



**HAL**  
open science

## Sequential Design for Microarray Experiments

Gilles Durrieu, Laurent Briollais

► **To cite this version:**

Gilles Durrieu, Laurent Briollais. Sequential Design for Microarray Experiments. Journal of the American Statistical Association, 2009, 104 (486), pp.650-660. 10.1198/jasa.2009.0135 . hal-00905356

**HAL Id: hal-00905356**

**<https://hal.science/hal-00905356v1>**

Submitted on 18 Nov 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sequential Design for Microarray Experiments

Gilles DURRIEU, and Laurent BRIOLLAIS\*

\***Author Footnote:** Gilles Durrieu is Associate Professor of Statistics at the University of Bordeaux 1, 33000 Bordeaux, France (email: g.durrieu@epoc.u-bordeaux1.fr); and Laurent Briollais is Research Scientist, Samuel Lunenfeld Research Institute, Toronto, ON M5G 1X5, Canada (email: laurent@mshri.on.ca). This work was supported by the Discovery Grant 93270-04 from the Natural Sciences and Engineering Research Council of Canada and the CNRS.

**Abstract.** A critical aspect in the design of microarray studies is the determination of the sample size necessary to declare genes differentially expressed across different experimental conditions. In this paper, we propose a sequential approach where the decision to stop the experiment depends on the accumulated microarray data. The study could stop whenever sufficient data have been accumulated to identify gene expression changes across several experimental conditions. The gene expression response is modeled by a robust linear regression model. We then construct a sequential confidence interval for the intercept of this model, which represents the median gene expression at a given experimental condition. We derive the stopping rule of the experiment for both continuous and discrete sequential approaches and give the asymptotic properties of the stopping variable. We demonstrate the desirable properties of our sequential approach, both theoretically and numerically. In our application to a study of hormone responsive breast cancer cell lines, we estimated the stopping variable for the sample size determination to be smaller than the actual sample size available to conduct the experiment. This means that we can obtain an accurate assessment of differential gene expression without compromising the cost and size of the study. Altogether, we anticipate that this approach could have an important contribution to microarray studies by improving the usual experimental designs and methods of analysis.

KEY WORDS: Microarray Study; Sample Size; Sequential Design; Robust Regression; Dose-response.

# 1 INTRODUCTION

Microarray chips and other high-throughput technologies have changed radically the nature of genomics assays by allowing the simultaneous screening of thousands of genes in a single experiment. Statistical inference about change patterns in microarrays has led to the competing applications of numerous statistical approaches and computing algorithms for several problems: class comparison, class discovery and class prediction (Simon 2003). A particular aspect of these experiments, which has received less attention, is the planning of the experimental design that should enable experimenters to have efficient and valid inference about expression profiling. Particular designs of interest for class comparison problems include two- and multi-class experiments, where the classes are the experimental conditions, for example different types of tumor tissue, stages of tumor progression, doses of an exposure, or time-points. There have been a number of papers discussing the general issues related to experimental designs (for reviews, see Churchill 2002; McShane, Shih and Michalowska 2003), the paradigm being to use fixed-sample size plans. The theory of experimental design can be used to optimize the information gained from an experiment (Kerr and Churchill 2001a,b; Glonek and Solomon 2004) and to determine the number of arrays required to conduct the experiment. However, this leads generally to very crude sample size estimates, whose determination depends on quantities such as the variability of gene expression across the different experimental conditions and the false discovery rate (FDR) (Benjamini and Hochberg (1995), that are generally unknown before the study (Pan, Win and Le 2002; Zien et al. 2003; Gadbury et al. 2004; Pawitan et al. 2005; Muller et al. 2004; Dobbin and Simon

2005). Therefore, there is no guarantee that the study will have sufficient efficiency to detect the expression profiles of interest. To address this problem, Warnes and Liu (2005) developed a procedure to estimate sample size that uses an estimate of the standard deviation for each gene based on control samples from existing studies. Tibshirani (2006) proposed a method for assessing sample sizes based on a permutation-based analysis of pilot data, which avoids strong parametric assumptions and allows prior information about the required quantities. The authors emphasized a two-stage approach but it remains unclear how the data from the pilot and main studies should be analyzed.

A more formal multistage strategy that does not rely on the distribution of primary measurements from other studies or from a pilot study is the sequential approach. In clinical trials, sequential designs are well-known to offer numerous ethical and economic advantages due to the possibility of early stopping either for futility or obvious advantage of a treatment (Jennison and Turnbull 2000). Indeed, the sample size which is unknown at the start of the study period is determined subsequently in part by the nature of the sequentially accumulating data. Such approaches could be particularly beneficial in the design of efficient microarray experiments for several reasons. First, many biologists use sequential designs implicitly by collecting an initial sample (a pilot study for example) and by performing a first analysis. In some experiments, additional samples will be collected and further analyses will be carried out, but without proper adjustment for the multiple analyses. Sequential approaches are also known to require less observations than fixed-sample size approaches to reach the same conclusion. This will be very advantageous for microarray studies both because of their cost (DNA chips retail for about \$1,000 each) and the difficulties to obtaining

sample material. Finally, determining an appropriate sample size to conduct a microarray experiment is often impossible and sequential approaches give a more rational solution by accumulating information to provide an estimate of the sample size, an estimate that can be given with its precision (because it is a random variable). Despite these obvious advantages, sequential designs have received very little attention in the field of microarray studies and more generally in the field of genomics.

To the best of our knowledge, only two papers discussed their application in the context of microarray studies. The first one introduced a sequential procedure for classification problems (Fu, Dougherty, Mallick and Carroll 2005). It provides stopping criteria that ensure with a certain level of confidence that at stopping the misclassification probability of any future subject into a particular experimental group will be smaller than a pre-determined threshold. Although interesting, this approach does not apply directly to class comparison problems. The second paper deals with class comparison problems and proposed the construction of nonparametric prediction intervals to identify differentially expressed genes (Gibbons et al. 2005). A fixed number of control samples (normal tissues) is first obtained, followed by sequential collection of the experimental samples (tumor tissues), one by one or by groups. The control samples are then used to construct a predictive interval for the mean gene expression or other summary statistic of the experimental samples. The procedure stops when the probability that this summary statistic calculated for one specific gene (or a group of genes) in the experimental samples is outside the prediction limits with high probability. This approach is not truly sequential because the probability of being outside the predictive bounds is determined for different sizes of the experimental samples but does

not depend on the accumulated information on gene expression data. To our knowledge, neither of these approaches has been applied in real situations.

In this paper, we propose a formal theory for sequential microarray design in the context of class comparison problems. Section 2 describes the main novelties: that the decision to stop the experiment really depends on the data that are accumulating, and that the statistical properties of the stopping rule proposed are thoroughly studied both theoretically and by simulation. Our approach builds on the general linear regression framework, which is a very common strategy for microarray studies. Besides the sequential aspect, we use nonparametric regression to provide robustness against outliers and distribution-free inference, which is particularly advantageous in small sample size microarray problems. The principle of our approach is the construction of confidence interval for the model parameter, corresponding to a particular level of gene expression. The case of correlated observations is also discussed and treated by adjustment of the data. The definition of the stopping rule of the sequential procedure for a given precision is presented in Section 3. We develop both the case of continuous and discrete monitoring and the asymptotic properties of the stopping variable. In section 4, we analyze gene expression profiles at various doses of estrogen exposure in breast cancer cell-lines. Additional finite properties of our approach, in particular the use of continuous vs. discrete sampling and the potential bias due to batch effects are studied in Section 5 by simulation. Finally, section 6 concludes with a general discussion of our approach and its feasibility and applicability in practice.

## 2 METHODS

### 2.1 The Linear Model Framework

Linear models have become a classical statistical framework for the analysis of two channels cDNA arrays, especially with the development of the statistical package LIMMA in *R* (Smyth 2004). For example, Kerr et al. (2001) used linear models and the analysis of variance (ANOVA) to estimate expression differences and assess the variability of their estimates. They assumed a fixed effects model for the log gene intensity which accounts for the variability due to the array, dye effect, treatment (or experimental condition) and gene effect. Differentially expressed genes are those that exhibit significant treatment by gene interactions, while normalization is effected by including a dye term in the model. The random error of the model represents variations due to unknown sources and is typically assumed to be normally distributed. Other authors have proposed linear models on a gene by gene basis, with a separate error for each gene, but still including normalization as part of their linear model (Jin et al. 2001; Wolfinger et al. 2001). Their models also include a random effect for the arrays. These modelling approaches have raised several questions, in part because the inference about gene effects is performed jointly with the normalization process and the estimation of multiple other effects (Yang and Speed 2002). In contrast, other linear models have been proposed for the log-ratios of intensity from experiments in which normalization has been carried out separately for each slide, typically using a non-linear adjustment, which could not be captured in a linear model (Yang and Speed 2002). The only terms that are included in these models relate to mRNA samples and their treatment. The general form of

the model is:

$$\mathbf{Y}_n = \mathbf{X}_n \boldsymbol{\beta} + \boldsymbol{\varepsilon}_n \tag{1}$$

where for all  $n \geq 1$ ,  $\mathbf{Y} \equiv \mathbf{Y}_n = (Y_1, \dots, Y_n)'$  is the  $\log_2$  ratio of gene expression from all slides,  $\mathbf{X} \equiv \mathbf{X}_n \equiv \mathbf{X}_{n \times p}$  is a known  $n \times p$  design matrix of experimental conditions,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  is the vector of unknown parameters to be estimated, and  $\boldsymbol{\varepsilon} \equiv \boldsymbol{\varepsilon}_n = (\varepsilon_1, \dots, \varepsilon_n)'$  is the error term. The principle of this model, in general, is to compare the means of the log-gene expression distribution under different conditions (treatments). If  $\boldsymbol{\varepsilon}$  follows a normal  $N(0, \mathbf{I})$  distribution, then ordinary least squares provides the maximum likelihood estimator of  $\boldsymbol{\beta}$ . However, it is clear from microarray data that the normality assumptions are often violated for interesting genes. Unusual probes and outlying probe level measurements often occur to upset normality. To avoid the distributional assumptions and to protect against outlying measurements, we propose a robust inferential method for model (1) by using quantile regression.

An overview of robust statistics literature can be found in Hampel, Ronchetti, Rousseeuw and Stahel (1986), Jurečková and Sen (1996) and Huber (2003). Our approach is based on regression quantile estimators of  $\boldsymbol{\beta}$  (Koenker and D'Orey 1987; Dodge and Jurečková 1995). Instead of focusing on the changes in the mean gene expression, the quantile regression approach allows one to test whether there is a change in the  $\tau$ th quantile of  $y$  for any given  $\tau \in (0,1)$ . When the conditional distributions of  $y$  are non-Gaussian, the mean might not be the best summary, and a change in distributions may not be detected.

Inference for linear quantile regression models has become a subject of intense investigation in the past years. Any solution of the following minimization problem ( $\mathbf{x}'_i$  denotes the  $i^{\text{th}}$

row of the matrix  $\mathbf{X}$ )

$$\widehat{\boldsymbol{\beta}}(\theta) \equiv \widehat{\boldsymbol{\beta}}^n(\theta) = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^n \rho_{\theta}(Y_i - \mathbf{x}'_i \boldsymbol{\beta}) \quad (2)$$

is called a  $\theta$ -regression quantile with  $\rho_{\theta}(x) = x(\theta - \mathbb{I}(x < 0))$  where  $\mathbb{I}(\mathcal{P})$  takes the value 1 or 0 depending on whether the condition  $\mathcal{P}$  is satisfied or not. Here we used the  $L^1$ -norm estimator, also known as *LAD* (*least absolute deviation*) estimator, obtained by taking  $\theta = 1/2$ , chosen as an alternative to least squares estimators in the presence of heavy-tailed distribution. It is known that this estimator performs better in the presence of a heavy-tailed distribution. Following Jurečková (1984), under regularity conditions, we have as  $n \rightarrow \infty$ ,

$$\sqrt{n} \left( \widehat{\beta}_1^n(\theta) - \beta_1 \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \frac{q^2(\theta)}{4} \right), \quad (3)$$

where  $q(\theta) = 1/f(Q(\theta))$  is the quantile density function with  $Q(\theta)$  denoting the quantile function. The asymptotic variance of  $\widehat{\boldsymbol{\beta}}^n(\theta)$  is unknown because it depends on the distribution function  $f$  of the error term in the model, which itself is unknown. So, we estimate this asymptotic variance using the regression invariant and scale equivariant kernel-type estimator of  $q(\theta)$  in model (1) based on the regression quantiles. This estimator of  $q(\theta)$ , for  $0 < \theta < 1$ , is defined by

$$\widehat{Z}_n(\theta) = \frac{1}{\nu_n^2} \int_0^1 \widehat{\beta}_1^n(w) k \left( \frac{\theta - w}{\nu_n} \right) dw,$$

where  $(\nu_n)_{n \geq 1}$  is a sequence of properly chosen bandwidths and  $k$  denotes the kernel function. The quality of a density estimate is now widely recognized to be primarily determined by the choice of bandwidth. The asymptotic behavior of this estimator has been studied by

Dodge and Jurečková (1995). It was proven that under regular conditions, as  $n \rightarrow \infty$ ,

$$\sqrt{n\nu_n} \left( \widehat{Z}_n(\theta) - q(\theta) \right) = O_p(1) \text{ and } \sqrt{n\nu_n} \left( \widehat{Z}_n(\theta) - q(\theta) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, q^2(\theta) \overline{K} \right), \quad (4)$$

where  $\overline{K} = \int K^2(x) dx$  with  $K(x) = \int_{-\infty}^x k(y) dy$ . Because of the choice of LAD estimator,  $\theta$  is fixed at 0.5.

## 2.2 Confidence interval of the model intercept

To construct the expected confidence interval of given bounded length, we estimate  $q(1/2)/2$  by

$$\widehat{W}_n(1/2) = \frac{\widehat{Z}_n(1/2)}{2}.$$

Using (4) with  $\theta = 1/2$  we have as  $n \rightarrow \infty$ ,

$$\widehat{W}_n(1/2) \xrightarrow{P} \frac{1}{2f(0)}.$$

So (3) and Slutsky's theorem imply

$$\frac{\sqrt{n} \left( \widehat{\beta}_1^n(1/2) - \beta_1 \right)}{\widehat{W}_n(1/2)} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \text{ as } n \rightarrow \infty.$$

Hence, a  $(1 - \alpha)\%$  confidence interval for  $\beta_1$  is given by

$$I_n = \left[ \widehat{\beta}_1^n(1/2) - \frac{z_{1-\alpha/2}}{\sqrt{n}} \widehat{W}_n(1/2), \widehat{\beta}_1^n(1/2) + \frac{z_{1-\alpha/2}}{\sqrt{n}} \widehat{W}_n(1/2) \right]. \quad (5)$$

## 2.3 Non Independent Observations

In section 2.1, we assumed that the errors are i.i.d. When it is not the case, we can replace the model (1) by

$$\mathbf{Y}_n = \mathbf{X}_n \boldsymbol{\beta} + \mathbf{h}(\boldsymbol{\lambda}, \mathbf{X}_n \boldsymbol{\beta}) \boldsymbol{\varepsilon}_n \quad (6)$$

where  $\varepsilon_n$  are i.i.d and independent of  $X_n$  and  $\mathbf{h}(\boldsymbol{\lambda}, \mathbf{X}_n\boldsymbol{\beta})$  models the correlations between the error terms, which can depend on the known variables and the new parameter  $\boldsymbol{\lambda}$  (Maronna, Martin and Yohai 2006). For example, in our application  $\boldsymbol{\lambda}$  represents the correlation between gene expressions within the same replicate. The same regression model (1) can then be applied to the following transformed variables

$$\mathbf{Y}'_n = \mathbf{Y}_n \mathbf{h}^{-1}(\boldsymbol{\lambda}, \mathbf{X}_n\boldsymbol{\beta}) \quad \text{and} \quad \mathbf{X}'_n = \mathbf{X}_n \mathbf{h}^{-1}(\boldsymbol{\lambda}, \mathbf{X}_n\boldsymbol{\beta}). \quad (7)$$

An example of function  $h$  is given in our application. In the following, we will assume we are working with the transformed variables if there is evidence for correlation between the observations (for example, observations within the same technical or biological replicate).

### 3 SEQUENTIAL APPROACH

The goal of our analyses is to detect a significant change in gene expression across various experimental conditions. From the asymptotic distribution of the intercept given in equation (4), it is straightforward to see that genes that are differentially expressed will have  $\beta_1$  significantly different from 0. Therefore, a sequential procedure can be designed such that experiment stops whenever gene expression changes resulting from exposure to several conditions have been measured with a desired precision. We describe more formally this concept based on the work of several authors (Jurečková 1991; Jurečková and Sen 1996; Hušková 1994 among others) in the following section. Our contribution consists in the use of a robust estimator of linear regression parameters without assuming a distribution for the error term.

### 3.1 Stopping rule for continuous monitoring

We construct a confidence interval  $I_n$  for the first component of the linear model based on a robust estimator of  $\beta_1$  which satisfies:

$$L_n \leq 2d \quad \text{and} \quad P_F(I_n \ni \beta_1) \geq 1 - \alpha, \quad (8)$$

where  $L_n$  denotes the length of  $I_n$ . Starting with an initial sample size  $n_0$ , the experiment stops when the number of replicates  $N_d$  is the smallest  $n > n_0$  such that the length  $L(N_d) < 2d$ .

The length  $L_n$  of the interval  $I_n$  in (5) satisfies

$$\sqrt{n} L_n = 2 z_{1-\alpha/2} \widehat{W}_n(1/2) \xrightarrow{P} \frac{z_{1-\alpha/2}}{f(0)} \text{ as } n \rightarrow \infty. \quad (9)$$

So, for a given sample size  $n$  and for a given  $\alpha$ , the length of the confidence interval  $I_n$  is a random variable. The sequential procedure consists of adding one new observation at a time (i.e. one technical replicate at the high dose in our experiment) until the stopping rule  $L_n \leq 2d$  ( $d > 0$  fixed) is satisfied.

Comparing  $I_n$  to  $I_n^* = [\widehat{\beta}_1^n(1/2) - d, \widehat{\beta}_1^n(1/2) + d]$  with fixed  $d > 0$  using (5), our stopping rule can be expressed as

$$n \geq \frac{z_{1-\alpha/2}^2 \widehat{W}_n^2(1/2)}{d^2}.$$

To avoid an erroneous determination of  $\widehat{W}_n^2(1/2)$ , we must choose an initial sample size  $n_0$  which is sufficiently large. Then, the stopping variable  $N_d$  satisfying the previous conditions on  $n$  can be defined by

$$N_d = \min \left\{ n \geq n_0 \mid n \geq \frac{z_{1-\alpha/2}^2 \widehat{W}_n^2(1/2)}{d^2} \right\}, \quad d > 0. \quad (10)$$

If we have a given sufficiently large sequence of observations, then  $N_d$  would be a nondecreasing function of  $d$ . The monotonicity of  $N_d$  follows directly from the definition of  $N_d$ .

### 3.2 Stopping rule for discrete monitoring: group sequential approach

For discrete monitoring (group sequential approach), the  $(1 - \alpha)$  confidence interval for  $\beta_1$  at the  $k$ th analysis is given by

$$I_{mk} = \left[ \hat{\beta}_1^{mk}(1/2) - \frac{C_k(\alpha)}{\sqrt{mk}} \widehat{W}_{mk}(1/2), \hat{\beta}_1^{mk}(1/2) + \frac{C_k(\alpha)}{\sqrt{mk}} \widehat{W}_{mk}(1/2) \right]. \quad (11)$$

$$I_{mk} = \left[ \hat{\beta}_1^{mk}(1/2) - d_k, \hat{\beta}_1^{mk}(1/2) + d_k \right]. \quad (12)$$

where  $m$  is the size of each group (assumed equal) and  $C_k(\alpha)$  is the critical value for the  $k$ th analysis.

Then, the new stopping variable  $N_{d_k}$  for the group sequential approach can be defined by

$$N_{d_k} = mk = \min \left\{ n \geq n_0 \mid n \geq \frac{C_k(\alpha)^2 \widehat{W}_n^2(1/2)}{d_k^2} \right\}, \quad d_k > 0. \quad (13)$$

In group sequential methods, the critical values  $C_k(\alpha)$  can be determined by using an  $\alpha$ -spending function (Jennison and Turnbull, 2000, Chap. 7). The principle is to partition the Type I error into  $K$  probabilities  $\pi_1, \pi_2, \dots, \pi_K$ , where  $K$  is the number of analyses performed, which sum to  $\alpha$ . Thus,  $\pi_k$  represents the probability of stopping at analysis  $k$  to reject  $H_0$  when this hypothesis is true, also termed the error spent at stage  $k$ . It can be determined using an error spending function which satisfies  $f(0)=0$  and  $f(t) = \alpha$  for  $t \geq 1$ . The value

$f(t)$  indicates the cumulative Type I error that is spent when a fraction  $t$  of the maximum anticipated information has been obtained, denoted  $I_{max}$ . When this maximum information has been reached, the sequential procedure will stop, either accepting or rejecting the null hypothesis.

The Type I error allocated to each analysis is

$$\pi_1 = f(I_1/I_{max}), \pi_k = f(I_k/I_{max}) - f(I_{k-1}/I_{max}), k = 2, 3, \dots, K \quad (14)$$

where  $I_k$  is the amount of information at analysis  $k$ . Without lack of generality, we took  $I_k = mk$  and  $I_{max} = mK$ .

The critical values need to satisfy the constraint

$$Pr\{|Z_1| < c_1, \dots, |Z_{k-1}| < c_{k-1}, |Z_k| \geq c_k | H_0\} = \pi_k \quad (15)$$

where  $|Z_k|$  is the standardized test statistic at the  $k$ th analysis. In our case, the  $Z_k$  statistics correspond to the  $\hat{\beta}_1^{mk}(1/2)$  divided by its standard deviation. The critical values can be determined numerically assuming a given  $\alpha$  – *spending function*. In the following, we consider O’Brien and Fleming test (1979). Another form of this approach has been proposed by Lan and DeMets (**Insert reference**) with the following spending function

$$f(t) = \min\{2 - 2\phi(z_{1-\alpha/2}/\sqrt{t}), \alpha\} \quad (16)$$

where  $t = I_k/I_{max}$  and  $\phi$  is the standard normal distribution.

### 3.3 Asymptotic properties of the stopping variable

Our first result shows that our construction of  $N_d$  fulfills the requirements (8). We first state our main result and the proofs are given in the appendix. These properties are satisfied

for the continuous and discrete cases and our notation below refers to the continuous case.

**Theorem 1.** *Under regularity conditions in the distribution function, the sequence of matrices  $\mathbf{X}_n$  and the kernel estimator, we have*

1.  $E_F(N_d) \rightarrow +\infty$  as  $d \rightarrow 0_+$ .
2.  $P_F(N_d < +\infty) = 1$  for any  $d > 0$ .
3.  $N_d/n_d \xrightarrow{P} 1$  as  $d \rightarrow 0_+$ , where  $n_d = z_{1-\alpha/2}^2 \sigma^2/d^2$  and  $\sigma = \sigma_{1/2} = 1/(2f(0))$ .
4.  $P_F(I_{N_d} \ni \beta_1) \geq 1 - \alpha$  for  $\alpha \in (0, 1)$  fixed as  $d \rightarrow 0_+$ .

The next theorem concerns the asymptotic behaviour of the stopping variable.

**Theorem 2.** *Let  $\eta \in (1/4, 1/3)$  and  $\nu_n$  be such that  $\nu_n = n^{2\eta-1}$ . Under regularity conditions we have as  $d \rightarrow 0_+$  :*

$$d^{1-2\eta} \left( \sqrt{N_d} - \frac{z_{1-\alpha/2} \sigma}{d} \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \sigma^{2(1-2\eta)} z_{1-\alpha/2}^{2(1-2\eta)} \int K^2(x) dx \right).$$

Proofs of theorems 1 and 2 require an intermediate result covering essentially the problem of uniform continuity in probability (Anscombe condition, Anscombe 1952) of the least-absolute regression estimator and the asymptotic kernel variance estimator. The proof of Theorem 2 is given in the Appendix. To avoid the presentation of lengthy mathematical results, the proof of the Anscombe condition is given as a technical report at posted at [www.amstat.org/publications/jasa/supplemental\\_materials](http://www.amstat.org/publications/jasa/supplemental_materials).

**Proof of Theorem 1.**

1) It follows by construction that  $N_d$  increases as  $d$  decreases ; this together with the fact that  $L_n > 0$  with probability 1 implies that

$$\forall m \geq 0, \quad \lim_{d \rightarrow 0_+} P_F \{N_d < m\} \leq \lim_{d \rightarrow 0_+} P_F \{L_{N_d} < 2d\} = 0,$$

and thus  $N_d \xrightarrow{P} \infty$  as  $d \rightarrow 0_+$ . So using the monotone convergence theorem we have:

$$\lim_{d \rightarrow 0} E_F(N_d) = E_F\left(\lim_{d \rightarrow 0} N_d\right) = \infty.$$

2) For every  $n \geq n_0$  we have:

$$P_F \{N_d > n\} \leq P_F \left\{ n < d^{-2} z_{1-\alpha/2}^2 \widehat{W}_n^2(1/2) \right\}.$$

From (4) we deduce that  $\frac{\widehat{W}_n(1/2)}{n} \xrightarrow{P} 0$  as  $n \rightarrow \infty$ , hence

$$P_F \{N_d = \infty\} = \lim_{n \rightarrow \infty} P_F \{N_d > n\} \leq \lim_{n \rightarrow \infty} P_F \{n < d^{-2} z_{1-\alpha/2}^2 \widehat{W}_n^2(1/2)\} = 0.$$

Therefore

$$P_F \{N_d = \infty\} = 0 \quad \text{and} \quad P_F \{N_d < +\infty\} = 1 \quad \text{for every } d > 0.$$

3) Note that with probability 1, we have

$$\frac{z_{1-\alpha/2}^2 \widehat{W}_{N_d}^2(1/2)}{d^2} \leq N_d < \max \left( n_0 + 1, \frac{z_{1-\alpha/2}^2 \widehat{W}_{N_d-1}^2(1/2)}{d^2} + 1 \right).$$

Recalling that  $n_d = \frac{z_{1-\alpha/2}^2 \sigma^2}{d^2}$ , we obtain

$$\frac{\widehat{W}_{N_d}^2(1/2)}{\sigma^2} < \frac{N_d}{n_d} < \max \left( \frac{n_0}{n_d}, \frac{\widehat{W}_{N_d-1}^2(1/2)}{\sigma^2} \right) + \frac{1}{n_d}. \quad (17)$$

Moreover, using the fact that

$$\widehat{W}_n(1/2) \xrightarrow{P} \sigma \text{ as } n \rightarrow \infty, \quad N_d \xrightarrow{P} \infty \text{ as } d \rightarrow 0_+$$

and uniform continuity in probability of  $\widehat{W}_n(1/2)$ , we obtain

$$\frac{\widehat{W}_{N_d}(1/2)}{\sigma} \xrightarrow{P} 1 \quad \text{and} \quad \frac{\widehat{W}_{N_d-1}(1/2)}{\sigma} \xrightarrow{P} 1 \text{ as } d \rightarrow 0_+. \quad (18)$$

By (18), as  $d \rightarrow 0_+$ , we have

$$\max \left( \frac{n_0}{n_d}, \frac{\widehat{W}_{N_d-1}^2(1/2)}{\sigma^2} \right) \longrightarrow \max(0, 1) = 1.$$

Finally, by (17) we obtain the expected result, i.e.

$$\frac{N_d}{n_d} \xrightarrow{P} 1 \text{ as } d \rightarrow 0_+.$$

4) Moreover, because  $\widehat{\beta}_1^n$  is asymptotically normal and uniformly continuous in probability,

we obtain,

$$\begin{aligned} \frac{\sqrt{N_d} \left( \widehat{\beta}_{N_d}(1/2) - \beta_1 \right)}{\widehat{W}_{N_d}(1/2)} &= \left( \frac{\sigma}{\widehat{W}_{N_d}(1/2)} \right) \left( \frac{N_d}{n_d} \right)^{1/2} \left( \frac{\sqrt{n_d} \left( \widehat{\beta}_{N_d}(1/2) - \beta_1 \right)}{\sigma} \right) \\ &= \left( \frac{\sigma}{\widehat{W}_{N_d}(1/2)} \right) \left( \frac{N_d}{n_d} \right)^{1/2} \left\{ \frac{\sqrt{n_d} \left( \widehat{\beta}_{n_d}(1/2) - \beta_1 \right)}{\sigma} + \frac{\sqrt{n_d} \left( \widehat{\beta}_{N_d}(1/2) - \widehat{\beta}_{n_d}(1/2) \right)}{\sigma} \right\} \end{aligned}$$

$$\xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } d \rightarrow 0_+.$$

As  $d \rightarrow 0_+$ , we have:

$$\lim_{d \rightarrow 0_+} P_F \left\{ \left| \widehat{\beta}_{N_d}(1/2) - \beta_1 \right| \leq z_{1-\alpha/2} \widehat{W}_{N_d}(1/2) N_d^{-1/2} \right\} = 1 - \alpha. \quad (19)$$

By (19) and (13) we finally have:

$$\lim_{d \rightarrow 0_+} P_F \left\{ \widehat{\beta}_{N_d}(1/2) - d \leq \beta_1 \leq \widehat{\beta}_{N_d}(1/2) + d \right\} \geq 1 - \alpha.$$

## 4 THE MICROARRAY EXPERIMENT

Recently, Lobenhofer et al. (2004) illustrated the interest of model-based approaches for microarray analysis in a study that tried to identify genes that respond to estrogen treatment and evaluated more particularly the doses of estrogen capable of inducing a transcriptional response in breast cancer cell lines. In terms of our methodology, this dataset possesses several major points of interest. First, we believe that their fixed sample size design is not optimal and the differentially expressed genes could have been detected using fewer replicates. Second, the design can be analyzed using discrete monitoring, the group can be a technical replicate ( $m=2$ ) or a biological replicate ( $m=4$ ). Due to the small sample size of this experiment, we decided to use the technical replicate as the group. Third, the observations can be correlated within each technical and biological replicate. Finally, this study was among the rare ones to clearly establish a dose-response effect in a gene expression experiment and therefore, the variable dose has a real scientific interest and needs to be modeled appropriately.

In this study, the gene expression levels of a hormone responsive breast cancer cell line (MCF-7) are measured after stimulation with various concentrations of estrogen, above (high-dose effect) and below (low-dose effect) normal physiologic levels and compared with the corresponding levels of expression in control samples (Lobenhofer et al. 2004). The data were downloaded from the website <http://dir.niehs.nih.gov/microarray/datasets/homepub.htm>.

The cell lines were treated with estrogen at four concentrations ( $10^{-8}$  M,  $10^{-10}$  M,

high doses) and ( $10^{-13}$  M,  $10^{-15}$  M, low doses) or with concentration-matched ethanol solvent (control samples). The RNA sample from each estrogen treatment and its corresponding control were compared using custom-made cDNA microarrays, ToxChip version 1.0, in a dye-swap fashion. Each chip had 1920 clones double-printed in 2 subarrays, so each clone was duplicated on the array. The dye swap and the duplicated spots lead to four technical replicates for each biological replicate. There were two biological replicates, so eight measurements at each dose. In this analysis, we focused on seven genes that were validated by real-time PCR. Five genes were confirmed upregulated and two were downregulated. These genes are known to be involved in different cancer pathways. Because the threshold dose can be observed below the level  $10^{-8}$  M, we just included in our analysis the 2 low doses and only 1 high dose ( $10^{-10}$  M). Thus, there were a total of 24 measurements available for each gene (3 doses x 8 measurements).

The data were pre-processed before analysis as described in Lobenhofer et al. (2004). In brief, the gridding quality of microarray images was checked for misalignment using the gridcheck function in the R/maanova package. Spots with higher background intensity than foreground were removed from the analysis. The normalization process included an intensity-lowess transformation and channel-mean centering of the data.

The data were then transformed to remove the possible correlation between the observations belonging to the same biological replicate, following the principle described in section 2.3. To estimate the function  $\mathbf{h}$  in (6, 7), we proceeded as follows using the pooled data of the seven genes. First, we estimated the residuals from the quantile regression model specified in (1) including the dose and dye effects as explanatory variables. Then, we estimated the

variance-covariance matrix of the residuals using a linear model with a compound symmetry correlation structure defined as a block-diagonal matrix in which each block is  $4 \times 4$  with 1 for the diagonal entries and  $\lambda$  for the off-diagonal entries.

Each block of dimension  $(4 \times 4)$  represents a biological replicate and we assumed the same correlation structure for each gene as well as the same correlation between two technical replicates. To analyze the data, we fitted the regression model (1) with the dose and dye effects as covariables using quantile regression on the transformed response and covariates. Then, we computed a robust confidence interval  $I_n$  of fixed length (sections 2-3) for the model intercept. The initial sample size  $n_0$  was fixed at 16, which corresponds to the number of measurements at the two low doses. Then we determine sequentially the number of groups of technical replicates at the high dose, where each group had two observations (duplicated spots). The experimental process stops for  $N_{d_k}$  the smallest  $n > n_0$  is such that the length of  $I_n$  ( $L_n < 2d$ ). The results are presented in table 1 for  $\alpha = 0.001$ , the value chosen by Lobenhofer et al. (2004) to adjust for multiple comparisons in their study.

For each gene, the precision parameter  $d_k$  ( $k = 1, \dots, 4$ ) is determined by the constraint on the stopping variable  $N_{d_k} = \frac{C_k^2(\alpha)\widehat{W}_n^2(1/2)}{d_k^2}$ . It is lower than 24 and can take only the following values 16, 18, 20, 22 or 24. The estimator  $\widehat{W}_n^2(1/2)$  denotes the kernel regression estimator of the standard deviation of the intercept parameter.

These results indicate the minimal number of technical replicates  $N_{d_k}$  required to estimate the gene expression level with a fixed precision. We observe that for a relatively small precision parameter ( $d_k$  values), the stopping value remains lower than 24 (the maximum number of technical replicates available for the 3 doses). Genes 1, 2, 3, 4 and 5 were signif-

icant upregulated differentially expressed ( $p < 0.001$ ) and genes 6 and 7 are downregulated differentially expressed. These seven genes were confirmed to be differentially expressed by quantitative PCR. In conclusion, our analysis showed that the experimenter could have stopped the experiment after collecting only 22 replicates and would have reached the same conclusion as in Lobenhofer et al. (2004).

## 5 SIMULATION STUDY

To better evaluate the performance of our sequential method, we conducted a small simulation study. Our goal was to assess more specifically the sensitivity of our sequential approach to three factors: (1) the use of continuous vs. discrete sequential procedures; (2) the distribution of the error term in the model; (3) the presence of a batch effect. The design of the simulation study mimicked the experimental design used in our application. The  $X$  and  $\varepsilon$  were generated independently, where  $X$  follows under discrete monitoring a discrete uniform distribution taking the values 0, 3 and 5 and under continuous monitoring a uniform distribution with values belonging to the interval (0, 5). In the discrete monitoring, each group included sequentially had a size of 8 observations and the group was added at one specific dose. We also considered the possibility of a batch effect, the value of which changed whenever a new group was added (discrete monitoring) or when eight observations were successively added (continuous monitoring). The data were generated from the following linear model:

$$Y = \beta_1 + \beta_2 Dose + \beta_3 Batch + \delta \varepsilon$$

where  $\beta_1$  is the intercept denoting the gene expression at an initial dose,  $\beta_2$  is the slope parameter coding for the change in gene expression with dose effect. The terms  $\beta_3$  and  $\delta$  correspond to the additive and multiplicative effects of the batch. The source of variation  $\epsilon$  is a random vector of independent and identically distributed errors with a distribution function which is either Normal  $N(0, \sigma_*^2)$  or Laplace  $L(0, \sigma_*^2)$ . The reason for choosing this latter distribution is to show the robustness of the  $L^1$  estimator in presence of long-tailed error distributions.

The regression parameters ( $\beta_1$ ,  $\beta_2$  and  $\beta_3$ ) were fixed respectively to 1,  $-0.2$  and  $0.1$  and  $\sigma_*$  was set to  $0.3$  and  $0.4$ , as observed in the application. The precision parameter  $d$  was determined such that the theoretical stopping rule value  $n_d = z_{1-\alpha/2}^2 \sigma^2 / d^2$  was equal to  $32$  or  $56$ , where  $\sigma^2$  denotes the theoretical asymptotic variance of the  $\beta_1$  estimator. In all simulations, we used  $\nu_n = n^{-\delta}$ , with  $\delta = 1 - 2\eta = 0.42 \in (1/3, 1/2)$ . In this situation, the conditions of the theorems giving the limiting behavior of the kernel estimator are satisfied. The Epanechnikov kernel was used but we obtained similar results with the Gaussian kernel. The initial sample size was fixed at  $n_0 = 16$  in all simulations and the number of Monte-Carlo simulations was  $1,000$ . Results are presented for the normal error distribution in Table 2 and for Laplace error distribution in Table 3.

**Continuous vs. discrete sequential procedure with no batch effect.** We obtained almost unbiased parameter estimates of the regression parameters when the procedure was continuous. The use of a discrete procedure does not alter the parameter estimates but the standard deviation of the estimates is slightly larger in that case. The standard deviation

of the intercept, estimated by the kernel procedure, is generally slightly underestimated compared to the theoretical value in both sequential procedures. The sample size required to stop the experiment is larger under the discrete procedure. In some cases, one or two additional groups (of eight observations) are needed to stop the experiment compared to the continuous procedure. The length of the confidence interval is inversely proportional to the stopping rule value and, consequently, it is smaller under the discrete procedure. The coverage probability was very close to the 99.9% nominal level in the continuous procedure but was lower than that in the discrete procedure. This is because more groups are used than expected, and therefore the critical values used in the group sequential procedure were not optimal.

**Batch effect.** The parameter estimates (intercept and slope) are not altered by the presence of the batch effect and are still almost unbiased. The additive effect of the batch is also well estimated in the model. The length of the confidence intervals is larger when there is a batch effect, which is expected because the multiplicative effect of the batch increased the variance of the error distribution in the regression model. The batch effect also affected the coverage probabilities which are smaller than the nominal 99.9% level even in the continuous case. The bias in the coverage probabilities is larger with the discrete sequential procedure.

**Normal vs. Laplace error distribution.** The results are very similar with normal and Laplace error terms. The parameter estimate is not very sensitive to the distribution of the error term, which shows the value of a robust estimator.

## 6 Discussion

Our paper establishes a general theoretical framework for sequential approaches in microarray experiments. Our theoretical developments as well as our real and simulated data demonstrate its advantages as compared to fixed-sample size approaches. More particularly, our application showed that we were able to replicate Lobenhofer et al.'s results (2004) and reach the same conclusions with fewer technical replicates. Sequential designs, in contrast to fixed-sample size designs, can achieve efficiency by making just enough measurements to evaluate the experimental endpoints with the desired precision. Numerically, we also proved that our approach is valid in finite samples and is robust to the choice of the error term distribution in the linear model. The use of discrete instead of continuous monitoring did not alter the parameter estimates of the model. However, for small sample size problems ( $n_d < 25$ ), as observed in our application, the use of discrete monitoring slightly increased the sample size needed to stop the experiment. Our sequential approach is therefore slightly liberal and some adjustment of the critical values are required to achieve the correct nominal type I error rate.

The major question related to this type of approach is its feasibility and applicability in practice. We partly answered this question through our application and small simulations. In the application, we sampled sequentially the technical replicates with two observations each. Sampling groups of biological replicates, each with four observations, would have been more relevant, but was not possible due to the relatively small sample size available for the experiment. The initial sample size was fixed to 16 and therefore the range of possible

stopping variables was quite limited. Despite this limitation, our sequential procedure was always able to converge and gave confidence intervals that were relatively precise for this type of problem. The small simulations confirmed the robustness of our approach to outliers and departure from normality. In many biological experiments, especially those involving technologies such as microarrays, data have generally a low signal-to-noise ratio that makes analysis more complex and very sensitive to outliers. Therefore, our approach warrants more general application in microarray studies. Our simulations also demonstrated the accuracy of the procedure to estimating the model parameters and its good behavior in the presence of batch effects.

Besides the statistical properties mentioned above, it is important to discuss the technical feasibility and applicability of the method. In a recent NCIC (National Cancer Institute of Canada) project, we planned to apply a novel microarray experimental design where four batches of eight microarrays will be sequentially realized with the aim to study tumor progression in prostate cancer patients. The first two batches allow one to perform a first screen of methylation profiles and draw hypotheses regarding the role of several genetic pathways. Each batch corresponds to a particular stage, so the next batches will make precise the role of these pathways in tumor progression. In our simulations, we introduced a method that took direct account of batch effects. Alternatively, one could adjust the data prior to analysis at the pre-processing step of the experiment (Johnson, Li and Rabinovic 2007). Because most array technologies are based on comparative hybridization, one can use the control sample for the adjustment. Therefore, we think batch effects are not a major problem to deal with, when present in the experiment.

At this stage of development, our approach can also be applied to find individual genes differentially expressed across several experimental conditions (treatment vs. control, different doses or time-points) and rank them according to their importance for the given class comparison problem. For example, a classical  $z$  test can be performed on the intercept and can serve as a criterion to rank the most important genes in the microarray experiment. We also plan an extension of this method to the joint analysis of multiple genes that could be differentially expressed in a coordinated manner. For example, the work from Gibbons et al. (2005) can be easily adapted to our method. We are also planning to integrate the correlation structure directly in the analysis rather than adjusting for it to gain some efficiency. Altogether, we anticipate that this approach could have a significant contribution to microarray data analysis by improving the usual experimental designs and methods of analysis.

## APPENDIX: PROOFS

Before demonstrating the proof of Theorem 2, we provide some auxiliary results on the asymptotic normality of  $N_d$ . We now study the asymptotic behavior of the stopping rule  $N_d$ . We assume that  $\nu_n = n^{2\eta-1}$  for some  $\eta \in (1/4, 1/3)$ .

**Lemma 1.** *Let  $\eta \in (1/4, 1/3)$  and  $\nu_n$  be such that  $\nu_n = n^{2\eta-1}$ . Under regularity conditions, we have as  $d \rightarrow 0_+$  :*

(i)

$$N_d^\eta \left( \widehat{W}_{N_d}(1/2) - \sigma \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \sigma^2 \overline{K} \right), \quad (20)$$

(ii)

$$N_d^\eta \left( \widehat{W}_{N_d-1}(1/2) - \sigma \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \sigma^2 \overline{K} \right), \quad (21)$$

(iii) for all  $\epsilon' > 0$  :

$$P \left\{ N_d^\eta \left| \left( \frac{\sqrt{N_d} - \sqrt{N_d-1}}{\sqrt{\overline{K}} \sigma} \right) \frac{d}{z_{1-\alpha/2}} \right| > \epsilon' \right\} \longrightarrow 0. \quad (22)$$

### Proof of Lemma 1.

(i) By (4) we obtain

$$\sqrt{n\nu_n} \left( \widehat{W}_n(1/2) - \sigma \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \sigma^2 \overline{K} \right). \quad (23)$$

Taking  $\nu_n = n^{2\eta-1}$ , we have

$$\sqrt{N_d \nu_{N_d}} \left( \widehat{W}_{N_d}(1/2) - \sigma \right) = \left( \frac{N_d}{z_{1-\alpha/2}^2 \sigma^2 / d^2} \right)^\eta \left( \frac{z_{1-\alpha/2}^2 \sigma^2}{d^2} \right)^\eta \left( \widehat{W}_{N_d}(1/2) - \sigma \right). \quad (24)$$

By the third result of Theorem 1 we have, as  $d \rightarrow 0_+$ ,

$$\frac{N_d^\eta}{\frac{z_{1-\alpha/2}^{2\eta} \sigma^{2\eta}}{d^{2\eta}}} \xrightarrow{P} 1 \quad (25)$$

and by (23), as  $d \rightarrow 0_+$ ,

$$P \left\{ n^\eta \left( \widehat{W}_n(1/2) - \sigma \right) \leq y \right\} \longrightarrow \Phi \left( \frac{y}{\sqrt{\overline{K}} \sigma} \right), \quad \forall y \in \mathbb{R}, \quad (26)$$

where  $\Phi$  denotes the cumulative distribution function of the standard normal distribution.

In addition to (25) and (26), the theorem of Anscombe (1952) allows us to conclude that,

$\forall y \in \mathbb{R}$ ,

$$P \left\{ \frac{\sigma^{2\eta} z_{1-\alpha/2}^{2\eta}}{d^{2\eta}} \left( \widehat{W}_{N_d}(1/2) - \sigma \right) \leq y \right\} \longrightarrow \Phi \left( \frac{y}{\sqrt{\overline{K}} \sigma} \right) \quad \text{as } d \rightarrow 0_+.$$

Therefore, as  $d \rightarrow 0_+$ , we have

$$\frac{\sigma^{2\eta} z_{1-\alpha/2}^{2\eta}}{d^{2\eta}} \left( \widehat{W}_{N_d}(1/2) - \sigma \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \overline{K} \sigma^2 \right). \quad (27)$$

Using (25), (27) and Slutsky's Theorem on (24), we obtain

$$N_d^\eta \left( \widehat{W}_{N_d}(1/2) - \sigma \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \sigma^2 \overline{K} \right) \text{ as } d \rightarrow 0_+,$$

(ii) Relation (21) is derived in the same way.

(iii) We now prove (22). For  $x > 1$ , we have  $0 \leq \sqrt{x} - \sqrt{x-1} \leq 1/(2\sqrt{x-1})$ . We thus have, for all  $\epsilon' > 0$ ,

$$\mathbb{P} \left\{ N_d^\eta \left( \frac{\sqrt{N_d} - \sqrt{N_d - 1}}{\sqrt{\overline{K}} \sigma} \right) \frac{d}{z_{1-\alpha/2}} > \epsilon' \right\} \leq \mathbb{P} \left\{ N_d^\eta \frac{d(N_d - 1)^{-1/2}}{2\sqrt{\overline{K}} z_{1-\alpha/2} \sigma} > \epsilon' \right\}.$$

Moreover, as  $d \rightarrow 0_+$ , we have

$$N_d^\eta \frac{d(N_d - 1)^{-1/2}}{\sqrt{\overline{K}} z_{1-\alpha/2} \sigma 2} \sim \frac{N_d^{\eta-1/2}}{2\sqrt{n_d} \sqrt{\overline{K}}}.$$

Consequently, by Theorem 1 (3), we have as  $d \rightarrow 0_+$ ,

$$\frac{N_d^{\eta-1/2}}{2\sqrt{n_d} \sqrt{\overline{K}}} \xrightarrow{P} 0.$$

**Proof of Theorem 2.** One one hand, as  $N_d = \min \left\{ n \geq n_0 \mid \widehat{W}_n(1/2) \leq \frac{\sqrt{n} d}{z_{1-\alpha/2}} \right\}$ , with  $d > 0$ , we have

$$\widehat{W}_{N_d}(1/2) \leq \frac{\sqrt{N_d} d}{z_{1-\alpha/2}},$$

and, therefore as  $d \rightarrow 0_+$ ,

$$\begin{aligned} & \limsup P \left\{ N_d^\eta \left( \frac{\sqrt{N_d} d}{z_{1-\alpha/2}} - \sigma \right) \frac{1}{\sqrt{K} \sigma} \leq y \right\} \\ & \leq \limsup P \left\{ N_d^\eta \left( \widehat{W}_{N_d}(1/2) - \sigma \right) \frac{1}{\sqrt{K} \sigma} \leq y \right\}. \end{aligned}$$

By Lemma 1 (i), we have as  $d \rightarrow 0_+$ ,

$$\limsup P \left\{ N_d^\eta \left( \frac{\sqrt{N_d} d}{z_{1-\alpha/2}} - \sigma \right) \frac{1}{\sqrt{K} \sigma} \leq y \right\} \leq \Phi(y). \quad (28)$$

On the other hand, according to the definition of the stopping rule  $N_d$ , we have

$$\widehat{W}_{N_d-1}(1/2) > \frac{\sqrt{N_d-1} d}{z_{1-\alpha/2}},$$

and therefore, as  $d \rightarrow 0_+$ ,

$$\begin{aligned} & \liminf P \left\{ N_d^\eta \left( \widehat{W}_{N_d-1}(1/2) - \sigma \right) \frac{1}{\sqrt{K} \sigma} \leq y \right\} \\ & \leq \liminf P \left\{ N_d^\eta \left( \frac{\sqrt{N_d-1} d}{z_{1-\alpha/2}} - \sigma \right) \frac{1}{\sqrt{K} \sigma} \leq y \right\}. \end{aligned}$$

Using Lemma 1 (ii) and Lemma 1 (iii), we obtain, as  $d \rightarrow 0_+$ ,

$$\liminf P \left\{ N_d^\eta \left( \frac{\sqrt{N_d} d}{z_{1-\alpha/2}} - \sigma \right) \frac{1}{\sqrt{K} \sigma} \leq y \right\} \geq \Phi(y). \quad (29)$$

Therefore, by (28) and (29), we have  $\forall y \in \mathbb{R}$

$$P \left\{ N_d^\eta \left( \frac{\sqrt{N_d} d}{z_{1-\alpha/2}} - \sigma \right) \frac{1}{\sqrt{K} \sigma} \leq y \right\} \longrightarrow \Phi(y) \text{ as } d \rightarrow 0_+.$$

Since  $d \sqrt{N_d} \xrightarrow{P} z_{1-\alpha/2} \sigma$ , the proof is complete.

## Acknowledgements

This study was supported by the Discovery Grant 293270-04 from the Natural Sciences and Engineering Research Council of Canada and CNRS. We would like to thank Dr. Cyr Emile M'lan, Prof. Jaromir Antoch, Prof. Jean-Marc Deshouillers and Prof. Jana Jurečková for their contributions on our manuscript.

## References

- Allison, D.B., Gadbury, G., Heo, M., Fernandez, J., Lee, C.K., Prolla, T.A., and Weindruch, R. (2002), “A mixture model approach for the analysis of microarray gene expression data,” *Computational Statistics and Data Analysis*, 39, 1–20.
- Anscombe, F. J. (1952), “Large sample theory of sequential estimation,” *Proceedings of the Cambridge Philosophical Society*, 48, 600–607.
- Benjamini, Y., and Hochberg, Y. (1995), “Controlling the false discovery rate: a practical and powerful approach to multiple testing,” *Journal of the Royal Statistical Society*, 57, 289–300.
- Churchill, G. A. (2002), “Fundamentals of experimental design for cDNA microarrays,” *Nature Genetics*, 32, 490–495.
- Cui, X., Kerr, M. K., and Churchill, G. A. (2002), “Data transformation for cDNA mi-

croarray data,” <http://www.jax.org/staff/churchill/absite/pubs/Transformation.pdf>.

Cui, X., and Churchill, G. A. (2003), “Statistical tests for differential expression in cDNA microarray experiments,” *Genome Biology*, 4(4),210.

Dobbin, K., and Simon, R. (2005), “Sample size determination in microarray experiments for class comparison and prognostic classification,” *Biostatistics*, 6, 27–38.

Dodge, Y., and Jurečková, J. (1995), “Estimation of quantile density function based on regression quantiles,” *Statistics and Probability Letters*, 23, 73–78.

Fu, W., Dougherty, E., Mallick, B., and Carroll, R. (2005), “How many samples are needed to build a classifier: a general sequential approach,” *Bioinformatics*, 21(1), 63-70.

Gadbury, G. L., Page, G. P., Edwards, J. W., Kayo, T., Prolla, T. A., Weindruch, R., Permana, P. A., Mountz, J., and Allison, D. B. (2004), “Power analysis and sample size estimation in the age of high dimensional biology: a parametric bootstrap approach and examples from microarray research,” *Statistical Methods in Medical Research*, 13, 325–338.

Gibbons, R. D., Bhaumik, D. K., Cox, D. R., Grayson, D. R., Davis, J. M., and Sharma, R. P. (2005), “Sequential prediction bounds for identifying differentially expressed genes in replicated microarray experiments,” *Journal of Statistical Planning and Inference*, 129, 19–37.

Glonek, G. F., and Solomon, P. J. (2004), “Factorial and time course designs for cDNA microarray experiments,” *Biostatistics*, 5(1), 89–111.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, J., and Stahel, W. A. (1986), “Robust

- Statistics: the approach based on influence functions,” New York: John Wiley.
- Huber, P. J. (2003), “Robust Statistics,” New York: Wiley series in probability and statistics.
- Hušková, M. (1994), “Some sequential procedures based on regression rank scores,” *Nonparametric Statistics*, 3, 285–298.
- Jennison, C., and Turnbull, B. W. (2000), “Group sequential methods with applications to clinical trials,” New York: Chapman and Hall.
- Jin, W., Riley, R. M., Wolfinger, R. D., White, K. P., Passador-Gurgel, G., and Gibson, G. (2001), “The contribution of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*,” *nature Genetics*, 29(4), 389–395.
- Johnson, W. E., Li, C., and Rabinovic, A. (2007), “Adjusting batch effects in microarray expression data using empirical Bayes methods,” *Biostatistics*, 8, 118–127.
- Jurečková, J. (1984), “Regression quantiles and trimmed least squares estimator under a general design,” *Kybernetika*, 20, 345–357.
- Jurečková, J. (1991), “Confidence sets and intervals,” in *Handbook of Sequential Analysis*, eds. B. K. Ghosh and P. K. Sen, pp. 269–282.
- Jurečková, J., and Sen, P. K. (1996), “Robust statistical procedures: Asymptotics and interrelations,” New York: John Wiley.
- Kerr, M. K., and Churchill, G. A. (2001a), “Experimental design for gene expression microarrays,” *Biostatistics*, 2, 183–201.
- Kerr, M. K., and Churchill, G. A. (2001b), “Statistical design and the analysis of gene expression microarray data,” *Genetical Research*, 77, 123–128.

- Koenker, R. W., and D'Orey, V. (1987), "Computing regression quantiles," *Journal of the Royal Statistical Society*, C49, 383–401.
- Lobenhofer, E. K., Cui, X., Bennett, L., Cable, P. L., Merrick, B. A., Churchill, G. A., and Afshari, C. A. (2004), "Exploration of low-dose estrogen effects: identification of no observed transcriptional effect level (NOTEL)," *Toxicologic Pathology*, 2, 482–492.
- Maronna R. A., Martin, R. M., and Yohai, V. J. (2006) "Robust Statistics: the approach based on influence functions," New York: John Wiley.
- McShane L. M., Shih J. H., and Michalowska, A. M., "Statistical issues in the design and analysis of gene expression microarray studies of animal models," *Journal of Mammary gland and Biological Neoplasia*, 8(3), 359–374.
- Muller, P., Parmigiani, G., Robert, C., and Rousseau, J. (2004), "Optimal sample size for multiple testing: The case of gene expression microarrays," *Journal of the American Statistical Association*, 99, 990–1001.
- Pan, W., Lin, J., and Le, C. T. (2002), "How many replicates of arrays are required to detect gene expression changes in microarray experiments? A mixture model approach," *Genome Biology*, 3(5), research 0022.1–0022.10.
- Pavlidis, P., Li, Q., and Noble, W. S. (2003), "The effect of replication on gene expression microarray experiments," *Bioinformatics*, 19, 1620–1627.
- Pawitan, Y., Michiels, S., Koscielny, S., Gusnanto, A., and Ploner, A. (2005), "False discovery rate, sensitivity and sample size for microarray studies," *Bioinformatics*, 21, 3017–3024.
- Rosenberg W. F. (2002), "Competing designs for phase I clinical trials: a review," *Statis-*

tics in *Medecine*, 21, 2757–2270.

Simon, R. M. (2003), “Using DNA microarrays for diagnostic and prognostic prediction,” *Expert Review of Molecular Diagnostics*, 3(5), 587–595.

Smyth, G. K. (2004), “Linear models and empirical Bayes methods for assessing differential expression in microarray experiments,” *Statistical applications in genetics and molecular biology*, 3(1), Article 3.

Tibshirani, R. (2006), “A simple method for assessing sample sizes in microarray experiments,” *BMC Bioinformatics*, 7, 106.

Thomas, J. G., Olson, J. M., Tapscott, S. J., and Zhao, L. P. (2001), “An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles,” *Genome Research*, 11, 1227–1236.

Tsai, C. A., Hsueh, H. M., and Chen, J. J. (2003), “Estimation of false discovery rates in multiple testing: application to gene microarray data,” *Biometrics*, 59, 1071–1081.

Warnes, G. R., and Liu, P. (2005), “Sample size estimation for microarray experiments,” [www.bioconductor.org/repository/devel/vignette/ssize.pdf](http://www.bioconductor.org/repository/devel/vignette/ssize.pdf).

Wolfinger, R. D., Gibson, G., Wolfinger, E. D., Bennett, L., Hamadeh, H., Bushel, P., Afshari, C., and Paules, R. S. (2001), “Assessing gene significance from cDNA microarray expression data via mixed models,” *Journal of Computational Biology*, 8(6), 625–637.

Yang, Y. H., and Speed, T. P. (2002) “Design issues for cDNA microarray experiments,” *Nature Review Genetics*, 3, 579–558.

Zhao, H. Y., Yue, P. Y., and Fang, K. T. (2004), “Identification of differentially expressed genes with multivariate outlier analysis,” *Journal of Biopharmaceutical Statistics*, 14(3),

629–646.

Zhou, Y. (2004), “Choice of designs and doses for early phase trials,” *Fundamentals Clin Pharma*, 18, 373–378.

Zien, A., Fluck, J., Zimmer, R., and Lengauer, T. (2003), “Microarrays: how many you need?,” *Journal of Computational Biology*, 10, 653–667.

Table 1: Results from group sequential discrete monitoring microarray experiment.

label	name	$d$	$N_d$	$\hat{\beta}_1^{N_d}$	$SD\left(\hat{\beta}_1^{N_d}\right)$	99.9% CI
<i>Gene 1</i>	SDF1	0.375	22	1.175	0.283	(0.976, 1.374)
<i>Gene 2</i>	MYB	0.410	20	0.971	0.457	(0.578,1.364)
<i>Gene 3</i>	CDC28	0.243	18	0.668	0.111	(0.582,0.755)
<i>Gene 4</i>	LRP8	0.480	22	0.479	0.547	(0.094,0.864)
<i>Gene 5</i>	CDC25A	0.265	18	0.081	0.042	(0.048,0.113)
<i>Gene 6</i>	TMP3	0.215	18	-0.369	0.257	(-0.569,-0.169)
<i>Gene 7</i>	PRKCZ	0.226	20	-0.445	0.260	(-0.669,-0.221)

NOTE:  $d$ : precision parameter;  $N_d$ : stopping variable estimator;  $\hat{\beta}_1^{N_d}$ : LAD intercept estimator;  $SD\left(\hat{\beta}_1^{N_d}\right)$ : kernel regression estimator of the standard deviation of  $\hat{\beta}_1^{N_d}$ ; CI: 99.9% Confidence Interval of the intercept.

Table 2: Simulation results with normal error in regression model (Standard deviations are in parenthesis)

$\sigma_*$	d	Batch	nd	Median $N_d$	Mean Intercept True=1	Mean Slope True= -0.2	Mean Batch True=0.1	$\sigma$ = $\sigma_{1/2}$	Mean $W_{N_d}$	CI 99.9%	CIL	CP
Continuous												
0.3	0.2187	0	32	32 (0.7611)	0.9970 (0.1351)	-0.1986 (0.0485)	– –	0.3760	0.3522 (0.1083)	(0.7835 , 1.2104)	0.4269	1.0000
0.3	0.2187	1	32	35 (2.9536)	0.9851 (0.2199)	-0.2017 (0.0678)	0.1070 (0.1117)	0.3760	0.3521 (0.1320)	(0.7625 , 1.2078)	0.4453	1.0000
0.4	0.2916	0	32	32 (0.7738)	0.9991 (0.1783)	-0.1995 (0.0644)	– –	0.5013	0.4756 (0.1480)	(0.7177 , 1.2806)	0.5629	1.0000
0.4	0.2916	1	32	34 (1.1264)	0.9994 (0.3309)	-0.2017 (0.0781)	0.1034 (0.1414)	0.5013	0.4715 (0.1818)	(0.6972 , 1.3016)	0.6044	0.9753
0.3	0.1653	0	56	48 (1.3801)	0.9968 (0.1159)	-0.1981 (0.0418)	– –	0.3760	0.3352 (0.1046)	(0.8269 , 1.1666)	0.3397	0.9988
0.3	0.1653	1	56	55 (2.4446)	0.9996 (0.2198)	-0.1980 (0.0545)	0.0989 (0.0950)	0.3760	0.3384 (0.1349)	(0.8099 , 1.1893)	0.3794	0.9782
0.4	0.2204	0	56	53 (1.4393)	0.9979 (0.1486)	-0.1988 (0.0539)	– –	0.5013	0.4568 (0.1413)	(0.7725 , 1.2233)	0.4508	0.9988
0.4	0.2204	1	56	52 (2.2060)	1.0025 (0.2804)	-0.1987 (0.0733)	0.0981 (0.1147)	0.5013	0.4418 (0.1824)	(0.7533 , 1.2517)	0.4984	0.9855
Discrete												
0.3	0.2187	0	32	40 (0.8261)	0.9896 (0.1590)	-0.1953 (0.0752)	– –	0.3760	0.3570 (0.1062)	(0.7866 , 1.1927)	0.4061	0.9876
0.3	0.2187	1	32	48 (1.2895)	0.9909 (0.2255)	-0.1954 (0.0820)	0.0994 (0.0680)	0.3760	0.3731 (0.1151)	(0.7803 , 1.2015)	0.4212	0.9517
0.4	0.2916	0	32	40 (0.8871)	0.9913 (0.2159)	-0.1989 (0.1033)	– –	0.5013	0.4711 (0.1428)	(0.7201 , 1.2624)	0.5423	0.9959
0.4	0.2916	1	32	48 (1.3299)	0.9749 (0.3133)	-0.1933 (0.1128)	0.1048 (0.0927)	0.5013	0.4971 (0.1531)	(0.6950 , 1.2548)	0.5598	0.9509
0.3	0.1653	0	56	56 (1.4688)	0.9995 (0.1363)	-0.1992 (0.0665)	– –	0.3760	0.3309 (0.1088)	(0.8329 , 1.1662)	0.3333	0.9948
0.3	0.1653	1	56	72 (2.3661)	0.9984 (0.2074)	-0.2000 (0.0732)	0.1017 (0.0631)	0.3760	0.3456 (0.1145)	(0.8188 , 1.1780)	0.3592	0.9684
0.4	0.2204	0	56	56 (1.4287)	0.9947 (0.1913)	-0.1954 (0.0913)	– –	0.5013	0.4404 (0.1396)	(0.7676 , 1.2217)	0.4541	0.9886
0.4	0.2204	1	56	72 (2.2734)	0.9923 (0.2614)	-0.1948 (0.0946)	0.0974 (0.0802)	0.5013	0.4621 (0.1533)	(0.7594 , 1.2252)	0.4658	0.9606

NOTE:  $\sigma_*$ : standard deviation of the error term; d: precision parameter; Batch (takes the value 1 and 0 according to whether the batch effect is included or not);  $n_d = z_{1-\alpha/2}^2 \sigma^2 / d^2$  is the theoretical stopping variable value in the continuous case and for the discrete case this is the sample size to calculate  $I_{max}$  and the corresponding critical values;  $\sigma = \sigma_{1/2} = 1/2 f(0)$  is the asymptotic standard deviation of the intercept regression estimator;  $W_{N_d}$ : kernel regression estimator of  $\sigma$ ;  $CI$ : mean 99.9% Confidence Interval of the intercept;  $CIL$ : Length of the CI;  $CP$ : Coverage Probability. The results are based on the simulation study with 1,000 replications.

Table 3: Simulation results with Laplace error in regression model (Standard deviations are in parenthesis)

$\sigma_*$	d	Batch	nd	Median $N_d$	Mean Intercept True=1	Mean Slope True= -0.2	Mean Batch True=0.1	$\sigma$ = $\sigma_{1/2}$	Mean $W_{N_d}$	CI 99.9%	CIL	CP
Continuous												
0.3	0.1234	0	32	31 (0.8613)	0.9948 (0.0656)	-0.1968 (0.0252)	– –	0.2121	0.1967 (0.0685)	(0.8819 , 1.1078)	0.2260	1.0000
0.3	0.1234	1	32	33 (0.9950)	0.9874 (0.1316)	-0.1969 (0.0332)	0.1085 (0.0620)	0.2121	0.1938 (0.0805)	(0.8626 , 1.1122)	0.2496	0.9844
0.4	0.1645	0	32	31 (0.8309)	0.9861 (0.0937)	-0.1931 (0.0361)	– –	0.2828	0.2562 (0.0903)	(0.8343 , 1.1379)	0.3036	1.0000
0.4	0.1645	1	32	33 (1.0535)	0.9711 (0.1671)	-0.1936 (0.0431)	0.1117 (0.0779)	0.2828	0.2603 (0.1083)	(0.8089 , 1.1333)	0.3244	0.9847
0.3	0.0933	0	56	48 (1.3714)	0.9900 (0.0577)	-0.1955 (0.0225)	– –	0.2121	0.1887 (0.0649)	(0.8983 , 1.0818)	0.1836	0.9990
0.3	0.0933	1	56	49 (2.4277)	0.9833 (0.1112)	-0.1970 (0.0313)	0.1098 (0.0480)	0.2121	0.1843 (0.0812)	(0.8779 , 1.0888)	0.2109	0.9702
0.4	0.1240	0	56	49 (1.5013)	0.9882 (0.0768)	-0.1943 (0.0311)	– –	0.2828	0.2507 (0.0877)	(0.8671 , 1.1094)	0.2423	1.0000
0.4	0.1240	1	56	44 (2.2752)	0.9858 (0.1418)	-0.1958 (0.0411)	0.1074 (0.0684)	0.2828	0.2431 (0.1087)	(0.8527 , 1.1190)	0.2663	0.9796
Discrete												
0.3	0.1234	0	32	40 (0.9797)	0.9934 (0.0764)	-0.1958 (0.0375)	– –	0.2121	0.2020 (0.0676)	(0.8870 , 1.0997)	0.2127	0.9959
0.3	0.1234	1	32	48 (1.3594)	0.9842 (0.1119)	-0.1946 (0.0446)	0.1047 (0.0357)	0.2121	0.2111 (0.0739)	(0.8703 , 1.0981)	0.2278	0.9692
0.4	0.1645	0	32	40 (0.9060)	0.9860 (0.1019)	-0.1908 (0.0529)	– –	0.2828	0.2640 (0.0897)	(0.8457 , 1.1263)	0.2806	0.9948
0.4	0.1645	1	32	48 (1.3292)	0.9771 (0.1555)	-0.1906 (0.0615)	0.1060 (0.0478)	0.2828	0.2785 (0.1015)	(0.8183 , 1.1359)	0.3176	0.9571
0.3	0.0933	0	56	64 (1.8276)	0.9907 (0.0677)	-0.1943 (0.0340)	– –	0.2121	0.1913 (0.0679)	(0.9015 , 1.0800)	0.1785	0.9907
0.3	0.0933	1	56	72 (2.4965)	0.9879 (0.1011)	-0.1965 (0.0397)	0.1042 (0.0353)	0.2121	0.2004 (0.0758)	(0.8890 , 1.0869)	0.1979	0.9638
0.4	0.1240	0	56	56 (1.5957)	0.9846 (0.0878)	-0.1917 (0.0431)	– –	0.2828	0.2467 (0.0867)	(0.8678 , 1.1013)	0.2335	0.9897
0.4	0.1240	1	56	72 (2.4501)	0.9746 (0.1356)	-0.1904 (0.0537)	0.1060 (0.0479)	0.2828	0.2685 (0.0977)	(0.8399 , 1.1092)	0.2693	0.9590

NOTE:  $\sigma_*$ : standard deviation of the error term; d: precision parameter; Batch (takes the value 1 and 0 according to whether the batch effect is included or not);  $n_d = \frac{z_{1-\alpha/2}^2 \sigma^2}{d^2}$  is the theoretical stopping variable value in the continuous case and for the discrete case this is the sample size to calculate  $I_{max}$  and the corresponding critical values;  $\sigma = \sigma_{1/2} = \frac{1}{2f(0)}$  is the asymptotic standard deviation of the intercept regression estimator;  $W_{N_d}$ : kernel regression estimator of  $\sigma$ ; CI: mean 99.9% Confidence Interval of the intercept; CIL: Length of the CI; CP: Coverage Probability. The results are based on the simulation study with 1,000 replications.

# Proof of the uniform continuity in probability condition of least-absolute regression estimator and asymptotic kernel variance estimator

We recall that an infinite sequence of random variables  $\{Y_n\}_{n \geq 1}$  is uniformly continuous in probability  $\iff \forall \epsilon > 0 \forall \eta > 0 \exists \nu \exists c > 0$  such that  $\forall n > \nu$

$$P \left\{ \max_{|n'-n| < cn} |Y_{n'} - Y_n| > \epsilon \right\} < \eta.$$

Furthermore, adopt the following regularity conditions related to the distribution function  $F$  (C1–C3), the sequence of matrices  $\mathbf{X}_n$  (C4–C7) and the estimators (C8–C10):

- $\mathbf{D}_n = n^{-1} \mathbf{X}'_n \mathbf{X}_n \equiv (d_{ij}^n)_{1 \leq i, j \leq p} \equiv (d_{ij})_{1 \leq i, j \leq p}$  and  $\mathbf{D}_n^{-1} \equiv (d_n^{ij})_{1 \leq i, j \leq p} \equiv (d^{ij})_{1 \leq i, j \leq p}$  the inverse matrix, which is supposed to exist.
- $\theta$  be a real number in  $(0, 1)$ .
- $\psi_\theta(x) = \theta - \mathbb{I}(x < 0)$  and  $\rho_\theta(x) = x\psi_\theta(x)$ , where  $\mathbb{I}(\mathcal{P})$  takes the value 1 and 0 depending on whether the condition  $\mathcal{P}$  is satisfied or not.
- $Q$  be the *quantile function* of  $\varepsilon_1$  assuming that  $Q(1/2) = 0$ .
- $q$  be the *quantile density function*, also known as the *sparsity function*, defined by

$$q(\theta) = Q'(\theta) = \frac{1}{f(Q(\theta))}.$$

- $\sigma_\theta = \sqrt{\theta(1-\theta)} q(\theta)$  and  $\sigma = \sigma_{1/2}$ .
- $z_{1-\alpha/2}$  be the  $\alpha$ -quantile of a  $N(0, 1)$  distribution.

- $\|\cdot\|$  be the Euclidean norm.

*C1.*  $F$  is twice continuously differentiable and its derivative function  $f$  is positive in  $\{x : 0 < F(x) < 1\}$ .

*C2.* There exist  $c_2 > 0$ ,  $\theta_0 \in (0, 1/2)$  and  $a \in (0, 1/4)$  such that  $q(\theta) \leq \frac{c_2}{(\theta(1-\theta))^{1+a}}$  for  $\theta \in (0, \theta_0) \cup (1 - \theta_0, 1)$ .

*C3.* There exist  $c_3$  and  $K$  positive numbers such that  $\left| \frac{f'(x)}{f(x)} \right| \leq c_3 |x|$  for  $|x| \geq K$ .

*C4.* Let  $a$  be the parameter introduced in condition *C2*. We assume  $\max_{1 \leq i \leq n} \|\mathbf{x}_i\| = O(n^{(2(b-a)-\delta)/(1+4b)})$  for some  $b > 0$  and  $\delta > 0$  such that  $0 < b - a < \epsilon/2$  as  $n \rightarrow \infty$ . From now on,  $\theta_n^* = n^{-1/(2(1+4b))}$ .

*C5.*  $\lim_{n \rightarrow \infty} \mathbf{D}_n = \mathbf{D}$  a positive definite matrix, (whence  $\mathbf{D}_n$  is non singular when  $n$  is large enough).

*C6.*  $n^{-1} \sum_{i=1}^n \|\mathbf{x}_i\|^4 = O(1)$  as  $n \rightarrow \infty$ .

*C7.*  $x_{i1} = 1$  for  $i = 1, \dots, n$ .

Let  $(\nu_n)_{n \geq 1}$  be a decreasing sequence of positive numbers (*bandwidth*) satisfying one or more of the following conditions:

*C8.*  $n\nu_n \rightarrow \infty$  as  $n \rightarrow \infty$ .

*C8'.*  $n\nu_n^2 \rightarrow \infty$  as  $n \rightarrow \infty$ .

*C9.*  $n\nu_n^3 \rightarrow 0$  as  $n \rightarrow \infty$ .

Our last condition relates to the kernel function.

C10. Let  $k$  denotes a continuous function with compact support such that  $\int k(v) dv = 0$  and

$$\int v k(v) dv = -1 ; \text{ we denote } K(x) = \int_{-\infty}^x k(y) dy \text{ and } \bar{K} = \int K^2(x) dx.$$

Condition C2 is used to control the heaviness of the tail of the distributions. All these conditions are realistic in many biological microarray experiments.

**Theorem 1.**

(i) Suppose that the conditions C1 – C7 are satisfied. Then the sequence

$$\left( \sqrt{n} \left( \widehat{\beta}_1^n(\theta) - \beta_1 - Q(\theta) \right) \right)_{n \geq 1}$$

is uniformly continuous in probability.

(ii) Suppose that the conditions C1 – C10 are satisfied. Then the sequence

$$\left( \sqrt{n} \nu_n \left( \widehat{W}_n(\theta) - \frac{q(\theta)}{2} \right) \right)_{n \geq 1}$$

is uniformly continuous in probability, where  $\widehat{W}_n(\theta)$  is the chosen estimator of  $q(\theta)/2$ .

The development of Gutenbrunner and Jurečková (1992) and Jurečková (1984) is not sufficient in our case. We need further results and prove that the remainder term in their development remains small when taking the supremum in the decomposition : this is the object of Lemma 1 and Theorem 2, on which the proof of Theorem 1 is based. We begin by

**Lemma 1.** Let  $C$  be a given constant,  $\gamma \in (0, a)$  (a defined in condition C2) and

$$B_n = \max \left( \frac{1}{n^{2(a-\gamma)/(1+4b)}}, \frac{1}{n^{(2-\gamma)(b-a)/(1+4b)}}, \frac{1}{n^{(b-\gamma)/(1+4b)}} \right). \quad (1)$$

Under assumptions C1 – C7, for every  $\mu > 0$ , there is  $A_\mu > 0$  such that for  $n \rightarrow \infty$ , we have

$$P \left\{ \sup |r_n(\boldsymbol{\beta}, \theta)| \geq A_\mu B_n : \|\boldsymbol{\beta}\| \leq C \sqrt{\log n}, \theta_n^* \leq \theta \leq 1 - \theta_n^* \right\} = O \left( \frac{1}{n^\mu} \right) \quad (2)$$

where

$$\begin{aligned} r_n(\boldsymbol{\beta}, \theta) &= \frac{1}{\sqrt{\theta(1-\theta)} \sigma_\theta} \sum_{i=1}^n \left[ \rho_\theta \left( \varepsilon_{i\theta} - \frac{\sigma_\theta \mathbf{x}_i' \boldsymbol{\beta}}{\sqrt{n}} \right) - \rho_\theta(\varepsilon_{i\theta}) \right] \\ &\quad + \frac{1}{\sqrt{n}(\theta(1-\theta))} \sum_{i=1}^n \mathbf{x}_i' \boldsymbol{\beta} \psi_\theta(\varepsilon_{i\theta}) - \frac{\boldsymbol{\beta}' \mathbf{D}_n \boldsymbol{\beta}}{2} \end{aligned}$$

and  $\varepsilon_{i\theta} = \varepsilon_i - Q(\theta)$  for  $i = 1, \dots, n$ .

**Proof.** We decompose the proof into two steps. In the first one we show that for any  $\lambda > 0$  we have

$$P \{ |r_n(\boldsymbol{\beta}, \theta)| \geq (\lambda + 1) B_n \} \leq \frac{1}{n^\lambda}. \quad (3)$$

In the second one, we show how to deduce (2) from (3).

We will denote  $C_n = C \sqrt{\log n}$ .

**Step I.**

First, we choose  $\theta \in [\theta_n^*, 1 - \theta_n^*]$  and  $\boldsymbol{\beta}$  such that  $\|\boldsymbol{\beta}\| \leq C_n$ . The Markov inequality implies

$$P \{ |r_n(\boldsymbol{\beta}, \theta)| \geq s_n \} \leq \exp(-us_n) M_{r_n}(u)$$

where  $M_{r_n}(u) = E(\exp(u |r_n(\boldsymbol{\beta}, \theta)|))$ , see Rogers and Williams (1994), p. 111.

Denote

$$\zeta_{ni} \equiv \zeta_{ni}(\boldsymbol{\beta}, \theta) = \frac{\sigma_\theta \mathbf{x}_i' \boldsymbol{\beta}}{\sqrt{n}}$$

and

$$R_i(\boldsymbol{\beta}, \theta) = \frac{1}{\sqrt{\theta(1-\theta)}\sigma_\theta} \left[ \rho_\theta \left( \varepsilon_{i\theta} - \frac{\sigma_\theta \mathbf{x}'_i \boldsymbol{\beta}}{\sqrt{n}} \right) - \rho_\theta(\varepsilon_{i\theta}) \right] \\ + \frac{1}{\sqrt{n}\theta(1-\theta)} \mathbf{x}'_i \boldsymbol{\beta} \psi_\theta(\varepsilon_{i\theta}) - \frac{(\mathbf{x}'_i \boldsymbol{\beta})^2}{2n}, \quad i = 1, \dots, n.$$

Replacing  $\rho_\theta(x) = |x| \{(1-\theta)\mathbb{I}(x < 0) + \theta\mathbb{I}(x > 0)\}$  in the previous equation, we have

$$R_i(\boldsymbol{\beta}, \theta) + \frac{(\mathbf{x}'_i \boldsymbol{\beta})^2}{2n} = \frac{1}{\sqrt{\theta(1-\theta)}\sigma_\theta} [(\varepsilon_{i\theta} - \zeta_{ni}) \mathbb{I}(\zeta_{ni} < \varepsilon_{i\theta} < 0) \\ + (\zeta_{ni} - \varepsilon_{i\theta}) \mathbb{I}(0 < \varepsilon_{i\theta} < \zeta_{ni})]. \quad (4)$$

So that for  $\zeta_{ni} > 0$  we have

$$R_i(\boldsymbol{\beta}, \theta) + \frac{(\mathbf{x}'_i \boldsymbol{\beta})^2}{2n} = \frac{1}{\sqrt{\theta(1-\theta)}\sigma_\theta} (\zeta_{ni} - \varepsilon_{i\theta}) \mathbb{I}(0 < \varepsilon_{i\theta} < \zeta_{ni}) \quad (5)$$

and for  $\zeta_{ni} < 0$  the similar relation holds

$$R_i(\boldsymbol{\beta}, \theta) + \frac{(\mathbf{x}'_i \boldsymbol{\beta})^2}{2n} = \frac{1}{\sqrt{\theta(1-\theta)}\sigma_\theta} (\varepsilon_{i\theta} - \zeta_{ni}) \mathbb{I}(\zeta_{ni} < \varepsilon_{i\theta} < 0). \quad (6)$$

**Remark.** Due to the similarity of relation (5) and (6), we will only treat the case of  $\zeta_{ni} > 0$ .

We have, for  $i = 1, \dots, n$ ,

$$\left| R_i(\boldsymbol{\beta}, \theta) + \frac{(\mathbf{x}'_i \boldsymbol{\beta})^2}{2n} \right| \leq \frac{1}{\sqrt{\theta(1-\theta)}\sqrt{n}} |\mathbf{x}'_i \boldsymbol{\beta}| \\ \leq \frac{1}{\sqrt{\theta(1-\theta)}\sqrt{n}} \max_{i=1, \dots, n} |\mathbf{x}'_i \boldsymbol{\beta}|. \quad (7)$$

We also have,

$$\frac{1}{\sqrt{\theta(1-\theta)}} \leq \frac{1}{\theta_n^*},$$

so that by condition  $C4$  , we obtain, as  $n \rightarrow \infty$

$$\begin{aligned} \left| R_i(\boldsymbol{\beta}, \theta) + \frac{(\mathbf{x}'_i \boldsymbol{\beta})^2}{2n} \right| &= O\left( \frac{n^{1/(2(1+4b))} C_n n^{(2(b-a)-\delta)/(1+4b)}}{\sqrt{n}} \right) \\ &= O\left( \frac{C_n}{n^{(2a+\delta)/(1+4b)}} \right), \end{aligned}$$

where the constant implied by the  $O$  symbol is uniform in  $\boldsymbol{\beta}$  and  $\theta$ .

From now on,  $c$  denotes an absolute positive constant. This proof requires some preliminary results :

**Lemma 2.** *Suppose that the conditions  $C1 - C7$  are satisfied. Then, for  $0 < u < \frac{n^{(2a+\delta)/(1+4b)}}{\sqrt{\log n}}$ , there is a positive constant  $c$  such that*

$$\log M_{R_i}(u) \leq uE(|R_i(\boldsymbol{\beta}, \theta)|) + cu^2 \text{Var}(R_i(\boldsymbol{\beta}, \theta)).$$

**Lemma 3.** *Under assumptions  $C1 - C7$ , as  $n \rightarrow \infty$ , we have*

$$\sum_{i=1}^n E(|R_i(\boldsymbol{\beta}, \theta)|) = O\left( \frac{1}{n^{2(b-a)/(1+4b)}} \right), \quad (8)$$

*uniformly for  $\theta \in (\theta_n^*, 1 - \theta_n^*)$  and  $\|\boldsymbol{\beta}\| \leq C_n$ .*

**Lemma 4.** *Under assumptions  $C1 - C7$ , as  $n \rightarrow \infty$ , we have*

$$\sum_{i=1}^n \text{Var}(R_i(\boldsymbol{\beta}, \theta)) = O\left( \frac{1}{n^{2b/(1+4b)}} \right)$$

*uniformly for  $\theta \in (\theta_n^*, 1 - \theta_n^*)$  and  $\|\boldsymbol{\beta}\| \leq C_n$ .*

**Remark.** An appendix sketches the proofs of these three lemmas.

We have, see Lemma 2,

$$\log(M_{r_n}(u)) \leq u \mathbb{E} \left( \sum_{i=1}^n |R_i(\boldsymbol{\beta}, \theta)| \right) + c u^2 \text{Var} \left( \sum_{i=1}^n R_i(\boldsymbol{\beta}, \theta) \right)$$

where  $r_n = \sum_{i=1}^n R_i$  and so

$$M_{r_n}(u) \leq \exp \left( u \sum_{i=1}^n \mathbb{E} (|R_i(\boldsymbol{\beta}, \theta)|) + c u^2 \sum_{i=1}^n \text{Var} (R_i(\boldsymbol{\beta}, \theta)) \right).$$

Let  $\lambda$  be any positive real number and let  $s_n = (\lambda + 1) B_n$  and  $u = \log n / B_n$ . We have  $u = O(n^{2a/(1+4b)})$  and so we may apply Lemma 2. It then follows from Lemmas 3 and 4 that there is a rank  $n_0$  independent of  $\theta$  and  $\boldsymbol{\beta}$  such that for all  $n \geq n_0$ ,

$$P \{ |r_n(\boldsymbol{\beta}, \theta)| \geq (\lambda + 1) B_n \} \leq \exp \left( -(\lambda + 1) \log n + K \frac{\log(n)}{B_n} \frac{1}{n^{2(b-a)/(1+4b)}} + K^2 \left( \frac{\log(n)}{B_n} \right)^2 \frac{1}{n^{2b/(1+4b)}} \right) \leq$$

$$\exp \left( -\lambda \log n - \log n + K \frac{\log n}{n^{\gamma(b-a)/(1+4b)}} + K \frac{\log^2 n}{n^{2\gamma/(1+4b)}} \right) \leq \frac{1}{n^\lambda}$$

for some  $K > 0$ . Thus, we proved (3).

## Step II.

In this step, we prove (2).

We choose  $n^5$  intervals of length  $1/n^5$  included in and covering  $[\theta_n^*, 1 - \theta_n^*]$  and  $O(n^{5p} \log n)$  balls of radius  $1/n^5$  included in and covering the ball  $\{\boldsymbol{\beta} : \|\boldsymbol{\beta}\| \leq C_n\}$ . We want to bound

$\Delta_i \equiv |R_i(\boldsymbol{\beta}_1, \theta_1) - R_i(\boldsymbol{\beta}_2, \theta_2)|$ . So, we define for  $l = 1, 2$ , the intervals  $J_l$  with end points  $Q(\theta_l)$  and  $\zeta_{ni}(\boldsymbol{\beta}_l, \theta_l) + Q(\theta_l)$ .

We further define

$$G_l(\varepsilon_i) = \frac{f(Q(\theta_l))}{\theta_l(1-\theta_l)} (\varepsilon_i - Q(\theta_l) - \zeta_{ni}(\boldsymbol{\beta}_l, \theta_l)). \quad (9)$$

Using (4), we have

$$\begin{aligned} R_i(\boldsymbol{\beta}_1, \theta_1) - R_i(\boldsymbol{\beta}_2, \theta_2) &= -\frac{(\mathbf{x}'_i \boldsymbol{\beta}_1)^2}{2n} + \frac{(\mathbf{x}'_i \boldsymbol{\beta}_2)^2}{2n} \\ &+ \frac{1}{\sqrt{\theta_1(1-\theta_1)} \sigma_{\theta_1}} [(\varepsilon_{i\theta_1} - \zeta_{ni}(\boldsymbol{\beta}_1, \theta_1)) \mathbb{I}(\zeta_{ni}(\boldsymbol{\beta}_1, \theta_1) < \varepsilon_{i\theta_1} < 0) \\ &+ (\zeta_{ni}(\boldsymbol{\beta}_1, \theta_1) - \varepsilon_{i\theta_1}) \mathbb{I}(0 < \varepsilon_{i\theta_1} < \zeta_{ni}(\boldsymbol{\beta}_1, \theta_1))] \\ &- \frac{1}{\sqrt{\theta_2(1-\theta_2)} \sigma_{\theta_2}} [(\varepsilon_{i\theta_2} - \zeta_{ni}(\boldsymbol{\beta}_2, \theta_2)) \mathbb{I}(\zeta_{ni}(\boldsymbol{\beta}_2, \theta_2) < \varepsilon_{i\theta_2} < 0) \\ &+ (\zeta_{ni}(\boldsymbol{\beta}_2, \theta_2) - \varepsilon_{i\theta_2}) \mathbb{I}(0 < \varepsilon_{i\theta_2} < \zeta_{ni}(\boldsymbol{\beta}_2, \theta_2))] \end{aligned}$$

with  $\varepsilon_{i\theta} = \varepsilon_i - Q(\theta)$ .

Thus

- if  $\zeta_{ni}(\boldsymbol{\beta}_1, \theta_1) > 0$  and  $\zeta_{ni}(\boldsymbol{\beta}_2, \theta_2) > 0$ , we have for  $\varepsilon_i \in (J_1 \cap J_2)$

$$R_i(\boldsymbol{\beta}_1, \theta_1) - R_i(\boldsymbol{\beta}_2, \theta_2) = -\frac{(\mathbf{x}'_i \boldsymbol{\beta}_1)^2}{2n} + \frac{(\mathbf{x}'_i \boldsymbol{\beta}_2)^2}{2n} + (G_2(\varepsilon_i) - G_1(\varepsilon_i)).$$

- If  $\zeta_{ni}(\boldsymbol{\beta}_1, \theta_1) < 0$  and  $\zeta_{ni}(\boldsymbol{\beta}_2, \theta_2) < 0$ , we have for  $\varepsilon_i \in (J_1 \cap J_2)$

$$R_i(\boldsymbol{\beta}_1, \theta_1) - R_i(\boldsymbol{\beta}_2, \theta_2) = -\frac{(\mathbf{x}'_i \boldsymbol{\beta}_1)^2}{2n} + \frac{(\mathbf{x}'_i \boldsymbol{\beta}_2)^2}{2n} + (G_1(\varepsilon_i) - G_2(\varepsilon_i)).$$

- If  $\zeta_{ni}(\boldsymbol{\beta}_1, \theta_1)$  and  $\zeta_{ni}(\boldsymbol{\beta}_2, \theta_2)$  have different signs, we have for  $\varepsilon_i \in (J_1 \cap J_2)$

$$R_i(\boldsymbol{\beta}_1, \theta_1) - R_i(\boldsymbol{\beta}_2, \theta_2) = -\frac{(\mathbf{x}'_i \boldsymbol{\beta}_1)^2}{2n} + \frac{(\mathbf{x}'_i \boldsymbol{\beta}_2)^2}{2n} + \Delta_i^*$$

where

$$\Delta_i^* = \frac{1}{\sqrt{\theta_1(1-\theta_1)}\sigma_{\theta_1}} (\zeta_{ni}(\boldsymbol{\beta}_1, \theta_1) - \varepsilon_{i\theta_1}) - \frac{1}{\sqrt{\theta_2(1-\theta_2)}\sigma_{\theta_2}} (\varepsilon_{i\theta_2} - \zeta_{ni}(\boldsymbol{\beta}_2, \theta_2))$$

if  $\zeta_{ni}(\boldsymbol{\beta}_1, \theta_1) > 0$  and

$$\Delta_i^* = -\frac{1}{\sqrt{\theta_1(1-\theta_1)}\sigma_{\theta_1}} (\varepsilon_{i\theta_1} - \zeta_{ni}(\boldsymbol{\beta}_1, \theta_1)) + \frac{1}{\sqrt{\theta_2(1-\theta_2)}\sigma_{\theta_2}} (\zeta_{ni}(\boldsymbol{\beta}_2, \theta_2) - \varepsilon_{i\theta_2})$$

if  $\zeta_{ni}(\boldsymbol{\beta}_1, \theta_1) < 0$ .

Therefore

$$\Delta_i = \left| -\frac{(\mathbf{x}'_i \boldsymbol{\beta}_1)^2}{2n} + \frac{(\mathbf{x}'_i \boldsymbol{\beta}_2)^2}{2n} + H(\varepsilon_i) \right|$$

where

$$H(\varepsilon_i) = \begin{cases} G_2(\varepsilon_i) - G_1(\varepsilon_i) & \text{if } \zeta_{ni}(\boldsymbol{\beta}_1, \theta_1) > 0 \text{ and } \zeta_{ni}(\boldsymbol{\beta}_2, \theta_2) > 0 \\ G_1(\varepsilon_i) - G_2(\varepsilon_i) & \text{if } \zeta_{ni}(\boldsymbol{\beta}_1, \theta_1) < 0 \text{ and } \zeta_{ni}(\boldsymbol{\beta}_2, \theta_2) < 0 \\ \Delta_i^* & \text{if } \zeta_{ni}(\boldsymbol{\beta}_1, \theta_1) \text{ and } \zeta_{ni}(\boldsymbol{\beta}_2, \theta_2) \text{ have different signs} \\ 0 & \text{otherwise.} \end{cases}$$

We now study each case separately :

1. When  $\zeta_{ni}(\boldsymbol{\beta}_1, \theta_1)$  and  $\zeta_{ni}(\boldsymbol{\beta}_2, \theta_2)$  have different signs, we have

$$\Delta_i \leq \left| -\frac{(\mathbf{x}'_i \boldsymbol{\beta}_1)^2}{2n} + \frac{(\mathbf{x}'_i \boldsymbol{\beta}_2)^2}{2n} \right| + |\Delta_i^*|,$$

where

$$|\Delta_i^*| \leq \max_{l=1,2} \left( \frac{f(Q(\theta_l)) |\zeta_{ni}(\boldsymbol{\beta}_l, \theta_l)|}{\theta_l(1-\theta_l)} \right) = \max_{l=1,2} \left( \frac{|\mathbf{x}'_i \boldsymbol{\beta}_l|}{\sqrt{n} \sqrt{\theta_l(1-\theta_l)}} \right)$$

$$\leq \frac{C_n |\mathbf{x}_i|}{\sqrt{n} \sqrt{\theta_l(1-\theta_l)}} \leq \frac{C_n n^{2(b-a)/(1+4b)}}{\sqrt{n} \sqrt{\theta_l(1-\theta_l)}}$$

and, as  $\theta_n^* \leq \theta \leq 1 - \theta_n^*$ , we have

$$|\Delta_i^*| \leq C_n \frac{n^{2(b-a)/(1+4b)} n^{1/(2(1+4b))}}{\sqrt{n}} = \frac{C_n}{n^{2a/(1+4b)}} \leq B_n,$$

when  $n$  is sufficiently large.

2. When  $\zeta_{ni}(\boldsymbol{\beta}_1, \theta_1)$  and  $\zeta_{ni}(\boldsymbol{\beta}_2, \theta_2)$  have the same signs, we have

$$\Delta_i \leq \left| -\frac{(\mathbf{x}'_i \boldsymbol{\beta}_1)^2}{2n} + \frac{(\mathbf{x}'_i \boldsymbol{\beta}_2)^2}{2n} \right| + |G_1(\varepsilon_i) - G_2(\varepsilon_i)|.$$

Using the definition of  $G_l$  ( $l = 1, 2$ ), see (9), we obtain

$$\begin{aligned} \Delta_i &\leq \left| -\frac{(\mathbf{x}'_i \boldsymbol{\beta}_1)^2}{2n} + \frac{(\mathbf{x}'_i \boldsymbol{\beta}_2)^2}{2n} \right| + \max_{l=1,2} \left\{ \frac{f(Q(\theta_l))}{\theta_l(1-\theta_l)} \right\} [|Q(\theta_2) - Q(\theta_1)| \\ &\quad + |\zeta_{ni}(\boldsymbol{\beta}_1, \theta_1) - \zeta_{ni}(\boldsymbol{\beta}_2, \theta_2)|]. \end{aligned}$$

Before studying each term separately, we give some preliminary elementary results. For

$\theta_n^* \leq \theta \leq 1 - \theta_n^*$ , we have by condition C2

$$\frac{1}{\theta(1-\theta)} \leq \frac{1}{(\theta_n^*)^2} \leq n$$

and

$$\frac{1}{\min_{\theta} (f(Q(\theta)))} \leq \frac{c}{(\theta(1-\theta))^{1+a}} \leq n^{1+a},$$

where  $a < 1/4$ . Therefore, we have

$$\frac{1}{\min_{\theta} (f(Q(\theta)))} \leq n^{1+1/4} = n^{5/4}.$$

- Study of  $|Q(\theta_1) - Q(\theta_2)|$ .

By the mean value theorem, there is  $\theta^* \in [\theta_1, \theta_2]$ , such that

$$\begin{aligned} |Q(\theta_1) - Q(\theta_2)| &= |(\theta_1 - \theta_2) Q'(\theta^*)| = \left| \frac{\theta_1 - \theta_2}{f(Q(\theta^*))} \right| \\ &\leq \left| \frac{\theta_1 - \theta_2}{\min_{\theta} (f(Q(\theta)))} \right| \leq |\theta_1 - \theta_2| n^{5/4}. \end{aligned}$$

We remember that we fixed intervals of length  $1/n^5$  included in and covering  $[\theta_n^*, 1 - \theta_n^*]$  and balls of radius  $1/n^5$  included in and covering the space  $\{\boldsymbol{\beta} : \|\boldsymbol{\beta}\| \leq C_n\}$ . Given this choice we have

$$|\theta_1 - \theta_2| \leq \frac{1}{n^5} \quad \text{and} \quad \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\| \leq \frac{1}{n^5}.$$

Therefore

$$|Q(\theta_1) - Q(\theta_2)| \leq \frac{1}{n^5} n^{5/4} = \frac{1}{n^{15/4}}. \quad (10)$$

- Study of  $|\zeta_{ni}(\boldsymbol{\beta}_1, \theta_1) - \zeta_{ni}(\boldsymbol{\beta}_2, \theta_2)|$ .

We note that for  $\|\boldsymbol{\beta}\| \leq C_n$  and  $\theta_n^* \leq \theta \leq 1 - \theta_n^*$

$$|\zeta_{ni}(\boldsymbol{\beta}_1, \theta_1) - \zeta_{ni}(\boldsymbol{\beta}_2, \theta_2)| = \left| \frac{\sigma_{\theta_1} \mathbf{x}'_i \boldsymbol{\beta}_1}{\sqrt{n}} - \frac{\sigma_{\theta_2} \mathbf{x}'_i \boldsymbol{\beta}_2}{\sqrt{n}} \right|,$$

whence

$$\begin{aligned}
|\zeta_{ni}(\boldsymbol{\beta}_1, \theta_1) - \zeta_{ni}(\boldsymbol{\beta}_2, \theta_2)| &\leq \frac{C_n \|\mathbf{x}_i\| \left| \sqrt{\theta_1(1-\theta_1)} - \sqrt{\theta_2(1-\theta_2)} \right|}{\sqrt{n} \min_{\theta} (f(Q(\theta)))} \\
&+ \frac{C_n \|\mathbf{x}_i\|}{\sqrt{n}} \left| \frac{1}{f(Q(\theta_1))} - \frac{1}{f(Q(\theta_2))} \right| + \frac{c \|\mathbf{x}_i\| \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|}{\sqrt{n} \min_{\theta} (f(Q(\theta)))} \\
&\leq C_n c \frac{n^{(2(b-a)-\delta)/(1+4b)} n^{2(1+a)/(1+4b)}}{\sqrt{n} n^5} + C_n c \frac{n^{(2(b-a)-\delta)/(1+4b)}}{\sqrt{n} n^{15/4}} \\
&+ c \frac{n^{(2(b-a)-\delta)/(1+4b)} n^{(1+a)/(1+4b)}}{\sqrt{n} n^5} = o\left(\frac{B_n}{n}\right).
\end{aligned}$$

Let  $\mathcal{I} \times \mathcal{B}$  be the product of a “small” interval in  $(\theta_n^*, 1 - \theta_n^*)$  and of a “small” ball in the space of the  $\boldsymbol{\beta}$ 's. For those  $i$ 's for which  $\epsilon_i \notin Q(\mathcal{I})$ , we have  $\Delta_i = o(B_n/n)$ ; thus their total contribution is  $o(B_n)$ .

If  $i_1, \dots, i_r$  are the indices for which  $\epsilon_i \in Q(\mathcal{I})$ , we have  $\Delta_i \leq r B_n$  for  $n$  sufficiently large. The quantity  $p = \int_{Q(\mathcal{I})} f(x) dx$  is the probability that  $\epsilon_i \in Q(\mathcal{I})$ ; since the random variables  $\epsilon_i$  are independent, we have

$$\mathbb{P}\{\epsilon_{i_1} \in Q(\mathcal{I}), \dots, \epsilon_{i_r} \in Q(\mathcal{I})\} = p^r.$$

Thus, we have (for  $n$  sufficiently large):

$$\mathbb{P}\left\{\sup_{(\boldsymbol{\beta}_1, \theta_1) \in \mathcal{I} \times \mathcal{B}} \sup_{(\boldsymbol{\beta}_2, \theta_2) \in \mathcal{I} \times \mathcal{B}} \left\{\sum_{i=1}^n \Delta_i\right\} \geq \lambda B_n\right\} \leq \sum_{r=\lambda}^n \binom{n}{r} p^r (1-p)^{n-r}.$$

We have also

$$p = \int_{Q(\theta_1)}^{Q(\theta_2)} f(x) dx.$$

Consequently, using (10), we obtain

$$p \leq \|f\|_\infty \int_{Q(\theta_1)}^{Q(\theta_2)} dx = \|f\|_\infty |Q(\theta_2) - Q(\theta_1)| = O(n^{-3.75}).$$

This implies

$$P \left\{ \sup_{(\boldsymbol{\beta}_1, \theta_1) \in \mathcal{I} \times \mathcal{B}} \sup_{(\boldsymbol{\beta}_2, \theta_2) \in \mathcal{I} \times \mathcal{B}} |r_n(\boldsymbol{\beta}_1, \theta_1) - r_n(\boldsymbol{\beta}_2, \theta_2)| \geq \lambda B_n \right\} \leq \quad (11)$$

$$\sum_{r=\lambda}^n n^r p^r = O((np)^\lambda) = O\left(\frac{1}{n^{2.75\lambda}}\right).$$

Let  $(\boldsymbol{\beta}^{(1)}, \theta^{(1)}), \dots, (\boldsymbol{\beta}^{(N)}, \theta^{(N)})$  be a family of chosen points, each in the small boxes  $\mathcal{I} \times \mathcal{B}$ 's. Their number  $N$  is at most  $n^{5p+6}$ . The probability that  $r_n(\boldsymbol{\beta}^{(k)}, \theta^{(k)})$  is larger than  $(\lambda + 1)B_n$  is at most  $n^{-\lambda}$  (cf. (3)); whence the probability that  $r_n$  is large at one of the points  $(\boldsymbol{\beta}^{(k)}, \theta^{(k)})_{1 \leq k \leq N}$  is at most  $n^{5p+6-\lambda}$ .

Since each pair  $(\boldsymbol{\beta}, \theta)$  with the condition  $\|\boldsymbol{\beta}\| \leq C_n$  and  $\theta_n^* \leq \theta \leq 1 - \theta_n^*$  is in a small box associated to one of the  $(\boldsymbol{\beta}^{(k)}, \theta^{(k)})$ , our second step (11) shows that  $r_n$  has, with a large probability, close values at  $(\boldsymbol{\beta}, \theta)$  and  $(\boldsymbol{\beta}^{(k)}, \theta^{(k)})$ .

Then, we obtain for  $\lambda > 5p + 6$  and sufficiently large  $n$ ,

$$\begin{aligned} & P \{ \sup |r_n(\boldsymbol{\beta}, \theta)| \geq (\lambda + 1)B_n + \lambda B_n : \|\boldsymbol{\beta}\| \leq C_n, \theta_n^* \leq \theta \leq \theta_n^* \} \\ &= O(n^{5p+6-\lambda} + n^{-2.75\lambda}) = O\left(\frac{1}{n^\mu}\right) \end{aligned}$$

since  $\mu = \lambda - (5p + 6) > 0$ , whence the result.

**Theorem 2.** Under conditions C1–C7, we have for every  $\mu > 0$ , there is  $A_\mu$  such that for  $n \rightarrow \infty$ , we have

$$P \left\{ \left\| \frac{\sqrt{n}}{\sigma_\theta} \left( \widehat{\boldsymbol{\beta}}^n(\theta) - \widetilde{\boldsymbol{\beta}}(\theta) \right) - \frac{1}{\sqrt{n} \sqrt{\theta(1-\theta)}} \mathbf{D}_n^{-1} \sum_{i=1}^n \mathbf{x}_i \psi_\theta(\varepsilon_{i\theta}) \right\| \geq A_\mu B_n \right\} = O\left(\frac{1}{n^\mu}\right).$$

**Proof.** As  $\widehat{\boldsymbol{\beta}}^n(\theta)$  minimizes  $\sum_{i=1}^n \rho_\theta(Y_i - \mathbf{x}'_i \boldsymbol{\beta})$  with respect to  $\boldsymbol{\beta}$ , the quantity

$$\mathbf{T}_{n\theta} := \frac{\sqrt{n}}{\sigma_\theta} \left( \widehat{\boldsymbol{\beta}}^n(\theta) - \widetilde{\boldsymbol{\beta}}(\theta) \right) \quad (12)$$

minimizes (with respect to  $\boldsymbol{\beta}$ ) the function

$$G_{n\theta}(\boldsymbol{\beta}) = \frac{1}{\sqrt{\theta(1-\theta)}} \frac{1}{\sigma_\theta} \sum_{i=1}^n \left[ \rho_\theta \left( \varepsilon_{i\theta} - \frac{\sigma_\theta \mathbf{x}'_i \boldsymbol{\beta}}{\sqrt{n}} \right) - \rho_\theta(\varepsilon_{i\theta}) \right].$$

In the sequel  $R_n = O^*(B_n)$  as  $n \rightarrow \infty$ , where  $(R_n)_n$  is a sequence of random variables and  $(B_n)_n$  is a sequence of numbers, means that for any  $\lambda > 0$  there is  $A_\lambda > 0$  and  $n_\lambda$  such that for all  $n \geq n_\lambda$

$$P \{ |R_n| \geq A_\lambda B_n \} \leq n^{-\lambda}.$$

Lemma 1 may be thus rewritten as,

$$\min_{\|\boldsymbol{\beta}\| \leq C_n} G_{n\theta}(\boldsymbol{\beta}) = \min_{\|\boldsymbol{\beta}\| \leq C_n} \left\{ -\boldsymbol{\beta}' \mathbf{Z}_{n\theta} + \frac{\boldsymbol{\beta}' \mathbf{D}_n \boldsymbol{\beta}}{2} \right\} + O^*(B_n)$$

uniformly in  $\theta_n^* \leq \theta \leq 1 - \theta_n^*$ , where

$$\mathbf{Z}_{n\theta} = \frac{1}{\sqrt{n} \sqrt{\theta(1-\theta)}} \sum_{i=1}^n \mathbf{x}_i \psi_\theta(\varepsilon_{i\theta}).$$

It will be necessary to give a probabilistic bound for  $\sup_{\theta_n^* \leq \theta \leq 1 - \theta_n^*} \mathbf{Z}_{n\theta}$ . By conditions C4, C6 and the fact that  $\theta_n^* > n^{-1}$  Theorem 1.2 of Shorack (1991) can be used. We thus have

$$\sup_{\theta_n^* \leq \theta \leq 1 - \theta_n^*} \mathbf{Z}_{n\theta} \leq c \sup_{\theta_n^* \leq s \leq 1 - \theta_n^*} \frac{\mathbf{W}(s)}{\sqrt{s(1-s)}} + O_p(1)$$

for some constant  $c$ , where  $\mathbf{W}(s)$  is a Brownian bridge. We apply the result (15) (page 599) of Shorack & Wellner (1986) with  $d_n = \theta_n^*$ ,  $e_n = 1 - \theta_n^*$  and obtain

$$\sup_{\theta_n^* \leq \theta \leq 1 - \theta_n^*} \mathbf{Z}_{n\theta} = O_p(\sqrt{\log \log n}). \quad (13)$$

Let  $\mathbf{U}_{n\theta} = \mathbf{D}_n^{-1} \mathbf{Z}_{n\theta}$ . We have

$$-\boldsymbol{\beta}' \mathbf{Z}_{n\theta} + \frac{\boldsymbol{\beta}' \mathbf{D}_n \boldsymbol{\beta}}{2} = \frac{(\boldsymbol{\beta} - \mathbf{U}_{n\theta})' \mathbf{D}_n (\boldsymbol{\beta} - \mathbf{U}_{n\theta})}{2} - \frac{\mathbf{U}_{n\theta}' \mathbf{D}_n \mathbf{U}_{n\theta}}{2}.$$

Since  $\mathbf{D}_n$  is positive definite, we have

$$\mathbf{U}_{n\theta} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left\{ -\boldsymbol{\beta}' \mathbf{Z}_{n\theta} + \frac{\boldsymbol{\beta}' \mathbf{D}_n \boldsymbol{\beta}}{2} \right\}$$

and

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left\{ -\boldsymbol{\beta}' \mathbf{Z}_{n\theta} + \frac{\boldsymbol{\beta}' \mathbf{D}_n \boldsymbol{\beta}}{2} \right\} = -\frac{\mathbf{Z}_{n\theta}' \mathbf{D}_n^{-1} \mathbf{Z}_{n\theta}}{2}.$$

We also obtain using (13),

$$\mathbf{U}_{n\theta} = O_p(\sqrt{\log \log n}).$$

uniformly on  $\theta_n^* \leq \theta \leq 1 - \theta_n^*$ .

Hence, applying Lemma 1 we have

$$\sup_{(\theta, \boldsymbol{\beta}) \in S} |r_n(\boldsymbol{\beta}, \theta)| =$$

$$\sup_{(\theta, \boldsymbol{\beta}) \in S} \left\{ \left| G_{n\theta}(\boldsymbol{\beta}) - \frac{(\boldsymbol{\beta} - \mathbf{U}_{n\theta})' \mathbf{D}_n (\boldsymbol{\beta} - \mathbf{U}_{n\theta})}{2} + \frac{\mathbf{U}_{n\theta}' \mathbf{D}_n \mathbf{U}_{n\theta}}{2} \right| \right\} = O^*(B_n),$$

where  $S = \{(\theta, \boldsymbol{\beta}) \in \mathbb{R} \times \mathbb{R}^p; \|\boldsymbol{\beta}\| \leq C \sqrt{\log n}, \theta_n^* \leq \theta \leq 1 - \theta_n^*\}$ .

Inserting  $\mathbf{U}_{n\theta}$  for  $\boldsymbol{\beta}$ , we have

$$\sup_{\theta_n^* \leq \theta \leq 1 - \theta_n^*} \left\{ G_{n\theta}(\mathbf{U}_{n\theta}) + \frac{\mathbf{U}_{n\theta}' \mathbf{D}_n \mathbf{U}_{n\theta}}{2} \right\} = O^*(B_n). \quad (14)$$

which implies that

$$\begin{aligned} \min_{\|\beta\| \leq C_n} G_{n\theta}(\beta) &= \min_{\|\beta\| \leq C_n} \left\{ \frac{(\beta - \mathbf{U}_{n\theta})' \mathbf{D}_n (\beta - \mathbf{U}_{n\theta}) - \mathbf{U}'_{n\theta} \mathbf{D}_n \mathbf{U}_{n\theta}}{2} \right\} \\ &= -\frac{\mathbf{Z}_{n\theta} \mathbf{D}_n^{-1} \mathbf{Z}_{n\theta}}{2}. \end{aligned}$$

Consider the ball  $B_{n\theta}$  with center  $\mathbf{U}_{n\theta}$  and radius  $\sqrt{B_n} |\log B_n| > 0$ . We wish to show that, with high probability the point  $\mathbf{T}_{n\theta}$  (cf. (12)) where the convex function  $G_{n\theta}$  is minimum is in  $B_{n\theta}$ . For that purpose, we show that, with high probability, the value of  $G_{n\theta}$  at  $\mathbf{U}_{n\theta}$  is less than the minimum of  $G_{n\theta}$  on the boundary of  $B_{n\theta}$ . In that case, the convexity of  $G_{n\theta}$  implies that the minimum of  $G_{n\theta}$  is attained inside  $B_{n\theta}$ . Let us denote by  $S_{n\theta}$  the boundary of  $B_{n\theta}$ . By Lemma 1, we have

$$\mathbb{P} \left\{ \sup_{\zeta \in S_{n\theta}} \left| G_{n\theta}(\mathbf{U}_{n\theta} + \zeta) - \frac{\zeta' \mathbf{D}_n \zeta}{2} + \frac{\mathbf{U}'_{n\theta} \mathbf{D}_n \mathbf{U}_{n\theta}}{2} \right| \geq (\lambda + 1) B_n \right\} \leq n^{-\lambda},$$

which leads to

$$\mathbb{P} \left\{ \sup_{\zeta \in S_{n\theta}} \left( G_{n\theta}(\mathbf{U}_{n\theta} + \zeta) - \frac{\zeta' \mathbf{D}_n \zeta}{2} + \frac{\mathbf{U}'_{n\theta} \mathbf{D}_n \mathbf{U}_{n\theta}}{2} \right) \geq (\lambda + 1) B_n \right\} \leq n^{-\lambda}. \quad (15)$$

By (14), we have

$$\mathbb{P} \left\{ \left| G_{n\theta}(\mathbf{U}_{n\theta}) + \frac{\mathbf{U}'_{n\theta} \mathbf{D}_n \mathbf{U}_{n\theta}}{2} \right| \geq (\lambda + 1) B_n \right\} \leq n^{-\lambda},$$

which implies

$$\mathbb{P} \left\{ G_{n\theta}(\mathbf{U}_{n\theta}) + \frac{\mathbf{U}'_{n\theta} \mathbf{D}_n \mathbf{U}_{n\theta}}{2} \geq (\lambda + 1) B_n \right\} \leq n^{-\lambda}. \quad (16)$$

By considering the complementary sets, we notice that

$$\left\{ \sup_{\boldsymbol{\zeta} \in S_{n\theta}} G_{n\theta}(\mathbf{U}_{n\theta} + \boldsymbol{\zeta}) \leq G_{n\theta}(\mathbf{U}_{n\theta}) \right\} \subset$$

$$\left\{ \left\{ \sup_{\boldsymbol{\zeta} \in S_{n\theta}} G_{n\theta}(\mathbf{U}_{n\theta} + \boldsymbol{\zeta}) \geq -\frac{\mathbf{U}'_{n\theta} \mathbf{D}_n \mathbf{U}_{n\theta}}{2} + (\lambda + 1) B_n \right\} \cup \right.$$

$$\left. \left\{ G_{n\theta}(\mathbf{U}_{n\theta}) \leq -\frac{\mathbf{U}'_{n\theta} \mathbf{D}_n \mathbf{U}_{n\theta}}{2} + (\lambda + 1) B_n \right\} \right\}.$$

Recall that  $\mathbf{D}_n$  tends to a positive definite matrix when  $n$  tends to infinity so that there is an absolute constant  $c$  such that  $\frac{\boldsymbol{\zeta}' \mathbf{D}_n \boldsymbol{\zeta}}{2} \geq c \|\boldsymbol{\zeta}\|^2$  when  $n$  is sufficiently large; we further notice that  $\|\boldsymbol{\zeta}\|^2 \geq B_n \log^2 B_n \geq 2(\lambda + 1) B_n$  for  $n \geq n_\lambda$ . This implies that

$$\left\{ G_{n\theta}(\mathbf{U}_{n\theta} + \boldsymbol{\zeta}) \geq -\frac{\mathbf{U}'_{n\theta} \mathbf{D}_n \mathbf{U}_{n\theta}}{2} + (\lambda + 1) B_n \right\} \subset$$

$$\left\{ \sup_{\boldsymbol{\zeta} \in S_{n\theta}} \left\{ G_{n\theta}(\mathbf{U}_{n\theta} + \boldsymbol{\zeta}) - \frac{\boldsymbol{\zeta}' \mathbf{D}_n \boldsymbol{\zeta}}{2} \right\} \geq -\frac{\mathbf{U}'_{n\theta} \mathbf{D}_n \mathbf{U}_{n\theta}}{2} - (\lambda + 1) B_n \right\}.$$

By using these inclusions and relations (15) and (16), we have

$$\mathbb{P} \left\{ \sup_{\boldsymbol{\zeta} \in S_{n\theta}} G_{n\theta}(\mathbf{U}_{n\theta} + \boldsymbol{\zeta}) \leq G_{n\theta}(\mathbf{U}_{n\theta}) \right\} \leq 2n^{-\lambda}.$$

By the convexity of the function  $G_{n\theta}$ , when  $\sup_{\boldsymbol{\zeta} \in S_{n\theta}} G_{n\theta}(\mathbf{U}_{n\theta} + \boldsymbol{\zeta}) > G_{n\theta}(\mathbf{U}_{n\theta})$ , the minimum of the function  $G_{n\theta}$  is attained inside the ball  $B_{n\theta}$ . We thus have

$$\mathbb{P} \left\{ \|\mathbf{T}_{n\theta} - \mathbf{U}_{n\theta}\| > \sqrt{B_n} |\log B_n| \right\} \leq 2n^{-\lambda},$$

and, since  $\sqrt{B_n} |\log B_n| < B_n$ , we obtain Theorem 2.

**Proof of Theorem 1.**

(i) Denote

$$\gamma_i = q(\theta) (\theta - \mathbb{I}(\varepsilon_i < Q(\theta))), \quad i = 1, \dots, n.$$

We have

$$\begin{aligned} \widehat{\beta}_1^{n'}(\theta) - \widehat{\beta}_1^n(\theta) &= \widehat{\beta}_1^{n'}(\theta) - \beta_1 - Q(\theta) + \beta_1 + Q(\theta) - \widehat{\beta}_1^n(\theta) \\ &= \frac{d_{n'}^{11}}{n'} \sum_{i=1}^{n'} \gamma_i - \frac{d_n^{11}}{n} \sum_{i=1}^n \gamma_i + (R_{n'}'(\theta) - R_n'(\theta)) \end{aligned}$$

where, according to Jurečková (1984),  $R_n'(\theta) = O_p(n^{-3/4})$ .

Let us assume, without loss of generality, that  $n' > n$ . Then

$$\begin{aligned} \widehat{\beta}_1^{n'}(\theta) - \widehat{\beta}_1^n(\theta) &= \frac{d_{n'}^{11}}{n'} \sum_{i=1}^n \gamma_i + \frac{d_{n'}^{11}}{n'} \sum_{i=n+1}^{n'} \gamma_i - \frac{d_n^{11}}{n} \sum_{i=1}^n \gamma_i + (R_{n'}'(\theta) - R_n'(\theta)) \\ &= \left( \frac{d_{n'}^{11}}{n'} - \frac{d_n^{11}}{n} \right) \sum_{i=1}^n \gamma_i + \frac{d_{n'}^{11}}{n'} \sum_{i=n+1}^{n'} \gamma_i + (R_{n'}'(\theta) - R_n'(\theta)). \end{aligned}$$

Let  $A_n = \left(1 - \frac{n d_{n'}^{11}}{n' d_n^{11}}\right) \frac{d_n^{11}}{n} \sum_{i=1}^n \gamma_i$  and  $S_n = \frac{d_{n'}^{11}}{n'} \sum_{i=n+1}^{n'} \gamma_i$ . Then

$$\widehat{\beta}_1^{n'}(\theta) - \widehat{\beta}_1^n(\theta) = S_n - A_n + (R_{n'}'(\theta) - R_n'(\theta)).$$

We shall first focus on the term  $S_n - A_n$ . We have for any fixed  $\lambda$ ,  $0 < \lambda < 1$ ,

$$\begin{aligned} P \left\{ \max_{|n'-n| < cn} |\sqrt{n}(S_n - A_n)| > \varepsilon \right\} &= P \left\{ \max_{|n'-n| < cn} |\sqrt{n}(S_n - A_n)| > \lambda\varepsilon + (1 - \lambda)\varepsilon \right\} \\ &\leq P \left\{ \max_{|n'-n| < cn} \sqrt{n} |A_n| > \lambda\varepsilon \right\} + P \left\{ \max_{|n'-n| < cn} \sqrt{n} |S_n| > (1 - \lambda)\varepsilon \right\} \\ &= P(E_n) + P(E_n'). \end{aligned}$$

We shall now study each term separately. Let us start with  $P(E_n)$ . We know from the assumptions about the model that  $\varepsilon_i$  are i.i.d. random variables, so that  $\mathbb{I}(\varepsilon_k < Q(\theta))$  follows a binomial distribution with parameters 1 and  $\theta$ , thus  $E(\gamma_i) = 0$  and  $\text{Var}(\gamma_i) = q^2(\theta)\theta(1-\theta)$ .

Then, by the central limit theorem

$$\frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_i}{q(\theta) \sqrt{\theta(1-\theta)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

and

$$\forall \eta > 0, \exists K, n_1 \text{ such that } \forall n \geq n_1, P \left\{ \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_i \right| \geq K \right\} \leq \eta/2.$$

$$\text{Denote } H_n = \left[ \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_i \right| \geq K \right].$$

For  $n \geq n_1$

$$\begin{aligned} P \left\{ \max_{|n'-n| < cn} \sqrt{n} |A_n| > \lambda \varepsilon \right\} &= P(E_n) = P(E_n \cap \overline{H_n}) + P(E_n \cap H_n) \\ &= P \left\{ \max_{|n'-n| < cn} \sqrt{n} |A_n| > \lambda \varepsilon, \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_i \right| < K \right\} + P(E_n \cap H_n) = I_1 + I_2. \end{aligned}$$

Realizing that  $E_n \cap H_n \subseteq H_n$ , then

$$I_2 = P(E_n \cap H_n) \leq P(H_n) = P \left\{ \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_i \right| \geq K \right\} \leq \eta/2.$$

For  $I_1$ , we notice that, when

$$\max_{|n'-n| < cn} \left( d_n^{11} - d_{n'}^{11} \frac{n}{n'} \right) \leq \frac{\lambda \varepsilon}{K}, \quad (17)$$

we cannot have at the same time  $E_n$  and  $\overline{H_n}$ . So,  $I_1 = 0$  when (17) is satisfied, and obviously

by 1 when (17) is not satisfied.

From *C5* we have

$$\forall \delta > 0, \exists n_0, \exists c_1 \text{ such that } \forall n \geq n_0 \max_{|n'-n| < c_1 n} \left| d_n^{11} - d_{n'}^{11} \frac{n}{n'} \right| < \delta.$$

Hence if we consider  $\delta = \frac{\lambda \varepsilon}{K}$  and  $c \leq c_1$ , we have for all  $n \geq \min(n_0, n_1)$  :

$$P(E_n) = I_1 + I_2 \leq 0 + \eta/2 = \eta/2. \quad (18)$$

We shall study now the behaviour of the term  $P(E'_n)$ , i.e.

$$P \left\{ \max_{|n'-n| < cn} \sqrt{n} \frac{d_{n'}^{11}}{n'} \left| \sum_{i=n+1}^{n'} \gamma_i \right| > (1-\lambda)\varepsilon \right\}. \quad (19)$$

By Kolmogorov (max) inequality, we have

$$P \left\{ \max_{|n'-n| < cn} \sqrt{n} \frac{d_{n'}^{11}}{n'} \left| \sum_{k=1}^{n'-n} \gamma_{k+n} \right| > (1-\lambda)\varepsilon \right\} \leq \frac{1}{(1-\lambda)^2 \varepsilon^2} \frac{n (d_{n'}^{11})^2}{n'^2} \sum_{k=1}^{n'-n} \text{Var}(\gamma_{k+n}) = \frac{1}{(1-\lambda)^2 \varepsilon^2} \frac{n (d_{n'}^{11})^2}{n'^2} (n'-n) \text{Var}(\gamma_1) \leq \quad (20)$$

$$\frac{1}{(1-\lambda)^2 \varepsilon^2} \frac{n (d_{n'}^{11})^2}{n'^2} cn \text{Var}(\gamma_1).$$

It is clear that  $|n' - n| < cn$  implies  $\left(\frac{n}{n'}\right)^2 < \frac{1}{(1-c)^2}$ .

Plugging all these results into (20), we have

$$P \left\{ \max_{|n'-n| < cn} \sqrt{n} \frac{d_{n'}^{11}}{n'} \left| \sum_{k=1}^{n'-n} \gamma_{k+n} \right| > (1-\lambda)\varepsilon \right\} \leq \frac{\theta(1-\theta)}{(1-\lambda)^2 \varepsilon^2} q^2(\theta) \frac{c}{(1-c)^2} (d_{n'}^{11})^2.$$

To find an upper bound for (19) it is sufficient to put  $k' = k + n$  in order to have

$$P \left\{ \max_{|n'-n| < cn} \sqrt{n} \frac{d_{n'}^{11}}{n'} \left| \sum_{k'=n+1}^{n'} \gamma_{k'} \right| > (1-\lambda)\varepsilon \right\} \leq \frac{\theta(1-\theta)}{(1-\lambda)^2 \varepsilon^2} \times \times q^2(\theta) \frac{c}{(1-c)^2} (d_{n'}^{11})^2. \quad (21)$$

It follows from *C5* that

$$d_{n'}^{11} \rightarrow d^{11} > 0 \text{ as } n' \rightarrow \infty,$$

so that there is  $n_0$  such that for all  $n \geq n_0$   $d_{n'}^{11} > d^{11}/2$ .

Let us choose for  $\theta$  fixed,  $\varepsilon$  and  $\eta$ . Then  $c$  can be chosen such that

$$\frac{c}{(1-c)^2} \leq \frac{\eta}{2} \frac{1}{(d^{11}/2)^2} \frac{(1-\lambda)^2 \varepsilon^2}{\theta(1-\theta)} \frac{1}{q^2(\theta)} \leq \frac{\eta}{2} \frac{1}{(d^{11}/2)^2} (1-\lambda)^2 \varepsilon' \quad (22)$$

where  $\varepsilon' = \frac{\varepsilon^2}{\theta(1-\theta)} \frac{1}{q^2(\theta)}$ .

Combining (21) and (22) we have

$$P \left\{ \max_{|n'-n| < cn} \sqrt{n} \frac{d_{n'}^{11}}{n'} \left| \sum_{k=n+1}^{n'} \gamma_k \right| > (1-\lambda)\varepsilon \right\} \leq \eta/2. \quad (23)$$

It is clear that (23) holds for any  $\lambda \in (0, 1)$ ,  $\lambda$  fixed. By (18) and (23) we have

$$P \left\{ \max_{|n'-n| < cn} |\sqrt{n}(S_n - A_n)| > \varepsilon \right\} \leq \eta/2 + \eta/2 = \eta$$

We now concentrate on the remainder term. As noticed before, further development is required to control the remainder term. Let  $(B_n)$  be the sequence defined in (1) and let  $R_n''(\theta)$  be the remainder term of the ‘‘improved’’ decomposition. We have

$$P \left\{ \max_{|n'-n| < cn} |R_{n'}''(\theta) - R_n''(\theta)| > \sqrt{B_n} |\log B_n| \right\} \leq \sum_{|n'-n| < cn} P \left\{ |R_{n'}''(\theta) - R_n''(\theta)| > \sqrt{B_n} |\log B_n| \right\}.$$

Since  $(B_n)$  tends to zero, for  $\varepsilon > 0$  fixed, there is  $n_1$  such that for all  $n \geq n_1$ ,  $\sqrt{B_n} |\log B_n| \leq \varepsilon$ . Thus, by Theorem 2, we have for every  $\mu > 1$  :

$$P \left\{ \max_{|n'-n| < cn} |R_{n'}''(\theta) - R_n''(\theta)| > \varepsilon \right\} \leq \frac{cn}{n^\mu} \rightarrow 0.$$

In general terms, given  $\varepsilon$  and  $\eta$ , we found  $c, \nu$  such that for all  $n > \nu$

$$P \left\{ \max_{|n'-n| < cn} |Y_{n'}(\theta) - Y_n(\theta)| > \varepsilon \right\} < \eta,$$

where  $Y_n(\theta) = \sqrt{n} \left( \widehat{\beta}_1^n(\theta) - \beta_1 - Q(\theta) \right)$ .

In this way Anscombe's condition for uniform continuity in probability of  $\widehat{\beta}_1^n$  is proved for the first term of  $\widehat{\beta}^n$  and  $\theta$  general.

(ii) The proof follows from Theorem 1.

We proceed in two steps: first collecting tools, then establishing the main Theorem.

**Proof of Lemma 2.** For  $0 < u < \frac{n^{(2a+\delta)/(1+4b)}}{\sqrt{\log n}}$  the quantity  $u R_i(\boldsymbol{\beta}, \theta)$  is bounded. Taylor expansion of the exponential function around  $E(u R_i(\boldsymbol{\beta}, \theta))$  gives

$$\exp(u |R_i(\boldsymbol{\beta}, \theta)|) = \exp(E(u |R_i(\boldsymbol{\beta}, \theta)|)) + (u |R_i(\boldsymbol{\beta}, \theta)| - E(u |R_i(\boldsymbol{\beta}, \theta)|))$$

$$\times \exp(E(u |R_i(\boldsymbol{\beta}, \theta)|)) + O\left((u |R_i(\boldsymbol{\beta}, \theta)| - E(u |R_i(\boldsymbol{\beta}, \theta)|))^2\right) \exp(E(u |R_i(\boldsymbol{\beta}, \theta)|)),$$

where the constant implied in the  $O$  symbol is absolute since the quantity  $u R_i(\boldsymbol{\beta}, \theta)$  is bounded (cf. 7).

By integration, we have

$$E(\exp(u |R_i(\boldsymbol{\beta}, \theta)|)) \leq \exp(u E(|R_i(\boldsymbol{\beta}, \theta)|)) (1 + c \text{Var}(u |R_i(\boldsymbol{\beta}, \theta)|)),$$

and since  $\text{Var}|Y| \leq \text{Var}Y$ , we have

$$E(\exp(u |R_i(\boldsymbol{\beta}, \theta)|)) \leq \exp(u E(|R_i(\boldsymbol{\beta}, \theta)|)) (1 + c \text{Var}(u R_i(\boldsymbol{\beta}, \theta))).$$

Taking the logarithm of both terms the claimed result follows.

**Proof of Lemma 3.** By (5) we have

$$\begin{aligned} E(R_i(\boldsymbol{\beta}, \theta)) &= -\frac{(\mathbf{x}'_i \boldsymbol{\beta})^2}{2n} + \frac{1}{\sqrt{\theta(1-\theta)} \sigma_\theta} E((\zeta_{ni} - \varepsilon_{i\theta}) \mathbb{I}(0 < \varepsilon_{i\theta} < \zeta_{ni})) \\ &= -\frac{(\mathbf{x}'_i \boldsymbol{\beta})^2}{2n} + \frac{1}{\sqrt{\theta(1-\theta)} \sigma_\theta} \int_0^{\zeta_{ni}} (\zeta_{ni} - z) f(z + Q(\theta)) dz. \end{aligned}$$

Now

$$\log(f(z + Q(\theta))) = \log(f(Q(\theta))) + \int_0^z \frac{d}{du} \log(f(u + Q(\theta))) du. \quad (24)$$

Applying condition *C3* and taking the exponential function in (24), we obtain

$$\begin{aligned} f(z + Q(\theta)) &= f(Q(\theta)) \exp\left(\int_0^z \frac{d}{du} \log(f(u + Q(\theta))) du\right) \\ &= f(Q(\theta)) \exp\left(\int_0^z O(u + |Q(\theta)|) du\right). \end{aligned} \quad (25)$$

Noticing that

$$\sigma_\theta |Q(\theta)| = \sqrt{\theta(1-\theta)} \frac{|Q(\theta)|}{f(Q(\theta))}.$$

Condition *C2* on quantile density function implies a similar upper bound for the quantile function itself; namely, we have

$$|Q(\theta)| \leq \frac{c}{(\theta(1-\theta))^a},$$

from where

$$\sqrt{\theta(1-\theta)} |Q(\theta)| \leq \frac{c}{(\theta(1-\theta))^{a-1/2}}.$$

Since  $q(\theta) = 1/f(Q(\theta))$ , we have, from condition *C2*;

$$\sqrt{\theta(1-\theta)} \frac{|Q(\theta)|}{f(Q(\theta))} \leq \frac{c}{(\theta(1-\theta))^{2a+1/2}},$$

so that, as  $n \rightarrow \infty$ ,

$$\zeta_{ni} |Q(\theta)| = O\left(\frac{1}{\sqrt{n} (\theta(1-\theta))^{2a+1/2}} n^{(2(b-a)-\delta)/(1+4b)} \sqrt{\log n}\right) \rightarrow 0.$$

Hence, using Taylor expansion in the neighbourhood of 0, we have

$$\exp\left(O\left(\int_0^z (u + |Q(\theta)|) du\right)\right) = 1 + O\left(\int_0^z (u + |Q(\theta)|) du\right), \quad (26)$$

where the constant in the right hand side is taken uniformly when  $z$  is positive and bounded.

Applying (25) and (26), we obtain

$$f(z + Q(\theta)) = f(Q(\theta)) \left(1 + O\left(\int_0^z (u + |Q(\theta)|) du\right)\right). \quad (27)$$

We remember that

$$\mathbb{E}(R_i(\boldsymbol{\beta}, \theta)) = -\frac{(\mathbf{x}'_i \boldsymbol{\beta})^2}{2n} + \frac{1}{\sqrt{\theta(1-\theta)} \sigma_\theta} \int_0^{\zeta_{ni}} (\zeta_{ni} - z) f(z + Q(\theta)) dz.$$

By (27) we obtain

$$\begin{aligned} \mathbb{E}(R_i(\boldsymbol{\beta}, \theta)) &\leq -\frac{(\mathbf{x}'_i \boldsymbol{\beta})^2}{2n} + \frac{1}{\sqrt{\theta(1-\theta)} \sigma_\theta} f(Q(\theta)) \times \\ &\times \left[ \int_0^{\zeta_{ni}} (\zeta_{ni} - z) dz + c \int_0^{\zeta_{ni}} (\zeta_{ni} - z) \left[ \int_0^z (u + |Q(\theta)|) du \right] dz \right]. \end{aligned}$$

By a simple calculation of the first integral on the right hand side of the previous inequality we obtain

$$\frac{1}{\sqrt{\theta(1-\theta)} \sigma_\theta} f(Q(\theta)) \int_0^{\zeta_{ni}} (\zeta_{ni} - z) dz = \frac{1}{\sigma_\theta^2} \left(\frac{\zeta_{ni}^2}{2}\right) = \frac{(\mathbf{x}'_i \boldsymbol{\beta})^2}{2n}.$$

Thus

$$\mathbb{E}(|R_i(\boldsymbol{\beta}, \theta)|) \leq \frac{c}{\sigma_\theta^2} \int_0^{\zeta_{ni}} (\zeta_{ni} - z) \left[ \int_0^z (u + |Q(\theta)|) du \right] dz.$$

Elementary calculations give us

$$\int_0^{\zeta_{ni}} (\zeta_{ni} - z) \left[ \int_0^z (u + |Q(\theta)|) du \right] dz = \frac{\zeta_{ni}^4 + 4 |Q(\theta)| \zeta_{ni}^3}{24},$$

from where

$$\begin{aligned} \mathbb{E} (|R_i(\boldsymbol{\beta}, \theta)|) &\leq \frac{c}{\sigma_\theta^2} \left( \frac{\sigma_\theta^4 (\mathbf{x}'_i \boldsymbol{\beta})^4}{24 n^2} + \frac{4 |Q(\theta)| \sigma_\theta^3 (\mathbf{x}'_i \boldsymbol{\beta})^3}{24 n^{3/2}} \right) \\ &\leq c \left( \frac{\sigma_\theta^2 (\mathbf{x}'_i \boldsymbol{\beta})^4}{n^2} + \frac{|Q(\theta)| \sigma_\theta (\mathbf{x}'_i \boldsymbol{\beta})^3}{n^{3/2}} \right). \end{aligned}$$

By *C2* we have

$$\mathbb{E} (|R_i(\boldsymbol{\beta}, \theta)|) \leq \frac{c |\mathbf{x}'_i \boldsymbol{\beta}|^4}{n^2 (\theta(1-\theta))^{1+2a}} + \frac{c |\mathbf{x}'_i \boldsymbol{\beta}|^3}{n^{3/2} (\theta(1-\theta))^{1/2+2a}}.$$

As we already mentioned, we have the same inequality when  $\zeta_{ni} < 0$ , which completes the proof of Lemma 3.

**Proof of Lemma 4.** By simple calculations, the variance of  $R_i(\boldsymbol{\beta}, \theta)$  given in (4) is

$$\begin{aligned} \text{Var} (R_i(\boldsymbol{\beta}, \theta)) &= \frac{1}{\theta(1-\theta) \sigma_\theta^2} \left[ \int_0^{|\zeta_{ni}|} (|\zeta_{ni}| - y)^2 f(y + |Q(\theta)|) dy \right. \\ &\quad \left. - \left( \int_0^{|\zeta_{ni}|} (|\zeta_{ni}| - y) f(y + |Q(\theta)|) dy \right)^2 \right]. \end{aligned}$$

As above, we can write the inequality

$$\begin{aligned} \text{Var} (R_i(\boldsymbol{\beta}, \theta)) &\leq \frac{1}{\theta(1-\theta) \sigma_\theta^2} \left\{ f(Q(\theta)) \int_0^{|\zeta_{ni}|} (|\zeta_{ni}| - z)^2 \left[ 1 + \int_0^z (y + |Q(\theta)|) dy \right] dz \right. \\ &\quad \left. + f^2(Q(\theta)) \left\{ \int_0^{|\zeta_{ni}|} (|\zeta_{ni}| - z) \left[ 1 + \int_0^z (y + |Q(\theta)|) dy \right] dz \right\}^2 \right\}. \end{aligned}$$

Therefore, applying the method used for the study of  $E(R_i(\boldsymbol{\beta}, \theta))$  we have

$$\sum_{i=1}^n \text{Var}(R_i(\boldsymbol{\beta}, \theta)) \leq \frac{c}{n^{3/2} \sqrt{\theta(1-\theta)}} \sum_{i=1}^n |\mathbf{x}'_i \boldsymbol{\beta}|^3 = O\left(\frac{1}{n^{2b/(1+4b)}}\right)$$

uniformly for  $\theta \in (\theta_n^*, 1 - \theta_n^*)$  and  $\|\boldsymbol{\beta}\| \leq C_n$ , which completes the proof of Lemma 4.

## References

- Anscombe, F. J. (1952), “Large sample theory of sequential estimation,” Proceedings of the Cambridge Philosophical Society, 48, 600–607.
- Gutenbrunner, C., and Jurečková, J. (1992), “Regression rank scores and regression quantiles,” The Annals of Statistics, 20, 305–330.
- Rogers, L. C. G., and Williams, D. (1994), “Diffusions, Markov processes, and martingales : foundations,” New York: John Wiley.
- Shorack, G., and Wellner, J. (1986), “Empirical Processes with Applications to Statistics,” New York: John Wiley.
- Shorack, G. (1991), “Embedding the finite sampling process at a rate,” Annals of Probability, 19, 826–842.