Exact two-sided confidence sets for a level set in simple 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

Abstract

Regression modeling is the workhorse of statistics. It is realized in recent years that one important aim in regression analysis may be the estimation of a level set of the regression function. The published work on this has thus far focused mainly on nonparametric regression, especially on point estimation. In Wan *et al.* (2022), exact upper and lower, but only conservative two-sided, confidence sets for a level set are constructed in linear regression. In this paper, exact two-sided confidence sets are constructed in simple linear regression. A simultaneity property of the exact two-sided confidence is also studied. An example is given to illustrate the method.

keywords: confidence sets; level set; linear regression; simultaneous confidence bands; statistical inference.

*Author names are in random order

1 INTRODUCTION

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; see, for example, Scott and Davenport (2007), Dau *et al.* (2020) and Wan *et al.* (2022) and the references therein. 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, aiming to construct $\hat{G}_\lambda$ to approximate G_λ using the observed data, 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. On the other hand, **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 Wan *et al.* (2022), confidence-set estimation of G_λ for linear regression is considered. It is shown that the problem is closely related to simultaneous confidence bands for the linear regression function, which 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. Specifically, one-sided lower and upper confidence sets for G_λ of exact $1-\alpha$ level can be constructed from the corresponding one-sided lower and upper simultaneous confidence bands for the regression function of exact $1-\alpha$ level. However, the two-sided confidence set for G_λ constructed from the corresponding two-sided exact $1-\alpha$ level simultaneous confidence band for the regression function is of conservative confidence level $1-\alpha$.

In this paper, a two-sided confidence set for G_λ of exact $1-\alpha$ level is constructed for simple linear regression. This confidence set is 'tighter' than the conservative confidence set given in Wan *et al.* (2022) in this case. As pointed out in Wan *et al.* (2022), the method can be directly extended to, for example, the generalized linear regression models, though the two-sided confidence set is of asymptotic $1-\alpha$ level since only asymptotic normality of the regression coefficients estimators is available in this case, instead of the exact normality in the linear regression case. It is also shown that the exact confidence sets constructed in this paper do not have the property of simultaneity in λ, which is different from the two-sided confidence sets constructed from the two-sided simultaneous confidence band.

The layout of the paper is as follows. The construction of the exact two-sided confidence set is given in Section 2. The property of simultaneity in λ is considered in Section 3. The method is illustrated with one example in Section 4. Section 5 contains a brief discussion on possible future research.

2 An Exact two-sided confidence set

In this section, an exact $1-\alpha$ level two-sided confidence set for G is constructed for simple linear regression model given by

$$Y = h(x) + e = \beta_0 + \beta_1 x + e\,,$$

where the independent errors $e_i = Y_i - h(x_i)$ have distribution $N(0, \sigma^2)$. From the observed sample of observations $(Y_i, x_i), i = 1, \cdots, n$, the usual estimator of $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ is given by $\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{Y}$ where X is the $n \times 2$ design matrix and $\mathbf{Y} = (Y_1, \cdots, Y_n)^T$. The usual estimator of the error variance σ^2 is denoted 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 $\nu = n-2$, and $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are independent.

Let $\tilde{\boldsymbol{x}} = (1, x)^T$, and the covariate region of interest be given by the interval $K = [l, u]$ where $-\infty \le l < u \le \infty$ are given numbers. Suppose the two-sided $1-\alpha$ simultaneous confidence band over the covariate region $x \in K = [l, u]$ is given by

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

where $g(x) = \sqrt{\tilde{\boldsymbol{x}}^T (X^T X)^{-1} \tilde{\boldsymbol{x}}}$, and $c_2 > 0$ is the critical constant to achieve the exact $1-\alpha$ confidence level. Define

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

It is shown in Wan *et al.* (2022) that

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

that is, $[\hat{G}_{\lambda,2l}, \hat{G}_{\lambda,2u}]$ is a two-sided confidence set for G_λ of at least $1-\alpha$ level.

Our aim in this section is to construct new and exact confidence set

$$\hat{G}_{\lambda,2u,N} = \left\{ x \in K:\ \tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} + c_{2N} \hat{\sigma} g(x) \ge \lambda \right\},\ \hat{G}_{\lambda,2l,N} = \left\{ x \in K:\ \tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} - c_{2N} \hat{\sigma} g(x) \ge \lambda \right\} \tag{3}$$

where $c_{2N} < c_2$ is a suitably chosen critical constant so that

$$\inf_{\boldsymbol{\beta}\in\Re^2,\,\sigma>0} \mathrm{P}\left\{ \hat{G}_{\lambda,2l,N} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u,N} \right\} = 1-\alpha\,, \tag{4}$$

that is, $[\hat{G}_{\lambda,2l,N}, \hat{G}_{\lambda,2u,N}]$ is a two-sided confidence set for G_λ of exact $1-\alpha$ level.

It is clear from (2) and (3) that the confidence sets $[\hat{G}_{\lambda,2l,N}, \hat{G}_{\lambda,2u,N}]$ and $[\hat{G}_{\lambda,2l}, \hat{G}_{\lambda,2u}]$ differ only in their critical constants. Since $c_{2N} < c_2$, $[\hat{G}_{\lambda,2l,N}, \hat{G}_{\lambda,2u,N}]$ is tighter than $[\hat{G}_{\lambda,2l}, \hat{G}_{\lambda,2u}]$ in the sense that $\hat{G}_{\lambda,2l} \subset \hat{G}_{\lambda,2l,N} \subset \hat{G}_{\lambda,2u,N} \subset \hat{G}_{\lambda,2u}$.

The main result of this section is given by the theorem below which gives an explicit expression for the minimum probability in (4).

Theorem 1. For $c_{2N} > 0$, we have

$$\begin{aligned} &\inf_{\boldsymbol{\beta}\in\Re^2,\,\sigma>0} \mathrm{P}\left\{ \hat{G}_{\lambda,2l,N} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u,N} \right\} \\ =\ & \mathtt{pt}(c_{2N},\nu) - \frac{\phi}{2\pi}\left(1 + c_{2N}^2/\nu\right)^{-\nu/2} - \frac{1}{\pi}\int_0^{(\pi-\phi)/2} \left(1 + \frac{c_{2N}^2}{\nu\sin^2(\phi/2+\theta)}\right)^{-\nu/2} d\theta \end{aligned} \tag{5}$$

where $\mathtt{pt}(\cdot,\nu)$ denotes the cumulative distribution function (cdf) of the t-distribution with ν degrees of freedom (df), and the angle $\phi \in (0,\pi]$ is defined in (10) below with $P = (X^T X)^{-1/2}$.

Proof. It will become clear from below that the probability in (4) does not depend on $\sigma > 0$, and so it is only necessary to study the minimum probability over $\boldsymbol{\beta} \in \Re^2$. To evaluate the minimum probability in (4) over $\boldsymbol{\beta} \in \Re^2$, the probability itself needs to be computed for various configurations of $\boldsymbol{\beta} \in \Re^2$ next. Below we consider four configurations of $\boldsymbol{\beta}$ which include all possible $\boldsymbol{\beta} \in \Re^2$. For a given set $A \subseteq K$, let A^c denote the complement set within K, i.e. $A^c = K\backslash A$.

First, consider **Configuration 1**: $\boldsymbol{\beta} \in \Re^2$ is such that the regression line $Y = \beta_0 + \beta_1 x$ crosses the level line $Y = \lambda$ at $x = m \in (l,u)$, $\beta_0 + \beta_1 x < \lambda$ for $x \in [l,m)$, and $\beta_0 + \beta_1 x > \lambda$

for $x \in (m, u]$. In this case, we have $G_\lambda = [m, u]$ and $G_\lambda^c = [l, m)$, and so

$$\begin{aligned}
& \mathrm{P}\left\{ \hat{G}_{\lambda,2l,N} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u,N} \right\} \qquad (6) \\
= \;& \mathrm{P}\left\{ G_\lambda^c \subseteq \hat{G}_{\lambda,2l,N}^c \text{ and } G_\lambda \subseteq \hat{G}_{\lambda,2u,N} \right\} \\
= \;& \mathrm{P}\left\{ \forall\, x \in G_\lambda^c : \tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} - c_{2N}\hat{\sigma} g(x) < \lambda \text{ and } \forall\, x \in G_\lambda : \tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} + c_{2N}\hat{\sigma} g(x) \geq \lambda \right\} \\
= \;& \mathrm{P}\Big\{ \forall\, x \in G_\lambda^c : \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - c_{2N}\hat{\sigma} g(x) < \lambda - \tilde{\boldsymbol{x}}^T \boldsymbol{\beta} \\
& \qquad \text{and } \forall\, x \in G_\lambda : \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + c_{2N}\hat{\sigma} g(x) \geq \lambda - \tilde{\boldsymbol{x}}^T \boldsymbol{\beta} \Big\} \\
\geq \;& \mathrm{P}\left\{ \forall\, x \in G_\lambda^c : \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - c_{2N}\hat{\sigma} g(x) < 0 \text{ and } \forall\, x \in G_\lambda : \tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + c_{2N}\hat{\sigma} g(x) \geq 0 \right\} \\
= \;& \mathrm{P}\left\{ \forall\, x \in G_\lambda^c : \frac{\tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\hat{\sigma}}{\sqrt{\tilde{\boldsymbol{x}}^T (X^T X)^{-1} \tilde{\boldsymbol{x}}}} < c_{2N} \text{ and } \forall\, x \in G_\lambda : \frac{\tilde{\boldsymbol{x}}^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\hat{\sigma}}{\sqrt{\tilde{\boldsymbol{x}}^T (X^T X)^{-1} \tilde{\boldsymbol{x}}}} \geq -c_{2N} \right\} \\
= \;& \mathrm{P}\left\{ \forall\, x \in G_\lambda^c : \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} < c_{2N} \text{ and } \forall\, x \in G_\lambda : \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} \geq -c_{2N} \right\} \\
= \;& \mathrm{P}\left\{ \forall\, x \in [l, m) : \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} < c_{2N} \text{ and } \forall\, x \in [m, u] : \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} \geq -c_{2N} \right\} \qquad (7)
\end{aligned}$$

where the second equality above follows directly from the definitions of $\hat{G}_{\lambda,2l,N}$ and $\hat{G}_{\lambda,2u,N}$ in (3), the inequality above follows directly from the definition of G_λ (and G_λ^c), $P = (X^T X)^{-1/2}$, and $\mathbf{T} = P^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/(\hat{\sigma}/\sigma)$ has a standard bivariate t distribution with covariance matrix I_2 and df ν (cf. Genz and Bretz, 2009). Also note that the inequality above approaches equality as the regression line $Y = \beta_0 + \beta_1 x$ approaches the level line $Y = \lambda$, and so the minimum of the probability in (6) over this **configuration** is given by the probability in (7).

For the probability in (7), the constraint $\frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} < c_{2N}$ for a given $P\tilde{\boldsymbol{x}} = P(1, x)^T = (\mathbf{p}_0, \mathbf{p}_1)(1, x)^T = \mathbf{p}_0 + x\mathbf{p}_1$ restricts $\mathbf{T}$ to the half plane on the same side as the origin and bounded by the straight line that is perpendicular to the vector $\mathbf{p}_0 + x\mathbf{p}_1$ and c_{2N} distance (in the direction of $\mathbf{p}_0 + x\mathbf{p}_1$) from the origin. Hence $\left\{ \forall\, x \in [l, m) : \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} < c_{2N} \right\}$ restricts $\mathbf{T}$ to the area that is bounded by the (thick) curve-line in red and contains the origin, which is depicted in Figure 1(a). The curve-line is formed by the arc $\left\{ \mathbf{T} : \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} = c_{2N} \text{ for } x \in [l, m) \right\}$ bounded by the vectors $\mathbf{p}_0 + l\mathbf{p}_1$ and $\mathbf{p}_0 + m\mathbf{p}_1$, the half line that is tangent to the arc at one end corresponding to the vector $\mathbf{p}_0 + l\mathbf{p}_1$ and extends from that end to infinity, and the half line that is tangent to the arc at the other end corresponding to the vector $\mathbf{p}_0 + m\mathbf{p}_1$ and extends from that end to infinity.

Similarly, $\left\{ \forall\, x \in [m, u] : \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} \geq -c_{2N} \right\}$ restricts $\mathbf{T}$ to the area that is bounded by the (thick) curve-line in blue and contains the origin, which is depicted in Figure 1(b). The curve-line is formed by the arc $\left\{ \mathbf{T} : \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} = -c_{2N} \text{ for } x \in [m, u] \right\}$ bounded by the vectors $-\mathbf{p}_0 - m\mathbf{p}_1$ and $-\mathbf{p}_0 - u\mathbf{p}_1$, the half line that is tangent to the arc at one end corresponding to the vector $-\mathbf{p}_0 - m\mathbf{p}_1$ and extends from the end to infinity, and the half line that is tangent to the arc at the other end corresponding to the vector $-\mathbf{p}_0 - u\mathbf{p}_1$ and extends from the end to infinity.

Hence all the constraints in (7) restrict $\mathbf{T}$ to the region bounded by the two (thick) curve-lines given in Figures 1(a) and 1(b), which is depicted in Figure 1(c). This region can be partitioned into four sub-regions: a half-stripe, two fans (shaded), and the remaining sub-region whose shape is depicted in Figure 1(d).

To calculate the probabilities of $\mathbf{T}$ in these regions, the following facts are used. First, the probability distribution of $\mathbf{T}$ is rotational invariant, that is, the probability of $\mathbf{T}$ in any given region is the same as the probability of $\mathbf{T}$ in the region that is resultant from rotating the given region around the origin by any angle. Second, let $(\|\mathbf{T}\|, \theta_{\mathbf{T}})$ be the polar coordinates of $\mathbf{T}$ then the cdf of $\|\mathbf{T}\|$ is given by

$$F_{\|\mathbf{T}\|}(x) = 1 - (1 + x^2/\nu)^{-\nu/2}, \quad x > 0, \tag{8}$$

$\theta_{\mathbf{T}}$ has a uniform distribution on the interval $[0, 2\pi)$, and $\|\mathbf{T}\|$ and $\theta_{\mathbf{T}}$ are independent random variables (cf. Liu, 2010, pp.18-19).

Since the pdf of $\mathbf{T} = (T_1, T_2)^T$ is rotational invariant, the probability of $\mathbf{T}$ in the half-stripe is the same as the probability of $\mathbf{T}$ in the half-stripe that is resulted from rotating the original half-stripe so that the end of the original half-stripe is on the T_1-axis. This probability is then equal to

$$\frac{1}{2}\mathrm{P}\left\{-c_{2N} < T_1 < c_{2N}\right\} = \mathrm{P}\left\{0 < T_1 < c_{2N}\right\} = \mathtt{pt}(c_{2N}, \nu) - 0.5 \tag{9}$$

where $\mathtt{pt}(\cdot, \nu)$ denotes the cdf of the univariate t distribution with df ν.

Again due to the rotational invariance of the $\mathbf{T}$ probability distribution, the probability of $\mathbf{T}$ in the two fans is equal to the probability of $\mathbf{T}$ in the big fan formed by these two fans (via rotating one fan to next to the other). The angle $\phi \in (0, \pi]$ between the two edges of the big fan is given by

$$\cos\phi = \left\{ P \begin{pmatrix} 1 \\ l \end{pmatrix} \right\}^T \left\{ P \begin{pmatrix} 1 \\ u \end{pmatrix} \right\} \Big/ \left\| P \begin{pmatrix} 1 \\ l \end{pmatrix} \right\| \left\| P \begin{pmatrix} 1 \\ u \end{pmatrix} \right\|. \tag{10}$$

Hence the probability of $\mathbf{T}$ in the big fan is equal to

$$\frac{\phi}{2\pi}\mathrm{P}\left\{\|\mathbf{T}\| < c_{2N}\right\} = \frac{\phi}{2\pi}\left(1 - (1 + c_{2N}^2/\nu)^{-\nu/2}\right) \tag{11}$$

where the equality above follows directly from (8).

For the probability of $\mathbf{T}$ in the remaining region, the remaining region can be rotated to the position so that it is symmetric about the T_1-axis as depicted in Figure 1(d). Then the probability of $\mathbf{T}$ in this region is equal to

$$\begin{aligned}
& 2\mathrm{P}\left\{ \theta_{\mathbf{T}} \in [0, (\pi-\phi)/2],\ \ (\cos((\pi-\phi)/2), \sin((\pi-\phi)/2))\, \mathbf{T} < c_{2N} \right\} \\
= \;& 2\mathrm{P}\left\{ \theta_{\mathbf{T}} \in [0, (\pi-\phi)/2],\ \ \|\mathbf{T}\| < c_{2N}/\cos((\pi-\phi)/2 - \theta_{\mathbf{T}}) \right\} \\
= \;& 2\int_0^{(\pi-\phi)/2} \frac{1}{2\pi} \mathrm{P}\left\{ \|\mathbf{T}\| < c_{2N}/\cos((\pi-\phi)/2-\theta) \right\} d\theta \\
= \;& \frac{1}{\pi}\int_0^{(\pi-\phi)/2} \left(1 - \left(1 + \frac{c_{2N}^2}{\nu \sin^2(\phi/2+\theta)} \right)^{-\nu/2} \right) d\theta
\end{aligned} \tag{12}$$

where the last equality follows directly from (8).

Adding the expressions (9-12) gives the probability in (7) equal to

$$\mathtt{pt}(c_{2N}, \nu) - \frac{\phi}{2\pi}\left(1 + c_{2N}^2/\nu\right)^{-\nu/2} - \frac{1}{\pi}\int_0^{(\pi-\phi)/2} \left(1 + \frac{c_{2N}^2}{\nu \sin^2(\phi/2+\theta)} \right)^{-\nu/2} d\theta\,. \tag{13}$$

It is clear that this expression depends only on X, n, c_{2N} and $[l, u]$, but not on the value $m \in (l, u)$ or $\sigma > 0$. Hence the minimum probability over $\boldsymbol{\beta}$ in **Configuration 1** is given by (13), which is also the expression given in the theorem.

Second, consider **Configuration 2**: $\boldsymbol{\beta} \in \Re^2$ is such that the regression line $Y = \beta_0 + \beta_1 x$ crosses the level line $Y = \lambda$ at $x = m \in (l, u)$, $\beta_0 + \beta_1 x > \lambda$ for $x \in [l, m)$, and $\beta_0 + \beta_1 x < \lambda$ for $x \in (m, u]$. An argument similar to that for **Configuration 1** above shows that the minimum probability over $\boldsymbol{\beta}$ in **Configuration 2** is also given by (13).

Third, consider **Configuration 3**: $\boldsymbol{\beta} \in \Re^2$ is such that the regression line $Y = \beta_0 + \beta_1 x$ is not below the level line $Y = \lambda$ over the interval $x \in [l, u]$. An argument similar to that for **Configuration 1** above shows that the minimum probability over $\boldsymbol{\beta}$ in **Configuration 3** is strictly larger than (13) (and attained when the regression line approaches the level line). The details are omitted to save space.

Finally, consider **Configuration 4**: $\boldsymbol{\beta} \in \Re^2$ is such that the regression line $Y = \beta_0 + \beta_1 x$ is not above the level line $Y = \lambda$ over the interval $x \in [l, u]$. An argument similar to that for **Configuration 1** above shows that the minimum probability over $\boldsymbol{\beta}$ in **Configuration 4** is the same as the minimum probability over $\boldsymbol{\beta}$ in **Configuration 3** and strictly larger than (13).

It is clear that the four configurations above form a partition of all possible $\boldsymbol{\beta} \in \Re^2$. The minimum probabilities for the four configurations given above therefore establish that the minimum probability in (4) is given by (13). This completes the proof of the theorem. ♣

The critical constant c_{2N} can therefore easily be solved numerically from the equation (4) with the minimum probability replaced with (13). The `R` code for computing c_{2N} and all the results in the example in the next section can be downloaded from the authors' websites.

3 Simultaneity in λ

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 $[\hat{G}_{\lambda,2l}, \hat{G}_{\lambda,2u}]$ or $[\hat{G}_{\lambda,2l,N}, \hat{G}_{\lambda,2u,N}]$ as a confidence set G_λ for each given value of λ. It is shown in Wan *et al.* (2022) that

$$\inf_{\boldsymbol{\beta}\in\Re^{p+1},\,\sigma>0} \mathrm{P}\left\{ \hat{G}_{\lambda,2l} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u}\ \forall\lambda\in\Re^1 \right\} \geq 1-\alpha\,, \tag{14}$$

which means that the joint confidence level of the confidence sets $[\hat{G}_{\lambda_1,2l}, \hat{G}_{\lambda_1,2u}]$, $[\hat{G}_{\lambda_2,2l}, \hat{G}_{\lambda_2,2u}]$, $\cdots$ for any sequence of λ-values $\lambda_1, \lambda_2, \cdots$ is at least $1-\alpha$.

The question is whether this property of "simultaneity in λ" also holds for the new confidence sets $[\hat{G}_{\lambda,2l,N}, \hat{G}_{\lambda,2u,N}]$. The answer is negative as shown in this section.

Consider the confidence set of the form

$$\hat{G}_{\lambda,2u,S} = \left\{ x\in K:\ \tilde{\boldsymbol{x}}^T\hat{\boldsymbol{\beta}} + c_{2S}\hat{\sigma}g(x) \geq \lambda \right\},\ \hat{G}_{\lambda,2l,S} = \left\{ x\in K:\ \tilde{\boldsymbol{x}}^T\hat{\boldsymbol{\beta}} - c_{2S}\hat{\sigma}g(x) \geq \lambda \right\}. \tag{15}$$

Note that the only difference between this confidence set, $[\hat{G}_{\lambda,2l}, \hat{G}_{\lambda,2u}]$ and $[\hat{G}_{\lambda,2l,N}, \hat{G}_{\lambda,2u,N}]$ is the critical constant used. Next we determine the critical constant c_{2S} so that

$$\inf_{\boldsymbol{\beta}\in\Re^{p+1},\,\sigma>0} \mathrm{P}\left\{ \hat{G}_{\lambda,2l,S} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u,S}\ \forall\lambda\in\Re^1 \right\} = 1-\alpha\,, \tag{16}$$

it is asserted in the theorem below that $c_{2S} = c_2$.

Theorem 2. The critical constant c_{2S} that satisfies (16) is given by $c_{2S} = c_2$, where c_2 is the critical constant of the two-sided simultaneous band in (1).

Proof. At the configuration $\boldsymbol{\beta} = (\beta_0, 0)^T$ where β_0 is a given constant, it is clear that

$$\begin{aligned}
&\left\{ \hat{G}_{\lambda,2l,S} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u,S}\ \forall\lambda\in\Re^1 \right\} \\
=\ & \cap_{\lambda\in\Re^1}\left\{ \hat{G}_{\lambda,2l,S} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u,S} \right\} \\
=\ & \left(\cap_{\lambda\leq\beta_0}\left\{ \hat{G}_{\lambda,2l,S} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u,S}\right\}\right) \cap \left(\cap_{\lambda>\beta_0}\left\{ \hat{G}_{\lambda,2l,S} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u,S}\right\}\right).
\end{aligned} \tag{17}$$

Now for each $\lambda \leq \beta_0$ we have $G_\lambda = [l, u]$ and $G_\lambda^c = \emptyset$. Hence an argument similar to that for establishing (7) from (6) in the last section implies that

$$\left\{ \hat{G}_{\lambda,2l,S} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u,S} \right\} \supseteq \left\{ \forall\, x \in [l, u] : \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} \geq -c_{2S} \right\},$$

and in particular the two sets above are equal at $\lambda = \beta_0$. It follows immediately that

$$\cap_{\lambda \leq \beta_0} \left\{ \hat{G}_{\lambda,2l,S} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u,S} \right\} = \left\{ \forall\, x \in [l, u] : \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} \geq -c_{2S} \right\}. \tag{18}$$

Next for each $\lambda > \beta_0$ we have $G_\lambda = \emptyset$ and $G_\lambda^c = [l, u]$. Again an argument similar to that for establishing (7) from (6) in the last section implies that

$$\left\{ \hat{G}_{\lambda,2l,S} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u,S} \right\} \supseteq \left\{ \forall\, x \in [l, u] : \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} < c_{2S} \right\},$$

and in particular the two sets are equal at $\lambda = \beta_0^+$; here β_0^+ denotes a number that is infinitesimally close to β_0 from right. It follows immediately that

$$\cap_{\lambda > \beta_0} \left\{ \hat{G}_{\lambda,2l,S} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u,S} \right\} = \left\{ \forall\, x \in [l, u] : \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} < c_{2S} \right\}. \tag{19}$$

The combination of (17-19) implies that, at $\boldsymbol{\beta} = (\beta_0, 0)^T$,

$$\left\{ \hat{G}_{\lambda,2l,S} \subseteq G_\lambda \subseteq \hat{G}_{\lambda,2u,S}\ \forall \lambda \in \Re^1 \right\} = \left\{ \forall\, x \in [l, u] : \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} \geq -c_{2S} \text{ and } \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} < c_{2S} \right\}.$$

Hence, in order to satisfy the simultaneity requirement in (16), c_{2S} cannot be smaller than the value that solve the equation

$$\mathrm{P}\left\{ \forall\, x \in [l, u] : \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} \geq -c_{2S} \text{ and } \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} < c_{2S} \right\} = 1 - \alpha\,.$$

Note on the other hand that c_2 of the two-sided simultaneous confidence band in (1) satisfies

$$\mathrm{P}\left\{ \forall\, x \in [l, u] : \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} \geq -c_2 \text{ and } \frac{\{P\tilde{\boldsymbol{x}}\}^T \mathbf{T}}{\|P\tilde{\boldsymbol{x}}\|} < c_2 \right\} = 1 - \alpha\,.$$

It follows therefore that c_{2S} cannot be smaller than c_2. This and the fact that the confidence sets $[\hat{G}_{\lambda,2l}, \hat{G}_{\lambda,2u}]$ constructed from the two-sided simultaneous confidence band do have the simultaneity property (14) imply $c_{2S} = c_2$, that is, $[\hat{G}_{\lambda,2l,S}, \hat{G}_{\lambda,2u,S}]$ must be given by $[\hat{G}_{\lambda,2l}, \hat{G}_{\lambda,2u}]$ in order to satisfy (16). This completes the proof of the theorem. ♣

In conclusion, the confidence sets $[\hat{G}_{\lambda,2l,S}, \hat{G}_{\lambda,2u,S}]$ must be the same as the confidence sets $[\hat{G}_{\lambda,2l}, \hat{G}_{\lambda,2u}]$ in order to satisfy the simultaneity property (16). Since $c_{2N} < c_2$, the confidence sets $[\hat{G}_{\lambda,2l,N}, \hat{G}_{\lambda,2u,N}]$ do not satisfy the simultaneity property.

4 An example

Kleinbaum, Kupper, Muller, and Nizam (1998, p. 192) provided a dataset on changes in systolic blood pressure ($y = SBP$, in mm Hg) with age ($x = age$, in years) for a group of 40 males. The data points and the fitted regression line, which is $Y = 110.04 + 0.96x$, are plotted in Figure 2(a). From the usual model diagnostic check, the simple linear regression model is a quite reasonable fit, with $R^2 = 0.7447$ and $\hat{\sigma} = 8.479$ on $\nu = 38$ df.

Since the minimum age $\min(x_i) = 18$ and the maximum age $\max(x_i) = 70$, we set $K = [l, u] = [18, 70]$ for illustration below. Suppose we are interested in identifying the ages within K for which the mean SPB is larger than or equal to 160, that is, the level set $G_{160} = \{\, x \in K : \beta_0 + \beta_1 x \geq 160 \,\}$ with $\lambda = 160$. The level line $Y = \lambda = 160$ is also plotted in Figure 2(a).

To compute the conservative confidence set $[\hat{G}_{\lambda,2l}, \hat{G}_{\lambda,2u}]$ in (2), we first need to compute the critical constant c_2 for the tw-sided simultaneous confidence band in (1). For $\alpha = 0.05$, c_2 is computed to be 2.514. The confidence band is plotted in Figure 2(b), given by the two curve-lines around the fitted regression line. The confidence set $[\hat{G}_{\lambda,2l}, \hat{G}_{\lambda,2u}]$ is plotted in the same figure, with $\hat{G}_{\lambda,2l} = [56.10, 70]$ indicated by the shorter line-segment, and $\hat{G}_{\lambda,2u} = [48.44, 70]$ indicated by the longer line-segment.

To compute the exact confidence set $[\hat{G}_{\lambda,2l,N}, \hat{G}_{\lambda,2u,N}]$ in (3), we first need to compute the critical constant c_{2N}. It is solved from equation (4), with the minimum probability given by expression (13). For $\alpha = 0.05$, c_{2N} is computed to be 2.327, which is about 7.5% smaller than $c_2 = 2.514$. The two functions $\tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} \pm c_{2N} \hat{\sigma} g(x)$ are plotted in Figure 2(c), given by the two long-dash curve-lines around the fitted regression line. The exact confidence set $[\hat{G}_{\lambda,2l,N}, \hat{G}_{\lambda,2u,N}]$ is plotted in the same figure, with $\hat{G}_{2l} = [55.75, 70]$ indicated by the shorter line-segment, and $\hat{G}_{\lambda,2u} = [48.70, 70]$ indicated by the longer line-segment.

As expected, $c_{2N} < c_2$. So the curve-lines $\tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} \pm c_{2N} \hat{\sigma} g(x)$ are closer to the fitted regression line than the curve-lines $\tilde{\boldsymbol{x}}^T \hat{\boldsymbol{\beta}} \pm c_2 \hat{\sigma} g(x)$, and $\hat{G}_{\lambda,2l} \subset \hat{G}_{\lambda,2l,N} \subset \hat{G}_{\lambda,2u,N} \subset \hat{G}_{\lambda,2u}$ even though the difference between the exact and conservative confidence sets is not that large in this example. This is demonstrated in Figure 2(d) where both the exact and conservative confidence sets are plotted.

If one is not fixated on one particular λ-value and would rather try several different λ-values, then one can construct confidence sets $[\hat{G}_{\lambda_1,2l}, \hat{G}_{\lambda_1,2u}], [\hat{G}_{\lambda_2,2l}, \hat{G}_{\lambda_2,2u}], \cdots$ for any sequence of $\lambda_1, \lambda_2, \cdots$. The result of Section 3 guarantees that the simultaneous confidence level of this sequence of confidence sets is still at least $1 - \alpha = 95\%$. However the simultaneous confidence level of the corresponding sequence of confidence sets $[\hat{G}_{\lambda_1,2l,N}, \hat{G}_{\lambda_1,2u,N}], [\hat{G}_{\lambda_2,2l,N}, \hat{G}_{\lambda_2,2u,N}], \cdots$ may be strictly less than $1 - \alpha = 95\%$ as pointed out in Section 3.

5 CONCLUSION AND DISCUSSION

Conservative two-sided confidence set for a level set is provided in Wan *et al.* (2022) for linear regression models, and based on a two-sided simultaneous confidence band. In this paper, an exact two-sided confidence set for a level set is constructed for a simple linear regression model. It is also shown that the exact two-sided confidence sets do not have the property of simultaneity in λ, while the confidence sets based on two-sided simultaneous confidence band do have the simultaneity property.

Construction of exact two-sided confidence sets for a level set for multiple linear or polynomial regressions is clearly of interest but much more challenging. It warrants further research.

It is also interesting to explore the construction of exact two-sided confidence sets using other forms of $g(x)$. While beyond the scope of this paper, it warrants further research.

References

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

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.

Genz, A. and Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities.* New York: Springer Lecture Notes in Statistics.

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.

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.

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

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.

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

Scott, C. and 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.

Wan, F., Liu, W. and Bretz, F. (2023). Confidence Sets for a level set in linear regression . *Statistics in Medicine*, to appear.

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. and 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.

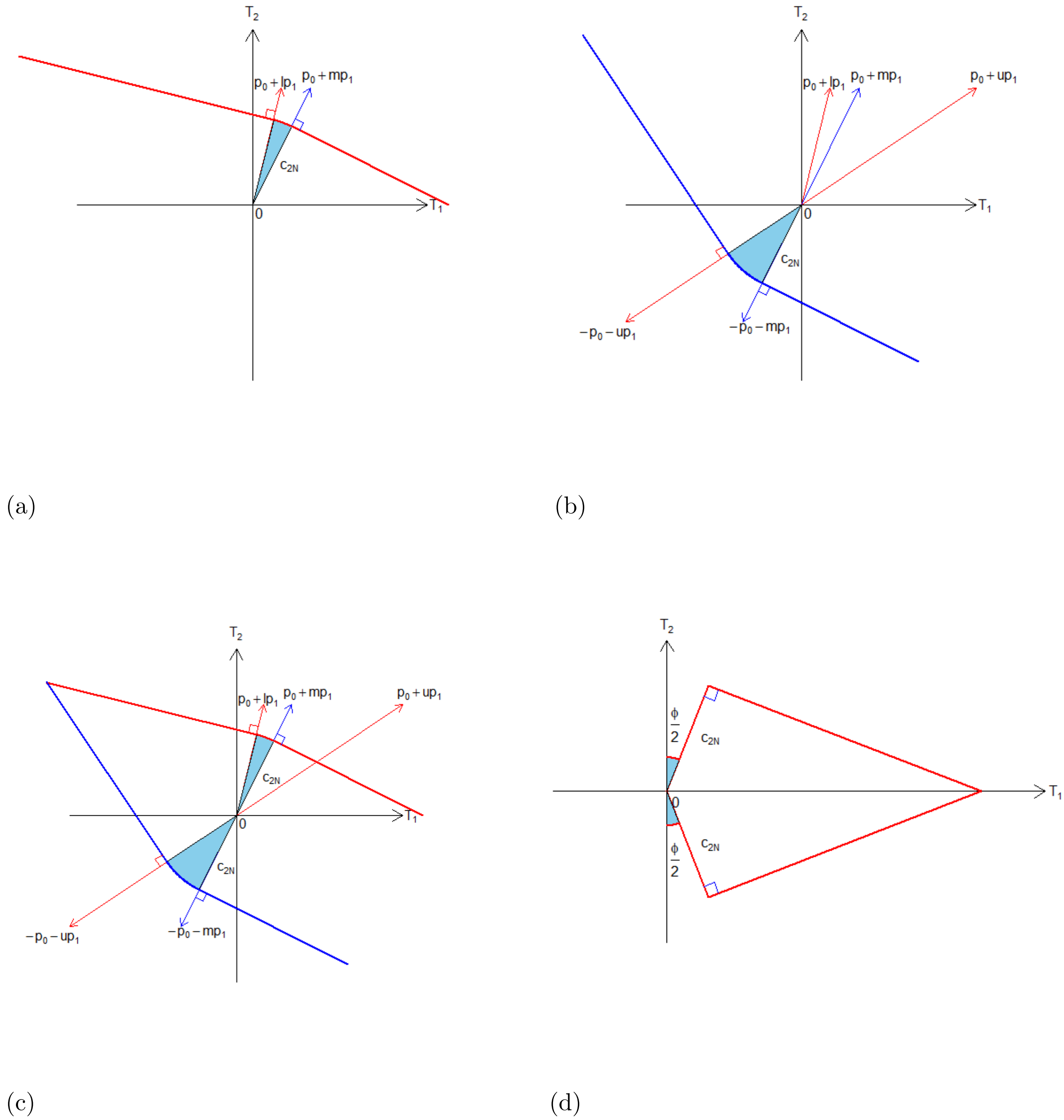
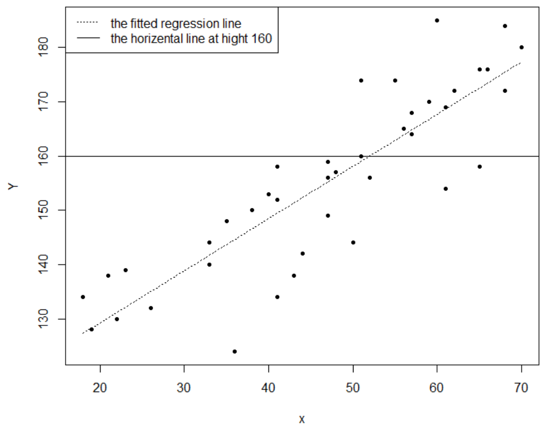
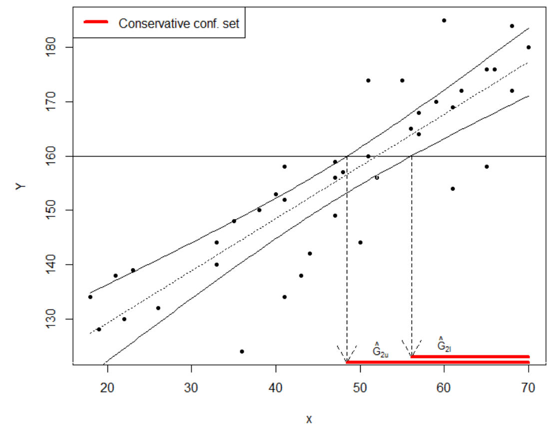


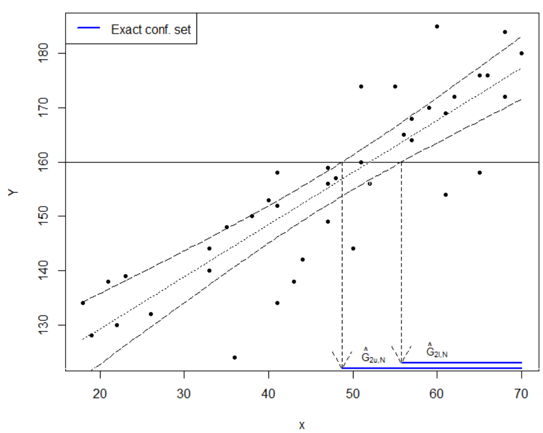
Figure 1: Computation of the probability in (6).



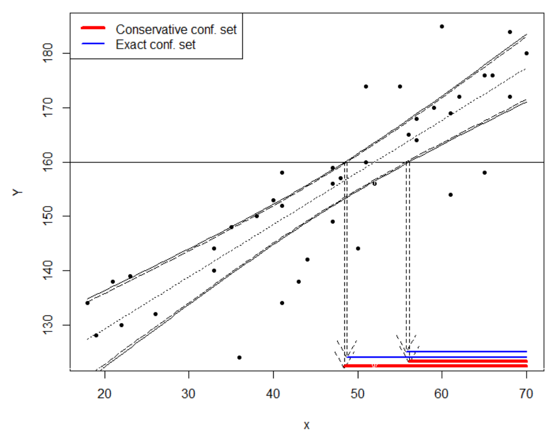
(a) Data points and fitted regression line



(b) Conservative confidence set $[\hat{G}_{\lambda,2l}, \hat{G}_{\lambda,2u}]$



(c) Exact confidence set $[\hat{G}_{\lambda,2l,N}, \hat{G}_{\lambda,2u,N}]$



(d) Both confidence sets

Figure 2: The 95% conservative and exact confidence sets in the example.