A Random Set Approach to Confidence Regions with Applications to the Effective Dose with Combinations of Agents

Hanna Jankowski[1]*, Xiang Ji[1], and Larissa Stanberry[2]
[1]Department of Mathematics and Statistics, York University, Toronto, ON, Canada
[2]Seattle Children's Research Institute, Seattle, WA, U.S.A.

January 31, 2014

Abstract

The effective dose is the pharmaceutical dosage required to produce a therapeutic response in a fixed proportion of the patients. When only one drug is considered, the problem is a univariate one and has been well-studied. However, in the multidimensional setting, i.e., in the presence of combinations of agents, estimation of the effective dose becomes more difficult. This study is focused on the plug-in logistic regression estimator of the multidimensional effective dose. We discuss consistency of such estimators, and focus on the problem of simultaneous confidence regions. We develop a bootstrap algorithm to estimate confidence regions for the multidimensional effective dose. Through simulation, we show that the proposed method gives 95% confidence regions which have better empirical coverage than the previous method for moderate to large sample sizes. The novel approach is illustrated on a cytotoxicity study on the effect of two toxins in the leukaemia cell line HL-60 and a decompression sickness study of the effects of the duration and depth of the dive.

Keywords: Multidimensional effective dose; drug combinations; Logistic regression; Plug-in estimation; Simultaneous confidence region.

1 Introduction

In the analysis of biological assays, one is often interested in the covariate, or combination of covariates, which yields a specific response. For example, when investigating the efficacy of a drug, the effective dose (ED) is the dose or amount of drug required to produce a therapeutic response in a desired proportion of the population under study. When this response is binary, a logistic linear

*Corresponding author: hkj@yorku.ca

regression model is a popular choice. That is, let p denote the probability of a positive response, and let z denote the drug dosage. Then the model is

$$\log(p/(1-p)) \quad = \quad \beta_0 + \beta_1 z,$$

and the dosage associated with a 50% response rate (ED_{50}) is the point $-\beta_0/\beta_1$. When a combination of drugs and/or covariates is being studied, the model becomes

$$\log(p/(1-p)) \quad = \quad \beta_0 + \sum_{i=1}^{k} \beta_i z_i.$$

If only a single z_i, say z_1, represents a drug and the remaining z_i represent covariates such as age or weight, then the median effective dose is $\mathrm{ED}_{50} = -(\beta_1)^{-1}(\beta_0 + \sum_{i=2}^{k} \beta_i z_i)$ when the patient-specific covariates are held fixed. Alternatively, multiple z_i can represent drug levels, and one is interested in the combinations of drugs or other agents required to yield a specific response. Such instances arise, for example, in Skarin et al. [1] where lymphoma treatments are considered or in Lang et al. [2]. Another example, considered extensively in [3, 4, 5, 6], is that of analysing the risk of decompression sickness among deep sea divers based on both the duration and pressure of the dive. In this setting, when combinations of agents are considered, the median effective dose is the set

$$\mathrm{ED}_{50} = \left\{ (z_1, z_2, \ldots, z_k) : \beta_0 + \textstyle\sum_{i=1}^{k} \beta_i z_i = 0 \right\}.$$

Li et al. [3] refer to this as the *multidimensional* effective dose. In the single drug setting (with or without covariates), the quantity of interest is univariate, and therefore much easier to handle. However, for combinations of drugs, the effective dose is a set or hyperplane and is considerably more complicated to analyse.

In practice, estimation of the mean response is commonly based on a logistic regression approach. We therefore focus on the plug-in logistic regression estimator of the effective dose. We provide conditions for consistency of the estimator and we study *simultaneous* confidence regions, or supersets, for the true effective dose. That is, we seek a confidence set C such that $P(C \supset \mathrm{ED}_{50}) \geq 95\%$, say. In the univariate setting, the variability of the estimator can be quantified by the standard error. In multidimensional settings, a confidence region provides a clear, intuitive way of quantifying and visualizing the variability of the estimator.

The first study of this problem was done in Carter et al. [7]. There, large sample simultaneous confidence regions were derived by inverting Scheffé's bounds for simultaneous confidence intervals [8]. This approach, however, is very conservative [7]. Further study of this method was carried out in Li et al. [3], who consider the estimation of both (1) a conditional single-dimensional effective dose in the presence of covariates and (2) an unconditional multidimensional effective dose in the linear logistic model. The conditional single-dimensional effective dose was also used by Chen [9] in the analysis of a dose-time response model. The extensive simulations of Li et al. [3] show how conservative the Scheffé-inversion method is: Empirical coverage probabilities of 95% confidence regions vary from 98% to 100% in the two-dimensional linear model.

In this work, we introduce an empirical bootstrap method to compute a confidence region for the multidimensional effective dose. We study the empirical coverage of the confidence regions through simulations (see Section 4.3 as well as the supplemental material), and we show that the

new method has better performance for medium to large sample sizes. Our choice of designs for the simulations is similar to that of Li et al. [3], in particular to allow for direct comparison. As compared to Li et al. [3], however, we extend the study to the non-linear regression setting. Notably, the over coverage of the Carter et al. [7] method is only increased for these cases, and therefore this is an important situation to consider. The methodology is illustrated on both simulated (Section 4.2) and real data examples (Section 5).

To our best knowledge, the works studying multidimensional effective dose estimation focus on confidence regions and do not discuss consistency [7, 3, 6, 4]. Here, we fill this gap, by providing necessary conditions for consistency of the plug-in estimators $\widehat{\mathrm{ED}}_{100p}$ and $\widehat{\mathrm{ED}}^{+}_{100p}$. As these are rather technical, the statement appears in Section A.3 of the Appendix.

2 Notation and Assumptions

Let $\mathcal{D} \subset \mathbb{R}^d$ denote the covariate domain. Assume $\mathcal{D}$ to be bounded and closed. Let Y be the observed binary response and consider estimating the sets

$$\mathrm{ED}^{+}_{100p} = \{\boldsymbol{z} \in \mathcal{D} : E[Y|Z = \boldsymbol{z}] \geq p\} \qquad \text{and} \qquad \mathrm{ED}_{100p} = \{\boldsymbol{z} \in \mathcal{D} : E[Y|Z = \boldsymbol{z}] = p\}$$

where $\boldsymbol{z} = \{z_1, \ldots, z_d\}$ denotes the *multidimensional vector* of observed covariates. We assume that the data can be modelled as a logistic regression with

$$\log\left(\frac{E[Y|Z = \boldsymbol{z}]}{1 - E[Y|Z = \boldsymbol{z}]}\right) \;=\; \beta_0 x_0 + \beta_1 x_1 + \ldots + \beta_k x_k = \beta^T \boldsymbol{x}. \tag{2.1}$$

Here, we use the notation $\boldsymbol{x} = \{x_0, x_1, \ldots, x_k\}^T$ to denote the function of the covariates $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{z}) : \mathbb{R}^d \mapsto \mathbb{R}^{k+1}$, where $k + 1 \geq d$. This allows us to emphasize the difference between the covariates $\boldsymbol{z}$ and how they are used in the model via $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{z})$. As an example, suppose that we have two covariates, z_1 and z_2, and that the right-hand side of (2.1) is

$$\beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_2^2.$$

Then we have $d = 2$ and four parameters ($k + 1 = 4$). In this case, the function $\boldsymbol{x}$ of the covariates is given as $\boldsymbol{x}(\boldsymbol{z}) = \{1, z_1, z_2, z_2^2\}^T$. We assume $\boldsymbol{x}(\boldsymbol{z})$ to be continuous on the domain $\mathcal{D}$.

Next, let $\mathrm{logit}(p) = \log(p/(1-p))$. Since $\mathrm{logit}(p)$ is an increasing function of p and $E[Y|Z = \boldsymbol{z}] = \mathrm{logit}^{-1}(\beta^T \boldsymbol{x})$, we may re-write

$$\mathrm{ED}^{+}_{100p} = \left\{\boldsymbol{z} \in \mathcal{D} : \beta^T \boldsymbol{x} \geq \mathrm{logit}(p)\right\} \quad \text{and} \quad \mathrm{ED}_{100p} = \left\{\boldsymbol{z} \in \mathcal{D} : \beta^T \boldsymbol{x} = \mathrm{logit}(p)\right\}.$$

We estimate these using the plug-in estimators

$$\widehat{\mathrm{ED}}^{+}_{100p} = \left\{\boldsymbol{z} \in \mathcal{D} : \widehat{\beta}_n^T \boldsymbol{x} \geq \mathrm{logit}(p)\right\} \quad \text{and} \quad \widehat{\mathrm{ED}}_{100p} = \left\{\boldsymbol{z} \in \mathcal{D} : \widehat{\beta}_n^T \boldsymbol{x} = \mathrm{logit}(p)\right\},$$

where $\widehat{\beta}_n$ is the maximum likelihood estimator of β.

Let $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ denote the observed covariates and let $\boldsymbol{x}_j = \boldsymbol{x}(\boldsymbol{z}_j), j = 1, \ldots, n$. Let X denote the design matrix, which is of size $n \times (k+1)$. With this notation, the maximum likelihood estimator $\widehat{\beta}_n$ is asymptotically normal with mean β and variance Σ/n, consistently estimated by $\widehat{\Sigma}_n/n$, where

$$\widehat{\Sigma}_n = n\left(X^T \mathrm{diag}\{\widehat{p}_1(1-\widehat{p}_1), \ldots, \widehat{p}_n(1-\widehat{p}_n)\}X\right)^{-1}, \tag{2.2}$$

with $\widehat{p}_j = \text{logit}^{-1}(\widehat{\beta}_n^T \boldsymbol{z}_j)$, for $j = 1, \ldots, n$.

One may also be interested in sets of the form

$$\begin{aligned} \text{ED}^{-}_{100p} &= \{\boldsymbol{z} \in \mathcal{D} : E[Y|Z = \boldsymbol{z}] \leq p\} \\ &= \{\boldsymbol{z} \in \mathcal{D} : E[1 - Y|Z = \boldsymbol{z}] \geq 1 - p\}. \end{aligned}$$

In this case, the methodology would be similar to ED^{+}_{100p} (by considering failures instead of successes as indicated above), and we therefore focus on ED^{+}_{100p} and ED_{100p} in what follows.

3 Calculating confidence regions

3.1 Scheffé's method

The existing methodology for simultaneous confidence regions is based on Scheffé's uniform contrast bounds. As shown in Carter et al. [7], $\sup_{\boldsymbol{x}} n[\boldsymbol{x}^T(\widehat{\beta}_n - \beta)]^2/\boldsymbol{x}^T\Sigma\boldsymbol{x} \leq Y_n$, where Y_n converges to Y, a $\chi^2(k+1)$ random variable. Let q_α denote the upper quantile of Y, a value such that $P(Y > q_\alpha) = \alpha$. The asymptotic confidence region for ED_{100p} is defined as

$$\begin{aligned} &\text{CR}_{\text{S},1-\alpha} \\ &= \left\{\boldsymbol{z} : n(\widehat{\beta}_n^T\boldsymbol{x} - \text{logit}(p))^2 \leq q_\alpha\, \boldsymbol{x}^T\Sigma\boldsymbol{x}\right\} \\ &= \left\{\boldsymbol{z} : \widehat{\beta}_n^T\boldsymbol{x} \geq \text{logit}(p) - \sqrt{q_\alpha\, \boldsymbol{x}^T\Sigma\boldsymbol{x}/n}\right\} \cap \left\{\boldsymbol{z} : \widehat{\beta}_n^T\boldsymbol{x} \leq \text{logit}(p) + \sqrt{q_\alpha\, \boldsymbol{x}^T\Sigma\boldsymbol{x}/n}\right\}. \quad (3.3) \end{aligned}$$

The asymptotic confidence region for ED^{+}_{100p} is then defined as

$$\text{CR}^{+}_{\text{S},1-\alpha} = \left\{\boldsymbol{z} : \widehat{\beta}_n^T\boldsymbol{x} \geq \text{logit}(p) - \sqrt{q_{2\alpha}\, \boldsymbol{x}^T\Sigma\boldsymbol{x}/n}\right\}.$$

Note that taking $q_{2\alpha}$ (instead of q_α) accounts for the fact the we only take one side of the square root. In practice, the covariance matrix Σ can be estimated as in (2.2). It is possible that the confidence region as defined above is empty, but this occurs rarely for larger sample sizes. As shown in Carter et al. [7], the probability that $\text{CR}^{+}_{\text{S},1-\alpha}$ covers ED^{+}_{100p} is at least $100(1-\alpha)\%$. As we show in the simulations which follow, this method (although asymptotically correct) is quite conservative, resulting in frequent and extensive over-coverage.

3.2 New algorithm

The novel approach to calculating a simultaneous confidence region introduced here is a bootstrap procedure. Our original motivation for this approach came from the setting $d = 1$, where we found that a simple bootstrap approach had the best overall performance. Indeed, when $d = 1$, our method is equivalent to univariate bootstrap methods. The new approach also has some appealing properties, and these are summarized in Proposition A.2 in the Appendix. In particular, the empirical quantile is invariant under rotations and equivariant under re-scaling.

Below, we describe the empirical parametric bootstrap algorithm to calculate the confidence region $\text{CR}^{+}_{1-\alpha}$ for ED^{+}_{100p}. A nonparametric approach may also be taken, although we do not give the details here.

Bootstrap-based confidence regions

Suppose that we have estimated $\widehat{\beta}_n$ and the variance $\widehat{\Sigma}_n$ in the logistic regression model based on an available data set with sample size n.

1. Generate $\widehat{\beta}^*_{n,i} \sim N(\widehat{\beta}_n, \widehat{\Sigma}_n/n)$ independent random variables for $i = 1, \ldots, B$.

2. For each $\widehat{\beta}^*_{n,i}$ calculate the effective dose estimate

$$\widehat{\mathrm{ED}}^{+*}_{100p,i} \;=\; \left\{ \boldsymbol{z} \in \mathcal{D} : \widehat{\beta}^{*T}_{n,i} \boldsymbol{x} \geq \mathrm{logit}(p) \right\}.$$

3. Compute

$$K = \bigcap_{i=1}^{B} \widehat{\mathrm{ED}}^{+*}_{100p,i},$$

and then calculate $\xi_i = \rho(\widehat{\mathrm{ED}}^{+*}_{100p,i}, K)$. Here, ρ denotes the Hausdorff distance on sets. A precise definition is given in the Appendix, see (A - 6).

4. Calculate γ_α, the $100(1-\alpha)$ sample percentile of the ξ_i observations. Finally, compute

$$\mathrm{CR}^{+}_{1-\alpha} \;=\; \bigcup_{i:\, \xi_i < \gamma_\alpha} \widehat{\mathrm{ED}}^{+*}_{100p,i}.$$

In the algorithm, if $K = \emptyset$, then the resulting confidence region is the empty set, and therefore uninformative. In practice, this becomes an issue for smaller sample sizes. In most cases, this happens because one of the sets $\widehat{\mathrm{ED}}^{+*}_{100p,i}$ is empty. We considered several solutions, and the optimal of these was to simply remove all empty $\widehat{\mathrm{ED}}^{+*}_{100p,i}$ from the algorithm apriori. This is done by appropriately modifying step two above.

The proportion of sets $\widehat{\mathrm{ED}}^{+*}_{100p,1}, \ldots, \widehat{\mathrm{ED}}^{+*}_{100p,B}$ which is contained in $\mathrm{CR}^{+}_{1-\alpha}$ is at least $\lfloor (1-\alpha)\, B \rfloor / B$. The confidence region thus creates a covering set which contains a desirable proportion of the observed bootstrapped samples. This assertion is formally stated in the Appendix (see Proposition A.2). We also prove that the bootstrap confidence region is invariant under rotations and satisfies a natural scaling property.

Finally, we note that to compute the $100(1-\alpha)\%$ confidence region for ED_{100p} we take

$$\mathrm{CR}_{1-\alpha} \;=\; \mathrm{CR}^{+}_{1-\alpha/2} \cap \mathrm{CR}^{-}_{1-\alpha/2},$$

where the algorithm for finding $\mathrm{CR}^{-}_{1-\alpha/2}$ is given in the Appendix. This "intersection" approach is similar to the standard CI for the mean, in a sense that the latter can be written as an intersection of the unbounded intervals with endpoints symmetric about the sample mean. Equation (3.3) shows that the same is also true for the confidence regions calculated as proposed in Carter et al. [7].

4 Simulation studies

4.1 Computation

It was noted in Li et al. [3, page 113] that their conditional univariate approach is simpler to implement than a multivariate approach ($\mathrm{CR}_{\mathrm{S},1-\alpha}$ in this paper). However, we do not find this to be the case, and $\mathrm{CR}_{\mathrm{S},1-\alpha}$ is quite easy to calculate in, for example, the R software. All computations and plots shown in this paper were done using R and/or Matlab. Matlab's imaging toolbox contains the function bwdist, which allows for easy computation of the Hausdorff distance, required for calculation of $\mathrm{CR}_{1-\alpha}$. This function is available both for $d = 2$ and $d = 3$, but not for higher dimensions. In all computations, including the simulations which follow, it was necessary to discretize the underlying domain. For $d = 2$, we used 401^2 pixels, and when $d = 3$, we used 101^3 voxels.

4.2 Method comparison

To visually compare the two approaches, we consider two parametric models:

$$E[Y|Z = \boldsymbol{z}] = -6 + 6z_1 + 6z_2, \tag{4.4}$$

$$E[Y|Z = \boldsymbol{z}] = -6 + 6z_1 + 6z_2 + 10z_1^2 + 3z_1z_2 + z_2^2. \tag{4.5}$$

For the linear model (4.4), the median effective dose is the straight line $\mathrm{ED}_{50} = \{(z_1, z_2) \in \mathcal{D} : z_2 = 1 - z_1\}$. For the quadratic model (4.5), it is the curve $\mathrm{ED}_{50} = \{(z_1, z_2) \in \mathcal{D} : 6z_1 + 6z_2 + 10z_1^2 + 3z_1z_2 + z_2^2 = 6\}$.

Examples of both confidence region methods for both models are shown in Figure 1. In these examples, and in the simulations in Section 4.3, we assume that data is collected according to a design with $\kappa = 36$ points uniformly spaced over the domain $\mathcal{D} = [0, 1]^2$ in a grid-like pattern (this is design one in Figure 6 of the Appendix). In Figure 1, for each point in the design, we assume $m = 10$ replicates, for an overall sample size of $n = \kappa\, m = 360$.

For the linear model, the shapes and sizes of the two methods appear quite similar. However, in the quadratic model, $\mathrm{CR}_{\mathrm{S},0.95}$ covers 0.803% of the domain while $\mathrm{CR}_{0.95}$ covers only 0.718%. The fact that $\mathrm{CR}_{0.95}$ is smaller than $\mathrm{CR}_{\mathrm{S},0.95}$ for the quadratic model is probably caused by the fact that Scheffé's bounds are conservative, and this is particularly emphasized with an increase in the difference between $k + 1$ (the number of parameters in the model) and d (the numbers of agents, or covariates). Indeed, the region $\mathrm{CR}_{0.95}$ is nearly 11% smaller than the region $\mathrm{CR}_{\mathrm{S},0.95}$.

Notably, the confidence regions show that there is considerable variability in the quadratic model, in that the function $\widehat{\beta}_n^T\boldsymbol{x}$ dips down with some frequency in the upper right corner of the domain. This is consistent with samples of $\widehat{\mathrm{ED}}_{50}$ (shown in Figure 8 in the supplementary material). Note how much additional information is revealed about the variability of $\widehat{\mathrm{ED}}_{50}$ through viewing the confidence region in this case. In particular the true function $f(\boldsymbol{z}) = \beta^T\boldsymbol{x}$ does not exhibit such "dips" on $\mathcal{D}$ (Figure 7 of the supplementary material), and neither does the function $\widehat{\beta}_n^T\boldsymbol{x}$ (not shown).

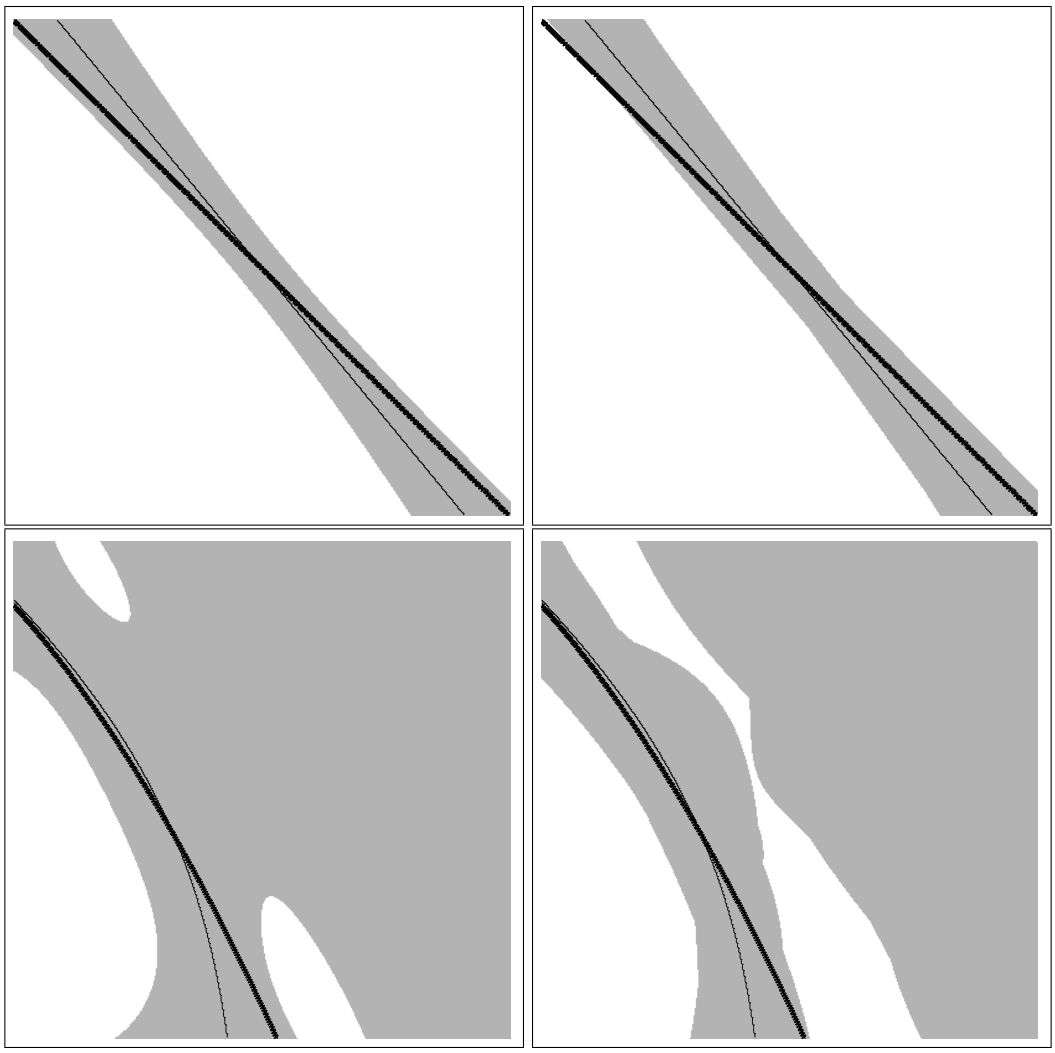

Figure 1: Confidence regions (gray) for ED_{50} (bold) under the linear model in the top (4.4) and the quadratic model in the bottom (4.5); $\widehat{\mathrm{ED}}_{50}$ is shown as the thin line. The region $\mathrm{CR}_{\mathrm{S},0.95}$ is shown on the left and the region $\mathrm{CR}_{0.95}$ is shown on the right.

4.3 Empirical coverage results

We next use simulations to study the behaviour of the new bootstrap algorithm, and to compare it with the method proposed in Carter et al. [7]. Each simulation is the result of 1000 Monte Carlo samples, and we also used $B = 1000$ in the bootstrap algorithm throughout. Whenever possible, the data in each simulation was the same for the different confidence regions.

We again consider both the linear model (4.4) and the quadratic model (4.5) with domain $\mathcal{D} = [0, 1]^2$. We assume that the true Σ is unknown. As in the previous section, we use design 1 in Figure 6 of the Appendix, with $\kappa = 36$. The value of m is chosen as $1, 10$ and 100 for three overall samples sizes of $n = 36, 360, 3600$. The results of the simulations are shown in Table 1. Note that we did not report the results for the quadratic model when $n = 36$. This is because there was too large a proportion of observations in "complete separation" [10] in this case.

Table 1: Empirical coverage probabilities of 95% confidence regions for ED_{100p} using the maximum likelihood estimate of Σ for $d = 2$. Results *not* statistically different from 0.95 are shown in bold. The corresponding sizes of the confidence regions, measured as the mean proportion of the domain covered by the region, are given in brackets. SCH denotes Scheffé's method, while CR denotes the new bootstrap approach.

		linear		quadratic	
n	p	SCH	CR	SCH	CR
36	.1	.989	.995	*	*
		(.47)	(.54)	*	*
	.5	1.00	.998	*	*
		(.93)	(.95)	*	*
	.9	.991	.993	*	*
		(.47)	(.53)	*	*
360	.1	.988	.977	1.00	.994
		(.18)	(.17)	(.34)	(.29)
	.5	.988	**.946**	1.00	.995
		(.19)	(.20)	(.51)	(.40)
	.9	.990	.985	.996	.973
		(.18)	(.19)	(.69)	(.66)
3600	.1	.995	.980	1.00	.971
		(.06)	(.05)	(.06)	(.05)
	.5	.986	**.944**	.999	.982
		(.06)	(.07)	(.06)	(.05)
	.9	.991	.973	.999	.975
		(.06)	(.06)	(.08)	(.07)

For the medium and large samples sizes ($n = 360, 3600$), we find that $\mathrm{CR}_{0.95}$ outperforms $\mathrm{CR}_{\mathrm{S},0.95}$: $\mathrm{CR}_{0.95}$ has better coverage with similar region size in the linear model, and better or equivalent coverage and smaller region size for the quadratic model. Specifically, when $n = 360$, empirical coverage decreased on average by 2% in the linear model and 1% in the quadratic model (these are absolute decreases, not relative ones); when $n = 3600$ empirical coverage decreased on average by 2.5% in the linear model and 2% in the quadratic model. The proportion of the domain covered decreased on average by 0.06 when $n = 360$ and by 0.01 when $n = 3600$ for the quadratic model, and remained the same, on average, for the linear model. When $n = 36$, however, $\mathrm{CR}_{\mathrm{S},0.95}$ and $\mathrm{CR}_{0.95}$ behave comparably. We believe that this is caused by the additional variability due to the estimation of the variance matrix. Much more extensive simulations for $d = 2$ are provided in the Supplementary Online Material. These include additional models and various designs. The results are similar to those presented here for the linear and quadratic models, but this does depend a little on the model and on the design.

We also considered two simple models for the higher dimensional $d = 3$ case: the true linear model was $\mathrm{logit}(p) = -6 + 3z_1 + 3z_2 + 3z_3$ and the true quadratic model was $\mathrm{logit}(p) = -6 +$

Table 2: Empirical coverage probabilities of 95% confidence regions for ED_{100p} for $d = 3$, assuming that Σ is known. Results *not* statistically different from 0.95 are shown in bold. The corresponding sizes of the confidence regions, measured as the mean proportion of the domain covered by the region, are given in brackets. SCH denotes Scheffé's method, while CR denotes the new bootstrap approach.

		linear		quadratic	
n	p	SCH	CR	SCH	CR
2160	.1	.991	.977	.997	**.937**
		(.17)	(.16)	(.16)	(.13)
	.5	.983	.921	.999	**.938**
		(.10)	(.09)	(.17)	(.14)
	.9	.997	.988	1.00	**.961**
		(.01)	(.01)	(.10)	(.08)

$3z_1 + 3z_2 + 3z_3 + z_1^2 + z_2^2 + z_3^2$. The design was similar to design one, in that data were observed on a uniform grid inside $\mathcal{D} = [0, 1]^3$ with $\kappa = 6^3$ and $m = 10$. The domain was the same for both models, and we considered only the setting when the true Σ is known (cf. additional simulations for $d = 2$ available in the Appendix). The results are shown in Table 2.

5 Examples

5.1 A decompression sickness study

Our first example comes from a decompression sickness (DCS) study from the University of Wisconsin, Madison. DCS is most often associated with diving, but can be experienced in other de-

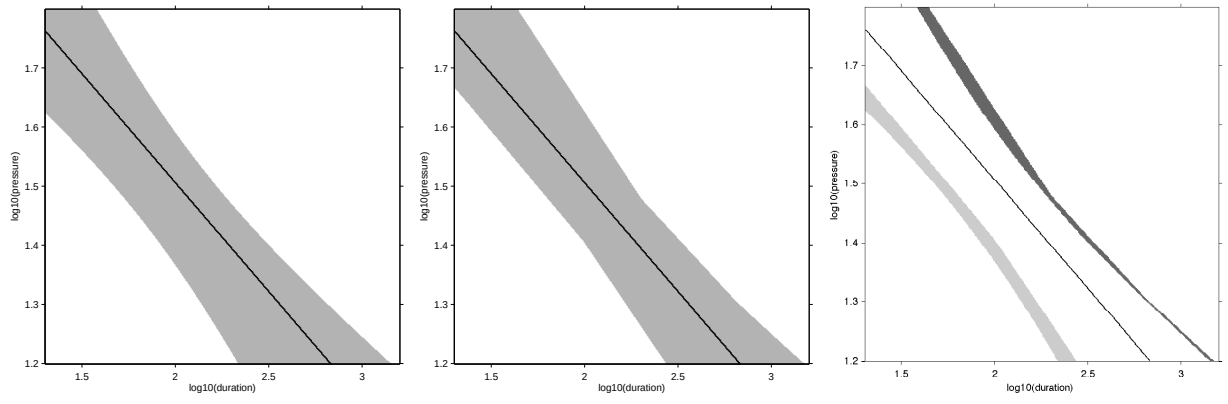

Figure 2: 95% confidence regions for ED_1 based on Scheffé's upper bound (left) and the new CRS quantiles (centre). The difference between the sets is shown on the left ($\mathrm{CR}_{\mathrm{S},0.95} \setminus \mathrm{CR}_{0.95}$ in light and $\mathrm{CR}_{0.95} \setminus \mathrm{CR}_{\mathrm{S},0.95}$ in dark gray). $\widehat{\mathrm{ED}}_1$ is superimposed throughout for reference.

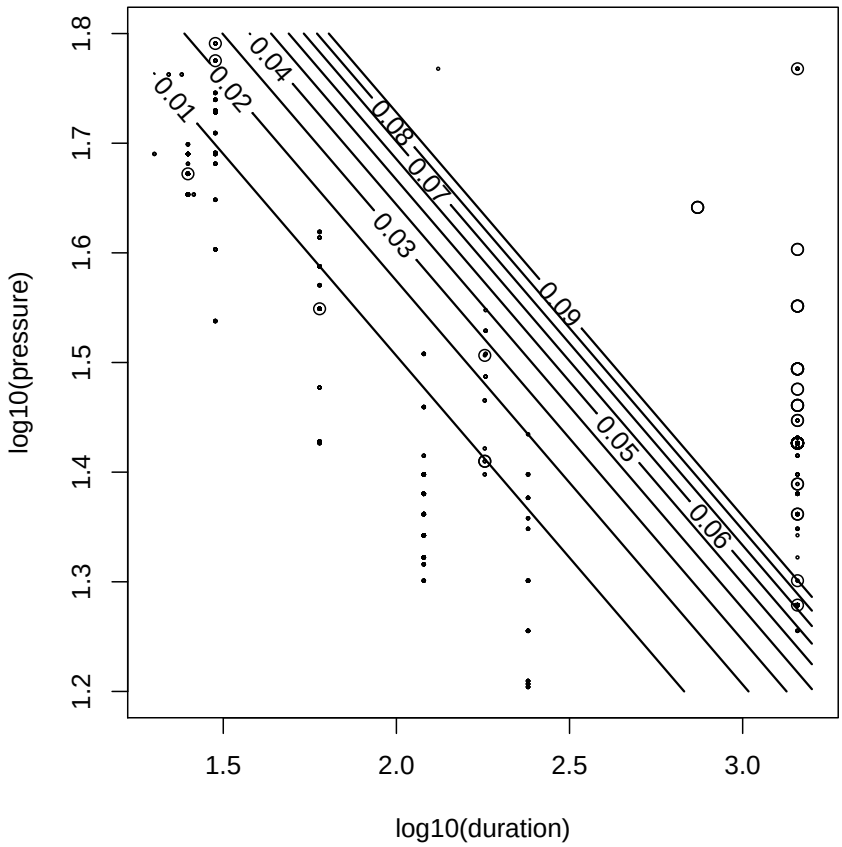

Figure 3: Estimates of ED_{100p} for different values of p for the DCS data. Data points with response equal to 0/1 are shown as closed/open circles.

pressurisation events such as caisson working (i.e. during the building of dams or tunnels) or unpressurized flight. In this study, sheep were exposed to a variety of exposure pressures and durations in a pressure chamber. DCS is caused by dissolved gasses vaporising on depressurization forming dangerous bubbles of gas throughout the body. The bubbles can form in different locations in the body, and therefore lead to a variety of symptoms as well as severities. In the study, central nervous system and respiratory DCS as well as limb bends and mortality were included in the recorded outcomes. The sample size is $n = 1108$. The data set has been considered extensively in Li et al. [3, 4, 5], Li and Wong [6] and we refer to these papers for further details on data collection and analysis. An alternative approach to the analysis was also presented in Li and Ma [11]. Here, we compare our methods to those of Li et al. [3] where a linear model was fit for the mortality response. The data has been updated several times in the past few years, and therefore our results differ slightly from those found in Li et al. [3].

When the survival response is considered, a simple linear model gives an adequate fit to the data. The effective dose (with death denoted as "success") is thus estimated as

$$\widehat{\mathrm{ED}}_{100p} \;=\; \{-27.64 + 3.78z_1 + 10.28z_2 = \mathrm{logit}(p)\},$$

where we take $\mathcal{D} = [1.3, 3.6] \times [1.2, 1.8]$ with z_1 and z_2 denoting the base 10 logarithm of duration

and pressure of the dive. The data and results for $\widehat{\mathrm{ED}}_{100p}$ are shown in Figure 3. Fix $p = 0.01$. From Theorem A.1 we know that, under a linear model, the plug-in estimator $\widehat{\mathrm{ED}}_1$ is consistent. We are also interested in the variability of $\widehat{\mathrm{ED}}_1$, and therefore we calculate simultaneous confidence regions for ED_1. Results for $\mathrm{CR}_{\mathrm{S},0.95}$ and $\mathrm{CR}_{0.95}$ are shown in Figure 2. The regions are similar in size, covering 30% and 29% of the domain respectively. However, $\mathrm{CR}_{0.95}$ appears more centred around $\widehat{\mathrm{ED}}_1$, particularly for large values of z_1.

5.2 Cytotoxicity in the leukemia cell line HL-60

Carter et al. [7] consider a cytotoxicity data set where the effect of two toxins, methylmethanesulfonate (MMS) and phorbol 12-myristate 13-acetate (PMA), on the human promyeloctic leukemia cell line HL-60 was evaluated. Both MMS and PMA have demonstrated carcinogenic properties, and it was of interest to understand their interactive properties. In the study, 16 treatments were considered with 83 to 98 observations per treatment, for a total sample size of 1436. The data and a detailed analysis is available in Carter et al. [7]. There, a logistic regression model was fit resulting in the plug-in estimate

$$\widehat{\mathrm{ED}}_{100p} \;=\; \{-1.330 - 0.084z_1 + 0.159z_2 + 0.003883z_1^2 - 0.001308z_2^2 = \mathrm{logit}(p)\},$$

with z_1, z_2 corresponding to MMS/10 and PMA, respectively. Several estimates for different values of p are shown in Figure 4. The appropriateness of the quadratic model is at first counterintuitive, however, due to lysis, certain treatments can become so toxic that cells become uncountable [7].

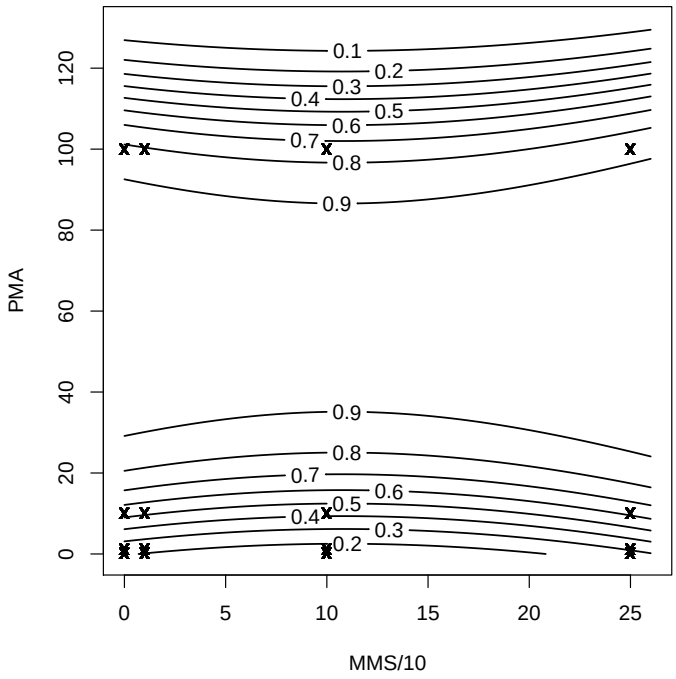

Figure 4: Estimates of ED_{100p} for different values of p for the cytotoxicity data. Locations where data was observed are marked with an "x" (the treatments had 83 to 98 observations each).

Thus, an increase in perceived survival at higher toxicities is an appropriate result in this experiment.

The fitted model is non-linear, however, we can preform a simple check for consistency as follows, based on the results given in Section A.3. The determinant of the Hessian matrix for the model $f(\boldsymbol{z}) = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_1^2 + \beta_4 z_2^2$ is equal to $4\beta_3\beta_4$. If this quantity is negative, then all critical points of $f(\boldsymbol{z})$ are saddle points, and we would therefore have consistency. Now, if $\beta_3 > 0$ and $\beta_4 < 0$, then $4\beta_3\beta_4 < 0$. From Carter et al. [7], the p-values for each of these tests are smaller than 0.0002 and 0.0001 respectively. Therefore, under the given model, the estimators $\widehat{\mathrm{ED}}_{100p}$ are consistent.

With such a large sample size, we would expect little variability in the values of $\widehat{\beta}_n$. However, it is not immediately clear how this translates to the variability of $\widehat{\mathrm{ED}}_{100p}$. We therefore calculated 95% confidence regions for ED_{50} using Scheffé's upper bound (Figure 5, left) and the new bootstrap algorithm (Figure 5, centre). The new method yields a tighter confidence band. Indeed, here, the percentage of the domain covered is 14.92%, whereas using Scheffé's upper bound 18.31% of the domain was covered. Figure 5 (right) shows that $\mathrm{CR}_{0.95}$ is almost entirely contained in $\mathrm{CR}_{\mathrm{S},0.95}$.

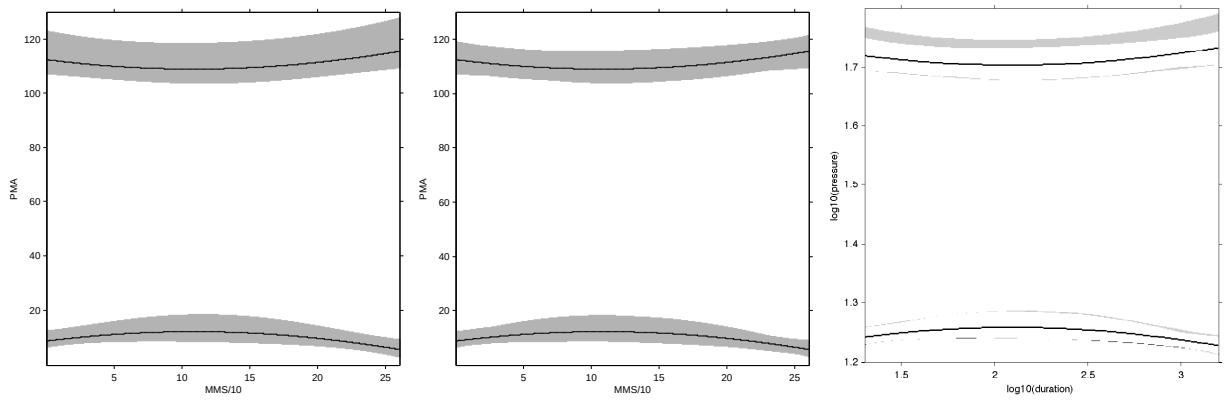

Figure 5: 95% confidence regions for ED_{50} based on Scheffé's upper bound (left) and the new method (centre). The difference between the sets is shown on the left ($\mathrm{CR}_{\mathrm{S},0.95} \setminus \mathrm{CR}_{0.95}$ in light and $\mathrm{CR}_{0.95} \setminus \mathrm{CR}_{\mathrm{S},0.95}$ in dark gray). $\widehat{\mathrm{ED}}_{50}$ is superimposed throughout for reference.

Carter et al. [7] also calculate confidence regions using Scheffé's method for $p = 0.4, 0.5$, and 0.6. These regions overlap, and hence Carter et al. [7] conclude that we "cannot confidently distinguish among the respective ED_{100p} sets." Although not shown, the regions in this case using the new method would also overlap. However, the estimates $\widehat{\mathrm{ED}}_{100p_1}$ and $\widehat{\mathrm{ED}}_{100p_2}$ are highly positively correlated, and therefore such an ad-hoc comparison is probably overly conservative. The question of how to correctly account for this correlation is also interesting and important, but is beyond the scope of this work.

6 Discussion

When studying confidence intervals in the univariate setting, one is typically looking for two properties: the confidence interval should be (1) as small as possible while (2) reaching the nominal

level, but without under-coverage. In our context, the observed data comes from a non-Euclidean space, so some additional properties are desirable: Confidence regions provide visual information about the variability of our estimates, and therefore the shape of the confidence regions should reflect the behaviour of the estimates and have the ability to detect local variability. For practitioners, it would also be desirable that the confidence region be relatively easy to calculate.

In this work we have focused on two methods of calculating a confidence region for a multidimensional effective dose. Both confidence regions $\mathrm{CR}_{\mathrm{S},0.95}$ and $\mathrm{CR}_{0.95}$ all nicely show the local variability of the effective dose estimators. However, $\mathrm{CR}_{\mathrm{S},0.95}$ systematically exhibits over-coverage. The new empirical bootstrap algorithm introduced here yields the region $\mathrm{CR}_{0.95}$ which outperforms $\mathrm{CR}_{\mathrm{S},0.95}$ in terms of empirical coverage and/or size of region for moderate and large samples sizes. For smaller sample sizes, practitioners may therefore prefer to use $\mathrm{CR}_{\mathrm{S},0.95}$ due to its simplicity. On the other hand, for medium to large sample sizes, such as those of the examples in Section 5, the proposed method provides confidence regions with better coverage properties.

Previous work on the problem of confidence regions for the effective dose is, to our best knowledge, rather limited. The region $\mathrm{CR}_{\mathrm{S},1-\alpha}$ was introduced in Carter et al. [7] and studied extensively in Li et al. [3] for the parametric model. Li et al. [3] also consider the problem of the conditional effective dose (i.e. the effective dose obtained after holding certain covariates fixed). Although our simulations do not explore this, the new method presented here can also be applied in this setting. We conjecture that empirical coverage will be similar, particularly for models without interactions, as fixing a covariate essentially changes the behaviour of the intercept in such cases. This conditional approach would be most appealing in higher dimensions, since relationships are more difficult to visualize when $d > 2$.

Li et al. [5] also study an extension of this approach to the case where $\widehat{f}_n$ is estimated using a semi-parametric model. They consider two methods: one is based on a theoretical bound of $\max n(\widehat{f}_n(\boldsymbol{z}) - f(\boldsymbol{z}))^2$ similar to the parametric case, and the second is based on empirical estimation of the theoretical quantiles of $\max n(\widehat{f}_n(\boldsymbol{z}) - f(\boldsymbol{z}))^2$ via a parametric bootstrap procedure. However, their theoretical bounds in this setting are incorrect (this is the reason for the discrepancy between the bootstrap and theoretical procedure noted in Li et al. [5, Section 4.1 and Figure 1]). The bootstrap approach, although quite conservative, is more promising. However, although the semi-parametric estimates are insensitive to the choice of bandwidth, the parametric bootstrap confidence regions increase in size and coverage as h decreases. This is not surprising as the variance depends on $\sqrt{nh}$ in the denominator. The question of optimal choice of h was not considered in Li et al. [3]. One could also try the region $\mathrm{CR}_{1-\alpha}$ in the semi-parametric setting, and we intend to study this problem thoroughly in a future work. Important developments in level set estimation in the nonparametric model with applications to the effective dose estimation problem are Mammen and Polonik [12], Mason and Polonik [13] (see also the references in Mason and Polonik [13]). In particular, Mammen and Polonik [12] study the problem of construction a confidence region for a density level set using kernel density estimators, which is closely related to the semi-parametric effective dose setting.

The confidence region $\mathrm{CR}_{\mathrm{S},0.95}$ is asymptotic, in that it uses the asymptotic distribution quantiles. For smaller sample sizes, one could potentially either bootstrap these quantiles or replace them with their non-asymptotic counterparts. Thus, in $\mathrm{CR}_{\mathrm{S},0.95}$ one could replace the asymptotic χ^2 quantile with its appropriate F distribution version [8]. However, these quantiles would be larger, and $\mathrm{CR}_{\mathrm{S},0.95}$ already exhibits over-coverage. Another potential avenue here would be to consider the method of Piegorsch and Casella [14] (see also Casella and Strawderman [15]), but we do not

explore this here. Notably, this modification would greatly increase the computational complexity of the regions $\mathrm{CR}_{\mathrm{S},0.95}$.

Of the confidence regions consider here, the new confidence region $\mathrm{CR}_{1-\alpha}$ is more computationally intensive and requires some simulations. On the other hand, the region $\mathrm{CR}_{\mathrm{S},1-\alpha}$ may be calculated directly. However, in all of the two-dimensional examples considered here, $\mathrm{CR}_{1-\alpha}$ took less then one minute to compute on a 2.4 GHz dual core Macintosh laptop. This could probably be reduced even further by studying more efficient programming techniques. Thus, the time requirements to calculate these regions do not carry great practical constraints. Matlab script calculating all the regions for the example of Section 5.2 will be made available online via the Dryad Digital Repository (`datadryad.org`).We intend to create an R version of this program in the near future, and make it available online, also via the Dryad Digital Repository.

A Appendix

A.1 Some additional notation

We denote Euclidean distance for $z, y \in \mathbb{R}^d$ as $|z-y|$. Then, for a set $A \subset \mathbb{R}^d$ and $\varepsilon > 0$ we define the dilation of A as $A^\varepsilon = \{z \in \mathbb{R}^d : |z-y| \leq \varepsilon, \text{ for some } y \in A\}$. The complement of a set A is denoted as A^c while the closure of a set A is denoted as $\overline{A}$. Finally, we define the Hausdorff distance between two sets

$$\rho(A,B) \;=\; \inf\{\varepsilon > 0 : A \subset B^\varepsilon,\, B \subset A^\varepsilon\}. \tag{A - 6}$$

A.2 Algorithm for confidence region for ED^-_{100p}

Bootstrap-based confidence region

Suppose that we have estimated $\widehat{\beta}_n$ and the variance $\widehat{\Sigma}_n$ in the logistic regression model based on an available data set with sample size n.

1. Generate $\widehat{\beta}^*_{n,i} \sim N(\widehat{\beta}_n, \widehat{\Sigma}_n/n)$ independent random variables for $i = 1, \ldots, B$.

2. For each $\widehat{\beta}^*_{n,i}$ calculate the effective dose estimate

$$\widehat{\mathrm{ED}}^{-*}_{100p,i} \;=\; \left\{\boldsymbol{z} \in \mathcal{D} : \widehat{\beta}^{*T}_{n,i}\boldsymbol{x} \leq \mathrm{logit}(p)\right\}.$$

3. Compute

$$K = \bigcap_{i=1}^{B} \widehat{\mathrm{ED}}^{-*}_{100p,i},$$

and then calculate $\xi_i = \rho(\widehat{\mathrm{ED}}^{-*}_{100p,i}, K)$. Here, ρ denotes the Hausdorff distance on sets (see (A - 6)).

4. Calculate γ_α, the $100(1-\alpha)$ sample percentile of the ξ_i observations. Finally, compute

$$\mathrm{CR}^-_{1-\alpha} \;=\; \bigcup_{i:\ \xi_i<\gamma_\alpha} \widehat{\mathrm{ED}}^{-\,*}_{100p,i}.$$

A.3 Consistency

Here, we provide conditions necessary of consistency, by extending the work of Molchanov [16] and Cuevas et al. [17] on level sets to the logistic regression setting. A very similar problem was recently studied in Jankowski and Stanberry [18], and the following result follows directly from the proof of Jankowski and Stanberry [18, Theorem 3.1] as well as consistency of the maximum likelihood estimator $\widehat{\beta}_n$. Following Molchanov [16], we say that a random set $\boldsymbol{A}_n$ is strongly consistent (for A) if for all compact sets K, $\lim_{n\to\infty} \rho(\boldsymbol{A}_n \cap K, A \cap K) = 0$ almost surely.

Proposition A.1. *Let β denote the value of the parameters in model* (2.1)*, and assume that $\boldsymbol{x}(\boldsymbol{z})$ is continuous. If $\widehat{\beta}_n \to \beta$ almost surely, then $\widehat{\mathrm{ED}}^{+}_{100p}$ is strongly consistent for ED^+_{100p}, if*

$$\overline{\{\boldsymbol{z} \in \mathcal{D} : \beta^T \boldsymbol{x} > \mathrm{logit}(p)\}} \;=\; \mathrm{ED}^+_{100p} \tag{A - 7}$$

holds. Furthermore, $\widehat{\mathrm{ED}}_{100p}$ is strongly consistent for ED_{100p}, if both (A - 7) *and*

$$\overline{\{\boldsymbol{z} \in \mathcal{D} : \beta^T \boldsymbol{x} < \mathrm{logit}(p)\}} \;=\; \{\boldsymbol{z} \in \mathcal{D} : \beta^T \boldsymbol{x} \le \mathrm{logit}(p)\} \;=\; \mathrm{ED}^-_{100p} \tag{A - 8}$$

hold.

The two conditions (A - 7) and (A - 8) can be re-stated mathematically in different ways; for example, (A - 7) is equivalent to ED^+_{100p} being regularly closed. In terms of the response surface, condition (A - 7) requires that $f(\boldsymbol{z}) = \beta^T \boldsymbol{x}(\boldsymbol{z})$ have no local maxima on ED_{100p} while condition (A - 8) requires that $f(\boldsymbol{z}) = \beta^T \boldsymbol{x}(\boldsymbol{z})$ have no local minima on ED_{100p}. Thus, in the linear model, we are guaranteed to have consistency if at least one β_i, $i = q, \ldots, k$ is non-zero (excluding the intercept term). For other models, one could perform a heuristic check of critical points of $f(\boldsymbol{z})$, if $\boldsymbol{x}$ is differentiable as a function of $\boldsymbol{z}$. As an example, consider the model $f(\boldsymbol{z}) = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_1 z_2$. The determinant of the Hessian matrix is $-\beta_3^2 < 0$. Thus, if β_3 is non-zero, any critical points of $f(\boldsymbol{z})$ are saddle points and therefore no local minima/maxima exist in the domain of f, and hence also on ED_{100p}.

A.4 Technical arguments

Proposition A.2. *Let $A_1, \ldots, A_B$ be a collection of nonempty subsets of $\mathbb{R}^d$.*

1. *The proportion of sets $\widehat{\mathrm{ED}}^{+\,*}_{100p,1}, \ldots, \widehat{\mathrm{ED}}^{+\,*}_{100p,B}$ which is contained in $\mathrm{CR}^+_{1-\alpha}$ is at least $\lfloor (1-\alpha)\, B \rfloor / B$.*

2. *Consider $\gamma_1 \le \gamma_2$. Then $\cup\{A_i, i = 1, \ldots, B : \rho(A_i, K_B) < \gamma_1\} \subseteq \cup\{A_i, i = 1, \ldots, B : \rho(A_i, K_B) < \gamma_2\}$.*

3. *Fix a rigid motion (a transformation consisting of rotations and translations)* $g \in E^+(d)$ *and let* $C_i = g(A_i)$. *Then*

$$g\left(\cup\{A_i, i = 1, \ldots, B : \rho(A_i, K_n) < \gamma\}\right) \quad = \quad \cup\{C_i, i = 1, \ldots, B : \rho(C_i, g(K_B)) < \gamma\}.$$

4. *Fix* $\alpha > 0$ *and let* $C_i = \alpha A_i$. *Then*

$$\cup\{C_i, i = 1, \ldots, B : \rho(C_i, \alpha K_B) < \alpha\gamma\} \quad = \quad \alpha\left\{\cup\{A_i, i = 1, \ldots, B : \rho(A_i, K_B) < \gamma\}\right\}.$$

Proof of Proposition A.2. 1. By definition, $\gamma_{q,B}$ is such that $\lfloor qB \rfloor$ of the observed sets satisfy $\rho(A_i, K_B) \leq \gamma_{q,B}$.

2. This again follows from the definition.

3. It is well known that $\rho(A, B) = \sup_{z \in \mathcal{D}} |d_A(z) - d_B(z)|$, where $d_A(\cdot)$ denotes the distance transform of A [19], and also that $d_{g(A)}(z) = d_A(g^{-1}(z))$ [20, see e.g.]. It follows that

$$\rho(g(A_i), g(K_B)) \quad = \quad \rho(A_i, K_B),$$

which implies the result.

4. The same argument as for rigid motions works for dilations by $\alpha > 0$.

□

Acknowledgements

The first author thanks Georges Monette from York University and Ruxandra Pinto from Sunnybrook Hospital for helpful discussions. All authors are grateful to Jialiang Li for sharing the DCS data set. Parts of this work were made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: www.sharcnet.ca) and Compute/Calcul Canada.

References

[1] Skarin, AT, Canellos, GP, Rosenthal, DS, Case, DC, MacIntyre, JM, Pinkus, GS, Moloney, WC, Frei, E. Improved prognosis of diffuse histiocytic and undifferentiated lymphoma by use of high dose methotrexate alternating with standard agents (M-BACOD). *Journal of Clinical Oncology* 1983; **1**:91–98.

[2] Lang, DR, Kurzepa, H, Cole, MS, Loper, JC. Malignant transformation of BALB/3T3 cells by residue organic mixtures from drinking water. *Journal of Environmental Pathology & Toxicology* 1980; **4**:41–54.

[3] Li, J, Zhang, C, Nordheim, E, Lehner, C. On the multivariate predictive distribution of multidimensional effective dose: a Bayesian approach. *Journal of Statistical Computation and Simulation* 2008; **78**(5):429–442.

[4] Li, J, Nordheim, E, Zhang, C, Lehner, C. Estimation and confidence regions for multi-dimensional effective dose. *Biometrical Journal* 2008; **50**(1):110–122.

[5] Li, J, Zhang, CM, Doksum, KA, Nordheim, EV. Simultaneous confidence intervals for semi-parametric logistic regression and confidence regions for the multi-dimensional effective dose. *Statistica Sinica* 2010; **20**(2):637659.

[6] Li, J, Wong, WK. Two-dimensional toxic dose and multivariate logistic regression, with application to decompression sickness. *Biostatistics* 2011; **12**(1):143155.

[7] Carter, W, Chinchilli, V, Wilson, J, Campbell, E, Kessler, F, Carchman, R. An asymptotic confidence region for the ED100p from the logistic response surface for a combination of agents. *The American Statistician* 1986; **40**(2):124–128.

[8] Scheffé, H. A method for judging all contrasts in the analysis of variance. *Biometrika* 1953; **40**:87–104.

[9] Chen, DG. Dose-time response cumulative multinomial generalized linear model. *J. Biopharm. Statist.* 2007; **17**(1):173–185.

[10] Albert, A, Anderson, JA. On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* 1984; **71**(1):1–10.

[11] Li, J, Ma, S. Interval-censored data with repeated measurements and a cured subgroup. *J. R. Stat. Soc. Ser. C. Appl. Stat.* 2010; **59**(4):693–705.

[12] Mammen, E, Polonik, W. Confidence regions for level sets. *Journal of Multivariate Analysis* 2013; **122**:202–214.

[13] Mason, DM, Polonik, W. Asymptotic normality of plug-in level set estimates. *Annals of Applied Probability* 2009; **19**:1108–1142.

[14] Piegorsch, WW, Casella, G. Confidence bands for logistic regression with restricted predictor variables. *Biometrics* 1988; **44**:739–750.

[15] Casella, G, Strawderman, WE. Confidence bands for linear regression with restricted predictor variables. *J. Amer. Statist. Assoc.* 1980; **75**(372):862–868.

[16] Molchanov, IS. A limit theorem for solutions of inequalities. *Scandinavian Journal of Statistics* 1998; **25**:235–242.

[17] Cuevas, A, González-Manteiga, W, Rodríguez-Casal, A. Plug-in estimation of general level sets. *Aust. N. Z. J. Stat.* 2006; **48**(1):7–19.

[18] Jankowski, H, Stanberry, L. Confidence regions for means of random sets using oriented distance functions. *Scandinavian Journal of Statistics* 2012; **39**(2):340–357.

[19] Delfour, MC, Zolésio, JP. Shape analysis via oriented distance functions. *J. Funct. Anal.* 1994; **123**(1):129–201.

[20] Jankowski, H, Stanberry, L. Expectations of random sets and their boundaries using oriented distance functions. *Journal of Mathematical Imaging and Vision* 2010; **36**(3):291–303.