Confidence sets for a level set in linear regression

Fang Wan[1], Wei Liu[2], Frank Bretz[3]

[1] Lancaster University, UK

[2] University of Southampton, UK

[3] Novartis Pharma AG, Switzerland

Summary

Regression modeling is the workhorse of statistics and there is a vast literature on estimation of the regression function. It is realized in recent years that in regression analysis the ultimate aim may be the estimation of a level set of the regression function, instead of the estimation of the regression function itself. The published work on estimation of the level set has thus far focused mainly on nonparametric regression, especially on point estimation. In this paper, the construction of confidence sets for the level set of linear regression is considered. In particular, $1-\alpha$ level upper, lower and two-sided confidence sets are constructed for the normal-error linear regression. It is shown that these confidence sets can be easily constructed from the corresponding $1-\alpha$ level simultaneous confidence bands. It is also pointed out that the construction method is readily applicable to other parametric regression models where the mean response depends on a linear predictor through a monotonic link function, which include generalized linear models, linear mixed models and generalized linear mixed models. Therefore the method proposed in this paper is widely applicable. Real example is used to illustrate the method.

keywords Confidence sets; linear regression; nonparametric regression; parametric regression; simultaneous confidence bands; statistical inference.

1 INTRODUCTION

Decompression sickness (DCS) is an injury caused by rapid change of pressure, such as during or after water dives. Mild DCS involves symptoms such as muscle or joint pain, while serious DCS can cause paralysis or death. It is important to understand the relationship between risk factors, such as the exposure pressure (depth) and exposure duration, and mortality rate, i.e., the chance of death due to serious DCS. Since adult sheep have a body mass similar to human they are considered to have a similar DCS susceptibility as humans. A sheep decompression trial is reported and studied in Li *et al.* (2008). In the paper, logistic regression is used to model the mortality rate p as a function of the two covariates exposure pressure (x_1) and exposure duration (x_2): $\log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$. Of particular interest are the values of $\boldsymbol{x} = (x_1, x_2)^T$ that correspond to a relatively low mortality rate, say 0.05, i.e., the set

$$\{\, \boldsymbol{x} \in K : \; \beta_0 + \beta_1 x_1 + \beta_2 x_2 = \log\left(0.05/(1-0.05)\right) \,\}$$

where $K \in \Re^2$ is a pre-specified region of $\boldsymbol{x}$. How to make inference about this set (that depends on the unknown parameters β_i), especially the construction of confidence sets, is the well known and well studied effective-dose problem; See, for example, Li *et al.* (2008), Tompsett *et al.* (2018), and the references therein for an overview.

A set of potentially more interest is

$$\{\, \boldsymbol{x} \in K : \; \beta_0 + \beta_1 x_1 + \beta_2 x_2 \leq \log\left(0.05/(1-0.05)\right) \,\},$$

since one would probably be more interested in identifying all the combinations of x_1 and

x_2 for which the mortality rate p is no more than the threshold 0.05. It becomes clear below that this is a level set. This motivates us to study the construction of confidence sets for a level set.

In general, let $Y = h(\boldsymbol{x}) + e$ where $Y \in \Re^1$ is the response, $\boldsymbol{x} \in \Re^p$ is the covariate (vector), h is the regression function, and e is the random error. In regression analysis, there is a vast literature on how to estimate the regression function h, based on the observed data $(Y_i, \boldsymbol{x}_i), i = 1, \ldots, n$. In recent years, it is realized that an important problem in regression is the inference of the λ-level set

$$G_\lambda = G_\lambda(h) = \{\boldsymbol{x} \in K : \ h(\boldsymbol{x}) \geq \lambda\}$$

where λ is a pre-specified number, and $K \subset \Re^p$ is a given covariate $\boldsymbol{x}$ region of interest. It is argued forcefully in Scott and Davenport (2007) that "In a wide range of regression problems, if it is worthwhile to estimate the regression function h, it is also worthwhile to estimate certain level sets. Moreover, these level sets may be of ultimate importance. And in many classification problems, labels are obtained by thresholding a continuous variable. Thus, estimating regression level sets may be a more appropriate framework for addressing many problems that are currently envisioned in other ways". Other than its application to the DCS problem alluded to above, one can envisage that, when considering a regression model of perinatal mortality rate on birth weight, it is interesting to identify the range of birth weight over which the perinatal mortality rate exceeds a certain λ. Further possible applications have been pointed out, for example, in Scott and Davenport (2007) and Dau *et al.* (2020). Inference of the level set G_λ is an important component of the more general field of subgroup analysis (cf. Wang et al., 2007, Herrera et al., 2011, Ting et al., 2020).

In nonparametric regression where h is not assumed to have a specific form, **point** estimation of G_λ aims to construct $\hat{G}_\lambda$ to approximate G_λ using the observed data. This has been considered by Cavalier (1997), Polonik and Wang (2005), Willett and Nowak (2007), Scott and Davenport (2007), Dau *et al.* (2020) and Reeve *et al.* (2021) among others. The main focus of these works is on large sample properties such as consistency and rate of convergence. Related work on estimation of level-sets of a nonparametric density function can be found in Hartigan (1987), Tsybakov (1997), Cadre (2006), Mason and Polonik (2009), Chen *et al.* (2017) and Qiao and Polonik (2019). **Confidence-set** estimation of G_λ aims to construct sets $\hat{G}_\lambda$ to contain or be contained in G_λ with a pre-specified confidence level $1-\alpha$. Large sample approximate $1-\alpha$ confidence-set estimation of G_λ is considered in Mammen and Polonik (2013).

In this paper confidence-set estimation of G_λ for linear regression is considered. It is shown that lower, upper and two-sided confidence-set estimators of G_λ can be easily constructed from the corresponding lower, upper and two-sided simultaneous confidence bands for a linear regression function. Simultaneous confidence bands for linear regression have been considered in Wynn and Bloomfield (1971), Naiman (1984, 1986), Piegorsch (1985a,b), Sun and Loader (1994), Liu and Hayter (2007) and numerous others; see Liu (2010) for an overview. It is also pointed out that the method can be directly extended to, for example, the generalized linear regression models (including the logistic regression for the DCS problem), though the confidence-set estimations are of asymptotic $1-\alpha$ level since the simultaneous confidence bands are of asymptotic $1-\alpha$ level in this case. A related problem is the confidence-set estimation of the maximum (or minimum) point of a linear regression model; see Wan *et al.* (2015, 2016) and the references therein.

The layout of the paper is as follows. The construction method of confidence-set estimators is given in Section 2. The method is illustrated with the DCS example in Section 3. Section 4 contains conclusions and a brief discussion. Finally the appendix sketches the proofs of two theorems in Section 2.

2 Method

The confidence sets for G_λ are constructed in this section. We first consider the normal-error linear regression model, the results of which can directly be extended to the generalized liner regression models, for example, by using the asymptotic normality of the estmator of the regression coefficients.

Let the normal-error linear regression model be given by

$$Y = h(\boldsymbol{x}) + e = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + e\,,$$

where the independent errors $e_i = Y_i - h(\boldsymbol{x}_i)$ have distribution $N(0, \sigma^2)$. From the observed sample of observations $(Y_i, \boldsymbol{x}_i), i = 1, \cdots, n$, the usual estimator of $\boldsymbol{\beta} = (\beta_0, \cdots, \beta_p)^T$ is given by $\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{Y}$ where X is the $n \times (p+1)$ design matrix and $\mathbf{Y} = (Y_1, \cdots, Y_n)^T$. The estimator of the error variance σ^2 is given by $\hat{\sigma}^2$. It is known that $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2 (X^T X)^{-1})$, $\hat{\sigma}^2 \sim \sigma^2 \chi^2_\nu / \nu$ with $\nu = n - p - 1$, and $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are independent. In order for both estimators

$\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ to be available, the sample size n must be at least $n \geq p+2$.

Let $\tilde{\boldsymbol{x}} = (1, \boldsymbol{x}^T)^T = (1, x_1, \cdots, x_p)^T$. Suppose the upper, lower and two-sided $1-\alpha$ simultaneous confidence bands over the covariate region $\boldsymbol{x} \in K$ are given, respectively, by

$$\mathrm{P}\left\{ \tilde{\boldsymbol{x}}^T \boldsymbol{\beta} \leq \tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} + c_1 \hat{\sigma} m(\boldsymbol{x}) \; \forall \boldsymbol{x} \in K \right\} = 1-\alpha \tag{1}$$

$$\mathrm{P}\left\{ \tilde{\boldsymbol{x}}^T \boldsymbol{\beta} \geq \tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} - c_1 \hat{\sigma} m(\boldsymbol{x}) \; \forall \boldsymbol{x} \in K \right\} = 1-\alpha \tag{2}$$

$$\mathrm{P}\left\{ \tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} - c_2 \hat{\sigma} m(\boldsymbol{x}) \leq \tilde{\boldsymbol{x}}^T \boldsymbol{\beta} \leq \tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} + c_2 \hat{\sigma} m(\boldsymbol{x}) \; \forall \boldsymbol{x} \in K \right\} = 1-\alpha \tag{3}$$

where $m(\boldsymbol{x}) = \sqrt{\tilde{\boldsymbol{x}}^T (X^T X)^{-1} \tilde{\boldsymbol{x}}}$ corresponding to the hyperbolic confidence bands, and $c_1 > 0$ and $c_2 > 0$ are the critical constants to achieve the exact $1-\alpha$ confidence level. Whilst another popular form is $m(\boldsymbol{x}) = 1$, corresponding to the constant-width confidence bands, the hyperbolic bands are often better than the constant-width band under various optimality criteria (see, e.g., Liu and Hayter, 2007, and the references therein) and so used throughout this paper. The critical constants c_1 and c_2 can be computed by using the method of Liu *et al.* (2005, 2008).

It is worth emphasizing that the three probabilities in (1-3) do not depend on the unknown parameters $\boldsymbol{\beta} \in \Re^{p+1}$ and $\sigma > 0$, and that $c_1 < c_2$.

From the simultaneous confidence bands in (1-3), define the confidence sets as

$$\hat{G}_{\lambda,1u} = \left\{ \boldsymbol{x} \in K : \; \tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} + c_1 \hat{\sigma} m(\boldsymbol{x}) \geq \lambda \right\}, \tag{4}$$

$$\hat{G}_{\lambda,1l} = \left\{ \boldsymbol{x} \in K : \; \tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} - c_1 \hat{\sigma} m(\boldsymbol{x}) \geq \lambda \right\}, \tag{5}$$

$$\hat{G}_{\lambda,2u} = \left\{ \boldsymbol{x} \in K : \; \tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} + c_2 \hat{\sigma} m(\boldsymbol{x}) \geq \lambda \right\}, \; \hat{G}_{\lambda,2l} = \left\{ \boldsymbol{x} \in K : \; \tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} - c_2 \hat{\sigma} m(\boldsymbol{x}) \geq \lambda \right\}. \tag{6}$$

The following theorem establishes that $\hat{G}_{\lambda,1u}$ is an upper, and $\hat{G}_{\lambda,1l}$ is a lower, confidence set for G_λ of exact $1-\alpha$ level, whilst $[\hat{G}_{\lambda,2l}, \hat{G}_{\lambda,2u}]$ is a two-sided confidence set for G_λ of at least $1-\alpha$ level. A proof is sketched in the appendix.

Theorem 1. We have

$$\inf_{\boldsymbol{\beta}\in\Re^{p+1},\,\sigma>0} \mathrm{P}\left\{ G_\lambda \subseteq \hat{G}_{\lambda,1u} \right\} = 1-\alpha\,, \tag{7}$$

$$\inf_{\boldsymbol{\beta}\in\Re^{p+1},\,\sigma>0} \mathrm{P}\left\{ \hat{G}_{\lambda,1l} \subseteq G_\lambda \right\} = 1-\alpha\,, \tag{8}$$

$$\inf_{\boldsymbol{\beta}\in\Re^{p+1},\,\sigma>0} \mathrm{P}\left\{ \hat{G}_{\lambda,2l} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u} \right\} \geq 1-\alpha\,. \tag{9}$$

From the definitions in (4-6), it is clear that each set $\hat{G}_{\cdot\cdot}$ is given by all the points in K at which the corresponding simultaneous confidence band is at least as high as the given threshold λ. Note that each set could be an empty set when λ is sufficiently large, and become K when λ is sufficiently small. Of course, each set cannot be larger than the given covariate set K from the definition. Since $c_1 > 0$ and $c_2 > 0$, it is clear that $\hat{G}_{\lambda,1l} \subseteq \hat{G}_{\lambda,1u}$ and $\hat{G}_{\lambda,2l} \subseteq \hat{G}_{\lambda,2u}$. Since $c_1 < c_2$, it is clear that $\hat{G}_{\lambda,1u} \subseteq \hat{G}_{\lambda,2u}$ and $\hat{G}_{\lambda,2l} \subseteq \hat{G}_{\lambda,1l}$. Hence $\hat{G}_{\lambda,2l} \subseteq \hat{G}_{\lambda,1l} \subseteq \hat{G}_{\lambda,1u} \subseteq \hat{G}_{\lambda,2u}$.

Intuitively, since the regression function $\tilde{\boldsymbol{x}}^T\boldsymbol{\beta}$ is bounded from above by the upper simultaneous confidence band $\tilde{\boldsymbol{x}}^T\hat{\boldsymbol{\beta}} + c_1\hat{\sigma}m(\boldsymbol{x})$ over the region $\boldsymbol{x} \in K$, the level set G_λ cannot be bigger than the set $\hat{G}_{\lambda,1u}$. Similarly, since the regression function $\tilde{\boldsymbol{x}}^T\boldsymbol{\beta}$ is bounded from below by the lower simultaneous confidence band $\tilde{\boldsymbol{x}}^T\hat{\boldsymbol{\beta}} - c_1\hat{\sigma}m(\boldsymbol{x})$ over the region $\boldsymbol{x} \in K$, the level set G_λ cannot be smaller than the set $\hat{G}_{\lambda,1l}$. Finally, since the regression function $\tilde{\boldsymbol{x}}^T\boldsymbol{\beta}$ is bounded, simultaneously, from below by the lower confidence band $\tilde{\boldsymbol{x}}^T\hat{\boldsymbol{\beta}} - c_2\hat{\sigma}m(\boldsymbol{x})$, and from above by the upper confidence band $\tilde{\boldsymbol{x}}^T\hat{\boldsymbol{\beta}} + c_2\hat{\sigma}m(\boldsymbol{x})$, over the region $\boldsymbol{x} \in K$, the level set G_λ must contain the set $\hat{G}_{\lambda,2l}$ and be contained in the set $\hat{G}_{\lambda,2u}$ simultaneously.

Instead of the level set G_λ, the set

$$M_\lambda = M_\lambda(h) = \{\boldsymbol{x} \in K : \; h(\boldsymbol{x}) \leq \lambda\} \tag{10}$$

may be of interest in some applications. In this case, one can consider the regression of $-Y$ on $\boldsymbol{x}$, given by $-Y = -h(\boldsymbol{x}) + (-e)$, and hence M_λ becomes the level set $G_{-\lambda}$ of the regression function $-h(\boldsymbol{x})$; see the DCS example in Section 3.

The confidence sets given in (4-6) for the normal-error linear regression can be generalized to other models that involve a linear predictor $\tilde{\boldsymbol{x}}^T\boldsymbol{\beta}$. In generalized linear models, linear mixed models and generalized linear mixed models (cf. McCulloch and Searle, 2001 and Faraway, 2016), for example, the mean response $E(Y)$ is often related to a linear predictor $\tilde{\boldsymbol{x}}^T\boldsymbol{\beta}$ by a given monotonic link function $L(\cdot)$, that is, $L[E(Y)] = \tilde{\boldsymbol{x}}^T\boldsymbol{\beta}$. Since $L(\cdot)$ is monotone, the set of interest $\{\, \boldsymbol{x} \in K : \; E(Y) \geq L_0\}$, for a given threshold L_0, becomes either $\{\, \boldsymbol{x} \in K : \; \tilde{\boldsymbol{x}}^T\boldsymbol{\beta} \geq \lambda \,\}$ or $\{\, \boldsymbol{x} \in K : \; \tilde{\boldsymbol{x}}^T\boldsymbol{\beta} \leq \lambda \,\}$, where $\lambda = L(L_0)$, depending on whether the function $L(\cdot)$ is increasing or decreasing. However, when the distribution of $\hat{\boldsymbol{\beta}}$ is asymptotically normal $N(\boldsymbol{\beta}, \hat{\Sigma})$, the simultaneous confidence bands of the forms in (1-3) are of approximate $1-\alpha$ level; see, e.g., Liu (2010, Chapter 8). As a result, the corresponding confidence sets of the forms in (4-6) are of approximate $1-\alpha$ level too. See the DCS example in the next section.

Now suppose that the value of λ is not pre-specified, that is, one might be interested in the confidence sets for G_λ for several different values of λ. Of course one can use the results above to construct a confidence set, $\hat{G}_{\lambda,1l}$ say, for each given value of λ. The question is "what is the joint confidence level of the confidence sets $\{\hat{G}_{\lambda_1,1l} \subseteq G_{\lambda_1}\}, \{\hat{G}_{\lambda_2,1l} \subseteq G_{\lambda_2}\}, \cdots$ for a sequence of λ-values $\lambda_1, \lambda_2, \cdots$?" The theorem below asserts that the joint confidence level is at least $1-\alpha$, the proof of which is also sketched in the appendix.

Theorem 2. We have

$$\inf_{\boldsymbol{\beta}\in\Re^{p+1},\,\sigma>0} \mathrm{P}\left\{ G_\lambda \subseteq \hat{G}_{\lambda,1u}\ \forall\lambda\in\Re^1 \right\} = 1-\alpha\,, \qquad (11)$$

$$\inf_{\boldsymbol{\beta}\in\Re^{p+1},\,\sigma>0} \mathrm{P}\left\{ \hat{G}_{\lambda,1l} \subseteq G_\lambda\ \forall\lambda\in\Re^1 \right\} = 1-\alpha\,, \qquad (12)$$

$$\inf_{\boldsymbol{\beta}\in\Re^{p+1},\,\sigma>0} \mathrm{P}\left\{ \hat{G}_{\lambda,2l} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u}\ \forall\lambda\in\Re^1 \right\} \geq 1-\alpha\,. \qquad (13)$$

3 Wisconsin-Madison sheep dive trial

In the Wisconsin-Madison sheep dive trial, 1108 dives were performed and recorded. Following Li *et al.* (2008), logistic regression is used to model the relationship between the mortality rate p and the two covariates x_1, the log base 10 exposure depth, and x_2, the log base 10 exposure duration: $\mathrm{logit}(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$; here $\mathrm{logit}(p) = \log(p/(1-p))$ which is monotone increasing in $p \in (0,1)$. Based on the recorded data, the MLE $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_1)^T$ is calculated to be $(-19.253, 14.196, 3.758)^T$ and the approximate covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\hat{\mathcal{I}}^{-1} = \begin{pmatrix} 5.0004779 & -5.5146133 & -0.8322975 \\ -5.5146133 & 7.1346280 & 0.7606109 \\ -0.8322975 & 0.7606109 & 0.1648114 \end{pmatrix}.$$

Hence $\hat{\boldsymbol{\beta}}$ has approximate normal distribution $N_3(\boldsymbol{\beta}, \hat{\mathcal{I}}^{-1})$.

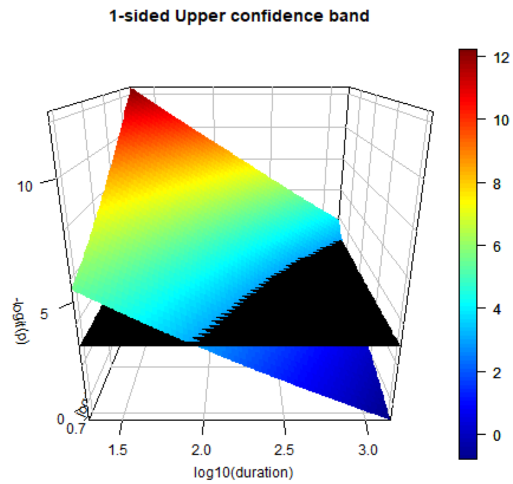
A major goal of this study as described in Li *et al.* (2008) was to determine the ranges of dive depth and duration that correspond to relatively low mortality rates. This knowledge will be invaluable for safety in human dives. Set $p = .05$ as a threshold of low mortality rate. From the recorded 1108 dives, the minimum value of x_1 is $\min(x_1) = 0.314$ the maximum value of x_1 is $\max(x_1) = 0.714$, $\min(x_2) = 1.301$ and $\max(x_2) = 3.158$. Hence it is important to identify the set

$$\{\, \boldsymbol{x} \in K : \ \beta_0 + \beta_1 x_1 + \beta_2 x_2 \le \text{logit}(0.05) \,\} = \left\{\, \boldsymbol{x} \in K : \ -\tilde{\boldsymbol{x}}^T \boldsymbol{\beta} \ge -\text{logit}(0.05) \,\right\}$$

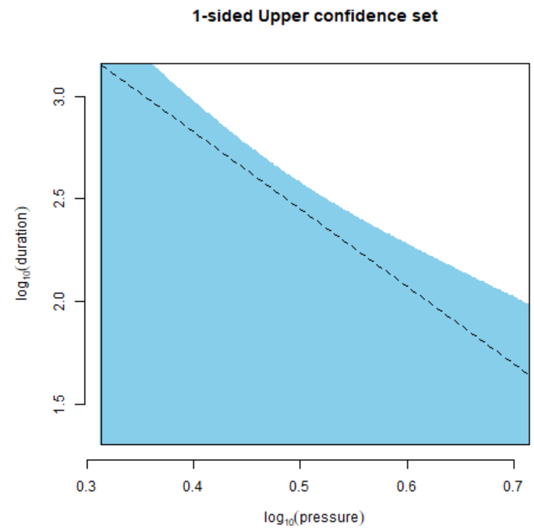
where $K = \{\, \boldsymbol{x} = (x_1, x_2)^T : 0.314 \le x_1 \le 0.714, 1.301 \le x_2 \le 3.158 \,\}$. The set above is just the λ-level set of the regression function $-\tilde{\boldsymbol{x}}^T \boldsymbol{\beta}$ with $\lambda = -\text{logit}(0.05)$. So the method of Section 2 can be used to construct the confidence sets in (4-6) for this G_λ.

From Section 2, simultaneous confidence bands for $-\tilde{\boldsymbol{x}}^T \boldsymbol{\beta}$ over $\boldsymbol{x} \in K$ need to be constructed first in order to construct the confidence sets. Note, however, only approximate $1 - \alpha$ confidence bands of the forms in (1-3), with $\hat{\sigma} = 1$, $\nu = \infty$ and $(X^T X)^{-1}$ replaced with $\hat{\mathcal{I}}^{-1}$, can be constructed by using the approximate normal distribution $N_3(\boldsymbol{\beta}, \hat{\mathcal{I}}^{-1})$ of $\hat{\boldsymbol{\beta}}$. Hence the confidence sets for G_λ are also of approximate $1 - \alpha$ level. For $1 - \alpha = 95\%$ and K given above, the critical values c_1 and c_2 are computed to be 2.483 and 2.728, respectively, by using the method of Liu *et al.* (2005) (see also Liu, 2010, Section 3.2).

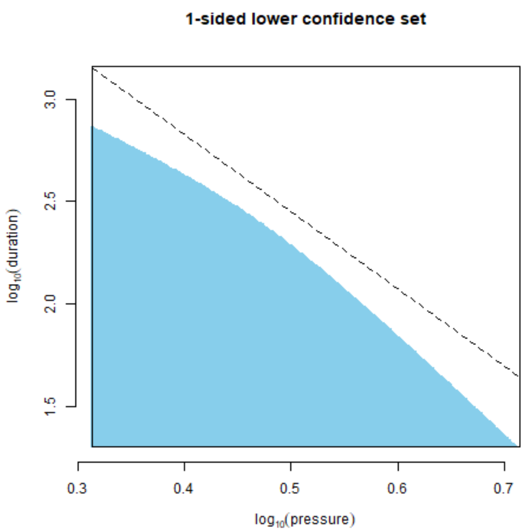
Figure 1(a) plots the 1-sided upper simultaneous confidence band for $-\tilde{\boldsymbol{x}}^T \boldsymbol{\beta}$ and the horizontal plane at height $\lambda = -\text{logit}(0.05)$ over the rectangular region K.



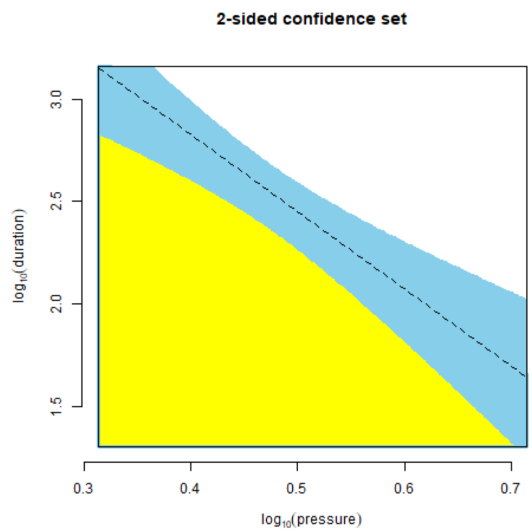
(a) 1-sided upper confidence band and the level plane



(b) 1-sided upper confidence set $\hat{G}_{\lambda,1u}$



(c) 1-sided lower confidence set $\hat{G}_{\lambda,1l}$



(d) 2-sided confidence set $[\hat{G}_{\lambda,2l}, \hat{G}_{\lambda,2u}]$

Figure 1: The 95% confidence sets in sheep dive example, given by the shaded regions.

Figure 1(b) plots the 1-sided upper confidence set $\hat{G}_{\lambda,1u}$, with the region K given by the rectangle in solid line and $-\tilde{\boldsymbol{x}}^T\boldsymbol{\beta} = -\text{logit}(0.05)$ given by the dashed line. Note that the curvilinear-boundary of $\hat{G}_{\lambda,1u}$ is given by the projection, to the $\boldsymbol{x}$-plane, of the intersection between the horizontal plane at height $-\text{logit}(0.05)$ and the 1-sided upper simultaneous confidence band over the region $\boldsymbol{x} \in K$ in Figure 1(a). The upper confidence set $\hat{G}_{\lambda,1u}$ tells us that, with 95% confidence level, within K only those dives with $\boldsymbol{x}$ in $\hat{G}_{\lambda,1u}$ may have mortality rate smaller than or equal to 0.05. Hence a dive with $\boldsymbol{x} \in K \backslash \hat{G}_{\lambda,1u}$ should be considered too dangerous in terms of mortality rate.

Similarly, Figure 1(c) plots the 1-sided lower confidence set $\hat{G}_{\lambda,1l}$ in the $\boldsymbol{x}$-plane. Note that the curvilinear-boundary of $\hat{G}_{1l}$ is given by the projection, to the $\boldsymbol{x}$-plane, of the intersection between the horizontal plane at height $-\text{logit}(0.05)$ and the 1-sided lower simultaneous confidence band for $-\tilde{\boldsymbol{x}}^T\boldsymbol{\beta}$ over the region K. The lower confidence set $\hat{G}_{\lambda,1l}$ tells us that, with 95% confidence level, dives with $\boldsymbol{x} \in \hat{G}_{\lambda,1l}$ have mortality rate smaller than or equal to 0.05. Hence these dives may be considered ‘safe’.

Figure 1(d) plots the two-sided confidence set $[\hat{G}_{\lambda,2l}, \hat{G}_{\lambda,2u}]$ in the $\boldsymbol{x}$-plane. Note that the curvilinear-boundaries of $[\hat{G}_{\lambda,2l}, \hat{G}_{\lambda,2u}]$ are given by the projection, to the $\boldsymbol{x}$-plane, of the intersection between the horizontal plane at height $-\text{logit}(0.05)$ and the two-sided confidence band for $-\tilde{\boldsymbol{x}}^T\boldsymbol{\beta}$ over the region K. The two-sided confidence set tells us that, with 95% confidence level, dives with $\boldsymbol{x} \in K \backslash \hat{G}_{\lambda,2u}$ are considered as dangerous, dives with $\boldsymbol{x} \in \hat{G}_{\lambda,2l}$ are considered as safe, and dives with $\boldsymbol{x} \in \hat{G}_{\lambda,2u}$ are possibly dangerous, in terms of mortality.

If one feels mortality rate 0.05 is too high, one may want to try 0.01, for example, and construct the corresponding confidence set $\hat{G}_{\lambda,2l}$. Indeed one can construct confidence sets

$\hat{G}_{\lambda_1,2l}, \hat{G}_{\lambda_2,2l}, \cdots$ for any sequence of $\lambda_1, \lambda_2, \cdots$. Theorem 2 guarantees that the simultaneous confidence level of this sequence of lower confidence sets is still $1-\alpha = 95\%$.

4 CONCLUSION AND DISCUSSION

In this paper, the construction of confidence sets for the level set of linear regression is considered. Upper, lower and two-sided confidence sets of level $1-\alpha$ are constructed for the normal-error linear regression. It is shown that these confidence sets are constructed from the corresponding $1-\alpha$ level simultaneous confidence bands. Hence these confidence sets and simultaneous confidence bands are closely related.

It is noteworthy that the sample size n only needs to satisfy $\nu = n-p-1 \geq 1$, i.e. $n \geq p+2$, so that the regression coefficients $\boldsymbol{\beta}$ and the error variance σ^2 can be estimated. So long as $n \geq p+2$, the theorem in Section 2 holds. A larger sample size n will make the confidence sets closer to the level set, which is similar to the usual confidence sets for the mean of a normally-distributed population. Hence the method for linear regression provided in this paper is much simpler than that for nonparametric regression and density level sets (cf. Mammend and Polonik, 2013, Chen *et al.*, 2017, Qiao and Polonik, 2019).

In Theorem 1 in Section 2, the minimum coverage probability over the whole parameter space $\boldsymbol{\beta} \in \Re^{p+1}$ and $\sigma > 0$ is sought since no assumption is made about any prior information on $\boldsymbol{\beta}$ or $\sigma > 0$. If it is known *a priori* that $\boldsymbol{\beta}$ and σ are in a restricted space, then the usual estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}$ should be replaced by the maximum likelihood estimators over the restricted space, and the minimum coverage probability should also be over this restricted space. This situation becomes more complicated and is beyond the scope of this paper.

It is also pointed out that the construction method is readily applicable to other parametric regression models where the mean response depends on a linear predictor through a monotonic link function. Examples are generalized linear models, linear mixed models and generalized linear mixed models. The illustrative example in Section 3 involves a generalized linear model. Therefore the method proposed in this paper is widely applicable.

We are unable to establish thus far whether the two-sided confidence set $[\hat{G}_{\lambda,2l}, \hat{G}_{\lambda,2u}]$ is of confidence level $1-\alpha$ exactly. Construction of a two-sided confidence set of exact confidence level $1-\alpha$ is clearly of interest and warrants further research. We are actively researching on this.

5 Acknowledgment

We thank Professor Jialiang Li of National University of Singapore for his help in preparing this paper.

6 Appendix

In this appendix proofs of the two theorems in Section 2 are sketched.

First, consider Theorem 1. For proving the statement in (7), we have

$$\begin{aligned}
& \left\{ G_\lambda \subseteq \hat{G}_{\lambda,1u} \right\} \\
= & \left\{ \forall \boldsymbol{x} \in G_\lambda : \boldsymbol{x} \in \hat{G}_{\lambda,1u} \right\} \\
= & \left\{ \forall \boldsymbol{x} \in G_\lambda : \ \tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} + c_1 \hat{\sigma} m(\boldsymbol{x}) \geq \lambda \right\} \\
= & \left\{ \forall \boldsymbol{x} \in G_\lambda : \ \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + c_1 \hat{\sigma} m(\boldsymbol{x}) \geq \lambda - \tilde{\boldsymbol{x}}^T \boldsymbol{\beta} \right\} \\
\supseteq & \left\{ \forall \boldsymbol{x} \in G_\lambda : \ \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + c_1 \hat{\sigma} m(\boldsymbol{x}) \geq 0 \right\} \\
\supseteq & \left\{ \forall \boldsymbol{x} \in K : \ \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + c_1 \hat{\sigma} m(\boldsymbol{x}) \geq 0 \right\}
\end{aligned}$$

where the second equation follows directly from the definition of $\hat{G}_{\lambda,1u}$, the first "$\supseteq$" follows directly from the definition of G_λ, and the second "$\supseteq$" follows directly from the fact that $G_\lambda \subseteq K$. It follows therefore

$$\mathrm{P}\left\{ G_\lambda \subseteq \hat{G}_{\lambda,1u} \right\} \geq \mathrm{P}\left\{ \forall \boldsymbol{x} \in K : \ \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + c_1 \hat{\sigma} m(\boldsymbol{x}) \geq 0 \right\} = 1 - \alpha \tag{14}$$

where the last equality is directly due to the fact that $\tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} + c_1 \hat{\sigma} m(\boldsymbol{x})$ is an upper simultaneous confidence band for $\tilde{\boldsymbol{x}}^T \boldsymbol{\beta}$ over $\boldsymbol{x} \in K$ of exact $1 - \alpha$ level, as given in (1).

Next we show that the minimum probability over $\boldsymbol{\beta} \in \Re^{p+1}$ and $\sigma > 0$ in statement (7) is $1 - \alpha$, attained at $\boldsymbol{\beta} = (\lambda, 0, \ldots 0)^T$. At $\boldsymbol{\beta} = (\lambda, 0, \ldots 0)^T$, we have $G_\lambda = K$ and $\lambda = \tilde{\boldsymbol{x}}^T \boldsymbol{\beta}$, and so

$$
\begin{aligned}
& \left\{ G_\lambda \subseteq \hat{G}_{\lambda,1u} \right\} \\
= & \left\{ \forall \boldsymbol{x} \in K : \boldsymbol{x} \in \hat{G}_{\lambda,1u} \right\} \\
= & \left\{ \forall \boldsymbol{x} \in K : \;\; \tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} + c_1 \hat{\sigma} m(\boldsymbol{x}) \geq \lambda \right\} \\
= & \left\{ \forall \boldsymbol{x} \in K : \;\; \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + c_1 \hat{\sigma} m(\boldsymbol{x}) \geq \lambda - \tilde{\boldsymbol{x}}^T \boldsymbol{\beta} \right\} \\
= & \left\{ \forall \boldsymbol{x} \in K : \;\; \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + c_1 \hat{\sigma} m(\boldsymbol{x}) \geq 0 \right\}
\end{aligned}
$$

which gives

$$
\mathrm{P}\left\{ G_\lambda \subseteq \hat{G}_{\lambda,1u} \right\} = \mathrm{P}\left\{ \forall \boldsymbol{x} \in K : \;\; \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + c_1 \hat{\sigma} m(\boldsymbol{x}) \geq 0 \right\} = 1 - \alpha \,. \tag{15}
$$

The combination of (14) and (15) proves the statement in (7).

Now we prove the statement in (8). For a given set $A \subseteq K$, let A^c denote the complement set within K, i.e. $A^c = K \backslash A$. We have

$$
\begin{aligned}
& \left\{ \hat{G}_{\lambda,1l} \subseteq G_\lambda \right\} \\
= & \left\{ G_\lambda^c \subseteq \hat{G}_{\lambda,1l}^c \right\} \\
= & \left\{ \forall \boldsymbol{x} \in G_\lambda^c : \boldsymbol{x} \in \hat{G}_{\lambda,1l}^c \right\} \\
= & \left\{ \forall \boldsymbol{x} \in G_\lambda^c : \ \tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} - c_1 \hat{\sigma} m(\boldsymbol{x}) < \lambda \right\} \\
= & \left\{ \forall \boldsymbol{x} \in G_\lambda^c : \ \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - c_1 \hat{\sigma} m(\boldsymbol{x}) < \lambda - \tilde{\boldsymbol{x}}^T \boldsymbol{\beta} \right\} \\
\supseteq & \left\{ \forall \boldsymbol{x} \in G_\lambda^c : \ \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - c_1 \hat{\sigma} m(\boldsymbol{x}) \leq 0 \right\} \\
\supseteq & \left\{ \forall \boldsymbol{x} \in K : \ \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - c_1 \hat{\sigma} m(\boldsymbol{x}) \leq 0 \right\}
\end{aligned}
$$

where the third equation follows directly from the definition of $\hat{G}_{\lambda,1l}$ (or $\hat{G}_{\lambda,1l}^c$), the first "$\supseteq$" follows directly from the definition of G_λ (or G_λ^c), and the second "$\supseteq$" follows directly from the fact that $G_\lambda^c \subseteq K$. It follows therefore

$$
\mathrm{P} \left\{ \hat{G}_{\lambda,1l} \subseteq G_\lambda \right\} \geq \mathrm{P} \left\{ \forall \boldsymbol{x} \in K : \ \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - c_1 \hat{\sigma} m(\boldsymbol{x}) \leq 0 \right\} = 1 - \alpha \qquad (16)
$$

where the last equality is directly due to the fact that $\tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} - c_1 \hat{\sigma} m(\boldsymbol{x})$ is a lower simultaneous confidence band for $\tilde{\boldsymbol{x}}^T \boldsymbol{\beta}$ over $\boldsymbol{x} \in K$ of exact $1 - \alpha$ level, as given in (2).

Next we show that the minimum probability over $\boldsymbol{\beta} \in \Re^{p+1}$ and $\sigma > 0$ in statement (8) is $1 - \alpha$, attained at $\boldsymbol{\beta} = (\lambda^-, 0, \ldots 0)^T$, where λ^- denotes a number that is infinitesimally smaller than λ. At $\boldsymbol{\beta} = (\lambda^-, 0, \ldots 0)^T$, we have $G_\lambda^c = K$ and so

$$\begin{aligned}
& \left\{ \hat{G}_{\lambda,1l} \subseteq G_\lambda \right\} \\
& \left\{ G_\lambda^c \subseteq \hat{G}_{\lambda,1l}^c \right\} \\
= & \left\{ \forall \boldsymbol{x} \in G_\lambda^c : \boldsymbol{x} \in \hat{G}_{\lambda,1l}^c \right\} \\
= & \left\{ \forall \boldsymbol{x} \in K : \boldsymbol{x} \in \hat{G}_{\lambda,1l}^c \right\} \\
= & \left\{ \forall \boldsymbol{x} \in K : \ \tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} - c_1 \hat{\sigma} m(\boldsymbol{x}) < \lambda \right\} \\
= & \left\{ \forall \boldsymbol{x} \in K : \ \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - c_1 \hat{\sigma} m(\boldsymbol{x}) < \lambda - \tilde{\boldsymbol{x}}^T \boldsymbol{\beta} \right\} \\
= & \left\{ \forall \boldsymbol{x} \in K : \ \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - c_1 \hat{\sigma} m(\boldsymbol{x}) < 0 \right\}
\end{aligned}$$

which gives

$$\mathrm{P} \left\{ \hat{G}_{\lambda,1l} \subseteq G_\lambda \right\} = \mathrm{P} \left\{ \forall \boldsymbol{x} \in K : \ \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - c_1 \hat{\sigma} m(\boldsymbol{x}) < 0 \right\} = 1 - \alpha \, . \qquad (17)$$

The combination of (16) and (17) proves the statement in (8).

The statement (9) can be proved by combining the arguments that establish (14) and (16) above to establish that

$$\left\{ \hat{G}_{\lambda,2l} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u} \right\} \supseteq \left\{ \forall \boldsymbol{x} \in K : \ -c_2 \hat{\sigma} m(\boldsymbol{x}) \leq \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) < c_2 \hat{\sigma} m(\boldsymbol{x}) \right\} ;$$

details are omitted here to save space. Unfortunately, a least favorable configuration of $\boldsymbol{\beta}$ that achieves the coverage probability $1 - \alpha$ cannot be identified in this case, and so $1 - \alpha$ is only a lower bound on the confidence level.

Now consider Theorem 2. For proving the statement in (11), we have

$$\begin{aligned}
& \Big\{ G_\lambda \subseteq \hat{G}_{\lambda,1u} \ \forall \lambda \in \Re^1 \Big\} \\
= \ & \cap_{\lambda \in \Re^1} \Big\{ G_\lambda \subseteq \hat{G}_{\lambda,1u} \Big\} \\
\supseteq \ & \cap_{\lambda \in \Re^1} \Big\{ \forall \boldsymbol{x} \in K: \ \tilde{\boldsymbol{x}}^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + c_1 \hat{\sigma} m(\boldsymbol{x}) \geq 0 \Big\} \qquad (18) \\
= \ & \Big\{ \forall \boldsymbol{x} \in K: \ \tilde{\boldsymbol{x}}^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + c_1 \hat{\sigma} m(\boldsymbol{x}) \geq 0 \Big\}
\end{aligned}$$

where the"$\supseteq$" in (18) follows directly from the proof of the statement in (7) above, and the second "=" follows directly since each set in (18) has nothing to do with λ. It follows therefore

$$\mathrm{P}\Big\{ G_\lambda \subseteq \hat{G}_{\lambda,1u} \ \forall \lambda \in \Re^1 \Big\} \geq \mathrm{P}\Big\{ \forall \boldsymbol{x} \in K: \ \tilde{\boldsymbol{x}}^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + c_1 \hat{\sigma} m(\boldsymbol{x}) \geq 0 \Big\} = 1 - \alpha\,. \qquad (19)$$

On the other hand, it is clear that $\Big\{ G_\lambda \subseteq \hat{G}_{\lambda,1u} \ \forall \lambda \in \Re^1 \Big\} \subseteq \Big\{ G_\lambda \subseteq \hat{G}_{\lambda,1u} \Big\}$ and so

$$\inf_{\boldsymbol{\beta} \in \Re^{p+1},\, \sigma > 0} \mathrm{P}\Big\{ G_\lambda \subseteq \hat{G}_{\lambda,1u} \ \forall \lambda \in \Re^1 \Big\} \leq \inf_{\boldsymbol{\beta} \in \Re^{p+1},\, \sigma > 0} \mathrm{P}\Big\{ G_\lambda \subseteq \hat{G}_{\lambda,1u} \Big\} = 1 - \alpha\,. \qquad (20)$$

The combination of (19) and (20) clearly gives the statement in (11).

The statements in (12-13) of Theorem 2 can be proved in a similar way, and so the details are omitted to save space.

References

Cadre, B. (2006). Kernel estimation of density level sets. *J. Multivariate Anal.*, 97 (4), 999 - 1023.

Cavalier, L. (1997). Nonparametric estimation of regression level sets. *Statistics*, 29 (2), 131 - 160.

Chen, Y.-C., Genovese, C.R., and Wasserman, L. (2017). Density level sets: Asymptotics, inference, and visualization. *J. Amer. Statist. Assoc.*, 112 (520), 1684 - 1696.

Dau, H. D., Laloë, T., and Servien, R. (2020). Exact asymptotic limit for kernel estimation of regression level sets. *Statistics and Probability Letters*, 161:108721.

Faraway, J. J. (2016). *Extending the Linear Model with R, 2nd Edition*. CRC Press.

Hartigan, J.A. (1987). Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.*, 82 (397), 267 - 270.

Herrera, F., Carmona, C. J., Gonzalez, P., and Del Jesus, M. J. (2011). An overview on subgroup discovery: foundations and applications. *Knowledge and Information Systems*, 29(3):495 - 525.

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

Liu, W. (2010). *Simultaneous Inference in Regression*. CRC Press.

Liu, W. and Hayter, A.J. (2007). Minimum area confidence set optimality for confidence bands in simple linear regression. *J. Amer. Statist. Assoc.*, **102**, 181-190.

Liu, W., Jamshidian, M., Zhang, Y., and Donnelly, J. (2005). Simulation-based simultaneous confidence bands in multiple linear regression with predictor variables constrained in intervals. *J. Comput. and Graph. Statist.*, **14**, 459-484.

Liu, W., Wynn, H.P., and Hayter, A.J. (2008). Statistical inferences for linear regression models when the covariates have functional relationships: polynomial regression. *J. Statist. Comput. and Simulat.*, **78**, 315-324.

Mammend, E., and Polonik, W. (2013). Confidence regions for level sets. *J. Multivariate Analysis*, 122, 202-214.

Mason, D.M., and Polonik, W. (2009). Asymptotic normality of plug-in level set estimates. *Ann. Appl. Probab.*, 19 (3), 1108 - 1142.

McCulloch, C. E., and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models.* Wiley.

Myers, R.H., Montgomery, D.C. and Vining, G.G. (2002). *Generalized Linear Models with Applications in Engineering and the Sciences.* Wiley.

Naiman, D.Q. (1984). Optimal simultaneous confidence bounds. *Ann. Statist.*, **12**, 702-715.

Naiman, D.Q. (1986). Conservative confidence bands in curvilinear regression. *Ann. Statist.*, **14**, 896-906.

Piegorsch, W.W. (1985a). Admissible and optimal confidence bands in simple linear regression. *Ann. Statist.*, **13**, 801-810.

Piegorsch, W.W. (1985b). Average-width optimality for confidence bands in simple linear-regression. *J. Amer. Statist. Assoc.*, **80**, 692-697.

Polonik, W. and Wang, Z. (2005). Estimation of regression contour clusters: an application of the excess mass approach to regression. *J. Multivariate Anal.*, 94, 227 - 249.

Qiao, W. and Polonik, W. (2019) Nonparametric confidence regions for level sets: statistical properties and geometry. Electronic Journal of Statistics, **13**, 985-1030.

Reeve, H.W.J., Cannings, T.I., and Samworth, R.J. (2021). Optimal subgroup selection. 2109.01077.pdf (arxiv.org).

Scott, C., Davenport, M. (2007). Regression level set estimation via cost-sensitive classification. *IEEE Trans. Signal Process.*, 55 (6, part 1), 2752 - 2757.

Sun, J., and Loader, C.R. (1994). Simultaneous confidence bands for linear regression and smoothing. *Ann. Statist.*, **22**, 1328-1346.

Ting, N., Cappelleri, J. C., Ho, S., and Chen, D.-G. (2020). *Design and Analysis of Subgroups with Biopharmaceutical Applications*. Springer.

Tompsett, D. M., Biedermann, S. and Liu, W. (2018). Simultaneous confidence sets for several effective doses. *Biometrical Journal*, **60**, 703-720.

Tsybakov, A.B. (1997). On nonparametric estimation of density level sets. *Ann. Statist.*, 25 (3), 948 - 969.

Wan, F., Liu, W., Han, Y., and Bretz, F. (2015). An exact confidence set for a maximum point of a univariate polynomial function in a given interval. *Technometrics*, 57(4), 559-565.

Wan, F., Liu, W., Bretz, F., and Han, Y. (2016). Confidence sets for optimal factor levels of a response surface. *Biometrics*, 72(4), 1285-1293.

Wang, R., Lagakos, S. W., Ware, J. H., Hunter, D. J., and Drazen, J. M. (2007). Statistics in medicine reporting of subgroup analyses in clinical trials. *New England Journal of Medicine*, 357(21): 2189 - 2194. PMID: 18032770.

Willett, R.M., Nowak, R.D. (2007). Minimax optimal level-set estimation. *IEEE Trans. Image Process.*, 16 (12), 2965 - 2979.

Wynn, H.P., and Bloomfield, P. (1971). Simultaneous confidence bands in regression analysis. *JRSS (B)*, **33**, 202-217.