Published in final edited form as:
Stat Probab Lett. 2010 December 1; 80(23-24): 1947–1953. doi:10.1016/j.spl.2010.08.024.

An adaptive nonparametric method in benchmark analysis for bioassay and environmental studies

Rabi Bhattacharya[1,*] and **Lizhen Lin**
Department of Mathematics, The University of Arizona, Tucson, AZ 85721, USA

Abstract

We present a novel nonparametric method for bioassay and benchmark analysis in risk assessment, which averages isotonic MLEs based on disjoint subgroups of dosages. The asymptotic theory for the methodology is derived, showing that the MISEs (mean integrated squared error) of the estimates of both the dose-response curve F and its inverse F^{-1} achieve the optimal rate $O(N^{-4/5})$. Also, we compute the asymptotic distribution of the estimate $\tilde{\zeta}_p$ of the effective dosage $\zeta_p = F^{-1}(p)$ which is shown to have an optimally small asymptotic variance.

Keywords

Monotone dose-response curve estimation; effective dosage; benchmark analysis; mean integrated square error; asymptotic normality

1. Introduction

The efficient estimation of effective dosage is an old but still very important problem in biology and medicine. In addition, concerns about the impact of pollutants in the environment have added a great sense of urgency to the development of good methods for the estimation of benchmarks in risk assessment (See, e.g., Piegorsch and Bailer (2005)). We present in this article the asymptotic theory of a new method. In a companion study based on extensive simulation and data analysis, to be presented elsewhere, it is shown that the method performs remarkably well even with small and moderate sample sizes (Bhattacharya and Lin (2010)).

Consider quantal dose-response experiments in bioassay where the response of a subject to a drug or a chemical agent is measured in a binary scale, 1 for response and 0 for non-response. Given a dosage x of the substance, let $F(x)$ be the probability of response. The function $x \mapsto F(x)$ is called the *dose-response curve*, and it is assumed to be *monotone increasing*. The *effective dosage* for a targeted response (probability) p is defined as the 'p-th quantile' ζ_p or ED_p,

$$\zeta_p = ED_p = F^{-1}(p), \quad 0 \le p \le 1; \quad F^{-1}(p) := \inf\{x : F(x) \ge p\}. \tag{1.1}$$

For the data, suppose that n_i subjects are given a dosage x_i ($i = 1, \ldots, m$), where $x_1 < \ldots < x_m$, with the total number of observations $N = \sum_{1 \le i \le m} n_i$. One may assume, without loss of

* Corresponding author rabi@math.arizona.edu (Rabi Bhattacharya), lizhen@math.arizona.edu (Lizhen Lin).
1 This research is supported by the NIH grant R21-ES016791 and NSF grant DMS 0806011.

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 the estimation of $F(x_i)$, $1 \le i \le m$, is

$$L(p_1, \ldots, p_m) = \prod_{i=1}^{m} p_i^{r_i} (1 - p_i)^{n_i - r_i} \quad (0 \le p_1 \le \ldots \le p_m \le 1); \quad [p_i := F(x_i)]. \tag{1.2}$$

The maximum likelihood estimator (MLE) of $(p_1, \ldots, p_m)$, under the monotonicity constraint, is given in Ayer et al.(1955) by the following *PAV*, or *pool-adjacent-violators algorithm* (Also see Barlow et al.(1972), p.73, and Cran (1980)):

$$\tilde{p}_i = \max_{0 \le u < ii \le v < m} \min \frac{\sum_{j=u}^{v} r_j}{\sum_{j=u}^{v} n_j} \qquad (1 \le i \le m). \tag{1.3}$$

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

$$\tilde{F}(x) = \begin{cases} \tilde{p}_i & \text{if} \quad x = x_i \\ \tilde{p}_i + \frac{\tilde{p}_{i+1} - \tilde{p}_i}{x_{i+1} - x_i}(x - x_i) & \text{if} \quad x_i < x \le x_{i+1}. \end{cases}$$

$\tilde{F}$ is a continuous function whose inverse is the estimate of ED_p as given by:

$$\widetilde{ED}_p = \begin{cases} x_1 & \text{if} \quad p \le \tilde{p}_1 \\ x_i + \frac{p - \tilde{p}_i}{\tilde{p}_{i+1} - \tilde{p}_i}(x_{i+1} - x_i) & \text{if} \quad \tilde{p}_i < p \le \tilde{p}_{i+1} \quad \text{for some i} \\ x_m & \text{if} \quad p > \tilde{p}_m, \end{cases} \tag{1.4}$$

if $\tilde{p}_{i+1} > \tilde{p}_i$ and, more generally, by $\tilde{F}^{-1}(p) = inf\left\{x : \tilde{F}(x) \ge p\right\}$.

From now on, we will assume, for simplicity, that there are m equidistant dosages and the same number n of i.i.d. $0 - 1$ valued observations at each dosage. Assume $n \to \infty$, $m \to \infty$ and

$$m = rs(n), \quad \text{with} \quad r \ge 1, \quad s(n) \quad \text{integers}, \tag{1.5}$$

$r \asymp \left(m^4/n\right)^{1/5}$ in Theorems 2.1, 2.2, 2.3 part (b), and $r \asymp \left(m^4/n\right)^{1/5} / (\log\log \quad n)^{6/5}$ in Theorem 2.3, 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 $\hat{p}_i$ 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:

$$[\text{Group}\quad 1] : (x_1, \widehat{p}_1), (x_{r+1}, \widehat{p}_{r+1}), (x_{2r+1}, \widehat{p}_{2r+1}), \quad \ldots, \quad (x_m, \widehat{p}_m);$$
$$[\text{Group}\quad 2] : (x_1, \widehat{p}_1), (x_2, \widehat{p}_2), (x_{r+2}, \widehat{p}_{r+2}), (x_{2r+2}, \widehat{p}_{2r+2}), \quad \ldots, \quad (x_m, \widehat{p}_m);$$
$$\cdots\cdots$$
$$[\text{Group}\quad r] : (x_1, \widehat{p}_1), (x_r, \widehat{p}_r), (x_{r+r}, \widehat{p}_{r+r}), (x_{2r+r}, \widehat{p}_{2r+r}), \quad \ldots, \quad (x_m, \widehat{p}_m). \tag{1.6}$$

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 $\left[\widetilde{F}(0), \widetilde{F}(1)\right]$. Compute

$$\widetilde{F} := (1/r) \sum_{1 \le j \le r} \widetilde{F}_j, \quad \widetilde{\zeta}_p := (1/r) \sum_{1 \le j \le r} \widetilde{\zeta_{p,j}}, \tag{1.7}$$

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

Among kernel based nonparametric methods for quantal bioassay, one may mention Müller and Schmitt (1988), Park and Park (2006), Dette et al. (2005) and Dette and Scheder (2010). A description of these methods may be found in the last two articles.

Remark 1.1. For the purpose of asymptotics, one may take the r groups in (1.6) 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, $\widehat{p}_i < \widehat{p}_{i+1} \quad \forall i$. Given $x \in (0, 1)$, if m, n are sufficiently large, and $m/r = o\left(\sqrt{n/\log \quad n}\right)$ (see (2.2)), x belongs to the domain of $\widetilde{F}_j \quad \forall j$, 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 (1.6), so that each has domain [0, 1].

We now provide a summary of the rest of the article. The asymptotic theory of the NAM is derived in Section 2. Theorem 2.1 proves that the estimate of the dose-response curve has a MISE attaining the optimal rate $O(N^{-4/5})$ under the assumptions that $f = F'$ is strictly positive, F'' is bounded and $m = o(n^{3/2}/(\log n)^{5/2})$. Theorem 2.2 provides the same optimal MISE rate for the estimate $p \to \widetilde{\zeta}_p$ of the quantile curve of interest, under the additional restriction $m/n^{2/3} \nrightarrow \infty$. Theorem 2.3 shows that ζ_p is asymptotically Normal around $E\widetilde{\zeta}_p$ with an asymptotic variance $O\left(N^{-4/5} \sqrt{\log \log \quad N}\right)$, under the same broad assumptions as in Theorem 2.1. However, for asymptotic Normality of $\widetilde{\zeta}_p$ around ζ_p, one needs the restriction $m = o(n^{2/3})$. For larger m, a bias correction of $\widetilde{\zeta}_p$ is thus called for. It will be shown in a companion paper (Bhattacharya and Lin (2010)), by extensive simulation and data analysis, that the method proposed here performs quite favorably in comparison with other leading nonparametric methods, including the new method due to Dette et al. (2005) and Dette and Scheder (2010)

2. Asymptotic Behavior

Let $\widehat{p}_i = r_i/n_i$ denote the sample proportion of responses to the dosage x_i ($i = 1, \ldots, m$). For simplicity, we assume in this section 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.

Theorem 2.1. *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.*

(a) The mean integrated squared error (MISE) of $\widetilde{F}$ *has the asymptotically optimal rate* $O(N^{-4/5})$ *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 also the MISE of* $\widetilde{F}$ *is* $O(N^{-4/5})$, *with a choice of r satisfying* $r \asymp \left(m^4/n\right)^{1/5}$.

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 $n > 1$,

$$P\left(|\widehat{p}_i - p_i| > c\sqrt{\log \quad n/n} \quad \text{for some i,} \quad i=1,\ldots,m\right) \leq c' N^{-2}. \tag{2.1}$$

It follows that if

$$m < (\theta/2c)\sqrt{n/\log \quad n}, \tag{2.2}$$

then

$$P\left(\widehat{p}_i \neq \widetilde{p}_i \quad \text{for some i,} \quad i=1,\ldots,m\right) \leq c'' N^{-2} \tag{2.3}$$

for some $c'' > 0$. Let B_n denote the union of the two sets within parentheses in (2.1) and (2.3). It is shown in Bhattacharya and Kong (2007), and simple to check using (2.1) to (2.3), that, on B_n^c, $\widehat{p}_i < \widehat{p}_{i+1} \quad \forall i$.

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

$$\begin{aligned}\widetilde{F}(x) &= \left(\frac{x_{i+1}-x}{x_{i+1}-x_i}\widehat{p}_i + \frac{x-x_i}{x_{i+1}-x_i}\widehat{p}_{i+1}\right)1_{B_n^c} + \widetilde{F}(x)\,1_{B_n} \\ &= \frac{x_{i+1}-x}{x_{i+1}-x_i}\widehat{p}_i + \frac{x-x_i}{x_{i+1}-x_i}\widehat{p}_{i+1} + \varepsilon_{n,1} \\ &= \widehat{p}_i + \frac{x-x_i}{x_{i+1}-x_i}\left(\widehat{p}_{i+1} - \widehat{p}_i\right) + \varepsilon_{n,1}\left(|\varepsilon_{n,1}| \leq 2 1_{B_n} = O_p\left(N^{-2}\right)\right).\end{aligned} \tag{2.4}$$

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

$$F(x) = F(x_i) + (x - x_i)F'(x^*) = p_i + (x - x_i)\left[\frac{F(x_{i+1}) - F(x_i)}{x_{i+1} - x_i} + \varepsilon(x)\right], |\varepsilon(x)| \leq M/m, \tag{2.5}$$

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

$$E\widetilde{F}(x) = \frac{x_{i+1} - x}{x_{i+1} - x_i} p_i + \frac{x - x_i}{x_{i+1} - x_i} p_{i+1} + O\left(N^{-2}\right) = p_i + \frac{x - x_i}{x_{i+1} - x_i}(p_{i+1} - p_i) + O\left(N^{-2}\right), \tag{2.6}$$

and

$$E\widetilde{F}(x) - F(x) = -(x - x_i)\varepsilon(x) + O\left(N^{-2}\right) = O\left(1/m^2\right). \tag{2.7}$$

From (2.4) and (2.5),

$$E\widetilde{F}(x) - F(x) = -(x - x_i)\varepsilon(x) + O\left(N^{-2}\right) = O\left(1/m^2\right). \tag{2.8}$$

and, by subtracting (2.7) from (2.8) one gets

$$\begin{aligned}\widetilde{F}(x) - E\widetilde{F}(x) &= \widehat{p}_i - p_i + \left[\widehat{p}_{i+1} - \widehat{p}_i - (p_{i+1} - p_i)\right]\tfrac{x - x_i}{x_{i+1} - x_i} + \varepsilon_{n,1} \\ &= \tfrac{x_{i+1} - x}{x_{i+1} - x_i}(\widehat{p}_i - p_i) + \tfrac{x - x_i}{x_{i+1} - x_i}(\widehat{p}_{i+1} - p_{i+1}) + \varepsilon_{n,1}.\end{aligned} \tag{2.9}$$

Hence

$$Var\left(\widetilde{F}(x)\right) = \left(\frac{x_{i+1} - x}{x_{i+1} - x_i}\right)^2 \frac{p_i(1 - p_i)}{n} + \left(\frac{x - x_i}{x_{i+1} - x_i}\right)^2 \frac{p_{i+1}(1 - p_{i+1})}{n} + O\left(N^{-2}\right) = O(1/n). \tag{2.10}$$

From (2.7) and (2.10) one obtains, on integration,

$$MISE\left(\widetilde{F}\right) = O(1/n) + O\left(1/m^4\right). \tag{2.11}$$

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

(b) First observe that the r groups in ((1.6)) 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 (2.2) 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 (2.7), (2.10))

$$MISE\left(\widetilde{F}\right) = MISE\left(1/r \sum_{1 \le j \le r} \widetilde{F}_j\right) = O(1/rn) + O\left(1/(m/r)^4\right). \tag{2.12}$$

The optimal choice of r is given by the relation $(m/r)^4 \asymp rn$ or, $r \asymp \left(m^4/n\right)^{1/5}$, yielding the optimal rate: $MISE\left(\widetilde{F}\right)=O\left((rn)^{-1}\right)=O\left(N^{-4/5}\right)$.

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

Theorem 2.2. *Assume the hypothesis of Theorem 2.1.*

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

(b) If $m/n^{1/4} \to \infty$, *but* $m/n^{2/3} \nrightarrow \infty$, *then* $MISE\left(\widetilde{\zeta}\right)=O\left(N^{-4/5}\right)$, with $r \asymp \left(m^4/n\right)^{1/5}$.

Proof. (a) For $m = O(n^{1/4})$, one may consider $r = 1$ in (1.7). 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,

$$F^{-1}(p) - \widetilde{F}^{-1}(p) = \widetilde{F}^{-1}(F(x)) - \widetilde{F}^{-1}(F(x)) . \tag{2.13}$$

First, consider, for an appropriate positive constant c_1,

$$p \in \left[p_i + c_1\sqrt{\log(n)/n}, p_{i+1} - c_1\sqrt{\log(n)/n}\right] = D_{n,i}, \tag{2.14}$$

say. Then on B_n^c, $F(x)$ and $\widetilde{F}(x)$ belong to $[\widehat{p}_i, \widehat{p}_{i+1}]$. Using (2.13), the linearity of $\widetilde{F}^{-1}$ on $[\widehat{p}_i, \widehat{p}_{i+1}]$ and (2.8), and writing

$$\delta_n=\widehat{p}_{i+1} - \widehat{p}_i - (p_{i+1} - p_i), \frac{1}{\widehat{p}_{i+1} - \widehat{p}_i} = \frac{1}{p_{i+1} - p_i}\left(1 - \frac{\delta_n}{\widehat{p}_{i+1} - \widehat{p}_i}\right) \tag{2.15}$$

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

$$\begin{aligned} F^{-1}(p) - \widetilde{F}^{-1}(p) &= \left[\widetilde{F}(x) - F(x)\right] \tfrac{x_{i+1}-x_i}{\widehat{p}_{i+1}-\widehat{p}_i} 1_{B_n^c} + \varepsilon_{n,2} \left(|\varepsilon_{n,2}| \le 1_{B_n} = O_p\left(N^{-2}\right)\right) \\ &= \left\{(\widehat{p}_i - p_i)\tfrac{x_{i+1}-x_i}{\widehat{p}_{i+1}-\widehat{p}_i} + \tfrac{x-x_i}{\widehat{p}_{i+1}-\widehat{p}_i}\delta_n - \varepsilon(x)(x-x_i)\tfrac{x_{i+1}-x_i}{\widehat{p}_{i+1}-\widehat{p}_i}\right\} 1_{B_n^c} + \varepsilon_{n,2} \\ &= \left\{(\widehat{p}_i - p_i)\tfrac{x_{i+1}-x_i}{p_{i+1}-p_i}\left(1 - \tfrac{\delta_n}{\widehat{p}_{i+1}-\widehat{p}_i}\right) + \tfrac{(x-x_i)\delta_n}{p_{i+1}-p_i}\left(1 - \tfrac{\delta_n}{\widehat{p}_{i+1}-\widehat{p}_i}\right) - \tfrac{\varepsilon(x)(x-x_i)(x_{i+1}-x_i)}{p_{i+1}-p_i}\left(1 - \tfrac{\delta_n}{\widehat{p}_{i+1}-\widehat{p}_i}\right)\right\} 1_{B_n^c} + \varepsilon_{n,2}. \\ &= \left\{(\widehat{p}_i - p_i)\tfrac{x_{i+1}-x_i}{p_{i+1}-p_i} + \tfrac{(x-x_i)\delta_n}{p_{i+1}-p_i} - \tfrac{\varepsilon(x)(x-x_i)(x_{i+1}-x_i)}{p_{i+1}-p_i}\right\} 1_{B_n^c} + \varepsilon_{n,3} + \varepsilon_{n,2}. \end{aligned} \tag{2.16}$$

Here

$$\varepsilon_{n,3} = -\left\{(\widehat{p}_i - p_i)\left(\frac{x_{i+1}-x_i}{p_{i+1}-p_i}\right)\frac{\delta_n}{\widehat{p}_{i+1}-\widehat{p}_i} + \left(\frac{x-x_i}{p_{i+1}-p_i}\right)\frac{\delta_n^2}{\widehat{p}_{i+1}-\widehat{p}_i} + \frac{\varepsilon(x)(x-x_i)(x_{i+1}-x_i)}{p_{i+1}-p_i}\left(\frac{-\delta_n}{\widehat{p}_{i+1}-\widehat{p}_i}\right)\right\} 1_{B_n^c}. \tag{2.17}$$

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

The expectation of (2.16) equals

$$F^{-1}(p) - E\widetilde{F}^{-1}(p) = -\frac{\varepsilon(x)(x-x_i)(x_{i+1}-x_i)}{p_{i+1}-p_i} + E\varepsilon_{n,3} + O\left(N^{-2}\right). \tag{2.18}$$

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

$$\begin{aligned} -E\left(\frac{(\widehat{p}_i-p_i)(x_{i+1}-x_i)\delta_n}{(p_{i+1}-p_i)(\widehat{p}_{i+1}-\widehat{p}_i)}1_{B_n^c}\right) &= -E\left((\widehat{p}_i-p_i)\frac{(x_{i+1}-x_i)\delta_n}{(p_{i+1}-p_i)^2}\left(1-\frac{\delta_n}{\widehat{p}_{i+1}-\widehat{p}_i}\right)1_{B_n^c}\right) \\ &= \frac{p_i(1-p_i)}{n}\frac{x_{i+1}-x_i}{(p_{i+1}-p_i)^2} + O\left(m^2/n^{3/2}\right) + O\left(N^{-2}\right) \\ &= O(m/n); \end{aligned} \tag{2.19}$$

$$\begin{aligned} &-E\left(\frac{(x-x_i)}{(p_{i+1}-p_i)}\frac{\delta_n^2}{(\widehat{p}_{i+1}-\widehat{p}_i)}1_{B_n^c}\right) = -\frac{(x-x_i)}{p_{i+1}-p_i}E\left(\frac{\delta_n^2}{\widehat{p}_{i+1}-\widehat{p}_i}1_{B_n^c}\right) \\ &= -\frac{(x-x_i)}{(p_{i+1}-p_i)^2}E\delta_n^2 + \frac{(x-x_i)}{(p_{i+1}-p_i)^2}E\frac{\delta_n^3}{\widehat{p}_{i+1}-\widehat{p}_i}1_{B_n^c} + O\left(N^{-2}\right) \\ &= -\frac{(x-x_i)}{(p_{i+1}-p_i)^2}\frac{\{p_i(1-p_i)+p_{i+1}(1-p_{i+1})\}}{n} + O\left(m^2/n^{3/2}\right); \end{aligned} \tag{2.20}$$

$$E\left(\frac{\varepsilon(x)(x-x_i)(x_{i+1}-x_i)}{p_{i+1}-p_i}\left(\frac{\delta_n}{\widehat{p}_{i+1}-\widehat{p}_i}\right)1_{B_n^c}\right) = \frac{\varepsilon(x)(x-x_i)(x_{i+1}-x_i)}{p_{i+1}-p_i}E\left\{\frac{\delta_n^2}{(p_{i+1}-p_i)(\widehat{p}_{i+1}-\widehat{p}_i)}\right\}1_{B_n^c} = O(1/n). \tag{2.21}$$

For the first relation in (2.21), use $\frac{\delta_n}{\widehat{p}_{i+1}-\widehat{p}_i} = \delta_n\left(1-\frac{\delta_n}{\widehat{p}_{i+1}-\widehat{p}_i}\right)/(p_{i+1}-p_i)$, and $E\delta_n = 0$. Hence the bias (of $\widetilde{F}^{-1}(p)$ as an estimator of $F^{-1}(p)$) is

$$\left|Bias\left(\widetilde{F}^{-1}(p)\right)\right| = O(m/n) + O\left(1/m^2\right). \tag{2.22}$$

Subtracting (2.18) from (2.16), one obtains

$$E\widetilde{F}^{-1}(p) - \widetilde{F}^{-1}(p) = \left\{(\widehat{p}_i-p_i)\frac{x_{i+1}-x}{p_{i+1}-p_i} + (\widehat{p}_{i+1}-p_{i+1})\frac{x-x_i}{p_{i+1}-p_i}\right\}1_{B_n^c} + O(m/n) + O_p\left(N^{-2}\right), \tag{2.23}$$

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

$$Var\left(\widetilde{F}^{-1}(p)\right) = O(1/n) + O\left(m^2/n^2\right). \tag{2.24}$$

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

$$F^{-1}(p) - \widetilde{F}^{-1}(p) = \left[\widetilde{F}(x) - F(x)\right]\frac{x_{i+1}-x_i}{\widehat{p}_{i+1}-\widehat{p}_i}\left(1+\varepsilon_{n,4}\right), \tag{2.25}$$

where $\varepsilon_{n,4} = O_p(1/m)$. Indeed, $|\varepsilon_{n,4}| \le c^{\nu}/m$ on B_n^c, for some $c^{\nu} > 0$. To establish (2.25), note first that if $p \in D_{n,i}^c \cap [p_i, p_{i+1}]$ then, although $\tilde{F}(x) \in [\hat{p}_i, \hat{p}_{i+1}]$ (since $x \in [x_i, x_{i+1}]$), it may happen that $F(x)$ belongs to $(\hat{p}_{i-1}, \hat{p}_i)$ or $(\hat{p}_{i+1}, \hat{p}_{i+2})$. On B_n^c, there is no other possibility.

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

$$\begin{aligned} F^{-1}(p) - \tilde{F}^{-1}(p) &\equiv \tilde{F}^{-1}\left(\tilde{F}(x)\right) - \tilde{F}^{-1}(F(x)) \\ &= \tilde{F}^{-1}\left(\tilde{F}(x)\right) - \tilde{F}^{-1}\left(\tilde{F}(x_i)\right) + \tilde{F}^{-1}\left(\tilde{F}(x_i)\right) - \tilde{F}^{-1}(F(x)) \\ &= \left(\tilde{F}(x) - \tilde{F}(x_i)\right)\left\{\frac{x_{i+1}-x_i}{\hat{p}_{i+1}-\hat{p}_i}\right\} + \left(\tilde{F}(x_i) - F(x)\right)\left\{\frac{x_i-x_{i-1}}{\hat{p}_i-\hat{p}_{i-1}}\right\}, \end{aligned} \tag{2.26}$$

in view of the linearity of $\tilde{F}^{-1}$ on both $[\hat{p}_i, \hat{p}_{i+1}]$ and $[\hat{p}_{i-1}, \hat{p}_i]$, 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^{ν}/m on B_n^c. The MISE of $\tilde{F}^{-1}$ is then given by

$$MISE\left(\tilde{F}^{-1}\right) = O\left(m^2/n^2\right) + O\left(1/m^4\right) + O(1/n). \tag{2.27}$$

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

$$MISE\left(\tilde{F}^{-1}\right) = O(1/n) = O\left(N^{-4/5}\right). \tag{2.28}$$

(b) Next consider the case $m/n^{1/4} \to \infty$, i.e., $n = o(N^{4/5})$. Since $MISE\left(\tilde{F}^{-1}\right) = O(1/n)$, it is of larger order than $N^{-4/5}$, and hence the estimation is suboptimal. In this case, again consider r groups of essentially disjoint equidistant dosages. Then the average $\tilde{\zeta}_p = \frac{1}{r}\sum_{j=1}^{r} \tilde{\zeta}_{p,j}$ has bias and variance (See (2.22) and (2.24)) given by

$$Bias\left(\tilde{\zeta}_p\right) = O(m/(rn)) + O\left((r/m)^2\right), \tag{2.29}$$

$$Var\left(\tilde{\zeta}_p\right) = O(1/(rn)) + O\left(m^2/\left(r^3 n^2\right)\right). \tag{2.30}$$

Assume that m is not very large, i.e., $\frac{m}{n^{2/3}} \nrightarrow \infty$. Then the optimal choice of r is $r \asymp \left(m^4/n\right)^{1/5}$, since the term $m/(rn)$ in (2.29) 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 $\tilde{\zeta}_p$, namely, $MISE\left(\tilde{\zeta}_p\right) = O(1/(rn)) = O\left((mn)^{-4/5}\right) = O\left(N^{-4/5}\right)$. □

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

Theorem 2.3. *Let* $p \in (0, 1)$. In addition to the hypothesis of Theorem 2.1, assume $m/n^{1/4} \to \infty$. *Then the following hold.*

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

$$\frac{\sqrt{n}\left(\tilde{\zeta}_p-\zeta_p\right)}{\delta(p)} \xrightarrow{\mathcal{L}} N\left(0, \frac{p(1-p)}{f^2(p)}\right), \tag{2.31}$$

where

$$\delta^2(p) \equiv \sum_{i=1}^{m-1}(x_{i+1}-x_i)^{-2}\left\{(x_{i+1}-x)^2+(x-x_i)^2\right\}1_{I_i}(p)$$
$$(I_i=[p_i, p_{i+1}) \quad for \quad 1 \le i \le m-2, I_{m-1}=[p_{m-1}, p_m]), \tag{2.32}$$

lies in [1/2,1], and $x = F^{-1}(p) = \zeta_p$.

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

$$\frac{\sqrt{rn}\left(\tilde{\zeta}_p-E\tilde{\zeta}_p\right)}{\bar{\delta}(p)} \xrightarrow{\mathcal{L}} N\left(0, \frac{p(1-p)}{f^2(p)}\right) \tag{2.33}$$

Here $\bar{\delta}^2(p)$ *is the average of the r quantities* $\delta_j^2(p)$, $1 \le j \le$ r, of the form (2.32), 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* $r \asymp \left(m^4/n\right)^{1/5}/(\log\log n)^{6/5}$,

$$\frac{\sqrt{rn}\left(\tilde{\zeta}_p-\zeta_p\right)}{\bar{\delta}(p)} \xrightarrow{\mathcal{L}} N\left(0, \frac{p(1-p)}{f^2(p)}\right). \tag{2.34}$$

Proof. (a) It follows from (2.16) (and (2.25)) that for $p \in [p_i, p_{i+1})$ one has, outside B_n,

$$\tilde{F}^{-1}(p)-F^{-1}(p)=-\left\{(\hat{p}_i-p_i)\frac{x_{i+1}-x}{p_{i+1}-p_i}+(\hat{p}_{i+1}-p_{i+1})\frac{x-x_i}{p_{i+1}-p_i}\right\}+O(m/n)+O\left(1/m^2\right) \tag{2.35}$$

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 (2.23), one has, outside B_n,

$$\tilde{F}^{-1}(p)-E\tilde{F}^{-1}(p)=-\left\{(\hat{p}_i-p_i)\frac{x_{i+1}-x}{p_{i+1}-p_i}+(\hat{p}_{i+1}-p_{i+1})\frac{x-x_i}{p_{i+1}-p_i}\right\}+O(m/n). \tag{2.36}$$

Using the analog of (2.36) for $\tilde{F}_j^{-1}(p)-E\tilde{F}_j^{-1}(p)$, one may apply Lyapunov's central limit theorem (See, e.g., Bhattacharya and Waymire (2007), p.103) to the *r* summands $\sqrt{n}\left(\tilde{F}_j^{-1}(p)-E\tilde{F}_j^{-1}(p)\right)$), $1 \le j \le r$, 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 $\sqrt{n}\frac{m/r}{n}=m/\left(r\sqrt{n}\right) \asymp m^{1/5}n^{1/5}/\sqrt{n}=m^{1/5}/n^{3/10} \to 0$ as $m = o(n^{3/2}/\log^{5/2} n)$, which also ensures that $m/r=o\left(\sqrt{n/\log n}\right)$ (See (2.2), (2.3)).

(c) One has (See (2.22))

$$\begin{aligned} Bias\left(\tilde{\zeta}_p\right) &= O\left(m/(rn)\right) + O\left(r^2/m^2\right), \\ \sqrt{rn}\,Bias\left(\tilde{\zeta}_p\right) &= O\left(m/\sqrt{rn}\right) + O\left(r^{5/2}\sqrt{n}/m^2\right) \to 0, \end{aligned} \quad (2.37)$$

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 $\tilde{\zeta}_p$, (2.34) follows from (2.33).

Remark 2.1. Note that (2.33) implies that, with $r \asymp \left(m^4/n\right)^{1/5}$, the asymptotic variance of $\tilde{\zeta}_p$ is $O(N^{-4/5})$.

Remark 2.2. Theorems 2.1-2.3 easily extend to the case of non-equal sample sizes n_i, $1 \le i \le m$, and non-equidistant dosages $x_1 < \ldots < x_m$, provided (1) 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 (2) 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.

Acknowledgments

The authors wish to thank Professor Walt Piegorsch for a careful reading of the paper and for helpful suggestions.

References

1. Ayer M, Brunk HD, Ewing GM, Reid WT, Silverman E. An empirical distribution function for sampling with incomplete information. Ann.Math.Statist 1955;26:641–647.
2. Barlow, RE.; Bartholomew, DJ.; Bremner, JM.; Brunk, HD. Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression. Wiley; London: 1972.
3. Bhattacharya R, Kong M. Consistency and asymptotic normality of the estimated effective dose in bioassay. Journal of Statistical Planning and Inference 2007;137:643–658.
4. Bhattacharya, R.; Lin, L. Nonparametric benchmark analysis in risk assessment: a comparative study by simulation and data analysis. 2010. preprint
5. Bhattacharya, R.; Waymire, EC. A Basic Course in Probability Theory. Springer; New York: 2007.
6. Cran GW. AS149 amalgamation of means in the case of simple ordering. Appl. Statist 1980;29(2): 209–211.
7. Dette H, Neumeyer N, Pliz KF. A note on nonparametric estimation of the effective dose in quantal bioassay. J.Amer.Statist.Assoc 2005;100:503–510.
8. Dette H, Scheder R. A finite sample comparison of nonparametric estimates of the effective dose in quantal bioassay. Journal of Statistical Computation and Simulation 2010;80(5):527–544.
9. Müller HG, Schmitt T. Kernel and probit estimation in quantal bioassay. J.Amer.Statist.Assoc 1988;83(403):750–759.
10. Park D, Park S. Parametric and nonparametric estimators of $ED_{100\alpha}$. Journal of Statistical Computation and Simulation 2006;76(8):661–672.
11. Piegorsch, WW.; Bailer, AJ. Analyzing Environmental Data. John Wiley & Sons; 2005.