*, 15(4):143–156.

Koenker, R., NG, P., and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, 81(4):673–680.

Kozubowski, T. J., Podgórski, K., and Rychlik, I. (2013). Multivariate generalized

laplace distribution and related random fields. *Journal of Multivariate Analysis*, 113:59–72.

Kunert, J. (1983). Optimal design and refinement of the linera model with applications to repeated measurements designs. *Ann. Statist.*, 11(1):247–257.

Kunert, J. and Martin, R. (2000). On the determination of optimal designs for an interference model. *The Annals of Statistics*, 28(6):1728–1742.

Kunert, J. and Stufken, J. (2002). Optimal crossover designs in a model with self and mixed carryover effects. *Journal of the American Statistical Association*, 97(459):898–906.

Lui, K.-J. (2016). *Crossover designs : testing, estimation, and sample size*. Statistics in practice. John Wiley & Sons.

Lum, K. and Gelfand, A. E. (2012). Spatial quantile multiple regression using the asymmetric laplace process. *Bayesian Analysis*, 7(2):235–258.

Padgett, L. (2014). *Repeated measurements and cross-over designs*. John Wiley & Sons.

Price, S. F., Payne, A. J., Howat, I. M., and Smith, B. E. (2011). Committed sea-level rise for the next century from greenland ice sheet dynamics during the past decade. *Proceedings of the National Academy of Sciences*, 108(22):8978–8983.

Roger Koenker, Victor Chernozhukov, X. H. L. P., editor (2017). *Handbook of Quantile Regression*. CRC Press.

Rutt, I. C., Hagdorn, M., Hulton, N., and Payne, A. (2009). The glimmer community ice sheet model. *Journal of Geophysical Research: Earth Surface*, 114(F2).

Santner, T. J., Williams, B. J., Notz, W., and Williams, B. J. (2003). *The design and analysis of computer experiments*, volume 1. Springer.

Sen, M. and Mukerjee, R. (1987). Optimal repeated measurements designs under interaction. *Journal of Statistical Planning and Inference*, 17:81 – 91.

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, pages 595–620.

Stufken, J. (1991). Some families of optimal and efficient repeated measurements designs. *Journal of Statistical Planning and Inference*, 27(1):75 – 83.

Stufken, J. (1996). 3 optimal crossover designs. In *Design and Analysis of Experiments*, volume 13 of *Handbook of Statistics*, pages 63 – 90.

Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, 93(441):228–237.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54(4):437–447.

Zarnitsyna, V. I., Huang, J., Zhang, F., Chien, Y.-H., Leckband, D., and Zhu, C. (2007). Memory in receptor-ligand-mediated cell adhesion. *Proceedings of the National Academy of Science, U.S.A.*, 104(46):18037–18042.