Supplementary materials for this article are available at https://doi.org/10.1007/s13253-021-00472-0.

Generalized Fiducial Inference for Threshold Estimation in Dose–Response and Regression Settings

Seungyong HWANG, Randy C. S. LAI, and Thomas C. M. LEE

In many biomedical experiments, such as toxicology and pharmacological dose–response studies, one primary goal is to identify a threshold value such as the minimum effective dose. This paper applies Fisher's fiducial idea to develop an inference method for these threshold values. In addition to providing point estimates, this method also offers confidence intervals. Another appealing feature of the proposed method is that it allows the use of multiple parametric relationships to model the underlying pattern of the data and hence, reduces the risk of model mis-specification. All these parametric relationships satisfy the qualitative assumption that the response and dosage relationship is monotonic after the threshold value. In practice, this assumption may not be valid but is commonly used in dose–response studies. The empirical performance of the proposed method is illustrated with synthetic experiments and real data applications. When comparing to existing methods in the literature, the proposed method produces superior results in most synthetic experiments and real data sets.

Supplementary materials accompanying this paper appear on-line.

Key Words: Confidence intervals; Markov chain Monte Carlo; Minimum effective dose; Model averaging; Uncertainty quantification.

1. INTRODUCTION

In pharmacological and toxicological dose–response experiments, a primary and important goal is to identify a threshold value for a specific purpose. For example, identifying the minimum effective dose (MED) of a medicine in the beginning stage of a dose–response experiment is important, as any dose level that is less than the identified level does not need to be considered in further studies. As another related example, the identification of the no-observed-adverse-effect-level (NOAEL) is equally important. It is because in general, as the

S. Hwang · R. C. S. Lai · T. C. M. Lee (✉) Department of Statistics, University of California at Davis, One Shields Avenue, Davis, CA 95616, USA (E-mail: *tcmlee@ucdavis.edu*)
S. Hwang (E-mail: *syhwang@ucdavis.edu*)
R. C. S. Lai (E-mail: *rcslai@ucdavis.edu*) .

Journal of Agricultural, Biological, and Environmental Statistics, Volume 27, Number 1, Pages 109–124
https://doi.org/10.1007/s13253-021-00472-0

dosage increases, the possibility of having adverse effects also increases. Therefore, once NOAEL is identified, one can minimize the adverse side effects from the medicine and also the cost of the experiments (Bretz et al. 2008; Guo and Li 2014; Kim et al. 2016; Schwartz et al. 2001; West and Kodell 2005; Coffey and Gennings 2007). In this context, the detection of MED or NOAEL can be considered as a threshold estimation problem (e.g., Mallik et al. 2011). In addition to dose–response studies, threshold estimation problems can also be found in environmental studies such as detecting changes in air pollutant levels (Holst et al. 1996).

Existing statistical methods for threshold estimation in dose–response experiments can be broadly grouped into two different categories (Bretz et al. 2008): multiple comparison methods (Williams 1972; Ruberg 1989; Hsu and Berger 1999; Tamhane and Logan 2002; Guo and Li 2014; Otava et al. 2017) and model-based methods (Cox 1987; Muggeo 2003; Schwartz et al. 2001; West and Kodell 2005; Coffey and Gennings 2007; Lutz and Lutz 2009; Mallik et al. 2011), while multiple comparison methods can be further divided into Bayesian (e.g., Guo and Li 2014; Otava et al. 2017) and frequentist (e.g., Williams 1972; Ruberg 1989; Hsu and Berger 1999; Tamhane and Logan 2002) approaches.

Multiple comparison methods require fewer assumptions, but they restrict the estimated MED to be one of the investigated levels, i.e., they will not produce any estimated MED value that is different from those dose levels that were experimented with—not even with interpolation between doses levels or similar techniques. On the other hand, model-based methods are capable of interpolating and extrapolating dose levels as they impose a stronger assumption that there exists a model that adequately captures the relationship between the dose levels and the response. Most drug development experiments assume that such a relationship is non-decreasing, which means the higher the dosage the stronger the therapeutical effects (Bretz et al. 2008; Lutz and Lutz 2009; Mallik et al. 2011). Of course, if the chosen model does not adequately capture the relationship, the estimated threshold value is unlikely to be close to the true value. This is a potential drawback of the model-based methods.

Both Bayesian (e.g., Guo and Li 2014; Otava et al. 2017) and frequentist (e.g., Williams 1972; Hsu and Berger 1999) multiple comparison methods typically define the hypotheses of interest according to the investigated dose levels. For example, if there are two dose levels, four hypotheses will be investigated: (i) none of the doses have any effect, (ii) only one dose has effect, (iii) both doses have the same effect, and (iv) the two doses have different effects. As the number of investigated dose levels increases, the number of hypotheses increases rapidly. Therefore, such methods (e.g., Williams 1972; Hsu and Berger 1999; Guo and Li 2014; Otava et al. 2017) only considered the case when the number of dose levels is small (4 or 5). The Bayesian methods typically assign prior probabilities to the hypotheses and make final conclusions using the posterior probabilities.

For model-based methods, if one believes the relationship between the dosage and response is monotonic, one can model this problem with a regression function that is constant to the left of a certain covariate value t^0, say, and is increasing to the right of t^0 (Bretz et al. 2008; Lutz and Lutz 2009; Mallik et al. 2011). Under this regression model, the main purpose is to make an inference about the value of t^0, the threshold value. In dose–response experiments, it can be a MED, which starts showing a discernible effect compared to a control. A notable example is the break point estimation method developed by Muggeo (2003).

With a parametric segmented regression model commonly used in the threshold estimation problem, this method first detects if a threshold t^0 exists, and if yes, it then provides the maximum likelihood estimate (MLE) as well as a confidence interval for t^0. Consequently, this method does not always produce an estimate for t^0. More recently, in Mallik et al. (2011) a method based on a novel p value framework was developed. The idea is to first conduct a sequence of tests that, for each covariate value, assesses whether the regression function is at its baseline, then calculate the corresponding p values, and lastly estimate t^0 by fitting a piecewise function to those calculated p values. This method, however, does not provide confidence intervals.

The main contribution of this paper is a new method for performing statistical inference for t^0. This method is model-based and has the following properties:

- It is neither Bayesian nor frequentist; rather, it follows the generalized fiducial inference methodology of Hannig et al. (2016).
- Model averaging is employed to avoid the above-mentioned danger of selecting a wrong model.
- Unlike most existing methods, it provides point estimates as well as confidence intervals for t^0.

To be more specific, four different parametric forms are used for model averaging: the first one is a step function with its jump point at t^0, the second one begins with a constant function until it hits t^0 and then has a linear trend afterward, while the third and fourth are similar, except they have, respectively, a quadratic and a cubic trend after t^0; some of these parametric forms were used by previous authors (more below). The generalized fiducial inference methodology is used to compute a probability density function for these four parametric forms. With such a density function, the proposed method pools multiple threshold estimates from each of the four parametric forms to form a single point estimate for t^0, as well as constructs a confidence interval for its value.

The rest of this paper is organized as follows. Section 2 provides a brief introduction of generalized fiducial inference, while Sect. 3 states the problem formulation. In Sect. 4, the empirical performance of the proposed method is illustrated under various experimental settings, while in Sect. 5 the proposed method is applied to two real datasets. Lastly, concluding remarks are offered Sect. 6.

2. A BRIEF INTRODUCTION OF GENERALIZED FIDUCIAL INFERENCE

Bayesian inference is an important statistical methodology that offers both point and uncertainty estimates for the unknown parameters via the calculations of posterior distributions. This provides more information compared to frequentist inference (Xie and Singh 2013). However, when no prior information is available, any naive application of the Bayesian methodology could cause concerns; see Efron (2013) for an example in which different priors lead to very different results. To address this issue, Fisher (1930) introduced fiducial inference, where the main idea is to transfer the randomness from the observed data

into the uncertainties of the parameter estimates via the use of a switching principle. This switching principle is also used in maximum likelihood estimation and will be described further below. The uncertainties of the parameter estimates are summarized in the form of a distribution function, termed generalized fiducial density, which plays a similar role as the posterior distribution in the Bayesian setting.

Soon it was realized by other researchers that Fisher's fiducial idea only works for univariate problems, and hence, it did not receive much attention from the statistics community for a few decades. However, there has been a recent resurgence of interest in different variants of Fisher's idea, including Dempster–Shafer theory (Dempster 2008), inferential models (Martin and Liu 2015), confidence distribution (Xie and Singh 2013), and more generally fusion learning Cheng et al. (2014). All these variants use similar but different ideas to extend Fisher's idea to multivariate cases. The particular variant that this paper considers is the so-called *generalized fiducial inference* (Hannig et al. 2016).

Generalized fiducial inference (GFI) repairs many drawbacks of Fisher's original proposal and has been applied to various inference problems with great success, including wavelet regression, ultrahigh-dimensional models, nonparametric survival function estimation; e.g., see Hannig and Lee (2009), Lai et al. (2015), Hannig et al. (2016), Cui and Hannig (2019) and references therein. GFI assumes that the data are generated from a data generating equation G and expresses the relationship between the data $\mathbf{Y}$ and the parameters $\boldsymbol{\theta}$ as

$$\mathbf{Y} = G(\mathbf{U}, \boldsymbol{\theta}), \tag{1}$$

where $\mathbf{U}$ is the random component whose distribution is assumed to be completely known. For many practical problems, it is the distribution of the (normalized) residuals. For the current problem, it is the distribution of u in (4) below, which is assumed to be $\mathcal{N}(0, 1)$. Similar to Fisher's maximum likelihood method, an important idea behind GFI is the so-called switching principle. That is, the roles of the random data $\mathbf{Y}$ and the deterministic parameters $\boldsymbol{\theta}$ are switched after $\mathbf{Y}$ is observed: the observed data, denoted as $\mathbf{y}$, are now treated as deterministic while $\boldsymbol{\theta}$ is treated as random. By using (1) and this switching principle, we can derive a probability distribution of $\boldsymbol{\theta}$ as follows.

Suppose for now, the data generating equation G is invertible for any $\mathbf{u} \in \mathbf{U}$ and any observed $\mathbf{y}$. That is, the inverse G^{-1} always exists:

$$\boldsymbol{\theta} = G^{-1}(\mathbf{y}, \mathbf{u}). \tag{2}$$

Recall that the distribution of $\mathbf{U}$ is completely known, thus one can generate a random sample $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}\}$ of $\mathbf{U}$ and plug them into (2) and obtain

$$\tilde{\boldsymbol{\theta}}_1 = G^{-1}(\mathbf{y}, \mathbf{u_1}), \quad \tilde{\boldsymbol{\theta}}_2 = G^{-1}(\mathbf{y}, \mathbf{u_2}), \quad \ldots, \quad \tilde{\boldsymbol{\theta}}_n = G^{-1}(\mathbf{y}, \mathbf{u_n}).$$

We shall call $\{\tilde{\boldsymbol{\theta}}_1, \ldots, \tilde{\boldsymbol{\theta}}_n\}$ a *fiducial sample* of $\boldsymbol{\theta}$, which can be used to form point estimates and confidence intervals for $\boldsymbol{\theta}$, as similar to a posterior sample in the Bayesian context. One can see that through G^{-1}, the randomness in the data $\mathbf{y}$ is transformed into the randomness in the parameter $\boldsymbol{\theta}$.

Of course in practice, the inverse function G^{-1} in (2) does not always exist, and Hannig (2013) proposed a solution to deal with this issue.

From the above discussion, one can see that there is a density function behind the fiducial sample $\{\tilde{\boldsymbol{\theta}}_1, \ldots, \tilde{\boldsymbol{\theta}}_n\}$. We shall call this density $r(\boldsymbol{\theta}|\mathbf{y})$ the *generalized fiducial density*. When $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ and $\mathbf{y} \in \mathbb{R}^n$, and under some regularity conditions, it can be shown that $r(\boldsymbol{\theta}|\mathbf{y})$ admits the following expression (Hannig 2013; Hannig et al. 2016):

$$r(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\mathbf{y}, \boldsymbol{\theta}) J(\mathbf{y}, \boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} f(\mathbf{y}, \boldsymbol{\theta}') J(\mathbf{y}, \boldsymbol{\theta}') \mathrm{d}\boldsymbol{\theta}'}, \tag{3}$$

where $f(\mathbf{y}, \boldsymbol{\theta})$ is the likelihood, $J(\mathbf{y}, \boldsymbol{\theta})$ is a Jacobian-like quantity that admits the expression:

$$J(\mathbf{y}, \boldsymbol{\theta}) = D\left(\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}} G(\mathbf{u}, \boldsymbol{\theta})|_{\mathbf{u}=G^{-1}(\mathbf{y}, \boldsymbol{\theta})}\right)$$

with

$$D(\mathbf{A}) = \begin{cases} |\det(\mathbf{A})| & \text{if } n = p \\ \sum_{i=(i_1,\ldots,i_p)} |\det(\mathbf{A})_i| & \text{if } n \neq p, i: \text{ index of } p\text{-tuples} \\ \{\det(\mathbf{A}^T\mathbf{A})\}^{1/2} & \text{if } n \neq p \text{ and } \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}} G(\mathbf{u}, \boldsymbol{\theta}) \text{ has continuous partial derivatives w.r.t } \boldsymbol{\theta}. \end{cases},$$

where $\mathbf{A} \in \mathbb{R}^{n \times p}$. The derivation of (3) is not straightforward, but it was obtained by applying the implicit function theorem and Jacobian transformation. Note also its similarity with the Jeffreys prior in the Bayesian methodology.

Depending on the problem settings, it may not be feasible to obtain closed-form expressions for $J(\mathbf{y}, \boldsymbol{\theta})$ or $\int_{\boldsymbol{\Theta}} f(\mathbf{y}, \boldsymbol{\theta}') J(\mathbf{y}, \boldsymbol{\theta}') \mathrm{d}\boldsymbol{\theta}'$. Especially, with the large sample size, the number of combinations of tuples is too large to consider all possible combinations. This problem can sometimes be solved by sampling the combinations of tuples and taking the mean of them (e.g., Hannig et al. 2014). Some other times, one could use Markov Chain Monte Carlo (MCMC) methods to generate fiducial samples without the need to evaluate the normalizing constant $\int_{\boldsymbol{\Theta}} f(\mathbf{y}, \boldsymbol{\theta}') J(\mathbf{y}, \boldsymbol{\theta}') \mathrm{d}\boldsymbol{\theta}'$. In fact, as described below, for the present problem we need to use MCMC to generate a fiducial sample.

3. GENERALIZED FIDUCIAL INFERENCE FOR THRESHOLD ESTIMATION

As mentioned before, many dose–response studies assume a monotonic relationship between the dose levels x and the pharmacological effects y of the medicine (e.g., Bretz et al. 2008, 2010; Pinheiro et al. 2006; Mallik et al. 2011). Hence, one can consider the following regression model:

$$y = \mu(x) + \sigma u \tag{4}$$

with

$$\mu(x) = b_0 \text{ if } x \leq t^0 \quad \text{and} \quad \mu(x) > b_0 \text{ if } x > t^0,$$

where t^0 is the threshold, $b_0 \in \mathbb{R}$ is the baseline, u is the error term (usually standard normal), and σ is the noise standard deviation. Also, we use the following to denote the observed data from a study:

$$y_{ij} = \mu(x_i) + \sigma u_{ij}, \quad i = 0, 1, \ldots, n, \quad j = 1, 2, \ldots, m_i,$$

where y_{ij} is the measured pharmacological effect of i-th dosage from subject j, x_i is the i-th dosage (x_0 represents control), n is the number of dosages, m_i is the number of subjects with i-th dosage, and $u_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$

The goal is to conduct statistical inference on t^0, which is the MED in this dose–response setting. As mentioned in the introduction, we assume a constant value (β_0) for the first portion of $\mu(x)$ (i.e., $\mu(x)$ for $x \leq t^0$) while we use four different non-decreasing functions to model $\mu(x)$ for $x > t^0$:

$$\mu_p(x) = \beta_0 + \beta_1 (x - t^0)^{p-1} \mathbf{I}_{\{x > t^0\}}, \quad p = 1, 2, 3, 4, \tag{5}$$

where $\mathbf{I}_E$ is the indicator function for the event E.

At least for the cases of $p = 1$ and 2, these functions have been used by previous authors (Muggeo 2003; Mallik et al. 2011) to model $\mu(x)$. We incorporate $p = 3$ and 4 with the expectation that (5) is rich enough to capture the behaviors of the responses in most practical situations.

To proceed with inference on t^0, we need to calculate the generalized fiducial density $r(\boldsymbol{\theta}|\mathbf{y})$ for the above setup, where $\boldsymbol{\theta} = \{t^0, p, \sigma, \beta_0, \beta_1\}$ and $\mathbf{y} = (y_{ij}; \forall i, j)$.

Due to the indicator function, the four regression functions above do not satisfy a smoothness assumption required by Hannig (2013), and therefore (3) cannot be applied. To overcome this issue, the indicator function is replaced by a sigmoid function:

$$\frac{1}{1 + \exp\{-\lambda(x - t^0)\}} \quad \text{with} \quad \lambda = \frac{2 \log C}{\min\{|\, x_k - x_l \,|; k \neq l\}}, \tag{6}$$

where C is a large constant that we set it to be $C = 10^3$. In the sigmoid function, λ controls the slope near t^0. See Figure S.1 for some examples of (5) with the indicator function replaced by the sigmoid function.

The data generating equation G for the current problem is:

$$y_{ij} = G(u_{ij}, \boldsymbol{\theta}) = \beta_0 + \beta_1 (x_i - t^0)^{p-1} \frac{1}{1 + \exp\{-\lambda(x_i - t^0)\}} + \sigma u_{ij}, \quad u_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1), \tag{7}$$

for $p = 1, 2, 3, 4$. Now with this data generating equation (7), one could attempt to calculate the generalized fiducial density $r(\boldsymbol{\theta}|\mathbf{y})$ using (3). However, the analytic evaluation of the normalizing constant $\int_{\boldsymbol{\Theta}} f(\mathbf{y}, \boldsymbol{\theta}') J(\mathbf{y}, \boldsymbol{\theta}') \mathrm{d}\boldsymbol{\theta}'$ in (3) is infeasible. Consequently, we need to resort to MCMC methods for generating fiducial samples for $\boldsymbol{\theta}$, as to be described below.

3.1. GENERATING FIDUCIAL SAMPLES

This subsection presents an MCMC algorithm for generating a fiducial sample for $\boldsymbol{\theta} = \{t^0, p, \sigma, \beta_0, \beta_1\}$. We first provide a summary of the algorithm, then followed by the details. The steps of the algorithm are:

1. generate a $\tilde{t}^0$ using (8);
2. given $\tilde{t}^0$ obtained in the previous step, generate a $\tilde{p}$ from (9);
3. given $\tilde{t}^0$ and $\tilde{p}$ obtained in the previous steps, generate a $\tilde{\sigma}$ from (10);
4. given $\tilde{t}^0$, $\tilde{p}$ and $\tilde{\sigma}$ obtained in the previous steps, generate $\tilde{\beta}_0$ and $\tilde{\beta}_1$ using (11);
5. accept $\boldsymbol{\theta}_{\text{new}} = (\tilde{t}^0, \tilde{p}, \tilde{\sigma}, \tilde{\beta}_0, \tilde{\beta}_1)$ if $h(\boldsymbol{\theta}_{\text{new}}, \boldsymbol{\theta}_{\text{old}})$ in (12) is larger than a realization from Uniform(0, 1); otherwise, set $\boldsymbol{\theta}_{\text{new}} = \boldsymbol{\theta}_{\text{old}}$.

Now, we present the detailed steps.

Generating t^0: To generate a value for t^0, we use the truncated normal distribution with x_0 and x_n as, respectively, the lower and upper bounds:

$$t^0_{\text{new}} \sim \mathcal{N}\left(t^0_{\text{old}}, \sigma_t^2\right), \quad x_0 \le t^0_{\text{old}} \le x_n, \quad x_0 \le t^0_{\text{new}} \le x_n, \quad \sigma_t^2 > 0, \tag{8}$$

where t^0_{old} is the t^0 value from the previous iteration, and σ_t^2 is used to perturb its value. In our practical implementation, we set $\sigma_t = (x_n - x_0)/10$ and chose the initial value of t^0 at random uniformly from $[x_0, x_n]$.

Generating p **given** t^0: Suppose now we have a value for t^0. We can calculate the conditional generalized fiducial density for each of the four models $\mu_p(x)$ in (5), indexed by $p = 1, 2, 3, 4$. Let $\boldsymbol{\xi} = \{\beta_0, \beta_1, \sigma\} \in \boldsymbol{\Xi}$ and denote the conditional generalized fiducial density as $r(p|t^0, \mathbf{y})$:

$$r(p|t^0, \mathbf{y}) = \frac{\int_{\boldsymbol{\Xi}} f_p(\mathbf{y}, \boldsymbol{\xi}_p) J_p(\mathbf{y}, \boldsymbol{\xi}_p) \mathrm{d}\boldsymbol{\xi}_p}{\sum_{p'=1}^4 \int_{\boldsymbol{\Xi}} f_{p'}(\mathbf{y}, \boldsymbol{\xi}_{p'}) J_{p'}(\mathbf{y}, \boldsymbol{\xi}_{p'}) \mathrm{d}\boldsymbol{\xi}_{p'}}.$$

Direct calculations give

$$J_p(\mathbf{y}, \boldsymbol{\xi}_p) = \sigma^{-1} |\det(\mathbf{X}_p' \mathbf{X}_p)|^{\frac{1}{2}} \mathrm{RSS}_p^{\frac{1}{2}}$$

and

$$\int_{\boldsymbol{\Xi}} f_p(\mathbf{y}, \boldsymbol{\xi}_p) J_p(\mathbf{y}, \boldsymbol{\xi}_p) \mathrm{d}\boldsymbol{\xi}_p = \Gamma\left(\frac{N}{2}\right) \left(\frac{1}{\pi}\right)^{\frac{N}{2}-1} \mathrm{RSS}_p^{\frac{1-N}{2}},$$

where $\mathbf{X}_p$ is the design matrix for $\mu_p(x)$, RSS_p is the corresponding residual sum of squares, and N is total number of observations ($N = \sum_i m_i$). Therefore,

$$r(p|t^0, \mathbf{y}) = \frac{\mathrm{RSS}_p^{\frac{1-N}{2}}}{\sum_{p'=1}^4 \mathrm{RSS}_{p'}^{\frac{1-N}{2}}}. \tag{9}$$

Thus, given a value for t^0, using (9) one can calculate the fiducial probabilities for the four models in (5), and from these probabilities, one can generate a model out of the four, or equivalently, generate a value for p.

Generating σ^2 given (t^0, p): Next, given t^0 and p, it can be shown that the conditional generalized fiducial distribution of σ^2 is inverse χ^2:

$$\sigma^2 \mid t^0, p, \mathbf{y} \sim \text{Inv-}\chi^2(n-3, s^2), \tag{10}$$

where

$$s^2 = (\mathbf{y} - \mathbf{X}_p\hat{\boldsymbol{\beta}}_p)^T(\mathbf{y} - \mathbf{X}_p\hat{\boldsymbol{\beta}}_p)/(n-3) \quad \text{with} \quad \hat{\boldsymbol{\beta}}_p = (\mathbf{X}_p^T\mathbf{X}_p)^{-1}\mathbf{X}_p^T\mathbf{y}.$$

Therefore, σ^2 can be sampled from (10).

Generating $\boldsymbol{\beta}_p = (\beta_0, \beta_1)$ given (t^0, p, σ): It is straightforward to show that the generalized fiducial distribution of $\boldsymbol{\beta}_p$ condition on t^0, p, σ^2 is

$$\boldsymbol{\beta}_p \mid t^0, p, \sigma^2, \mathbf{y} \sim \mathcal{N}\left(\hat{\boldsymbol{\beta}}_p, \sigma^2(\mathbf{X}_p^T\mathbf{X}_p)^{-1}\right), \tag{11}$$

from which a value for $\boldsymbol{\beta}_p$ can be sampled from.

Accept/Reject: Executing the above steps will produce a value for $\boldsymbol{\theta} = (t^0, p, \sigma, \beta_0, \beta_1)$ which we need to decide if it should be accepted or rejected. To do so, we need the generalized fiducial density $r(\boldsymbol{\theta}|\mathbf{y})$:

$$r(\boldsymbol{\theta}|\mathbf{y}) \propto r(p|t^0, \mathbf{y}) \exp\left\{-\frac{(\mathbf{y} - \mathbf{X}_p\boldsymbol{\beta}_p)^T(\mathbf{y} - \mathbf{X}_p\boldsymbol{\beta}_p)}{2\sigma^2}\right\} \frac{1}{\sigma} \sum_i |\det(\mathbf{A}_p)_i|,$$

where

$$\mathbf{A}_p = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ 1 & \frac{(x_i - t^0)^{p-1}}{1+\exp\{-\lambda(x_i - t^0)\}} & -\left(\frac{\beta_{p,1}(p-1)(x_i - t^0)^{(p-2)}}{1+\exp\{-\lambda(x_i - t^0)\}} + \frac{\lambda\beta_{p,1}(x_i - t^0)^{(p-1)}\exp\{-\lambda(x_i - t^0)\}}{[1+\exp\{-\lambda(x_i - t^0)\}]^2}\right) & u_{ij} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}_{(N\times 4)}.$$

To speed up the computations, we use the sampling technique proposed in Hannig et al. (2014) to approximate the summation $\sum_i |\det(\mathbf{A}_p)_i|$.

Denote those new parameter values generated using the above steps as $\boldsymbol{\theta}_{\text{new}}$, and those parameter values in the previous MCMC iteration as $\boldsymbol{\theta}_{\text{old}}$. Let $T(\boldsymbol{\theta}_{\text{new}}|\boldsymbol{\theta}_{\text{old}})$ be the proposal distribution which is the product of (10), (11) and (8). If the ratio

$$h(\boldsymbol{\theta}_{\text{new}}, \boldsymbol{\theta}_{\text{old}}) = \frac{r(\boldsymbol{\theta}_{\text{new}}|\mathbf{y})T(\boldsymbol{\theta}_{\text{old}}|\boldsymbol{\theta}_{\text{new}})}{r(\boldsymbol{\theta}_{\text{old}}|\mathbf{y})T(\boldsymbol{\theta}_{\text{new}}|\boldsymbol{\theta}_{\text{old}})} \tag{12}$$

is larger than a realization from Uniform(0, 1), we accept $\boldsymbol{\theta}_{\text{new}}$. Otherwise, we use $\boldsymbol{\theta}_{\text{old}}$ again for the next iteration.

Repeating the above procedure, a fiducial sample for $\boldsymbol{\theta}$ can be generated, and the methods implemented by Plummer et al. (2006) can be applied to determine if the sample size is

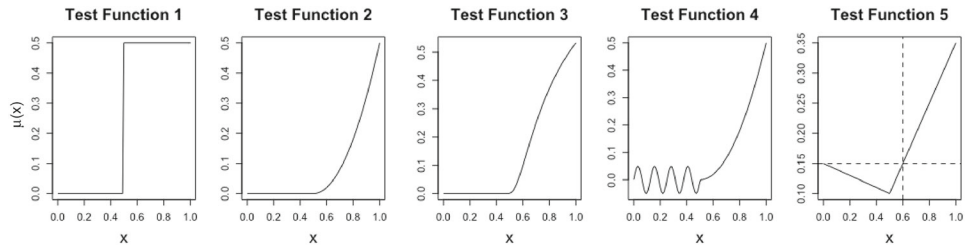

Figure 1. Plots of the five test functions used in the numerical experiments .

large enough. Once a sufficiently large fiducial sample is available, statistical inference can then be conducted. For example, one could use the average of all $\tilde{t}^0$'s as an estimate of t^0. Also, for a 95% confidence interval, one could take the quantiles 2.5% and 97.5 % of the $\tilde{t}^0$'s.

Lastly, we stress that in our approach more than one model (see (5)) are used to generate the above fiducial sample $\tilde{t}^0$'s, and therefore reducing the risk of incorrect model selection.

4. NUMERICAL EXPERIMENTS

This section studies the empirical performance of the proposed method through a sequence of numerical experiments. In particular, the proposed method is compared to two existing methods in the literature: the MLE break point estimation method of Muggeo (2003) and the p value method of Mallik et al. (2011). Further details of these two methods can be found in the introduction. In sequel, we shall refer the proposed method as GFI, the method of Mallik et al. (2011) as p value, and the method of Muggeo (2003) as MLE. The code for GFI can be downloaded from https://github.com/vic-dragon/GFI_threshold

Altogether five different test functions were used for $\mu(x)$:

- Test Function 1: $\mu(x) = 0.5 \times \mathbf{I}_{\{x \geq 0.5\}}$,
- Test Function 2: $\mu(x) = 2(x - 0.5)^2 \times \mathbf{I}_{\{x \geq 0.5\}}$,
- Test Function 3: $\mu(x) = \exp\left\{-\frac{\log(2)}{2(x-0.45)}\right\} \times \mathbf{I}_{\{x \geq 0.5\}}$,
- Test Function 4: $\mu(x) = 0.05\sin(50x)\mathbf{I}_{\{x<0.5\}} + 2(x - 0.5)^2 \times \mathbf{I}_{\{x \geq 0.5\}}$,
- Test Function 5: $\mu(x) = 0.1 - 0.1(x - 0.5)\mathbf{I}_{\{x<0.5\}} + 0.5(x - 0.5) \times \mathbf{I}_{\{x \geq 0.5\}}$.

These functions are displayed in Fig. 1. Notice that only the first two functions satisfy the parametric assumption (5). Also notice that all five functions share the same threshold value $t^0 = 0.5$, although their shapes are quite different. However, for Test Function 5, one should treat $t^0 = 0.6$ as the threshold value (instead of 0.5), as it is the point that starts showing a discernible effect compared to the baseline (Agathokleous et al. 2019). Some of these test functions have been used by other authors (e.g., Muggeo 2003; Mallik et al. 2011).

The data were generated as follows. First, we set $x_i = i/(n + 1)$ for $i = 1, \ldots, n$. Then, we calculated $y_{ij} = \mu(x_i) + \epsilon_{ij}$, where $\mu(x)$ is one of the five test functions,

$\epsilon_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, for $j = 1, \ldots, m$ and $i = 1, \ldots, n$. Two values were used for $\sigma = \{0.1, 0.3\}$, while five different combinations of values were used for $(m, n) = \{(6, 5), (10, 10), (10, 20), (20, 50), (50, 100)\}$. We also investigated the performances of the methods under heteroscedastic noise: we generated data with the gradually increasing noise variance $(\sigma + 0.2(x - 0.5)\mathbf{I}_{\{x \geq 0.5\}})$.

For each combination of experimental parameters, 200 data sets were generated. For each of these data sets, the three methods mentioned above were applied to estimate the threshold value $t^0 = 0.5$ for Test Functions 1 to 4 and $t^0 = 0.6$ for Test Function 5. The averaged biases and mean-squared-errors (MSEs) were calculated. Also, for the GFI and MLE methods, we constructed 95% confidence intervals for the threshold value. The empirical coverage rates (ECRs) and their averaged widths of these confidence intervals were calculated. The results can be found in Tables 1, 2, S.1, and S.2. Note that the p value method does not produce confidence intervals and hence, is omitted for this comparison.

With the largest sample size that we tested ($n = 50, m = 100$), the GFI method typically took less than 2 min to process one data set with a Dell PowerEdge R360 server, 2015 model. The acceptance rates of the MCMC algorithm depend on the experimental configurations. They ranged from 0.022 to 0.572 for homoscedastic noise, and from 0.024 to 0.559 for heteroscedastic noise. More detailed information can be found in Tables S.3 and S.4.

Recall that the MLE method does not always produce an estimate $\hat{t}^0$ for t^0 (see Sect. 1). Thus, all the results in the previous tables are based on those cases where $\hat{t}^0$ were available. The percentages of time that MLE produced estimates are given in Tables S.5 and S.6. These percentages ranged from 42% to 100%, with most cases larger than 90%.

From the experimental results, the following empirical conclusions can be drawn. First, in terms of point estimation, no method is uniformly the best. The relative performances of the three methods depend largely on the test function. Nevertheless, it seems GFI provided the best results for a majority of the experimental configurations (72% of the experimental configurations for homoscedastic noise and in 66% of the experimental configurations for heteroscedastic noise), and that p value suffered from bias issues. For confidence intervals, one cannot conclude if GFI or MLE is uniformly better, although GFI gave better results in 88% of the experimental configurations for homoscedastic noise and in 96% of the experimental configurations for heteroscedastic noise. Therefore, in practice, our recommendation is that, if nothing is known about the shape of the underlying function, GFI would be the first method to apply.

Recall that the proposed GFI method uses four different parametric functions to model the underlying true function (see (5)), and that the fiducial samples for p were sampled from the conditional distribution (9). Figures S.2 to S.5 report the relative percentages of the sampled values of p. As Test Functions 1 and 2 are special cases of (5), one can see that when m and n are large, GFI has a high tendency of selecting the true model.

5. APPLICATIONS WITH REAL DATA

In this section, two real data sets are analyzed with the proposed GFI method. We note that for one data set the monotonicity is in the opposite direction to model (5), while for the

Table 1. Simulation results for the homoscedastic noise cases with $\sigma = 0.1$

Test function	(n,m)	Bias			MSE			ECR	
		GFI	MLE	p value	GFI	MLE	p value	GFI	MLE
1	**(6,5)**	**0.15 (0.18)**	6.47 (1.59)	−6.15 (0.88)	**0.06 (0.01)**	2.54 (0.14)	1.90 (0.29)	**0.98 (0.11)**	0.65 (0.46)
	(10,10)	**0.03 (0.10)**	−3.28 (1.60)	−2.56 (0.50)	**0.02 (0.00)**	5.16 (0.13)	0.56 (0.13)	**0.98 (0.06)**	0.10 (0.27)
	(10,20)	**−0.08 (0.09)**	−3.43 (1.59)	−2.78 (0.48)	**0.02 (0.00)**	5.18 (0.10)	0.53 (0.11)	**0.97 (0.06)**	0.02 (0.20)
	(20,50)	**−0.01 (0.05)**	3.37 (1.66)	−1.19 (0.26)	**0.00 (0.00)**	5.59 (0.05)	0.15 (0.03)	**0.98 (0.03)**	0.00 (0.09)
	(50,100)	**0.00 (0.02)**	−0.01 (1.78)	−0.47 (0.11)	**0.00 (0.00)**	5.97 (0.02)	0.03 (0.01)	**0.96 (0.01)**	0.00 (0.04)
2	**(6,5)**	11.94 (0.48)	7.90 (0.78)	**7.21 (1.01)**	1.89 (0.09)	**1.66 (0.10)**	2.55 (0.23)	**0.94 (0.47)**	0.58 (0.32)
	(10,10)	**2.94 (0.86)**	14.18 (0.46)	8.42 (0.70)	**1.57 (0.13)**	2.42 (0.13)	1.69 (0.14)	**0.96 (0.52)**	0.16 (0.14)
	(10,20)	**−0.56 (0.72)**	14.13 (0.37)	6.68 (0.64)	**1.03 (0.10)**	2.26 (0.11)	1.27 (0.09)	**0.98 (0.43)**	0.06 (0.10)
	(20,50)	**−3.96 (0.40)**	15.23 (0.16)	6.91 (0.35)	**0.47 (0.04)**	2.37 (0.05)	0.72 (0.04)	**0.94 (0.24)**	0.00 (0.04)
	(50,100)	**−0.92 (0.17)**	16.05 (0.08)	6.44 (0.21)	**0.07 (0.02)**	2.59 (0.02)	0.50 (0.02)	**0.95 (0.07)**	0.00 (0.02)
3	**(6,5)**	7.57 (0.66)	1.44 (0.52)	**−1.09 (0.96)**	1.44 (0.08)	**0.55 (0.05)**	1.86 (0.26)	0.84 (0.40)	**0.89 (0.25)**
	(10,10)	**−0.35 (0.44)**	0.60 (0.27)	1.66 (0.63)	0.39 (0.05)	**0.15 (0.02)**	0.82 (0.13)	**0.94 (0.25)**	0.92 (0.13)
	(10,20)	**0.42 (0.24)**	0.99 (0.16)	0.87 (0.58)	0.11 (0.02)	**0.06 (0.01)**	0.67 (0.10)	0.91 (0.12)	**0.96 (0.09)**
	(20,50)	**0.51 (0.06)**	0.52 (0.07)	2.06 (0.30)	**0.01 (0.00)**	0.01 (0.00)	0.22 (0.03)	0.91 (0.03)	**0.94 (0.04)**
	(50,100)	**0.29 (0.03)**	0.32 (0.03)	1.79 (0.17)	**0.00 (0.00)**	0.00 (0.00)	0.09 (0.01)	0.83 (0.01)	**0.90 (0.02)**
4	**(6,5)**	14.69 (0.34)	9.64 (0.66)	**−6.03 (1.71)**	2.39 (0.09)	**1.70 (0.10)**	6.15 (0.41)	**0.91 (0.45)**	0.55 (0.30)
	(10,10)	**1.20 (0.81)**	13.71 (0.56)	10.44 (0.52)	**1.32 (0.12)**	2.51 (0.15)	1.62 (0.11)	**0.99 (0.52)**	0.23 (0.15)
	(10,20)	**−2.54 (0.64)**	13.43 (0.41)	9.27 (0.39)	**0.87 (0.08)**	2.14 (0.11)	1.17 (0.08)	**0.99 (0.42)**	0.09 (0.11)
	(20,50)	**−2.66 (0.32)**	14.02 (0.17)	7.41 (0.27)	**0.27 (0.03)**	2.02 (0.05)	0.69 (0.04)	**0.97 (0.22)**	0.00 (0.04)
	(50,100)	**−0.21 (0.08)**	14.79 (0.09)	6.70 (0.18)	**0.01 (0.01)**	2.20 (0.03)	0.51 (0.02)	**0.97 (0.04)**	0.00 (0.02)
5	**(6,5)**	**−1.06 (0.45)**	−9.14 (0.84)	−48.36 (0.15)	**0.42 (0.05)**	1.98 (0.18)	23.43 (0.15)	**0.99 (0.46)**	0.83 (0.50)
	(10,10)	**−2.32 (0.68)**	−10.53 (0.76)	−52.38 (0.11)	**0.99 (0.09)**	2.24 (0.20)	27.46 (0.12)	**0.94 (0.51)**	0.56 (0.25)
	(10,20)	**−5.45 (0.69)**	−10.37 (0.51)	−52.13 (0.10)	**1.24 (0.11)**	1.58 (0.13)	27.20 (0.11)	**0.92 (0.46)**	0.44 (0.19)
	(20,50)	**−6.63 (0.19)**	−10.10 (0.17)	−55.92 (0.07)	**0.51 (0.03)**	1.08 (0.03)	31.28 (0.08)	**0.18 (0.14)**	0.01 (0.08)
	(50,100)	**−5.32 (0.07)**	−9.93 (0.07)	−58.31 (0.03)	**0.29 (0.01)**	0.99 (0.01)	34.00 (0.03)	**0.00 (0.04)**	0.00 (0.04)

Reported are the averages ($\times 10^2$) and standard errors ($\times 10^2$, in parentheses) of the biases and the MSEs for the threshold estimates obtained by the GFI, MLE, and p value methods (smallest values are bolded). Also reported are the empirical coverage rates (ECRs) and the average widths (in parentheses) of the 95% confidence intervals produced by the GFI and MLE methods (ECRs that are closest to the nominal target rate of 95% are bolded). Recall that n is the number of dosages and m is the number of subjects for each dosage

Table 2. Simulation results for the homoscedastic noise cases with $\sigma = 0.3$

Test function	(n,m)	Bias			MSE			ECR	
		GFI	MLE	p value	GFI	MLE	p value	GFI	MLE
1	**(6,5)**	**−0.82 (0.27)**	2.87 (1.22)	−6.65 (0.91)	**0.16 (0.02)**	1.59 (0.12)	2.11 (0.31)	**0.98 (0.25)**	0.89 (0.88)
	(10,10)	**−0.22 (0.14)**	−3.58 (1.59)	−2.99 (0.51)	**0.04 (0.01)**	4.65 (0.20)	0.60 (0.13)	**1.00 (0.10)**	0.58 (0.49)
	(10,20)	**−0.13 (0.12)**	−1.32 (1.62)	−3.51 (0.54)	**0.03 (0.00)**	4.90 (0.18)	0.70 (0.16)	**0.97 (0.08)**	0.4 (0.35)
	(20,50)	**0.00 (0.06)**	0.33 (1.77)	−1.57 (0.32)	**0.01 (0.00)**	6.21 (0.18)	0.23 (0.06)	**0.97 (0.03)**	0 (0.17)
2	**(50,100)**	**0.02 (0.02)**	−0.29 (1.78)	−0.75 (0.13)	**0.00 (0.00)**	5.96 (0.07)	0.04 (0.01)	**0.95 (0.01)**	0 (0.09)
	(6,5)	**2.39 (0.60)**	3.04 (1.16)	16.07 (1.2)	**0.78 (0.07)**	1.55 (0.13)	5.45 (0.31)	**0.97 (0.50)**	0.87 (0.86)
	(10,10)	12.71 (0.71)	**8.94 (1.26)**	16.20 (0.78)	**2.63 (0.17)**	3.63 (0.22)	3.83 (0.21)	**0.92 (0.58)**	0.55 (0.40)
	(10,20)	12.38 (0.83)	**10.89 (0.95)**	13.57 (0.77)	2.91 (0.18)	**2.90 (0.19)**	3.01 (0.17)	**0.87 (0.56)**	0.53 (0.31)
	(20,50)	**−1.37 (0.72)**	15.13 (0.45)	12.16 (0.46)	**1.05 (0.10)**	2.68 (0.14)	1.89 (0.09)	**0.98 (0.49)**	0.09 (0.13)
	(50,100)	**−4.59 (0.51)**	15.98 (0.21)	10.32 (0.30)	**0.72 (0.06)**	2.64 (0.07)	1.24 (0.06)	**0.94 (0.29)**	0.00 (0.06)
3	**(6,5)**	6.91 (0.50)	**1.62 (0.88)**	4.59 (1.09)	**0.98 (0.07)**	1.18 (0.10)	2.56 (0.25)	**0.98 (0.47)**	0.89 (0.69)
	(10,10)	6.15 (0.68)	**−0.80 (0.96)**	6.08 (0.66)	1.31 (0.10)	1.81 (0.17)	**1.24 (0.13)**	**0.89 (0.5)**	0.78 (0.37)
	(10,20)	3.57 (0.73)	**−0.54 (0.68)**	4.05 (0.60)	1.20 (0.09)	0.92 (0.11)	**0.89 (0.07)**	**0.87 (0.45)**	0.82 (0.28)
	(20,50)	−1.45 (0.34)	**0.20 (0.23)**	3.89 (0.37)	0.24 (0.05)	**0.11 (0.01)**	0.43 (0.04)	**0.99 (0.21)**	0.95 (0.13)
	(50,100)	**0.23 (0.08)**	0.38 (0.10)	3.83 (0.20)	**0.01 (0.00)**	0.02 (0.00)	0.23 (0.01)	0.91 (0.04)	**0.94 (0.06)**
4	**(6,5)**	**2.36 (0.63)**	4.32 (1.06)	14.33 (1.38)	**0.84 (0.07)**	1.43 (0.13)	5.84 (0.33)	**0.97 (0.50)**	0.87 (0.80)
	(10,10)	12.47 (0.67)	**9.69 (1.10)**	17.47 (0.65)	**2.45 (0.17)**	3.11 (0.21)	3.90 (0.21)	**0.91 (0.58)**	0.62 (0.42)
	(10,20)	11.92 (0.81)	**10.65 (0.87)**	14.84 (0.60)	2.72 (0.18)	**2.56 (0.19)**	2.92 (0.18)	**0.89 (0.56)**	0.60 (0.32)
	(20,50)	**−0.68 (0.68)**	13.99 (0.46)	12.72 (0.39)	**0.92 (0.09)**	2.38 (0.14)	1.92 (0.09)	**0.99 (0.48)**	0.15 (0.13)
	(50,100)	**−2.97 (0.45)**	14.86 (0.23)	10.54 (0.28)	**0.49 (0.04)**	2.31 (0.07)	1.26 (0.06)	**0.95 (0.28)**	0.01 (0.06)
5	**(6,5)**	−9.89 (0.58)	**−7.91 (1.14)**	−24.94 (1.72)	**1.65 (0.14)**	2.01 (0.24)	12.12 (0.73)	0.91 (0.48)	**0.93 (0.97)**
	(10,10)	**−2.97 (0.71)**	−10.40 (1.42)	−40.59 (1.27)	**1.10 (0.15)**	4.35 (0.36)	19.67 (0.77)	**0.96 (0.57)**	0.70 (0.56)
	(10,20)	**0.40 (0.63)**	−10.02 (1.19)	−50.01 (0.52)	**0.78 (0.09)**	3.53 (0.31)	25.55 (0.38)	**0.97 (0.55)**	0.70 (0.48)
	(20,50)	**−4.99 (0.78)**	−10.49 (0.77)	−55.9 (0.07)	**1.45 (0.12)**	2.27 (0.27)	31.26 (0.07)	**0.93 (0.52)**	0.58 (0.25)
	(50,100)	**−8.19 (0.32)**	−9.67 (0.25)	−58.32 (0.03)	**0.87 (0.08)**	1.05 (0.05)	34.02 (0.04)	**0.39 (0.25)**	0.09 (0.11)

Reported are the averages ($\times 10^2$) and standard errors ($\times 10^2$, in parentheses) of the biases and the MSEs for the threshold estimates obtained by the GFI, MLE, and p value methods (smallest values are bolded). Also reported are the empirical coverage rates (ECRs) and the average widths (in parentheses) of the 95% confidence intervals produced by the GFI and MLE methods (ECRs that are closest to the nominal target rate of 95% are bolded). Recall that n is the number of dosages and m is the number of subjects for each dosage

other data set the covariate values x_i's are reversed. Nevertheless, the proposed method can be applied with minor modifications.

The first data set is related to a dose response experiment originated from a biological study of the effect of the 1-methyl-3-butylimidazolium tetrafluoroborate treatment on the IPC-81 rat cell lines; see Ranke et al. (2004). The purpose of this study is to assess environmental hazards through measuring toxicity of ionic liquids on the mammalian cell lines. It is of particular interest to estimate the lethal dose level which makes cell lines stop responding. The cell responses were measured as percentages of cell viability compared to control, taken at different dose levels (measured in μM) of the treatment. The sample size is $N = n \times m = 9 \times 9 = 81$. The GFI method provides a threshold value estimate of $\hat{t}^0 = 5.22\ \mu M$, with a 95% confidence interval as $(4.92\ \mu M, 5.34\ \mu M)$. The MLE method provides a threshold value estimate of $\hat{t}^0 = 5.10\ \mu M$, with a 95% confidence interval as $(4.58\ \mu M, 5.62\ \mu M)$, while the p value method provides a threshold value estimate of $\hat{t}^0 = 5.52\ \mu M$. The observed data, as well as these results, are displayed in the left panel of Fig. 2. One can see that, after the GFI estimated threshold $5.22\ \mu M$, the cell culture dose not show any significant response. Also, the GFI point estimate for t^0 is not in the middle of the confidence interval. Instead, it is closer to the right side, which is very reasonable when inspecting the left panel of Fig. 2. Lastly, as 100% of the fiducial sampled values from $p = 2$, it suggests that the underlying relationship can be well modeled with a piecewise linear function.

The second data set concerns a light detection and ranging (LiDAR) experiment. LiDAR is a technique for measuring the concentrations of various atmospheric particulates. Very often, it is used to monitor the air pollution status in an area. This second data set contains atmospheric atomic mercury measurements taken at the geothermal power plan Bella Vista in Italy (e.g., Holst et al. 1996; Mallik et al. 2011). There are 221 observations. The predictor is the distance from the measuring equipment to the source that reflected the light, with values ranging from 390 m to 720 m. The response is the logarithm of the received light ratio from two different wavelengths; see the right panel of Fig. 2. The decreasing trend starting around range=540 m indicates the existence of mercury. And it is of interest in knowing exactly when this decreasing trend started. The GFI method gives a threshold value estimate of $\hat{t}^0 = 543.55$ m, with a 95% confidence interval as (541.47 m, 548.92 m). The MLE method provides a threshold value estimate of $\hat{t}^0 = 544.66$ m, with a 95% confidence interval as (538.04 m, 551.29 m), while the p value method provides a threshold value estimate of $\hat{t}^0 = 541$ m. These results are also displayed in the right panel of Fig. 2. As similar to above, the point estimate from GFI is not in the middle of the confidence interval. In addition, 24% and 76% of the fiducial sampled values for p are 1 and 2, respectively. This suggests that the underlying relationship may not be simply modeled by either a piecewise step function or linear function.

From Fig. 2, for both real data examples, one can see that the GFI and MLE methods gave very similar point estimates for t^0, while the GFI confidence intervals are narrower than those from MLE. The point estimates from the p value method are outside the GFI confidence intervals: it seems to suggest that the p value estimates are biased, which are in agreement with the simulation results.

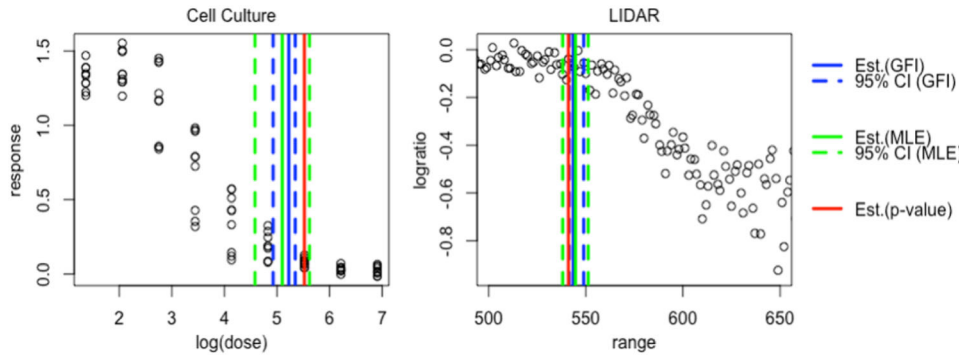

Figure 2. Results obtained by applying the three methods GFI (Blue), MLE (Green), and p value (Red) to two real data sets. Circles are the observed data, while the solid lines and the dashed lines represent the estimates and the 95% confidence intervals, respectively .

6. CONCLUDING REMARKS

In this paper, we developed a method for estimating threshold under the a general setting used in many biomedical studies such as toxicology and dose–response studies. The proposed method is based on the generalized fiducial inference framework. Two noteworthy features of the proposed method are (i), in addition to point estimates, it also provides confidence intervals for the parameters of interest, and (ii) it allows the possibility of having multiple parametric models to capture the underlying pattern of the data so that the risk of model mis-specification is reduced. Results from numerical experiments and applications to real data suggest that the proposed method possesses excellent empirical properties.

Some extensions of the current work are worth considering. One possibility would be to adopt a richer set of non-decreasing functions to model $\mu(x)$, as opposed to (5). This could be done quite straightforwardly. Another possibility would be to allow the occurrence of multiple change points, where the number of the change points is unknown a priori. However, this extension could cause some complications, as for example, one potentially would need to handle the issue that different fiducial samples could give different number of change points. At the moment, if there is no change point, our numerical experience suggests that our method will often return x_1 as the estimated threshold, while its performance is less predictable if there are multiple change points.

One could also investigate the theoretical properties of the proposed method. For example, if one is willing to make the unrealistic assumption that t^0 is given, one can show that as the sample size increases, the probability of selecting the correct model from (5) goes to 1, which leads to the following proposition.

Proposition 1. *Assume the technical conditions in* Gao et al. (2020) *and that* t^0 *is given. Denote the true value of* p *in* (5) *as* p^0. *Then, the fiducial probability* (9) *follows* $r(p^0|t^0, \mathbf{y}) \to 1$ *as* $n \to \infty$.

This proposition can be proved following the arguments of Gao et al. (2020). Essentially, it involves showing that $\mathrm{RSS}_{p'}/\mathrm{RSS}_{p^0} \to 0$ for any $p' \neq p^0$, which leads to

$$r(p^0|t^0, \mathbf{y}) = \frac{\mathrm{RSS}_{p^0}^{\frac{1-N}{2}}}{\sum_{p'=1}^{4} \mathrm{RSS}_{p'}^{\frac{1-N}{2}}} = \frac{1}{1 + \sum_{p' \neq p^0} \left(\mathrm{RSS}_{p'}/\mathrm{RSS}_{p^0}\right)^{\frac{1-N}{2}}} \to 1 \quad \text{as} \quad n \to \infty.$$

With Proposition 1, the results of Hannig et al. (2016) also guarantee that inferences made by GFI enjoy asymptotically correct frequentist properties, i.e., confidence intervals constructed by $r(\boldsymbol{\theta}|\mathbf{y})$ will achieve the nominated coverage levels when the sample size is large enough.

However, the case of unknown t^0 is substantially more complicated and we plan to report our results in a future paper.

ACKNOWLEDGEMENTS

The authors are most grateful to the reviewers, the associate editor, and the editor for their constructive and helpful comments that led to a much improved version of the paper. The fund was provided by National Science Foundation (Grant Nos. DMS-1811405, DMS-1811661, DMS-1916125 and CCF-1934568)

Declarations

Conflict of interest The author declare that they have no conflict of interest.

Human and animal rights The work presented in this article does not involve human participants and/or animals.

Informed consent No informed consent is needed as the data sets used in the paper are publicly available.

[Received July 2020. Revised August 2021. Accepted August 2021. Published Online September 2021.]

REFERENCES

Agathokleous E, Belz RG, Calatayud V, De Marco A, Hoshika Y, Kitao M, Saitanis CJ, Sicard P, Paoletti E, Calabrese EJ (2019) Predicting the effect of ozone on vegetation via linear non-threshold (lnt), threshold and hormetic dose-response models. Sci Total Environ 649:61–74

Bretz F, Dette H, Pinheiro JC (2010) Practical considerations for optimal designs in clinical dose finding studies. Stat Med 29:731–742

Bretz F, Hsu J, Pinheiro J, Liu Y (2008) Dose finding-a challenge in statistics. Biom J 50:480–504

Cheng JQ, Liu RY, Xie M-G (2014) Fusion learning. Statistics Reference Online, Wiley StatsRef, pp 1–8

Coffey T, Gennings C (2007) The simultaneous analysis of mixed discrete and continuous outcomes using nonlinear threshold models. J Agric Biol Environ Stat 12:55

Cox C (1987) Threshold dose-response models in toxicology. Biometrics 43:511–523

Cui Y, Hannig J (2019) Nonparametric generalized fiducial inference for survival functions under censoring. Biometrika 106:501–518. https://doi.org/10.1093/biomet/asz016

Dempster AP (2008) The Dempster–Shafer calculus for statisticians. Int J Approx Reason 48:365–377

Efron B (2013) Bayes' theorem in the 21st century. Science 340:1177–1178

Fisher RA (1930) Inverse probability. In: Mathematical proceedings of the Cambridge philosophical society, vol 26. Cambridge University Press, pp 528–535

Gao Q, Lai RCS, Lee TCM, Li Y (2020) Uncertainty quantification for high-dimensional sparse nonparametric additive models. Technometrics 62:513–524

Guo B, Li Y (2014) Bayesian designs of phase II oncology trials to select maximum effective dose assuming monotonic dose-response relationship. BMC Med Res Methodol 14:95

Hannig J (2013) Generalized fiducial inference via discretization. Stat Sin 23:489–514

Hannig J, Iyer H, Lai RCS, Lee TCM (2016) Generalized fiducial inference: a review and new results. J Am Stat Assoc 111:1346–1361

Hannig J, Lai RCS, Lee TCM (2014) Computational issues of generalized fiducial inference. Comput Stat Data Anal 71:849–858

Hannig J, Lee TCM (2009) Generalized fiducial inference for wavelet regression. Biometrika 96:847–860

Holst U, Hossjer O, Bjorklund C, Ragnarson P, Edner H (1996) Locally weighted least squared kernel regression and statistical evaluation of lidar measurements. Environmetrics 7:401–416

Hsu JC, Berger RL (1999) Stepwise confidence intervals without multiplicity adjustment for dose response and toxicity studies. J Am Stat Assoc 94:468–482

Kim SB, Bartell SM, Gillen DL (2016) Inference for the existence of hormetic dose-response relationships in toxicology studies. Biostatistics 17:523–536. https://doi.org/10.1093/biostatistics/kxw004

Lai RCS, Hannig J, Lee TCM (2015) Generalized fiducial inference for ultrahigh-dimensional regression. J Am Stat Assoc 110:760–772

Lutz WK, Lutz RW (2009) Statistical model to estimate a threshold dose and its confidence limits for the analysis of sublinear dose-response relationships, exemplified for mutagenicity data. Mutat Res Genetic Toxicol Environ Mutagenesis 678:118–122

Mallik A, Sen B, Banerjee M, Michailidis G (2011) Threshold estimation based on a p-value framework in dose-response and regression settings. Biometrika 98:887–900

Martin R, Liu C (2015) Inferential models: reasoning with uncertainty, vol 145. CRC Press, Boca Raton

Muggeo VM (2003) Estimating regression models with unknown break-points. Stat Med 22:3055–3071

Otava M, Shkedy Z, Hothorn LA, Talloen W, Gerhard D, Kasim A (2017) Identification of the minimum effective dose for normally distributed data using a Bayesian variable selection approach. J Biopharm Stat 27:1073–1088

Pinheiro JC, Bretz F, Branson M (2006) Analysis of dose–response studies–modeling approaches. In: Dose finding in drug development. Springer, pp 146–171

Plummer M, Best N, Cowles K, Vines K (2006) CODA: convergence diagnosis and output analysis for MCMC. R news 6:7–11

Ranke J, Mölter K, Stock F, Bottin-Weber U, Poczobutt J, Hoffmann J, Ondruschka B, Filser J, Jastorff B (2004) Biological effects of imidazolium ionic liquids with varying chain lengths in acute vibrio fischeri and WST-1 cell viability assays. Ecotoxicol Environ Saf 58:396–404

Ruberg SJ (1989) Contrasts for identifying the minimum effective dose. J Am Stat Assoc 84:816–822

Schwartz PF, Gennings C, Teuschler LK, Fariss MW (2001) Optimizing the precision of toxicity threshold estimation using a two-stage experimental design. J Agric Biol Environ Stat 6:409

Tamhane AC, Logan BR (2002) Multiple test procedures for identifying the minimum effective and maximum safe doses of a drug. J Am Stat Assoc 97:293–301

West RW, Kodell RL (2005) Changepoint alternatives to the noael. J Agric Biol Environ Stat 10:197

Williams DA (1972) The comparison of several dose levels with a zero dose control. Biometrics 28:519–531

Xie M-G, Singh K (2013) Confidence distribution, the frequentist distribution estimator of a parameter: A review. Int Stat Rev 81:3–39