Nonparametric Inference for Bioassay

by
LIZHEN LIN

A Dissertation Submitted to the Faculty of the
Department of Mathematics
In Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy
In the Graduate College
The University of Arizona

2 0 1 2

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Lizhen Lin entitled
Nonparametric Inference for Bioassay
and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

_____ Date: 04/09/12
Rabi Bhattacharya

_____ Date: 04/09/12
Tom Kenneday

_____ Date: 04/09/12
Walter Piegorsch

_____ Date: 04/09/12
Joseph Watkins

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____ Date: 04/09/12
Rabi Bhattacharya

Statement by Author

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Signed: _____ Lizhen Lin _____

Acknowledgments

I am indebted to my advisor Professor Rabi Bhattacharya for his guidance, support and encouragement over the years. Having Professor Rabi Bhattacharya as my advisor is one of the luckiest things that have happened to my life. My advisor has not just taught me a great deal of mathematics and statistics, he has also greatly influenced me with his great personality and the kind of generosity he has been showing to many others. I can not imagine having a better and more generous advisor. I hope one day I can live up to his example.

I am very grateful for Professor Tom Kennedy and Professor Walter Piegorsch for their constant support and encouragement. I also wish to thank both of them and Professor Joseph Watkins for being on my thesis committee and for their helpful suggestions.

Dedication

To my parents Lin Hanghong and Hu Caifang and to my brother Lin Yuan.

Table of Contents

Table of Contents—*Continued*

List of Figures

Abstract

This thesis proposes some new model independent or nonparametric methods for estimating the dose-response curve and the effective dosage curve in the context of bioassay. The research problem is also of importance in environmental risk assessment and other areas of health sciences. It is shown in the thesis that our new nonparametric methods while bearing optimal asymptotic properties also exhibit strong finite sample performance. Although our specific emphasis is on bioassay and environmental risk assessment, the methodology developed in this dissertation applies broadly to general order restricted inference.

Chapter 1
Introduction

1.1 Introduction

Regression analysis is a very important branch of statistics which concerns the relationship between two variables y (dependent variable) and x (independent variable). The typical setup is

$$y = F(x) + \epsilon, \tag{1.1.1}$$

where $F(x)$ is the *regression function* and ϵ is some zero-mean independent error. Given data (x_i, y_i) $(i = 1, \cdots, N)$, the main object is to make inference about $F(x)$ such as estimation of $F(x)$ and testing hypotheses about $F(x)$. There are various tools for estimating the regression function, from parametric, semiparametric models depending on the assumptions on the error and the regression function, to nonparametric methods. There is a growing interest in estimating the regression function $F(x)$ under some shape constraint, the most common of which is the *monotone constraint*. The statistical theory of *isotonic regression* deals with such regression problems, where there is an order restriction on $F(x)$. In this case $F(x)$ is estimated by minimizing an objective function which is a weighted sum of squares of the differences between observed data and fitted values (see Barlow et. al (1972) and Robertson et al. (1988)). The present thesis focuses on the monotone constrained regression problem where the dependent variable y is measured in *binary scale* and records the response of a subject exposed to some level x representing the level of a substance such as a chemical drug, an environmental pollutant, exposure to an occupational hazard, etc. The regression function $F(x)$ is the *probability of response* of a subject at dosage level x. The motivation for this research problem mainly comes from the area of bioassay where one is interested in measuring the potency or the nature of a substance by

studying its effects on living matter. In quantal bioassay, the analysis of the effects of the substance heavily relies on proper estimation of the *dose-response curve* $F(x)$ or, more importantly, its inverse-the *effective dosage curve*. Such procedures are essential in the development of new drugs, for example. The research problem also finds applications in environmental risk assessment, where one is generally interested in finding safe levels of a pollutant or toxicant with a small given probability of response (such as cancer) of an exposed subject. Thus, it is of immense importance in the health sciences.

The research problem is an old one. For the most part, however, the old literature focuses on parametric models, and there have been legitimate concerns about model misspecification which occurs more often than not. As will be shown in Section 1.3, the specification of the wrong model leads to inconsistent estimates. In contrast, nonparametric methods in the present context are of relatively recent origin and they provide asymptotically correct inference. Therefore, this thesis centers on model-independent, or nonparametric, methodologies in estimating $F(x)$ and its inverse.

Although our specific emphasis is on bioassay and environmental risk assessment, the methodology developed in this dissertation applies broadly to general order restricted inference. In the following section we survey some of the landmarks of the nonparametric theory lying at the root of the new methodology presented in this thesis.

1.2 Isotonic Regression and Isotonic Density Estimation

The *isotonic regression problem* is to find an estimate F^*, say, of the true regression function F in (1.1.1) at observed points $0 \leq x_1 \leq x_2 \leq \cdots \leq x_N$, such that F^* minimizes, for given weights $w(x_i)$ $(1 \leq i \leq N)$, the weighted sum of squares

$$\sum_{i=1}^{N} (y_i - F(x_i))^2 w(x_i), \tag{1.2.1}$$

over the class of all F satisfying $F(x_1) \leq F(x_2) \leq \cdots \leq F(x_N)$. The solution of this problem is expressed in terms of so-called *Pool-Adjacent-Violators*, or *PAV*, algorithm (see Ayer et al. (1955)) as

$$F^*(x_j) = \max_{\{s:x_s \leq x_j\}} \min_{\{t:x_j \leq x_t\}} \frac{\sum_{\{i:x_s \leq x_i \leq x_t\}} w(x_i) y_i}{\sum_{\{i:x_s \leq x_i \leq x_t\}} w(x_i)}, \ (j = 1, 2, \cdots, N). \tag{1.2.2}$$

This value $F^*(x_j)$ may be alternately described as the left-hand slope at x_j of the *greatest convex minorant* of the graph of (W_j, G_j), $0 \leq j \leq N$, with $W_j = \sum_{i=1}^{j} w(x_i)$ and $G_j = \sum_{i=1}^{j} y_i w(x_i)$, linearly interpolated in (W_j, W_{j+1}) $(j = 0, 1, \cdots, N-1)$. Here one takes $x_0 = 0 = y_0$ (See, e.g., Robertson (1988), pp. 7,8).

Note that one may derive the isotonic regression of a *decreasing* F, by simply reversing 'max' and 'min' in (1.2.2). In this case $F^*(x_j)$ is the right-hand slope of the smallest concave majorant of the linearly interpolated graph of (W_j, G_j), $0 \leq j \leq N$.

In the dose-response problem in bioassay, let '1' stand for response of a subject and '0' stand for non-response. Given some dosages of a chemical agent $x_1 < \ldots < x_m$, suppose that n_i subjects are exposed to a dosage x_i $(i = 1, \ldots, m)$. The number of responses observed at dosage x_i is recorded as r_i $(i = 1, \ldots, m)$. One has

$$F(x) = P\left(\text{response at } x\right), \tag{1.2.3}$$

which is the probability of observing a 1 at dosage x_i and $x \to F(x)$ is called the *dose-response curve*. One defines the *effective dosage curve* as

$$ED_p = F^{-1}(p) \tag{1.2.4}$$

for any level of response probability $p \in [0, 1]$. Given the data (x_i, r_i), the purpose is to come up with efficient estimates of $F(x)$ and ED_p. The isotonic regression estimate, with weights n_i as the number of exposed subjects at dosages x_i, is given by

$$F^*(x_j) = \max_{\{s:x_s \leq x_j\}} \min_{\{t:x_j \leq x_t\}} \frac{\sum_{i:x_s \leq x_i \leq x_t} r_i}{\sum_{i:x_s \leq x_i \leq x_t} n_i} \ (1 \leq j \leq m). \tag{1.2.5}$$

This estimate is also the *nonparametric maximum likelihood estimate of* F (at x_j, $1 \le j \le m$) (See Brunk (1955)).

In an apparently different direction, consider the problem of nonparametric estimation of a monotone decreasing density f on $[0, \infty)$, based on *i.i.d.* data ordered $X_1 \le X_2 \le \cdots \le X_N$. It was proved by Grenander (1956) (also see Brunk (1958) and Van Eeden (1958)) that the MLE of f is given by

$$\widehat{f}_N(x) = \begin{cases} \sup_{v>x} \inf_{u<x} \dfrac{F_N(v) - F_N(u)}{v-u} & \text{for } 0 = X_0 < x < X_N, \\ \widehat{f}_N(x) = 0 & \text{otherwise.} \end{cases} \tag{1.2.6}$$

Here F_N is the empirical distribution function: $F_N(x) = \dfrac{1}{N} \max\{i : X_i \le x\}$. A beautiful result of Prakasa Rao (1969) is the following theorem. For its statement, consider a *two-sided standard Brownnian motion* $\{B_t : -\infty < t < \infty\}$ with $B(0) = 0$.

Theorem 1.2.1. *Suppose* f *is decreasing and differentiable on* $[0, \infty)$ *and that* $f'(\zeta) \neq 0$ *for some* $\zeta > 0$. *Then*

$$N^{1/3} \left(\frac{f(\zeta)}{2}(-f'(\zeta)) \right)^{-1/3} \left[\widehat{f}_N(\zeta) - f(\zeta) \right] \xrightarrow{L} V, \tag{1.2.7}$$

where V *is the slope at* $t = 0$ *of the smallest concave majorant of the process*

$$\left\{ B(t) - t^2 : -\infty < t < \infty \right\}.$$

If one compares (1.2.2) and (1.2.6), it may not come as a complete surprise that a result analogous to Theorem 1.2.1 holds for the isotonic estimate $F^*(x)$ in (1.2.2). To state this result due to Brunk (1970) (also see Wright (1981)), let H_N be the distribution function of the design points, $H_N(x) = \dfrac{1}{N} \#\{i : x_i \le x\}$, with x lying in an interval I. Assume that there exists a distribution function H, continually differentiable in a neighborhood $(z_0 - \delta, z_0 + \delta)$, $\delta > 0$, of a point z_0 of interest. Assume $H'(x) > 0$ on $(z_0 - \delta, z_0 + \delta)$, and that

$$\sup_{x \in (z_0 - \delta, z_0 + \delta)} |H_N(x) - H_N(z_0)| = o\left(N^{-1/3}\right) \text{ as } N \to \infty. \tag{1.2.8}$$

Further assume that there is a positive and continuous function w such that $\phi(x) = \sigma^2(x)w(x)$ is continuous and bounded on $(z_0 - \delta, z_0 + \delta)$, where $\sigma^2(x)$ is the variance of y_i when the independent variable has the value x. Assume further that

$$\sum_{i=1}^{N} w(x_i)(y_i - F(x_i))/\left(\sum_{i=1}^{N} \phi(x_i)\right)^{1/2} \xrightarrow{L} N(0,1), \text{ as } N \to \infty. \tag{1.2.9}$$

Theorem 1.2.2. *Suppose F is nondecreasing on I and continuously differentiable on $(z_0 - \delta, z_0 + \delta)$, $F'(z_0) > 0$. Then under the above assumptions,*

$$\left(\frac{2H'(z_0)}{\sigma^2(z_0)F'(z_0)}\right)^{1/3} N^{1/3}\left[F^*(z_0) - F(z_0)\right] \xrightarrow{L} U, \text{ as } N \to \infty, \tag{1.2.10}$$

where U is the slope at $t = 0$ of the greatest convex minorant of the process $\{B_t + t^2 : -\infty < t < \infty\}$, with $B(t)$, $-\infty < t < \infty$, as the two-sided standard Brownian motion, $B(0) = 0$.

Remark 1.2.3. The assumption (1.2.8) holds, in particular, if the difference $x_i - x_{i-1}$ $(i = 1, \cdots, N)$ are equidistant on $I = [a, b]$, with $x_0 = a$, $x_N = b$. In this case H is the uniform distribution function on $[a, b]$. The assumption (1.2.9) then holds if $\sigma^2(\cdot)$ and $w(\cdot)$ are continuous and positive on $[a, b]$. These assumptions also hold if there are m design points $a < x_1 < x_2 < \cdots < x_m = b$ with $x_i - x_{i-1} = \dfrac{1}{m}$ $\forall i$, $x_0 = a$, and for each x_i there are n observations y, provided m is of larger order than $N^{1/3}$. Here $N = mn$. The assumptions of equidistant design points and equal number of observations at each point may be greatly relaxed, and the results will still apply.

We next turn to the dose-response problem and the PAV estimate (1.2.5), with equidistant design points $a \le x_1 < x_2 < \cdots < x_m \le b$ on $[a, b]$, with n observations at each point, so that $N = mn$. It follows from Bhattacharya and Kong (2007), referred to as B-K (2007) (also see Kong (2004)), that, if F is twice continuously differentiable and $F' > 0$ everywhere then

$$F^*(x) \text{ is asymptotically Normal } N(F(x), \nu_N),\ \nu_N = O(n^{-1}),\ (x \in [x_1, x_m]), \tag{1.2.11}$$

provided $n^{1/4} << m << n^{1/2}$. Here F^* is linearly interpolated on (x_{i-1}, x_i) $\forall i = 1, \cdots, m$ $(x_0 = a)$. Since $N = mn$, this requires

$$N^{1/5} << m << N^{1/3}, \tag{1.2.12}$$

and, correspondingly, the asymptotic variance $\nu_n = O(m/N)$ satisfies

$$N^{-4/5} << \nu_N << N^{-2/3}. \tag{1.2.13}$$

The requirements imposed on m in this case and that in Theorem 1.2.2 are exclusive of each other.

Note that the optimal design for the interpolated PAV estimate in B-K (2007) is achieved with m slightly larger than $O(N^{1/5})$, in which case the asymptotic variance is only a bit larger than $O(N^{-4/5})$, and it is essentially the optimal order of the asymptotic variance in nonparametric estimation of a twice differentiable curve $x \to F(x)$ (see, e.g., Tsybakov (2010)). This motivates the *nonparametric adaptive method NAM* developed in this dissertation. We described this new method very briefly below.

If $m >> N^{1/5}$, then the NAM divides the m dosages into r groups of adjacent dosages, each containing about m/r dosages. The corresponding r PAV estimates are then averaged to provide the *NAM estimate* $\widetilde{F}$. The value of r may be obtained as that which minimizes the estimate of the *mean integrated squared error*, or *MISE*, of $\widetilde{F}$. It is shown that, with this choice of r, $\widetilde{F}$ is essentially optimal:

$$\widetilde{F}(x) \text{ is asymptotically Normal } N(F(x), \delta_N),\ \delta_N = O(N^{-4/5}(\log N)^{\alpha}) \text{ for some } \alpha > 0, \tag{1.2.14}$$

provided m lies in the range

$$N^{1/5} << m << N^{3/5}. \tag{1.2.15}$$

For this range of values of m, which includes a substantial part of the range of $m >> N^{1/3}$ in Theorem 1.2.2, NAM provides asymptotically optimal inference for F, under the stated assumptions.

The main goal in this thesis, however, is the estimation of the inverse curve F^{-1} and, in particular, $\zeta_p = F^{-1}(p)$ for a reasonably small $p \in (0,1)$. The inverse of $\widetilde{F}$ provides this estimate $\widetilde{\zeta}_p = \widetilde{F}^{-1}(p)$. Once again it is shown that the optimal rate of asymptotic variance is attained by $\widetilde{\zeta}_p$ (as in (1.2.14)) for a smaller range of values of m, namely,

$$N^{1/5} << m << N^{2/5}. \tag{1.2.16}$$

In Chapter 6, we outline a method which extends the range of m beyond (1.2.15), or (1.2.16). When $m > N^{3/5} \log N$, we divide the distinct dosages x_i $(i = 1, \cdots, m)$ into approximately m/k groups. We estimate the response at the 'middle point' of each of the groups (with k dosages) by applying the PAV method to the sample proportions in these groups. The estimates between two 'middle points' are obtained by linear interpolation. The estimates of the dose-response curve are again shown to be optimal. Note that the method in Chapter 6 and the NAM are two different methods since, for NAM, the dosages that are r dosages apart are grouped together while for our method in Chapter 6, k neighboring dosages form a group.

In the next section of this chapter, we describe the consequences of model-misspecification in parametric estimation in general. Chapter 2 reviews the family of methods that are kernel based, especially one due to Müller and Schimidt (1988) and a leading new method, which is termed the *DNP method*, due to Dette et al. (2005) and Dette and Scheder (2010). Chapter 3 is devoted to monotone spline curve estimation. A related work is due to Kong and Eubank (2007). In Chapter 4, we describe the methods that are based on isotonic MLEs (maximum likelihood estimators). Due to the connection between the isotonic MLEs and the solutions to the isotonic regression problems, a detailed discussion of isotonic regression is given in the first section which helps establish this connection. In section 2 of Chapter 4, the B-K method due to Bhattacharya and Kong (2007) is described in detail. Then we describe our newly developed non-parametric methodology, *NAM*, and study its precise asymptotics which turn out to

be optimal. In Chapter 5, we carry out a comprehensive comparative study of NAM with all other major existing nonparametric methods in the area including kernel methods. The NAM is shown to perform best in the majority of the cases of practical significance. In Chapter 6, we describe a new method for estimating the dose-response curve and effective dosage curve together for a range of design points not covered by the NAM. The NAM and this method essentially cover the entire range of appropriate designs. We also show that the methods we have developed in this dissertation apply immediately to the general isotonic regression problem (1.1.1).

Although the context of bioassay and environmental risk assessment seems very specific, the research in this area applies to broader fields of nonparametric estimation of curves under isotonic constraints, as described above. In addition, it has the potential to be extended to higher-dimension regressors x. Examples of this type include the estimation of probability of response of subjects under the joint action of two environmental pollutants or the example of drug-interaction where there are more than one drug.

1.3 Parametric Estimation and Model Misspecification

Traditional parametric estimation assumes that the data come from some parametric distribution F_θ with the parameter θ from some finite-dimensional parameter space Θ, say $\Theta \subset \mathbb{R}^k$. The distribution F_θ is completely determined if the parameter θ is specified. Thus estimation of the distribution based on the data boils down to estimating the parameter θ, which is generally obtained by maximum likelihood estimation. The theory of parametric estimation is mathematically elegant and parametric models are known to have certain advantages such as its straightforwardness to interpret. If the correct model is specified, which is extremely rare in real applications, then it can be shown that the estimates obtained in terms of maximum likelihood estimation are asymptotically optimal in the sense that the variance of an estimate attains its

lower bound, which is the Cramér-Rao bound (See Bhattacharya and Patrangenaru (2013)). In general, however, it is difficult to justify the choice of the model. Model misspecification almost always occurs in real applications and often results in inconsistent estimates. This is one of the main motivations for alternative inference using nonparametric methodologies.

In the following, we aim to make precise the consequences of model misspecifications. It is shown that in the particular application of estimating the α-th quantile $F^{-1}(\alpha)$, incorrect specification of the parametric model for $F(x)$ indeed leads to inconsistent estimates.

Consider a parametric family of strictly positive densities $f(x;\theta)$, $x \in I$, and $\theta \in \Theta$, where I is an interval and the parameter space Θ is an open subset of R^k for some $k \geq 1$. Suppose a random sample $X_1, ..., X_n$ is drawn from a distribution with density $g(x)$.

We first consider the asymptotic result of the case **(A)** when the correct model is specified. **(A)** First assume g belongs to a parametric family, say, $g(x) = f(x;\theta)$. Denote $\hat{\theta}_n$ as the MLE of θ, and let $F(;\theta)$ and $G(x)$ be the distribution functions with respective densities $f(x;\theta)$ and $g(x)$. Under broad assumptions, $\hat{\theta}_n$ is asymptotically Normal $N(\theta_0, \Sigma/n)$, i.e., $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$ converges in distribution to $N(0, \Sigma)$, where Σ is the inverse of the Fisher information matrix which is defined as $\mathcal{I}(\theta) = \left(E\left(\frac{\partial}{\partial\theta_i}\ln f(X;\theta)\frac{\partial}{\partial\theta_j}\ln f(X;\theta)\right)\right)$ (See Bhattacharya and Patrangenaru (2013)).

The one-parameter case is given by the following theorem:

Theorem 1.3.1. *Let $f(x;\theta)$ be the strictly positive density of X on $\aleph$ and $\forall\theta \in \Theta$, $P_\theta(\aleph) = 1$. Assume also the following conditions:*

(1) $\theta \to f(x;\theta)$ is thrice continuously differentiable on Θ, $\forall\ x \in \aleph$;

(2) $\int_\aleph \frac{d}{d\theta}f(x;\theta)dx = 0 (\equiv \frac{d}{d\theta}\int_\aleph f(x;\theta)dx)$, $\int_\aleph \frac{d^2}{d\theta^2}f(x;\theta)dx = 0 (\equiv \frac{d^2}{d\theta^2}\int_\aleph f(x;\theta)dx)$;

(3) $0 < \mathcal{I}(\theta) := \mathrm{E}\left\{ \left[\frac{\partial}{\partial\theta} \ln f(X_1;\theta)\right]^2 \middle| \theta \right\} < \infty \ \forall\ \theta \in \Theta.$

(4) For each $\theta_0 \in \Theta$*, there exists an* $\epsilon > 0$ *such that* $| \frac{d^3 \log(f(x;\theta))}{d\theta^3} |^3 \leq g(x), \ \forall\ \theta \in [\theta_0 - \epsilon, \theta_0 + \epsilon]$*, where* $\int_{\aleph} g(x) f(x;\theta_0) dx < \infty$*.*

(5) The log *likelihood equation (writing* $l(\theta) := \sum_{j=1}^{n} log f(X_j : \theta)$*),* $\frac{dl(\theta)}{d\theta} = 0$ *has a consistent solution* $\hat{\theta}_n$*.*

Then $\hat{\theta}_n$ *is asymptotically Normal* $N(\theta_0, \frac{1}{nI(\theta_0)})$*, i.e.,* $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$ *converges in distribution to* $N\left(0, \frac{1}{I(\theta_0)}\right)$ *if* θ_0 *is the true parameter value.*

Under correct model specification, a Corollary of the above theorem implies that the α-th quantile of the estimated distribution is a consistent estimator of the true quantile. The Corollary is stated as follows.

Corollary 1.3.2. *Assume the hypothesis of Theorem 1.3.1, given* $0 < \alpha < 1$*, denote the MLE of the* α*-th quantile as* $\hat{q}_{\alpha,n} = F^{-1}(\alpha, \hat{\theta}_n)$ *of* $F(;\hat{\theta}_n)$*. Let the true quantile of* F *be* $q_\alpha = F^{-1}(\alpha, \theta_0)$*. Then* $\sqrt{n}(\hat{q}_{\alpha,n} - q_\alpha)$ *converges in distribution to* $N(0, \delta_1)$ *where* $\delta_1 = \left(\frac{dF^{-1}}{d\theta}\right)^2_{\theta_0} \left(\frac{1}{I(\theta_0)}\right)$*.*

Proof. Take the second-order Taylor approximation of $F^{-1}(\alpha, \theta)$ at θ_0 as

$$F^{-1}(\alpha, \theta) = F^{-1}(\alpha, \theta_0) + \frac{dF^{-1}}{d\theta} \mid_{\theta_0} (\theta - \theta_0) + \frac{dF^{-1}}{d\theta} \mid_{\theta^*} (\theta - \theta_0)^2$$

for some θ^* between θ and θ_0. One has

$$F^{-1}(\alpha, \hat{\theta}_n) = F^{-1}(\alpha, \theta_0) + \frac{dF^{-1}}{d\theta} \mid_{\theta_0} (\hat{\theta}_n - \theta_0) + \frac{dF^{-1}}{d\theta} \mid_{\theta^*} (\hat{\theta}_n - \theta_0)^2.$$

$$\sqrt{n}\left(F^{-1}(\alpha, \hat{\theta}_n) - F^{-1}(\alpha, \theta_0)\right) = \sqrt{n}\left(\frac{dF^{-1}}{d\theta} \mid_{\theta_0} (\hat{\theta}_n - \theta_0)\right) + \sqrt{n}\left(\frac{dF^{-1}}{d\theta} \mid_{\theta^*} (\hat{\theta}_n - \theta_0)^2\right).$$

So from the above results, under broad conditions, $\hat{\theta}_n$ is asymptotically Normal $N\left(\theta_0, \frac{1}{nI(\theta_0)}\right)$, $\sqrt{n}\left(F^{-1}(\alpha, \hat{\theta}_n) - F^{-1}(\alpha, \theta_0)\right)$ converges in distribution $N\left(0, \left(\frac{dF^{-1}}{d\theta} \mid_{\theta_0}\right)^2 \frac{1}{I(\theta_0)}\right)$.

□

We now state a theorem on the asymptotic distribution of a nonparametric estimator of the quantile based on the empirical distributions.

Theorem 1.3.3. *Let $\hat{\zeta}_{\alpha,n} = X_{([n\alpha])}$ be the nonparametric estimator of the α-th quantile q_α based on i.i.d observations $X_1, X_2, ..., X_n$ with a density f. Let $X_{(1),n}, X_{(2),n}, ..., X_{(n),n}$ be the ordering of the sample observations. Then $\sqrt{n}\left(\hat{\zeta}_{\alpha,n} - q_\alpha\right)$ converges in distribution to $N(0, \delta_2)$ with $\delta_2 = \dfrac{\alpha(1-\alpha)}{f^2(q_\alpha)}$.*

Proof. Assume that F is continuous, the solution q_α of $F(x) = \alpha$ is unique, F is continuously differentiable in a neighborhood of q_α and $F'(q_\alpha) = f(q_\alpha) > 0$.

Fix $z \in (-\infty, \infty)$ and write 1_A for the indicator of A.

$$
\begin{aligned}
&Prob(\sqrt{n}(\hat{\zeta}_{\alpha,n} - q_\alpha) \le z) = Prob(\hat{\zeta_\alpha} \le q_\alpha + \frac{z}{\sqrt{n}}) \\
&= Prob(\#\text{ of observations among } X_1, ..., X_n \text{ which are} \le q_\alpha + \frac{z}{\sqrt{n}} \text{ is} \ge [n\alpha]) \\
&= Prob\left(\sum_j^n 1_{\{X_j \le q_\alpha + \frac{z}{\sqrt{n}}\}} \ge [n\alpha]\right) \\
&= Prob\left[\frac{1}{\sqrt{n}}\left(\sum_j^n 1_{\{X_j \le q_\alpha + \frac{z}{\sqrt{n}}\}} - nF(q_\alpha + \frac{z}{\sqrt{n}})\right) \ge \frac{1}{\sqrt{n}}\left([n\alpha] - nF(q_\alpha + \frac{z}{\sqrt{n}}\right)\right].
\end{aligned}
$$

Now apply Lyapounov's central limit theorem (see Bhattacharya and Waymire (2007)) to the (triangular array of) random variables $\frac{1}{\sqrt{n}} 1_{\{X_j \le q_\alpha + \frac{z}{\sqrt{n}}\}} (j = 1, \cdots, n; n = 1, 2, \cdots)$. One has

$$
\begin{aligned}
\sigma_{n,j}^2 &:= \text{VAR}\left(\frac{1}{\sqrt{n}} 1_{\{X_j \le q_\alpha + \frac{z}{\sqrt{n}}\}}\right) \\
&= \frac{1}{n} F\left(q_\alpha + \frac{z}{\sqrt{n}}\right)\left(1 - F\left(q_\alpha + \frac{z}{\sqrt{n}}\right)\right) \\
&= \frac{(\alpha + o(1))(1 - \alpha + o(1))}{n} \\
&= \frac{\alpha(1-\alpha) + o(1)}{n}
\end{aligned}
$$

where $o(1)$ is a term which goes to zero as $n \to \infty$. It holds that

$$\rho_{n,j} := E\left|\frac{1}{\sqrt{n}}1_{\{X_j \le q_\alpha + \frac{z}{\sqrt{n}}\}} - \frac{F(q_\alpha + \frac{z}{\sqrt{n}})}{\sqrt{n}}\right|^3 \le \frac{1}{n^{\frac{3}{2}}}.$$

Now write

$$s_n^2 = \sum_{j=1}^n \sigma_{n,j}^2, \quad \rho_{3,n} = \sum_{j=1}^n \rho_{n,j}.$$

Then

$$\frac{\rho_{3,n}}{s_n^3} \le \frac{1}{\sqrt{n}}\left(\alpha(1-\alpha) + o(1)\right)^{-\frac{3}{2}} \to 0$$

as $n \to \infty$.

Therefore, Lyapounov's central limit theorem applies and one has

$$Prob\left[\frac{\frac{1}{\sqrt{n}}\left(\sum_j^n 1_{\{X_j \le q_\alpha + \frac{z}{\sqrt{n}}\}} - nF(q_\alpha + \frac{z}{\sqrt{n}})\right)}{s_n} \ge \frac{\frac{1}{\sqrt{n}}\left([n\alpha] - nF(q_\alpha + \frac{z}{\sqrt{n}})\right)}{s_n}\right]$$
$$= Prob\left(Z \ge \frac{\frac{1}{\sqrt{n}}([n\alpha] - nF(q_\alpha + \frac{z}{\sqrt{n}}))}{s_n}\right) + o(1)$$

where Z is a standard normal random variable. Now

$$\frac{\frac{1}{\sqrt{n}}([n\alpha] - nF(q_\alpha + \frac{z}{\sqrt{n}}))}{s_n} = \frac{\frac{1}{\sqrt{n}}([n\alpha] - n\{\alpha + \frac{z}{\sqrt{n}}f(q_\alpha) + o(\frac{1}{\sqrt{n}})\})}{s_n}$$
$$= \frac{-zf(q_\alpha)}{\sqrt{\alpha(1-\alpha)}} + o(1).$$

Therefore, the right side is

$$Prob(Z \ge \frac{-zf(q_\alpha)}{\sqrt{\alpha(1-\alpha)}} + o(1)) \to Prob(Z \ge \frac{-zf(q_\alpha)}{\sqrt{\alpha(1-\alpha)}})$$

$$= Prob(\frac{\sqrt{\alpha(1-\alpha)}}{f(q_\alpha)}Z \ge -z)$$

$$= Prob(\frac{\sqrt{\alpha(1-\alpha)}}{f(q_\alpha)}Z \le z).$$

Hence, $\sqrt{n}(\hat{\zeta}_{\alpha,n} - q_\alpha)$ converges in distribution to $N(0, \delta_2)$ with $\delta_2 = \frac{\alpha(1-\alpha)}{f^2(q_\alpha)}$ as $n \to \infty$. □

In this case when the correct model is specified, one has $\delta_2 > \delta_1$. That is, the nonparametric estimator $\hat{\zeta}_{\alpha,n}$ is asymptotically less efficient than the MLE $\hat{q}_{\alpha,n}$. This follows from the fact that the MLE asymptotically attains the Cramér-Rao lower bound for variances among all asymptotically unbiased estimators of $F^{-1}(\alpha, \theta_0)$.

Now one considers the case **(B)** when the model is not correctly specified which likely occurs more often than not in real applications. In the case of estimating the quantiles, model misspecification can thus lead to inconsistency.

(B) Suppose that the true density g does not belong to the parametric family $f(\cdot, \theta)$, $\theta \in \Theta$, and $g(x) > 0$ for all $x \in I$. One can show that under this assumption, $F(\cdot, \hat{\theta}_n)$ is not a consistent estimator of G. For this case $\hat{\theta}_n$, called an $M-estimator$, is obtained by maximizing the likelihood function, and $F(\cdot, \hat{\theta}_n)$, converges to $F(\cdot, \theta_0)$ for some θ_0. But since $g(x)$ does not belong to the parametric family, i.e, $g \neq f(\cdot, \theta)$ for any θ, $F(x, \hat{\theta}_n)$ does not converge to $G(x)$ for all $x \in I$.

It follows that $\hat{q}_{\alpha,n}$ is not a consistent estimator of q_α, for all α.

To consider more closely model misspecification, write the average misspecified log likelihood as

$$\frac{1}{n} \sum_{1 \leq i \leq n} \log f(X_i, \theta). \tag{1.3.1}$$

Its true expectation is

$$\int \log f(x, \theta) g(x) dx. \tag{1.3.2}$$

That is, view $\hat{\theta}_n$ as the $M-estimator$ which maximizes the sum in (1.3.1) (by setting its gradient w.r.t θ equal to zero) and compute the expected derivation of the sum from the maximizer of the integral in (1.3.2). By a general theorem for M-estimators (see, e.g., Van Der Vaart (1998), p.62) $\hat{\theta}_n$ converges to a value θ_0 which minimizes the *Kullback-Leibler (K-L) divergence*

$$D(f(\cdot, \theta), g) = \int \log[g(x)/f(x, \theta)] g(x) dx = \int \log[g(x)] g(x) dx - \int \log f(x; \theta) g(x) dx. \tag{1.3.3}$$

Here $D(f,g) := \int \log[g/f]g$, or its square root, is not a distance in the class of probability densities on I (note that $D(f,g)$ is not equal $D(g,f)$). One can show that $\sqrt{D(f,g)}$ is larger than the L^2-distance between $\sqrt{f}$ and $\sqrt{g}$ which is the square of the *Hellinger distance* between f and g, as is shown using the elementary relations $\log y = 2\log(\sqrt{y}) \leq 2(\sqrt{y}-1)$ for all $y \geq 0$. Then we have

$$-D(f,g) \leq 2\int [\sqrt{(f/g)}-1]g = 2\int \sqrt{(fg)} - 2 = -\int (\sqrt{f}-\sqrt{g})^2. \qquad (1.3.4)$$

Thus, as n increases, $f(\cdot,\hat{\theta}_n)$ approaches $f(\cdot,\theta_0)$, which is the best approximation of g in the sense of minimizing the K-L divergence. By (1.3.4) the L^2 distance between $\sqrt{f(.,\theta)}$ and $\sqrt{g}$ is smaller than $\sqrt{D(f,g)}$. Hence the K-L divergence is not zero unless g is $f(\cdot,\theta_0)$, i.e., unless g belongs to the specified parametric family.

In summary, if the model is correctly specified, then the MLE provides the asymptotically optimal estimator. However, if the model is misspecified, the MLE is not even consistent, although it provides the best prediction that minimizes the K-L divergence. Therefore, in the presence of model misspecification one seeks to mainly use nonparametric inference methodologies which are always asymptotically correct. As will be shown in the coming chapters, our proposed nonparametric methods while bearing asymptotically optimal properties also exhibit strong finite sample performance (Bhattacharya and Lin (2010), Bhattacharya and Lin (2011)).

Chapter 2

Nonparametric estimation: Kernel based

2.1 Introduction

For a general regression model $y_i = F(x_i)+\epsilon_i$, with ϵ_i as zero-mean, independent errors and data $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, there is a substantial literature for estimating $F(x)$ based on kernel functions. The idea underlying kernel estimation is to estimate $F(x)$ for some x as a moving averaging of the data points near x. The closer the data points are to x, the larger the weights that are assigned to those corresponding y values. The precise weights are governed by the chosen kernel function and its bandwidth, the definition and properties of which are given in the next section.

Note that kernel estimation can be adapted to the case of estimating $F(x)$ when it is shape constrained, e.g., in estimating the dose-response and effective dosage curves, $F(x)$ is assumed to be *monotone increasing*. The *DNP* method due to Dette et al. (2005) and Dette and Scheder (2010), which will be studied in detail in Section 2.3.1, is one of the leading kernel-based methods. Another method, due to Gasser and Müller (1988) is introduced in the following section.

2.2 Kernel Method of Gasser and Müller

2.2.1 Kernel Method for General Nonparametric Regression

For a general regression model $y_i = F(x_i) + \epsilon_i$ described above, assume without loss of generality that $x \in [0, 1]$. Then the standard *Nadaraya-Watson kernel estimate* of the curve $F(x)$ is given as

$$\widehat{F}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) y_i}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)}. \tag{2.2.1}$$

Clearly, $\widehat{F}(x)$ is the weighted average of the values y_i $(i = 1, \cdots, n)$ where the weight assigned to the point y_i is given by $\frac{K\left(\frac{x-x_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)}$.

As one can see, h, which is termed the *bandwidth* or smoothing parameter, plays a major role in assigning the weights which in turn affects the value of the estimator. It is assumed that h will converge to zero as the sample size increases. One in general picks a small $h > 0$ according the General Cross Validation algorithm as we will describe later.

The kernel function K is a continuous and symmetric density satisfying the following conditions:

$$\int_{-1}^{1} K(u)du = 1, \int_{-1}^{1} uK(u)du = 0,$$
$$M_2 = \int_{-1}^{1} u^2 K(u)du \neq 0, V = \int_{-1}^{1} K(u)^2 du < \infty.$$

There are different variations of the estimator, depending on the choice of the kernel function and the form of the weights. But the underlying idea is similar, that is, the estimate is a weighted average of the available data points. One of the most popular forms is due to Gasser and Müller (1984) which has the following structure

$$\widehat{F}(x) = \sum_{i=1}^{n} \left(\frac{1}{h} \int_{s_{i-1}}^{s_i} K\left(\frac{x-s}{h} \right) ds \right) y_i. \tag{2.2.2}$$

where s_i is taken to be $\frac{1}{2}(x_i + x_{i+1})$ for $1 \leq i \leq n-1$ and $s_0 = 0$ and $s_n = 1$.

Let $\widehat{y} = [\widehat{F}(x_1), \cdots, \widehat{F}(x_n)]'$. One can write

$$\widehat{\boldsymbol{y}} = S_h \boldsymbol{y}$$

where $\boldsymbol{y} = [y_1, \cdots, y_n]'$. Here the (i,j)-th element of the matrix S_h is given by $h^{-1} \int_{s_{j-1}}^{s_j} K\left(\frac{x_i - s}{h} \right) ds$. The bandwidth is chosen by minimizing the GCV criterion

$$GCV(h) = n^{-1} \sum_{i}^{n} \left(y_i - \widehat{F}(x_i) \right)^2 / \left[1 - n^{-1} tr(S_h) \right]^2, \tag{2.2.3}$$

where $tr(S_h)$ is the trace of the matrix S_h. Note that the idea of finding the bandwidth h in minimizing the $GCV(h)$ is roughly the same as picking an h that minimizes the *integrated mean squared error* (MISE) which measures the overall fit of an estimator to the entire curve.

For the choice of the kernel function, one can pick the one which minimizes the risk function. Gasser and Müller (1979) proves the quadratic kernel is optimal with two sign changes in $[-1, 1]$. To avoid oscillations, it is assumed that the kernel is positive on $[-1, 1]$. Then the optimal kernel is $\frac{3}{4}(1-u^2) I_{[-1,1]}(u)$. There are some other choices such as the Gaussian kernel, the uniform kernel and the biweight kernel. When the sample is large, the choice of the kernel function does not make much difference in terms of performance.

The asymptotic properties for the kernel estimator of Gasser and Müller (1984) are explored in the following Lemmas, Corollaries and Theorems. We follow Eubank (1999) for most of the proof. We assume that the design for the x_i's are uniform with $x_i = \frac{2i-1}{2n}$ and $Var(\epsilon_i) = \sigma^2$ for $i = 1, \cdots, n$.

Lemma 2.2.1. *If $F \in C^1[0,1]$, then*

$$E(\widehat{F}(x)) = \frac{1}{h}\int_0^1 K\left(h^{-1}(x-s)\right) F(s)ds + O(n^{-1}). \tag{2.2.4}$$

Proof. Note that $E(y_i) = F(x_i)$ for $i = 1, \cdots n$. Then one has

$$E\left(\widehat{F}(x)\right) = \sum_i^n F(x_i)\left(h^{-1}\int_{s_{i-1}}^{s_i} K\left(\frac{x-s}{h}\right) ds\right). \tag{2.2.5}$$

Write

$$\frac{1}{h}\int_0^1 K\left(h^{-1}(x-s)\right) F(s)ds = \sum_{i=1}^n \left(h^{-1}\int_{s_{i-1}}^{s_i} K\left(\frac{x-s}{h}\right) F(s)ds\right).$$

Since $F(x)$ is continuous and the integral $\int_{s_{i-1}}^{s_i} K\left(\frac{x-s}{h}\right) ds$ is positive over $[s_{i-1}, s_i]$ thus does not change sign over $[s_{i-1}, s_i]$, one can apply the mean value theorem for

the integral which yields the following:

$$\sum_{i=1}^{n}\left(h^{-1}\int_{s_{i-1}}^{s_i}K\left(\frac{x-s}{h}\right)F(s)ds\right)=\sum_{i=1}^{n}F(\epsilon_i)\left(h^{-1}\int_{s_{i-1}}^{s_i}K\left(\frac{x-s}{h}\right)ds\right) \quad (2.2.6)$$

where ϵ_i is some value between s_{i-1} and s_i.

Taking the difference between (2.2.5) and (2.2.6), one has

$$\left|E(\widehat{F}(x))-\frac{1}{h}\int_0^1 K\left(h^{-1}(x-s)\right)F(s)ds\right|=\frac{1}{h}\Big|\sum_{i}^{n}(F(x_i)-F(\epsilon_i))\int_{s_{i-1}}^{s_i}K\left(\frac{x-s}{h}\right)ds)\Big|.$$

Now let $u=\dfrac{x-s}{h}$, $u_i=\dfrac{x-s_i}{h}$ so that $\int_{s_{i-1}}^{s_i}K\left(\frac{x-s}{h}\right)ds=-h\int_{u_{i-1}}^{u_i}K(u)du$. Then one has for the difference in (2.2.7),

$$\begin{aligned}\frac{1}{h}\Big|\sum_{i}^{n}(F(x_i)-F(\epsilon_i))\int_{s_{i-1}}^{s_i}K\left(\frac{x-s}{h}\right)ds)\Big| &\le n^{-1}\max_{x\in[0,1]}\left|F^{'}(x)\right|\sum_{i}^{n}\Big|\int_{u_{i-1}}^{u_i}K(u)du\Big| \\ &\le n^{-1}\max_{x\in[0,1]}\left|F^{'}(x)\right|\max_{u\in[-1,1]}|K(u)|\sum_{u_i\in[-1,1]}|u_i-u_{i-1}| \\ &= O(1/n). \quad (2.2.7)\end{aligned}$$

□

Corollary 2.2.2. *If $F\in C^2[0,1]$ then*

$$E\left(\widehat{F}(x)\right)-F(x)=\frac{h^2}{2}F^{''}(x)M_2+o(h^2)+O(n^{-1}). \quad (2.2.8)$$

Proof. According to Lemma 2.2.1,

$$E(\widehat{F}(x))=\frac{1}{h}\int_0^1 K\left(h^{-1}(x-s)\right)F(s)ds+O(n^{-1}).$$

Let $\dfrac{x-s}{h}=u$; then

$$\frac{1}{h}\int_0^1 K\left(h^{-1}(x-s)\right)F(s)ds=\int K(u)F\left(x-hu\right)du.$$

Expand $F(x-hu)$ at the point x to achieve

$$\begin{aligned}\int K(u)F(x-hu) &= \int(K(u)F(x)-hF^{'}(x)uK(u)+\frac{F^{''}(x)}{2h^2}u^2K(u)+o(h^2u^2))du \\ &= F(x)+\frac{h^2}{2}F^{''}(x)M_2+o(h^2), \quad (2.2.9)\end{aligned}$$

where the integrals are over the interval $\left[\frac{x-1}{h}, \frac{x}{h}\right]$. Then it follows

$$E(\widehat{F}(x)) - F(x) = \frac{h^2}{2} F''(x) M_2 + o(h^2) + O(n^{-1}).$$

□

Lemma 2.2.3.

$$VAR\left(\widehat{F}(x)\right) = \frac{\sigma^2}{nh^2} \int_0^1 K^2\left(h^{-1}(x-u)\right) du + O\left((nh)^{-2}\right). \tag{2.2.10}$$

Proof.

$$VAR\left(\widehat{F}(x)\right) = \frac{\sigma^2}{nh^2} \sum_i^n \left[\int_{s_{i-1}}^{s_i} K(h^{-1}(x-u)) du\right]^2. \tag{2.2.11}$$

Use the mean value theorem for the integral to achieve

$$\sum_{i=1}^n \left[\int_{s_{i-1}}^{s_i} K(h^{-1}(x-u)) du\right]^2 - n^{-1} \int_0^1 K^2\left(h^{-1}(x-u)\right) du \tag{2.2.12}$$

$$= \sum_{i=1}^n \left[\int_{s_{i-1}}^{s_i} K(h^{-1}(x-u)) du\right]^2 - \sum_{i=1}^n n^{-1} \int_{s_{i-1}}^{s_i} K^2\left(h^{-1}(x-u)\right) du$$

$$= \sum_{i=1}^n \left[(s_i - s_{i-1}) K\left(\frac{x-\theta_i}{h}\right)\right]^2 - \sum_{i=1}^n n^{-1}(s_i - s_{i-1}) K^2\left(h^{-1}(x-\epsilon_i)\right) \tag{2.2.13}$$

$$= \sum_{i=1}^n (s_i - s_{i-1}) \left[(s_i - s_{i-1}) K^2\left(\frac{x-\theta_i}{h}\right) - n^{-1} K^2\left(\frac{x-\epsilon_i}{h}\right)\right],$$

where θ_i and ϵ_i are values between s_{i-1} and s_i, $i = 1, \ldots, n$.

Since the design is uniform with length $\frac{1}{2n}$, i.e $x_i = \dfrac{2i-1}{2n}$, then

$$\sum_{i=1}^n (s_i - s_{i-1}) \left[(s_i - s_{i-1}) K^2\left(\frac{x-\theta_i}{h}\right) - n^{-1} K^2\left(\frac{x-\epsilon_i}{h}\right)\right] \tag{2.2.14}$$

$$= n^{-2} \Big| \sum_{i=1}^n \left[K^2\left(\frac{x-\theta_i}{h}\right) - K^2\left(\frac{x-\epsilon_i}{h}\right)\right] \Big| \leq n^{-2} \sum_{i=1}^n \Big| K^2(u_i) - K^2(v_i) \Big|,$$

where $u_i = \dfrac{x-\theta_i}{h}$ and $v_i = \dfrac{x-\epsilon_i}{h}$. Since K is compactly supported and has continuous first derivative over $[-1,1]$,

$$\left|K^2(u_i) - K^2(v_i)\right| \leq C\left|u_i - v_i\right|$$

for some constant C. Then

$$\begin{aligned} n^{-2}\sum_i^n \left|K^2(u_i) - K^2(v_i)\right| &\leq n^{-2}C \sum_{u_i,v_i\in[-1,1]} \left|u_i - v_i\right| \leq n^{-2}(nh)^{-1}O(nh) \\ &= O(n^{-2}), \end{aligned} \tag{2.2.15}$$

since the cardinality of the set of u_i, v_i which fall in $[-1,1]$ is $O(nh)$. Thus comparing the disparity with (2.2.11), we have

$$VAR\left(\widehat{F}(x)\right) = \frac{\sigma^2}{nh^2}\int_0^1 K^2\left(h^{-1}(x-u)\right)du + O\left((nh)^{-2}\right).$$

For the mean integrated square error,

$$\begin{aligned} &E\left(\int_0^1 \left(\widehat{F}(x) - F(x)\right)\right)^2 = \int_0^1 \left(VAR\left(\widehat{F}(x)\right) + bias^2\left(\widehat{F}(x)\right)\right)dx \\ &= \int_0^1 \left(\frac{h^2}{2}F''(x)M_2 + o\left(h^2\right) + O(n^{-1})\right)^2 dx \\ &+ \int_0^1 \left(\frac{\sigma^2}{nh^2}dx\int_0^1 K^2\left(h^{-1}(x-u)\right)du + O\left((nh)^{-2}\right)\right)dx. \end{aligned} \tag{2.2.16}$$

Minimizing this with respect to the bandwidth h, shows that $h \sim n^{-1/5}$ and the MISE $\sim n^{-4/5}$. □

Theorem 2.2.4. *(Müller and Schmitt 1988) Let $nh^5 \to \gamma^2$ for some $\gamma > 0$. Then*

$$(nh)^{1/2}\left(\widehat{F}(x) - F(x)\right) \xrightarrow{L} N\left(\gamma F^{(2)}(x)M_2, F(x)(1-F(x)V)\right)$$

where $V = \int (K(u))^2\,du$.

Proof. The Theorem is proven by verifying the Lindeberg condition for asymptotic normality. Note that if $h = o(n^{-1/5})$, say $h = cn^{-1/5}/\log n$ $(n > 1)$, then $\sqrt{nh}(E\widehat{F}(x) - F(x)) \to 0$, as $n \to \infty$. □

Making use of the asymptotic normality of the estimator, a confidence interval of level $1-\alpha$ for $F(x)$ can be constructed as

$$[\widehat{F}(x) - \phi^{-1}(1-\alpha/2)\eta, \widehat{F}(x) + \phi^{-1}(1-\alpha/2)\eta]$$

where

$$\eta^2 = \widehat{F}(x)\left(1-\widehat{F}(x)\right)\sum_{i}^{n}\left(\frac{1}{h}\int_{s_{i-1}}^{s_i} K\left(\frac{x-s}{h}\right)ds\right)^2, \tag{2.2.17}$$

since one can show $nh\eta^2 \to F(x)(1-F(x))$ in probability as $n \to \infty$.

2.2.2 The Shape-constrained Case

Now one wishes to apply the above kernel estimators to our bioassay problem, where there are typically some dosage levels $x_1 < x_2 < \ldots < x_m$ and n_i $(1 \le i \le m)$ subjects are exposed to each dosage level with r_i responses. The responses are measured in a binary scale with 1 for response and 0 for non-response. Here we denote $\hat{p}_i = \frac{r_i}{n_i}$, which is the proportion of responses at level x_i. The *dose-response curve* $F(x)$ at x is defined to be the *probability of observing* 1 *at dosage* x, which is assumed to be *monotone increasing* with the biological background in mind.

Note that the estimate $\widehat{F}(x)$ defined above is not necessarily increasing. Therefore it cannot be directly applied in estimating the dose-response curve or the effective dosage curve. Some sort of 'monotonization' is required. A monotonized version of the kernel estimate for the effective dosage ED_α (Gasser and Müller (1988)) is given by

$$\widetilde{ED}(\alpha) = \left\{(F^{-1}(\alpha), \alpha) \mid \alpha \in [0,1], \quad F^{-1}(\alpha) = \frac{1}{2}\left(\inf M_\alpha + \sup M_\alpha\right)\right\} \tag{2.2.18}$$

where $M_\alpha = \left\{x : \hat{F}(x) = \alpha\right\}$.

The asymptotic results of the estimates $\widetilde{ED}_\alpha$ are illustrated in the following two Theorems (See Gasser and Müller (1988)).

Theorem 2.2.5. *Assume that* $ED_\alpha \in [\delta, 1-\delta]$. *Also assume the kernel function* K *and its derivative* K' *are Lipschitz continuous and* $\lim_{n\to\infty} \log n/nh^{2j+1} = 0$ *for* $j = 0, 1$, *then* $\widetilde{ED}_\alpha \to ED_\alpha$ *a.s. as* $n \to \infty$. *The rate of convergence is given by*

$$\sup_\alpha \left|\widetilde{ED}_\alpha - ED_\alpha\right| = O((\log n/nh)^{1/2} + h^2) \text{ a.s.} \tag{2.2.19}$$

Theorem 2.2.6. *Assume that* $ED_\alpha \in [\delta, 1-\delta]$ *and* $nh^5 \to \gamma^2$ *as* $n \to \infty$, *and also assume the kernel function* K *and its derivative* K' *are Lipschitz continuous. Then*

$$(nh)^{1/2}\left(\widetilde{ED}_\alpha - ED_\alpha\right) \to N\left(\frac{\gamma F''(ED_\alpha)B}{F'(ED_\alpha)}, \frac{\alpha(1-\alpha)}{F'(ED_\alpha)^2}\right) \text{ in distribution} \tag{2.2.20}$$

as $n \to \infty$.

2.3 The DNP Method.

2.3.1 The DNP Method and its Asymptotics

In this section, we introduce another important kernel-based method in estimating the effective dosage curve, namely, the *DNP method* (following the terminology in Dette and Scheder (2010)). One of the advantages of this method is that it gives an estimator of the effective dosage curve without having to invert an estimated does-response curve.

As before, we have data (x_i, r_i) where the x_i's are the dosages and the r_i's are the sample proportions of response out of n_i subjects at dosage levels x_i. Dette et al. (2005) assume that $n_i = 1$ for all i. Let the response to x_i be y_i (=0 or 1) and, for simplicity, assume $0 = x_1 < x_2 < \ldots < x_m = 1$ to be equidistant. The local linear estimator is obtained first by finding the solution to the following minimization problem: for a small $h > 0$, find the minimizer of

$$\min_{\beta_1,\beta_2} \sum_{i=1}^m K\left(\frac{x - x_i}{h}\right)(y_i - \beta_1 - \beta_2(x_i - x))^2, \tag{2.3.1}$$

where $K(x)$ is a symmetric density on the real line $\mathbb{R}$ with a finite second moment, and h is the bandwidth. The estimator $\hat{\beta}_1(x)$ of β_1 is the estimator $\widehat{F}(x)$ for $F(x)$. The motivation comes from the fact that the standard kernel estimator of $F(x)$, namely,

$$\sum_{i=1}^{m} y_i K\left(\frac{x-x_i}{h}\right) / \sum_{i=1}^{m} K\left(\frac{x-x_i}{h}\right)$$

is obtained by minimizing (w.r.t θ)

$$\sum_{i=1}^{m} K\left(\frac{x-x_i}{h}\right)(y_i-\theta)^2.$$

Thus it is natural to think of a local polynomial approximation of the objective function. In particular, one may use a local linear approximation, as in (2.3.1).

The α-th quantile $ED_\alpha = F^{-1}(\alpha)$ is then estimated as

$$\widehat{ED}_\alpha = \int_0^1 \int_{-\infty}^{\alpha} \frac{1}{h_d} K_d\left(\frac{\hat{F}(x)-u}{h_d}\right) du\, dx, \tag{2.3.2}$$

where h_d is small. Here K_d is a symmetric kernel with the same properties as K (e.g, $K_d = K$). But h and h_d are not of the same order, as shown below. Note that as $h_d \downarrow 0$, $\widehat{ED}_\alpha$ converges to ED_α. To understand this, observe that $\frac{1}{h_d} K_d\left(\frac{\hat{F}(x)-u}{h_d}\right) du$ converges to the Dirac measure $\delta_{\hat{F}(x)}(du)$, as $h_d \downarrow 0$, so that the inner integral converges to the indicator function $1_{[\hat{F}(x)\le\alpha]}(x)$. The outer integral of this limit is the Lebesgue measure of the set $\{x : \hat{F}(x) \le \alpha\}$ which equals the length of this interval , namely $\hat{F}^{-1}(\alpha)$. This method of monotonization of a function $\widehat{F}$ is called *monotone or measure-preserving rearrangement* in Hardy et al. (1952).

The following lemma and theorem due to Dette et al.(2005) provide the asymptotic behavior of the estimate $\widehat{ED}_\alpha$.

Lemma 2.3.1. *Let the dose-response curve be twice continuously differentiable and strictly increasing. Make the following assumptions:*

i. $h \to 0, h_d \to 0$

ii. $mh_d \to \infty, mh \to \infty$
iii. $\lim_{h_d \to 0, h \to 0} \frac{h}{h_d} = \infty$
iv. $mh^5 = O(1)$
v. $\frac{1}{mh\, h_d^2} = o(1)$
Then one has, for any $\alpha \in (F(0), F(1))$ *with* $F'(F^{-1}(\alpha)) > 0$,

$$E[\widehat{ED}_\alpha] = b_{h,h_d} + o(h_d^2) + o(h^2)$$

and

$$VAR(\widehat{ED}_\alpha) = \frac{s(\alpha)^2}{mh} + o(\frac{1}{mh}),$$

where

$$b_{h,h_d} = F^{-1}(\alpha) + \kappa_2(K_d) h_d^2 (F^{-1})''(\alpha) - \kappa_2(K) h^2 \left(\frac{F''}{F'}(F^{-1}(\alpha)) \right)$$

and

$$s^2(\alpha) = \frac{\alpha(1-\alpha)}{(F'(F^{-1}(\alpha)))^2} \int_{-\infty}^{\infty} K^2(u) du,$$

with the constant $\kappa_2(K)$ *given by*

$$\kappa_2(K) = \frac{1}{2} \int_{-\infty}^{\infty} v^2 K(v)\, dv.$$

Thus the asymptotic bias is given by

$$E[\widehat{ED}_\alpha - F^{-1}(\alpha)] = -\kappa_2(K) h^2 (\frac{F''}{F'}(F^{-1}(\alpha))) + o(h^2). \qquad \square$$

Theorem 2.3.2. *If the dose-response curve is strictly increasing and the same conditions as in lemma 2.3.1 hold, then one has, for any* $\alpha \in (F(0), F(1))$ *with* $F'(F^{-1}(\alpha)) > 0$,

$$\sqrt{mh}(\widehat{ED}_\alpha - b_{h,h_d}) \Rightarrow N(0, s^2(\alpha))$$

where b_{h,h_d} *and the variance* $s^2(\alpha)$ *are defined as in the previous lemma.* $\square$

Remark 2.3.3. As in nonparametric curve estimation in general, observe that the bias $b_{h,h_d} - F^{-1}(\alpha)$ is negligible only if $\sqrt{mh}h^2 \to 0$, i.e., if $h = o(m^{-1/5})$. With this correction, the asymptotic variance of $\widehat{ED}_\alpha$ is of larger order than $m^{-4/5} \equiv N^{-4/5}$.

The above method is extended to the case $n_i \geq 1$ for all i in Dette and Scheder (2010), by replacing y_i by r_i in (2.3.1), and letting $\widehat{F}(x) = \dfrac{\widehat{\beta}_1(x)}{n}$. Note that this method is designed for cases when there are more dosage levels than the number of observations (say $n = 1$ for example) at each dosage level. A comparison study is carried out by Dette and Scheder (2010) among the DNP method, the kernel based method due to Müller and Schmitt (1988), a local smoothing estimator due to Park and Park (2006) and the B-K method due to Bhattacharya and Kong (2007), in which the data are simulated from selected parametric models and the comparison is carried out in terms of the MISE (mean integrated squared error) of the simulated data. It seems that the DNP method outperforms other methods. We will describe in the coming chapters the B-K method which belongs to a different class of methods that are not kernel based.

2.3.2 Application of the DNP Method to the Space Shuttle Disaster Problem.

As an illustration, we apply the DNP method to estimating the probability of O-ring failure under certain temperatures for the space shuttle disaster problem. As remarked above, the DNP method is perhaps more suitable for the design or situations where there are more 'levels' then the number of observations for each level, which is the case for the space shuttle data. We will first describe the statistical problem.

It was determined that in the disaster of the space shuttle Challenger, the explosion of the shuttle was the result of O-ring failure, a splitting of a ring of rubber that seals different parts of the external rocket motors together. The flight accident was believed to be caused by the unusually cold weather (31^0F at the time of the launch).

Previous O-ring failure data along with temperature at launch time are given in the Tables 2.3.2 and 2.3.2 below (in increasing order of temperature) for 23 prior flights, with the flight numbers denoting the time order of the launch. The numbers 0 and 1 indicate "no O-ring failure" and " O-ring failure" respectively. The object of the study is to determine the probability of O-ring failure under certain temperatures and carry out inference in terms of confidence intervals. Specifically, given the response probability 0.9 of "no O-ring failure", or probability 0.1 of "O-ring failure", we aim to estimate the corresponding temperature using the DNP method. We denote this target temperature as $T_{0.1}$. We also wish to calculate the bootstrap 95% confidence interval on $T_{0.1}$. The results are compared with those obtained by fitting a standard logistic regression to the data.

After calculation, the estimate of $T_{0.1}$ for the data in Tables 2.3.2-2.3.2 is given as 76.4760. The 95% bootstrap confidence interval for this temperature given the level of probability of 'no O- ring failure' of 0.9 is given by [66.8798, 81]. For the 95% *one sided* confidence limit, the lower bound is 69.3642, which is the 50th quantile of 1000 bootstrap estimates of $T_{0.1}$.

We also fit a logistic regression model to the data, and the point estimate of $T_{0.1}$ is 74.2344, obtained by maximum likelihood estimation. The 95% bootstrap confidence interval by fitting the logistic model to the bootstrap data is given by [64.3721 , 80.37]. A standalone *one-sided* bootstrap confidence limit (the lower limit) is given as 66.9217.

TABLE 2.1. Data from prior flights

Flight #	14	9	23	10	1	5	13	15	4	3	8	17
Failure:	1	1	1	1	0	0	0	0	0	0	0	0
Temperature	53	57	58	63	66	67	67	67	68	69	70	70

TABLE 2.2. Data from prior flights

Flight #	2	11	6	7	16	21	19	22	12	20	18
Failure:	1	1	0	0	0	1	0	0	0	0	0
Temperature	70	70	72	73	75	75	76	76	78	79	81

Chapter 3

Nonparametric inference based on splines

3.1 Nonparametric Monotone Spline Smoothing

There is another important nonparametric method for quantal bioassay based on splines; see, e.g., Silverman (1985), Wahba (1990), Kelly and Rice (1990), Kong and Eubank (2007). Given data (x_i, y_i) $(i = 1, \cdots, m)$ from the model $y_i = F(x_i) + \epsilon_i$ with $x_1 \leq \cdots \leq x_m$ (ϵ_i's independent, mean 0), the general smoothing spline problem is to find the estimate $\widehat{F}_h$ of the function F that minimizes the objective function (over the class of twice differentiable functions)

$$\mathcal{J} = \frac{1}{m}\sum_{i=1}^{m}(y_i - F(x_i))^2 w_i + h\int_{x_1}^{x_m} F''(x)^2 dx, \tag{3.1.1}$$

where w_i are positive weights and h is a smoothing parameter which controls the trade off between the smoothness of the curve and the fidelity to the data.

The existence and characterizations of the solution to (3.1.1) are given in references such as Eubank (1999) or Wahba (1990). The idea of showing the existence of such solutions is to tie the minimization problem to a regularization problem in a reproducing kernel Hilbert space. In this special case, the solution turns out to be a natural cubic spline with knots at $x_1, x_2, \cdots, x_m$.

When F is *assumed to be monotone,* which is the case of interest to our dose-response problem, a complete characterization of the solution is still unknown, though Utreras (1985) shows the existence of the solution and gives a partial characterization. We adopt the approach by Kong and Eubank (2007) which assumes that F is a linear combination of a B-spline basis, and the monotonicity of the estimate is guaranteed by imposing a monotone constraint on the coefficients. Here we proceed to give a detailed description of the approach.

Given a *knot sequence* $\tau_1 < \tau_2 < .... < \tau_n$, functions in the spline basis $\{B_{j,4}\}$ are defined recursively as follows:

$$B_{j,1}(x) = \begin{cases} 1 & \tau_j \leq x \leq \tau_{j+1} \\ 0 & \text{otherwise} \end{cases}$$

$$B_{j,l}(x) = \frac{x - \tau_j}{\tau_{j+l-1} - \tau_j} B_{j,l-1}(x) + \frac{\tau_{j+l} - x}{\tau_{j+l} - \tau_{j+1}} B_{j+1,l-1}(x) \ \ (l = 2, 3, 4).$$

Note that such a definition can be extended to higher degrees if needed, i.e., by letting $l = 1, 2, \cdots, k$. Here we focus on the cubic spline case, i.e., when $k = 4$.

The collection of B-Splines forms a basis for the spline function. There are different systems of basis functions used by other authors for smoothing splines, such as exponential functions, tangent functions, and cosine and sine functions. But B-splines are locally very flexible, for example, compared to polynomial bases. The latter have one serious limitation: the lack of flexibility in the sense that changing the behavior of F near one value x_1 has radical implications for the behavior for any other value x_2. Other than its local smooth properties, the space spanned by the cubic spline basis includes all polynomials of degree 3. (See De Boor (2001)).

Here are some properties of the $B-$spline basis ($k = 4$):

(1) Support: $B_{j,k}$ is positive on the interior of the interval (τ_j, τ_{j+k}) and vanishes outside it.

(2) On the interval (τ_j, τ_{j+1}), at most $k + 1$ B-splines functions are non-zero; these are $B_{j-k,k}$, $B_{j-k+1,k}$, $\cdots$, $B_{j,k}$.

(3) Partition of unity: $\sum_j B_{j,k}(x) = 1$ for any $x \in [\tau_1, \tau_N]$.

$m + 2$ of such B-Splines $B_{j,4}$ ($j = -2, -1, 1, \cdots, m$) are used with knots taken to be $x_{-2}, x_{-1}, x_0,\ x_1, x_2, \cdots, x_m, x_{m+1}, x_{m+2}, x_{m+3}$. The first three and the last three knots are taken to be equidistant with $x_2 - x_1 = x_1 - x_0 = x_0 - x_{-1} = x_{-1} - x_{-2}$ and

$x_m - x_{m-1} = x_{m+1} - x_m = x_{m+2} - x_{m+1} = x_{m+3} - x_{m+2}$. Therefore, $F(x)$ on $[x_1, x_m]$ can be written as a linear combination of this B-spline basis, which is

$$F(x) = \sum_{j=-2}^{m-1} \beta_j B_{j,4}(x), \tag{3.1.2}$$

Note that

$$F'(x) = \sum_{j=-2}^{m} 3\frac{\beta_j - \beta_{j-1}}{x_{j+3} - x_j} B_{j,3}(x).$$

Therefore, it is clear $F(x)$ a is monotone function under the constraint that $\beta_{-2} \leq \beta_{-1} \leq \cdots \leq \beta_{N-1}$. Then, the problem boils down to finding the solutions to the system

$$\begin{cases} \min \quad \frac{1}{m}\sum_{i=1}^{m}(y_i - F(x_i))^2 w_i + h\int_{x_1}^{x_m} F''(x)^2 dx, \\ \text{subject to} \quad F(x) = \sum_{j=-2}^{m-1} \beta_j B_{j,4}(x) \text{ with } \beta_{-2} \leq \beta_{-1} \leq \cdots \leq \beta_{N-1}. \end{cases} \tag{3.1.3}$$

The existence of such a solution can be easily seen when the above system is rewritten in matrix form. The problem can be solved using quadratic programming with software packages such as Matlab.

It seems that there does not exist an adequate asymptotic distribution theory for splines to construct confidence intervals, rates of MISE, etc. Indeed, the derivation of precise asymptotic inference with splines is difficult (see, e.g., Kelly and Rice (1990) and Eubank (1999)).

For constructing confidence intervals using monotone spline estimates, Kong and Eubank (2007) proposed a form of parametric Bayesian inference for deriving confidence intervals. In the article Bhattacharya and Lin (2012), bootstrapping is proposed to construct confidence intervals which makes the procedure *fully nonparametric*. Since the estimated spline curve is smooth, bootstrapping is valid here. For the smoothing parameter h, the ideal estimate of the smoothing parameter is given by the GCV (generalized cross validation) algorithm.

We now show that the set of all monotone B-splines of order k is dense in the space of all Lipchitzian monotone functions on [0,1]. The case of cubic splines follows when $k = 4$.

We prove the theorem for the general case, not just for the cubic spline basis. The case of cubic splines follows when $k = 4$.

First denote

$$S_\tau^m = \left\{ \sum_i \beta_i B_{i,k},\ \beta_1 \le \beta_2 \le \cdots \le \beta_J,\ \beta_i \in \mathbb{R},\ i = 1, \cdots, J \right\}. \tag{3.1.4}$$

Theorem 3.1.1. *Let F_0 be any Lipchitzian monotone function on [0,1]. Let S_τ^m be the space of monotone splines with the knots sequence $\{\tau_1, \cdots, \tau_N\}$ as defined above. Then there exists an element $PF_0 \in S_\tau^m$ such that* $\sup |F_0 - P(F_0)| \le C\ J^{-1} \sup_{x \ne y} \dfrac{|F_0(x) - F_0(y)|}{|x - y|}$ *for some constant $C > 0$.*

Proof. Given $t_1 \le t_2 \cdots \le t_J$ in [0,1], let $PF_0 = \sum_{i=1}^J F_0(t_i) B_i$. One can verify that $PF_0 \in S_\tau^m$ since F_0 is monotone.

For any value $x \in [\tau_j, \tau_{j+1})$,

$$PF_0(x) = \sum_{i=j+1-k}^{j} F_0(t_i) B_i(x)$$

since only $B_{j+1-k}, \cdots, B_j$ are supported in $[\tau_j, \tau_{j+1})$. Note that for any $x \in [\tau_j, \tau_{j+1})$, due to the partition property of unity of B-splines, one has $\sum_{i=j+1-k}^{j} B_i(x) = 1$, which implies that

$$F_0(x) = F_0(x) \sum_{i=j+1-k}^{j} B_i(x) = \sum_{i=j+1-k}^{j} F_0(x) B_i(x).$$

Then it follows that

$$F_0(x) - PF_0(x) = \sum_{i=j+1-k}^{j} (F_0(x) - F_0(t_i)) B_i(x).$$

Therefore,

$$\begin{aligned} |F_0(x) - PF_0(x)| &\le \sum_{i=j+1-k}^{j} |(F_0(x) - F_0(t_i))| B_i(x) \\ &\le \max \left\{ |F_0(x) - F_0(t_i)|,\ j - k + 1 \le i \le j \right\}. \end{aligned}$$

Such a bound clearly depends on the choice of the values $t_1, \cdots, t_J$. Note that the support of B_i is $[\tau_i, \tau_{i+k})$. One can choose t_i to be close to the support of B_i, e.g., let $t_i = \tau_{i+k/2}$ and define $\tau_{i+k/2} := (\tau_{i+(k-1)/2} + \tau_{i+(k+1)/2})/2$ when k is odd. Under such choices, one has

$$\max\left\{|F_0(x) - F_0(t_i)|\,,\ j-k+1 \le i \le j\right\} \le \max\left\{|F_0(x') - F_0(y')|\right\},$$
$$\text{where } x', y' \in [\tau_{j+1-k/2}, \tau_{j+1}],\ \text{or } x', y' \in [\tau_j, \tau_{j+k/2}).$$

Note that the end point $\tau_{j+1-k/2} = \tau_{j-k+1+k/2}$ is the leftmost point of $t_i, j-k+1 \le i \le j$ and $\tau_{j+k/2}$ the similar rightmost point.

It follows that

$$\begin{aligned}\sup|F_0 - P(F_0)| &\le \lfloor (k+1)/2 \rfloor h \sup_{x' \ne y'} \frac{|F_0(x') - F_0(y')|}{|x' - y'|} \\ &\le \text{const } J^{-1} \sup_{x' \ne y'} \frac{|F_0(x') - F_0(y')|}{|x' - y'|},\end{aligned}$$

where h is the maximum of the mesh sizes $\tau_{j+1} - \tau_j$. Such bounds can be made arbitrarily small by increasing the number of knots.

□

Chapter 4

Nonparametric inference based on isotonic MLEs

4.1 Introduction

In this chapter, we introduce two nonparametric methodologies for estimating the dose-response curve and effective dosage curves based on *isotonic maximum likelihood estimators* which are the MLEs under a monotone constraint on the dose response function $F(x)$. There is a deep connection between isotonic MLEs and the solutions to the general *isotonic regression* problem. In the following section, we will first give a review of the theory of isotonic regression and characterizations of the solutions, and prove the theorem that links the solutions of the isotonic regression problem to isotonic MLEs.

In Sections 4.3 and 4.4, the two above mentioned nonparametric methods are introduced in detail and the asymptotic results are provided with detailed proofs. A comprehensive finite sample study is carried out with all the results displayed in the next chapter.

4.2 Isotonic Regressions

The problem of isotonic regressions is concerned with finding an estimate of the regression function under some order restriction. The theory of isotonic regression is very important in order restricted statistical inference. We refer to the books by Robertson et al. (1988) and Barlow et al. (1972), which give rather comprehensive treatments of the literature in the area of isotonic regression. All the theorems in this section are based on their results.

Definition 4.2.1. A binary relation $\prec$ on X is a *simple order* on X if

1. it is *reflexive*: $x \prec x$ for $x \in X$;
2. it is *transitive*: $x, y, z \in X$, $x \prec y$ and $y \prec z$ imply $x \prec z$;
3. it is *antisymmetric*: $x, y \in X$, $x \prec y$ and $y \prec x$ imply $x = y$; and
4. every *two elements* of X are *comparable*: $x, y \in X$ implies that either $x \prec y$ or $y \prec x$.

A binary relation $\precsim$ is called a *partial-order* if it is *reflexive*, *transitive* and *antisymmetric*. A binary relation $\lesssim$ is called a *quasi-order* if it is *reflexive* and *transitive*. It is clear from the definition that a simple order is also a partial order which in turn is a quasi-order.

Definition 4.2.2. Consider a finite set $X = \{x_1, \cdots, x_m\}$ with the quasi-order $x_1 \lesssim x_2 \lesssim \cdots \lesssim x_m$. A real-valued function f on X is *isotonic* with respect to the quasi order if x, $y \in X$ and $x \lesssim y$ imply $f(x) \leq f(y)$.

We give the following definition on isotonic regression.

Definition 4.2.3. Let g be a given function on X and w a given positive function on X. An isotonic function g^* on X is an *isotonic regression* of g *with weights* w *with respect to the quasi-ordering* $x_1 \lesssim x_2 \lesssim \cdots \lesssim x_m$ if it minimizes, over the class of all isotonic functions f on X, the sum

$$\sum_{x \in X} \left(g(x) - f(x)\right)^2 w(x). \tag{4.2.1}$$

A series of theorems on characterizing the solutions g^* under the general quasi-ordering case will be given. Using these one can show that (4.2.8) below is indeed the solution to the minimization problem (4.2.1) under the case of simple ordering.

Theorem 4.2.4. *An isotonic regression g^* of g with weights w is an isotonic function on X with respect to $\lesssim$ and satisfies*

$$\sum_{x\in X}[g(x)-g^*(x)][g^*(x)-f(x)]w(x)\geq 0 \tag{4.2.2}$$

and

$$\sum_{x\in X}[g(x)-f(x)]^2w(x)\geq\sum_{x\in X}[g(x)-g^*(x)]^2w(x)+\sum_{x\in X}[g^*(x)-f(x)]^2w(x) \tag{4.2.3}$$

for every isotonic function f on X. Conversely, if an isotonic function u satisfies

$$\sum_{x}[g(x)-u(x)][u(x)-f(x)]w(x)\geq 0 \tag{4.2.4}$$

for every isotonic function f on X then u is an isotonic regression of g with weights w. There is at most one such isotonic function.

Proof. First note that the set of isotonic functions is *convex*, i.e., if f_1 and f_2 are isotonic functions then so are $\alpha f_1+(1-\alpha)f_2$ for any $\alpha\in[0,1]$. Since $(1-\alpha)g^*+\alpha f$ is also an isotonic function,

$$\sum_{x\in X}\left(g(x)-((1-\alpha)g^*(x)+\alpha f(x))\right)^2w(x)$$

attaints its minimum at $\alpha=0$. One can view the above weighted sum as a quadratic function of α with $\alpha\in[0,1]$, so that its derivative at $\alpha=0$ is non-negative. Since the derivative is just

$$2\sum_{x\in X}[g(x)-g^*(x)][g^*(x)-f(x)]w(x),$$

we have shown (4.2.2).

Write

$$\sum_{x\in X}\left(g(x)-f(x)\right)^2w(x)=\sum_{x\in X}\left(g(x)-g^*(x)+g^*(x)-f(x)\right)^2w(x),$$

expand the right hand side and use the result of (4.2.2); then (4.2.3) follows.

Now suppose $u(x)$ is an isotonic function that satisfies $\sum_x [g(x) - u(x)][u(x) - f(x)]w(x) \geq 0$ for all isotonic functions f, we want to show that $u(x)$ is an isotonic function of g with weights w and that it is unique. Since $\sum_x [g(x) - u(x)][u(x) - f(x)]w(x) \geq 0$ for all isotonic functions f, one has as above

$$\sum_{x \in X} (g(x) - f(x))^2 w(x) \geq \sum_{x \in X} (g(x) - u(x))^2 w(x) + \sum_{x \in X} (u(x) - f(x))^2 w(x),$$

which implies that $\sum_{x \in X} (g(x) - u(x))^2 w(x) \leq \sum_{x \in X} (g(x) - f(x))^2 w(x)$ for all isotonic functions f. Therefore, $u(x)$ is indeed an isotonic regression of g with weights $w(x)$.

Now we are left with showing that $u(x)$ is unique. Assume to the contrary that there are two isotonic regressions u_1 and u_2 of g. Then by (4.2.2), one has

$$\sum_{x \in X} [g(x) - u_1(x)][u_1(x) - f(x)]w(x) \geq 0,$$

and

$$\sum_{x \in X} [g(x) - u_2(x)][u_2(x) - f(x)]w(x) \geq 0,$$

for all isotonic functions f. Let $f = u_2$ in the first equation, and u_1 for the second equation, then add. Simplifying, one has

$$-\sum_x [u_1(x) - u_2(x)]^2 w(x) \geq 0.$$

Therefore, $u_1(x) = u_2(x)$ for all x. □

Before proving the following theorem, let us remark that the set of isotonic functions is also a *cone*, that is, if f is an isotonic function, so is cf for any c non-negative.

Theorem 4.2.5. *An isotonic function u on X is the isotonic regression of g with weights w if and only if*

$$\sum_x [g(x) - u(x)]u(x)w(x) = 0 \tag{4.2.5}$$

and

$$\sum_x [g(x) - u(x)]f(x)w(x) \leq 0 \tag{4.2.6}$$

for all isotonic f. The isotonic regression g^* satisfies also

$$\sum_x g(x)w(x) = \sum_x g^*(x)w(x). \tag{4.2.7}$$

Proof. If $u(x)$ satisfies (4.2.5) and (4.2.6), then $u(x)$ satisfies also (4.2.4), so that u is an isotonic regression of g and satisfies (4.2.3) by Theorem 4.2.4. To see conversely that $u = g^*$ satisfies (4.2.5), set $f(x) = cg^*(x)$ in (4.2.2), first with a constant $c > 1$ and then with a positive constant $c < 1$. To prove (4.2.7), let $u = g^*$ in (4.2.6) and let f be first the constant 1, then the constant -1. □

Another property of the class of isotonic functions is that it is a *lattice*. To say the class of isotonic function on X is a *lattice* is simply to say that if f and g are isotonic functions on X, so are $f \wedge g$ and $f \vee g$.

Theorem 4.2.6. *If g_1 and g_2 are isotonic functions on X such that $g_1(x) \leq g(x) \leq g_2(x)$ for $x \in X$, then if g^* is an isotonic regression of g, one has $g_1(x) \leq g^*(x) \leq g_2(x)$ for $x \in X$. In particular, if a and b are constants such that $a \leq g(x) \leq b$ for $x \in X$, then also $a \leq g^*(x) \leq b$ for all $x \in X$.*

Proof. Let f be isotonic and let $h = f \vee g_1$, then by the lattice property of the set of isotonic functions, h is also isotonic. Further, if $f(x) \geq g_1(x)$ for a particular $x \in X$, then $h(x) = f(x)$ so that $g(x) - f(x) = g(x) - h(x)$; while if $f(x) < g_1(x)$, then $0 \leq g(x) - h(x) = g(x) - g_1(x) < g(x) - f(x)$. Therefore, for all $x \in X$, $[g(x) - h(x)]^2 \leq [g(x) - f(x)]^2$. As a result,

$$\sum_x [g(x) - h(x)]^2 w(x) \leq \sum_x [g(x) - f(x)]^2 w(x).$$

Letting $f = g^*$, it then follows that $h = g^*$ which implies $g^* \leq g_1$. By similar arguments, one can show that $g^*(x) \leq g_2(x)$ for all x. □

Corollary 4.2.7. *An isotonic regression of g exists.*

Proof. Since X is a finite set, g is bounded with some maximum b and minimum a over X. By the above theorem, it suffices to show that there exists a minimum of $\sum_x [g(x) - f(x)]^2 w(x)$ with $f(x) \in [a, b]$ for all $x \in X$. Then the conclusion follows since the objective function is minimized over the compact set $\{(a_1, \cdots, a_k) : a \leq a_1 \leq \cdots a_k \leq b\}$ of Euclidean space, where k is the number of elements in X. □

Definition 4.2.8. The average of g over a set A with weights w is defined by

$$Av(A) = \sum_{x \in A} g(x)w(x) / \sum_{x \in A} w(x)$$

for $A \neq \emptyset$, $A \subset X$.

Lemma 4.2.9. *If c is a real number and if the subset $[g^* = c]$ of X on which g^* takes the value c is not empty, then $c = Av([g^* = c])$.*

Proof. Write

$$\sum_x [g(x) - g^*(x)]^2 w(x) = \sum_{[g^* \neq c]} [g(x) - g^*(x)]^2 w(x) + \sum_{[g^* = c]} [g(x) - c)]^2 w(x).$$

Now consider $\sum_{[g^*=c]} [g(x) - t)]^2 w(x)$ as a function of t which attains its minimum at $t = Av([g^* = c])$. One claims that $c = Av([g^* = c])$, or else one can find another isotonic function $\tilde{g}$ which coincides with g^* on $[g^* \neq c]$ but different from g^* on $[g^* = c]$. To do so, let $\tilde{g}$ have a value, say c' on this set which is a little closer to $Av([g^* = c])$ than c, but does not violate isotonicity. Then $\tilde{g}$ has a smaller weighted minimum than g^*, a contradiction. □

Theorem 4.2.10. *For an arbitrary real valued function ψ on the reals,*

$$\sum_x [g(x) - g^*(x)] \psi[g^*(x)] w(x) = 0.$$

Proof. This theorem is an extension of the special case we proved before. One can write

$$\sum_x [g(x) - g^*(x)]\psi[g^*(x)]w(x) = \sum_c \psi(c) \sum_{[g^*=c]} [g(x) - c]w(x),$$

for by the above Lemma

$$\sum_{[g^*=c]} [g(x) - c]w(x) = 0$$

for all real numbers c. □

Making use of the theorems above, we will prove two different characterizations of the solutions g^* with one in terms of the max min formula and the other in terms of the left slope of the *greatest convex minorant* (GCM) of the *cumulative sum diagram* (CSD).

Definition 4.2.11. Given a quasi-order $\lesssim$ on X, a subset L of X is a *lower set* with respect to the quasi-order $\lesssim$ if $y \in L$, $x \in X$, $x \lesssim y$ imply $x \in L$. A subset U of X is an *upper set* if $x \in U$, $y \in X$, and $x \lesssim y$ imply $y \in U$. Let $\mathcal{L}$ denote the class of all lower sets and $\mathcal{U}$ the class of all upper sets.

One characterization of the solution g^* under the simple ordering is given by the following max min formula

$$g^*(x_i) = \max_{s \le i} \min_{t \ge i} \frac{\sum_{r=s}^t g(x_r)w(x_r)}{\sum_{r=s}^t w(x_r)}. \tag{4.2.8}$$

We first prove the following theorem which is Theorem 1.4.3 in Robertson et al. (1988) for the quasi-ordering case where the simple ordering case follows as a special case. Then we will prove that (4.2.8) is indeed the solution of the minimization problem (4.2.1).

Theorem 4.2.12. *If g^* is the isotonic regression of G then for each real number a,*

$$Av(L \cap [g^* \ge a]) \ge a,$$

$$Av(L \cap [g^* > a]) > a,$$

$$Av(U \cap [g^* \leq a]) \leq a,$$

and

$$Av(U \cap [g^* < a]) \leq a,$$

for any $L \in \mathcal{L}$, and $U \in \mathcal{U}$ such that the set over which the weighted average is taken is not empty. Here $Av(W)$ is the weighted average over the elements in the set W.

Proof. First observe that, for any a, b such that $-\infty \leq a < b \leq \infty$, one has

$$\sum_{x \in [a < g^* < b]} [g(x) - g^*(x)]w(x) = 0$$

which follows from the Lemma which says

$$\sum_{g^* = c} [g(x) - g^*(x)]w(x) = 0.$$

Also note that $-1_L(\cdot)$ is an isotonic function, given a lower set L. Then it follows that

$$-\sum_{x \in L} [g(x) - g^*(x)]w(x) = \sum_{x \in X} [g(x) - g^*(x)][-1_L(x)]w(x) \leq 0,$$

as a consequence of Theorem 4.2.5. Note also that $U^c \in \mathcal{L}$ if $U \in \mathcal{U}$, so that -1_{U^c} is isotonic. Now suppose $a \in \mathbb{R}$ and $U \cap [g^* < a] \neq \emptyset$. Then

$$\begin{aligned} \sum_{U \cap [g^* < a]} [g(x) - a]w(x) &< \sum_{U \cap [g^* < a]} [g(x) - g^*(x)]w(x) \\ &= \sum_{x \in [g^* < a]} [g(x) - g^*(x)]w(x) - \sum_{U^c \cap [g^* < a]} [g(x) - g^*(x)]w(x) \leq 0, \end{aligned}$$

which implies that

$$\sum_{U \cap [g^* < a]} g(x)w(x) < a \sum_{U \cap [g^* < a]} w(x)$$

proving that

$$Av(U \cap [g^* < a]) < a.$$

For the case $g^* \leq a$, the same argument is used, but with the strict inequality $<$ replaced by $\leq$. Similar arguments can be used to show the rest of the inequalities. $\square$

Now we are ready to prove the following theorem that gives the closed form for the isotonic regression g^* under the general quasi-ordering in terms of the max min formula.

Theorem 4.2.13. *For $x \in X$, one has*

$$\begin{aligned} g^*(x) &= \max_{U:x\in U} \min_{L:x\in L} Av(L\cap U) \\ &= \max_{U:x\in U} \min_{L:L\cap U\neq\emptyset} Av(L\cap U) \\ &= \min_{L:x\in L} \max_{U:x\in U} Av(L\cap U) \\ &= \min_{L:x\in L} \max_{U:L\cap U\neq\emptyset} Av(L\cap U). \end{aligned}$$

Here $L \in \mathcal{L}$ and $U \in \mathcal{U}$.

Proof. Let $g^*(x) = a$. Note that $[g^* \le a] \in \mathcal{L}$ and $[g^* \ge a] \in \mathcal{U}$. By Theorem 4.2.12,

$$\max_{U\mathcal{U},x\in U} \min_{L\in\mathcal{L},x\in L} Av(L\cap U) \le \max_{U\in\mathcal{U},x\in U} Av\left([g^*\le a]\cap U\right) \le a, \tag{4.2.9}$$

and

$$\max_{U\mathcal{U},x\in U} \min_{L\in\mathcal{L},x\in L} Av(L\cap U) \ge \min_{L\in\mathcal{L},x\in L} Av\left([g^*\ge a]\cap L\right) \ge a, \tag{4.2.10}$$

proving the result $\max_{U\mathcal{U},x\in U} \min_{L\in\mathcal{L},x\in L} Av(L\cap U) = g^*(x)$.

Next

$$\max_{U\in\mathcal{U}} \min_{L\cap U\neq\emptyset} Av(L\cap U) \le \max_{U\in\mathcal{U}} Av([g^*\le a]\cap U) = g^* \le a$$

by (4.2.9). Also

$$\max_{U\in\mathcal{U}} \min_{L\cap U\neq\emptyset} Av(L\cap U) \ge \min_{L\cap[g^*\ge a]\neq\emptyset} Av(L\cap[g^*\ge a]) \ge a,$$

by Theorem 4.2.12. Therefore,

$$\max_{U\in\mathcal{U},x\in U} \min_{L\cap U\neq\emptyset} Av(L\cap U) = a = g^*(x).$$

□

Remark 4.2.14. When X is equipped with a simple ordering, then the above Theorem implies that

$$g^*(x_i) = \max_{s \le i} \min_{t \ge i} \frac{\sum_{r=s}^{t} g(x_r)w(x_r)}{\sum_{r=s}^{t} w(x_r)},$$

which is exactly (4.2.8).

We now proceed to establish the connection between solutions to the isotonic regression problem and MLEs for certain problems. The ordered binomial case, which is applied to our problem of estimating the dose-response and effective dosage curve, will be a special case of a general theorem.

Let Φ be a convex function which is finite on an interval I containing the range of the function g and infinite elsewhere, and let ϕ be its derivative. Then ϕ is nondecreasing. For any u, v, let

$$\Delta(u, v) = \Phi(u) - \Phi(v) - (u - v)\phi(v), \text{ if } u, v \in I \tag{4.2.11}$$

and $\Delta(u, v) = \infty$, if $v \in I, u \notin I$. Note that Δ is non-negative. By definition, one can derive the following

$$\Delta(r, t) = \Delta(r, s) + \Delta(s, t) + (r - s)[\phi(s) - \phi(t)], \text{ if } s, t \in I. \tag{4.2.12}$$

The following theorem says the isotonic regression g^* of g minimizes another objective function $\sum_x \Delta[g(x), f(x)]w(x)$ which is equivalent to maximizing $\sum_x(\Phi[f(x)] + [g(x) - f(x)]\phi[f(x)])w(x)$.

Theorem 4.2.15. *If f is isotonic on X and if the range of f is in I then*

$$\sum_x \Delta[g(x), f(x)]w(x) \ge \sum_x \Delta[g(x), g^*(x)]w(x) + \sum_x \Delta[g^*(x), f(x)]w(x). \tag{4.2.13}$$

Consequently g^* minimizes

$$\sum_x \Delta[g(x), f(x)]w(x) \tag{4.2.14}$$

in the class of isotonic f with range in I, and maximizes

$$\sum_x (\Phi[f(x)] + [g(x) - f(x)]\phi[f(x)])w(x).$$

The minimizing (maximizing) function is unique if Φ is strictly convex.

Proof. Let $r = g(x)$, $t = f(x)$, $s = g^*(x)$ in (4.2.12), so that the last term becomes $[g(x) - g^*(x)]\{\phi[g^*(x)] - \phi[f(x)]\}$. Since ϕ is nondecreasing, $\phi(f(x))$ is isotonic when f is, and then

$$\sum_x [g(x) - g^*(x)]\phi[f(x)]w(x) \leq 0$$

by Theorem 4.2.5. Also

$$\sum_x [g(x) - g^*(x)]\phi[g^*(x)]w(x) = 0$$

by Theorem 4.2.10. The first conclusion above follows easily from (4.2.12). The second conclusion follows since $\Delta \geq 0$. The last conclusion follows by noting that the first term in $\Delta(g(x), f(x))$ is $\Phi(g(x))$ which doesn't depend on f, and, therefore, minimizing $\Delta[g(x), f(x)]w(x)$ is equivalent to maximizing the last quantity. The uniqueness follows from the fact that the second part of (4.2.13) is strictly positive unless $f = g^*$. □

Now we are ready to apply the above Theorem to our dose-response curve estimation problem where one can show the MLEs are equivalent to the solutions of some isotonic regression problem.

Theorem 4.2.16. *Let*

$$\widetilde{p}_i = \widetilde{F}(x_i) = \max_{0 \leq u < i} \min_{i \leq v < m} \frac{\sum_{j=u}^{v} r_j}{\sum_{j=u}^{v} n_j} \qquad (1 \leq i \leq m). \tag{4.2.15}$$

Then the $\widetilde{p}_i$'s are the nonparametric MLEs of the likelihood function

$$L(F(x_1), \ldots, F(x_m)) = \prod_{i=1}^{m} F(x_i)^{r_i}(1 - F(x_i))^{n_i - r_i} \tag{4.2.16}$$

under the monotonicity constraint $F(x_1) \leq F(x_2) \cdots \leq F(x_m)$.

Proof. Here we use slightly different notation for our dose-response problem in order to be consistent with the notation used for isotonic regression. Let $X = \{x_1, \cdots, x_m\}$. Suppose $n(x)$ subjects are exposed to each dosage $x \in X$, where $r(x)$ of these respond and $n(x) - r(x)$ of them don't. Let $\bar{y}(x) = \dfrac{r(x)}{n(x)}$. Then the maximum estimator of the dose-response function $F(x)$ is given by maximizing the likelihood function

$$\prod_{x \in X} [F(x)]^{r(x)} [1 - F(x)]^{n(x) - r(x)} \tag{4.2.17}$$

and the negative log-likelihood can be written as

$$-\sum_{x} \{\bar{y}(x) \log F(x) + [1 - \bar{y}(x)] \log[1 - F(x)]\}\, n(x). \tag{4.2.18}$$

Now let $\Phi(u) = u \log u + (1 - u) \log(1 - u),\ 0 < u < 1$. Then one has

$$\Delta(g, F) = g \log g + (1 - g) \log(1 - g) - [g \log F + (1 - g) \log(1 - F)].$$

Note that maximizing the log-likelihood is equivalent to minimizing

$$\begin{aligned} \Delta(y(x), F(x)]n(x) = \sum_{x} \{&\bar{y}(x) \log \bar{y}(x) + (1 - \bar{y}(x)) \log(1 - \bar{y}(x)) \\ &- \bar{y}(x) \log F(x) - [1 - \bar{(}x)] \log[1 - F(x)]\} n(x), \end{aligned} \tag{4.2.19}$$

since the first component $\sum_x \{\bar{y}(x) \log \bar{y}(x) + (1 - \bar{y}(x)) \log(1 - \bar{y}(x))\}$ does not depend on F.

Then by Theorem 4.2.15, the isotonic regression $F^*(x)$ of $\bar{y}(x)$ with weights $n(x)$ also turns out to be the MLE of $F(x)$. Since we have shown that the solutions given by the max min formula above are the isotonic regression of $\bar{y}(x)$, those max min solutions are also the MLEs over the constraint that F is monotone (isotonic). We will make use of the result in the following section. □

4.3 The B-K Estimate and its Asymptotic Behavior

As in the introduction for the data, suppose that n_i subjects are recorded at a dosage x_i $(i = 1, \ldots, m)$, where $x_1 < \ldots < x_m$. One may assume, without loss of generality,

that $0 = x_1 < \ldots < x_m = 1$. The number of responses observed at dosage x_i is r_i $(i = 1, \ldots, m)$. The likelihood function for estimation of $p_i = F(x_i)$, $1 \leq i \leq m$, is

$$L(p_1, \ldots, p_m) = \prod_{i=1}^{m} p_i^{r_i}(1 - p_i)^{n_i - r_i} \quad (0 \leq p_1 \leq p_2 \leq \ldots \leq p_m \leq 1). \tag{4.3.1}$$

As it is shown in the previous section, the MLE of $(p_1, \ldots, p_m)$ under the monotonicity constraint is given by Ayer et al.(1955) as the following *PAV*, or *pool-adjacent-violators algorithm* (also see Barlow et al.(1972), p.73, and Cran (1980)):

$$\widetilde{p}_i = \max_{0 \leq u \leq i} \min_{i \leq v \leq m} \frac{\sum_{j=u}^{v} r_j}{\sum_{j=u}^{v} n_j} \qquad (1 \leq i \leq m). \tag{4.3.2}$$

The $\widetilde{p}_i$ may also be obtained by *isotonic regression*, i.e., as the solution of the weighted least squares problem:

$$(\widetilde{p}_1, \ldots, \widetilde{p}_m) = \arg\min \sum (r_i - n_i p_i)^2 n_i \tag{4.3.3}$$

where the minimization is over the set $0 \leq p_1 \leq p_2 \ldots \leq p_m \leq 1$; (see Barlow et al. (1972), section 1.4).

In order to estimate the entire dose-response curve, Bhattacharya and Kong (2007) proposed an estimate $\widetilde{F}(x)$ of $F(x)$, by taking $\widetilde{F}(x)$ to be $\widetilde{p}_i$ at x_i and then linear interpolating in the interval (x_i, x_{i+1})

$$\widetilde{F}(x) = \begin{cases} \widetilde{p}_i & \text{if } x = x_i \\ \widetilde{p}_i + \dfrac{\widetilde{p}_{i+1} - \widetilde{p}_i}{x_{i+1} - x_i}(x - x_i) & \text{if } x_i < x \leq x_{i+1} \end{cases}$$

It is a continuous function whose inverse is the estimate of ED_p as given by:

$$\widetilde{ED}_p = \begin{cases} x_1 & \text{if } p \leq \widetilde{p}_1 \\ x_i + \dfrac{p - \widetilde{p}_i}{\widetilde{p}_{i+1} - \widetilde{p}_i}(x_{i+1} - x_i) & \text{if } \widetilde{p}_i < p \leq \widetilde{p}_{i+1} \text{ for some i} \\ x_m & \text{if } p > \widetilde{p}_m \end{cases} \tag{4.3.4}$$

Precise conditions for consistency of such an estimator and for proper confidence intervals for ED_p were derived in Bhattacharya and Kong (2007), as described by the following theorems.

Theorem 4.3.1. *For a given $p \in (0,1)$, assume that the p-th quantile ED_p is the unique solution of $F(x) = p$. Also assume that there exists a γ such that*

$$\theta_\gamma := \min_{[ED_p-\gamma, ED_p+\gamma]} F'(x) > 0. \tag{4.3.5}$$

Then

$$\widetilde{\widehat{ED}}_p \to ED_p \quad a.s$$

as $n_l \to \infty$, $m \to \infty$, and $m \le M\left(\frac{n_l^{1/2}}{(\log n_l)^{1/2}}\right)$, where $n_l = \min\{n_i : 1 \le i \le m\}$ and $M = \frac{(b-a)\theta_\gamma}{4((7+\epsilon_0)/6)^{1/2}}$ for some $\epsilon_0 > 0$. $\square$

Proof. The main ingredient of the proof is that under certain conditions on the orders between the number m of dosages and sample size n at each level, the sample proportion $\widehat{p}_i$ are also monotone.

Assume $\zeta_p \in (a,b)$ and $[a,b] \in [\zeta_p - \gamma,\ \zeta_p + \gamma]$. Without any essential loss of generality, assume

$$x_0 = a, x_m = b, d(m) = \frac{b-a}{m}$$

and

$$x_i = x_0 + i \cdot d(m),\ i = 0, \cdots, k.$$

There exists some interval $[x_i, x_{i+1})$ such that $\zeta_p \in [x_i, x_{i+1})$. Let $i_0 = i$ or $i+1$ such that $|\zeta_p - x_{i_0}| \le d(m)/2$, i.e., x_{i_0} is the dose which is closest to ζ_p. By the assumption of the uniqueness of ζ_p, the following holds:

$$p \in (p_{i_0-1}, p_{i_0+1}),\ \ \min\left\{|\zeta_p - x_{i_0-1}|,\ |\zeta_p - x_{i_0+1}|\right\} \ge \frac{d(m)}{2}. \tag{4.3.6}$$

Let $\delta = \min\left\{(p - P_{i_0-1})/2, (P_{i_0+1} - p)/2\right\}$. For large m, say $m > 2(b-a)\gamma$, one has $[x_{i_0-1}, x_{i_0+1}] \subset [\zeta_p - \gamma, \zeta_p + \gamma]$. By assumption (4.3.5), using (4.3.6),

$$\delta \ge \frac{d(m)\theta_\gamma}{4}.$$

Now applying Bernstein's Lemma (Serfling, 1980, p.95), one has

$$P\left(\left|\frac{r_i}{n_i}-p_i\right|\right) \le 2\exp\left(-\frac{n_i^2t^2}{2n_ip_i(1-p_i)+2/3cn_it}\right) \le 2\exp\left(-\frac{n_i^2t^2}{1/2n_i+2/3n_i}\right) = 2\exp\left(-\frac{6n_it^2}{7}\right) \tag{4.3.7}$$

for any dose x_i and for any n_i. Here $c = \max\{1-p_i, p_i\} \le 1$. If we take $t(n_l) = ((7+\epsilon_0)/6 \cdot (\log n_l)/n_l)^{1/2}$, we will have

$$P\left(\left|\frac{r_i}{n_i}-p_i\right| \ge t(n_l)\right) \le \exp\left(-\frac{6}{7}n_i \cdot \frac{7+\epsilon_0}{6} \cdot \frac{\log n_l}{n_l}\right) \le 2\exp\left(-\frac{7+\epsilon_0}{7} \cdot \log n_l\right) = 2n_l^{-(1+\epsilon_0/7)}.$$

Let $d(m)\theta_\gamma \ge t(n_l)$, that is,

$$m \le \frac{(b-a)\theta_\gamma}{4\left((7+\epsilon_0)/6 \cdot (\log n_l)/n_l\right)^{1/2}} = \frac{(b-a)\theta_\gamma n_l^{1/2}}{4\left((7+\epsilon_0)/6 \cdot (\log n_l)\right)^{1/2}}. \tag{4.3.8}$$

Then one has

$$P\left(\left|\frac{r_i}{n_i}-p_i\right| \ge \delta(m,n)\right) \le P\left(\left|\frac{r_i}{n_i}-p_i\right| \ge \frac{d(m)\theta_\gamma}{4}\right) \le P\left(\left|\frac{r_i}{n_i}-p_i\right| \ge t(n_l)\right) \le 2n_l^{-(1+\epsilon_0)/7}. \tag{4.3.9}$$

Then by the Borel-Cantelli lemma,

$$P\left(\left|\frac{r_i}{n_i}-p_i\right| \ge \frac{d(m)\theta_\gamma}{4} i.o.\right) = 0,$$

which says

$$P\left(\left|\frac{r_i}{n_i}-p_i\right| < \frac{d(m)\theta_\gamma}{4} \text{ for all } n_l \text{ but finitely many } n_l\right) = 1.$$

Let $A = \left\{\omega : \left|\frac{r_i}{n_i}-p_i\right| < \frac{d(m)\theta_\gamma}{4} \text{ for all } n_l \text{ but finitely many } n_l\right\}$, then for any $\omega \in A$, there exists $N_1(\omega)$ such that for $n_l \ge N_1(\omega)$,

$$\frac{r_i}{n_i} < p_i + \frac{d(m)\theta_\gamma}{4} \le p_i + \frac{p_{i+1}-p_i}{4} = p_{i+1} - \frac{3(p_{i+1}-p_i)}{4} < p_{i+1} - \frac{p_{i+1}-p_i}{4} \le p_{i+1} - \frac{d(m)\theta_\gamma}{4} < \frac{r_{i+1}}{n_{i+1}}.$$

Then one has

$$P\left(\widetilde{p}_i = \frac{r_i}{n_i} \text{ for all} l; i\right) \to 1 \text{ as } n_l \to \infty. \tag{4.3.10}$$

Also by the Borel-Cantelli Lemma, one has

$$P\left(\left|\frac{r_{i_0-1}}{n_{i_0-1}} - p_{i_0-1}\right| < \delta(m,n) \text{ for all but finitely many } n_l\right) = 1.$$

Let

$$A_{-1} = \left\{\omega : \left|\frac{r_i}{n_i} - p_{i_0-1}\right| < \delta(m,n) \text{ for all but finitely many } n_l\right\},$$

and

$$A_1 = \left\{\omega : \left|\frac{r_{i_0+1}}{n_{i_0+1}} - p_{i_0+1}\right| < \delta(m,n) \text{ for all but finitely many } n_l\right\}.$$

Then one has $P(A_{-1} \cap A_1 \cap A) = 1$. For $\omega \in A_{-1} \cap A_1 \cap A$, one has $\widetilde{p}_i = \dfrac{r_i}{n_i}$,when $n_l \geq N(\omega)$,

$$\begin{aligned}\widetilde{p}_{i_0-1} = \frac{r_{i_0-1}}{n_{i_0-1}} &< p_{i_0-1} + \delta(m,n) \leq p_{i_0-1} + \frac{p - p_{i_0-1}}{2} = \frac{p + p_{i_0-1}}{2} < p \\ &< \frac{p}{2} + \frac{p_{i_0+1}}{2} = p_{i_0+1} - \frac{p_{i_0+1} - p}{2} \leq p_{i_0+1} - \delta(m,n) < \frac{r_{i_0+1}}{n_{i_0+1}} = \widetilde{p}_{i_0+1},\end{aligned}$$

which implies

$$\widetilde{p}_{i_0-1} < p < p_{i_0+1}.$$

Therefore,

$$|\widetilde{\zeta_p} - \zeta_p| \leq 2d(m) \to 0 \; a.s. \text{ as } n_l \to \infty, m \to \infty \text{ and } m \leq M\frac{n_l^{1/2}}{(\log n_l)^{1/2}},$$

where $M = (b-a)\theta_\gamma/4((7+\epsilon_0)/6)^{1/2}$ for some $\epsilon_0 > 0$ which now completes the proof. □

The following is the main result in Bhattacharya and Kong (2007).

Theorem 4.3.2. *Assume that for a given $p \in (0,1)$, the dose-response function $F(x)$ is twice differentiable in a neighborhood of ED_p, the first derivative $f(x)$ of $F(x)$ has a lower bound and the second derivative of $F(x)$ is bounded in the neighborhood of*

ED_p. That is, there exists $\gamma > 0$ such that (4.3.5) *holds in Theorem 4.3.1, and there exists some number M_1 such that*

$$|F''(x)| = |f'(x)| \leq M_1$$

for $x \in [ED_p - \gamma, ED_p + \gamma]$. Also assume that there exists some constant $c \geq 1$ such that

$$\frac{n_u}{n_l} \leq c$$

for all the experiments, where $n_u = \max\{n_i, 1 \leq i \leq m\}$, $n_l = \min\{n_i : 1 \leq i \leq m\}$. Then

$$\sqrt{n_{i_0}}(\widetilde{ED}_p - ED_p) \to N\left(0, \frac{p(1-p)}{(f(ED_p))^2}\right)$$

as $n_l \to \infty$ and $n_l^{1/4}(\log\log n_l)^{1/2} \leq m \leq M(\frac{n_l^{1/2}}{(\log n_l)^{1/2}})$, where $M = \frac{(b-a)\theta_\gamma}{4((7+\epsilon_0)/6)^{1/2}}$ as in Theorem 3.1 *for some $\epsilon_0 > 0$, and n_{i_0} is the number of subjects assigned to the dosage level x_{i_0} which is closest to ED_p.* □

Proof. A more general result is proved in the next section (Theorem 4.5.4). We remark here that a minor correction seems to be needed in the variance expression of the asymptotic Normal distribution. □

4.4 The NAM Method

4.4.1 The NAM Estimates

In this subsection, we shall describe the *New Adaptive Method*, termed *NAM*, and derive *NAM* estimates in detail. We shall keep the same assumptions as above for the data, i.e., n_i subjects are given at a dosage x_i $(i = 1, \ldots, m)$, where $0 = x_1 < \ldots < x_m = 1$ is assumed without loss of generality. The number of responses observed at dosage x_i is r_i $(i = 1, \ldots, m)$. Recall that the B-K estimates of $F(x)$ are taken to be $\widetilde{p}_i$, which are the isotonic MLEs at dosages x_i, and the estimates between x_i and x_{i+1}

are obtained by linear interpolation in the interval (x_i, x_{i+1}). Also the B-K estimates of $\zeta_p = ED_p$ are taken to be the inverse of $\widetilde{F}(x)$ which are the B-K estimates of $F(x)$.

From now on, we will assume, for simplicity, that there are m equidistant dosages and the same number n of i.i.d. binary observations at each dosage. Assume $n \to \infty, m \to \infty$ and

$$m = rs(n), \text{ with } r \geq 1, \, s(n) \text{ integers}, \tag{4.4.1}$$

$r \asymp (m^4/n)^{1/5}$ in Theorems 4.5.2, 4.5.3, 4.5.4(b) and $r \asymp (m^4/n)^{1/5}/(\log\log n)^{6/5}$ in Theorem 4.5.4, part (c). Here $f(m,n) \asymp g(m,n)$ means that the ratio of the two sides are bounded away from zero and infinity.

Let $\widehat{p}_i$ denote the observed proportion of 1's at dosage x_i. Divide the observed proportions and dosages into r groups, and consider the following application of the PAV algorithm to each of the r groups of levels below:

$$\begin{aligned}
&[\text{Group 1}] : \widehat{p}_1, \ \widehat{p}_{r+1}, \ \widehat{p}_{2r+1}, \ \ldots, \widehat{p}_{(s(n)-1)r+1}, \ \widehat{p}_{s(n)r}, \ \widehat{p}_m; \\
&[\text{Group 2}] : \widehat{p}_1, \ \widehat{p}_2, \ \widehat{p}_{r+2}, \ \widehat{p}_{2r+2}, \ \ldots, \widehat{p}_{(s(n)-1)r+2}, \ \widehat{p}_m; \\
&[\text{Group 3}] : \widehat{p}_1, \ \widehat{p}_3, \ \widehat{p}_{r+3}, \ \widehat{p}_{2r+3}, \ \ldots, \widehat{p}_{(s(n)-1)r+3}, \ \widehat{p}_m; \\
&\qquad\qquad \vdots \\
&[\text{Group r}] : \widehat{p}_1, \ \widehat{p}_r, \ \widehat{p}_{r+r}, \ \widehat{p}_{2r+r}, \ \ldots, \widehat{p}_{(s(n)-2)r+r}, \ \widehat{p}_m.
\end{aligned} \tag{4.4.2}$$

Note that Group 2 through Group $r-1$ each has $s(n)+2$ levels, while Groups 1 and r each has $s(n)+1$ levels. Also, except for the smallest and the largest levels (with proportions $\widehat{p}_1$ and $\widehat{p}_m$), the sets of levels covered by them are disjoint. Together, they comprise all the different $m = rs(n)$ dosages.

By linear interpolation, each Group j ($j = 1, \ldots, r$) provides an estimate $\widetilde{F}_j$ of the dose-response curve F on $[0,1]$, and an estimate $\widetilde{\zeta}_{p,j}$ of F^{-1}. Note that while F^{-1} is defined on $[F(0), F(1)]$, $\widetilde{\zeta}_{p,j}$ and $\widetilde{\zeta}_p$ below are defined on $[\widetilde{F}(0), \widetilde{F}(1)]$. Compute

$$\widetilde{F} := (1/r) \sum_{1 \le j \le r} \widetilde{F}_j, \quad \widetilde{\zeta}_p := (1/r) \sum_{1 \le j \le r} \widetilde{\zeta_{p,j}}, \tag{4.4.3}$$

and choose the values of r for which the estimated MISEs of $\widetilde{F}$ and $\widetilde{\zeta}$ are the smallest. These are the *NAM estimates* of F and F^{-1}.

Remark 4.4.1. For purposes of asymptotics, one may take the r groups in (4.4.2) to be disjoint, omitting $\widehat{p}_m$ from Group 1, $\widehat{p}_1$ and $\widehat{p}_m$ from Groups 2 through $r-1$, and $\widehat{p}_1$ from Group r. As is shown in Bhattacharya and Kong (2007), outside a set B_n of negligible probability, $\widehat{p}_i < \widehat{p}_{i+1}$ $\forall i$. Given $x \in (0,1)$, if m, n are sufficiently large, and $m/r = o(\sqrt{n/\log n})$ (see (4.5.6)), x belongs to the domain of $\widetilde{F}_j$ $\forall j$, even if the curve $\widetilde{F}_j$ is constructed with common points removed. Outside B_n, the curves so obtained would coincide, on their respective domains, with the curves constructed after adjoining the end points. On the other hand, for relatively small sample sizes one needs to construct $\widetilde{F}_j$ with the groupings (4.4.2), so that each has domain $[0,1]$.

Remark 4.4.2. The method of forming disjoint groups of dosages, and averaging over estimates from them as described above, may formally be applied to other procedures as well. However, one can show that asymptotic rates do not improve if one applies this to kernel-based methods. One may check this phenomenon without difficulty for general kernel-based curve estimation. This is also confirmed by finite sample simulations, which we omit.

Remark 4.4.3. Note that in (4.4.3), $\widetilde{\zeta}$ is not the same as $\widetilde{F}^{-1}$. An alternative estimate of F^{-1} is given by $\widetilde{F}^{-1}$. Simulations show that the performances of $\widetilde{\zeta}$ and $\widetilde{F}^{-1}$ are not significantly different. The derivation of the asymptotics for $\widetilde{F}^{-1}$, however, appears to be more delicate than that for $\widetilde{\zeta}$.

4.5 Asymptotics of the NAM Estimates.

The asymptotic theory supporting the NAM is derived in this subsection. Theorem 4.5.2 proves that the estimate of the dose-response curve has a MISE attaining the optimal rate $O\left(N^{-4/5}\right)$ under the assumptions that $f = F'$ is strictly positive, F'' is bounded and $m = o\left(n^{3/2}/(\log n)^{5/2}\right)$. Theorem 4.5.3 provides the same optimal MISE rate for the estimate $p \to \widetilde{\zeta}_p$ of the quantile curve of interest, under the additional restriction $m/n^{2/3} \not\to \infty$. Theorem 4.5.4 shows that $\widetilde{\zeta}_p$ is asymptotically Normal around $E\widetilde{\zeta}_p$ with an asymptotic variance $O\left(N^{-4/5}\sqrt{\log\log N}\right)$, under the same broad assumptions as in Theorem 4.5.2. However, for asymptotic Normality of $\widetilde{\zeta}_p$ around ζ_p, one requires the restriction $m = o(n^{2/3})$.

Let $\widehat{p}_i = r_i/n_i$ denote the sample proportion of responses at dosage x_i $(i = 1, \ldots, m)$. For simplicity, we assume in the proof that $n_i = n$ for all i and that $x_{i+1} - x_i = 1/m$ for $i = 1, \ldots, m-1$. Let $N = mn$ denote the total number of observations.

Before proving Theorem 4.5.2, one first makes the following remark.

Remark 4.5.1. In nonparametric regression, under the assumption that $F(x)$ has bounded second derivative, the min max risk of mean square error is of order $N^{-4/5}$ where N is the total sample size. See Stone (1980) for a proof, which we will omit here.

We use the following assumption throughout unless stated otherwise.

(A) The ratio of $min\{n_i : 1 \le i \le m\}$ to $max\{n_i : 1 \le i \le m\}$ is bounded away from zero, and the ratio of $min\{x_{i+1} - x_i : 1 \le i \le m-1\}$ to $max\{x_{i+1} - x_i : 1 \le i \le m-1\}$ is bounded away from zero.

However, for simplicity the proofs are given for the case $n_i = n \; \forall \; i$ and $x_i - x_{i-1} = \frac{1}{m}$ $(x_0 = 0)$.

Theorem 4.5.2. *Assume that the dose-response function F on $[0,1]$ is twice differentiable, $f = F'$ has a positive lower bound θ and that F'' is bounded. Also assume* ***(A)****.*

(a) The mean integrated squared error (MISE) of $\widetilde{F}$ has the asymptotically optimal rate $O\left(N^{-4/5}\right)$ as $N \to \infty$, if $r = O(1)$, $m \asymp n^{1/4}$.

(b) If $m/n^{1/4} \to \infty$, $m = o(n^{3/2}/(\log n)^{5/2})$, then the MISE of $\widetilde{F}$ is $O\big(N^{-4/5}\big)$, with a choice of r satisfying $r \asymp (m^4/n)^{1/5}$.

(c) If $r = O(1)$, $m/n^{1/4} \to \infty$, then

$$\frac{\sqrt{n}(\widetilde{F}(x) - F(x))}{\Delta(x)} \xrightarrow{L} N\left(0, F(x)(1-F(x))\right), \tag{4.5.1}$$

where

$$\Delta^2(x) = \sum_{i=1}^{m} (x_{i+1} - x_i)^{-2} \left\{(x_{i+1} - x)^2 + (x - x_i)^2\right\} 1_{I_i}(x) \tag{4.5.2}$$

$$(I_i = [x_i, x_{i+1}) \text{ for } 1 \le i \le m-2, I_{m-1} = [x_{m-1}, x_m]).$$

(d) (i) If $m = o\left(n^{3/2}/(\log n)^{5/2}\right)$, then with $r \asymp (m^4/n)^{1/5}$,

$$\frac{\sqrt{rn}(\widetilde{F}(x) - E\widetilde{F}(x))}{\bar{\Delta}(x)} \xrightarrow{L} N\left(0, F(x)(1-F(x))\right), \tag{4.5.3}$$

where $\bar{\Delta}(x)$ is the average of $\Delta_i(x)$ $(i = 1, \cdots, r)$ of the form $\Delta(x)$ given in (4.5.2)*, corresponding to the r groups each with m/r dosages.*

(ii) If $m = o\left(n^{3/2}/\left(\log^{5/2} n(\log\log n)^2\right)\right)$, and with a slightly smaller number of groups $r \asymp (m^4/n)^{1/5}/(\log\log n)^{2/5}$, then

$$\frac{\sqrt{rn}(\widetilde{F}(x) - F(x))}{\bar{D}(x)} \xrightarrow{L} N\left(0, F(x)(1-F(x))\right). \tag{4.5.4}$$

Proof. (a) It follows from Bernstein's inequality, as in the proof of Theorem 1 in Bhattacharya and Kong (2007), that there exist appropriate positive constants c, c' such that for $n > 1$,

$$P\left(\mid \widehat{p}_i - p_i \mid > c\sqrt{\log n/n} \text{ for some i, } i = 1, \ldots, m\right) \le c' N^{-2}. \tag{4.5.5}$$

It follows that if

$$m < (\theta/2c)\sqrt{n/\log n}, \tag{4.5.6}$$

then

$$P\left(\widehat{p}_i \neq \widetilde{p}_i \text{ for some i, } i = 1, \ldots, m\right) \leq c'' N^{-2} \tag{4.5.7}$$

for some $c'' > 0$. Let B_n denote the union of the two sets within parentheses in (4.5.5) and (4.5.7). It is shown in Bhattacharya and Kong (2007), and simple to check using (4.5.5) to (4.5.7), that, on B_n^c, $\widehat{p}_i < \widehat{p}_{i+1}$ for all i.

Let $x \in [x_i, x_{i+1}]$. By linearity of $\widetilde{F}$ on $[x_i, x_{i+1}]$,

$$\begin{aligned}
\widetilde{F}(x) &= \left(\frac{x_{i+1} - x}{x_{i+1} - x_i}\widehat{p}_i + \frac{x - x_i}{x_{i+1} - x_i}\widehat{p}_{i+1}\right) 1_{B_n^c} + \widetilde{F}(x) 1_{B_n} \\
&= \frac{x_{i+1} - x}{x_{i+1} - x_i}\widehat{p}_i + \frac{x - x_i}{x_{i+1} - x_i}\widehat{p}_{i+1} + \varepsilon_{n,1} \\
&= \widehat{p}_i + \frac{x - x_i}{x_{i+1} - x_i}(\widehat{p}_{i+1} - \widehat{p}_i) + \varepsilon_{n,1} \ \left(\mid \varepsilon_{n,1} \mid \leq 21_{B_n} = O_p(N^{-2})\right).
\end{aligned} \tag{4.5.8}$$

Also, for some $x^* \in [x_i, x_{i+1}]$,

$$\begin{aligned}
F(x) &= F(x_i) + (x - x_i)F'(x^*) \\
&= p_i + (x - x_i)\Big[\frac{F(x_{i+1}) - F(x_i)}{x_{i+1} - x_i} + \varepsilon(x)\Big], \mid \varepsilon(x) \mid \leq M/m,
\end{aligned} \tag{4.5.9}$$

where $\varepsilon(x) = F'(x^*) - F'(x^{**})$ for some x^*, x^{**} lying in $[x_i, x_{i+1}]$, and $M = sup\{|F''(x)| : 0 \leq x \leq 1\}$. Thus, noting that F, $\widetilde{F}$ are bounded by one,

$$\begin{aligned}
E\widetilde{F}(x) &= \frac{x_{i+1} - x}{x_{i+1} - x_i}p_i + \frac{x - x_i}{x_{i+1} - x_i}p_{i+1} + O(N^{-2}) \\
&= p_i + \frac{x - x_i}{x_{i+1} - x_i}(p_{i+1} - p_i) + O(N^{-2}),
\end{aligned} \tag{4.5.10}$$

and

$$E\widetilde{F}(x) - F(x) = -(x - x_i)\varepsilon(x) + O(N^{-2}) = O\left(1/m^2\right). \tag{4.5.11}$$

From (4.5.8) and (4.5.9),

$$\begin{aligned}\widetilde{F}(x) - F(x) &= \widehat{p}_i - p_i + [\widehat{p}_{i+1} - \widehat{p}_i - (p_{i+1} - p_i)] \frac{x - x_i}{x_{i+1} - x_i} - \varepsilon(x)(x - x_i) + \varepsilon_{n,1} \\ &= \frac{x_{i+1} - x}{x_{i+1} - x_i} (\widehat{p}_i - p_i) + \frac{x - x_i}{x_{i+1} - x_i} (\widehat{p}_{i+1} - p_{i+1}) - \varepsilon(x)(x - x_i) + \varepsilon_{n,1},\end{aligned} \tag{4.5.12}$$

and, by subtracting (4.5.11) from (4.5.12) one achieves

$$\begin{aligned}\widetilde{F}(x) - E\widetilde{F}(x) &= \widehat{p}_i - p_i + [\widehat{p}_{i+1} - \widehat{p}_i - (p_{i+1} - p_i)] \frac{x - x_i}{x_{i+1} - x_i} + \varepsilon_{n,1} \\ &= \frac{x_{i+1} - x}{x_{i+1} - x_i}(\widehat{p}_i - p_i) + \frac{x - x_i}{x_{i+1} - x_i}(\widehat{p}_{i+1} - p_{i+1}) + \varepsilon_{n,1}.\end{aligned} \tag{4.5.13}$$

Hence

$$\begin{aligned}Var\left(\widetilde{F}(x)\right) &= \left(\frac{x_{i+1} - x}{x_{i+1} - x_i}\right)^2 \frac{p_i(1 - p_i)}{n} + \left(\frac{x - x_i}{x_{i+1} - x_i}\right)^2 \frac{p_{i+1}(1 - p_{i+1})}{n} + O(N^{-2}) \\ &= O\left(1/n\right).\end{aligned} \tag{4.5.14}$$

From (4.5.11) and (4.5.14) one obtains, on integration,

$$MISE(\widetilde{F}) = O\left(1/n\right) + O\left(1/m^4\right). \tag{4.5.15}$$

If $m \asymp n^{1/4}$, then the MISE attains its optimal rate (noting that $mn = N$, or $n^{5/4} = O(N)$),

$$MISE(\widetilde{F}) = O\left(1/n\right) = O\left(N^{-4/5}\right). \tag{4.5.16}$$

(b) First observe that the r groups in (4.4.2) are essentially disjoint. Inclusion of $(x_1, \widehat{p}_1)$ and $(x_m, \widehat{p}_m)$ in each group ensures that $\widetilde{F}_j$ $(j = 1, \ldots, r)$ is defined on all of $[0, 1]$. Note the strict inequality $\widehat{p}_j < \widehat{p}_{j+1}$, $\forall$ j on B_n^c, since the assumption $m = o(n^{3/2}/(\log n)^{5/2})$ implies that (4.5.6) holds with m/r in place of m.

If one has $m/n^{1/4} \to \infty$, then using r essentially disjoint groups, and averaging, one has (see (4.5.11), (4.5.14))

$$MISE(\widetilde{F}) = MISE\Big(1/r \sum_{1 \le j \le r} \widetilde{F}_j\Big) = O\left(1/rn\right) + O\left(1/(m/r)^4\right). \tag{4.5.17}$$

The optimal choice of r is given by the relation $(m/r)^4 \asymp rn$ or, $r^5 \asymp m^4/n$, or, $r \asymp (m^4/n)^{1/5}$, yielding the optimal rate:

$$MISE(\widetilde{F}) = O\left((rn)^{-1}\right) = O\left(N^{-4/5}\right). \tag{4.5.18}$$

(c) By (4.5.11), it follows that $\sqrt{n}(E\widetilde{F}(x) - F(x)) \to 0$ since $\sqrt{n}/m^2 \to 0$ as $m/n^{1/4} \to \infty$. Note that $Var(\widetilde{F}(x))$ is given by (4.5.14). Then by the CLT and Slutsky's Theorem, the result follows.

(d) (i) As in (b), since the r groups are essentially disjoint, if $m/r = o\left(\sqrt{n/\log n}\right)$, then the same calculations for $\sqrt{n}\left(\widetilde{F_j}(x) - E\widetilde{F_j}(x)\right)$ $(1 \le j \le r)$ of each group go through with m replaced by m/r. In order to derive the asymptotic distribution, one applies Lyapunov's CLT to the r summands $\sqrt{n}(\widetilde{F}(x) - E\widetilde{F}(x))$, $1 \le j \le r$. The mean of the summands is zero and the variance can be calculated using the variance formula in (4.5.13) which is bounded away from zero and infinity. The third moment is also bounded. Therefore, the Lyapunov conditions apply. When $r \asymp (m^4/n)^{1/5}$ and $m = o(n^{3/2}/(\log n)^{5/2})$, $m/r = o\left(\sqrt{n/\log n}\right)$ holds.

(ii) Note that

$$Bias(\widetilde{F}) = O(r^2/m^2).$$

Then, with $r \asymp (m^4/n)^{1/5}/(\log\log n)^{2/5}$,

$$\sqrt{nr}Bias(\widetilde{F}) = O\left(r^{5/2}\sqrt{n}/m^2\right) = O((m^2/\sqrt{n}/\log\log n)\sqrt{n}/m^2) \to 0.$$

With $m = o\left(n^{3/2}/(\log n)^{5/2}(\log\log n)^2\right)$, $m/r = o\left(\sqrt{n/\log n}\right)$ holds. By similar arguments in part(i), the results follow. □

We now turn to the estimation of the curve F^{-1}.

Theorem 4.5.3. *Assume the hypothesis of Theorem 4.5.2 and* ***(A)****.*

(a) If $m \asymp O\left(n^{1/4}\right)$ *then, with* $r = 1$, $\widetilde{\zeta} = \widetilde{F}^{-1}$, *one has* $MISE(\widetilde{\zeta}) = O\left(N^{-4/5}\right)$.

(b) If $m/n^{1/4} \to \infty$, *but* $m/n^{2/3} \not\to \infty$, *then* $MISE\left(\widetilde{\zeta}\right) = O\left(N^{-4/5}\right)$, *with* $r \asymp (m^4/n)^{1/5}$.

Proof. (a) For $m = O\left(n^{1/4}\right)$, one may consider $r = 1$ in (4.4.3). Then $\widetilde{\zeta} = \widetilde{F}^{-1}$. Let $p \in [p_i, p_{i+1}]$, so that $x = F^{-1}(p) \in [x_i, x_{i+1}]$. Then, on B_n^c,

$$F^{-1}(p) - \widetilde{F}^{-1}(p) = \widetilde{F}^{-1}(\widetilde{F}(x)) - \widetilde{F}^{-1}(F(x)). \tag{4.5.19}$$

First, consider, for an appropriate positive constant c_1,

$$p \in \left[p_i + c_1\sqrt{\log(n)/n},\ p_{i+1} - c_1\sqrt{\log(n)/n}\right] = D_{n,i}\ , \tag{4.5.20}$$

say. Then on B_n^c, $F(x)$ and $\widetilde{F}(x)$ belong to $[\widehat{p}_i, \widehat{p}_{i+1}]$. Using (4.5.19), the linearity of $\widetilde{F}^{-1}$ on $[\widehat{p}_i, \widehat{p}_{i+1}]$ and (4.5.12), and writing

$$\bigtriangledown_n = \widehat{p}_{i+1} - \widehat{p}_i - (p_{i+1} - p_i),\ \frac{1}{\widehat{p}_{i+1} - \widehat{p}_i} = \frac{1}{p_{i+1} - p_i}\left(1 - \frac{\bigtriangledown_n}{\widehat{p}_{i+1} - \widehat{p}_i}\right) \tag{4.5.21}$$

on B_n^c, we get the following relation, noting that $F^{-1}(p) - \widetilde{F}^{-1}(p)$ is bounded by 1:

$$\begin{aligned}
F^{-1}(p) - \widetilde{F}^{-1}(p) &= \left[\widetilde{F}(x) - F(x)\right]\frac{x_{i+1} - x_i}{\widehat{p}_{i+1} - \widehat{p}_i}1_{B_n^c} + \varepsilon_{n,2}\ \left(|\ \varepsilon_{n,2}\ | \leq 1_{B_n} = O_p(N^{-2})\right) \\
&= \left\{(\widehat{p}_i - p_i)\,\frac{x_{i+1} - x_i}{\widehat{p}_{i+1} - \widehat{p}_i} + \frac{x - x_i}{\widehat{p}_{i+1} - \widehat{p}_i}\,\bigtriangledown_n - \varepsilon(x)(x - x_i)\frac{x_{i+1} - x_i}{\widehat{p}_{i+1} - \widehat{p}_i}\right\}1_{B_n^c} \\
&\quad + \varepsilon_{n,2} \\
&= \left\{(\widehat{p}_i - p_i)\frac{x_{i+1} - x_i}{p_{i+1} - p_i}\left(1 - \frac{\bigtriangledown_n}{\widehat{p}_{i+1} - \widehat{p}_i}\right) + \frac{(x - x_i)\bigtriangledown_n}{p_{i+1} - p_i}\left(1 - \frac{\bigtriangledown_n}{\widehat{p}_{i+1} - \widehat{p}_i}\right)\right. \\
&\quad \left. - \frac{\varepsilon(x)(x - x_i)(x_{i+1} - x_i)}{p_{i+1} - p_i}\left(1 - \frac{\bigtriangledown_n}{\widehat{p}_{i+1} - \widehat{p}_i}\right)\right\}1_{B_n^c} + \varepsilon_{n,2}. \\
&= \left\{(\widehat{p}_i - p_i)\frac{x_{i+1} - x_i}{p_{i+1} - p_i} + \frac{(x - x_i)\bigtriangledown_n}{p_{i+1} - p_i} - \frac{\varepsilon(x)(x - x_i)(x_{i+1} - x_i)}{p_{i+1} - p_i}\right\}1_{B_n^c} \\
&\quad + \varepsilon_{n,3} + \varepsilon_{n,2}.
\end{aligned} \tag{4.5.22}$$

Here

$$\begin{aligned}
\varepsilon_{n,3} = -\Big\{&(\widehat{p}_i - p_i)\left(\frac{x_{i+1} - x_i}{p_{i+1} - p_i}\right)\frac{\bigtriangledown_n}{\widehat{p}_{i+1} - \widehat{p}_i} + \left(\frac{x - x_i}{p_{i+1} - p_i}\right)\frac{\bigtriangledown_n^2}{\widehat{p}_{i+1} - \widehat{p}_i} \\
&+ \frac{\varepsilon(x)(x - x_i)(x_{i+1} - x_i)}{p_{i+1} - p_i}\left(\frac{-\bigtriangledown_n}{\widehat{p}_{i+1} - \widehat{p}_i}\right)\Big\}1_{B_n^c}.
\end{aligned} \tag{4.5.23}$$

Note that, on B_n^c, $\mid \widehat{p}_i - p_i \mid < c\,(\log n/n)^{1/2} = \varepsilon_n$, say, $\forall i$, so that $\widehat{p}_{i+1} - \widehat{p}_i \geq p_{i+1} - p_i - 2\varepsilon_n > \theta/2m$ for all sufficiently large n.

The expectation of (4.5.22) equals

$$F^{-1}(p) - E\widetilde{F}^{-1}(p) = -\frac{\varepsilon(x)(x - x_i)(x_{i+1} - x_i)}{p_{i+1} - p_i} + E\varepsilon_{n,3} + O(N^{-2}). \tag{4.5.24}$$

Now $E\varepsilon_{n,3}$ is the sum of the following:

$$\begin{aligned}
&- E\left(\frac{(\widehat{p}_i - p_i)(x_{i+1} - x_i)\bigtriangledown_n}{(p_{i+1} - p_i)(\widehat{p}_{i+1} - \widehat{p}_i)} 1_{B_n^c}\right) \\
&= -E\left((\widehat{p}_i - p_i)\,\frac{(x_{i+1} - x_i)\bigtriangledown_n}{(p_{i+1} - p_i)^2}\left(1 - \frac{\bigtriangledown_n}{\widehat{p}_{i+1} - \widehat{p}_i}\right) 1_{B_n^c}\right) \\
&= \frac{p_i(1 - p_i)}{n}\frac{x_{i+1} - x_i}{(p_{i+1} - p_i)^2} + O\left(m^2/n^{3/2}\right) + O\left(N^{-2}\right) \\
&= O(m/n);
\end{aligned} \tag{4.5.25}$$

$$\begin{aligned}
&- E\left(\frac{(x - x_i)}{(p_{i+1} - p_i)}\frac{\bigtriangledown_n^2}{(\widehat{p}_{i+1} - \widehat{p}_i)} 1_{B_n^c}\right) = -\frac{(x - x_i)}{p_{i+1} - p_i} E\left(\frac{\bigtriangledown_n^2}{\widehat{p}_{i+1} - \widehat{p}_i} 1_{B_n^c}\right) \\
&= -\frac{(x - x_i)}{(p_{i+1} - p_i)^2} E\bigtriangledown_n^2 + \frac{(x - x_i)}{(p_{i+1} - p_i)^2} E\frac{\bigtriangledown_n^3}{\widehat{p}_{i+1} - \widehat{p}_i} 1_{B_n^c} + O\left(N^{-2}\right) \\
&= -\frac{(x - x_i)}{(p_{i+1} - p_i)^2}\frac{\{p_i(1 - p_i) + p_{i+1}(1 - p_{i+1})\}}{n} + O\left(m^2/n^{3/2}\right);
\end{aligned} \tag{4.5.26}$$

$$\begin{aligned}
&E\left(\frac{\varepsilon(x)(x - x_i)(x_{i+1} - x_i)}{p_{i+1} - p_i}\left(\frac{\bigtriangledown_n}{\widehat{p}_{i+1} - \widehat{p}_i}\right) 1_{B_n^c}\right) \\
&= \frac{\varepsilon(x)(x - x_i)(x_{i+1} - x_i)}{p_{i+1} - p_i} E\left\{\frac{\bigtriangledown_n^2}{(p_{i+1} - p_i)(\widehat{p}_{i+1} - \widehat{p}_i)}\right\} 1_{B_n^c} = O\left(1/n\right).
\end{aligned} \tag{4.5.27}$$

For the first relation in (4.5.27), use $\frac{\bigtriangledown_n}{\widehat{p}_{i+1} - \widehat{p}_i} = \bigtriangledown_n \left(1 - \frac{\bigtriangledown_n}{\widehat{p}_{i+1} - \widehat{p}_i}\right)/(p_{i+1} - p_i)$, and $E\bigtriangledown_n = 0$. Hence the bias (of $\widetilde{F}^{-1}(p)$ as an estimator of $F^{-1}(p)$) is

$$\left|Bias\left(\widetilde{F}^{-1}(p)\right)\right| = O(m/n) + O(1/m^2). \tag{4.5.28}$$

Subtracting (4.5.24) from (4.5.22), one obtains

$$\begin{aligned}
E\widetilde{F}^{-1}(p) - \widetilde{F}^{-1}(p) = &\left\{(\widehat{p}_i - p_i)\frac{x_{i+1} - x}{p_{i+1} - p_i} + (\widehat{p}_{i+1} - p_{i+1})\frac{x - x_i}{p_{i+1} - p_i}\right\} 1_{B_n^c} \\
&+ O(m/n) + O_p(N^{-2}),
\end{aligned} \tag{4.5.29}$$

noting that $E\left(\varepsilon_{n,3}^2\right) \leq c'''m^2/n^2$ for some constant c'''. The term $O_p(N^{-2})$ is bounded by $c^{iv}1_{B_n}$, for some constant c^{iv}. Therefore,

$$Var(\widetilde{F}^{-1}(p)) = O(1/n) + O(m^2/n^2). \tag{4.5.30}$$

It is relatively simple to check that the contribution from $D_{n,i}^c$, $1 \leq i \leq m-1$, to $MISE\left(\widetilde{\zeta}\right)$ is negligible compared to that from $D_n = \bigcup_{1\leq i\leq m-1} D_{n,i}$. It is useful, however, to show that for all $p \in [0,1]$, one has on B_n^c, the relation

$$F^{-1}(p) - \widetilde{F}^{-1}(p) = \left[\widetilde{F}(x) - F(x)\right] \frac{x_{i+1} - x_i}{\widehat{p}_{i+1} - \widehat{p}_i} \left(1 + \varepsilon_{n,4}\right), \tag{4.5.31}$$

where $\varepsilon_{n,4} = O_p(1/m)$. Indeed, $\mid \varepsilon_{n,4} \mid \leq c^v/m$ on B_n^c, for some $c^v > 0$. To establish (4.5.31), note first that if $p \in D_{n,i}^c \bigcap [p_i, p_{i+1}]$ then, although $\widetilde{F}(x) \in [\widehat{p}_i, \widehat{p}_{i+1}]$ (since $x \in [x_i, x_{i+1}]$), it may happen that $F(x)$ belongs to $(\widehat{p}_{i-1}, \widehat{p}_i)$ or $(\widehat{p}_{i+1}, \widehat{p}_{i+2})$. On B_n^c, there is no other possibility.

Now if $F(x) \in (\widehat{p}_{i-1}, \widehat{p}_i)$, e.g., then, recalling that $x = F^{-1}(p)$,

$$\begin{aligned} F^{-1}(p) - \widetilde{F}^{-1}(p) &\equiv \widetilde{F}^{-1}\left(\widetilde{F}(x)\right) - \widetilde{F}^{-1}\left(F(x)\right) \\ &= \widetilde{F}^{-1}\left(\widetilde{F}(x)\right) - \widetilde{F}^{-1}\left(\widetilde{F}(x_i)\right) + \widetilde{F}^{-1}\left(\widetilde{F}(x_i)\right) - \widetilde{F}^{-1}\left(F(x)\right) \\ &= \left(\widetilde{F}(x) - \widetilde{F}(x_i)\right)\left\{\frac{x_{i+1} - x_i}{\widehat{p}_{i+1} - \widehat{p}_i}\right\} + \left(\widetilde{F}(x_i) - F(x)\right)\left\{\frac{x_i - x_{i-1}}{\widehat{p}_i - \widehat{p}_{i-1}}\right\}, \end{aligned} \tag{4.5.32}$$

in view of the linearity of $\widetilde{F}^{-1}$ on both $[\widehat{p}_i, \widehat{p}_{i+1}]$ and $[\widehat{p}_{i-1}, \widehat{p}_i]$, but with different slopes (given in curly brackets). But the second slope differs from the first by an amount $\varepsilon_{n,4}$ which is easily shown to be no more than c^v/m on B_n^c. The MISE of $\widetilde{F}^{-1}$ is then given by

$$MISE\left(\widetilde{F}^{-1}\right) = O(m^2/n^2) + O(1/m^4) + O(1/n). \tag{4.5.33}$$

Once again, the optimal choice of m is $m \asymp n^{1/4}$, and then the MISE has the optimal rate

$$MISE(\widetilde{F}^{-1}) = O\left(1/n\right) = O\left(N^{-4/5}\right). \tag{4.5.34}$$

(b) Next consider the case $m/n^{1/4} \to \infty$, i.e., $n = o\left(N^{4/5}\right)$. Since $MISE(\widetilde{F}^{-1}) = O\left(1/n\right)$, it is of larger order than $N^{-4/5}$, and hence the estimator is suboptimal. In this case, again consider r groups of essentially disjoint equidistant dosages. Then the average $\widetilde{\zeta}_p = \frac{1}{r}\sum_{j=1}^{r}\widetilde{\zeta}_{p,j}$ has bias and variance (see (4.5.28) and (4.5.30)) given by

$$Bias(\widetilde{\zeta}_p) = O\left(m/(rn)\right) + O\left((r/m)^2\right), \tag{4.5.35}$$

and

$$Var(\widetilde{\zeta}_p) = O\left(1/(rn)\right) + O\left(m^2/(r^3n^2)\right). \tag{4.5.36}$$

Assume that m is not very large, i.e., $\frac{m}{n^{2/3}} \not\to \infty$. Then the optimal choice of r is $r \asymp (m^4/n)^{1/5}$, since the term $m/(rn)$ in (4.5.35) is not of larger order than $(r/m)^2$, and one equates the orders of $1/(rn)$ and $(r/m)^4$ to get the optimal r. This yields the optimal MISE of $\widetilde{\zeta}_p$, namely,

$$MISE\left(\widetilde{\zeta}_p\right) = O\left(1/(rn)\right) = O\left((mn)^{-4/5}\right) = O\left(N^{-4/5}\right).$$

□

Finally, we arrive at the asymptotic distribution of $\widetilde{\zeta}_p$.

Theorem 4.5.4. *Let $p \in (0,1)$. In addition to the hypothesis of Theorem 4.5.2, assume $m/n^{1/4} \to \infty$. Also assume **(A)**. Then the following hold.*

(a) With $r = 1$ and $\widetilde{\zeta}_p = \widetilde{F}^{-1}(p)$, if $m < (2/(c\theta))(n/\log n)^{1/2}$, then

$$\frac{\sqrt{n}(\widetilde{\zeta}_p - \zeta_p)}{\bigtriangledown(p)} \xrightarrow{\mathcal{L}} N\left(0, \frac{p(1-p)}{f^2(p)}\right), \tag{4.5.37}$$

where

$$\bigtriangledown^2(p) \equiv \sum_{i=1}^{m}(x_{i+1} - x_i)^{-2}\left\{(x_{i+1} - x)^2 + (x - x_i)^2\right\}1_{I_i}(p) \tag{4.5.38}$$

$$(I_i = [p_i, p_{i+1}) \text{ for } 1 \le i \le m-2, I_{m-1} = [p_{m-1}, p_m]),$$

lies in [1/2,1].

(b) If $m = o(n^{3/2}/\log^{5/2} n)$*, then with* $r \asymp (m^4/n)^{1/5}$*,*

$$\frac{\sqrt{rn}(\widetilde{\zeta}_p - E\widetilde{\zeta}_p)}{\overline{\bigtriangledown}(p)} \xrightarrow{\mathcal{L}} \left(0, \frac{p(1-p)}{f^2(p)}\right) \tag{4.5.39}$$

Here $\overline{\bigtriangledown}^2(p)$ *is the average of the* r *quantities* $\bigtriangledown_j^2(p)$*,* $1 \le j \le r$*, of the form (4.5.38), one for each subgroup with* m/r *dosages at a distance of* r/m *from each other.*

(c) If $m = o(n^{2/3}/\log\log n)$*, then with* $r \asymp (m^4/n)^{1/5}/(\log\log n)^{6/5}$*,*

$$\frac{\sqrt{rn}(\widetilde{\zeta}_p - \zeta_p)}{\overline{\bigtriangledown}(p)} \xrightarrow{\mathcal{L}} \left(0, \frac{p(1-p)}{f^2(p)}\right). \tag{4.5.40}$$

Proof. (a) It follows from (4.5.22) (and (4.5.31)) that for $p \in [p_i, p_{i+1})$ one has, outside B_n,

$$\begin{aligned}\widetilde{F}^{-1}(p) - F^{-1}(p) = &-\Big\{(\widehat{p}_i - p_i)\frac{x_{i+1} - x}{p_{i+1} - p_i} + (\widehat{p}_{i+1} - p_{i+1})\frac{x - x_i}{p_{i+1} - p_i}\Big\} \\ &+ O(m/n) + O(1/m^2).\end{aligned} \tag{4.5.41}$$

Multiplying the two sides by $\sqrt{n}$, and noting that $m/\sqrt{n} \to 0$, $\sqrt{n}/m^2 \to 0$, the desired Normal approximation holds.

(b) By (4.5.29), one has, outside B_n,

$$\widetilde{F}^{-1}(p) - E\widetilde{F}^{-1}(p) = -\Big\{(\widehat{p}_i - p_i)\frac{x_{i+1} - x}{p_{i+1} - p_i} + (\widehat{p}_{i+1} - p_{i+1})\frac{x - x_i}{p_{i+1} - p_i}\Big\} + O(m/n). \tag{4.5.42}$$

Using the analog of (4.5.42) for $\widetilde{F}_j^{-1}(p) - E\widetilde{F}_j^{-1}(p)$, one may apply Lyapunov's central limit theorem (see, e.g., Bhattacharya and Waymire (2007), p.103) to the r summands $\sqrt{n}(\widetilde{F}_j^{-1}(p) - E\widetilde{F}_j^{-1}(p))$, $1 \le j \le r$, and with m/r for m, to get the desired result. Note that the summands have zero means, variances bounded away from zero and infinity, and bounded third moments, since $\sqrt{n}\frac{m/r}{n} = m/(r\sqrt{n}) \asymp$

$m^{1/5}n^{1/5}/\sqrt{n} = m^{1/5}/n^{3/10} \to 0$ as $m = o\left(n^{3/2}/\log^{5/2} n\right)$, which also ensures that $m/r = o\left(\sqrt{n/\log n}\right)$ (see (4.5.6), (4.5.7)).

(c) One has (see (4.5.28))

$$Bias(\widetilde{\zeta}_p) = O(m/(rn)) + O(r^2/m^2), \tag{4.5.43}$$
$$\sqrt{rn}Bias(\widetilde{\zeta}_p) = O(m/\sqrt{rn}) + O(r^{5/2}\sqrt{n}/m^2) \to 0,$$

since $m/\sqrt{rn} = \left(\frac{m}{n^{2/3}}\log\log n\right)^{3/5} \to 0$, and $\frac{r^{5/2}\sqrt{n}}{m^2} = O\left(\left(\frac{m^2}{\sqrt{n}}/(\log\log n)^3\right)\frac{\sqrt{n}}{m^2}\right) \to 0$.

Hence subtracting the bias from $\widetilde{\zeta}_p$, (4.5.40) follows from (4.5.39). □

Remark 4.5.5. Note that (4.5.40) implies that, with $r \asymp (m^4/n)^{1/5}/\sqrt{\log\log N}$, $\widetilde{\zeta}_p$ has the asymptotic distribution $N(\zeta_p, \nu_n)$ where $v_n = O\left(N^{-4/5}(\log\log N)^{6/5}\right)$.

Remark 4.5.6. It may be noted (see (4.5.35) and (4.5.36)) that the contribution of the bias to the asymptotic variance of $\widetilde{\zeta}_p$ around the true value ζ_p increases as r increases. Hence a bias correction is advisable when r is relatively large.

Chapter 5

Simulation Results and Data Examples

In this chapter, we focus on evaluating the finite sample performances of different methods. A comprehensive comparison study is first carried out among the parametric estimates, the DNP estimates, the B-K estimates and our NAM estimates in terms of both simulated data and actual data examples. Comparisons are mainly in terms of estimated mean integrated square error(MISE) and bootstrap confidence intervals. As the results show, the NAM estimates outperform the other estimates in the majority of cases. Most of the results can be also found in Bhattacharya and Lin (2011).

5.1 Simulation Study: True Bias-Corrected MISEs in A Class of Models

First recall the MISE (mean integrated square error) is defined as

$$E\left(\int\left(\widehat{G}_p-\zeta_p\right)^2 dp\right)$$

given that $\widehat{G}_p$ is an estimate for the true curve ζ_p. The MISE can be written as

$$\begin{aligned} E\left(\int\left(\widehat{G}_p-\zeta_p\right)^2 dp\right) &= \int\left(E\widehat{G}_p-\zeta_p\right)^2 dp+\int VAR\left(\widehat{G}_p\right) dp \\ &= \int Bias\left(\widehat{G}_p\right)^2 dp+\int VAR\left(\widehat{G}_p\right) dp. \end{aligned} \tag{5.1.1}$$

In this short subsection, we compute bias-corrected MISEs or integrated variance, for the DNP and NAM estimates for a class of models considered by Dette and Scheder (2010) with $m = 10$, $n = 5$. The computations in Table 5.1 show that NAM corresponds to $r = 3$ in the population, as it does in the samples. The models are listed as follows:

(1) Linear model: $F(x) = \begin{cases} 2x & \text{if } 0 \le x \le 0.3 \\ 0.4x + 0.48 & \text{if } 0.3 \le x \le 0.8 \\ x & \text{if } 0.8 \le x \le 1 \end{cases}$

(2) Normal Model 1 : $F(x) = \Phi\left((x-\mu)/\sigma\right)$, with $\mu = 0.5$, $\sigma = 0.5$.

(3) Normal Model 2: $F(x) = \Phi\left((x-\mu)/\sigma\right)$, with $\mu = 0.5$, $\sigma = 0.1$.

(4) Weibull Model: $F(x) = 1 - exp\left(-x^{\gamma}\right)$, with $\gamma = 0.52876$.

(5) Cauchy Model: $F(x) = 1/2 + 1/\pi \arctan\left((x-\mu)/\sigma\right)$, with $\mu = 0.15$, $\sigma = 0.05$.

(6) Beta Model: $F'(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(1-x\right)^{\beta-1} x^{\alpha-1}$, with $\alpha = 2$, $\beta = 3$.

(7) Logistic Model: $F(x) = 1/(1 + exp\left(5 - 15x\right))$.

TABLE 5.1. Integrated variance with m=10, n=5.

	B-K	r=2	r=3	DNP	NAM
Linear Model:	0.0140	0.0109	0.0051	0.0088	0.0051(r=3)
Normal1 Model:	0.0192	0.0150	0.0077	0.0130	0.0077(r=3)
Normal2 Model	0.0028	0.0021	0.0010	0.0014	0.0010(r=3)
Weibull Model:	0.0263	0.0210	0.0127	0.0210	0.0127(r=3)
Cauchy Model:	0.0051	0.0045	0.0027	0.0030	0.0027(r=3)
Beta Model:	0.0076	0.0055	0.0028	0.0039	0.0028(r=3)
Logistic Model:	0.0036	0.0025	0.0015	0.0017	0.0015(r=3)

5.2 Simulation Study: Comparison of Confidence Intervals.

In this section, 95% confidence intervals are obtained by applying different methods to data simulated from different models as described below.

For data simulated from the logistic model $F(x) = \frac{1}{1+(\exp(-\alpha-\beta x))}$, we take $\alpha = -20$, $\beta = 10$. The confidence intervals for the MLEs of different quantiles are obtained by fitting the logistic model to data. Denote by $(\hat{\alpha}, \hat{\beta})$ the MLE of (α, β). Let v_{11}, v_{22}

be the asymptotic variances of $\widehat{\alpha}$ and $\widehat{\beta}$, respectively, and v_{12} the covariance of the two. Also write $d_p = \log(p/(1-p))$, $ED_p = (d_p - \alpha)/\beta$. Then the MLE $\widehat{ED}_p$ for ED_p is given as $(d_p - \hat{\alpha})/\hat{\beta}$, and the 95% CI for ED_p is given by using the Fieller method as $[\widehat{ED}_p - 1.96\sqrt{\hat{w}}/\hat{\beta},\ \widehat{ED}_p + 1.96\sqrt{\hat{w}}/\hat{\beta}]$, where $\hat{w} = v_{11} + 2v_{12}\widehat{ED}_p + v_{22}(\widehat{ED}_p)^2$ (see Piegorsch and Bailer (2005), pp. 30, 39-40).

We take samples from the probit model $F(x) = \Phi((x-\mu)/\sigma)$ with $\mu = 0.5$ and $\sigma = 0.3$.

Simulations from the Beta distribution $F(x) = \int_0^x ((B(\alpha,\beta))^{-1} t^{\alpha-1}(1-t)^{\beta-1})dt$ $(0 \le x \le 1)$ are carried out with $\alpha = 2$ and $\beta = 3$.

The last set of simulations are from the Weibull model $F(x) = 1 - exp(-(x/\alpha)^{\beta})$ with $\alpha = 2$ and $\beta = 1.5$.

The computations of the MLEs of the parameters of the models are carried out on the Matlab, starting with some properly chosen easily computable consistent estimators as initial values.

Bootstrapping is used for constructing bias-corrected confidence intervals for the DNP and the NAM estimates and the MLEs of the last three models. For the NAM, the value of r in each case is the one among $r = 1, 2, 3$ for which the bootstrap estimate of $MISE(\widetilde{\zeta}_p)$ is the smallest.

Simulation studies were carried out for computing confidence bands for the cases $m = 5$, 10, and $n = 5$, 10, 25,50, with samples from each of the four models-Logistic, Probit, Beta and Weibull, as mentioned above. Here we display the graphs of confidence bands are provided for the DNP, NAM and MLE, only for data obtained from the Probit and Weibull models, but for other models, similar results follow. In the graphs, the blue line represents the true curve of the quantiles, while the green lines represent 95% confidence intervals for the MLE obtained by bootstrapping. The red lines with circles represent the confidence intervals obtained by NAM with $r = 2$ for $m = 5$ and $r = 3$ for $m = 10$, and the dark dashed lines represent the confidence bounds of DNP estimates. When the bias is not corrected, the confidence bands of

the DNP and NAM still include the true curve in almost all cases. Here we include only the bias corrected figures for the DNP and NAM, and, of course, Tables 5.1-5.16 are the same whether the bias is corrected or not.

A number of significant features are revealed by the graphs and tables. First, both the DNP and the NAM tend to perform better than the MLE for small values of n such as $n = 5$. The MLE also seems to be more bias prone in small samples than the two nonparametric methods. Secondly, the DNP and the NAM are more or less on parallel in performance with small n ($n = 5$). This makes sense because the DNP was originally devised for large m and small n (such as $n = 1$) (see Dette at al. (2005)). For larger n ($n = 10$ or more), the NAM outperforms the DNP in almost all cases. A succinct comparison of the two nonparametric methods, by the lengths of confidence intervals, for all four models (Logistic, Probit, Beta and Weibull) is provided in Table 5.18. Here an entry (i, j) means that for i dose levels the NAM has shorter intervals than the DNP, and for j dosages the DNP has shorter intervals, for the particular model and particular values of (m, n) as indicated.

Although we do not show the confidence intervals for the B-K method ($r = 1$) in order to avoid cluttering up of the graphs, our computations show that they are wider than those of the DNP and NAM in most cases.

FIGURE 5.1. [Probit]

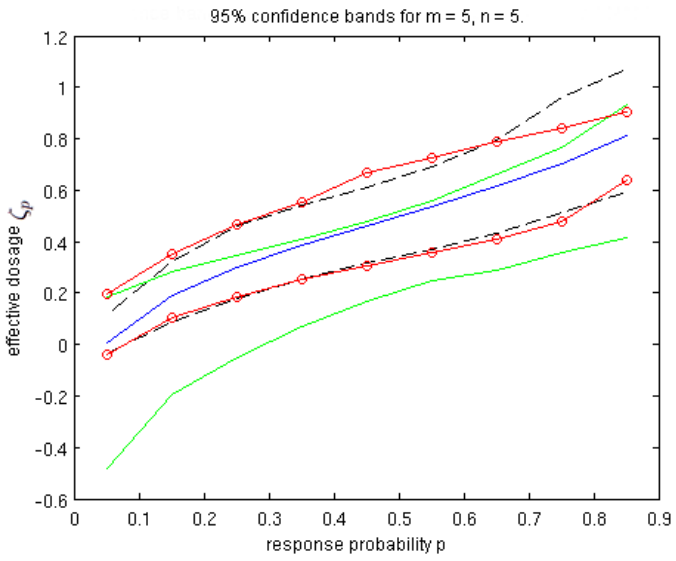


TABLE 5.2. The Length of Confidence Intervals for DNP and NAM (r=2) for m=5,n=5

DNP:	0.1487	0.2320	0.2779	0.2871	0.2944	0.3224	0.3601	0.4471	0.4750
NAM:	0.2331	0.2479	0.2861	0.3009	0.3627	0.3681	0.3795	0.3597	0.2624

TABLE 5.3. The Length of Confidence Intervals for DNP and NAM (r=2) for m=5,n=10

DNP:	0.2159	0.3341	0.4126	0.4264	0.3729	0.3423	0.3422	0.4062	0.2959
NAM:	0.2144	0.2835	0.3498	0.3716	0.3611	0.3404	0.3265	0.4057	0.2604

FIGURE 5.2. [Probit]

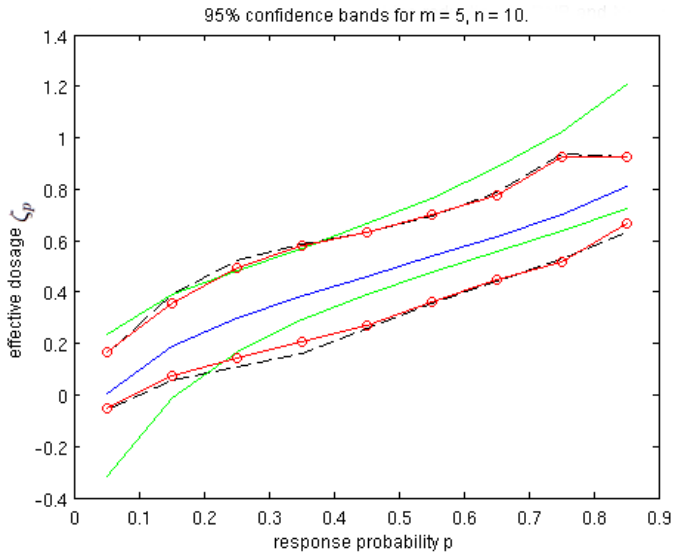


TABLE 5.4. The Length of Confidence Intervals for DNP and NAM (r=2) for m=5,n=25

DNP:	0.2382	0.2785	0.2533	0.2119	0.2286	0.2034	0.2014	0.2269	0.2443
NAM:	0.1575	0.2329	0.2091	0.1857	0.1884	0.1817	0.1797	0.1939	0.1954

FIGURE 5.3. [Probit]

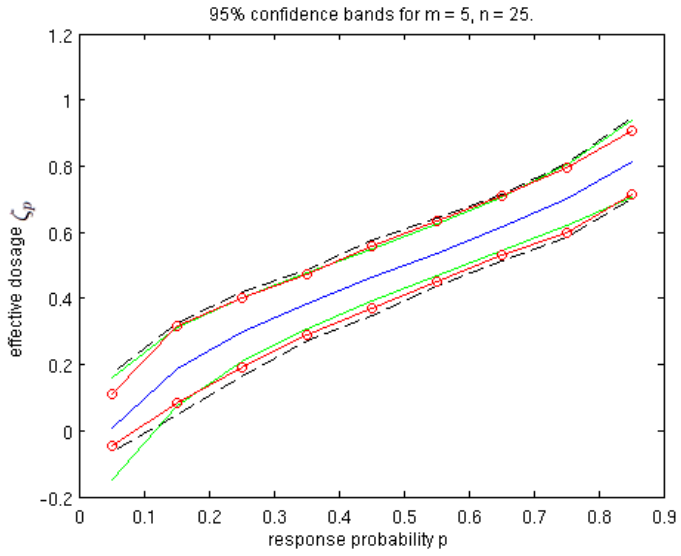


FIGURE 5.4. [Probit]

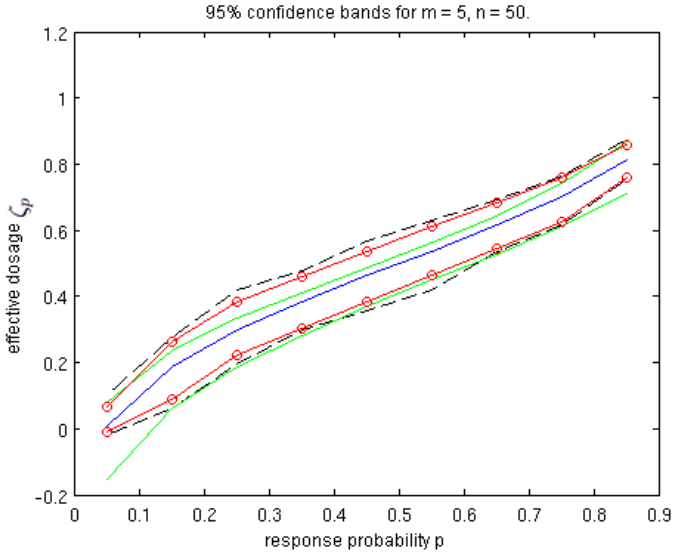


TABLE 5.5. The length of confidence intervals for DNP and NAM (r=2) for m=5,n=50

DNP:	0.1206	0.2162	0.2257	0.1806	0.2102	0.2098	0.1580	0.1478	0.1194
NAM:	0.0767	0.1753	0.1607	0.1560	0.1505	0.1481	0.1389	0.1343	0.0982

FIGURE 5.5. [Probit]

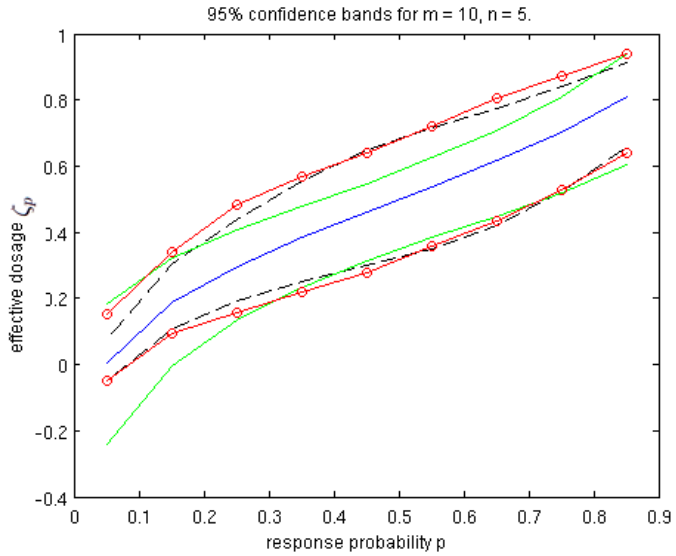


TABLE 5.6. The Length of Confidence Intervals for DNP and NAM (r=3) for m=10,n=5

DNP:	0.1238	0.1962	0.2452	0.3028	0.3492	0.3638	0.3565	0.3171	0.2571
NAM:	0.2030	0.2497	0.3244	0.3479	0.3625	0.3627	0.3691	0.3409	0.2962

FIGURE 5.6. [Probit]

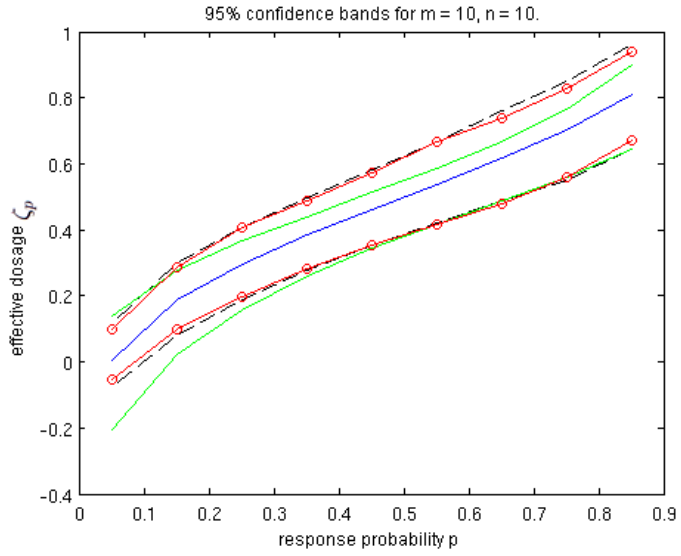


TABLE 5.7. The Length of Confidence Intervals for DNP and NAM (r=3) for m=10,n=10

DNP:	0.1916	0.2217	0.2194	0.2188	0.2293	0.2444	0.2735	0.2962	0.3149
NAM:	0.1534	0.1904	0.2139	0.2072	0.2229	0.2519	0.2580	0.2700	0.2671

TABLE 5.8. The Length of Confidence Intervals for DNP and NAM (r=3) for m=10,n=25

DNP:	0.1380	0.2559	0.2370	0.2489	0.2447	0.1917	0.1216	0.1506	0.2003
NAM:	0.1227	0.1853	0.2108	0.2197	0.1695	0.1396	0.1215	0.1169	0.1274

TABLE 5.9. The length of confidence intervals for DNP and NAM (r=3) for m=10,n=50

DNP:	0.1602	0.1931	0.1580	0.1354	0.1305	0.1746	0.2017	0.1498	0.0924
NAM:	0.1186	0.1134	0.1128	0.1035	0.1021	0.1031	0.1075	0.1108	0.0983

FIGURE 5.7. [Probit]

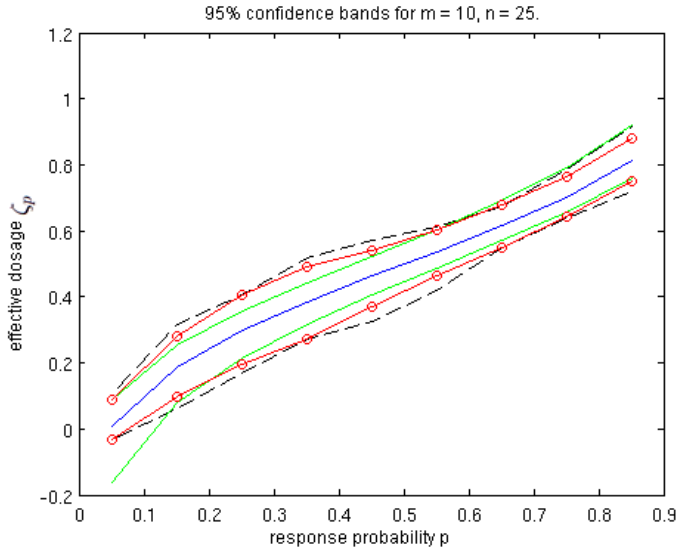


FIGURE 5.8. [Probit]

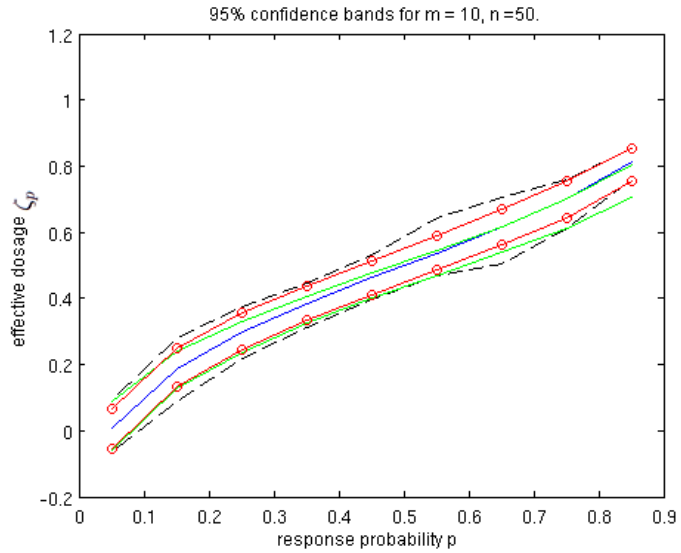


FIGURE 5.9. [Weibull]

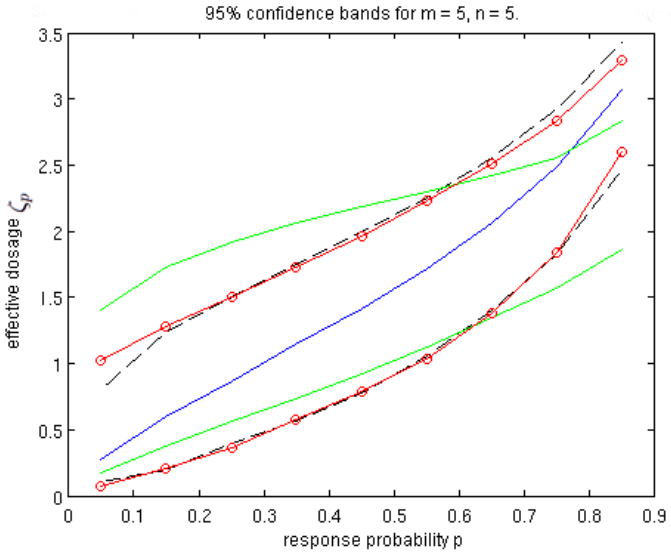


TABLE 5.10. The Length of Confidence Intervals for DNP and NAM (r=2) for m=5,n=5

DNP:	0.6969	1.0458	1.1129	1.1841	1.2247	1.1968	1.1445	1.0868	0.9663
NAM:	0.9469	1.0721	1.1421	1.1569	1.1716	1.1863	1.1274	0.9948	0.6963

TABLE 5.11. The Length of Confidence Intervals for DNP and NAM (r=2) for m=5,n=10

DNP:	0.2949	0.5102	0.6378	0.7579	0.8538	1.0105	1.0884	1.2061	1.2235
NAM:	0.1982	0.4228	0.5281	0.6325	0.7405	0.9112	1.0303	1.1790	1.2932

FIGURE 5.10. [Weibull]

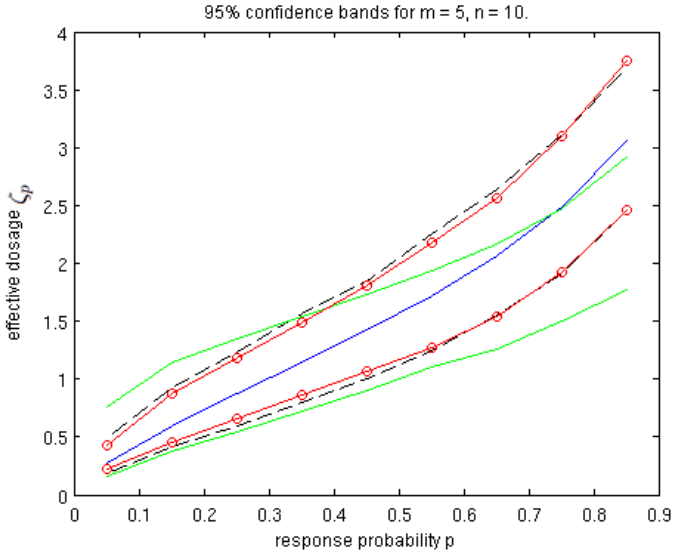


TABLE 5.12. The Length of Confidence Intervals for DNP and NAM (r=2) for m=5,n=25

DNP:	0.1768	0.4657	0.5843	0.6694	0.8776	0.8382	0.7878	0.8514	1.3161
NAM:	0.1219	0.3443	0.4422	0.5409	0.6491	0.6968	0.7236	0.8510	0.8382

FIGURE 5.11. [Weibull]

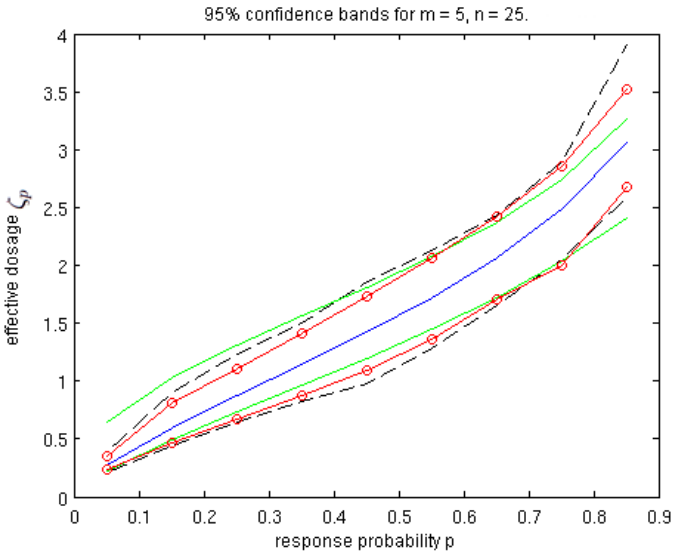


FIGURE 5.12. [Weibull]

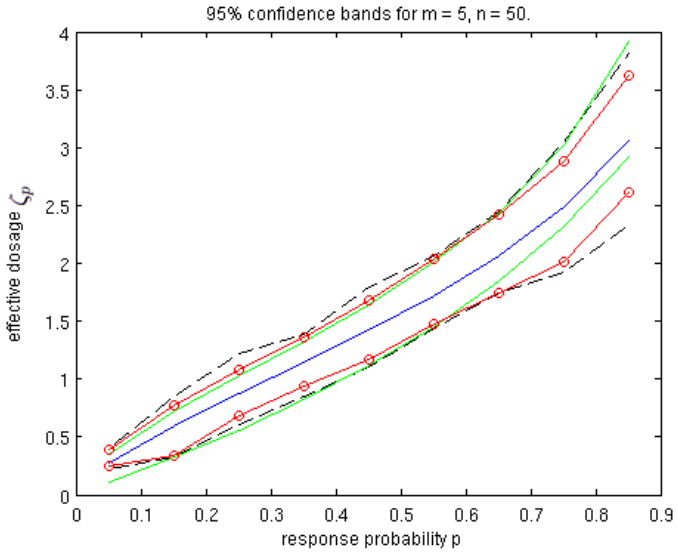


TABLE 5.13. The length of confidence intervals for DNP and NAM (r=2) for m=5,n=50

DNP:	0.1768	0.5299	0.6163	0.5316	0.6977	0.6190	0.7105	1.1189	1.4846
NAM:	0.1432	0.4354	0.4041	0.4221	0.5095	0.5638	0.6736	0.8739	1.0095

FIGURE 5.13. [Weibull]

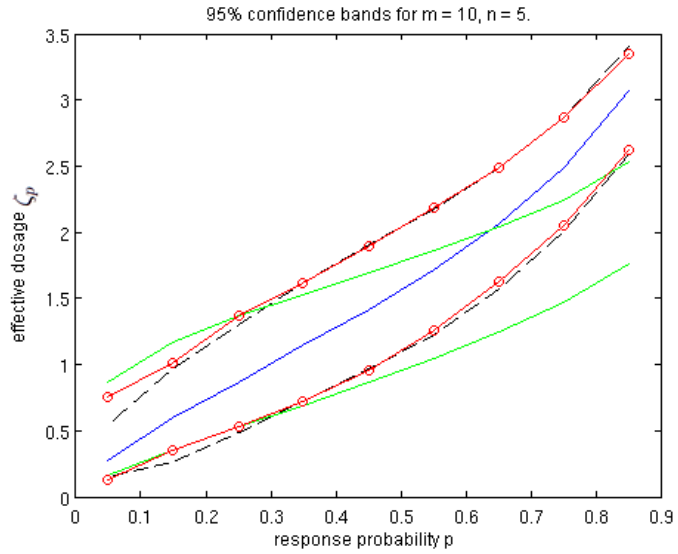


TABLE 5.14. The Length of Confidence Intervals for DNP and NAM (r=3) for m=10,n=5

DNP:	0.3918	0.7013	0.8197	0.8952	0.9395	0.9487	0.9247	0.8603	0.8180
NAM:	0.6207	0.6573	0.8481	0.8931	0.9406	0.9233	0.8597	0.8192	0.7286

FIGURE 5.14. [Weibull]

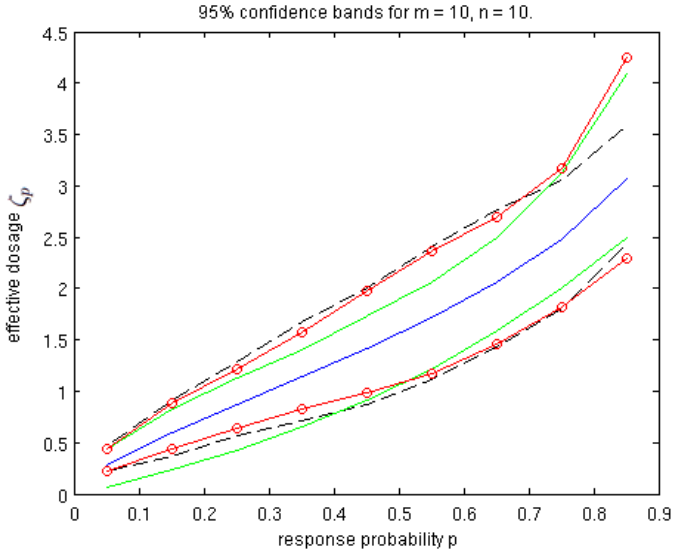


TABLE 5.15. The Length of Confidence Intervals for DNP and NAM (r=3) for m=10,n=10

DNP:	0.2442	0.5456	0.7215	0.9547	1.1352	1.2921	1.3359	1.2411	1.1389
NAM:	0.2163	0.4406	0.5700	0.7428	0.9899	1.1956	1.2320	1.3621	1.9545

TABLE 5.16. The Length of Confidence Intervals for DNP and NAM (r=3) for m=10,n=25

DNP:	0.1969	0.5101	0.7398	0.7254	0.6308	0.6309	0.8897	1.0056	0.8166
NAM:	0.1608	0.3385	0.4566	0.5475	0.5700	0.5768	0.6072	0.7201	0.8915

TABLE 5.17. The length of confidence intervals for DNP and NAM (r=3) for m=10,n=50

DNP:	0.1850	0.4590	0.6483	0.5250	0.4681	0.7402	0.9252	0.7402	0.7703
NAM:	0.1589	0.3223	0.3255	0.3734	0.4168	0.4837	0.5507	0.5769	0.5607

FIGURE 5.15. [Weibull]

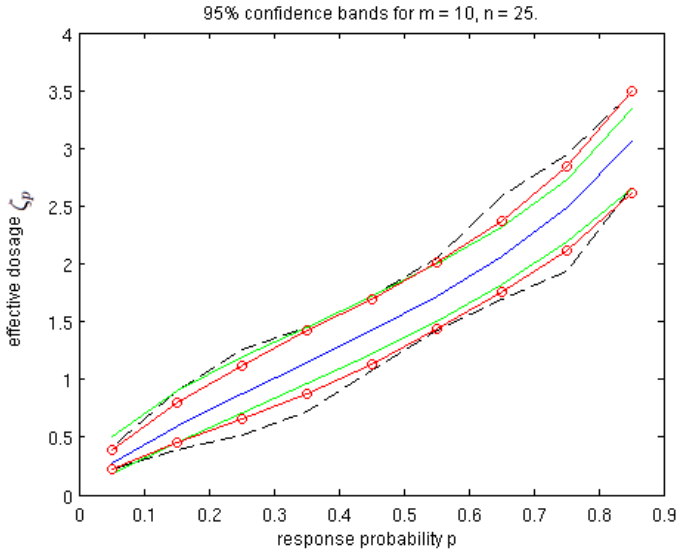


FIGURE 5.16. [Weibull]

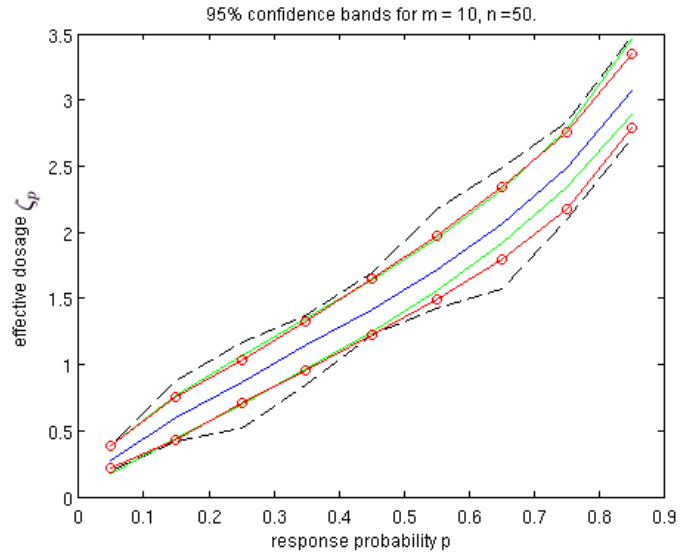


TABLE 5.18. Comparison by Number of Shorter Confidence Intervals

		Model			
m	n	Logistic	Probit	Beta	Weibull
5	5	(5,4)	(2,7)	(5,4)	(6,3)
5	10	(6,3)	(9,0)	(9,0)	(8,1)
5	25	(9,0)	(9,0)	(9,0)	(9,0)
5	50	(9,0)	(9,0)	(9,0)	(9,0)
10	5	(4,5)	(2,7)	(5,4)	(6,3)
10	10	(7,2)	(8,1)	(7,2)	(7,2)
10	25	(8,1)	(9,0)	(7,2)	(8,1)
10	50	(9,0)	(8,1)	(9,0)	(9,0)

An entry (i, j) in Table 5.18 indicates that for the estimation of i effective dosages (for the particular model, and particular combination of (m, n) for the cell), the NAM provides shorter confidence intervals than the DNP, and for j of them the DNP has shorter confidence intervals.

5.3 Real Data Examples

5.3.1 Data Example I: Cancer Remission Data

In this subsection, in a real data example, estimates of $ED_{0.5}$ (also denoted as ED_{50}) are obtained by fitting parametric models and also using nonparametric methods such as the B-K, NAM and DNP. The source of the data is Lee (1974). This cancer remission data set has 27 observations (binary variable with 1 for remission of cancer). There are 14 'dose levels' between 8 and 38 for the explanatory variable labeling index (LI). This labeling index measures proliferative activity of cells after a patient receives an injection of tritiated thymidine, representing the percentage of cells that are labeled. The number of observations are 1 for 6 levels, 2 for 3 levels and 3 for 5 levels. We are interested in estimating the quantile curve and, in particular, the effective dosage $ED_{0.5}$.

Lee (1974) fitted the logistic model to the data, while Dette and Scheder (2010)

fitted Cauchy and Weibull models. The maximum likelihood estimators for $ED_{0.5}$ are 26.05, 23.65 and 26.09, respectively, under the three models. The estimated curves are given by $\widehat{F}(x) = \frac{1}{1+exp(-3.777+0.145x)}$ for the Logistic model, $\widehat{F}(x) = 1 - exp\left(-(0.00028x)^{2.3954}\right)$ for the Weibull model and $\widehat{F}(x) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-23.6474}{6.2391}\right)$ for the Cauchy model.

For the nonparametric estimates of $ED_{0.5}$, Dette and Scheder (2010) computed the DNP estimate as 20.35229, and the B-K estimate as 8. After checking repeatedly, our calculations show that the B-K estimate is actually 17.8462. Also, for an estimate of $ED_{0.5}$ using the NAM method proposed in this paper, we obtained 21.2 with $r = 2$ and 25.1667 with $r = 3$. Since the number of dosages is much larger here than the numbers of responses at the individual level, we would prefer $r = 3$ for the NAM. Note that the NAM estimate is then closer to all the MLE's than the DNP estimate. The estimates of $ED_{0.5}$ are given in Table 5.19 for the different methods. The estimates of ED_p with $p = 0.1$, 0.2 ,0.3, 0.4, 0.5, 0.6, 0.7, 0,8, 0.9, 1 for both parametric and nonparametric methods are illustrated in Figure 5.17. Asymptotic comparisons are not feasible here using bootstrapping, since the numbers of responses n_i are very small. This data example is a follow up of the calculations carried out in Dette and Scheder (2010), and it is only meant to illustrate that the NAM may be applied even with such small sample sizes.

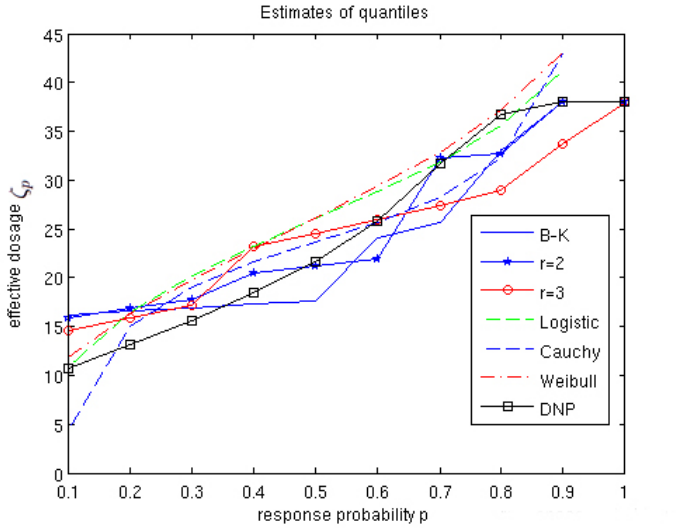
TABLE 5.19. The estimates of ED_{50} using different methods.

Methods:	B-K	r=2	r=3	DNP	Logistic	Cauchy	Weibull
$\widehat{ED_{50}}$:	17.8462	21.2	25.1667	20.35229	26.5	23.65	26.09

5.3.2 Data Example II: Data on Insect Mortality

The logistic model together with the nonparametric methods B-K, NAM (r=2, 3) and DNP, are used to fit the data in this example which records kills of Tribolium confusum following 5-hour exposures to known concentrations of Carbon Disulphide.

FIGURE 5.17.



The data are due to Strand and may be found in Bliss (1935). This data set has 8 dosages from 49.06 mg/litre to 76.04 mg/litre, with around 30 insects for each dosage.

The estimates of ED_p for $p = 0.1, 0.2, \ldots, 0.9$ are obtained and the estimated quantile curves are illustrated in Figure 5.18.

The MISE is calculated for the logistic model, probit model and Weibull models and for each of the nonparametric methods, through bootstrapping. The NAM with $r = 3$ yields the smallest estimated MISE 0.9130 compared to 1.0413 of the Probit model and 1.9374 of DNP as shown in Table 5.20.

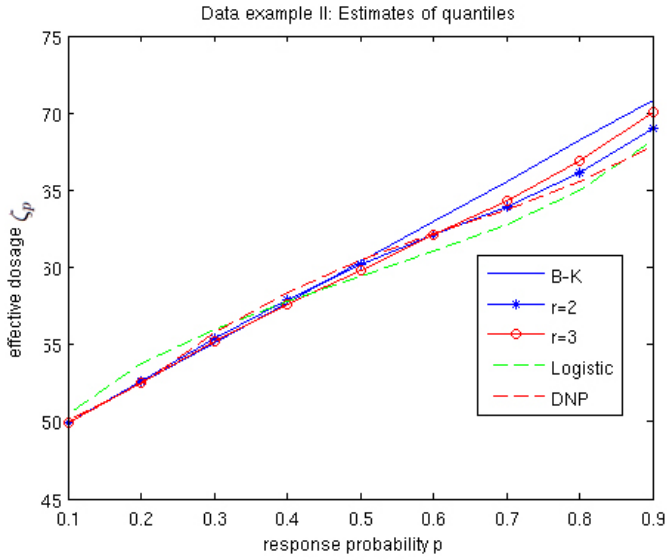
TABLE 5.20. Estimated MISEs of the estimates of the B-K, DNP, NAM and MLEs.

Methods:	B-K	r=2	r=3	DNP	NAM (r=3)	Logistic	Probit	Weibull
MISE:	2.2090	1.3258	0.9130	1.9374	0.9130	1.2355	1.0413	1.2077

5.3.3 Data Example III: the Toxicity of Rotenone

In this example using data from Martin (1942), we investigate the relation between the dose of rotenone (mg/l) and the response of the toxin exposed insects. The percentage of insects dead or seriously affected out of about 50 insects are recorded

FIGURE 5.18.



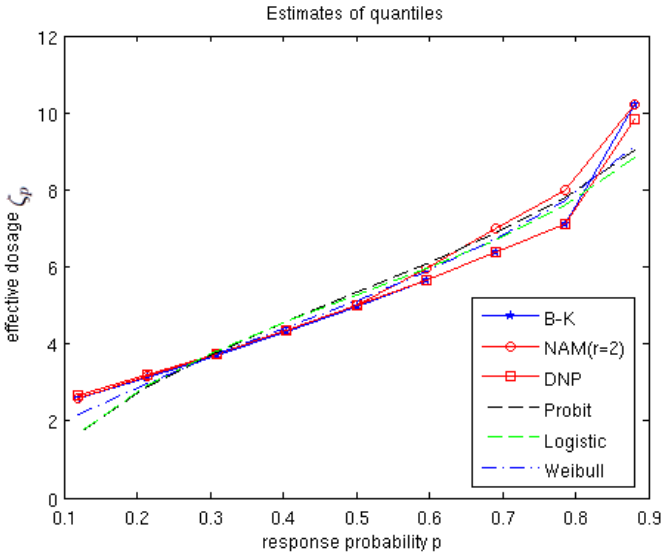
for each of the 5 dosages from 2.6mg/l to 10.2mg/l. The parametric estimates are obtained by fitting the Logistic, Probit and Weibull models to data together with the nonparametric estimates of the B-K, NAM and DNP. The results are shown in Figure 5.19.

Comparisons are carried out among the different methods by their estimated MISEs through bootstrapping as shown in Table 5.21. The Weibull models yields the smallest estimated MISE 0.1626 followed by that of the NAM estimates which is 0.1888.

TABLE 5.21. Estimated MISEs of the estimates of the B-K, DNP, NAM and MLEs.

Methods:	B-K	r=2	DNP	NAM (r=2)	Logistic	Probit	Weibull
MISE:	0.4070	0.1888	0.3512	0.1888	0.2066	0.1908	0.1626

FIGURE 5.19.


Estimates of quantiles
effective dosage ζ_p
response probability p
B-K
NAM(r=2)
DNP
Probit
Logistic
Weibull
0
2
4
6
8
10
12
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

Chapter 6

Another Method for Monotone Curve Estimation

6.1 Introduction

In this Chapter we develop another method for the estimation of the dose-response curve F and its inverse F^{-1} which applies to large values of m (the number of dosages) not covered by *NAM*. We also show that this new method, as well as the NAM described in Chapter 4, may be extended to general isotonic regression effectively. The method of grouping described here was also used by Wright (1982). Our results are proved under milder hypotheses and are, therefore, more widely applicable.

6.2 The Methodology

First consider the dose-response problem with observations (x_i, y_i), $1 \leq i \leq N$, where $a = x_0 \leq x_1 \leq \cdots \leq x_N = b$, with m *distinct dosages* $z_1 < z_2 < \cdots < z_m$. We set $z_1 = a$, and assume for simplicity that $z_i - z_{i-1} = \frac{b-a}{m}$ $\forall i = 1, \cdots, m$. Here $m \to \infty$, $m \leq N$. Let k be a positive integer, $k < m$. We divide the m dosages into approximate m/k groups of dosages $(z_1, z_2, \cdots, z_k)$, $(z_{k+1}, \cdots, z_{2k})$, $\cdots$, etc. If k is odd, then $\bar{z}_1 = z_{(k+1)/2}$, $\bar{z}_2 = z_{k+(k+1)/2}$, $\cdots$, $\bar{z}_j = z_{(j-1)k+(k+1)/2}$, $\cdots$, etc. If k is even, then $\bar{z}_1 = \dfrac{z_{k/2} + z_{k/2+1}}{2}$, $\bar{z}_j = \dfrac{z_{(j-1)k+k/2} + z_{(j-1)k+\frac{k}{2}+1}}{2}$, $\cdots$.

Write

$$\bar{F}_j = \frac{1}{k} \sum_{i=1}^{k} F(z_{(j-1)k+i}). \tag{6.2.1}$$

Then, for the case k is odd, assuming F is twice continuously differentiable on

$[a, b]$,

$$\begin{aligned}\bar{F}_j - F(\bar{z}_j) &= \frac{1}{k}\sum_{i=1}^{k}\left\{F(z_{(j-1)k+i}) - F(\bar{z}_j)\right\} \\ &= \frac{1}{k}\sum_{i=1}^{k}\left\{\frac{b-a}{m}\left(i - \frac{k+1}{2}\right)F'(\bar{z}_j) + O\left(\frac{i-\frac{k+1}{2}}{m}\right)^2\right\} \\ &= \frac{1}{km^2}\sum_{i=1}^{k} O\left(i - \frac{k+1}{2}\right)^2 = O\left(\frac{k^2}{m^2}\right). \end{aligned} \tag{6.2.2}$$

Now define

$$\begin{aligned}\bar{Y}_{j,i} &= \frac{1}{n}y_{(j-1)k+i)}\ (i = 1, \cdots k; j = 1, 2, \cdots), \\ \widehat{\bar{F}}_j &= \frac{1}{k}\sum_{i=1}^{k}\bar{Y}_{j,i}. \end{aligned} \tag{6.2.3}$$

Our new estimate of $F(\bar{z}_j)$ is $\widetilde{\bar{F}}_j$ at design points $\bar{z}_j$ ($j = 1, \cdots, m/k$ (approximately)), where $\widetilde{\bar{F}}_j$ ($j = 1, \cdots,$) is derived by the PAV algorithm applied to $(\bar{z}_j, \widehat{\bar{F}}_j)$, $1 \le j \le m/k$.

Next,

$$\begin{aligned}\widehat{\bar{F}}_j - F(\bar{z}_j) &= \frac{1}{k}\sum_{i=1}^{k}\left(\bar{Y}_{j,i} - \bar{F}_j\right) + \bar{F}_j - F(\bar{z}_j) \\ &= \frac{1}{k}\sum_{i=1}^{k}\left(\bar{Y}_{j,i} - F(z_{(j-1)k+i})\right) + \bar{F}_j - F(\bar{z}_j), \end{aligned} \tag{6.2.4}$$

so that, using (6.2.2)

$$\begin{aligned}E\left(\widehat{\bar{F}}_j - F(\bar{z}_j)\right)^2 &= \frac{1}{k^2}\sum_{i=1}^{k} Var(\bar{Y}_{j,i}) + (\bar{F}_j - F(\bar{z}_j))^2 \\ &= O\left(\frac{1}{nk}\right) + O\left(\frac{k^4}{m^4}\right). \end{aligned} \tag{6.2.5}$$

The optimal rate of k is, therefore,

$$k \asymp \left(\frac{m^4}{n}\right)^{1/5} = m/N^{1/5}, \tag{6.2.6}$$

in which case

$$E(\widehat{\bar{F}}_j - F(\bar{z}_j))^2 = O\left(\frac{N^{1/5}}{m^4}\right) = O\left(N^{-4/5}\right). \tag{6.2.7}$$

Assuming $F' > 0$ on $[a,b]$, one has $F(\bar{z}_{j+1}) - F(\bar{z}_j) \geq c(\bar{z}_{j+1} - \bar{z}_j) = c\frac{(b-a)k}{m} > c'N^{-1/5}$, where $c = \min\{F'(x), q \leq x \leq b\}$, and c' is another positive constant. Noting that $\bar{F}_j = E\widehat{\bar{F}}_j$, and using (6.2.2), one has $\bar{F}_{j+1} - \bar{F}_j > c''N^{-1/5}$ for some $c'' > 0$. From Bernstein's inequality (Serfling (1980), p.95), it also follows that

$$\begin{aligned} P\left(\widehat{\bar{F}}_j > \widehat{\bar{F}}_{j+1} \text{ for some } j\right) &\leq P\left(\left|\widehat{\bar{F}}_j - E\widehat{\bar{F}}_j\right| \geq \frac{c''N^{-1/5}}{2} \text{ for some } j\right) \\ &\leq 2\exp\left\{\frac{-k^2(c''N^{-1/5})/2)^2}{2\sum_{i=1}^{k} var(\bar{Y}_{j,i}) + \frac{2}{3}k(c''N^{-1/5}/2)}\right\} \\ &\leq 2(m/k)\exp\left\{\frac{-k^2(c''N^{-1/5}/2)^2}{2k/4n + \frac{2}{3}k(c''N^{-1/5}/2)}\right\}. \end{aligned} \tag{6.2.8}$$

Since $\frac{c}{a+b} \geq \min\left\{\frac{c}{2a}, \frac{c}{2b}\right\}$ $(c > 0, a > 0, b > 0)$, and k is of larger order than $N^{-1/5}\log N$ (since $\frac{k}{m} \asymp N^{-1/5}$, and $m \geq N^{3/5}\log N$), it follows that

$$\begin{aligned} P\left(\widehat{\bar{F}}_j > \widehat{\bar{F}}_{j+1} \text{ for some } j\right) &\leq \frac{m}{k}\exp\{-\min(2k(c''N^{-1/5}/2)^2, 2k(c''N^{-1/5}/2)\} \\ &= 2\frac{m}{k}\exp\left\{-c'''kN^{-2/5}\right\} \leq 2\frac{m}{k}\exp\left\{-\log N^{1/5+2}\right\} \\ &= O(N^{-2}). \end{aligned} \tag{6.2.9}$$

Since $\sum_{N=1}^{\infty} N^{-2} < \infty$, it follows that $\widehat{\bar{F}}_j$ are monotone increasing in j, outside a set of probability $O(N^{-2})$. Hence $\widetilde{\bar{F}}_j = \widehat{\bar{F}}_j \ \forall\ j$, outside a set of negligible probability $O(N^{-2})$. We now define $\widetilde{\bar{F}}$ on $[a,b]$, by letting $\widetilde{\bar{F}}_{\bar{z}_j} = \widetilde{\bar{F}}_j$, and linearly interpolating in (z_j, z_{j+1}) for all j.

Theorem 6.2.1. *Assume F is twice continuously differentiable on $[a,b]$, $F' > 0$ on $[a,b]$, and the design points are equidistant $z_i - z_{i-1} = (b-a)/m$, $m > \alpha N^{3/5}\log N$ for an appropriate constant $\alpha > 0$. Then the following hold:*

(a) The MISE of $\widetilde{\bar{F}}$ as an estimate of F is of optimal order $O\left(N^{-4/5}\right)$, for a choice of k satisfying (6.2.6).

(b)

$$\frac{\widetilde{\bar{F}}(z) - E\widetilde{\bar{F}}(z)}{\sqrt{\nu_N(z)}} \xrightarrow{L} N(0,1),\ \forall z \in (a,b),$$

where $\nu_N(z) = Var(\widetilde{\bar{F}}(z)) \leq c' N^{-4/5}$, *for some* $c' > 0$.

(c) If one chooses k *such that* m/k *is of larger order than* $N^{1/5}$,

$$\frac{\widetilde{\bar{F}}(z) - F(z)}{\sqrt{\nu_N(z)}} \xrightarrow{L} N(0,1),\ \forall z \in (a,b),$$

with $\nu_N(z) >> N^{-4/5}$. *In particular, with* $m/k \asymp N^{1/5} \log\log N$, $\nu_N(z) = O\left(N^{-4/5} \log\log N\right)$.

Proof. We have already proved (a). The proof of (b) follows Lyapunov's CLT, as in the proof of Theorem 4.5.4(b). Part (c) follows from part (b), noting that under given assumptions on m/k, $E\widetilde{\bar{F}}(z) - F(z) = o\left(k^2/m^2\right) = o(N^{-2/5}) = o(\sqrt{\nu_N(z)})$. □

Remark 6.2.2. The assumption $m \geq \alpha N^{3/5} \log N$ in Theorem 6.2.1 complements the assumption $m = o(N^{3/5})$ in Theorem 4.5.2.

Remark 6.2.3. The assumption of equidistributed z_i's may be relaxed to boundedness of the ratio $\max_i \{z_i - z_{i-1}\} / \min_i \{z_i - z_{i-1}\}$, as $N \to \infty$, $m \to \infty$. Similarly, the assumption of constant $n_i = n\ \forall i$ may be replaced by the boundedness of the ratio $\max_i n_i / \min_i n_i$.

Remark 6.2.4. In applications with real data the group size k may be decided on the basis of estimates of the MISE for a range of reasonable values of k.

6.3 Extension to General Regression

Consider now the general isotonic regression problem described in Chapter 1, with observations (x_i, y_i), $y_i = F(x_i) + \epsilon_i$, $x_1 \leq x_2 \leq \cdots \leq x_N$. Assume ϵ_i's are independent, $E\epsilon_i = 0$, and ϵ_i is bounded. Then Theorems 4.5.2-4.5.4 of chapter 4 extend

almost verbatim with essentially the same proofs. One does not actually need the ϵ_i's to be bounded to apply Bernstein's inequality, but only that ϵ_i has a finite moment generating function (m.g.f) in a neighborhood of 0 (see, e.g., Loéve (1977)). We state the theorems here.

Let $\sigma^2(x)$ denote the variance of $y = F(x) + \epsilon$, and assume that $\sigma^2(x)$ and the weight function $w(x)$ are both positive and continuous on [0,1]. Let $0 < x_1 < \cdots < x_m = 1$ be equidistant design points, with n observations at each. Let S_i be the sum of the observations at x_m $(i = 1, \cdots, m)$. Let $\widetilde{F}$ denote the linear interpolation of

$$\widetilde{F}(x_j) = \max_{\{s:x_s \le x_j\}} \min_{\{t:x_j \le x_t\}} \frac{\sum_{i:x_s \le x_i \le x_t} w(x_i) S_i/n}{\sum_{i:x_s \le x_i \le x_t} w(x_i)} \ (j = 1, \cdots, m).$$

Theorem 6.3.1. *With the same assumptions on F made for the dose-response function, the conclusion of Theorems 4.5.2-4.5.4 and 6.21 hold, with the variance function $\sigma^2(x)$ in place of $F(x)(1 - F(x))$, provided ϵ has a finite moment generating function (mgf) in a neighborhood of zero.*

Proof. The proofs are essentially the same as those in the cited theorems with $F(x_i)$ for p_i, $\dfrac{S_i}{n}$ for $\widehat{p}_i$, etc. □

References

[] Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T. and Silverman, E. *An empirical distribution function for sampling with incomplete information.* Ann. Math. Statist. **26** (1955), 641–647.

[] Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. *Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression.* Wiley, London, (1972).

[] Bhattacharya, R. and Kong, M. *Consistency and asymptotic normality of the estimated effective dose in bioassay.* J. Statist. Plan. Inf. **137** (2007), 643-658.

[] Bhattacharya, R and Lin, L. *An adaptive nonparametric method in benchmark analysis for bioassay and environmental studies.* Statist. Probab. Letters. **80** (2010), 1947-1953.

[] Bhattacharya, R and Lin, L. *Nonparametric benchmark analysis in risk assessment: a comparative study by simulation and data analysis.* Sankhya, Series B, **73** (2011), Issue 1, 144-163.

[] Bhattacharya, R and Waymire, E. C. *A Basic Course in Probability Theory.* Universitext, Springer (2007).

[] Bhattacharya, R and Patrangenaru, V. *A Course in Mathematical Statistics and Large Sample Theory.* Springer, Statistics Series. New York, USA (2013). To appear.

[] Bliss, C.I. *The calculations of dose-mortality curve (with an appendix by Fisher, R.A.).* Ann. Appl. Bio. **22**(1935), 134–167. (Table IV).

[] Brunk, H. D. *Maximum likelihood estimates of monotone parameters.* Ann. math. Statist., **26**(1955), 607-616.

[] Brunk, H. D. *On the estimation of parameters restricted by inequalities.* Ann. math. Statist., **29**(1958), 437–454.

[] Brunk, H. D. *Estimation of isotonic regression (with discussion by Ronald Pyke).* In M.L. Puri (Ed), Nonparametric Techniques in Statistical Inference, Cambridge Unviersity Press, 177–197 (1970).

[] Dette, H. and Scheder, R. *A finite sample comparison of nonparametric estimates of the effective dose in quantal bioassay.* J. Statist. Comput. Simulation. **80 (5)**(2011), 527–544.

[] Dette, H , Neumeyer, N. and Pliz, K.F. *A note on nonparametric estimation of the effective dose in quantal bioassay.* J.Amer.Statist.Assoc. **100**(2005), 503-510.

[] de Boor, C. *A Practical Guide to Splines* Springer-Verlag, 2001.

[] Efron, B. and Tibshirani, R.J. *An Introduction to the Bootstrap.* Chapman & Hall, London, 2003.

[] Eubank, R.L. *Nonparametric Regression and Spline Smoothing.* 2nd Edition. Marcel Dekker, New York, 1999.

[] Gasser, T. and Müller, H. G. *Estimating regression functions and their derivatives by the kernel method.* Scandinavian Journal of Statistics. **11** (1984), 171–185.

[] Grenander, U. *On the theory of mortality measurement. Part II..* Skand. Akt. **39** (1956), 125–153.

[] Hardy, G.H., Littlewood, G.E., and Pólya, G. *Inequality.* Cambridge University Press (1988).

[] Kelly, C, and Rice, J. *Monotone smoothing with application to dose response curves and the assessment of synergism.* Biometrics **46** (1990), 1071–1085.

[] Kong, M. *Nonparametric techniques in bioassay.* Dissertation (2004), Indiana University.

[] Kong, M. and Eubank, R.L. *Monotone smoothing with application to dose-response curve.* Comm. Statist. Simulation Comput. **35** (2006), no. 4, 991-1004,

[] Lee, E.T. *A computer program for linear logistic regression analysis.* Comput. Programs Biomed. **4** (1974), pp 80–92.

[] Loéve, M. *Probability Theory I (v. 1).* Graduate Text in Mathematics (1977).

[] Martin, J .T. *The problem of the evaluation of rotenone-containing plants. VI. The toxicity of l-elliptone and of poisons applied jointly, with further observations on the rotenone equivalent method of assessing the toxicity of derris root.* Ann. Appl.Biol., **30** (1940), 293–300.

[] Morgan, B.J.T. *Analysis of Quantal Response Data.* Monographs on Statistics and Applied Probability, vol. **46,** (1992).

[] Müller, H.G. and Schmitt, T. *Kernel and probit estimation in quantal bioassay.* J.Amer.Statist.Assoc. **83** (1988), pp. 750–759.

[] Nadaraya, E. A. *On Estimating Regression.* Theory of Probability and its Applications **9**(1)(1964), 141-142.

[] Nitcheva, D.K, Piegorsch, W.W, and West, R.W. *On use of the multistage dose-response model for assessing laboratory animal carcinogenicity.* Regul. Toxic. Pharmaco. **48**(2007), 135–147.

[] Park, D. and Park, S. *Parametric and nonparametric estimators of* $ED_{100\alpha}$. J. Statist. Comput. Simulation. **76**(2006), No. 8. 661–672

[] Piegorsch, W.W. and Bailer, A.J. *Analyzing Environmental Data.* John Wiley & Sons, 2005.

[] Prakasa Rao, B.L.S. *Estimation of a unimodal density.* Sankhya (A), **31**(1969), 23–36.

[] Robertson, T, Wright, F.T. and Dykstra, R.L. *Order Restricted Statistical Inference.* Wiley Series in Probability and Statistics (1988).

[] Serfling, R. J. *Approximation Theorems of Mathematical Statistics.* Wiley, New York (1980).

[] Silverman, B.W. *Some aspects of the spline smoothing approach to non-parametric regression curve fitting* Journal of the Royal Statistical Society B, **47** (1985), 1–52.

[] Stone, C.J. *Optimal rates of convergence for nonparametric estimators.* Ann. Statist. **8.**(1980), 1348–1360.

[] Tsybakov, A.B. *Introduction to Nonparametric Estimation.* Springer Series in Statistics (2010).

[] U.S.EPA. *Exposure Factors Handbook (Final Report).* U.S. Environmental Protection Agency, Washington, DC, EPA/600/P-95/002F a-c, 1997.

[] Utreras, F.I. *Smoothing noisy data under monotonicity constraints: existence, characterization and convergence rates.* Numer. Math. **74** (1985), 611–625.

[] van Eeden, C. *Testing and estimating ordered parameters of probability distributions.* Ph.D. Thesis, 1958.

[] Van Der Vaart, A.W. *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics (1998).

[] Wahba, G. *Spline Models for Observational Data.* CBMS-NSF Series , **59** (1990). SIAM, Philadelphia, PA.

[] Watson, G.S. *Smooth regression analysis.* Sankhya, Series A, **26** (1964), 359–372.

[] Wright, F.T. *The asymptotic behavior of monotone regression estimates.* Annals of Statistics, Vol. **9**, 443–448 (1981).

[] Wright, F.T. *Monotone regression estimates for grouped observations.* Annals of Statistics, Vol. **10**, 278-286 (1982).