MAGISTERARBEIT / MASTER'S THESIS

Titel der Magisterarbeit / Title of the Master's Thesis

# Conditional expected length of post-selection confidence intervals

verfasst von / submitted by

Danijel Kivaranovic, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magister der Sozial- und Wirtschaftswissenschaften (Mag. rer. soc. oec.)

Wien, 2017 / Vienna, 2017

| | |
|---|---|
| Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet: | A 066 951 |
| Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet: | Magisterstudium Statistik |
| Betreut von / Supervisor: | Univ.-Prof. Mag. Dr. Hannes Leeb |

# Contents

# 1 Introduction

It is common practice in applied statistics that the model fitted to the data is selected by a data-dependent model selector. The same data is, then, used to draw inferences from the selected model. Traditional theory of statistical inference, however, assumes that the model generating the data is a priori known. Therefore, the use of "naive" confidence intervals (that is, intervals that do not take the model selection into account) can lead to highly misleading interpretations as the actual coverage probability of the true underlying parameter can be significantly below the nominal level; see among others Kabaila and Leeb (2006); Leeb and Pötscher (2005, 2006*a*,*b*, 2008); Pötscher (1991); Kabaila (1998).

Berk et al. (2013) proposed the so-called PoSI-intervals, which guarantee user-specified minimum coverage probability irrespective of the model selector that was applied to obtain the model. Yet, these authors reformulated the coverage target of the confidence interval; instead of targeting the true parameter, a data-dependent target is defined as the quantity of interest (see Section 3 for the definition of the data-dependent target). As the PoSI-intervals offer protection against any possible model selector, they are typically too wide and very conservative (in the sense that the actual coverage probability of the data-dependent target is larger than the nominal level). Leeb et al. (2015) compared the coverage probabilities of the PoSI-intervals (and variations of it) with the "naive" intervals in several scenarios. While the PoSI-intervals were always conservative, the actual coverage probability of the "naive" interval was quite close to the nominal level, irrespective of the true parameter. Thus, it seems that the "bad" coverage behavior of the "naive" interval when targeting the true parameter is not present to the same extent when targeting the data-dependent target. Computation time of the PoSI-intervals increases exponentially with the number of variables under consideration; therefore, exact calculation of the PoSI-intervals is only feasible if the number of candidate models is small.

In contrast to the PoSI-intervals, which are valid after any model selector, Lee et al. (2016) recently proposed confidence intervals that have exact coverage probability of the data-dependent target as defined in Berk et al. (2013), conditional on the model that was selected by the lasso. Let denote by $\{\hat{M} = M\}$ the set where the model $M$ is selected and by $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ the set where the model $M$ and the sign vector $s_M$ of the non-zero lasso estimates is selected (see Section 3 for the exact definition). It is shown in Lee et al. (2016) that the event $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ can be characterized as a polyhedron and therefore the set $\{\hat{M} = M\}$ as a union of polyhedra. Assuming a Gaussian model, these authors developed a general framework for valid inference, conditional on polyhedral constraints. Two types of intervals were proposed: a confidence interval that has exact coverage probability conditional on $\{\hat{M} = M\}$ and another confidence interval that has exact coverage probability conditional on $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$. It was argued in Lee et al. (2016) that the latter confidence interval is typically wider as one conditions on a smaller set. However, the computation time of the former increases exponentially with the number of variables selected while the computation time of the latter increases only linearly. It is criticized in Bachoc et al. (2016) that the coverage probability of the confidence intervals is only exact if the tuning parameter is fixed, while when the tuning parameter is chosen by cross-validation (which is typically the case in applications), the actual coverage probability can fall below the nominal level. Further, it is shown in a simulation in Bachoc et al. (2016) that in certain scenarios the confidence intervals by Lee et al. (2016) are considerably wider than the PoSI-intervals, even though the PoSI-intervals offer protection against any model selector while the confidence intervals in Lee et al. (2016) are only valid after model selection with the lasso. Certain optimality properties of the general framework developed in Lee et al. (2016) are considered in Fithian et al. (2017). The same framework is used in Tibshirani et al. (2016) for valid inference after stepwise regression and least angle regression.

This thesis deals with the length properties of the confidence intervals proposed by Lee et al. (2016) and is structured as follows: Section 2 discusses confidence intervals for the mean parameter of a normal distribution that is truncated to a fixed set. This is an essential first step, as the methodology in Lee et al. (2016) highly depends on a normal distribution that is truncated to a random set. It will be shown that the conditional expected length of practically any confidence interval with exact conditional coverage is $+\infty$ if the truncation set is at least bounded from below or above. On the other hand, if the truncation set is unbounded, then the length of the confidence intervals is bounded and an upper bound for the conditional expected length is derived. Section 3 provides a detailed description of the approach in Lee et al. (2016). Further, it will be shown that the random truncation sets obtained by conditioning on the smaller set $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ are at least bounded from below or above with probability 1. On the other hand, the random truncation sets obtained by conditioning on the larger set $\{\hat{M} = M\}$ are unbounded with probability 1. Using the results of Section 2, we will then show that the confidence interval with exact conditional coverage on $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ has infinite conditional expected length, while the confidence interval with exact conditional coverage on $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ has finite conditional expected length. Conclusions are drawn in Section 4.

# 2 Expected length of confidence intervals for the mean in the truncated normal case

## 2.1 The truncated normal distribution

Let $\phi(x)$ denote the probability density function (p.d.f.) and $\Phi(x)$ the cumulative distribution function (c.d.f.) of the standard normal distribution. Let $X$ be a normally distributed random variable with mean $\theta \in \mathbb{R}$ and variance $\sigma^2 \in (0, +\infty)$. The variance $\sigma^2$ is assumed to be known. The p.d.f. of $X$ is denoted by $\phi_{\theta,\sigma^2}(x)$ and the c.d.f. is denoted by $\Phi_{\theta,\sigma^2}(x)$. The definition of $\phi_{\theta,\sigma^2}(x)$ and $\Phi_{\theta,\sigma^2}(x)$ is given below:

$$\phi_{\theta,\sigma^2}(x) = \frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right) \quad \text{and} \quad \Phi_{\theta,\sigma^2}(x) = \Phi\left(\frac{x-\theta}{\sigma}\right). \tag{2.1}$$

We write $\mathbb{P}_{\theta,\sigma^2}(\cdot)$ and $\mathbb{E}_{\theta,\sigma^2}(\cdot)$ to denote the corresponding probability measure and expectation operator, respectively. Let $S = \bigcup_{i=1}^{K}(a_i, b_i)$, with $K < +\infty$ and $-\infty \leq a_1 < b_1 < \cdots < a_K < b_K \leq +\infty$, denote the truncation set. For ease of discussion, we will refer to any set $(a, b)$, where $-\infty \leq a < b \leq +\infty$, as an interval. Note that the dependence of $S$ on $a_1, b_1, \ldots, a_K, b_K$ is suppressed in the notation. The distribution of $X$ conditional on $S$ is called truncated normal distribution and we write

$$X|\{X \in S\} \sim TN(\theta, \sigma^2, S). \tag{2.2}$$

The probability of the conditioning event $\{X \in S\}$ is strictly positive because for all $k = 1, \ldots, K$, the interval $(a_k, b_k)$ has positive Lebesgue measure. The terms "$X|\{X \in S\}$", "$X$ conditional on $S$", and "$X$ truncated to $S$" will be used interchangeably. The corresponding conditional probability measure and conditional expectation operator are denoted by $\mathbb{P}_{\theta,\sigma^2}(\cdot|X \in S)$ and by $\mathbb{E}_{\theta,\sigma^2}(\cdot|X \in S)$, respectively. The distribution of $X$ truncated to the closure of $S$, that is, to the union of the closed intervals $[a_k, b_k]$, is the same as the distribution of $X|\{X \in S\}$, since the set of endpoints $\{a_1, b_1, \ldots, a_K, b_K\}$ is a Lebesgue null set. Clearly, conditioning on the union of $S$ with any set of Lebesgue measure zero (e.g. a set of countably many points) does not change the conditional distribution of $X$. The special case where $X$ is truncated to a single interval is obtained by setting $K = 1$. If $K = 1$, $a_1 = -\infty$ and $b_1 = +\infty$, then $X|\{X \in S\}$ is distributed as $X$, this means normally distributed. The truncation set $S$ is bounded from below and above if both $a_1 > -\infty$ and $b_K < +\infty$; it is bounded on one side if either $a_1 > -\infty$ or $b_K < +\infty$ and it is unbounded if both $a_1 = -\infty$ and $b_K = +\infty$.

We write $f^S_{\theta,\sigma^2}(x)$ for the p.d.f. and $F^S_{\theta,\sigma^2}(x)$ for the c.d.f. of the truncated normal distribution. Then the p.d.f. $f^S_{\theta,\sigma^2}(x)$ is defined as follows:

$$f^S_{\theta,\sigma^2}(x) = \frac{\frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right)}{\sum\limits_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)} \mathbb{1}_S(x). \tag{2.3}$$

The c.d.f $F^S_{\theta,\sigma^2}(x)$ is defined as follows:

$$F^S_{\theta,\sigma^2}(x) = \begin{cases} 0 & \text{if } x \leq a_1 \\ \dfrac{\Phi\left(\frac{x-\theta}{\sigma}\right) - \Phi\left(\frac{a_k-\theta}{\sigma}\right) + \sum\limits_{i=1}^{k-1} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\sum\limits_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)} & \text{if } a_k < x < b_k \text{ for } k = 1, \ldots, K \\ \dfrac{\sum\limits_{i=1}^{k} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\sum\limits_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)} & \text{if } b_k \leq x \leq a_{k+1} \text{ for } k = 1, \ldots, K-1 \\ 1 & \text{if } x \geq b_K. \end{cases} \tag{2.4}$$

If $k = 1$, then the expression $\sum\limits_{i=1}^{k-1} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)$ is to be interpreted as 0. We adopt the convention that $\Phi(-\infty) = 0$ and $\Phi(+\infty) = 1$. Thus, if $K = 1$, $a_1 = -\infty$ and $b_K = +\infty$, then it follows that $f^S_{\theta,\sigma^2}(x) = \phi_{\theta,\sigma^2}(x)$ and $F^S_{\theta,\sigma^2}(x) = \Phi_{\theta,\sigma^2}(x)$. The support of the p.d.f. $f^S_{\theta,\sigma^2}(x)$ is the closure of the set $S$. The denominator of $f^S_{\theta,\sigma^2}(x)$ is simply the probability of the event $\{X \in S\}$, that is, $P_{\theta,\sigma^2}(X \in S)$ and does not depend on $x$. As the c.d.f. $F^S_{\theta,\sigma^2}(x)$ has density $f^S_{\theta,\sigma^2}(x)$, it follows that $F^S_{\theta,\sigma^2}(x)$ is continuous and monotonically increasing in $x$ for every $\theta \in \mathbb{R}$ and every $\sigma^2 \in (0, +\infty)$. For every $k = 1, \ldots, K$, the interval $(a_k, b_k)$ has positive Lebesgue measure and for every $x \in (a_k, b_k)$, the p.d.f $f^S_{\theta,\sigma^2}(x)$ is strictly positive; therefore it follows that $F^S_{\theta,\sigma^2}(x)$ is strictly monotonically increasing in $x$ on the interval $(a_k, b_k)$ for every $k = 1, \ldots, K$ and constant elsewhere. A comparison of the p.d.f. and c.d.f.

of a truncated normal distribution (solid line) and of a normal distribution (dashed line) is illustrated in Figure 2.1.

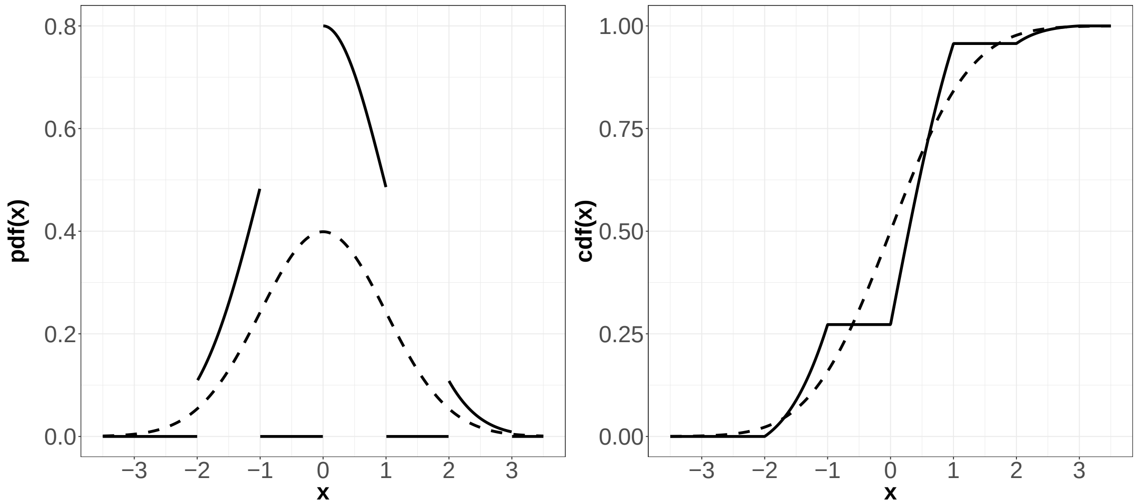


Figure 2.1: In the left panel the p.d.f. and in the right panel the c.d.f. of a truncated normal distribution (solid line) and a normal distribution (dashed line) are shown. The parameters of the normal distribution were $\theta = 0$ and $\sigma^2 = 1$ and the truncation set $S$ was set as follows: $S = (-2, 1) \cup (0, 1) \cup (2, 3)$.

## 2.2 Confidence intervals with exact conditional coverage probability

Let $\alpha$ with $0 < \alpha < 1$ be a fixed constant. Consider two functions $L, U : \mathbb{R} \to \mathbb{R}$ such that $L(x) \leq U(x)$ for all $x \in \mathbb{R}$. A confidence interval $[L(X), U(X)]$ for the parameter $\theta$ has exact coverage probability of $1 - \alpha$ if

$$\mathbb{P}_{\theta,\sigma^2}(L(X) \leq \theta \leq U(X)) = 1 - \alpha \quad \text{for all } \theta \in \mathbb{R},\ \sigma^2 \in (0, +\infty). \tag{2.5}$$

This means that, irrespective of the true parameter $\theta$ and $\sigma^2$, the coverage probability of $[L(X), U(X)]$ is exactly $1 - \alpha$. However, it does not necessarily follow that the coverage probability of $[L(X), U(X)]$ conditional on the set $S$, that is,

$$\mathbb{P}_{\theta,\sigma^2}(L(X) \leq \theta \leq U(X) | X \in S) \tag{2.6}$$

is also equal to $1 - \alpha$ for all $\theta \in \mathbb{R}$ and $\sigma^2 \in (0, \sigma^2)$. In fact, it is shown in the following example that the conditional coverage probability of a confidence interval with exact coverage probability of $1-\alpha$ can be exactly 0 for some values of $\theta$ and $\sigma^2$.

**Example 2.1.** Consider the standard confidence interval $[L(X), U(X)]$ for the mean parameter $\theta$ of the normally distributed random variable $X$, that is, the functions $L(x)$ and $U(x)$ are defined as follows:

$$L(x) = x - \sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right) \quad \text{and} \quad U(x) = x - \sigma\Phi^{-1}\left(\frac{\alpha}{2}\right). \tag{2.7}$$

Here $\Phi^{-1}(x)$ denotes the inverse function of $\Phi(x)$. The (unconditional) coverage probability of the confidence interval $[L(X), U(X)]$ is:

$$\begin{aligned}
\mathbb{P}_{\theta,\sigma^2}(L(X) \leq \theta \leq U(X)) &= \mathbb{P}_{\theta,\sigma^2}\left(\theta + \sigma\Phi^{-1}\left(\frac{\alpha}{2}\right) \leq X \leq \theta + \sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right) \\
&= \Phi_{\theta,\sigma^2}\left(\theta + \sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right) - \Phi_{\theta,\sigma^2}\left(\theta + \sigma\Phi^{-1}\left(\frac{\alpha}{2}\right)\right) \\
&= 1-\alpha \qquad \text{for all } \theta \in \mathbb{R},\ \sigma^2 \in (0,+\infty).
\end{aligned} \tag{2.8}$$

Therefore, the confidence interval $[L(X), U(X)]$ has exact coverage probability of $1-\alpha$. Conditional on the set $S$, $X$ is truncated normally distributed with c.d.f. $F^S_{\theta,\sigma^2}(x)$. Therefore, the conditional coverage probability of the confidence interval $[L(X), U(X)]$ is given by

$$\begin{aligned}
\mathbb{P}_{\theta,\sigma^2}(L(X) \leq \theta \leq U(X) | X \in S) &= \mathbb{P}_{\theta,\sigma^2}\left(\theta + \sigma\Phi^{-1}\left(\frac{\alpha}{2}\right) \leq X \leq \theta + \sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right) | X \in S\right) \\
&= F^S_{\theta,\sigma^2}\left(\theta + \sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right) - F^S_{\theta,\sigma^2}\left(\theta + \sigma\Phi^{-1}\left(\frac{\alpha}{2}\right)\right).
\end{aligned} \tag{2.9}$$

In contrast to the unconditional case, the probability in the last equation typically depends on the parameter $\theta$ and $\sigma^2$. If the truncation set $S$ is bounded from above (that is if $b_K < +\infty$), then, conditional on $S$, $U(X)$ is also bounded by $b_K - \sigma\Phi^{-1}\left(\frac{\alpha}{2}\right)$ with probability 1. Hence, it follows that if the underlying parameter $\theta$ is larger than $b_K - \sigma\Phi^{-1}\left(\frac{\alpha}{2}\right)$, then the conditional coverage probability of the interval $[L(X), U(X)]$ must be 0. The conditional coverage probability as a function of $\theta$ for a particular truncated distribution is shown in Figure 2.2.

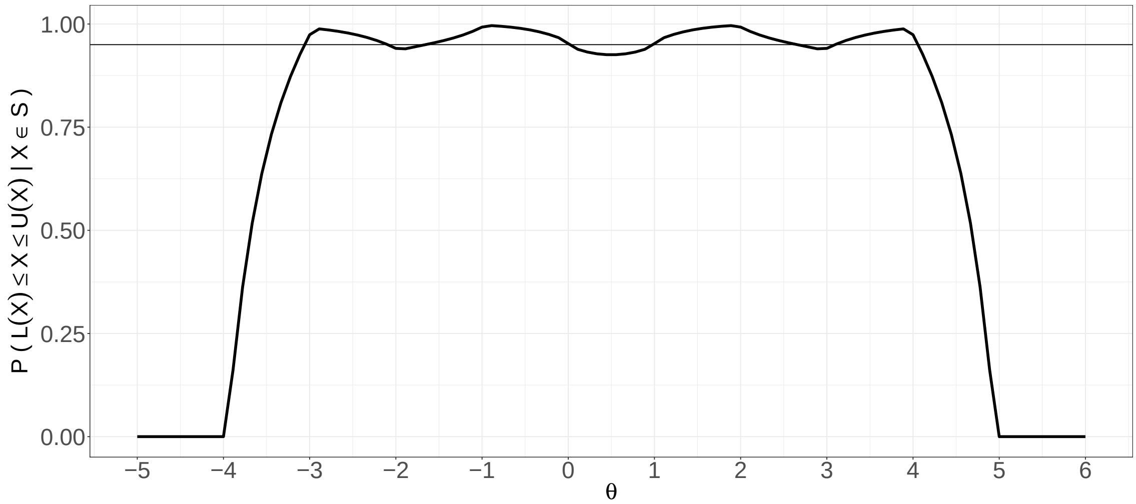


Figure 2.2: The conditional coverage probability of the standard confidence interval $[L(X), U(X)]$ as a function of the parameter $\theta$ when $X|\{X \in S\} \sim TN(\theta, 1, S)$ where $S = (-2, -1) \cup (0, 1) \cup (2, 3)$. The $\alpha$ was set to 0.05. Observe that the coverage probability converges to 0 as $\theta \to -\infty$ and $\theta \to +\infty$.

This example demonstrates that the condition in (2.5) is insufficient if one is interested in the conditional coverage probability of $\theta$. Consider the following definition which is the conditional analog of a confidence interval with exact coverage probability of $1 - \alpha$.

**Definition 2.2.** *A confidence interval* $[L(X), U(X)]$ *has exact conditional coverage probability of* $1 - \alpha$ *on the set* $S$ *if*

$$\mathbb{P}_{\theta,\sigma^2}(L(X) \leq \theta \leq U(X)|X \in S) = 1 - \alpha \quad \textit{for all } \theta \in \mathbb{R},\ \sigma^2 \in (0, +\infty). \tag{2.10}$$

Note that $\theta$ is not to be confused with the mean of $X$ conditional on $S$, that is, $E_{\theta,\sigma^2}(X|X \in S)$. The conditional mean is, in general, not equal to $\theta$. In the following we will show how to find a confidence interval with exact conditional coverage probability of $1 - \alpha$ on the set $S$. By definition, $X$ conditional on $S$ is truncated normally distributed with the continuous c.d.f. $F^S_{\theta,\sigma^2}(x)$. Therefore, the probability integral transformation implies that, conditional on $S$, $F^S_{\theta,\sigma^2}(X)$ is uniformly distributed on the unit interval, that is,

$$F^S_{\theta,\sigma^2}(X)|\{X \in S\} \sim U(0, 1). \tag{2.11}$$

Thus, conditional on $S$, the distribution of $F^S_{\theta,\sigma^2}(X)$ does not the depend on the unknown parameter $\theta$. For this reason, $F^S_{\theta,\sigma^2}(X)$ can be used as a test statistic for testing a hypothesized value of $\theta$. Typically, one rejects the null hypothesis (i.e. that the hypothesized value $\theta$ is the true parameter), if the test statistic $F^S_{\theta,\sigma^2}(X)$ is either too small (e.g. $F^S_{\theta,\sigma^2}(X) \le \frac{\alpha}{2}$) or too large (e.g. $F^S_{\theta,\sigma^2}(X) \ge 1 - \frac{\alpha}{2}$). Since $F^S_{\theta,\sigma^2}(X)$ is uniformly distributed under the null hypothesis, the acceptance probability (that is the probability of not rejecting the test) is:

$$\mathbb{P}_{\theta,\sigma^2}\left(\frac{\alpha}{2} \le F^S_{\theta,\sigma^2}(X) \le 1 - \frac{\alpha}{2} | X \in S\right) = 1 - \alpha \quad \text{for all } \theta \in \mathbb{R},\ \sigma^2 \in (0, +\infty). \tag{2.12}$$

A confidence set $C_S(x)$ is now obtained by inverting the previous test:

$$C_S(x) = \left\{\theta \in \mathbb{R} : \frac{\alpha}{2} \le F^S_{\theta,\sigma^2}(x) \le 1 - \frac{\alpha}{2}\right\} \quad \text{for all } x \in \mathbb{R}. \tag{2.13}$$

For all $\theta \in \mathbb{R}$ it follows that if $x \le a_1$, then $F^S_{\theta,\sigma^2}(x) = 0$, and if $x \ge b_K$, then $F^S_{\theta,\sigma^2}(x) = 1$. Therefore, if either $x \le a_1$ or $x \ge b_K$, it follows that $C_S(x) = \emptyset$. By definition of the set $C_S(x)$ it follows that $\theta \in C_S(x)$ if and only if $F^S_{\theta,\sigma^2}(x) \in \left[\frac{\alpha}{2}, 1 - \frac{\alpha}{2}\right]$. We know that the conditional probability of the latter event equals $1-\alpha$, thus it follows that

$$\mathbb{P}_{\theta,\sigma^2}\left(\theta \in C_S(X)\ |X \in S\right) = 1 - \alpha \quad \text{for all } \theta \in \mathbb{R},\ \sigma^2 \in (0, +\infty). \tag{2.14}$$

This means that the confidence set $C_S(X)$ has exact conditional coverage probability of $1-\alpha$ on the set $S$. The following lemma guarantees that the set $C_S(x)$ is, in fact, an interval for all x in the interval $(a_1, b_K)$, where $a_1$ denotes the leftmost boundary point and $b_K$ denotes the rightmost boundary point of the truncation set $S$.

**Lemma 2.3.** *For all $x \in (a_1, b_K)$, it follows that*

*(i)* $F^S_{\theta,\sigma^2}(x)$ *is continuous in* $\theta$*;*

*(ii)* $F^S_{\theta,\sigma^2}(x)$ *is strictly monotonically decreasing in* $\theta$*;*

*(iii)* $\lim\limits_{\theta\to+\infty} F^S_{\theta,\sigma^2}(x) = 0$*;*

*(iv)* $\lim\limits_{\theta\to-\infty} F^S_{\theta,\sigma^2}(x) = 1$.

The proof of this lemma requires several technical arguments and is postponed to Section 2.5. A direct consequence of Lemma 2.3 is that for every $x \in (a_1, b_K)$ and every $q$ with $0 < q < 1$, the equation

$$F^S_{\theta,\sigma^2}(x) = q \tag{2.15}$$

has a unique solution in $\theta$. This, in turn, implies that the set $C_S(x)$ is non-empty for all $x \in (a_1, b_K)$. As $F^S_{\theta,\sigma^2}(x)$ is strictly decreasing in $\theta$ by Lemma 2.3 (ii), it follows that the set $C_S(x)$ must be an interval for all $x \in (a_1, b_K)$. For every $x \in (a_1, b_K)$, let denote by $L_S(x)$ the lower bound and by $U_S(x)$ the upper bound of the set $C_S(x)$, that is, $C_S(x) = [L_S(x), U_S(x)]$. The continuity of $F^S_{\theta,\sigma^2}(x)$ as a function of $\theta$ (Lemma 2.3 (i)) implies that the lower bound $L_S(x)$ is the unique solution to the following equation:

$$F^S_{L_S(x),\sigma^2}(x) = 1 - \frac{\alpha}{2} \quad \text{for all } x \in (a_1, b_K). \tag{2.16}$$

On the other hand, again by continuity of $F^S_{\theta,\sigma^2}(x)$ as a function of $\theta$, it follows that the upper bound $U_S(x)$ is the unique solution to the following equation:

$$F^S_{U_S(x),\sigma^2}(x) = \frac{\alpha}{2} \quad \text{for all } x \in (a_1, b_K). \tag{2.17}$$

Since we are only interested in the coverage probability conditional on $S$, we simply set $L_S(x) = U_S(x) = 0$ for all $x \notin (a_1, b_K)$. As $F^S_{\theta,\sigma^2}(x)$ is strictly decreasing as a function of $\theta$ (Lemma 2.3 (ii)), it follows that

$$L_S(x) < U_S(x) \quad \text{for all } x \in (a_1, b_K). \tag{2.18}$$

Part (iii) and (iv) of Lemma 2.3 imply that any level $\alpha$ with $0 < \alpha < 1$ can be chosen such that the functions $L_S(x)$ and $U_S(x)$ are well-defined.

**Theorem 2.4.** *Let $L_S(x)$ and $U_S(x)$ be defined as in (2.16) and (2.17), respectively. Then the confidence interval $[L_S(X), U_S(X)]$ satisfies*

$$\mathbb{P}_{\theta,\sigma^2}\left(L_S(X) \le \theta \le U_S(X) | X \in S\right) = 1 - \alpha \quad \textit{for all } \theta \in \mathbb{R},\ \sigma^2 \in (0, +\infty). \tag{2.19}$$

*Proof.* We already showed in (2.14) that $C_S(X)$ has conditional coverage probability $(1 - \alpha)$ for all $\theta \in \mathbb{R}$ and $\sigma^2 \in (0, +\infty)$. By construction, we have $C_S(x) =$

$[L_S(x), U_S(x)]$ for all $x \in (a_1, b_K)$ (and especially for all $x \in S$). Therefore, also $[L_S(X), U_S(X)]$ has conditional coverage probability $(1 - \alpha)$ for all $\theta \in \mathbb{R}$ and $\sigma^2 \in (0, +\infty)$. $\square$

Thus it follows that the confidence interval $[L_S(X), U_S(X)]$ has exact conditional coverage probability of $1 - \alpha$ on the set $S$. In particular, the probability of not covering $\theta$ is equally allocated to both ends, that is,

$$\mathbb{P}_{\theta,\sigma^2}\left(L_S(X) > \theta \mid X \in S\right) = \mathbb{P}_{\theta,\sigma^2}\left(U_S(X) < \theta \mid X \in S\right) = \frac{\alpha}{2} \quad \text{for all } \theta \in \mathbb{R},\ \sigma^2 \in (0, +\infty). \tag{2.20}$$

Therefore, we say that the lower and upper confidence bound $L_S(X)$ and $U_S(X)$, respectively, have conditional confidence level $1 - \frac{\alpha}{2}$ on the set $S$. Corollary 3.5.1 in Lehmann and Romano (2006) implies that $L_S(X)$ and $U_S(X)$ are the uniform most accurate confidence bounds at confidence level $1 - \frac{\alpha}{2}$ in the sense that for any confidence bounds $L'_S(X)$ and $U'_S(X)$ that have conditional confidence level $1 - \frac{\alpha}{2}$ it follows that

$$\mathbb{P}_{\theta,\sigma^2}\Big(L_S(X) \leq \theta' \mid X \in S\Big) \leq \mathbb{P}_{\theta,\sigma^2}\Big(L'_S(X) \leq \theta' \mid X \in S\Big) \quad \text{for all } \theta, \theta' \in \mathbb{R} \tag{2.21}$$

$$\mathbb{P}_{\theta,\sigma^2}\Big(U_S(X) \geq \theta' \mid X \in S\Big) \leq \mathbb{P}_{\theta,\sigma^2}\Big(U'_S(X) \geq \theta' \mid X \in S\Big) \quad \text{for all } \theta, \theta' \in \mathbb{R}. \tag{2.22}$$

This means that $L_S(X)$ and $U_S(X)$ minimize the probability of covering any "wrong" parameter $\theta'$. However, it does not follow that $[L_S(X), U_S(X)]$ is the optimal confidence interval in the sense that $[L_S(X), U_S(X)]$ has minimal (expected) length. The condition that $\alpha$ is equally allocated to the two tails is not necessary to obtain a confidence interval with exact $1 - \alpha$ conditional coverage probability. In the following, we will derive a whole class of confidence intervals with exact conditional coverage probability of $(1 - \alpha)$ which includes $[L_S(X), U_S(X)]$ as a special case. Fix a $t$ with $0 < t < 1$. The set $C_{S,t}(x)$ is defined as follows:

$$C_{S,t}(x) = \{\theta \in \mathbb{R} : (1 - t)\alpha \leq F^S_{\theta,\sigma^2}(x) \leq 1 - t\alpha\} \quad \text{for all } x \in \mathbb{R}. \tag{2.23}$$

Lemma 2.3 implies that $C_{S,t}(x)$ is an interval for every $x \in (a_1, b_K)$. The lower bound of $C_{S,t}(x)$ is denoted by $L_{S,t}(x)$ and the upper bound is denoted by $U_{S,t}(x)$.

Again by Lemma 2.3 (ii), $L_{S,t}(x)$ must satisfy

$$F^S_{L_{S,t}(x),\sigma^2}(x) = 1 - t\alpha \quad \text{for all } x \in (a_1, b_K). \tag{2.24}$$

and $U_{S,t}(x)$ must satisfy

$$F^S_{U_{S,t}(x),\sigma^2}(x) = (1 - t)\alpha \quad \text{for all } x \in (a_1, b_K). \tag{2.25}$$

For $x \notin (a_1, b_K)$, we set $L_{S,t}(x) = U_{S,t}(x) = 0$. Lemma 2.3 (ii) also implies that $L_{S,t}(x) < U_{S,t}(x)$ for all $x \in (a_1, b_K)$. If $t = \frac{1}{2}$, it follows that $C_{S,t}(x) = C_S(x)$, $L_{S,t}(x) = L_S(x)$ and $U_{S,t}(x) = U_S(x)$.

**Corollary 2.5.** *For all $t$ with $0 < t < 1$, let $L_{S,t}(x)$ and $U_{S,t}(x)$ be defined as in (2.24) and (2.25), respectively. Then the confidence interval $[L_{S,t}(X), U_{S,t}(X)]$ satisfies*

$$\mathbb{P}_{\theta,\sigma^2}\Big(L_{S,t}(X) \leq \theta \leq U_{S,t}(X) |\ X \in S\Big) = 1 - \alpha \quad \textit{for all } \theta \in \mathbb{R},\ \sigma^2 \in (0, +\infty). \tag{2.26}$$

*Proof.* By construction, the set $C_{S,t}$ must satisfy

$$\mathbb{P}_{\theta,\sigma^2}\Big(\theta \in C_{S,t}(X) |\ X \in S\Big) = 1 - \alpha \quad \text{for all } \theta \in \mathbb{R},\ \sigma^2 \in (0, +\infty). \tag{2.27}$$

For all $x \in (a_1, b_K)$, we have that $C_{S,t}(x) = [L_{S,t}(x), U_{S,t}(x)]$. Hence, the result follows. $\square$

This means that $[L_{S,t}(X), U_{S,t}(X)]$ also has exact conditional coverage probability of $1 - \alpha$ on the set $S$ for all $t$ with $0 < t < 1$. Corollary 3.5.1 in Lehmann and Romano (2006) implies that $L_{S,t}(X)$ and $U_{S,t}(X)$ are the uniform most accurate confidence bounds at confidence level $1 - t\alpha$ and confidence level $1 - (1 - t)\alpha$, respectively. However, among the confidence intervals $\{[L_{S,t}(X), U_{S,t}(X)]\}_{0<t<1}$, preliminary investigations suggest that no confidence interval is uniformly narrower than the others, that is, there is no $t^*$ such that $U_{S,t^*}(x) - L_{S,t^*}(x) \leq U_{S,t}(x) - L_{S,t}(x)$ $(t \neq t^*)$ for all $x \in \mathbb{R}$.

Consider the standard boundary functions $L(x)$ and $U(x)$ as defined in (2.7). In the following, the length of the interval $[L_S(x), U_S(x)]$ will be compared with the standard interval $[L(x), U(x)]$. We showed that if the truncation set $S$ is bounded, then the boundedness of $S$ implies that, conditional on $S$, the confidence interval $[L(X), U(X)]$ is bounded with probability 1. Therefore, conditional on $S$, the coverage probability of the confidence interval $[L(X), U(X)]$ is always 0 for large values of $\theta$. Thus we are expecting that the confidence interval $[L_S(X), U_S(X)]$ will be significantly longer to compensate in a certain sense for the boundedness of the set $S$. Figure 2.3 compares $L_S(x)$ and $U_S(x)$ with $L(x)$ and $U(x)$ when the truncation set $S$ is bounded. The length of the interval $[L(x), U(x)]$ is constantly equal to $2\sigma\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$. However, it appears that both $L_S(x)$ and $U_S(x)$ diverge to $+\infty$ as $x \to b_K$ and both diverge to $-\infty$ as $x \to a_1$. In fact, the Figure 2.3 suggests that also the length $U_S(x) - L_S(x)$ diverges as $x$ converges to the out-most boundaries, that is, to $a_1$ or $b_K$. This observation is not new; in the pre-testing literature this was already reported by Meeks and D'Agostino (1983) for a single truncation interval that is bounded on one side. Interestingly, when $x$ is close to the inner boundary points, the interval $[L_S(x), U_S(x)]$ is even narrower than $[L(x), U(x)]$. This means that $[L_S(x), U_S(x)]$ is not uniformly wider than $[L(x), U(x)]$, in general.

However, if the truncation set $S$ is unbounded, then, conditional on $S$, $L(X)$ and $U(X)$ are clearly not bounded in probability. Thus, conditional on $S$, $[L(X), U(X)]$ also covers large values of $\theta$ with positive probability. Therefore, one would expect that the confidence interval $[L_S(X), U_S(X)]$ does not need to become extremely wide to cover $\theta$ with exact conditional probability of $1 - \alpha$. Figure 2.4 compares $L_S(x)$ and $U_S(x)$ with $L(x)$ and $U(x)$ when the truncation set $S$ is unbounded. In contrast to the bounded case, it appears that the lower bound $L_S(x)$ and the upper bound $U_S(x)$ converge to the standard bounds $L(x)$ and $U(x)$ as $x \to -\infty$ and $x \to +\infty$. Therefore, it seems that also the length of the interval $[L_S(x), U_S(x)]$ is bounded and converges to $2\sigma\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ as $x \to -\infty$ and $x \to +\infty$.

These observations are not restricted to $[L_S(X), U_S(X)]$ but can also be observed for any confidence interval $[L_{S,t}(X), U_{S,t}(X)]$ with $t \in (0, 1)$. In the following, only the confidence interval $[L_S(X), U_S(X)]$ (that is, the case $t = \frac{1}{2}$) will be considered but the results also hold for any $[L_{S,t}(X), U_{S,t}(X)]$ with $t \in (0, 1)$.

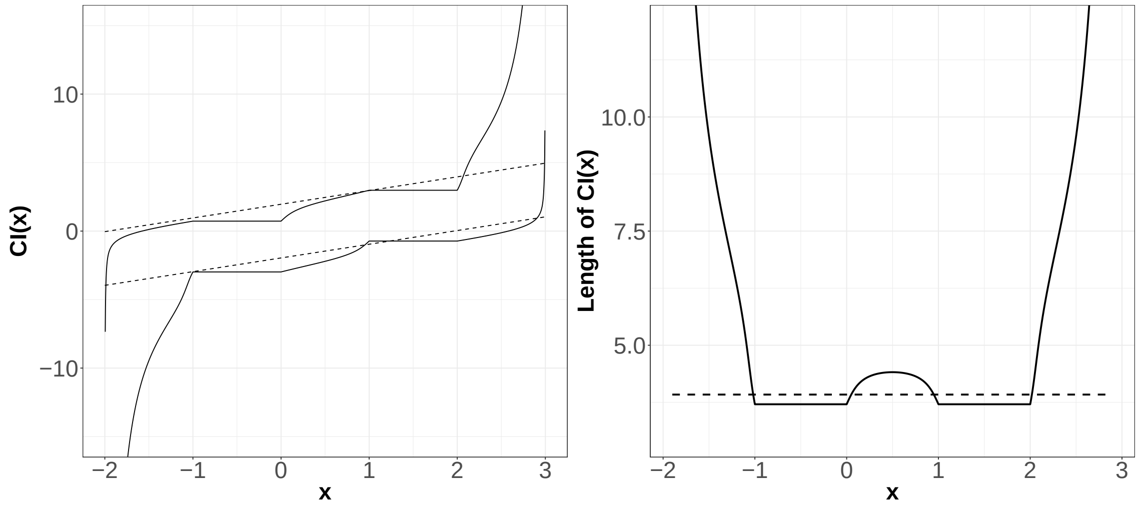


Figure 2.3: In the left panel, $L_S(x)$ and $U_S(x)$ as defined in (2.16) and (2.17) (solid) are compared with the boundary functions of the standard confidence interval $L(x)$ and $U(x)$ as defined in (2.7) (dashed) when the truncation set $S$ is bounded. In the right panel, the corresponding length of the intervals is shown. The parameters were set as follows: $\sigma^2 = 1$, $S = (-2, -1) \cup (0, 1) \cup (2, 3)$ and $\alpha = 0.05$. Observe that the length of the interval $[L_S(x), U_S(x)]$ diverges as $x$ converges to the out-most boundaries.

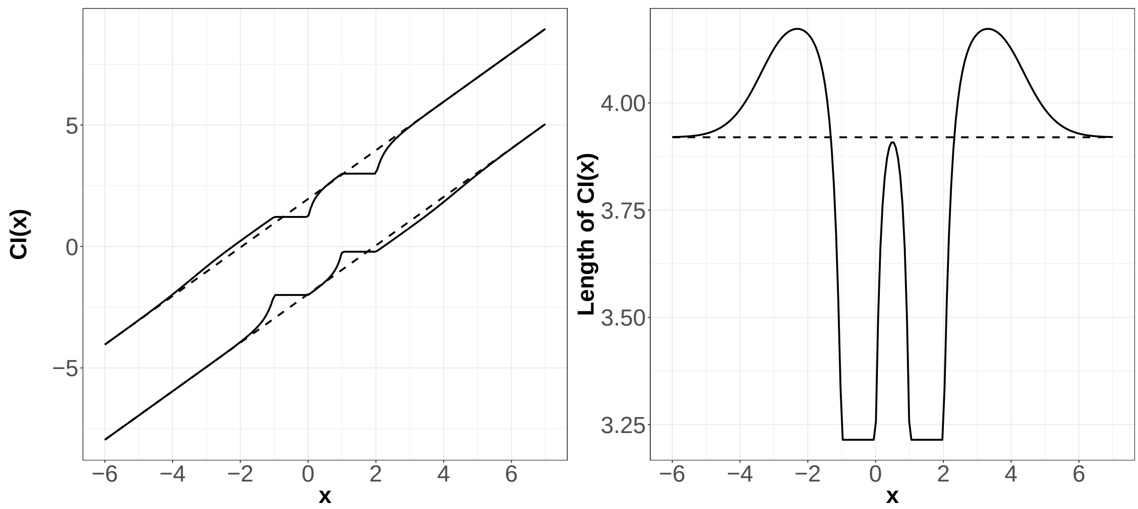


Figure 2.4: In the left panel, $L_S(x)$ and $U_S(x)$ as defined in (2.16) and (2.17) (solid) are compared with the boundary functions of the standard confidence interval $L(x)$ and $U(x)$ as defined in (2.7) (dashed) when the truncation set $S$ is unbounded. In the right panel, the corresponding length of the intervals is shown. The parameters were set as follows: $\sigma^2 = 1$, $S = (-\infty, -1) \cup (0, 1) \cup (2, +\infty)$ and $\alpha = 0.05$. Observe that $L_S(x)$ and $U_S(x)$ converge to the standard bounds $L(x)$ and $U(x)$ as $x \to +\infty$ and $x \to -\infty$.

## 2.3 Conditional expected length when the truncation set is bounded

In Figure 2.3, where the truncation set $S$ is bounded, it appears that the length of the interval $[L_S(x), U_S(x)]$ diverges as $x \to b_K$ or $x \to a_1$. An interesting question that arises is: what it is the asymptotic behavior of $U_S(x) - L_S(x)$ as $x \to b_K$ or $x \to a_1$? The answer to this question is given in the following lemma.

**Lemma 2.6.** *Let $q$ with $0 < q < 1$ be a constant. Let the function $B_S(x)$ be defined by the equation*

$$F^S_{B_S(x),\sigma^2}(x) = q \quad \textit{for all } x \in (a_1, b_K). \tag{2.28}$$

*For all $x \leq a_1$ and all $x \geq b_K$, set $B_S(x) = 0$. Then it follows that*

*(i) if $b_K < +\infty$, then $\lim_{x \nearrow b_K} (b_K - x)B_S(x) = -\sigma^2 \log(q)$;*

*(ii) if $a_1 > -\infty$, then $\lim_{x \searrow a_1} (a_1 - x)B_S(x) = -\sigma^2 \log(1 - q)$.*

The proof of the lemma is postponed to Section 2.6. Observe that the asymptotic behavior of the function $B_S(x)$ as $x \nearrow b_K$ only depends on $b_K$, while it is completely independent of the remaining boundary points $a_1, b_1, \ldots, b_{K-1}, a_K$. Analogously, the asymptotic behavior of the function $B_S(x)$ only depends on $a_1$ as $x \searrow a_1$. Note that if $q = 1 - \frac{\alpha}{2}$, it follows that $B_S(x) = L_S(x)$ for all $x \in (a_1, b_K)$. Therefore, the previous lemma implies that

$$\lim_{x \nearrow b_K} (b_K - x)L_S(x) = -\sigma^2 \log\left(1 - \frac{\alpha}{2}\right) \quad \text{if } b_K < +\infty \tag{2.29}$$

$$\lim_{x \searrow a_1} (a_1 - x)L_S(x) = -\sigma^2 \log\left(\frac{\alpha}{2}\right) \quad \text{if } a_1 > -\infty. \tag{2.30}$$

On the other hand, if $q = \frac{\alpha}{2}$, it follows that $B_S(x) = U_S(x)$ for all $x \in (a_1, b_K)$. Therefore, the previous lemma implies that

$$\lim_{x \nearrow b_K} (b_K - x)U_S(x) = -\sigma^2 \log\left(\frac{\alpha}{2}\right) \quad \text{if } b_K < +\infty \tag{2.31}$$

$$\lim_{x \searrow a_1} (a_1 - x)U_S(x) = -\sigma^2 \log\left(1 - \frac{\alpha}{2}\right) \quad \text{if } a_1 > -\infty. \tag{2.32}$$

Set $C(\alpha) = -\log\left(\frac{\frac{\alpha}{2}}{1-\frac{\alpha}{2}}\right)$. Observe that $C(\alpha)$ is strictly positive for all $\alpha \in (0, 1)$. Then the following asymptotic bound is obtained for the length of the interval

$[L_S(x), U_S(x)]$:

$$\lim_{x \nearrow b_K} (b_K - x)(U_S(x) - L_S(x)) = \sigma^2 C(\alpha) \quad \text{if } b_K < +\infty \tag{2.33}$$

$$\lim_{x \searrow a_1} (a_1 - x)(U_S(x) - L_S(x)) = -\sigma^2 C(\alpha) \quad \text{if } a_1 > -\infty. \tag{2.34}$$

This means that if the truncation set is bounded from above, then $U_S(x) - L_S(x)$ is asymptotically equivalent to the function $\frac{1}{b_K - x}$ as $x \nearrow b_K$. Analogously, if $S$ is bounded from below, then $U_S(x) - L_S(x)$ is asymptotically equivalent to the function $\frac{-1}{a_1 - x}$ as $x \searrow a_1$. In the case where $S$ is a single interval that is bounded on one side, this property of $U_S(x) - L_S(x)$ was already noted by Ohman Strickland and Casella (2003) in the context of group sequential testing.

In the unconditional case, we know that the length of the standard confidence interval $[L(X), U(X)]$, where $L(x)$ and $U(x)$ are defined as in (2.7), is constant. One might now be interested in the conditional expected length of the confidence interval $[L_S(X), U_S(X)]$ to see how much longer the confidence $[L_S(X), U_S(X)]$ is on average, compared to the standard interval $[L(X), U(X)]$.

**Theorem 2.7.** *Let the function $L_S(x)$ and the function $U_S(x)$ be defined as in (2.16) and (2.17), respectively. Suppose that at least one of the inequalities $a_1 > -\infty$ and $b_K < +\infty$ holds. Then it follows that, conditional on $S$, the expected length of the confidence interval $[L_S(X), U_S(X)]$ is $+\infty$, that is,*

$$\mathbb{E}_{\theta,\sigma^2}(U_S(X) - L_S(X) | X \in S) = +\infty \quad \textit{for all } \theta \in \mathbb{R},\ \sigma^2 \in (0, +\infty). \tag{2.35}$$

*Proof.* Without loss of generality, we assume that $b_K < +\infty$. Fix an $\epsilon > 0$ such that $\sigma^2 C(\alpha) - \epsilon > 0$. Such an $\epsilon$ always exists since $C(\alpha)$ is strictly positive. By equation (2.33), there exists a $x_\epsilon$ with $a_K < x_\epsilon < b_K$ such that for all $x \in (x_\epsilon, b_K)$ the following inequality holds:

$$(b_K - x)(U_S(x) - L_S(x)) > \sigma^2 C(\alpha) - \epsilon \quad \text{for all } x \in (x_\epsilon, b_K). \tag{2.36}$$

$X|\{X \in S\}$ is truncated normally distributed with p.d.f. $f^S_{\theta,\sigma^2}(x)$. Note that on the interval $(a_K, b_K)$, the p.d.f. $f^S_{\theta,\sigma^2}(x)$ is strictly positive. We set $D = \inf_{x \in (x_\epsilon, b_K)} f^S_{\theta,\sigma^2}(x)$. As both $x_\epsilon$ and $b_k$ are finite, it follows that $D > 0$. By inequality (2.18), the interval length $U_S(x) - L_S(x)$ is strictly positive for all $x \in (a_1, b_K)$. Therefore, the conditional expected length is bounded from below as follows:

$$\begin{aligned}
\mathbb{E}_{\theta,\sigma^2}(U_S(X) - L_S(X)|X \in S) &= \int_{-\infty}^{+\infty} (U_S(x) - L_S(x)) f^S_{\theta,\sigma^2}(x) dx \\
&\geq \int_{x_\epsilon}^{b_K} (U_S(x) - L_S(x)) f^S_{\theta,\sigma^2}(x) dx \\
&\geq D \int_{x_\epsilon}^{b_K} (U_S(x) - L_S(x)) dx \\
&\geq D(\sigma^2 C(\alpha) - \epsilon) \int_{x_\epsilon}^{b_K} \frac{1}{b_K - x} dx \\
&= +\infty,
\end{aligned} \tag{2.37}$$

where we used the strict positivity of $U_S(X) - L_S(X)$ in the first inequality, we replaced $f^S_{\theta,\sigma^2}(x)$ by its infimum on $(x_\epsilon, b_K)$ in the second inequality and we used inequality (2.36) in the third inequality. $\square$

The crucial argument in the proof is that the density $f^S_{\theta,\sigma^2}(x)$ is strictly positive for all $x \in (a_K, b_K)$ such that the function $\frac{1}{b_K - x}$ is the dominant term in the integral. This means that even though $U_S(x) - L_S(x)$ is finite for every $x \in S$, the interval length tends to be extremely wide.

## 2.4 Conditional expected length when the truncation set is unbounded

The length of the interval $[L_S(x), U_S(x)]$ diverges when $b_K < +\infty$ and $x \nearrow b_K$ (and analogously when $a_1 > -\infty$). Even worse, by Theorem 2.7 it follows that, conditional on the event $S$, the expected length of $[L_S(X), U_S(X)]$ is $+\infty$. However, Figure 2.4 suggests that if $S$ is unbounded, then $L_S(x)$ and $U_S(x)$ converge to the standard boundary functions $L(x)$ and $U(x)$ as $x \to +\infty$ or $x \to -\infty$. In the following, a formal and precise statement of this observation is given.

**Lemma 2.8.** *Let $q$ with $0 < q < 1$ be a constant. Let the function $B_S(x)$ be defined as in (2.28). Set $B(x) = x - \sigma\Phi^{-1}(q)$. Then it follows that* *(i) If* $b_K = +\infty$*, then* $\lim_{x \to +\infty} B(x) - B_S(x) = 0$*;*

*(ii) If* $a_1 = -\infty$*, then* $\lim_{x \to -\infty} B(x) - B_S(x) = 0$.

The proof of this lemma is postponed to Section 2.7. Analogous to the bounded case, the asymptotic behavior of the function $B_S(x)$ depends solely on the out-most boundary points $a_1$ and $b_K$. Consider the boundary functions $L(x)$ and $U(x)$ of the standard confidence interval as defined in (2.7). If $q = 1 - \frac{\alpha}{2}$, it follows that $B_S(x) = L_S(x)$ and $B(x) = L(x)$. On the other hand, if $q = \frac{\alpha}{2}$, it follows that $B_S(x) = U_S(x)$ and $B(x) = U(x)$. The functions $L(x)$ and $U(x)$ are linear in $x$. Thus it follows that also the boundary functions $L_S(x)$ and $U_S(x)$ are asymptotically linear in $x$. The length of the standard interval, that is, $U(x) - L(x)$, is constantly equal to $2\sigma\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ for all $x \in \mathbb{R}$. Therefore, the interval length $U_S(x) - L_S(x)$ must also converge to the same constant as $x \to +\infty$ (if $b_K = +\infty$) or as $x \to -\infty$ (if $a_1 = -\infty$):

$$\lim_{x \to +\infty} U_S(x) - L_S(x) = \lim_{x \to +\infty} U(x) - L(x) = 2\sigma\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \quad \text{if } b_K = +\infty \tag{2.38}$$

$$\lim_{x \to -\infty} U_S(x) - L_S(x) = \lim_{x \to +\infty} U(x) - L(x) = 2\sigma\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \quad \text{if } a_1 = -\infty. \tag{2.39}$$

Suppose that both $a_1 = -\infty$ and $b_K = +\infty$ hold. By definition of the functions $L_S(x)$ and $U_S(x)$, it follows that $U_S(x) - L_S(x) > 0$ holds for all x in $(a_1, b_K) = (-\infty, +\infty)$. From the previous two equations, it follows that for every $\epsilon > 0$ there exists a $x_\epsilon > 0$ such that

$$U_S(x) - L_S(x) < 2\sigma\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \epsilon \quad \text{for all } |x| > x_\epsilon. \tag{2.40}$$

Lemma 2.15 in Section 2.6 implies that both $L_S(x)$ and $U_S(x)$ are continuous on $(-\infty, +\infty)$; as a consequence, $U_S(x) - L_S(x)$ is also continuous on $(-\infty, +\infty)$. Thus it follows that on the compact interval $[-x_\epsilon, x_\epsilon]$, the interval length $U_S(x) - L_S(x)$ must be bounded. Thus we showed that $U_S(x) - L_S(x)$ is bounded for all $x \in \mathbb{R}$. Hence it follows that the expected length of the confidence interval $[L_S(X), U_S(X)]$, conditional on $S$, must be finite. The following theorem gives an upper bound for the conditional expected length:

**Theorem 2.9.** *Let $L_S(x)$ and $U_S(x)$ be defined as in (2.16) and (2.17), respectively. Suppose that $K > 1$, $a_1 = -\infty$ and $b_K = +\infty$. Let $C_{\theta,\sigma^2}(b_1, a_K)$ be defined as follows:*

$$C_{\theta,\sigma^2}(b_1, a_K) = 2(a_K - b_1) + \sigma \frac{\phi\left(\frac{b_1-\theta}{\sigma}\right)}{\Phi\left(\frac{b_1-\theta}{\sigma}\right)} + \sigma \frac{\phi\left(\frac{a_K-\theta}{\sigma}\right)}{1 - \Phi\left(\frac{a_K-\theta}{\sigma}\right)} + 2\sigma\Phi^{-1}\left(1 - \frac{\alpha}{2}\right). \quad (2.41)$$

*Then, conditional on $S$, the expected length of the confidence interval $[L_S(X), U_S(X)]$ is bounded by $C_{\theta,\sigma^2}(b_1, a_K)$, that is,*

$$\mathbb{E}_{\theta,\sigma^2}(U_S(X) - L_S(X) \; |X \in S) < C_{\theta,\sigma^2}(b_1, a_K), \quad (2.42)$$

*Further, $C_{\theta,\sigma^2}(b_1, a_K)$ is strictly decreasing as a function of $b_1$ and strictly increasing as a function of $a_K$.*

Note that $K = 1$ was excluded. In this case it follows that $S = \mathbb{R}$, which means that $F^S_{\theta,\sigma^2}(x) = \Phi_{\theta,\sigma^2}(x)$ and therefore $L_S(x) = L(x)$ and $U_S(x) = U(x)$. Thus it follows that $\mathbb{E}_{\theta,\sigma^2}(U_S(X) - L_S(X)|X \in S) = \mathbb{E}_{\theta,\sigma^2}(U(X) - L(X)) = 2\sigma\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$. The proof of the theorem is given in Section 2.8. The constant $C_{\theta,\sigma^2}(b_1, a_K)$ is definitely not the tightest upper bound that is achievable, but it is sufficient for the purposes of Section 3.4.

To compare the conditional expected length of $[L_S(X), U_S(X)]$ with the upper bound $C_{\theta,\sigma^2}(b_1, a_K)$, we estimated the conditional expected length by Monte Carlo simulations. We considered several values of $\theta$ in the interval $[-6, 7]$. We set $\sigma^2 = 1$. For each $\theta$, 20000 independent observations were sampled from the $TN(\theta, 1, S)$-distribution, where $S = (-\infty, -1) \cup (0, 1) \cup (2, +\infty)$. For each of the 20000 observations, we calculated the length of the confidence interval. The average of the 20000 interval lengths is the estimate of the conditional expected length. In the left panel of Figure 2.5, the Monte Carlo estimates of the conditional expected length of $[L_S(X), U_S(X)]$ (solid) are compared to the upper bound $C_{\theta,\sigma^2}(b_1, a_K)$ (dashed). In the right panel of Figure 2.5, the conditional expected length is compared to the length of the standard interval (dashed). We see that the upper bound $C_{\theta,\sigma^2}(b_1, a_K)$ is very loose. It appears that $C_{\theta,\sigma^2}(b_1, a_K)$ diverges while the conditional expected length seems to converge to the length of the standard interval as $\theta \to +\infty$ or $\theta \to -\infty$. The right panel also suggests that the conditional expected length can even be less than the length of the standard interval for certain values of $\theta$.

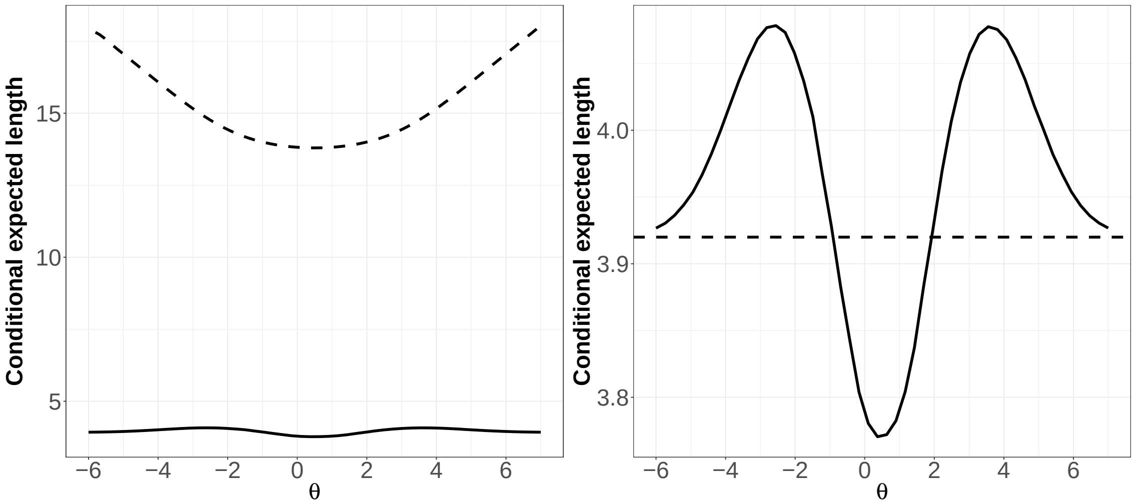


Figure 2.5: In the left panel, the Monte Carlo estimate of $\mathbb{E}_{\theta,\sigma^2}(U_S(X)-L_S(X)\,|X \in S)$ (solid) is compared to the upper bound $C_{\theta,\sigma^2}(b_1, a_K)$ in (2.41) (dashed). In the right panel the conditional expected length (solid) is compared to the length of the standard interval (dashed). To estimate the conditional expected length, 20000 independent observations were sampled from the $TN(\theta, 1, S)$-distribution, where $S = (-\infty, -1) \cup (0, 1) \cup (2, +\infty)$.

## 2.5 Proof of Lemma 2.3

Before Lemma 2.3 is proved, a few auxiliary results are required. Throughout the section, we suppose that $\sigma^2 \in (0, +\infty)$. All results hold for any $\sigma^2 \in (0, +\infty)$.

**Lemma 2.10.** *For all $a, b$ with $-\infty \leq a < b \leq +\infty$, the following holds:*

$$\lim_{\theta \to +\infty} \frac{\Phi\left(\frac{a-\theta}{\sigma}\right)}{\Phi\left(\frac{b-\theta}{\sigma}\right)} = 0. \tag{2.43}$$

*Proof.* If $a = -\infty$, then it follows that $\Phi\left(\frac{a-\theta}{\sigma}\right) = 0$ for all $\theta \in \mathbb{R}$, while $\Phi\left(\frac{b-\theta}{\sigma}\right)$ is strictly positive for all $\theta \in \mathbb{R}$. Therefore, $\lim\limits_{\theta \to +\infty} \frac{\Phi\left(\frac{a-\theta}{\sigma}\right)}{\Phi\left(\frac{b-\theta}{\sigma}\right)} = \lim\limits_{\theta \to +\infty} 0 = 0$. If $b = +\infty$, then it follows that $\Phi\left(\frac{b-\theta}{\sigma}\right) = 1$ for all $\theta \in \mathbb{R}$, while $\Phi\left(\frac{a-\theta}{\sigma}\right)$ converges to 0 as $\theta \to +\infty$. Therefore, $\lim\limits_{\theta \to +\infty} \frac{\Phi\left(\frac{a-\theta}{\sigma}\right)}{\Phi\left(\frac{b-\theta}{\sigma}\right)} = \lim\limits_{\theta \to +\infty} \Phi\left(\frac{a-\theta}{\sigma}\right) = 0$.

Now suppose that $-\infty < a < b < +\infty$. Note that $a - b < 0$. Both $\Phi\left(\frac{a-\theta}{\sigma}\right)$ and

$\Phi\left(\frac{b-\theta}{\sigma}\right)$ converge to zero as $\theta \to +\infty$. Therefore, L'Hospital's rule is used to obtain:

$$\begin{aligned}
\lim_{\theta\to+\infty} \frac{\Phi\left(\frac{a-\theta}{\sigma}\right)}{\Phi\left(\frac{b-\theta}{\sigma}\right)} &= \lim_{\theta\to+\infty} \frac{-\frac{1}{\sigma}\phi\left(\frac{a-\theta}{\sigma}\right)}{-\frac{1}{\sigma}\phi\left(\frac{b-\theta}{\sigma}\right)} \\
&= \lim_{\theta\to+\infty} e^{-\frac{1}{2\sigma^2}\left(a^2-2a\theta+\theta^2-b^2+2b\theta-\theta^2\right)} \\
&= \lim_{\theta\to+\infty} e^{\frac{1}{\sigma^2}(a-b)\theta} e^{-\frac{1}{2\sigma^2}(a^2-b^2)} = 0.
\end{aligned} \tag{2.44}$$

□

**Corollary 2.11.** *Let $-\infty \leq a_1 < b_1 < \cdots < a_K < b_K \leq +\infty$. Then it follows that*

$$\lim_{\theta\to+\infty} \frac{\sum_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\Phi\left(\frac{b_K-\theta}{\sigma}\right)} = 1. \tag{2.45}$$

*Proof.* Rearranging the terms yields

$$\begin{aligned}
\frac{\sum_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\Phi\left(\frac{b_K-\theta}{\sigma}\right)} &= \frac{\Phi\left(\frac{b_K-\theta}{\sigma}\right) - \Phi\left(\frac{a_K-\theta}{\sigma}\right) + \sum_{i=1}^{K-1} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\Phi\left(\frac{b_K-\theta}{\sigma}\right)} \\
&= 1 + \frac{-\Phi\left(\frac{a_K-\theta}{\sigma}\right) + \sum_{i=1}^{K-1} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\Phi\left(\frac{b_K-\theta}{\sigma}\right)} \\
&= 1 - \frac{\Phi\left(\frac{a_K-\theta}{\sigma}\right)}{\Phi\left(\frac{b_K-\theta}{\sigma}\right)} + \sum_{i=1}^{K-1} \frac{\Phi\left(\frac{b_i-\theta}{\sigma}\right)}{\Phi\left(\frac{b_K-\theta}{\sigma}\right)} - \frac{\Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\Phi\left(\frac{b_K-\theta}{\sigma}\right)}.
\end{aligned} \tag{2.46}$$

By assumption, $b_K$ is strictly larger than $a_1, b_1, \ldots, b_{K-1}, a_K$. Therefore, Lemma 2.10 implies that

$$\frac{\Phi\left(\frac{b_i-\theta}{\sigma}\right)}{\Phi\left(\frac{b_K-\theta}{\sigma}\right)} \to 0 \quad \text{as} \quad \theta \to +\infty \quad \text{for all } i = 1, \ldots, K-1, \tag{2.47}$$

and that

$$\frac{\Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\Phi\left(\frac{b_K-\theta}{\sigma}\right)} \to 0 \quad \text{as} \quad \theta \to +\infty \quad \text{for all } i = 1, \ldots, K. \tag{2.48}$$

As there are only finitely many terms, it follows that

$$\lim_{\theta\to+\infty} \frac{\sum_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\Phi\left(\frac{b_K-\theta}{\sigma}\right)} = 1. \tag{2.49}$$

□

Recall that the truncation set $S$ was defined as follows:

$$S = \bigcup_{i=1}^{K} (a_i, b_i), \tag{2.50}$$

where $K < +\infty$ and $-\infty \leq a_1 < b_1 < \cdots < a_K < b_K \leq +\infty$. The reflection of the set $S$, denoted by $-S$, is defined as follows:

$$-S = \{x \in \mathbb{R} : -x \in S\} = \bigcup_{i=1}^{K} (-b_i, -a_i). \tag{2.51}$$

Observe that $-S$ is again a union of disjoint intervals.

**Lemma 2.12.** *Let $S$ and $-S$ be defined as in (2.50) and (2.51), respectively. Then it follows that*

$$F^S_{\theta,\sigma^2}(x) = 1 - F^{-S}_{-\theta,\sigma^2}(-x) \quad \textit{for all } x \in \mathbb{R}. \tag{2.52}$$

*Proof.* Note that the endpoints of $-S$ are ordered as follows:

$$-\infty \leq -b_K < -a_K < \cdots < -b_1 < -a_1 \leq +\infty. \tag{2.53}$$

Thus the function $F^{-S}_{\theta,\sigma^2}(x)$ is defined as follows:

$$F^{-S}_{\theta,\sigma^2}(x) = \begin{cases} 0 & \text{if } x \leq -b_K \\ \dfrac{\Phi\left(\frac{x-\theta}{\sigma}\right) - \Phi\left(\frac{-b_k-\theta}{\sigma}\right) + \sum\limits_{i=k+1}^{K} \Phi\left(\frac{-a_i-\theta}{\sigma}\right) - \Phi\left(\frac{-b_i-\theta}{\sigma}\right)}{\sum\limits_{i=1}^{K} \Phi\left(\frac{-a_i-\theta}{\sigma}\right) - \Phi\left(\frac{-b_i-\theta}{\sigma}\right)} & \text{if } -b_k < x < -a_k \text{ for } k = 1, \ldots, K \\ \dfrac{\sum\limits_{i=k}^{K} \Phi\left(\frac{-a_i-\theta}{\sigma}\right) - \Phi\left(\frac{-b_i-\theta}{\sigma}\right)}{\sum\limits_{i=1}^{K} \Phi\left(\frac{-a_i-\theta}{\sigma}\right) - \Phi\left(\frac{-b_i-\theta}{\sigma}\right)} & \text{if } -a_k \leq x \leq -b_{k-1} \text{ for } k = 2, \ldots, K \\ 1 & \text{if } x \geq -a_1. \end{cases} \tag{2.54}$$

If $k = K$, then the expression $\sum\limits_{i=k+1}^{K} \Phi\left(\frac{-a_i-\theta}{\sigma}\right) - \Phi\left(\frac{-b_i-\theta}{\sigma}\right)$ is to be interpreted as zero. Now consider 4 cases.

First case: $x \leq a_1$. For all $x \leq a_1$, the equation $F^S_{\theta,\sigma^2}(x) = 0$ holds. From the definition of $F^{-S}_{-\theta,\sigma^2}(-x)$ in (2.54), we see that also $1 - F^{-S}_{-\theta,\sigma^2}(-x) = 1 - 1 = 0$ is satisfied for all $x \leq a_1$.

Second case: $x \geq b_K$. For all $x \geq b_K$, the equation $F^S_{\theta,\sigma^2}(x) = 1$ holds. Again from the definition of $F^{-S}_{-\theta,\sigma^2}(-x)$ in (2.54), also $1 - F^{-S}_{-\theta,\sigma^2}(-x) = 1 - 0 = 1$ is satisfied for all $x \geq b_K$.

Third case: $x \in S$. For every $x \in S$, there exists a $k$ (depending on $x$) with $1 \leq k \leq K$ such that $a_k < x < b_k$. This means that $F^S_{\theta,\sigma^2}(x)$ satisfies

$$F^S_{\theta,\sigma^2}(x) = \frac{\Phi\left(\frac{x-\theta}{\sigma}\right) - \Phi\left(\frac{a_k-\theta}{\sigma}\right) + \sum_{i=1}^{k-1} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\sum_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}. \tag{2.55}$$

Note that the function $\Phi(x)$ satisfies the following symmetry property: $\Phi(x) = 1 - \Phi(-x)$ for all $x \in \mathbb{R}$. Applying this property to the right hand side of the equation yields

$$\begin{aligned}
F^S_{\theta,\sigma^2}(x) &= \frac{\left(1 - \Phi\left(\frac{-x+\theta}{\sigma}\right)\right) - \left(1 - \Phi\left(\frac{-a_k+\theta}{\sigma}\right)\right) + \sum_{i=1}^{k-1} \left(1 - \Phi\left(\frac{-b_i+\theta}{\sigma}\right)\right) - \left(1 - \Phi\left(\frac{-a_i+\theta}{\sigma}\right)\right)}{\sum_{i=1}^{K} \left(1 - \Phi\left(\frac{-b_i+\theta}{\sigma}\right)\right) - \left(1 - \Phi\left(\frac{-a_i+\theta}{\sigma}\right)\right)} \\
&= \frac{\Phi\left(\frac{-a_k+\theta}{\sigma}\right) - \Phi\left(\frac{-x+\theta}{\sigma}\right) + \sum_{i=1}^{k-1} \Phi\left(\frac{-a_i+\theta}{\sigma}\right) - \Phi\left(\frac{-b_i+\theta}{\sigma}\right)}{\sum_{i=1}^{K} \Phi\left(\frac{-a_i+\theta}{\sigma}\right) - \Phi\left(\frac{-b_i+\theta}{\sigma}\right)}.
\end{aligned} \tag{2.56}$$

We add and subtract 1 on the right hand side and rearrange the terms to obtain

$$\begin{aligned}
F^S_{\theta,\sigma^2}(x) &= 1 - 1 + \frac{\Phi\left(\frac{-a_k+\theta}{\sigma}\right) - \Phi\left(\frac{-x+\theta}{\sigma}\right) + \sum_{i=1}^{k-1} \Phi\left(\frac{-a_i+\theta}{\sigma}\right) - \Phi\left(\frac{-b_i+\theta}{\sigma}\right)}{\sum_{i=1}^{K} \Phi\left(\frac{-a_i+\theta}{\sigma}\right) - \Phi\left(\frac{-b_i+\theta}{\sigma}\right)} \\
&= 1 + \frac{-\left(\sum_{i=1}^{K} \Phi\left(\frac{-a_i+\theta}{\sigma}\right) - \Phi\left(\frac{-b_i+\theta}{\sigma}\right)\right) + \Phi\left(\frac{-a_k+\theta}{\sigma}\right) - \Phi\left(\frac{-x+\theta}{\sigma}\right) + \sum_{i=1}^{k-1} \Phi\left(\frac{-a_i+\theta}{\sigma}\right) - \Phi\left(\frac{-b_i+\theta}{\sigma}\right)}{\sum_{i=1}^{K} \Phi\left(\frac{-a_i+\theta}{\sigma}\right) - \Phi\left(\frac{-b_i+\theta}{\sigma}\right)} \\
&= 1 + \frac{\Phi\left(\frac{-b_k+\theta}{\sigma}\right) - \Phi\left(\frac{-x+\theta}{\sigma}\right) - \sum_{i=k+1}^{K} \Phi\left(\frac{-a_i+\theta}{\sigma}\right) - \Phi\left(\frac{-b_i+\theta}{\sigma}\right)}{\sum_{i=1}^{K} \Phi\left(\frac{-a_i+\theta}{\sigma}\right) - \Phi\left(\frac{-b_i+\theta}{\sigma}\right)} \\
&= 1 - \frac{\Phi\left(\frac{-x+\theta}{\sigma}\right) - \Phi\left(\frac{-b_k+\theta}{\sigma}\right) + \sum_{i=k+1}^{K} \Phi\left(\frac{-a_i+\theta}{\sigma}\right) - \Phi\left(\frac{-b_i+\theta}{\sigma}\right)}{\sum_{i=1}^{K} \Phi\left(\frac{-a_i+\theta}{\sigma}\right) - \Phi\left(\frac{-b_i+\theta}{\sigma}\right)} \\
&= 1 - F^{-S}_{-\theta,\sigma^2}(-x).
\end{aligned} \tag{2.57}$$

Fourth case: $x \in (a_1, b_K) \setminus S$. For every $x \in (a_1, b_K) \setminus S$, there exists a $k$ (depending on $x$) with $1 \leq k \leq K - 1$ such that $b_k \leq x \leq a_{k+1}$. Then $F^S_{\theta,\sigma^2}(x)$

satisfies

$$F^S_{\theta,\sigma^2}(x) = \frac{\sum\limits_{i=1}^{k} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\sum\limits_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}. \tag{2.58}$$

Rearranging the right hand side of the equation as in the previous case yields $F^S_{\theta,\sigma^2}(x) = 1 - F^{-S}_{-\theta,\sigma^2}(-x)$ for all $x \in (a_1, b_K) \setminus S$. □

*Proof of Lemma 2.3.* Proof of Part (i): There are 2 cases that need to be considered

First case: $x \in S$: For all $x \in S$, there exists a $k$ (depending on $x$) with $1 \leq k \leq K$ such that $a_k < x < b_k$. This means that $F^S_{\theta,\sigma^2}(x)$ satisfies

$$F^S_{\theta,\sigma^2}(x) = \frac{\Phi\left(\frac{x-\theta}{\sigma}\right) - \Phi\left(\frac{a_k-\theta}{\sigma}\right) + \sum\limits_{i=1}^{k-1} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\sum\limits_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}. \tag{2.59}$$

Clearly, both the numerator and the denominator on the right hand side of the previous equation are continuous functions in $\theta$ on $\mathbb{R}$. Also, the denominator is strictly positive for all $\theta \in \mathbb{R}$. Therefore, $F^S_{\theta,\sigma^2}(x)$ is continuous in $\theta$ on $\mathbb{R}$ for all $x \in S$.

Second case: $x \in (a_1, b_K) \setminus S$. Similar arguments as in the first case imply that $F^S_{\theta,\sigma^2}(x)$ is continuous in $\theta$ on $\mathbb{R}$ for all $x \in (a_1, b_K) \setminus S$.

Proof of Part (ii) (as in Lee et al. (2016)): Set $P(\theta) = \sum\limits_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)$ and note that $f^S_{\theta,\sigma^2}(x) = \frac{1}{P(\theta)} \frac{1}{\sigma} \phi\left(\frac{x-\theta}{\sigma}\right) \mathbb{1}_S(x)$. Observe that $f^S_{\theta,\sigma^2}(x)$ is strictly positive for all $x \in S$. Let $\theta_0, \theta_1 \in \mathbb{R}$ with $\theta_0 < \theta_1$. As $f^S_{\theta,\sigma^2}(x)$ is strictly positive for all $x \in S$, the likelihood ratio can be calculated for all $x \in S$:

$$\begin{aligned}
\frac{f^S_{\theta_1,\sigma^2}(x)}{f^S_{\theta_0,\sigma^2}(x)} &= \frac{\frac{1}{\sigma}\phi\left(\frac{x-\theta_1}{\sigma}\right)}{\frac{1}{\sigma}\phi\left(\frac{x-\theta_0}{\sigma}\right)} \frac{P(\theta_0)}{P(\theta_1)} \\
&= e^{-\frac{1}{2\sigma^2}\left(x^2 - 2x\theta_1 + \theta_1^2 - x^2 + 2x\theta_0 - \theta_0^2\right)} \frac{P(\theta_0)}{P(\theta_1)} \\
&= e^{\frac{1}{\sigma^2}(\theta_1-\theta_0)x} e^{-\frac{1}{2\sigma^2}(\theta_1^2-\theta_0^2)} \frac{P(\theta_0)}{P(\theta_1)}.
\end{aligned} \tag{2.60}$$

As $\theta_1$ is strictly larger than $\theta_0$, it follows that the likelihood ratio is strictly increasing in $x$ on the set $S$. Fix a $x \in (a_1, b_K)$. For every $x_0, x_1 \in S$ with $x_0 < x < x_1$ (such an $x_0$ and $x_1$ always exists since $S$ is an open set), the monotone likelihood ratio implies that

$$\frac{f^S_{\theta_1,\sigma^2}(x_0)}{f^S_{\theta_0,\sigma^2}(x_0)} < \frac{f^S_{\theta_1,\sigma^2}(x_1)}{f^S_{\theta_0,\sigma^2}(x_1)} \quad \text{for all } x_0, x_1 \in S \text{ with } x_0 < x < x_1 \tag{2.61}$$

and equivalently

$$f^S_{\theta_1,\sigma^2}(x_0) f^S_{\theta_0,\sigma^2}(x_1) < f^S_{\theta_1,\sigma^2}(x_1) f^S_{\theta_0,\sigma^2}(x_0) \quad \text{for all } x_0, x_1 \in S \text{ with } x_0 < x < x_1. \tag{2.62}$$

If either $x_0$ or $x_1$ is not in $S$, both sides in the line above are 0. Hence, the following inequality holds for all $x_0, x_1 \in \mathbb{R}$ with $x_0 < x < x_1$:

$$f^S_{\theta_1,\sigma^2}(x_0) f^S_{\theta_0,\sigma^2}(x_1) \leq f^S_{\theta_1,\sigma^2}(x_1) f^S_{\theta_0,\sigma^2}(x_0) \quad \text{for all } x_0, x_1 \in \mathbb{R} \text{ with } x_0 < x < x_1. \tag{2.63}$$

The intersection $(-\infty, x) \cap S$ has positive Lebesgue measure because $x$ is an interior point of $(a_1, b_K)$. Thus for all $x_0 \in (-\infty, x) \cap S$ and all $x_1 \in S$ with $x < x_1$, the previous inequality is strict. Therefore, integrating $x_0$ over $(-\infty, x)$ in the previous inequality yields

$$\int_{-\infty}^{x} f^S_{\theta_1,\sigma^2}(x_0) f^S_{\theta_0,\sigma^2}(x_1) dx_0 < \int_{-\infty}^{x} f^S_{\theta_1,\sigma^2}(x_1) f^S_{\theta_0,\sigma^2}(x_0) dx_0 \quad \text{for all } x_1 \in S \text{ with } x < x_1,$$

$$f^S_{\theta_0,\sigma^2}(x_1) F^S_{\theta_1,\sigma^2}(x) < f^S_{\theta_1,\sigma^2}(x_1) F^S_{\theta_0,\sigma^2}(x) \quad \text{for all } x_1 \in S \text{ with } x < x_1. \tag{2.64}$$

Integrating now $x_1$ over $(x, +\infty)$ (note again that the intersection $(x, +\infty) \cap S$ has positive Lebesgue measure) results in:

$$\left(1 - F^S_{\theta_0,\sigma^2}(x)\right) F^S_{\theta_1,\sigma^2}(x) < \left(1 - F^S_{\theta_1,\sigma^2}(x)\right) F^S_{\theta_0,\sigma^2}(x). \tag{2.65}$$

Therefore, $F^S_{\theta_1,\sigma^2}(x) < F^S_{\theta_0,\sigma^2}(x)$ holds for all $\theta_0 < \theta_1$ and all $x \in (a_1, b_K)$. This means that $F^S_{\theta,\sigma^2}(x)$ is strictly decreasing as a function of $\theta$ for all $x \in (a_1, b_K)$.

Proof of Part (iii): There are 2 cases that need to be considered.

First case: $x \in S$. For every $x \in S$, there exists a $k$ (depending on $x$) with $1 \leq k \leq K$ such that $a_k < x < b_k$. Therefore, $F^S_{\theta,\sigma^2}(x)$ satisfies

$$F^S_{\theta,\sigma^2}(x) = \frac{\Phi\left(\frac{x-\theta}{\sigma}\right) - \Phi\left(\frac{a_k-\theta}{\sigma}\right) + \sum_{i=1}^{k-1} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\sum_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}. \tag{2.66}$$

Note that $x < b_K$ holds for all $x \in S$ since the set $S$ is open. Corollary 2.11 implies that there exists a function $\epsilon(\theta)$ with $\epsilon(\theta) \to 0$ as $\theta \to +\infty$ such that

$$\sum_{i=1}^{K} \Phi\left(\frac{b_i - \theta}{\sigma}\right) - \Phi\left(\frac{a_i - \theta}{\sigma}\right) = (1 - \epsilon(\theta))\Phi\left(\frac{b_K - \theta}{\sigma}\right). \tag{2.67}$$

Therefore, $F^S_{\theta,\sigma^2}(x)$ can be rewritten as follows:

$$F^S_{\theta,\sigma^2}(x) = \frac{1}{1-\epsilon(\theta)} \frac{\Phi\left(\frac{x-\theta}{\sigma}\right) - \Phi\left(\frac{a_k-\theta}{\sigma}\right) + \sum_{i=1}^{k-1} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\Phi\left(\frac{b_K-\theta}{\sigma}\right)}, \tag{2.68}$$

Since $\epsilon(\theta) \to 0$ as $\theta \to +\infty$, it follows that the first factor on the right hand side converges to 1. As $b_K$ is strictly larger than $a_1, b_1, \ldots, a_k, x$, Lemma 2.10 implies that the second factor on the right hand side converges to 0 as $\theta \to +\infty$. Thus $\lim\limits_{\theta\to+\infty} F^S_{\theta,\sigma^2}(x) = 0$ follows for all $x \in S$.

Second case: $x \in (a_1, b_K) \setminus S$. For every $x \in (a_1, b_K) \setminus S$, there exists a $k$ (depending on $x$) with $1 \le k \le K-1$ such that $b_k \le x \le a_{k+1}$. Then $F^S_{\theta,\sigma^2}(x)$ satisfies

$$F^S_{\theta,\sigma^2}(x) = \frac{\sum_{i=1}^{k} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\sum_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}. \tag{2.69}$$

By applying the same argument as before, we obtain $\lim\limits_{\theta\to+\infty} F^S_{\theta,\sigma^2}(x) = 0$ for all $x \in (a_1, b_K) \setminus S$.

Proof of Part (iv): Let denote by $-S$ the reflection of the set $S$ as defined in (2.51). We first use Lemma 2.12 and afterwards Part (iii) of Lemma 2.3 to obtain

$$\lim_{\theta\to-\infty} F^S_{\theta,\sigma^2}(x) = \lim_{\theta\to-\infty} 1 - F^{-S}_{-\theta,\sigma^2}(-x) = 1 - \lim_{\theta\to+\infty} F^{-S}_{\theta,\sigma^2}(-x) = 1 \quad \text{for all } x \in (a_1, b_K). \tag{2.70}$$

$\square$

## 2.6 Proof of Lemma 2.6

Again, before Lemma 2.6 is proved, some auxiliary results are required. Throughout the section, we suppose that $\sigma^2 \in (0, +\infty)$. All results hold for any $\sigma^2 \in (0, +\infty)$.

**Lemma 2.13.** *For all $q$ with $0 < q < 1$ and all $b \in \mathbb{R}$, it follows that*

$$\lim_{\theta\to+\infty} \frac{q\Phi\left(\frac{b-\theta}{\sigma}\right)}{\Phi\left(\frac{b-\theta+\frac{\sigma^2\log(q)}{\theta}}{\sigma}\right)} = 1. \tag{2.71}$$

*Proof.* Both $q\Phi\left(\frac{b-\theta}{\sigma}\right)$ and $\Phi\left(\frac{b-\theta+\frac{\sigma^2\log(q)}{\theta}}{\sigma}\right)$ converge to 0 as $\theta \to +\infty$. Therefore, L'Hospital's rule is used to obtain:

$$\begin{aligned}
\lim_{\theta\to+\infty} \frac{q\Phi\left(\frac{b-\theta}{\sigma}\right)}{\Phi\left(\frac{b-\theta+\frac{\sigma^2\log(q)}{\theta}}{\sigma}\right)} &= \lim_{\theta\to+\infty} \frac{\frac{-q}{\sigma}}{\frac{-1-\frac{\sigma^2\log(q)}{\theta^2}}{\sigma}} \frac{\phi\left(\frac{b-\theta}{\sigma}\right)}{\phi\left(\frac{b-\theta+\frac{\sigma^2\log(q)}{\theta}}{\sigma}\right)} \\
&= \lim_{\theta\to+\infty} \frac{q}{1+\frac{\sigma^2\log(q)}{\theta^2}} e^{-\frac{1}{2\sigma^2}\left((b-\theta)^2-(b-\theta)^2-2(b-\theta)\frac{\sigma^2\log(q)}{\theta}-\left(\frac{\sigma^2\log(q)}{\theta}\right)^2\right)} \\
&= \lim_{\theta\to+\infty} \frac{q}{1+\frac{\sigma^2\log(q)}{\theta^2}} e^{-\frac{1}{2\sigma^2}\left(-2b\frac{\sigma^2\log(q)}{\theta}+2\sigma^2\log(q)-\left(\frac{\sigma^2\log(q)}{\theta}\right)^2\right)} \\
&= \lim_{\theta\to+\infty} \frac{q}{1+\frac{\sigma^2\log(q)}{\theta^2}} e^{-\log(q)} e^{-\frac{1}{2\sigma^2}\left(-2b\frac{\sigma^2\log(q)}{\theta}-\left(\frac{\sigma^2\log(q)}{\theta}\right)^2\right)} \\
&= q\quad \frac{1}{q}\quad 1 \\
&= 1.
\end{aligned} \tag{2.72}$$

$\square$

**Corollary 2.14.** *Let $q(\theta)$ be a function such that $\lim_{\theta\to+\infty} q(\theta) = q$. For every $\delta \in \mathbb{R}$ such that $q + \delta > 0$, it follows that*

$$\lim_{\theta\to+\infty} \frac{q(\theta)\Phi\left(\frac{b-\theta}{\sigma}\right)}{\Phi\left(\frac{b-\theta+\frac{\sigma^2\log(q+\delta)}{\theta}}{\sigma}\right)} = \frac{q}{q+\delta}. \tag{2.73}$$

*Proof.* The same arguments as in the proof of Lemma 2.13 are used to show that

$$\lim_{\theta\to+\infty} \frac{\Phi\left(\frac{b-\theta}{\sigma}\right)}{\Phi\left(\frac{b-\theta+\frac{\sigma^2\log(q+\delta)}{\theta}}{\sigma}\right)} = \frac{1}{q+\delta}. \tag{2.74}$$

As $\lim_{\theta\to+\infty} q(\theta) = q$ and as multiplication is a continuous mapping, the result follows.

$\square$

**Lemma 2.15.** *Let $q$ with $0 < q < 1$ be a constant. Let the function $B_S(x)$ be defined as in (2.28). Then $B_S(x)$ is continuous on the interval $(a_1, b_K)$, strictly increasing on $(a_k, b_k)$ for all $k = 1, \ldots, K$ and constant on $[b_k, a_{k+1}]$ for every $k = 1, \ldots, K-1$.*

*Proof.* By definition, $B_S(x)$ satisfies

$$F^S_{B_S(x),\sigma^2}(x) = q \quad \text{for all } x \in (a_1, b_K). \tag{2.75}$$

Recall some properties of the c.d.f. $F^S_{\theta,\sigma^2}(x)$: As $F^S_{\theta,\sigma^2}(x)$ has density $f^S_{\theta,\sigma^2}(x)$ with respect to Lebesgue measure, it follows that $F^S_{\theta,\sigma^2}(x)$ is continuous and monotonically increasing in $x$ on $\mathbb{R}$. For every $k = 1, \ldots, K$ and every $x \in (a_k, b_k)$, it holds that $f^S_{\theta,\sigma^2}(x) > 0$; therefore, $F^S_{\theta,\sigma^2}(x)$ is strictly increasing in $x$ on $(a_k, b_k)$ for every $k = 1, \ldots, K$. For every $k = 1, \ldots, K-1$ and every $x \in [b_k, a_{k+1}]$, it holds that $f^S_{\theta,\sigma^2}(x) = 0$; therefore $F^S_{\theta,\sigma^2}(x)$ is constant in $x$ on $[b_k, a_{k+1}]$ for every $k = 1, \ldots, K-1$.

We first show that $B_S(x)$ is monotonically increasing in $x$ on $(a_1, b_K)$. For any $x, y \in (a_1, b_K)$, where $x < y$, the inequality $F^S_{\theta,\sigma^2}(x) \leq F^S_{\theta,\sigma^2}(y)$ holds for all $\theta \in \mathbb{R}$. Suppose that $B_S(x) > B_S(y)$ holds. Then it follows that $F^S_{B_S(x),\sigma^2}(x) < F^S_{B_S(y),\sigma^2}(x) \leq F^S_{B_S(y),\sigma^2}(y)$, where, in the first inequality, we used the fact that $F^S_{\theta,\sigma^2}(x)$ is strictly decreasing in $\theta$ for all $x \in (a_1, b_K)$ (Lemma 2.3 (ii)) and, in the second inequality, we used the fact that $F^S_{\theta,\sigma^2}(x)$ is increasing in $x$. However, by equation (2.75), also $F^S_{B_S(x),\sigma^2}(x) = F^S_{B_S(x),\sigma^2}(x) = q$ must be satisfied for all $x, y \in (a_1, b_K)$. This clearly contradicts the previous inequality, therefore we can conclude that $B_S(x)$ is monotonically increasing in $x$ on $(a_1, b_K)$.

Now we show that $B_S(x)$ is strictly increasing in $x$ on $(a_k, b_k)$ for every $k = 1, \ldots, K$. Fix an arbitrary $k$ with $1 \leq k \leq K$. For any $x, y \in (a_k, b_k)$, where $x < y$, the inequality $F^S_{\theta,\sigma^2}(x) < F^S_{\theta,\sigma^2}(y)$ holds for all $\theta \in \mathbb{R}$. Suppose that $B_S(x) \geq B_S(y)$. Then it follows that $F^S_{B_S(x),\sigma^2}(x) \leq F^S_{B_S(y),\sigma^2}(x) < F^S_{B_S(y),\sigma^2}(y)$, where, in the first inequality, we used the fact that $F^S_{\theta,\sigma^2}(x)$ is decreasing in $\theta$ for all $x \in (a_1, b_K)$ (Lemma 2.3 (ii)) and, in the second inequality, we used the fact that $F^S_{\theta,\sigma^2}(x)$ is strictly increasing in $x$ on $(a_k, b_k)$. However, by equation (2.75), also $F^S_{B_S(x),\sigma^2}(x) = F^S_{B_S(x),\sigma^2}(x) = q$ must be satisfied for all $x, y \in (a_k, b_k)$. This again contradicts the previous inequality, therefore it follows that $B_S(x)$ is strictly increasing in $x$ on the interval $(a_1, b_K)$.

Analogous arguments can be used to show that $B_S(x)$ is constant in $x$ on the interval $[b_k, a_{k+1}]$ for every $k = 1, \ldots, K-1$.

It remains to show that $B_S(x)$ is continuous in $x$ on $(a_1, b_K)$. To proof this, we first need to show that $F^S_{\theta,\sigma^2}(x)$ is continuous in $(x, \theta)$ on $(a_1, b_K) \times \mathbb{R}$. This means that for all $(x, \theta) \in (a_1, b_K) \times \mathbb{R}$ and for any sequence $\{(x_n, \theta_n)\}_{n \geq 1}$ with $(x_n, \theta_n) \to (x, \theta)$ as $n \to +\infty$, we need to show that $F^S_{\theta_n,\sigma^2}(x_n) \to F^S_{\theta,\sigma^2}(x)$ as $n \to +\infty$. To proof this intermediate step, there again 2 cases that need to be considered.

First case: $(x, \theta) \in S \times \mathbb{R}$. For all $x \in S$, there exists a $k$ (depending on $x$) with $1 \leq k \leq K$ such that $a_k < x < b_k$ and, therefore, $(x, \theta) \in (a_k, b_k) \times \mathbb{R}$. Therefore, $F^S_{\theta,\sigma^2}(x)$ satisfies

$$F^S_{\theta,\sigma^2}(x) = \frac{\Phi\left(\frac{x-\theta}{\sigma}\right) - \Phi\left(\frac{a_k-\theta}{\sigma}\right) + \sum_{i=1}^{k-1} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\sum_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)} \tag{2.76}$$

Since $(a_k, b_k)$ is an open interval and since $(x_n, \theta_n) \to (x, \theta)$ as $n \to +\infty$, it follows that $(x_n, \theta_n) \in (a_k, b_k) \times \mathbb{R}$ for all sufficiently large $n$. This implies that, for all sufficiently large $n$, $F^S_{\theta_n,\sigma^2}(x_n)$ satisfies

$$F^S_{\theta_n,\sigma^2}(x_n) = \frac{\Phi\left(\frac{x_n-\theta_n}{\sigma}\right) - \Phi\left(\frac{a_k-\theta_n}{\sigma}\right) + \sum_{i=1}^{k-1} \Phi\left(\frac{b_i-\theta_n}{\sigma}\right) - \Phi\left(\frac{a_i-\theta_n}{\sigma}\right)}{\sum_{i=1}^{K} \Phi\left(\frac{b_i-\theta_n}{\sigma}\right) - \Phi\left(\frac{a_i-\theta_n}{\sigma}\right)} \tag{2.77}$$

Clearly, the numerator in (2.77) converges to the numerator in (2.76) and, at the same time, the denominator in (2.77) converges to the denominator in (2.76) as $n \to +\infty$. Therefore, $F^S_{\theta_n,\sigma^2}(x_n)$ converges to $F^S_{\theta,\sigma^2}(x)$ as $n \to +\infty$ for all $(x, \theta) \in S \times \mathbb{R}$.

Second case: $(x, \theta) \in ((a_1, b_K) \setminus S) \times \mathbb{R}$. For every $x \in (a_1, b_K) \setminus S$, there exists a $k$ (depending on $x$) with $1 \leq k \leq K-1$ such that $b_k \leq x \leq a_{k+1}$ and, therefore, $(x, \theta) \in [b_k, a_{k+1}] \times \mathbb{R}$. We will only elaborate the more complicated case where $x = a_{k+1}$. Almost identical arguments can be used in the remaining cases. Observe that $F^S_{\theta,\sigma^2}(a_{k+1})$ satisfies

$$F^S_{\theta,\sigma^2}(a_{k+1}) = \frac{\sum_{i=1}^{k} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)}{\sum_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)} \tag{2.78}$$

Since $(x_n, \theta_n) \to (a_{k+1}, \theta)$ as $n \to +\infty$, it follows that, for all sufficiently large $n$, either $(x_n, \theta_n) \in [b_k, a_{k+1}] \times \mathbb{R}$ or $(x_n, \theta_n) \in (a_{k+1}, b_{k+1}) \times \mathbb{R}$.

Consider any non-finite subsequence $\{(x_{n_j}, \theta_{n_j})\}_{j \geq 1}$ (if there exists one) that satisfies $(x_{n_j}, \theta_{n_j}) \in [b_k, a_{k+1}] \times \mathbb{R}$ for all $j \geq 1$. Since $F^S_{\theta,\sigma^2}(x)$ is constant in $x$ on $[b_k, a_{k+1}]$, it follows that $F^S_{\theta_{n_j},\sigma^2}(x_{n_j}) = F^S_{\theta_{n_j},\sigma^2}(a_{k+1})$ for all $j \geq 1$. As $F^S_{\theta,\sigma^2}(x)$ is continuous in $\theta$ for all $x \in (a_1, b_K)$ (Lemma 2.3 (i)), it follows that $F^S_{\theta_{n_j},\sigma^2}(a_{k+1})$ converges to $F^S_{\theta,\sigma^2}(a_{k+1})$ as $j \to +\infty$. This, in turn, implies that $F^S_{\theta_{n_j},\sigma^2}(x_{n_j})$ converges to $F^S_{\theta,\sigma^2}(a_{k+1})$ as $j \to +\infty$.

Now consider any non-finite subsequence $\{(x_{n_j}, \theta_{n_j})\}_{j \geq 1}$ (if there exists one) that satisfies $(x_{n_j}, \theta_{n_j}) \in (a_{k+1}, b_{k+1}) \times \mathbb{R}$ for all $j \geq 1$. This implies that, for all $j \geq 1$, $F^S_{\theta_{n_j},\sigma^2}(x_{n_j})$ satisfies

$$F^S_{\theta_{n_j},\sigma^2}(x_{n_j}) = \frac{\Phi\left(\frac{x_{n_j}-\theta_{n_j}}{\sigma}\right) - \Phi\left(\frac{a_{k+1}-\theta_{n_j}}{\sigma}\right) + \sum_{i=1}^{k} \Phi\left(\frac{b_i-\theta_{n_j}}{\sigma}\right) - \Phi\left(\frac{a_i-\theta_{n_j}}{\sigma}\right)}{\sum_{i=1}^{K} \Phi\left(\frac{b_i-\theta_{n_j}}{\sigma}\right) - \Phi\left(\frac{a_i-\theta_{n_j}}{\sigma}\right)} \tag{2.79}$$

Clearly, the numerator in (2.79) converges to the numerator in (2.78) and, at the same time, the denominator in (2.79) converges to the denominator in (2.78) as $j \to +\infty$. Therefore, $F^S_{\theta_{n_j},\sigma^2}(x_{n_j})$ converges to $F^S_{\theta,\sigma^2}(a_{k+1})$ as $j \to +\infty$.

We proved the convergence of $F^S_{\theta_{n_j},\sigma^2}(x_{n_j})$ to $F^S_{\theta,\sigma^2}(a_{k+1})$ for every non-finite subsequence $\{(x_{n_j}, \theta_{n_j})\}_{j \geq 1}$, where $x_{n_j}$ converges to $a_{k+1}$ either from below or above. This implies that $F^S_{\theta_n,\sigma^2}(x_n)$ converges to $F^S_{\theta,\sigma^2}(a_{k+1})$ as $n \to +\infty$.

Finally, it will be proved that $B_S(x)$ is continuous in $x$ on $(a_1, b_K)$. Let $x \in (a_1, b_K)$ and suppose that the sequence $\{x_n\}_{n \geq 1}$ converges to $x$ as $n \to +\infty$. Consider any non-finite subsequence $\{x_{n_j}\}_{j \geq 1}$ (if there exists one) that converges to $x$ from below as $j \to +\infty$. In the beginning of this proof we showed that $B_S(x)$ is monotonically increasing in $x$ on $(a_1, b_K)$, therefore, the sequence $\{B_S(x_{n_j})\}_{j \geq 1}$ is monotonically increasing. Since $x_{n_j} \leq x$ for all $j \geq 1$, it follows that $B_S(x_{n_j}) \leq B_S(x)$ for all $j \geq 1$. As $B_S(x)$ satisfies equation (2.75) for a $q \in (0, 1)$, it follows that $B_S(x)$ must be finite. Thus the sequence $\{B_S(x_{n_j})\}_{j \geq 1}$ is bounded and, therefore, must converge to a limit, which we denote by $C$. We previously showed that $F^S_{\theta,\sigma^2}(x)$ is continuous in $(x,\theta)$ on $(a_1,b_K)\times\mathbb{R}$. This implies that the sequence $\{(x_{n_j},B_S(x_{n_j}))\}_{j\geq 1}$ satisfies

$$\lim_{j\to+\infty} F^S_{B_S(x_{n_j}),\sigma^2}(x_{n_j}) = F^S_{C,\sigma^2}(x). \tag{2.80}$$

As $F^S_{\theta,\sigma^2}(x)$ is strictly decreasing in $\theta$ for all $x\in(a_1,b_K)$, it follows that $C=B_S(x)$; this means that $B_S(x_{n_j})$ converges to $B_S(x)$ as $j\to+\infty$.

Almost identical arguments can be used to show that $B_S(x_{n_j})$ converges to $B_S(x)$ as $j\to+\infty$ for any non-finite subsequence $\{x_{n_j}\}_{j\geq 1}$ (if there exists one) that converges to $x$ from above as $j\to+\infty$. Thus we can conclude that also $B_S(x_n)$ converges to $B_S(x)$ as $n\to+\infty$. □

**Lemma 2.16.** *Let $B_S(x)$ be defined as in (2.28). Then the following holds:*

*(i)* $\lim_{x\nearrow b_K} B_S(x) = +\infty$

*(ii)* $\lim_{x\searrow a_1} B_S(x) = -\infty.$

*Proof.* Proof of Part (i): Suppose that the sequence $\{x_n\}_{n\geq 1}$ converges to $b_K$ as $n\to+\infty$. By Lemma 2.15, the function $B_S(x)$ is monotonically increasing, therefore, $B_S(x_n)$ must converge to a (not necessarily finite) limit $C$ as $n\to+\infty$. Suppose that $C<+\infty$. Then, for all $\epsilon>0$ such that $b_K-\epsilon\in(a_1,b_K)$, it follows that

$$\liminf_{n\to+\infty} F^S_{B_S(x_n),\sigma^2}(x_n) \geq \liminf_{n\to+\infty} F^S_{B_S(x_n),\sigma^2}(x_n-\epsilon) = F^S_{C,\sigma^2}(b_K-\epsilon), \tag{2.81}$$

where the inequality follows from the fact that $F^S_{\theta,\sigma^2}(x)$ is increasing in $x$ and the equality follows from the continuity of $F^S_{\theta,\sigma^2}(x)$ in $(x,\theta)$ on $(a_1,b_K)\times\mathbb{R}$ (shown as an intermediate result in the proof of Lemma 2.15). As the $\epsilon$ can be chosen arbitrarily small (such that $F^S_{C,\sigma^2}(b_K-\epsilon)$ gets arbitrarily close to 1), it follows that $\liminf_{n\to+\infty} F^S_{B_S(x_n),\sigma^2}(x_n)=1$. However, by definition, $B_S(x)$ satisfies $F^S_{B_S(x),\sigma^2}(x)=q$ for a $q\in(0,1)$ and for all $x\in(a_1,b_k)$. This is clearly a contradiction, thus, it follows that $C=+\infty$.

Proof of Part (ii): Almost identical arguments can used as in the first part of the proof. □

*Proof of Lemma 2.6.* Proof of Part (i): Let $B_S(x)$ be defined as in (2.28). This means that $B_S(x)$ satisfies $F^S_{B_S(x),\sigma^2}(x)=q$ for all $x\in(a_1,b_K)$. For $x$ sufficiently close to $b_K$, it follows that $a_K < x < b_K$. Therefore, suppose that $x \in (a_K, b_K)$. Then the equation $F^S_{B_S(x),\sigma^2}(x) = q$ can be written as follows:

$$\frac{\Phi\left(\frac{x-B_S(x)}{\sigma}\right) - \Phi\left(\frac{a_K-B_S(x)}{\sigma}\right) + \sum_{i=1}^{K-1} \Phi\left(\frac{b_i-B_S(x)}{\sigma}\right) - \Phi\left(\frac{a_i-B_S(x)}{\sigma}\right)}{\sum_{i=1}^{K} \Phi\left(\frac{b_i-B_S(x)}{\sigma}\right) - \Phi\left(\frac{a_i-B_S(x)}{\sigma}\right)} = q \tag{2.82}$$

Multiplying both sides of the equation by $\sum_{i=1}^{K} \Phi\left(\frac{b_i-B_S(x)}{\sigma}\right) - \Phi\left(\frac{a_i-B_S(x)}{\sigma}\right)$ yields

$$\begin{aligned}&\Phi\left(\frac{x-B_S(x)}{\sigma}\right) - \Phi\left(\frac{a_K-B_S(x)}{\sigma}\right) + \sum_{i=1}^{K-1} \Phi\left(\frac{b_i-B_S(x)}{\sigma}\right) - \Phi\left(\frac{a_i-B_S(x)}{\sigma}\right)\\ &= q\sum_{i=1}^{K} \Phi\left(\frac{b_i-B_S(x)}{\sigma}\right) - \Phi\left(\frac{a_i-B_S(x)}{\sigma}\right).\end{aligned} \tag{2.83}$$

The equation is now rearranged such that only $\Phi\left(\frac{x-B_S(x)}{\sigma}\right)$ remains on the left hand side:

$$\begin{aligned}\Phi\left(\frac{x-B_S(x)}{\sigma}\right) &= q\Phi\left(\frac{b_K-B_S(x)}{\sigma}\right)\\ &\quad - (1-q)\left(-\Phi\left(\frac{a_K-B_S(x)}{\sigma}\right) + \sum_{i=1}^{K-1} \Phi\left(\frac{b_i-B_S(x)}{\sigma}\right) - \Phi\left(\frac{a_i-B_S(x)}{\sigma}\right)\right).\end{aligned} \tag{2.84}$$

On the right hand side, we see that $b_K$ of the first term is strictly larger than the values $a_1, b_1, \ldots, b_{K-1}, a_K$ of the second term. As $B_S(x) \to +\infty$ as $x \nearrow b_K$ (Lemma 2.16 (i)), we can use Lemma 2.10 which implies that

$$\lim_{x \nearrow b_K} \frac{q\Phi\left(\frac{b_K-B_S(x)}{\sigma}\right) - (1-q)\left(-\Phi\left(\frac{a_K-B_S(x)}{\sigma}\right) + \sum_{i=1}^{K-1} \Phi\left(\frac{b_i-B_S(x)}{\sigma}\right) - \Phi\left(\frac{a_i-B_S(x)}{\sigma}\right)\right)}{q\Phi\left(\frac{b_K-B_S(x)}{\sigma}\right)} = 1. \tag{2.85}$$

This means that there exists a function $q(x)$ with $q(x) \to q$ as $x \nearrow b_K$ such that

$$\begin{aligned}q(x)\Phi\left(\frac{b_K-B_S(x)}{\sigma}\right) &= q\Phi\left(\frac{b_K-B_S(x)}{\sigma}\right)\\ &\quad - (1-q)\left(-\Phi\left(\frac{a_K-B_S(x)}{\sigma}\right) + \sum_{i=1}^{K-1} \Phi\left(\frac{b_i-B_S(x)}{\sigma}\right) - \Phi\left(\frac{a_i-B_S(x)}{\sigma}\right)\right).\end{aligned} \tag{2.86}$$

Substituting equation (2.86) into (2.84) yields

$$\begin{aligned}
\Phi\left(\frac{x - B_S(x)}{\sigma}\right) &= q(x)\Phi\left(\frac{b_K - B_S(x)}{\sigma}\right) \\
\frac{x - B_S(x)}{\sigma} &= \Phi^{-1}\left(q(x)\Phi\left(\frac{b_K - B_S(x)}{\sigma}\right)\right) \\
x &= B_S(x) + \sigma\Phi^{-1}\left(q(x)\Phi\left(\frac{b_K - B_S(x)}{\sigma}\right)\right) \\
b_K - x &= b_K - B_S(x) - \sigma\Phi^{-1}\left(q(x)\Phi\left(\frac{b_K - B_S(x)}{\sigma}\right)\right) \\
(b_K - x)B_S(x) &= B_S(x)\left(b_K - B_S(x) - \sigma\Phi^{-1}\left(q(x)\Phi\left(\frac{b_K - B_S(x)}{\sigma}\right)\right)\right).
\end{aligned} \tag{2.87}$$

Since $B_S(x) \to +\infty$ (Lemma 2.16 (i)) and $q(x) \to q$ as $x \nearrow b_K$, we can use Corollary 2.14 which implies the following for all $\delta \in \mathbb{R}$, such that $q + \delta > 0$:

$$\lim_{x \nearrow b_K} \frac{q(x)\Phi\left(\frac{b_K - B_S(x)}{\sigma}\right)}{\Phi\left(\frac{b_K - B_S(x) + \frac{\sigma^2 \log(q+\delta)}{B_S(x)}}{\sigma}\right)} = \frac{q}{q+\delta}. \tag{2.88}$$

For all $\delta > 0$, it follows that $\frac{q}{q+\delta}$ is strictly less than 1. On the other hand, for all $\delta < 0$, such that $q + \delta > 0$ (such a $\delta$ always exists as $q \in (0, 1)$), it follows that $\frac{q}{q+\delta}$ is strictly greater than 1. This implies that for all $x$ sufficiently close to $b_K$, the numerator on the left hand side is strictly smaller than the denominator if $\delta > 0$, and strictly larger if $\delta < 0$ (for a $\delta$ sufficiently close to 0). Thus it follows that for every $\delta > 0$ small enough such that $q - \delta > 0$, the following inequality holds for all $x$ sufficiently close to $b_K$:

$$\Phi\left(\frac{b_K - B_S(x) + \frac{\sigma^2 \log(q-\delta)}{B_S(x)}}{\sigma}\right) < q(x)\Phi\left(\frac{b_K - B_S(x)}{\sigma}\right) < \Phi\left(\frac{b_K - B_S(x) + \frac{\sigma^2 \log(q+\delta)}{B_S(x)}}{\sigma}\right). \tag{2.89}$$

Thus inequality (2.89) and the last line of equation (2.87) yield the following for all

$x$ sufficiently close to $b_K$:

$$B_S(x)\left(b_K - B_S(x) - \sigma\Phi^{-1}\left(\Phi\left(\frac{b_K - B_S(x) + \frac{\sigma^2\log(q+\delta)}{B_S(x)}}{\sigma}\right)\right)\right)$$
$$< (b_K - x)B_S(x) < B_S(x)\left(b_K - B_S(x) - \sigma\Phi^{-1}\left(\Phi\left(\frac{b_K - B_S(x) + \frac{\sigma^2\log(q-\delta)}{B_S(x)}}{\sigma}\right)\right)\right). \tag{2.90}$$

Observe that the previous inequality simplifies to

$$-\sigma^2\log(q+\delta) < (b_K - x)B_S(x) < -\sigma^2\log(q-\delta) \quad \text{for all } x \text{ sufficiently close to } b_K. \tag{2.91}$$

Since $\delta$ can be chosen arbitrarily small and the logarithm is a continuous function, it follows that $\lim_{x\nearrow b_K}(b_K - x)B_S(x) = -\sigma^2\log(q)$.

Proof of Part (ii): Let denote by $-S$ the reflection of the set $S$ as defined in (2.51). Lemma 2.12 implies that $F^S_{\theta,\sigma^2}(x) = 1 - F^{-S}_{-\theta,\sigma^2}(-x)$ for all $x, \theta \in \mathbb{R}$. As $B_S(x)$ satisfies the equation $F^S_{B_S(x),\sigma^2}(x) = q$ for all $x \in (a_1, b_K)$, it follows that

$$F^{-S}_{-B_S(x),\sigma^2}(-x) = 1 - q \quad \text{for all } x \in (a_1, b_K). \tag{2.92}$$

Now observe that $-a_1$ is the rightmost boundary point of the set $-S$ and that $-a_1 < +\infty$. Furthermore, $x \searrow a_1$ implies that $-x \nearrow -a_1$. Thus part (i) of Lemma 2.6 implies that

$$\lim_{x\searrow a_1}(-a_1 + x)(-B_S(x)) = -\log(1-q). \tag{2.93}$$

□

## 2.7 Proof of Lemma 2.8

Proof of Part (i): Let $b_K = +\infty$. This implies that $\Phi\left(\frac{b_K-\theta}{\sigma}\right) = 1$ for all $\theta \in \mathbb{R}$. Let $B_S(x)$ be defined as in (2.28). This means that $B_S(x)$ satisfies $F^S_{B_S(x),\sigma^2}(x) = q$ for all $x \in (a_1, b_K)$. Suppose that $a_K < x < +\infty$. Then the equation $F^S_{B_S(x),\sigma^2}(x) = q$ can be written as follows:

$$\frac{\Phi\left(\frac{x-B_S(x)}{\sigma}\right) - \Phi\left(\frac{a_K-B_S(x)}{\sigma}\right) + \sum_{i=1}^{K-1}\Phi\left(\frac{b_i-B_S(x)}{\sigma}\right) - \Phi\left(\frac{a_i-B_S(x)}{\sigma}\right)}{1 - \Phi\left(\frac{a_K-B_S(x)}{\sigma}\right) + \sum_{i=1}^{K-1}\Phi\left(\frac{b_i-B_S(x)}{\sigma}\right) - \Phi\left(\frac{a_i-B_S(x)}{\sigma}\right)} = q. \tag{2.94}$$

The function $N(x)$ is defined as follows:

$$N(x) = -\Phi\left(\frac{a_K - B_S(x)}{\sigma}\right) + \sum_{i=1}^{K-1} \Phi\left(\frac{b_i - B_S(x)}{\sigma}\right) - \Phi\left(\frac{a_i - B_S(x)}{\sigma}\right), \tag{2.95}$$

Substituting (2.95) in (2.94) first and then rearranging the equation yields

$$\begin{aligned} \frac{\Phi\left(\frac{x-B_S(x)}{\sigma}\right) + N(x)}{1+N(x)} &= q \\ \Phi\left(\frac{x-B_S(x)}{\sigma}\right) + N(x) &= q + qN(x) \\ \Phi\left(\frac{x-B_S(x)}{\sigma}\right) &= q - (1-q)N(x). \end{aligned} \tag{2.96}$$

Set $B(x) = x - \sigma\Phi^{-1}(q)$. Further, rearranging the previous equation yields

$$\begin{aligned} x - B_S(x) &= \sigma\Phi^{-1}(q-(1-q)N(x)) \\ x - \sigma\Phi^{-1}(q) - B_S(x) &= -\sigma\left(\Phi^{-1}(q) - \Phi^{-1}(q-(1-q)N(x))\right) \\ B(x) - B_S(x) &= -\sigma\left(\Phi^{-1}(q) - \Phi^{-1}(q-(1-q)N(x))\right). \end{aligned} \tag{2.97}$$

Since $B_S(x) \to +\infty$ as $x \to +\infty$ (Lemma 2.16 (i)) and by equation (2.95), it follows that $N(x) \to 0$ as $x \to +\infty$. Therefore, by the continuity of the function $\Phi^{-1}$, the right hand side in the last line of the previous equation converges to 0 as $x \to +\infty$. Thus it follows that

$$\lim_{x\to+\infty} B(x) - B_S(x) = 0. \tag{2.98}$$

Proof of Part (ii): Suppose that $a_1 = -\infty$. Let denote by $-S$ the reflection of the set $S$ as defined in (2.51). Lemma 2.12 implies that $F^S_{\theta,\sigma^2}(x) = 1 - F^{-S}_{-\theta,\sigma^2}(-x)$ for all $x, \theta \in \mathbb{R}$. As $B_S(x)$ satisfies the equation $F^S_{B_S(x),\sigma^2}(x) = q$ for all $x \in (-\infty, b_K)$ it follows that

$$F^{-S}_{-B_S(x),\sigma^2}(-x) = 1 - q \quad \text{for all } x \ \in (-\infty, b_K). \tag{2.99}$$

Since $S$ is unbounded from below, it follows that the set $-S$ is unbounded from above. Moreover, if $x \to -\infty$, then also $-x \to +\infty$. Thus Lemma 2.8 (i) implies that

$$\lim_{x\to-\infty} -x - \sigma\Phi^{-1}(1-q) + B_S(x) = 0. \tag{2.100}$$

Note that $\sigma\Phi^{-1}(1-q) = -\sigma\Phi^{-1}(q)$; therefore the previous equation can be rewritten as follows:

$$\begin{aligned}\lim_{x\to-\infty} -x + \sigma\Phi^{-1}(q) + B_S(x) = 0 \\ \lim_{x\to-\infty} -B(x) + B_S(x) = 0.\end{aligned} \tag{2.101}$$

## 2.8 Proof of Theorem 2.9

Before the theorem is proved, another technical result is required. Throughout the section, we suppose that $\sigma^2 \in (0, +\infty)$. All results hold for any $\sigma^2 \in (0, +\infty)$.

**Lemma 2.17.** *The function $\frac{\phi(x)}{\Phi(x)}$ is strictly decreasing and the function $\frac{\phi(x)}{1-\Phi(x)}$ is strictly increasing.*

*Proof.* Set $m(x) = \frac{\phi(x)}{\Phi(x)}$. Observe that $m(-x) = \frac{\phi(-x)}{\Phi(-x)} = \frac{\phi(x)}{1-\Phi(x)}$. Thus it is sufficient to show that $m(x)$ is strictly decreasing. $m(x)$ is strictly positive for all $x \in \mathbb{R}$. The derivative $m'(x)$ is given by

$$m'(x) = -x\frac{\phi(x)}{\Phi(x)} - \frac{\phi(x)^2}{\Phi(x)^2} = -(x + m(x))m(x). \tag{2.102}$$

Thus $m'(x) < 0$ holds for all $x \geq 0$. It remains to show that $m'(x) < 0$ holds for all $x < 0$. Assume that $X$ is a standard normally distributed random variable. This means that $X$ has p.d.f. $\phi(x)$. Let denote by $\phi'(x)$ the derivative of $\phi'(x)$ and note the following symmetry property: $\phi'(x) = -x\phi(x)$. The expectation of $X$ conditional on $\{X < x\}$ is given by

$$\begin{aligned}\mathbb{E}(X|X < x) &= \int_{-\infty}^{x} u\frac{\phi(u)}{\Phi(x)}du = \frac{1}{\Phi(x)}\int_{-\infty}^{x} u\phi(u)du \\ &= -\frac{1}{\Phi(x)}\int_{-\infty}^{x} \phi'(u)du = -\frac{\phi(x)}{\Phi(x)} = -m(x).\end{aligned} \tag{2.103}$$

The inequality $\mathbb{E}(X|X < x) < x$ is trivially satisfied for all $x \in \mathbb{R}$. This implies that $0 < x+m(x)$ holds for all $x < 0$. Hence, it follows that $m'(x) = -(x+m(x))m(x) < 0$ for all $x < 0$. $\square$

*Proof of Theorem 2.9.* Suppose that $a_1 = -\infty$ and $b_K = +\infty$. This implies that $\Phi\left(\frac{a_1-\theta}{\sigma}\right) = 0$ and $\Phi\left(\frac{b_K-\theta}{\sigma}\right) = 1$ for all $\theta \in \mathbb{R}$. Further, suppose that $K > 1$, i.e., that the truncation set $S$ is at least a union of two disjoint open intervals. This implies that $-\infty < b_1 < a_K < +\infty$. Note that $(a_1, b_K) = \mathbb{R}$. Let the function $L_S(x)$ and the function $U_S(x)$ be defined as in (2.16) and (2.17), respectively. This means that $L_S(x)$ satisfies $F^S_{L_S(x),\sigma^2}(x) = 1 - \frac{\alpha}{2}$ for all $x \in \mathbb{R}$ and that $U_S(x)$ satisfies $F^S_{U_S(x),\sigma^2}(x) = \frac{\alpha}{2}$ for all $x \in \mathbb{R}$. We define the functions $N_L(x)$ and $N_U(x)$ as follows:

$$N_L(x) = -\Phi\left(\frac{a_K - L_S(x)}{\sigma}\right) + \sum_{i=1}^{K-1} \Phi\left(\frac{b_i - L_S(x)}{\sigma}\right) - \Phi\left(\frac{a_i - L_S(x)}{\sigma}\right), \tag{2.104}$$

$$N_U(x) = -\Phi\left(\frac{a_K - U_S(x)}{\sigma}\right) + \sum_{i=1}^{K-1} \Phi\left(\frac{b_i - U_S(x)}{\sigma}\right) - \Phi\left(\frac{a_i - U_S(x)}{\sigma}\right). \tag{2.105}$$

As $a_K$ is strictly larger than $a_1, b_1, \ldots, a_{K-1}, b_{K-1}$, it follows that $N_L(x) < 0$ and $N_U(x) < 0$ hold for all $x \in \mathbb{R}$. To obtain an upper bound for $U_S(x) - L_S(x)$, we derive an upper bound for $U_S(x)$ and a lower bound for $L_S(x)$.

Upper bound: For all $x \in (a_K, +\infty)$, the equation $F^S_{U_S(x),\sigma^2}(x) = \frac{\alpha}{2}$ can be written as follows:

$$\frac{\Phi\left(\frac{x - U_S(x)}{\sigma}\right) + N_U(x)}{1 + N_U(x)} = \frac{\alpha}{2}. \tag{2.106}$$

Rearranging the previous equation yields:

$$x - U_S(x) = \sigma\Phi^{-1}\left(\frac{\alpha}{2} - (1 - \frac{\alpha}{2})N_U(x)\right) \quad \text{for all } x \in (a_K, +\infty). \tag{2.107}$$

$N_U(x) < 0$ for all $x \in \mathbb{R}$ implies that $-(1 - \frac{\alpha}{2})N_U(x) > 0$ for all $x \in \mathbb{R}$. The function $\Phi^{-1}$ is strictly increasing; therefore, it follows that

$$\begin{aligned} x - U_S(x) &> \sigma\Phi^{-1}\left(\frac{\alpha}{2}\right) \quad \text{for all } x \in (a_K, +\infty), \\ U_S(x) &< x - \sigma\Phi^{-1}\left(\frac{\alpha}{2}\right) \quad \text{for all } x \in (a_K, +\infty). \end{aligned} \tag{2.108}$$

Hence, $x - \sigma\Phi^{-1}\left(\frac{\alpha}{2}\right)$ is an upper bound of $U_S(x)$ for all $x \in (a_K, +\infty)$. Lemma 2.15 implies that $U_S(x)$ is strictly monotonically increasing; hence, for all $x \in (-\infty, a_K]$, the function $U_S(x)$ can be bounded from above as follows:

$$U_S(x) \leq U_S(a_K) < a_K - \sigma\Phi^{-1}\left(\frac{\alpha}{2}\right) \quad \text{for all } x \in (-\infty, a_K]. \tag{2.109}$$

Note that equation (2.106) and (2.107) and inequality (2.108) are still valid in the case $x = a_K$ (simply replace each $x$ by $a_K$), thus, the second inequality in the previous line is strict.

Lower bound: For all $x \in (-\infty, b_1)$, the equation $F^S_{L_S(x),\sigma^2}(x) = 1 - \frac{\alpha}{2}$ can be written as follows:

$$\frac{\Phi\left(\frac{x-L_S(x)}{\sigma}\right)}{1+N_L(x)} = 1 - \frac{\alpha}{2}. \tag{2.110}$$

Rearranging the previous equation yields:

$$x - L_S(x) = \sigma\Phi^{-1}\left((1-\frac{\alpha}{2})(1+N_L(x)\right) \quad \text{for all } x \in (-\infty, b_1). \tag{2.111}$$

Again, note that $N_L(x) < 0$ for all $x \in \mathbb{R}$ implies that $1 + N_L(x) < 1$ holds for all $x \in \mathbb{R}$. Therefore, by the monotonicity of $\Phi^{-1}$ it follows that

$$\begin{aligned} x - L_S(x) &< \sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right) \quad \text{for all } x \in (-\infty, b_1), \\ L_S(x) &> x - \sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right) \quad \text{for all } x \in (-\infty, b_1). \end{aligned} \tag{2.112}$$

Thus it follows that $x - \sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$ is a lower bound of $L_S(x)$ for all $x \in (-\infty, b_1)$. $L_S(x)$ is monotonically increasing by Lemma 2.15, therefore, for all $x \in [b_1, +\infty)$, the function $L_S(x)$ can be bounded from below by:

$$L_S(x) \geq L_S(b_1) > b_1 - \sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right) \quad \text{for all } x \in [b_1, +\infty). \tag{2.113}$$

Note again that equation (2.110) and (2.111) and inequality (2.112) are still valid in the case $x = b_1$ (simply replace each $x$ by $b_1$), thus, the second inequality in the previous line is strict.

As we derived an upper bound for $U_S(x)$ and a lower bound for $L_S(x)$ for all $x \in R$, we can now bound the interval length $U_S(x) - L_S(x)$ from above as follows (note that $\Phi^{-1}\left(1-\frac{\alpha}{2}\right) = -\Phi^{-1}\left(\frac{\alpha}{2}\right)$):

$$U_S(x) - L_S(x) < (a_K - x) + 2\sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right) \quad \text{for all } x \in (-\infty, b_1), \tag{2.114}$$

$$U_S(x) - L_S(x) < (a_K - b_1) + 2\sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right) \quad \text{for all } x \in [b_1, a_K], \tag{2.115}$$

$$U_S(x) - L_S(x) < (x - b_1) + 2\sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right) \quad \text{for all } x \in (a_K, +\infty). \tag{2.116}$$

Conditional on the set $S$, $X$ is truncated normally distributed with p.d.f. $f^S_{\theta,\sigma^2}(x)$.

Thus the conditional expected length is bounded from above by:

$$\mathbb{E}_{\theta,\sigma^2}(U_S(X) - L_S(X) \mid X \in S) = \int_{-\infty}^{+\infty} (U_S(x) - L_S(x)) f^S_{\theta,\sigma^2}(x)dx \tag{2.117}$$

$$< \int_{-\infty}^{b_1} (a_K - x) f^S_{\theta,\sigma^2}(x)dx + \int_{b_1}^{a_K} (a_K - b_1) f^S_{\theta,\sigma^2}(x)dx + \int_{a_K}^{+\infty} (x - b_1) f^S_{\theta,\sigma^2}(x)dx + 2\sigma\Phi^{-1}\left(1 - \frac{\alpha}{2}\right). \tag{2.118}$$

The integrals in the previous lines can be easily calculated, but we will derive a larger upper bound that is simpler to handle. Note that $\sum_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right) > \Phi\left(\frac{b_1-\theta}{\sigma}\right)$ and $\sum_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right) > 1 - \Phi\left(\frac{a_K-\theta}{\sigma}\right)$ holds for all $b_1, a_K \in \mathbb{R}$. Therefore, it follows that

$$f^S_{\theta,\sigma^2}(x) = \frac{\frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right)}{\sum_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)} < \frac{\frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right)}{\Phi\left(\frac{b_1-\theta}{\sigma}\right)} \quad \text{for all } x < b_1 \tag{2.119}$$

$$f^S_{\theta,\sigma^2}(x) = \frac{\frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right)}{\sum_{i=1}^{K} \Phi\left(\frac{b_i-\theta}{\sigma}\right) - \Phi\left(\frac{a_i-\theta}{\sigma}\right)} < \frac{\frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right)}{1 - \Phi\left(\frac{x-\theta}{\sigma}\right)} \quad \text{for all } x > a_K. \tag{2.120}$$

The first integral in (2.118) can be bounded from above as follows: Note that $a_K - x > 0$ holds for all $y \in (-\infty, b_1)$. Thus it follows that

$$\int_{-\infty}^{b_1} (a_K - x) f^S_{\theta,\sigma^2}(x)dx < \int_{-\infty}^{b_1} (a_K - x) \frac{\frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right)}{\Phi\left(\frac{b_1-\theta}{\sigma}\right)} dx = a_K - \int_{-\infty}^{b_1} x \frac{\frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right)}{\Phi\left(\frac{b_1-\theta}{\sigma}\right)} dx. \tag{2.121}$$

Note that the integral on the right hand side is simply the expectation of $X$ conditional on $\{X < b_1\}$, that is, $\mathbb{E}_{\theta,\sigma^2}(X|X < b_1)$. The integral is solved by substitution: Set $u = \frac{x-\theta}{\sigma}$. Observe that $x = \sigma u + \theta$, and note that $\phi'(u) = -u\phi(u)$. Therefore, we obtain

$$\begin{aligned} \int_{-\infty}^{b_1} x \frac{\frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right)}{\Phi\left(\frac{b_1-\theta}{\sigma}\right)} dx &= \int_{-\infty}^{\frac{b_1-\theta}{\sigma}} (\sigma u + \theta) \frac{\phi(u)}{\Phi\left(\frac{b_1-\theta}{\sigma}\right)} du = \theta + \sigma \int_{-\infty}^{\frac{b_1-\theta}{\sigma}} u \frac{\phi(u)}{\Phi\left(\frac{b_1-\theta}{\sigma}\right)} du \\ &= \theta - \sigma \int_{-\infty}^{\frac{b_1-\theta}{\sigma}} \frac{\phi'(u)}{\Phi\left(\frac{b_1-\theta}{\sigma}\right)} du = \theta - \sigma \frac{\phi\left(\frac{b_1-\theta}{\sigma}\right)}{\Phi\left(\frac{b_1-\theta}{\sigma}\right)}. \end{aligned} \tag{2.122}$$

Thus the first integral in (2.118) can be bounded from above as follows:

$$\int_{-\infty}^{b_1} (a_K - x) f^S_{\theta,\sigma^2}(x)dx < a_K - \theta + \sigma \frac{\phi\left(\frac{b_1-\theta}{\sigma}\right)}{\Phi\left(\frac{b_1-\theta}{\sigma}\right)}. \tag{2.123}$$

For the second integral in (2.118), we obtain:

$$\int_{b_1}^{a_K} (a_K - b_1) f^S_{\theta,\sigma^2}(x)dx = (a_K - b_1)(F^S_{\theta,\sigma^2}(a_K) - F^S_{\theta,\sigma^2}(b_1)) < a_K - b_1. \tag{2.124}$$

For the third integral in (2.118), arguments analogous to the ones of the first integral in (2.118) are used to obtain

$$\int_{a_K}^{+\infty} (x - b_1) f^S_{\theta,\sigma^2}(x)dx < \int_{a_K}^{+\infty} (x - b_1) \frac{\frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right)}{1 - \Phi\left(\frac{a_K-\theta}{\sigma}\right)} dx = -b_1 + \theta + \sigma \frac{\phi\left(\frac{a_K-\theta}{\sigma}\right)}{1 - \Phi\left(\frac{a_K-\theta}{\sigma}\right)}. \tag{2.125}$$

Thus it follows that the conditional expected length is bounded from above by:

$$\begin{aligned}
&\mathbb{E}_{\theta,\sigma^2}(U_S(X) - L_S(X)|X \in S) \\
&\quad < a_K - \theta + \sigma \frac{\phi\left(\frac{b_1-\theta}{\sigma}\right)}{\Phi\left(\frac{b_1-\theta}{\sigma}\right)} + a_K - b_1 - b_1 + \theta + \sigma \frac{\phi\left(\frac{a_K-\theta}{\sigma}\right)}{1 - \Phi\left(\frac{a_K-\theta}{\sigma}\right)} + 2\sigma\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \\
&\quad = 2(a_K - b_1) + \sigma \frac{\phi\left(\frac{b_1-\theta}{\sigma}\right)}{\Phi\left(\frac{b_1-\theta}{\sigma}\right)} + \sigma \frac{\phi\left(\frac{a_K-\theta}{\sigma}\right)}{1 - \Phi\left(\frac{a_K-\theta}{\sigma}\right)} + 2\sigma\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \\
&\quad = C_{\theta,\sigma^2}(b_1, a_K).
\end{aligned} \tag{2.126}$$

Lemma 2.17 implies that $\frac{\phi\left(\frac{b_1-\theta}{\sigma}\right)}{\Phi\left(\frac{b_1-\theta}{\sigma}\right)}$ is strictly decreasing in $b_1$ and that $\frac{\phi\left(\frac{a_K-\theta}{\sigma}\right)}{1-\Phi\left(\frac{a_K-\theta}{\sigma}\right)}$ is strictly increasing in $a_K$. Thus it follows that $C_{\theta,\sigma^2}(b_1, a_K)$ is strictly decreasing in $b_1$ and strictly increasing in $a_K$. $\square$

# 3 Post-selection inference with the lasso

## 3.1 The framework

Let $Y$ denote the $n$-dimensional response vector from a multivariate normal distribution:

$$Y \sim \mathrm{N}(\mu, \sigma^2 I_n), \quad \mu \in \mathbb{R}^n,\ \sigma^2 \in (0, +\infty). \tag{3.1}$$

The variance $\sigma^2$ is assumed to be known. Let $W = (w_1, \ldots, w_p)$, with $w_i \in \mathbb{R}^n$ for each $i = 1, \ldots, p$, be the non-stochastic regression matrix. We assume that $w_j \neq 0$ for all $j = 1, \ldots, p$. It is not necessary to assume that $p \leq n$ (i.e. that the number of parameters is less than or equal to the number of observations). All subsets of the full model are collected in $\mathcal{M}$, that is, $\mathcal{M} = \{M : M \subseteq \{1, \ldots, p\}\}$. The full model $\{1, \ldots, p\}$ is denoted by $M_F$ and the empty model is denoted by $\emptyset$. For any $M = \{i_1, \ldots, i_k\} \in \mathcal{M} \setminus \emptyset$, we set $W_M = (w_{i_1}, \ldots, w_{i_k})$. Analogously, for any vector $v \in \mathbb{R}^p$, we set $v_M = (v_{i_1}, \ldots, v_{i_k})'$. We will also use the above notation if $M$ is the empty set. The context will help to clarify how expressions containing $W_M$ and $v_M$ are to be interpreted. The complement of the model $M$ (i.e. the variables not in $M$) is denoted by $M^c = M_F \setminus M$. The cardinality of the model $M$ is denoted by $|M|$.

The lasso estimator, denoted by $\hat{\beta}(y)$, is a minimizer of the least squares problem with an additional penalty on the absolute size of the regression coefficients (Frank and Friedman (1993), Tibshirani (1996)):

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|y - W\beta\|_2^2 + \lambda \|\beta\|_1, \quad y \in \mathbb{R}^n,\ \lambda > 0. \tag{3.2}$$

The tuning parameter $\lambda$ controls the amount of shrinkage that is applied to the regression parameters. The lasso has the property that some coefficients of $\hat{\beta}(y)$ are exactly shrunk to 0 (note that this is not necessarily true for all $y \in \mathbb{R}^n$); for this reason it is viewed as, both, a regularization procedure and a model selection procedure. A minimizer of the function always exists, but it is not necessarily unique. A sufficient condition for the uniqueness of $\hat{\beta}(y)$ for all $y \in \mathbb{R}^n$ is that the columns of $W$ are in general position (Tibshirani (2013), Lemma 3).

**Definition 3.1.** *The matrix $W$ has columns in general position if for any set $\{i_1, \ldots, i_k\} \in \mathcal{M}$ with $k \leq \min(n,p)$ and for any sign vector $(\sigma_1, \ldots, \sigma_k)' \in \{-1,1\}^k$, the affine span of the points $\sigma_1 w_{i_1}, \ldots, \sigma_k w_{i_k}$ does not contain any element of $\{\pm w_i : i \neq i_1, \ldots, i_k\}$. This means that for any $\gamma \in \mathbb{R}^k$ such that $\sum_{i=1}^{k} \gamma_i = 1$, the following must hold for all for all $(\sigma_1, \ldots, \sigma_k)' \in \{-1,1\}^k$:*

$$w_i \neq \gamma_1 \sigma_1 w_{i_1} + \cdots + \gamma_k \sigma_k w_{i_k} \quad \text{for all } i \neq i_1, \ldots, i_k, \tag{3.3}$$

$$-w_i \neq \gamma_1 \sigma_1 w_{i_1} + \cdots + \gamma_k \sigma_k w_{i_k} \quad \text{for all } i \neq i_1, \ldots, i_k. \tag{3.4}$$

We, henceforth, assume that the columns of $W$ are in general position to ensure the uniqueness of the lasso estimator $\hat{\beta}(y)$. This assumption is a relatively mild condition as it is shown that if the entries of the matrix $W$ are drawn from a distribution with continuous p.d.f., then the columns of $W$ are in general positions with probability 1 (Tibshirani (2013), Lemma 4).

The model selected by the lasso, denoted by $\hat{M}(y)$, is defined as the set of indices of the nonzero coefficients of $\hat{\beta}(y)$, that is,

$$\hat{M}(y) = \left\{ j : \hat{\beta}_j(y) \neq 0 \right\}. \tag{3.5}$$

Under the assumption that the columns of $W$ are in general position, the function $\hat{M}(y)$ satisfies the following properties (Tibshirani (2013), Lemma 2 and Lemma 3):

**Lemma 3.2.** *Let the columns of $W$ be in general position. Then for all $y \in \mathbb{R}^n$, it follows that $|\hat{M}(y)| \leq \min(n,p)$ and that the matrix $W_{\hat{M}(y)}$ has full rank, i.e. rank$\left(W_{\hat{M}(y)}\right) = |\hat{M}(y)|$.*

If the empty model is selected, that is if $\hat{M}(y) = \emptyset$, then we adopt the convention that the rank of $W_{\hat{M}(y)}$ is 0 and therefore $W_{\hat{M}(y)}$ has full rank. The lemma implies that any model $M \in \mathcal{M}$ with $|M| > \min(n,p)$ is never selected by the lasso. Let $\mathcal{M}^+$ denote the subset of models $M \in \mathcal{M}$ that have positive selection probability, that is,

$$\mathcal{M}^+ = \left\{ M \in \mathcal{M} : \mathbb{P}_{\mu,\sigma^2}(\hat{M}(Y) = M) > 0 \right\}. \tag{3.6}$$

As the measures $\mathbb{P}_{\mu_0,\sigma_0^2}$ and $\mathbb{P}_{\mu_1,\sigma_1^2}$ are equivalent with respect to the null sets for any $\mu_0, \mu_1 \in \mathbb{R}$ and $\sigma_0^2, \sigma_1^2 \in (0,\infty)$ (in the sense that a set is a $\mathbb{P}_{\mu_0,\sigma_0^2}$-null set if and only if it is a $\mathbb{P}_{\mu_1,\sigma_1^2}$-null set), it follows that $\mathcal{M}^+$ does not depend on $\mu$ and $\sigma^2$. It will be shown in Section 3.2, the empty set $\emptyset$ is always an element of $\mathcal{M}^+$. Note that by Lemma 3.2 it follows that $|M| \leq \min(n,p)$ and that $W_M$ has full rank for all $M \in \mathcal{M}^+$. The latter property is particularly convenient as it implies that the invertability of $W_M' W_M$ is assured for all $M \in \mathcal{M}^+ \setminus \emptyset$.

For all $M \in \mathcal{M} \setminus \emptyset$, we set $\mathcal{S}_M = \{-1,1\}^{|M|}$; we also set $\mathcal{S}_\emptyset = \{0\}$. Let denote by $\hat{s}_{\hat{M}}(y)$ the sign vector of the non-zero elements of the lasso estimator $\hat{\beta}(y)$, that is,

$$\hat{s}_{\hat{M}}(y) = \operatorname{sign}\left(\hat{\beta}_{\hat{M}(y)}(y)\right). \tag{3.7}$$

If $\hat{M}(y) = \emptyset$, then $\hat{\beta}_{\hat{M}(y)}(y)$ is to be interpreted as 0 and therefore $\hat{s}_{\hat{M}}(y) = 0$. As shown in Lee et al. (2016), it is substantially easier to first characterize the set $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\} = \{y \in \mathbb{R}^n : \hat{M}(y) = M, \hat{s}_{\hat{M}}(y) = s_M\}$ for all $s_M \in \mathcal{S}_M$. Then the set $\{\hat{M} = M\} = \{y \in \mathbb{R}^n : \hat{M}(y) = M\}$ is simply obtained by taking the union of the sets $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ over all $s_M \in \mathcal{S}_M$:

$$\{\hat{M} = M\} = \bigcup_{s_M \in \mathcal{S}_M} \{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}. \tag{3.8}$$

Just as not all models $M \in \mathcal{M}$ are selected with positive probability, not every sign vector $s_M \in \mathcal{S}_M$ needs to be observed in the model $M$ (see Example 3.16). Thus let us define the set $\mathcal{S}_M^+$ as the collection of sign vectors $s_M$ that have positive observation probability in the model $M$, that is,

$$\mathcal{S}_M^+ = \{s_M \in \mathcal{S}_M : \mathbb{P}_{\mu,\sigma^2}(\hat{M}(Y) = M, \hat{s}_{\hat{M}}(Y) = s_M) > 0\} \quad \text{for all } M \in \mathcal{M}. \tag{3.9}$$

Again, the equivalence of $\mathbb{P}_{\mu_0,\sigma_0^2}$ and $\mathbb{P}_{\mu_1,\sigma_1^2}$ with respect to null sets for any $\mu_0, \mu_1 \in \mathbb{R}$ and $\sigma_0^2, \sigma_1^2 \in (0,\infty)$ implies that $\mathcal{S}_M^+$ does not depend on $\mu$ and $\sigma^2$. Clearly, it holds that $\mathcal{S}_M^+$ is non-empty if and only if $M \in \mathcal{M}^+$.

The coverage target of the confidence intervals, denoted by $\beta^M$, depends on the selected model $M$. For all $M \in \mathcal{M}^+ \setminus \emptyset$, $\beta^M$, is defined as follows:

$$\beta^M = \underset{b^M \in \mathbb{R}^{|M|}}{\arg\min} \|\mu - W_M b^M\|_2^2 = (W_M' W_M)^{-1} W_M' \mu \quad \text{for all } M \in \mathcal{M}^+. \tag{3.10}$$

If $M = \emptyset$, then $\beta^M$ is to be interpreted as 0. For all $M \in \mathcal{M}^+ \setminus \emptyset$, $W_M$ has full rank (Lemma 3.2) and, therefore, $\beta^M$ is well-defined. The target $\beta^M$ is chosen such that $W_M\beta^M$ is the best linear approximation of the mean $\mu$ in the model $M$ with respect to quadratic loss. Note that for some models $M \in \mathcal{M} \setminus \mathcal{M}^+$ an analogous target is not even defined, as it is not guaranteed that $W_M$ has full rank. Thus confidence intervals are defined only for models $M \in \mathcal{M}^+$. Confidence intervals are only calculated for the target $\beta^{\hat{M}(y)}$, i.e. only for the target of the model that was selected by the lasso. If $\hat{M}(y) = \emptyset$, then inference is trivial as the target is always defined as 0. However, for all $M \in \mathcal{M}^+ \setminus \emptyset$, inference is non-trivial. Two types of confidence intervals were proposed by Lee et al. (2016): The first interval, denoted by $[L_{M,j}(Y), U_{M,j}(Y)]$, is defined for all $M \in \mathcal{M}^+ \setminus \emptyset$ and all $j \in \{1, \dots, |M|\}$ and satisfies

$$\mathbb{P}_{\mu,\sigma^2}\left(L_{M,j}(Y) \leq \beta_j^M \leq U_{M,j}(Y) | \hat{M} = M\right) = 1 - \alpha \quad \text{for all } \mu \in \mathbb{R}^n, \sigma^2 \in (0, +\infty). \tag{3.11}$$

The second interval, denoted by $[L_{M,s_M,j}(Y), U_{M,s_M,j}(Y)]$, is defined for all $M \in \mathcal{M}^+ \setminus \emptyset$, all $s_M \in \mathcal{S}_M^+$ and all $j \in \{1, \dots, |M|\}$ and satisfies

$$\mathbb{P}_{\mu,\sigma^2}\left(L_{M,s_M,j}(Y) \leq \beta_j^M \leq U_{M,s_M,j}(Y) | \hat{M} = M, \hat{s}_{\hat{M}} = s_M\right) = 1 - \alpha \quad \text{for all } \mu \in \mathbb{R}^n, \sigma^2 \in (0, +\infty). \tag{3.12}$$

Thus $[L_{M,j}(Y), U_{M,j}(Y)]$ and $[L_{M,s_M,j}(Y), U_{M,s_M,j}(Y)]$ have exact conditional coverage probability on the set $\{\hat{M} = M\}$ and $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$, respectively. Note that the $j$-element of the model $M$ is not necessarily equal to the $j$-element of a different model $M'$.

Finally, set

$$\eta_{M,j} = W_M(W_M' W_M)^{-1} e_j \quad \text{for all } M \in \mathcal{M}^+ \setminus \emptyset,\ j = 1, \dots, |M|, \tag{3.13}$$

where $e_j$ is the $j$th standard basis vector of $\mathbb{R}^{|M|}$. Observe that $\eta_{M,j}Y$ is an unbiased estimator of the $j$th coefficient of the target $\beta^M$, i.e., $\mathbb{E}(\eta_{M,j}'Y) = \beta_j^M$. To construct confidence intervals for $\beta_j^M$ that have exact conditional coverage of $1 - \alpha$, it will be essential to understand the distribution of

$$\eta_{M,j}'Y | \{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\} \tag{3.14}$$

and

$$\eta_{M,j}'Y | \{\hat{M} = M\}. \tag{3.15}$$

## 3.2 Characterization of the selection event

In the following, a few properties of convex functions are derived that will be required for the characterization of the set $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$. The subdifferential of a convex function is defined as follows:

**Definition 3.3.** *Let $m \geq 1$ and let $f : \mathbb{R}^m \to \mathbb{R}$ be a convex function. A vector $g \in \mathbb{R}^m$ is called subgradient of $f$ at $y_0$ if for all $y \in \mathbb{R}^m$*

$$f(y) \geq f(y_0) + g'(y - y_0). \tag{3.16}$$

*The set of all subgradients of $f$ at $y_0$ is denoted by $\partial f(y_0)$. $\partial f(y_0)$ is called the subdifferential of $f$ at $y_0$. We say that $f$ is subdifferentiable everywhere if $\partial f(y) \neq \emptyset$ for all $y \in \mathbb{R}^m$.*

If $y$ is an interior point of the domain of a real-valued convex function $f$, then $f$ is subdifferentiable at $y$ (see Zeidler (1984) (Theorem 47.A)). This means that any real-valued convex function $f$ with domain $\mathbb{R}^m$ is subdifferentiable everywhere. Clearly, any convex function that is differentiable everywhere is also subdifferentiable everywhere and the subdifferential at each point is just a singleton. An example of a function that is subdifferentiable everywhere is given below.

**Example 3.4.** Consider the convex function $f : y \to |y|$, $y \in \mathbb{R}$. For all $y \neq 0$, $f(y)$ is differentiable and the derivative is $-1$ if $y < 0$ and $1$ if $y > 0$. At $y = 0$, the function $f$ is not differentiable. However, any $g \in \mathbb{R}$ that satisfies

$$|y| \geq gy \quad \text{for all } y \in \mathbb{R} \tag{3.17}$$

is a subgradient of $f$ at $y = 0$. It is easy to see that $g$ is a subgradient of the function $f$ at $y = 0$ if and only if $g \in [-1, 1]$. Thus the function $f$ is subdifferentiable everywhere and the subdifferential $\partial f(y)$ is defined as follows:

$$\partial f(y) = \begin{cases} \{-1\} & \text{if } y < 0 \\ [-1, 1] & \text{if } y = 0 \\ \{1\} & \text{if } y > 0. \end{cases} \tag{3.18}$$

The next result directly follows from the definition of a subdifferential.

**Lemma 3.5.** *Let $m \geq 1$ and let $f : \mathbb{R}^m \to \mathbb{R}$ be a convex function. Suppose that there exists at least one minimizer of $f$. Then $y_0$ is a minimizer of $f$ if and only if $0 \in \partial f(y_0)$*

*Proof.* Suppose $0 \in \partial f(y_0)$, then it follows that

$$f(y) \geq f(y_0) \quad \text{for all } y \in \mathbb{R}^n. \tag{3.19}$$

Thus $y_0$ is a minimizer of $f$. On the other hand, if $y_0$ is a minimizer, then the previous inequality always holds by the definition of a minimizer and therefore $0 \in f(y_0)$ follows. $\square$

The condition $0 \in \partial f(y_0)$ is also referred to as the Karush-Kuhn-Tucker (KKT) conditions. The next result shows that the sum rule for differentiable real-valued functions can also be extended to subdifferential functions.

**Lemma 3.6.** *Let $m \geq 1$ and let $f_1, f_2 : \mathbb{R}^m \to \mathbb{R}$ be two convex functions. Let $\kappa_1, \kappa_2 \in \mathbb{R}$ be two constants. Then $g \in \partial(\kappa_1 f_1 + \kappa_2 f_2)(y)$ if and only if $g = \kappa_1 g_1 + \kappa_2 g_2$ for a $g_1 \in \partial f_1(y)$ and a $g_2 \in \partial f_2(y)$.*

From the definition of subdifferentials it, immediately, follows that $\kappa_1 g_1 + \kappa_2 g_2$ is a subgradient of $(\kappa_1 f_1 + \kappa_2 f_2)(y)$, yet the opposite direction is more difficult to proof. We refer to Zeidler (1984) (Theorem 47.B) for a proof of the lemma.

Now, consider again the lasso estimator $\hat{\beta}(y)$, which is the minimizer of the function $L(\beta)$, where $L(\beta)$ is defined as follows:

$$L(\beta) = \frac{1}{2}\|y - W\beta\|_2^2 + \lambda\|\beta\|_1. \tag{3.20}$$

Both the squared $l_2$-norm and the $l_1$-norm are convex functions. Therefore, it follows that $L(\beta)$ is convex. Additionally, the squared $l_2$-norm is differentiable everywhere (and therefore subdifferentiable everywhere) and the $l_1$-norm is subdifferentiable everywhere. Since the lasso estimator $\hat{\beta}(y)$ is a minimizer of $L(\beta)$, Lemma 3.5 implies that $0 \in \partial L(\hat{\beta}(y))$. Lemma 3.6 implies that for every $y \in \mathbb{R}^n$ there exists a subgradient of $\|\beta\|_1$ at the point $\hat{\beta}(y)$, denoted by $\hat{s}(y)$, such that the KKT conditions $0 \in \partial L(\hat{\beta}(y))$ can be written as

$$W'(W\hat{\beta}(y) - y) + \lambda \hat{s}(y) = 0 \quad \text{for all } y \in \mathbb{R}^n. \tag{3.21}$$

Note that the function $\hat{s}(y)$ is well-defined, since $\hat{\beta}(y)$ is unique for all $y \in \mathbb{R}^n$ and therefore $\hat{s}(y)$ must be the unique subgradient of $\|\beta\|_1$ at the point $\hat{\beta}(y)$ that satisfies the previous equation.[1] The condition $\hat{s}(y) \in \partial\|\hat{\beta}(y)\|_1$ implies that for each $j = 1, \dots, p$, the $j$th element $\hat{s}_j(y)$ must satisfy

$$\hat{s}_j(y) \in \begin{cases} \{\text{sign}(\hat{\beta}_j(y))\} & \text{if } \hat{\beta}_j(y) \neq 0 \\ [-1, 1] & \text{if } \hat{\beta}_j(y) = 0. \end{cases} \tag{3.22}$$

Before we state the following theorem, we introduce further notation. For all $M \in \mathcal{M} \setminus \emptyset$ such that $W_M$ has full rank, let denote by $P_M$ the projection matrix of $W_M$, that is,

$$P_M = W_M(W_M' W_M)^{-1} W_M'. \tag{3.23}$$

For any scalar $c$ and any vector $v$, the expressions $v + c$, $v < c$ and $v \leq c$ are to be interpreted component-wise. Further, for any two vectors $v_1$ and $v_2$ of the same dimension, the inequalities $v_1 < v_2$ and $v_1 \leq v_2$ are to be interpreted component-wise.

**Theorem 3.7.** *For every $M \in \mathcal{M} \setminus \{\emptyset, M_F\}$ such that $W_M$ has full rank and for every $s_M \in \mathcal{S}_M$, let $A^0_{M,s_M}$ and $A^1_{M,s_M}$ be defined as*

$$A^0_{M,s_M} = \frac{1}{\lambda} \begin{pmatrix} W'_{M^c}(I_n - P_M) \\ -W'_{M^c}(I_n - P_M) \end{pmatrix} \tag{3.24}$$

$$A^1_{M,s_M} = -diag(s_M)(W_M' W_M)^{-1} W_M', \tag{3.25}$$

*and let $b^0_{M,s_M}$ and $b^1_{M,s_M}$ be defined as*

$$b^0_{M,s_M} = \begin{pmatrix} 1 - W'_{M^c} W_M (W_M' W_M)^{-1} s_M \\ 1 + W'_{M^c} W_M (W_M' W_M)^{-1} s_M \end{pmatrix} \tag{3.26}$$

$$b^1_{M,s_M} = -\lambda\, diag(s_M)(W_M' W_M)^{-1} s_M. \tag{3.27}$$

[1] In fact, the subgradient $\hat{s}(y)$ is well-defined even if the solution of the lasso is not unique. See Lemma 1 in Tibshirani (2013) and the subsequent discussion.

*For $M = \emptyset$ and $M = M_F$ (and $W = W_{M_F}$ has full rank), see Remark 3.9 and 3.10 for the definition of $A^0_{M,s_M}$, $A^1_{M,s_M}$, $b^0_{M,s_M}$ and $b^1_{M,s_M}$. Then, for all $M \in \mathcal{M}$ such that $W_M$ has full rank and for every $s_M \in \mathcal{S}_M$ it follows that*

$$\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\} = \left\{ \begin{array}{ccc} A^0_{M,s_M} y & \leq & b^0_{M,s_M} \\ A^1_{M,s_M} y & < & b^1_{M,s_M} \end{array} \right\}. \tag{3.28}$$

**Remark 3.8.** The inequality involving $A^0_{M,s_M}$ and $b^0_{M,s_M}$ is also referred to as the inactive constraint as the inequality is deduced from the constraints of the zero coefficients of $\hat{\beta}(y)$. On the other hand, the inequality involving $A^1_{M,s_M}$ and $b^1_{M,s_M}$ is also referred to as the active constraint as the inequality is deduced from the constraints of the non-zero coefficients of $\hat{\beta}(y)$.

**Remark 3.9.** The active constraint is trivial for the characterization of the set $\{\hat{M} = \emptyset, \hat{s}_{\hat{M}} = 0\}$, since for any $y$ in this set we have that $\hat{\beta}(y) = 0$. Therefore, $A^1_{\emptyset,0}$ and $b^1_{\emptyset,0}$ are to be interpreted as follows:

$$A^1_{\emptyset,0} = (0, \ldots, 0) \quad \text{and} \quad b^1_{\emptyset,0} = 1, \tag{3.29}$$

such that $A^1_{\emptyset,0} y < b^1_{\emptyset,0}$ is trivially satisfied for all $y \in \mathbb{R}^n$. The inactive constraint is in this case to be interpreted as follows:

$$A^0_{\emptyset,0} = \frac{1}{\lambda} \begin{pmatrix} W' \\ -W' \end{pmatrix} \quad \text{and} \quad b^0_{\emptyset,0} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{3.30}$$

**Remark 3.10.** If $W_{M_F} = W$ has full rank, then the inactive constraint is trivial for the characterization of the set $\{\hat{M} = M_F, \hat{s}_{\hat{M}} = s_{M_F}\}$ for any $s_{M_F} \in \mathcal{S}_{M_F}$, as for all $y$ in this set, we have that all coefficients of $\hat{\beta}(y)$ are non-zero. Therefore, $A^0_{M_F,s_{M_F}}$ and $b^0_{M_F,s_{M_F}}$ are to be interpreted as follows:

$$A^0_{M_F,s_{M_F}} = (0, \ldots, 0) \quad \text{and} \quad b^0_{M_F,s_{M_F}} = 1 \quad \text{for all } s_{M_F} \in \mathcal{S}_{M_F}, \tag{3.31}$$

such that $A^0_{M_F,s_{M_F}} y \leq b^0_{M_F,s_{M_F}}$ is trivially satisfied for all $y \in \mathbb{R}^n$. The matrix $A^1_{M,s_M}$ and the vector $b^1_{M,s_M}$ are defined as in (3.25) and (3.27) for $M = M_F$ and for $s_M = s_{M_F}$.

*Proof of Theorem 3.7 (as in Lee et al. (2016)).* The KKT conditions are necessary and sufficient; therefore, it follows that $y \in \{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ if and only if $y$ satisfies the conditions (3.21) and (3.22) and also $\text{sign}(\hat{\beta}_M(y)) = s_M$ and $\hat{\beta}_j(y) = 0$ for all $j \notin M$. Thus, to characterize $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$, it is equivalent to characterize the following set:

$$\{y \in \mathbb{R}^n : y \text{ satisfies (3.21) and (3.22)}, \text{sign}(\hat{\beta}_M(y)) = s_M, \hat{\beta}_j(y) = 0 \text{ for all } j \notin M\}. \tag{3.32}$$

There are 3 cases that need to be considered.

First case: $M = \emptyset$. All $y$ in the set (3.32) satisfy $\hat{\beta}(y) = 0$ and, by convention, $\text{sign}(\hat{\beta}_\emptyset(y)) = s_\emptyset = 0$. Substituting $\hat{\beta}(y) = 0$ into condition (3.21) yields

$$-W'y + \lambda\hat{s}(y) = 0. \tag{3.33}$$

Substituting $\hat{\beta}(y) = 0$ into condition (3.22) yields:

$$\hat{s}(y) \in [-1, 1]^p. \tag{3.34}$$

The set of all $y \in \mathbb{R}$ that satisfy the previous two conditions can be written as follows:

$$\{-1 \leq \frac{1}{\lambda}W'y \leq 1\} = \left\{\frac{1}{\lambda}\begin{pmatrix} W' \\ -W' \end{pmatrix} y \leq \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\} = \{A^0_{\emptyset,0}y \leq b^0_{\emptyset,0}\} \tag{3.35}$$

Clearly, the set (3.32) is a subset of the set (3.35). On the other hand, fix an arbitrary $y$ in the set (3.35). Set $\hat{s}(y) = \frac{1}{\lambda}W'y \in [-1, 1]^p$ and $\hat{\beta}(y) = 0$. Then, $\hat{s}(y)$ and $\hat{\beta}(y)$ satisfy the KKT-conditions in conditions (3.21) and (3.22). Therefore $\hat{\beta}(y)$ is the unique solution and it follows that $y$ is an element of the set (3.32).

Second case: $M \in \mathcal{M} \setminus \{\emptyset, M_F\}$ such that $W_M$ has full rank. Fix an arbitrary $s_M \in \mathcal{S}_M$. All $y$ in the set (3.32) satisfy $\text{sign}(\hat{\beta}_M(y)) = s_M$ and $\hat{\beta}_j(y) = 0$ for all $j \notin M$. This implies that the condition (3.21) can be written as follows:

$$W'_M(W_M\hat{\beta}_M(y) - y) + \lambda s_M = 0 \tag{3.36}$$

$$W'_{M^c}(W_M\hat{\beta}_M(y) - y) + \lambda\hat{s}_{M^c}(y) = 0. \tag{3.37}$$

The condition (3.22) can be written as follows:

$$\operatorname{sign}(\hat{\beta}_M(y)) = s_M \tag{3.38}$$

$$\hat{s}_{M^c}(y) \in [-1,1]^{p-|M|}. \tag{3.39}$$

Note that (3.36) and (3.38) correspond to the active constraint while (3.37) and (3.39) correspond to the inactive constraint. An equivalent formulation of condition (3.36) is obtained by solving for $\hat{\beta}_M(y)$:

$$\hat{\beta}_M(y) = (W_M' W_M)^{-1}(W_M' y - \lambda s_M). \tag{3.40}$$

As, by assumption, $W_M$ has full rank, the matrix $W_M' W_M$ is invertible. Solving equation (3.37) for $\hat{s}_{M^c}(y)$ and then substituting equation (3.40) into it yields

$$\begin{aligned}\hat{s}_{M^c}(y) &= -\frac{1}{\lambda} W_{M^c}'(W_M \hat{\beta}_M(y) - y) \\ &= -\frac{1}{\lambda} W_{M^c}' \left( W_M (W_M' W_M)^{-1} W_M' y - \lambda W_M (W_M' W_M)^{-1} s_M - y \right) \\ &= -\frac{1}{\lambda} W_{M^c}' \left( -\lambda W_M (W_M' W_M)^{-1} s_M - (I_n - P_M) y \right) \\ &= W_{M^c}' W_M (W_M' W_M)^{-1} s_M + \frac{1}{\lambda} W_{M^c}' (I_n - P_M) y.\end{aligned} \tag{3.41}$$

The condition (3.38) is equivalent to the condition that the product of the $j$th element of $\hat{\beta}_M(y)$ and the $j$th element of $s_M$ is strictly positive for all $j = 1, \ldots, |M|$. Therefore, by (3.38) and (3.40), the set of all $y \in \mathbb{R}^n$ that satisfies the active constraints can be written as follows:

$$\begin{aligned}\{\operatorname{sign}(\hat{\beta}_M(y)) = s_M\} &= \{\operatorname{diag}(s_M)\hat{\beta}_M(y) > 0\} \\ &= \{\operatorname{diag}(s_M)(W_M' W_M)^{-1}(W_M' y - \lambda s_M) > 0\} \\ &= \{-\operatorname{diag}(s_M)(W_M' W_M)^{-1} W_M' y < -\lambda \operatorname{diag}(s_M)(W_M' W_M)^{-1} s_M\} \\ &= \{A^1_{M,s_M} y < b^1_{M,s_M}\}.\end{aligned} \tag{3.42}$$

By (3.39) and (3.41), the set of all $y \in \mathbb{R}^n$ that satisfies the inactive constraints can be formulated as follows:

$$
\begin{aligned}
&\{\hat{s}_{M^c}(y) \in [-1,1]^{p-|M|}\} = \{-1 \le \hat{s}_{M^c}(y) \le 1\} \\
&\quad = \{-1 \le W'_{M^c} W_M (W'_M W_M)^{-1} s_M + \frac{1}{\lambda} W'_{M^c}(I_n - P_M) y \le 1\} \\
&\quad = \{-1 - W'_{M^c} W_M (W'_M W_M)^{-1} s_M \le \frac{1}{\lambda} W'_{M^c}(I_n - P_M) y \le 1 - W'_{M^c} W_M (W'_M W_M)^{-1} s_M\} \\
&\quad = \left\{ \frac{1}{\lambda} \begin{pmatrix} W'_{M^c}(I_n - P_M) \\ -W'_{M^c}(I_n - P_M) \end{pmatrix} y \le \begin{pmatrix} 1 - W'_{M^c} W_M (W'_M W_M)^{-1} s_M \\ 1 + W'_{M^c} W_M (W'_M W_M)^{-1} s_M \end{pmatrix} \right\} \\
&\quad = \{A^0_{M,s_M} y \le b^0_{M,s_M}\}.
\end{aligned}
\tag{3.43}
$$

It is clear that the set (3.32) is a subset of

$$
\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\} = \left\{ \begin{array}{ccc} A^0_{M,s_M} y & \le & b^0_{M,s_M} \\ A^1_{M,s_M} y & < & b^1_{M,s_M} \end{array} \right\}. \tag{3.44}
$$

On the other hand, fix an arbitrary $y$ in the previous set. Let $\hat{s}(y)$ be defined by $\hat{s}_M(y) = s_M$ and $\hat{s}_{M^c}(y)$ as in equation (3.41). Let $\hat{\beta}(y)$ be defined by $\hat{\beta}_M(y)$ as in equation (3.40) and $\hat{\beta}_{M^c}(y) = 0$. Then, $\hat{s}(y)$ and $\hat{\beta}(y) = 0$ satisfy the KKT-conditions in conditions (3.21) and (3.22). Therefore $\hat{\beta}(y)$ is the unique solution and it follows that $y$ is an element of the set (3.32).

Third case: $M = M_F$ and $W_{M_F} = W$ has full rank. This case is almost identical to the second case, except that the inactive constraint is trivial since $M_F^c$ is empty.

□

The theorem in particular holds for all models $M \in \mathcal{M}^+$ as by Lemma 3.2 the matrix $W_M$ has full rank for all $M \in \mathcal{M}^+$. However, the theorem is not necessarily applicable to all models $M \in \mathcal{M} \setminus \mathcal{M}^+$, as in this case $W_M$ may not have full rank (not even if $|M| \le \min(n,p)$). The theorem implies that all sets $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ that have positive selection probability are polyhedral sets. The following corollary to Theorem 3.7 shows that the selection probability of the empty set is always positive and therefore the set $\{\hat{M} = \emptyset\}$ is always a polyhedron.

**Corollary 3.11.** *For every regression matrix $W$ with columns in general position, the empty set is selected with positive probability, that is, $\emptyset \in \mathcal{M}^+$.*

*Proof.* By equivalence of the measure $\mathbb{P}_{\mu,\sigma}$ and the Lebesgue measure, it is sufficient to show that $\{\hat{M} = \emptyset\}$ has positive Lebesgue measure. From the previous theorem, it follows that

$$\{\hat{M} = \emptyset\} = \{\hat{M} = \emptyset, \hat{s}_{\hat{M}} = 0\} = \{-1 \leq -\frac{1}{\lambda}W'y \leq 1\}. \tag{3.45}$$

Let denote by $W_{i,j}$ the element of the $i$th row and the $j$th column of the matrix $W$. Set $C = \max_{i,j}|n\frac{1}{\lambda}W_{i,j}|$. Note that $C > 0$. By choice of the constant $C$, it follows that the set $\{-\frac{1}{C} < y < \frac{1}{C}\}$ is a subset of the set $\{-1 \leq -\frac{1}{\lambda}W'y \leq 1\}$ . The set $\{-\frac{1}{C} < y < \frac{1}{C}\}$ has positive Lebesgue measure; hence, $\{-1 \leq -\frac{1}{\lambda}W'y \leq 1\}$ also has positive Lebesgue measure. Therefore, it follows that

$$\mathbb{P}_{\mu,\sigma^2}(\hat{M} = \emptyset) > 0 \quad \text{for all } \mu \in (-\infty, +\infty),\ \sigma^2 \in (0, +\infty). \tag{3.46}$$

This implies that $\emptyset \in \mathcal{M}^+$. $\square$

For any model $M \in \mathcal{M}^+$, the selection set $\{\hat{M} = M\}$ is simply the union of the polyhedral sets over all sign vectors $s_M \in \mathcal{S}_M$.

**Corollary 3.12.** *Let $A^0_{M,s_M}$, $A^1_{M,s_M}$, $b^0_{M,s_M}$ and $b^1_{M,s_M}$ be defined as before. Then for all $M \in \mathcal{M}^+$*

$$\{\hat{M} = M\} = \bigcup_{s_M \in \mathcal{S}_M} \left\{ \begin{array}{ccc} A^0_{M,s_M}y & \leq & b^0_{M,s_M} \\ A^1_{M,s_M}y & < & b^1_{M,s_M} \end{array} \right\}. \tag{3.47}$$

**Remark 3.13.** In Lee et al. (2016), the function $\hat{M}(y)$ is defined as the equicorrelation set (Tibshirani (2013)), that is, $\hat{M}(y) = \{j : |\hat{s}_j(y)| = 1\}$. Proposition 4.2 in Lee et al. (2016) states that the set $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ is an open polyhedron (i.e. no faces are included); however, the result is not exactly correct. In the definition of $\hat{s}(y)$ in (3.22), we see that it is possible that $\hat{\beta}_j(y) = 0$ but $|\hat{s}_j(y)| = 1$ for a $y \in \mathbb{R}^n$. One can easily provide an example where this is actually the case. Thus for such a $y$, it follows that $j \in \hat{M}(y)$ but $\hat{\beta}_j(y) = 0$. However, in the proof of Proposition 4.2, it is assumed that $\text{sign}(\hat{\beta}_M(y)) = s_M$ holds for all $y \in \{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$. Fortunately, it follows by the next lemma that the set of $y \in \{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ such that $\text{sign}(\hat{\beta}_M(y)) \neq s_M$ is a null set with respect to the measure $\mathbb{P}_{\mu,\sigma}$ and therefore all the subsequent results on the conditional distributions in Lee et al. (2016) are still valid for all $M \in \mathcal{M}^+$ and all $s_M \in \mathcal{S}_M^+$.

By Theorem 3.7, for all $M \in \mathcal{M}^+$, the set $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ is a polyhedron where some faces are included and others are not. The faces included in the set $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ are expressed by the set $\{A^0_{M,s_M} y \leq b^0_{M,s_M}\}$. In the next lemma, we show that these faces are null sets with respect to the measure $\mathbb{P}_{\mu,\sigma}$. Note that for all $M \in \mathcal{M}^+ \setminus M_F$, $A^0_{M,s_M}$ is a $(2|M^c| \times n)$-dimensional matrix and $b^0_{M,s_M}$ is a $(2|M^c|)$-dimensional vector (see the Definition in (3.24) and (3.26) for all $M \in \mathcal{M}^+ \setminus \{\emptyset, M_F\}$ and in (3.30) for $M = \emptyset$). For $M_F \in \mathcal{M}^+$, $A^0_{M,s_M}$ is a $(1 \times n)$-dimensional matrix and $b^0_{M,s_M}$ is a scalar (see Definition in (3.31)). The $j$th row of $A^0_{M,s_M}$ is denoted by $(A^0_{M,s_M})_j$ and the $j$th element of $b^0_{M,s_M}$ is denoted by $(b^0_{M,s_M})_j$.

**Lemma 3.14.** *For all $M \in \mathcal{M}^+$ and all $s_M \in \mathcal{S}_M$, the faces of $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ are $\mathbb{P}_{\mu,\sigma}$-null sets. That is, the set*

$$\left\{(A^0_{M,s_M})_j y = (b^0_{M,s_M})_j\right\} \tag{3.48}$$

*is a $\mathbb{P}_{\mu,\sigma}$-null set for all $j = 1, \ldots, 2|M^c|$ (if $M \in \mathcal{M}^+ \setminus M_F$) or for $j = 1$ (if $M = M_F$).*

*Proof.* By equivalence of the measure $\mathbb{P}_{\mu,\sigma}$ and the Lebesgue measure, it is again sufficient to show that the set $\left\{(A^0_{M,s_M})_j y = (b^0_{M,s_M})_j\right\}$ is a null set with respect to Lebesgue measure. There are 3 cases that need to be considered.

First case: $M = M_F$. This case is trivial, as by definition in (3.31), $A^0_{M_F,s_{M_F}}$ is to be interpreted as $(0, \ldots, 0)$ and $b^0_{M_F,s_{M_F}}$ is to be interpreted as 1.

Before we continue with the remaining 2 cases, we note that for all $M \in \mathcal{M}^+ \setminus M_F$, the matrix $A^0_{M,s_M}$ satisfies the following symmetry property: $(A^0_{M,s_M})_j = -(A^0_{M,s_M})_{j+|M^c|}$ for all $j = 1, \ldots, |M^c|$ (see the definition in (3.24) and for all $M \in \mathcal{M}^+ \setminus \{\emptyset, M_F\}$ and in (3.30) for $M = \emptyset$). Therefore, it will be sufficient sufficient to consider only the first $|M^c|$ rows in the remaining 2 cases.

Second case: $M = \emptyset$. For every $j \in \{1, \ldots, |M^c|\}$, we have that $(A^0_{\emptyset,0})_j = w_j'$ and $(b^0_{\emptyset,0})_j = 1$. We assumed that $w_j \neq 0$. Hence, the set $\{(A^0_{\emptyset,0})_j y = (b^0_{\emptyset,0})_j\}$ is a hyperplane of $\mathbb{R}^n$ and, therefore, a Lebesgue null set.

Third case: $M \in \mathcal{M}^+ \setminus \{\emptyset, M_F\}$. Fix a $j \in \{1, \ldots, |M^c|\}$. If $(A^0_{M,s_M})_j \neq 0$, then the set $\{(A^0_{M,s_M})_j y = (b^0_{M,s_M})_j\}$ is again a hyperplane of $\mathbb{R}^n$ and therefore a Lebesgue null set. Now suppose that $(A^0_{M,s_M})_j = 0$. From the definition of $A^0_{M,s_M}$ in (3.24), it follows that $(A^0_{M,s_M})_j = 0$ if and only if the $j$th column of $W_{M^c}$, denoted by $W_{M^c,j}$ is in the column space of $W_M$. Set $M = \{i_1, \ldots, i_k\}$, where $k = |M|$. Then, $W_{M^c,j}$ must be of the following form:

$$W_{M^c,j} = \gamma_{0,1} w_{i_1} + \cdots + \gamma_{0,k} w_{i_k} \quad \text{for a } \gamma_0 \in \mathbb{R}^k. \tag{3.49}$$

Let denote by $s_{M,j}$ the $j$th element of $s_M$. Recall that the columns of $W$ are in general position (see Definition 3.1). Therefore, for all $s_M \in \mathcal{S}_M$, the following must be satisfied:

$$W_{M^c,j} \neq \gamma_1 s_{M,1} w_{i_1} + \cdots + \gamma_k s_{M,k} w_k \quad \text{for all } \gamma \in \mathbb{R}^k \text{ such that } \sum_{i=1}^{k} \gamma_i = 1, \tag{3.50}$$

$$-W_{M^c,j} \neq \gamma_1 s_{M,1} w_{i_1} + \cdots + \gamma_k s_{M,k} w_k \quad \text{for all } \gamma \in \mathbb{R}^k \text{ such that } \sum_{i=1}^{k} \gamma_i = 1. \tag{3.51}$$

Set $\gamma = (\gamma_{0,1} s_{M,1}, \ldots, \gamma_{0,k} s_{M,k})' \in \mathbb{R}^k$ and suppose that $\sum_{i=1}^{k} \gamma_i = \gamma_0' s_M = 1$. Then it follows that

$$\begin{aligned} \gamma_1 s_{M,1} w_{i_1} + \cdots + \gamma_k s_{M,k} w_k &= \gamma_{0,1} s_{M,1}^2 w_{i_1} + \cdots + \gamma_{0,k} s_{M,k}^2 w_k \\ &= \gamma_{0,1} w_{i_1} + \cdots + \gamma_{0,k} w_k \\ &= W_{M^c,j}, \end{aligned} \tag{3.52}$$

where the first equality follows by the definition of $\gamma$, the second equality follows from the trivial fact that $s_{M,i}^2 = 1$ for all $i = 1, \ldots, k$ and all $s_M \in \mathcal{S}_M$ and the third equality follows from equation (3.49). This obviously contradicts equation (3.50), thus, it follows that $\gamma_0' s_M \neq 1$. A similar argument (set $\gamma = -(\gamma_{0,1} s_{M,1}, \ldots, \gamma_{0,k} s_{M,k})'$ and assume that $\sum_{i=1}^{k} \gamma_i = 1$ holds) implies that also $\gamma_0' s_M \neq -1$ must hold. Therefore, it follows that the $j$th component of $b^0_{M,s_M}$ (see Definition of $b^0_{M,s_M}$ in (3.26)) satisfies

$$\begin{aligned} (b^0_{M,s_M})_j &= 1 + W'_{M^c,j} W_M (W'_M W_M)^{-1} s_M \\ &= 1 + \gamma_0' W'_M W_M (W'_M W_M)^{-1} s_M \\ &= 1 + \gamma_0' s_M \neq 0. \end{aligned} \tag{3.53}$$

Hence, the set $\{(A^0_{M,s_M})_j y = (b^0_{M,s_M})_j\}$ is empty. $\square$

**Remark 3.15.** The set

$$\left\{ \begin{pmatrix} A^0_{M,s_M} \\ A^1_{M,s_M} \end{pmatrix} y < \begin{pmatrix} b^0_{M,s_M} \\ b^1_{M,s_M} \end{pmatrix} \right\} \tag{3.54}$$

coincides with the open polyhedron of Proposition 4.2 in Lee et al. (2016). For all $y$ in this set, it follows that if $\hat{\beta}_j(y) = 0$, then also $|\hat{s}_j(y)| < 1$ for all $j = 1, \ldots, p$. Since the previous lemma implies that the faces are null sets with respect to $\mathbb{P}_{\mu,\sigma}$ it follows that the probability of the set

$$\bigcup_{M \in \mathcal{M}^+} \bigcup_{s_M \in \mathcal{S}^+_M} \left\{ \begin{pmatrix} A^0_{M,s_M} \\ A^1_{M,s_M} \end{pmatrix} y < \begin{pmatrix} b^0_{M,s_M} \\ b^1_{M,s_M} \end{pmatrix} \right\} \tag{3.55}$$

is 1. Hence, it follows that set $\{y \in \mathbb{R}^n : \hat{\beta}_j(y) = 0 \text{ and } |\hat{s}_j(y)| = 1 \text{ for a } j = 1, \ldots, p\}$ is also a null set with respect to $\mathbb{P}_{\mu,\sigma}$.

We conclude this section with an illustration of the sets $\{\hat{M} = M\}$ and $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ for a particular regression matrix $W$.

**Example 3.16.** Consider the following regression matrix $W$:

$$W = \begin{pmatrix} 1 & 0 & 2 & 0.1 \\ 0 & 1 & 2 & 0 \end{pmatrix}. \tag{3.56}$$

This means that $n = 2$ and $p = 4$ and therefore the number of regressors is larger than the number of observations. The tuning parameter $\lambda$ is set to 1. The columns of $W$ are in general position. The lasso estimator $\hat{\beta}(y)$ is, therefore, unique for all $y \in \mathbb{R}^2$. By Lemma 3.2, the models $\{1,2,3\}$, $\{1,2,4\}$, $\{1,3,4\}$, $\{2,3,4\}$ and $\{1,2,3,4\}$ are never selected. The sets $\{\hat{M} = M\}$ and $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ are illustrated in Figure 3.1. The signs of the non-zero lasso estimates are denoted by a $+$ and a $-$. We see that all sets $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ which have positive selection probability are of polyhedral form. The empty model is a single polyhedron located at the origin. The other sets $\{\hat{M} = M\}$ which have positive selection probabilities are a union of polyhedra. For the models $\{1\}$, $\{2\}$ and $\{3\}$, both a positive and a negative sign is possible. For the model $\{1,2\}$ only the sign patterns $\{-,+\}$ and $\{+,-\}$ are possible while for the models $\{1,3\}$ and $\{2,3\}$ only the sign patterns $\{-,-\}$ and $\{+,+\}$ are possible. None of the selected models include the fourth variable, because $w_4$ is a scalar multiple of $w_1$, where the scalar is less than 1. Due to the penalty constraint on the size of the regression coefficients, the lasso therefore always prefers $w_1$.

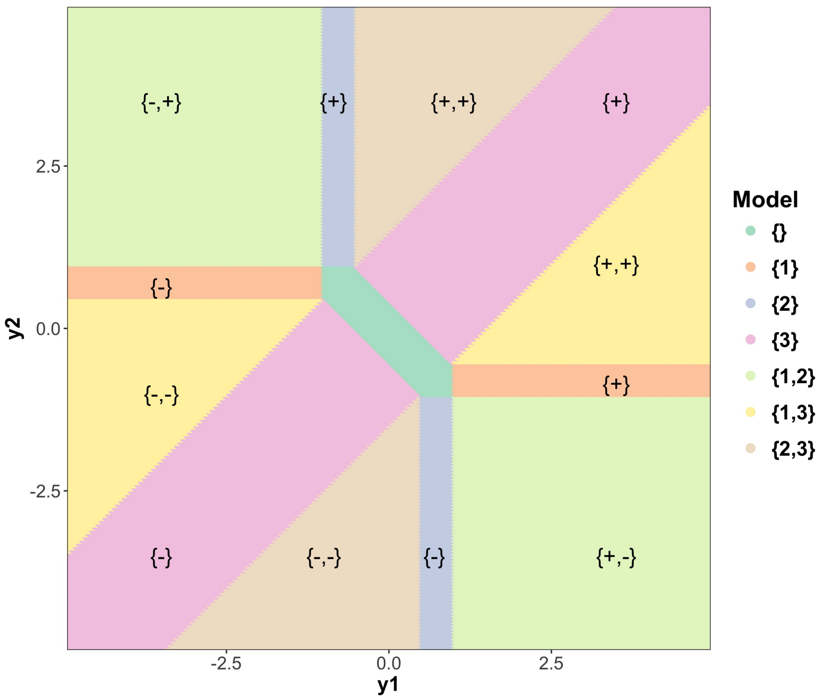


Figure 3.1: The selection sets $\{\hat{M} = M\}$ and $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ are shown for the regression matrix $W$ in (3.56). The signs of the non-zero regression coefficients of $\hat{\beta}(y)$ are denoted by a $+$ and a $-$. Note that not every sign pattern is possible for each model $M$. Also, the fourth variable is never selected.

## 3.3 Confidence intervals of Lee et al. (2016)

For all $M \in \mathcal{M}^+$ and all $s_M \in \mathcal{S}_M$, the matrix $A_{M,s_M}$ and the vector $b_{M,s_M}$ are defined as follows:

$$A_{M,s_M} = \begin{pmatrix} A^0_{M,s_M} \\ A^1_{M,s_M} \end{pmatrix} \quad \text{and} \quad b_{M,s_M} = \begin{pmatrix} b^0_{M,s_M} \\ b^1_{M,s_M} \end{pmatrix} \quad \text{for all } M \in \mathcal{M}^+,\ s_M \in \mathcal{S}_M. \tag{3.57}$$

Lemma 3.14 implies that the faces of the polyhedron $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ are null sets with respect to $\mathbb{P}_{\mu,\sigma^2}$. Note that the sets $\{\hat{M} = M\}$ and $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ have positive probability for all $M \in \mathcal{M}^+$ and all $s_M \in \mathcal{S}_M^+$. Therefore, it follows that

$$Y|\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\} \stackrel{d}{=} Y|\{A_{M,s_M} Y < b_{M,s_M}\} \quad \text{for all } M \in \mathcal{M}^+,\ s_M \in \mathcal{S}_M^+ \tag{3.58}$$

and

$$Y|\{\hat{M} = M\} \stackrel{d}{=} Y|\bigcup_{s_M \in \mathcal{S}_M} \{A_{M,s_M} Y < b_{M,s_M}\} \quad \text{for all } M \in \mathcal{M}^+. \tag{3.59}$$

Thus to study the distributions in (3.14) and (3.15), it is required to understand the distribution of

$$\eta' Y \mid \{AY < b\}, \tag{3.60}$$

where $\eta$ is the $n$-dimensional vector of interest, $A$ is an arbitrary matrix, $b$ an arbitrary vector and the set $\{Ay < b\}$ has positive probability.

**Lemma 3.17.** *Set $Z(y) = (I_n - c\eta')y$ with $c = \frac{\eta}{\eta'\eta}$ and let the functions $\mathcal{V}^-(z)$, $\mathcal{V}^+(z)$ and $\mathcal{V}^0(z)$ (all from $\mathbb{R}^n \to \mathbb{R} \cup \{-\infty, +\infty\}$) be defined as follows:*

$$\mathcal{V}^-(z) = \begin{cases} \max\limits_{i:(Ac)_i<0} \frac{b_i-(Az)_i}{(Ac)_i} & \textit{if there is a } i : (Ac)_i < 0 \\ -\infty & \textit{else} \end{cases} \tag{3.61}$$

$$\mathcal{V}^+(z) = \begin{cases} \min\limits_{i:(Ac)_i>0} \frac{b_i-(Az)_i}{(Ac)_i} & \textit{if there is a } i : (Ac)_i > 0 \\ +\infty & \textit{else} \end{cases} \tag{3.62}$$

$$\mathcal{V}^0(z) = \begin{cases} \min\limits_{i:(Ac)_i=0} b_i - (Az)_i & \textit{if there is a } i : (Ac)_i = 0 \\ 1 & \textit{else.} \end{cases} \tag{3.63}$$

*Then, it follows that*

$$\{Ay < b\} = \{\mathcal{V}^-(Z(y)) < \eta' y < \mathcal{V}^+(Z(y)), \mathcal{V}^0(Z(y)) > 0\}. \tag{3.64}$$

*Further, $Z(Y)$ is independent of $\eta' Y$.*

*Proof as in Lee et al. (2016).* The expected value of $Y$ conditional on $\eta'Y$ is given by

$$\mathbb{E}(Y|\eta'Y) = \mu + \frac{\eta}{\eta'\eta}(\eta'Y - \eta'\mu). \tag{3.65}$$

Note that $\mathbb{E}(Y|\eta'Y)$ is linear in $Y$. $Y - \mathbb{E}(Y|\eta'Y)$ is uncorrelated with $\eta'Y$ and, since $Y$ is normally distributed, it follows that $Y - \mathbb{E}(Y|\eta'Y)$ and $\eta'Y$ are also independent. Let $Z(y)$ be defined by

$$Z(y) = y - \frac{\eta}{\eta'\eta}\eta'y = (I_n - c\eta')y, \tag{3.66}$$

where $c = \frac{\eta}{\eta'\eta}$. Clearly, $Z(Y)$ is also independent of $\eta'Y$. An equivalent formulation of the set $\{Ay < b\}$ is given in the following equation:

$$\begin{aligned} \{Ay < b\} &= \{Ay - Ac\eta'y < b - Ac\eta'y\} = \{A(I_n - c\eta')y < b - Ac\eta'y\} \\ &= \{AZ(y) < b - Ac\eta'y\} = \{(Ac)_i\eta'y < b_i - (AZ(y))_i \text{ for all } i\}. \end{aligned} \tag{3.67}$$

The functions $\mathcal{V}^-(z)$, $\mathcal{V}^+(z)$ and $\mathcal{V}^0(z)$ in (3.61)-(3.63) are now obtained by a case distinction of the coefficients $(Ac)_i$. All negative coefficients determine $\mathcal{V}^-(z)$, all positive coefficients determine $\mathcal{V}^+(z)$ and all zero coefficients determine $\mathcal{V}^0(z)$. If all coefficient are non-negative, $\mathcal{V}^-(z)$ is set to $-\infty$; analogously, if all coefficients are non-positive, $\mathcal{V}^+(z)$ is set to $+\infty$. And if $(Ac)_i \neq 0$ for all $j$, then we set $\mathcal{V}^0(z)$ to 1. Therefore,

$$\{Ay < b\} = \{\mathcal{V}^-(Z(y)) < \eta'y < \mathcal{V}^+(Z(y)), \mathcal{V}^0(Z(y)) > 0\}, \tag{3.68}$$

as claimed. □

For every $y \in \{Ay < b\}$, it is clear that $\mathcal{V}^-(Z(y)) < \mathcal{V}^+(Z(y))$ and $\mathcal{V}^0(Z(y)) > 0$ must hold. However, for any $y \notin \{Ay < b\}$, there are several cases that need to be distinguished. Consider the inequality $(Ac)_i\eta'y < b_i - (AZ(y))_i$ from the previous proof for an arbitrary $i$. Suppose that the inequality is violated, that is $(Ac)_i\eta'y \geq b_i - (AZ(y))_i$ holds. If $(Ac)_i = 0$, then it follows that $\mathcal{V}^0(Z(y)) \leq 0$. If $(Ac)_i > 0$, then it follows that $\mathcal{V}^+(Z(y)) \leq \eta'y$. However, note that it does not necessarily follow that $\mathcal{V}^-(Z(y)) \geq \mathcal{V}^+(Z(y))$. And if $(Ac)_i < 0$, then it follows that $\mathcal{V}^-(Z(y)) \geq \eta'y$ and again it does not necessarily follow that $\mathcal{V}^-(Z(y)) \geq \mathcal{V}^+(Z(y))$.

Suppose that there is also a $j$ such that $(Ac)_j \eta' y \geq b_j - (AZ(y))_j$. If $(Ac)_i > 0$ and $(Ac)_j < 0$, then it also follows that $\mathcal{V}^-(Z(y)) \geq \mathcal{V}^+(Z(y))$.

In the proof of the previous lemma, we can see that the function $Z(y)$ only depends on the direction of interest $\eta$ while the functions $\mathcal{V}^-(z)$, $\mathcal{V}^+(z)$ and $\mathcal{V}^0(z)$ additionally depend on the matrix $A$ and the vector $b$. Thus to construct the confidence intervals, it is necessary to define the function $Z(y)$ for each $\eta_{M,j}$ and the functions $\mathcal{V}^-(z)$, $\mathcal{V}^+(z)$ and $\mathcal{V}^0(z)$ for each model $M$, each sign vector $s_M$ and each direction of interest $\eta_{M,j}$. Note that $\eta_{M,j}$ was only defined for all $M \in \mathcal{M}^+ \setminus \emptyset$, as inference is trivial in the case $M = \emptyset$. Now, for all $M \in \mathcal{M}^+ \setminus \emptyset$, all $s_M \in \mathcal{S}_M$ and all $j = \{1, \ldots, |M|\}$, the functions $Z_{M,j}(y)$, $\mathcal{V}^-_{M,s_M,j}(z)$, $\mathcal{V}^+_{M,s_M,j}(z)$ and $\mathcal{V}^0_{M,s_M,j}(z)$ are defined as follows:

$$Z_{M,j}(y) = (I_n - c_{M,j}\eta'_{M,j})y \quad \text{with} \quad c_{M,j} = \frac{\eta_{M,j}}{\eta'_{M,j}\eta_{M,j}} \tag{3.69}$$

and

$$\mathcal{V}^-_{M,s_M,j}(z) = \begin{cases} \max\limits_{i:(A_{M,s_M}c_{M,j})_i<0} \frac{(b_{M,s_M})_i - (A_{M,s_M}z)_i}{(A_{M,s_M}c_{M,j})_i} & \text{if there is a } i : (A_{M,s_M}c_{M,j})_i < 0 \\ -\infty & \text{else} \end{cases} \tag{3.70}$$

$$\mathcal{V}^+_{M,s_M,j}(z) = \begin{cases} \min\limits_{i:(A_{M,s_M}c_{M,j})_i>0} \frac{(b_{M,s_M})_i - (A_{M,s_M}z)_i}{(A_{M,s_M}c_{M,j})_i} & \text{if there is a } i : (A_{M,s_M}c_{M,j})_i > 0 \\ +\infty & \text{else} \end{cases} \tag{3.71}$$

$$\mathcal{V}^0_{M,s_M,j}(z) = \begin{cases} \min\limits_{i:(A_{M,s_M}c_{M,j})_i=0} (b_{M,s_M})_i - (A_{M,s_M}z)_i & \text{if there is a } i : (A_{M,s_M}c_{M,j})_i = 0 \\ 1 & \text{else.} \end{cases} \tag{3.72}$$

The previous lemma implies that the distribution of $\eta'_{M,j}Y|\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ is of the form

$$\eta'_{M,j}Y|\{\mathcal{V}^-_{M,s_M,j}(Z_{M,j}(Y)) < \eta'_{M,j}Y < \mathcal{V}^+_{M,s_M,j}(Z_{M,j}(Y)), \mathcal{V}^0_{M,s_M,j}(Z_{M,j}(Y)) > 0\} \tag{3.73}$$

and the distribution of $\eta'_{M,j}Y|\{\hat{M}=M\}$ can be written as:

$$\eta'_{M,j}Y| \bigcup_{s_M \in \mathcal{S}_M} \{\mathcal{V}^-_{M,s_M,j}(Z_{M,j}(Y)) < \eta'_{M,j}Y < \mathcal{V}^+_{M,s_M,j}(Z_{M,j}(Y)), \mathcal{V}^0_{M,s_M,j}(Z_{M,j}(Y)) > 0\}, \tag{3.74}$$

where $Z_{M,j}(Y)$ is independent of $\eta'_M Y$. These conditional distributions look similar to a normally distributed random variable that is truncated to a single interval and to a union of intervals, respectively. But the truncation intervals are random variables, instead of fixed quantities, and are independent of $\eta'_{M,j}Y$. However, in the next lemma, it will be shown that, by additionally conditioning on the observed value of $Z_{M,j}(Y)$, the conditional distribution of $\eta'_{M,j}Y$ is, in fact, a truncated normal distribution. Before we state the lemma, further notation is introduced. For all $M \in \mathcal{M}^+ \setminus \emptyset$, all $s_M \in \mathcal{S}_M$ and all $j \in \{1,\dots,|M|\}$, set $\mathcal{Z}_{M,s_M,j} = \{z \in \mathbb{R}^n : \text{there exists a } y \in \{A_{M,s_M}y < b_{M,s_M}\} \text{ such that } z = Z_{M,j}(y)\}$. Further, set $\mathcal{Z}_{M,j} = \bigcup_{s_M \in \mathcal{S}_M} \mathcal{Z}_{M,s_M,j} = \{z \in \mathbb{R}^n : \text{there exists a } y \in \bigcup_{s_M \in \mathcal{S}_M} \{A_{M,s_M}y < b_{M,s_M}\} \text{ such that } z = Z_{M,j}(y)\}$.

**Theorem 3.18.** *For all $M \in \mathcal{M}^+ \setminus \emptyset$, all $s_M \in \mathcal{S}^+_M$ and all $j \in \{1,\dots,|M|\}$, let the set $S_{M,s_M,j}(z)$ be defined as follows:*

$$S_{M,s_M,j}(z) = \begin{cases} \left(\mathcal{V}^-_{M,s_M,j}(z), \mathcal{V}^+_{M,s_M,j}(z)\right) & \text{if } \mathcal{V}^-_{M,s_M,j}(z) < \mathcal{V}^+_{M,s_M,j}(z) \\ \emptyset & \text{else.} \end{cases} \tag{3.75}$$

*Then $S_{M,s_M,j}(z)$ is an interval for all $z \in \mathcal{Z}_{M,s_M,j}$ and*

$$\eta'_{M,j}Y \mid \{\hat{M}=M, \hat{s}_{\hat{M}} = s_M, Z_{M,j}(Y) = z\} \sim TN(\beta^M_j, \sigma^2\eta'_{M,j}\eta_{M,j}, S_{M,s_M,j}(z)). \tag{3.76}$$

*Proof.* It is first shown that $S_{M,s_M,j}(z)$ is an interval for all $z \in \mathcal{Z}_{M,s_M,j}$: By definition of $\mathcal{Z}_{M,s_M,j}$, for every $z \in \mathcal{Z}_{M,s_M,j}$ there exists a $y \in \{A_{M,s_M}y < b_{M,s_M}\}$ such that $Z_{M,j}(y) = z$. For every $y \in \{A_{M,s_M}y < b_{M,s_M}\}$, the inequality $\mathcal{V}^-_{M,s_M,j}(Z_{M,j}(y)) < \mathcal{V}^+_{M,s_M,j}(Z_{M,j}(y))$ holds. Therefore, also for every $z \in \mathcal{Z}_{M,s_M,j}$, the inequality $\mathcal{V}^-_{M,s_M,j}(z) < \mathcal{V}^+_{M,s_M,j}(z)$ holds.

As $\{A_{M,s_M}y < b_{M,s_M}\} \subseteq \{\hat{M}=M, \hat{s}_{\hat{M}} = s_M\}$ and as for every $z \in \mathcal{Z}_{M,s_M,j}$ there exists a $y \in \{A_{M,s_M}y < b_{M,s_M}\}$ such that $Z_{M,j}(y) = z$, it follows that the conditioning set $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M, Z_{M,j}(y) = z\}$ is non-empty. Observe that $\eta'_{M,j}Y$ is normally distributed with mean $\eta'_{M,j}\mu = \beta_j^M$ and variance $\sigma^2\eta'_{M,j}\eta_{M,j}$. Then it follows that

$$
\begin{aligned}
&\eta'_{M,j}Y \mid \ \{\hat{M} = M, \hat{s}_{\hat{M}} = s_M, Z_{M,j}(Y) = z\} \\
&\stackrel{d}{=} \eta'_{M,j}Y \mid \{\mathcal{V}^-_{M,s_M,j}(Z_{M,j}(Y)) < \eta'_{M,j}Y < \mathcal{V}^+_{M,s_M,j}(Z_{M,j}(Y)), \mathcal{V}^0_{M,s_M,j}(Z_{M,j}(Y)) > 0, Z_{M,j}(Y) = z\} \\
&\stackrel{d}{=} \eta'_{M,j}Y \mid \{\mathcal{V}^-_{M,s_M,j}(z) < \eta'_{M,j}Y < \mathcal{V}^+_{M,s_M,j}(z), \mathcal{V}^0_{M,s_M,j}(z) > 0, Z_{M,j}(Y) = z\} \\
&\stackrel{d}{=} \eta'_{M,j}Y \mid \{\mathcal{V}^-_{M,s_M,j}(z) < \eta'_{M,j}Y < \mathcal{V}^+_{M,s_M,j}(z)\} \\
&\sim TN(\beta_j^M, \sigma^2\eta'_{M,j}\eta_{M,j}, S_{M,s_M,j}(z)),
\end{aligned}
\tag{3.77}
$$

where the first equality holds by equation (3.73), $Z_{M,j}(Y) = z$ is substituted to obtain the second equality and the third equality follows from the independence of $Z_{M,j}(Y)$ and $\eta'_{M,j}Y$. □

By definition of the set $\mathcal{Z}_{M,s_M,j}$, it follows that $Z_{M,j}(y) \in \mathcal{Z}_{M,s_M,j}$ for almost all $y \in \{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$. Only if $y$ is on the boundary of the set $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ (that is the faces of the polyhedron), it does not necessarily follow that $Z_{M,j}(y) \in \mathcal{Z}_{M,s_M,j}$. However, by Lemma 3.14, the faces of the polyhedron are null sets so that this case is negligible. This means that for almost all $y \in \{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$, the set $S_{M,s_M,j}(Z_{M,j}(y))$ is, in fact, an interval.

As a corollary of the previous theorem, the distribution of $\eta'_{M,j}Y$ conditional on $\{\hat{M} = M, Z_{M,j}(Y) = z\}$ is also a truncated normal distribution for all $z \in \mathcal{Z}_{M,j}$.

**Corollary 3.19.** *For all $M \in \mathcal{M}^+ \setminus \emptyset$ and all $j \in \{1, \ldots, |M|\}$, let the set $S_{M,j}(z)$ be defined as follows*

$$S_{M,j}(z) = \bigcup_{s_M \in \mathcal{S}_M} S_{M,s_M,j}(z). \tag{3.78}$$

*Then, for all $z \in \mathcal{Z}_{M,j}$, $S_{M,j}(z)$ is a finite union of intervals and*

$$\eta'_{M,j}Y \mid \ \{\hat{M} = M, Z_{M,j}(Y) = z\} \sim TN(\beta_j^M, \sigma^2\eta'_{M,j}\eta_{M,j}, S_{M,j}(z)) \tag{3.79}$$

*Proof.* We only need to show that $S_{M,j}(z)$ is a finite union of of intervals (or at least a single interval). The remaining arguments are identical to the ones in the proof of Theorem 3.18.

By definition of $S_{M,j}(z)$, for all $z \in S_{M,j}(z)$, there exists $y \in \bigcup_{s_M \in \mathcal{S}_M} \{A_{M,s_M} y < b_{M,s_M}\}$ such that $Z_{M,j}(y) = z$. This implies that there exists a $s_M \in \mathcal{S}_M$ such that $y \in \{A_{M,s_M} y < b_{M,s_M}\}$ and $Z_{M,j}(y) = z$. As $\mathcal{V}^{-}_{M,s_M,j}(Z_{M,j}(y)) < \mathcal{V}^{+}_{M,s_M,j}(Z_{M,j}(y)$ for all $y \in \{A_{M,s_M} y < b_{M,s_M}\}$, it follows that $S_{M,s_M,j}(z)$ is an interval and, therefore, $S_{M,j}(z)$ is at least a single interval. $S_{M,j}(z)$ is trivially a finite union of intervals, as the cardinality of $\mathcal{S}_M$ is finite.

□

As for every $y \in \{\hat{M} = M\}$ there exists a $s_M \in \mathcal{S}_M$ such that $y \in \{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$, it follows that for almost all $y \in \{\hat{M} = M\}$ (except for the faces of the polyhedra that are null sets by Lemma 3.14) there exists a set $S_{M,s_M,j}(Z_{M,j}(y))$ that is, in fact, an interval. Therefore, for almost all $y \in \{\hat{M} = M\}$, the set $S_{M,j}(Z_{M,j}(y))$ is, in fact, a union of intervals (or at least a single interval).

**Remark 3.20.** Simulations of the sets $S_{M,s_M,j}(z)$ suggest that $\{S_{M,s_M,j}(z)\}_{s_M \in \mathcal{S}_M}$ is a collection of sets of pairwise disjoint sets for all $M \in \mathcal{M}^+ \setminus \emptyset$, all $j = 1, \ldots, |M|$ and all $z \in \mathbb{R}^n$. This would imply that the set $S_{M,j}(z)$ is a union of disjoint intervals $S_{M,s_M,j}(z)$. Yet, we have not proved or disproved this assumption so far.

The following lemma implies that the set $S_{M,s_M,j}(z)$ is bounded either from below and/or above for all $z \in \mathbb{R}^n$ (and therefore especially for all $z \in \mathcal{Z}_{M,s_M,j}$). On the other hand, the lemma also implies that the set $S_{M,j}(z)$ is unbounded for all $z \in \mathbb{R}^n$ (and therefore especially for all $z \in \mathcal{Z}_{M,j}$). This lemma is a key result for the proofs of the upcoming theorems in Section 3.4.

**Lemma 3.21.** *For all $M \in \mathcal{M}^+ \setminus \emptyset$ and all $j \in \{1, \ldots, |M|\}$, there exists a $s^-_{M,j} \in \mathcal{S}_M$ and a $s^+_{M,j} \in \mathcal{S}_M$, such that*

$$\mathcal{V}^{-}_{M,s^-_{M,j},j}(z) = -\infty \quad \textit{and} \quad \mathcal{V}^{+}_{M,s^+_{M,j},j}(z) = +\infty \quad \textit{for all } z \in \mathbb{R}^n. \tag{3.80}$$

*Further, for every $s_M \in \mathcal{S}_M$ either*

$$\infty < \mathcal{V}^{-}_{M,s_M,j}(z) \quad \textit{for all } z \in \mathbb{R}^n \tag{3.81}$$

*or*

$$\mathcal{V}^{+}_{M,s_M,j}(z) < +\infty \quad \textit{for all } z \in \mathbb{R}^n \tag{3.82}$$

*or both.*

*Proof.* From the definition of $\mathcal{V}^-_{M,s_M,j}(z)$ in (3.70), we see that $\mathcal{V}^-_{M,s_M,j}(z) = -\infty$ if and only if $(A_{M,s_M} c_{M,j})_i \geq 0$ for all $i$. Analogously, $\mathcal{V}^+_{M,s_M,j}(z) = +\infty$ if and only if $(A_{M,s_M} c_{M,j})_i \leq 0$ for all $i$. Consequently, $\mathcal{V}^-_{M,s_M,j}(z) = -\infty$ and, at the same time, $\mathcal{V}^+_{M,s_M,j}(z) = +\infty$ if and only if $(A_{M,s_M} c_{M,j})_i = 0$ for all $i$. Now, $A_{M,s_M}$ is defined as

$$A_{M,s_M} = \begin{pmatrix} A^0_{M,s_M} \\ A^1_{M,s_M} \end{pmatrix}, \tag{3.83}$$

where $A^0_{M,s_M}$ and $A^1_{M,s_M}$ are defined in (3.24) and (3.25). The definition of $c_{M,j}$ is given in (3.69) and of $\eta_{M,j}$ in (3.13). For $A^0_{M,s_M} c_{M,j}$, we get:

$$\begin{aligned} A^0_{M,s_M} c_{M,j} &= \frac{1}{\lambda} \begin{pmatrix} W'_{-M}(I_n - P_M) \\ -W'_{-M}(I_n - P_M) \end{pmatrix} \frac{\eta_{M,j}}{\eta'_{M,j}\eta_{M,j}} \\ &= \frac{1}{\lambda \eta'_{M,j}\eta_{M,j}} \begin{pmatrix} W'_{-M}(I_n - P_M) W_M (W'_M W_M)^{-1} e_j \\ -W'_{-M}(I_n - P_M) W_M (W'_M W_M)^{-1} e_j \end{pmatrix} \\ &= \frac{1}{\lambda \eta'_{M,j}\eta_{M,j}} \begin{pmatrix} W'_{-M}\left(W_M (W'_M W_M)^{-1} - W_M (W'_M W_M)^{-1}\right) e_j \\ -W'_{-M}\left(W_M (W'_M W_M)^{-1} - W_M (W'_M W_M)^{-1}\right) e_j \end{pmatrix} \\ &= 0. \end{aligned} \tag{3.84}$$

Hence, $A^0_{M,s_M} c_{M,j} = 0$ holds for all $s_M \in \mathcal{S}_M$. Note that if $M = M_F$, then $A^0_{M_F,s_{M_F}}$ is to be interpreted as $(0, \ldots, 0)$ and therefore $A^0_{M_F,s_{M_F}} c_{M_F,j} = 0$ also holds. On the other hand, for $A^1_{M,s_M} c_{M,j}$, we obtain:

$$\begin{aligned} A^1_{M,s_M} c_{M,j} &= -\mathrm{diag}(s_M)(W'_M W_M)^{-1} W'_M c_{M,j} \\ &= \mathrm{diag}(s_M)(-1)(W'_M W_M)^{-1} W'_M \frac{\eta_{M,j}}{\eta'_{M,j}\eta_{M,j}} \\ &= \mathrm{diag}(s_M) \underbrace{(W'_M W_M)^{-1} e_j \frac{-1}{\eta'_{M,j}\eta_{M,j}}}_{=v} \\ &= \mathrm{diag}(s_M) v. \end{aligned} \tag{3.85}$$

Note that in the last line of the previous equation, every element of $v$ is multiplied by either $+1$ or $-1$. Thus $A^1_{M,s_M} c_{M,j} = 0$ if and only if $v = 0$. The vector $v$ is the $j$th column of $(W_M'W_M)^{-1}$ multiplied by the factor $\frac{-1}{\eta_{M,j}'\eta_{M,j}}$. Since $W_M'W_M$ has full rank it follows that $(W_M'W_M)^{-1}e_j \neq 0$ and therefore $v \neq 0$. This implies that for any $s_M \in \mathcal{S}_M$, there exists a $i$ such that either $(A^1_{M,s_M}c_{M,j})_i > 0$ or $(A^1_{M,s_M}c_{M,j})_i < 0$. Therefore, either

$$-\infty < \mathcal{V}^-_{M,s_M,j}(z) \quad \text{for all } z \in \mathbb{R}^n \tag{3.86}$$

or

$$\mathcal{V}^+_{M,s_M,j}(z) < +\infty \quad \text{for all } z \in \mathbb{R}^n \tag{3.87}$$

or both. As every element of $v$ is multiplied by either $+1$ or $-1$, it follows that there exists a $s^-_{M,j} \in \mathcal{S}_M$ such that $A^1_{M,s^-_{M,j}}c_{M,j} \geq 0$ holds. This implies that

$$\mathcal{V}^-_{M,s^-_{M,j},j}(z) = -\infty \quad \text{for all } z \in \mathbb{R}^n. \tag{3.88}$$

Analogously, we can find a $s^+_{M,j} \in \mathcal{S}_M$ such that $A^1_{M,s^+_{M,j}}c_{M,j} \leq 0$ and therefore

$$\mathcal{V}^+_{M,s^+_{M,j},j}(z) = +\infty \quad \text{for all } z \in \mathbb{R}^n. \tag{3.89}$$

□

From the previous lemma it follows that, for all $z \in \mathcal{Z}_{M,s_M,j}$, the set $S_{M,s_M,j}(z) = \left(\mathcal{V}^-_{M,s_M,j}(z), \mathcal{V}^+_{M,s_M,j}(z)\right)$ is at least bounded from below or above. On the other hand, for all $z \in \mathbb{R}^n$ there exists a $s^-_{M,j} \in \mathcal{S}_M$ and a $s^+_{M,j} \in \mathcal{S}_M$ such that $\mathcal{V}^-_{M,s^-_{M,j},j}(z) = -\infty$ and $\mathcal{V}^+_{M,s^+_{M,j},j}(z) = +\infty$. Note that Lemma 3.21 also implies that $\mathcal{V}^+_{M,s^-_{M,j},j}(z) < +\infty$ and $\mathcal{V}^-_{M,s^+_{M,j},j}(z) > -\infty$ hold for all $z \in \mathbb{R}^n$. Since $S_{M,j}(z) = \bigcup_{s_M \in \mathcal{S}_M} S_{M,s_M,j}(z)$, the unboundedness of $S_{M,j}(z)$ follows for all $z \in \mathbb{R}$. Clearly, this also means that the set $S_{M,j}(z)$ is non-empty for all $z \in \mathbb{R}^n$.

Recall that the c.d.f. of a $N(\theta, \sigma^2)$-distributed random variable $X$ truncated to the set $S$ was denoted by $F^S_{\theta,\sigma^2}(x)$ (see Definition 2.4). By Theorem 3.18, it therefore follows that, conditional on $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M, Z_{M,j}(Y) = z\}$, $\eta_{M,j}'Y$ has c.d.f. $F^{S_{M,s_M,j}(z)}_{\beta^M_j,\sigma^2\eta_{M,j}'\eta_{M,j}}(\eta_{M,j}'y)$ for all $M \in \mathcal{M}^+ \setminus \emptyset$ and all $s_M \in \mathcal{S}^+_M$. By Corollary 3.19, it follows that, conditional on $\{\hat{M} = M, Z_{M,j}(Y) = z\}$, $\eta_{M,j}'Y$ has c.d.f. $F^{S_{M,j}(z)}_{\beta^M_j,\sigma^2\eta_{M,j}'\eta_{M,j}}(\eta_{M,j}'y)$ for all $M \in \mathcal{M}^+ \setminus \emptyset$.

**Theorem 3.22.** *Let the set $S_{M,j}(z)$ be defined as in (3.78). For all $M \in \mathcal{M}^+ \setminus \emptyset$ and all $j \in \{1, \ldots, |M|\}$, let $L_{M,j}(y)$ and $U_{M,j}(y)$ satisfy*

$$F^{S_{M,j}(Z_{M,j}(y))}_{L_{M,j}(y),\sigma^2\eta'_{M,j}\eta_{M,j}}(\eta'_{M,j}y) = 1 - \frac{\alpha}{2} \quad \text{for all } y \in \bigcup_{s_M \in \mathcal{S}_M} \{A_{M,s_M}y < b_{M,s_M}\} \tag{3.90}$$

*and*

$$F^{S_{M,j}(Z_{M,j}(y))}_{U_{M,j}(y),\sigma^2\eta'_{M,j}\eta_{M,j}}(\eta'_{M,j}y) = \frac{\alpha}{2} \quad \text{for all } y \in \bigcup_{s_M \in \mathcal{S}_M} \{A_{M,s_M}y < b_{M,s_M}\}. \tag{3.91}$$

*For any $y \notin \bigcup_{s_M \in \mathcal{S}_M} \{A_{M,s_M}y < b_{M,s_M}\}$, set $L_{M,j}(y) = U_{M,j}(y) = 0$. Then it follows that*

$$\mathbb{P}_{\mu,\sigma^2}\left(L_{M,j}(Y) \le \beta^M_j \le U_{M,j}(Y) | \hat{M} = M\right) = 1 - \alpha \quad \text{for all } \mu \in \mathbb{R}^n,\ \sigma^2 \in (0, +\infty). \tag{3.92}$$

*Proof.* Corollary 3.19 implies that, for all $z \in \mathcal{Z}_{M,j}$, $\eta'_{M,j}Y$, conditional on $\{\hat{M} = M, Z_{M,j}(Y) = z\}$, is truncated normally distributed with c.d.f. $F^{S_{M,j}(z)}_{\beta^M_j,\sigma^2\eta'_{M,j}\eta_{M,j}}(\eta'_{M,j}y)$. Existence and uniqueness of the boundary functions $L_{M,j}(y)$ and $U_{M,j}(y)$ are guaranteed by Lemma 2.3. By Theorem 2.4, it follows that for all $z \in \mathcal{Z}_{M,j}$ the confidence interval $[L_{M,j}(Y), U_{M,j}(Y)]$ has exact conditional coverage probability of $1 - \alpha$ on the set $\{\hat{M} = M, Z_{M,j}(Y) = z\}$, that is,

$$\mathbb{P}_{\mu,\sigma^2}\left(L_{M,s_M,j}(Y) \le \beta^M_j \le U_{M,s_M,j}(Y) | \hat{M} = M, Z_{M,j}(Y) = z\right) = 1 - \alpha. \tag{3.93}$$

For almost all $y \in \{\hat{M} = M\}$ (aside from the faces that are null sets by Lemma 3.14), it follows that $Z_{M,j}(y) \in \mathcal{Z}_{M,j}$. Thus it follows that

$$\begin{aligned}
&\mathbb{P}_{\mu,\sigma^2}\left(L_{M,s_M,j}(Y) \le \beta^M_j \le U_{M,s_M,j}(Y) | \hat{M} = M\right) \\
&= \mathbb{E}_{\mu,\sigma^2}\left(\mathbb{P}_{\mu,\sigma^2}\left(L_{M,s_M,j}(Y) \le \beta^M_j \le U_{M,s_M,j}(Y) | \hat{M} = M, Z_{M,j}(Y)\right) | \hat{M} = M\right) \\
&= \mathbb{E}_{\mu,\sigma^2}\left(1 - \alpha | \hat{M} = M\right) \\
&= 1 - \alpha.
\end{aligned} \tag{3.94}$$

□

Note that the set $S_{M,j}(z)$ is the union over all possible sign vectors $s_M \in \mathcal{S}_M$. However, there are $2^{|M|}$ sign vectors in total. It may be computationally feasible to compute the set $S_{M,j}(z)$ if $|M|$ is small, but for large $|M|$, the task is infeasible. Lee et al. (2016) therefore suggested another confidence interval that has exact conditional coverage probability of $1-\alpha$ on the set $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$.

**Theorem 3.23.** *Let the set $S_{M,s_M,j}(z)$ be defined as in (3.75). For all $M \in \mathcal{M}^+ \setminus \emptyset$, all $s_M \in \mathcal{S}_M^+$ and all $j \in \{1, \ldots, |M|\}$, let $L_{M,s_M,j}(y)$ and $U_{M,s_M,j}(y)$ satisfy*

$$F^{S_{M,s_M,j}(Z_{M,j}(y))}_{L_{M,s_M,j}(y),\sigma^2\eta'_{M,j}\eta_{M,j}}(\eta'_{M,j}y) = 1 - \frac{\alpha}{2} \quad \text{for all } y \in \{A_{M,s_M}y < b_{M,s_M}\} \tag{3.95}$$

*and*

$$F^{S_{M,s_M,j}(Z_{M,j}(y))}_{U_{M,s_M,j}(y),\sigma^2\eta'_{M,j}\eta_{M,j}}(\eta'_{M,j}y) = \frac{\alpha}{2} \quad \text{for all } y \in \{A_{M,s_M}y < b_{M,s_M}\}. \tag{3.96}$$

*For any $y \notin \{A_{M,s_M}y < b_{M,s_M}\}$, set $L_{M,s_M,j}(y) = U_{M,s_M,j}(y) = 0$. Then it follows that*

$$\mathbb{P}_{\mu,\sigma^2}\left(L_{M,s_M,j}(Y) \leq \beta_j^M \leq U_{M,s_M,j}(Y) | \hat{M} = M, \hat{s}_{\hat{M}} = s_M\right) = 1 - \alpha \tag{3.97}$$

*for all $\mu \in \mathbb{R}^n$ and all $\sigma^2 \in (0, +\infty)$.*

*Proof.* The proof is identical to the proof of the previous theorem. The only change is that Theorem 3.18 is used instead of Corollary 3.19. $\square$

**Remark 3.24.** Note that the confidence interval $[L_{M,s_M,j}(Y), U_{M,s_M,j}(Y)]$ is only constructed for sign vectors $s_M \in \mathcal{S}_M^+$, that is, only for sets $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ that have positive observation probability.

The set $S_{M,s_M,j}(z)$ is easily computed as only the sign vector $s_M$ is required for the computation. Yet, it was noted in Lee et al. (2016) that the confidence interval $[L_{M,s_M,j}(Y), U_{M,s_M,j}(Y)]$ will be in general wider than the confidence interval $[L_{M,j}(Y), U_{M,j}(Y)]$.

## 3.4 Conditional expected length of the confidence intervals

In Section 2.3 and 2.4, it was shown that the boundedness of the truncation set $S$ is crucial for the conditional expected length of confidence intervals with exact conditional coverage probability of $1-\alpha$. If the truncation set is either bounded from below and/or above, then Theorem 2.7 implies that the confidence interval $[L_S(X, U_S(X)]$, where $L_S(x)$ and $U_S(x)$ are defined as in (2.16) and (2.17), has infinite conditional expected length. By Lemma 3.21, the truncation set $S_{M,s_M,j}(z)$ is either bounded from below and/or above for all $z \in \mathcal{Z}_{M,s_M,j}$. Applying Theorem 2.7 to the confidence interval $[L_{M,s_M,j}(Y), U_{M,s_M,j}(Y)]$ yields the following result.

**Theorem 3.25.** *Let $L_{M,s_M,j}(y)$ and $U_{M,s_M,j}(y)$ be defined as in (3.95) and (3.96), respectively. Then for all $M \in \mathcal{M}^+ \setminus \emptyset$, all $s_M \in \mathcal{S}_M^+$ and all $j \in 1,\dots,|M|$, it follows that the expected length of the confidence interval $[L_{M,s_M,j}(Y), U_{M,s_M,j}(Y)]$, conditional on $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$, is $+\infty$, that is,*

$$\mathbb{E}_{\mu,\sigma^2}\left(U_{M,s_M,j}(Y) - L_{M,s_M,j}(Y) | \hat{M} = M, \hat{s}_{\hat{M}} = s_M\right) = +\infty \quad \textit{for all } \mu \in \mathbb{R}^n,\ \sigma^2 \in (0,+\infty). \tag{3.98}$$

*Proof.* Theorem 3.18 implies that, for all $s_M \in \mathcal{S}_M^+$ and all $z \in \mathcal{Z}_{M,j}$, $\eta'_{M,j}Y$, conditional on $\{\hat{M} = M, Z_{M,j}(Y) = z\}$, is truncated normally distributed with c.d.f. $F^{S_{M,s_M,j}(z)}_{\beta_j^M, \sigma^2 \eta'_{M,j}\eta_{M,j}}(\eta'_{M,j}y)$. By Lemma 3.21, it follows that the truncation set $S_{M,s_M,j}(z)$ is either bounded from below and/or above. Thus Theorem 2.7 implies that

$$\mathbb{E}_{\mu,\sigma^2}\left(U_{M,s_M,j}(Y) - L_{M,s_M,j}(Y) | \hat{M} = M, \hat{s}_{\hat{M}} = s_M, Z_{M,j}(Y) = z\right) = +\infty \quad \text{for all } z \in \mathcal{Z}_{M,s_M,j}. \tag{3.99}$$

For almost all $y \in \{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ (aside from the faces that are null sets by Lemma 3.14), it follows that $Z_{M,j}(y) \in \mathcal{Z}_{M,s_M,j}$. Thus it follows that

$$\begin{aligned}
&\mathbb{E}_{\mu,\sigma^2}\left(U_{M,s_M,j}(Y) - L_{M,s_M,j}(Y) | \hat{M} = M, \hat{s}_{\hat{M}} = s_M\right) \\
&= \mathbb{E}_{\mu,\sigma^2}\left(\mathbb{E}_{\mu,\sigma^2}\left(U_{M,s_M,j}(Y) - L_{M,s_M,j}(Y) | \hat{M} = M, \hat{s}_{\hat{M}} = s_M, Z_{M,j}(Y)\right) | \hat{M} = M, \hat{s}_{\hat{M}} = s_M\right) \\
&= \mathbb{E}_{\mu,\sigma^2}\left(+\infty | \hat{M} = M, \hat{s}_{\hat{M}} = s_M\right) \\
&= +\infty.
\end{aligned} \tag{3.100}$$

□

On the other hand, Lemma 3.21 implies that the truncation set $S_{M,j}(z)$ is unbounded for all $z \in \mathcal{Z}_{M,j}$. If the truncation set is unbounded, then Theorem 2.9 implies that the confidence interval $[L_S(X, U_S(X)]$, where $L_S(x)$ and $U_S(x)$ are defined as in (2.16) and (2.17), has finite conditional expected length. Applying Theorem 2.9 to the confidence interval $[L_{M,j}(Y), U_{M,j}(Y)]$ yields the following result.

**Theorem 3.26.** *Let $L_{M,j}(y)$ and $U_{M,j}(y)$ be defined as in (3.90) and (3.91), respectively. Then for all $M \in \mathcal{M}^+ \setminus \emptyset$ and all $j \in 1, \ldots, |M|$ it follows that the expected length of the confidence interval $[L_{M,j}(Y), U_{M,j}(Y)]$, conditional on $\{\hat{M} = M\}$, is finite, that is,*

$$\mathbb{E}_{\mu,\sigma^2}\left(U_{M,j}(Y) - L_{M,j}(Y)|\hat{M} = M\right) < +\infty \quad \text{for all } \mu \in \mathbb{R}^n,\ \sigma^2 \in (0, +\infty). \tag{3.101}$$

*Proof.* Corollary 3.19 implies that, for all $z \in \mathcal{Z}_{M,j}$, $\eta'_{M,j}Y$, conditional on $\{\hat{M} = M, Z_{M,j}(Y) = z\}$, is truncated normally distributed with c.d.f. $F^{S_{M,j}(z)}_{\beta^M_j, \sigma^2\eta'_{M,j}\eta_{M,j}}(\eta'_{M,j}y)$. We use $\sigma^2_{M,j}$ as shorter notation for $\sigma^2\eta'_{M,j}\eta_{M,j}$. Recall that $S_{M,j}(z)$ is defined as follows

$$S_{M,j}(z) = \bigcup_{s_M \in \mathcal{S}_M} \left(\mathcal{V}^-_{M,s_M,j}(z), \mathcal{V}^+_{M,s_M,j}(z)\right) \tag{3.102}$$

By Corollary 3.19 and Lemma 3.21, the truncation set $S_{M,j}(z)$ is finite union of intervals that is unbounded for all $z \in \mathcal{Z}_{M,j}$. Also by Lemma 3.21, there exists a $s^-_{M,j} \in \mathcal{S}_M$ such that $\mathcal{V}^-_{M,s^-_{M,j},j}(z) = -\infty$ and $\mathcal{V}^+_{M,s^-_{M,j},j}(z) < +\infty$, and a $s^+_{M,j} \in \mathcal{S}_M$ such that $\mathcal{V}^+_{M,s^+_{M,j},j}(z) = +\infty$ and $\mathcal{V}^-_{M,s^+_{M,j},j}(z) > -\infty$. First an upper bound of

$$\mathbb{E}_{\mu,\sigma^2}\left(U_{M,j}(Y) - L_{M,j}(Y)|\hat{M} = M, Z_{M,j}(Y) = z\right) \tag{3.103}$$

is derived for each $z \in \mathcal{Z}_{M,j}$. Recall that $S_{M,j}(z) = \bigcup_{s_M \in \mathcal{S}_M} S_{M,s_M,j}(z)$ where $S_{M,s_M,j}(z)$ was defined in (3.75). As mentioned in Remark 3.20, we cannot assume that the sets $S_{M,s_M,j}(z)$ are disjoint. Therefore, as an intermediate step, we need to distinguish two cases. The first case is that $S_{M,j}(z)$ is a union of possibly overlapping intervals that does not equal to $\mathbb{R}$. The second case is that $S_{M,j}(z) = \mathbb{R}$.

First case: A finite union overlapping intervals can again be written as a union of disjoint intervals. Define $D_1(z)$ and $D_2(z)$ as follows:

$$D_1(z) = \inf\{t \in \mathbb{R} : t \notin S_{M,j}(z)\} \quad \text{and} \quad D_2(z) = \sup\{t \in \mathbb{R} : t \notin S_{M,j}(z)\}. \tag{3.104}$$

Since the set $S_{M,j}(z)$ is unbounded from below and above, it follows that $-\infty < D_1(z) < D_2(z) < +\infty$. If the sets $S_{M,s_M,j}(z)$ are disjoint, then it follows that $\mathcal{V}^+_{M,s^-_{M,j},j}(z) = D_1(z)$ and $\mathcal{V}^-_{M,s^+_{M,j},j}(z) = D_2(z)$, otherwise, the following inequality is satisfied:

$$\mathcal{V}^+_{M,s^-_{M,j},j}(z) \leq D_1(z) \quad \text{and} \quad \mathcal{V}^-_{M,s^+_{M,j},j}(z) \geq D_2(z). \tag{3.105}$$

Theorem 2.9 implies that the conditional expectation in (3.103) is bounded by $C_{\beta^M_j,\sigma^2_{M,j}}(D_1(z), D_2(z))$, where the function $C_{\beta^M_j,\sigma^2_{M,j}}$ is defined in (2.41). In addition, by Theorem 2.9, the function $C_{\beta^M_j,\sigma^2_{M,j}}$ is strictly decreasing in the first argument and strictly increasing in the second argument. Thus it follows that $C_{\beta^M_j,\sigma^2_{M,j}}(D_1(z), D_2(z)) \leq C_{\beta^M_j,\sigma^2_{M,j}}(\mathcal{V}^+_{M,s^-_{M,j},j}(z), \mathcal{V}^-_{M,s^+_{M,j},j}(z))$. Define the function $B_1(t)$ and $B_2(t)$ for all $t \in \mathbb{R}$ as follows:

$$B_1(t) = 2|t| + \sigma_{M,j}\frac{\phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right)}{1-\Phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right)} \tag{3.106}$$

$$B_2(t) = 2|t| + \sigma_{M,j}\frac{\phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right)}{\Phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right)}. \tag{3.107}$$

Note that $B_1(t) > 0$ and $B_2(t) > 0$ hold for all $t \in \mathbb{R}$. It will be convenient to use an even larger bound than $C_{\beta_j^M,\sigma^2_{M,j}}(\mathcal{V}^+_{M,s^-_{M,j},j}(z),\mathcal{V}^-_{M,s^+_{M,j},j}(z))$:

$$\begin{aligned}
&C_{\beta_j^M,\sigma^2_{M,j}}(\mathcal{V}^+_{M,s^-_{M,j},j}(z),\mathcal{V}^-_{M,s^+_{M,j},j}(z)) = 2(\mathcal{V}^-_{M,s^+_{M,j},j}(z)-\mathcal{V}^+_{M,s^-_{M,j},j}(z))\\
&\quad+\sigma_{M,j}\frac{\phi\left(\frac{\mathcal{V}^+_{M,s^-_{M,j},j}(z)-\beta_j^M}{\sigma_{M,j}}\right)}{\Phi\left(\frac{\mathcal{V}^+_{M,s^-_{M,j},j}(z)-\beta_j^M}{\sigma_{M,j}}\right)}+\sigma_{M,j}\frac{\phi\left(\frac{\mathcal{V}^-_{M,s^+_{M,j},j}(z)-\beta_j^M}{\sigma_{M,j}}\right)}{1-\Phi\left(\frac{\mathcal{V}^-_{M,s^+_{M,j},j}(z)-\beta_j^M}{\sigma_{M,j}}\right)}+2\sigma_{M,j}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\\
&\quad\le 2|\mathcal{V}^-_{M,s^+_{M,j},j}(z)|+2|\mathcal{V}^+_{M,s^-_{M,j},j}(z)|+\sigma_{M,j}\frac{\phi\left(\frac{\mathcal{V}^+_{M,s^-_{M,j},j}(z)-\beta_j^M}{\sigma_{M,j}}\right)}{\Phi\left(\frac{\mathcal{V}^+_{M,s^-_{M,j},j}(z)-\beta_j^M}{\sigma_{M,j}}\right)}\\
&\quad+\sigma_{M,j}\frac{\phi\left(\frac{\mathcal{V}^-_{M,s^+_{M,j},j}(z)-\beta_j^M}{\sigma_{M,j}}\right)}{1-\Phi\left(\frac{\mathcal{V}^-_{M,s^+_{M,j},j}(z)-\beta_j^M}{\sigma_{M,j}}\right)}+2\sigma_{M,j}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\\
&\quad= B_1(\mathcal{V}^-_{M,s^+_{M,j},j}(z))+B_2(\mathcal{V}^+_{M,s^-_{M,j},j}(z))+2\sigma_{M,j}\Phi^{-1}\left(1-\frac{\alpha}{2}\right).
\end{aligned}\tag{3.108}$$

Thus it follows that the conditional expectation in (3.103) is bounded by $B_1(\mathcal{V}^-_{M,s^+_{M,j},j}(z))+B_2(\mathcal{V}^+_{M,s^-_{M,j},j}(z))+2\sigma_{M,j}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$.

Second case: In this case, the equation $F^{S_{M,j}(z)}_{\beta_j^M,\sigma^2_{M,j}}(\eta'_{M,j}y)=\Phi_{\beta_j^M,\sigma^2_{M,j}}(\eta'_{M,j}y)$ is satisfied. Therefore, we simply obtain:

$$\begin{aligned}
&\mathbb{E}_{\mu,\sigma^2}\left(U_{M,j}(Y)-L_{M,j}(Y)|\hat{M}=M,Z_{M,j}(Y)=z\right)=2\sigma_{M,j}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\\
&\quad< B_1(\mathcal{V}^-_{M,s^+_{M,j},j}(z))+B_2(\mathcal{V}^+_{M,s^-_{M,j},j}(z))+2\sigma_{M,j}\Phi^{-1}\left(1-\frac{\alpha}{2}\right).
\end{aligned}\tag{3.109}$$

The inequality in the previous line is trivially satisfied as both $B_1(t)$ and $B_2(t)$ are strictly positive functions.

Thus from the previous two cases we can conclude that

$$\begin{aligned}
&\mathbb{E}_{\mu,\sigma^2}\left(U_{M,j}(Y)-L_{M,j}(Y)|\hat{M}=M,Z_{M,j}(Y)=z\right)\\
&\quad< B_1(\mathcal{V}^-_{M,s^+_{M,j},j}(z))+B_2(\mathcal{V}^+_{M,s^-_{M,j},j}(z))+2\sigma_{M,j}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\quad\text{for all } z\in\mathcal{Z}_{M,j}.
\end{aligned}\tag{3.110}$$

Note that the upper bound is strictly positive for all $z \in \mathbb{R}$, as both $B_1(t)$ and $B_2(t)$ are strictly positive for all $t \in \mathbb{R}$. Therefore, the expected length of the confidence interval $[L_{M,j}(Y), U_{M,j}(Y)]$, conditional on $\{\hat{M} = M\}$, is bounded from above as follows:

$$
\begin{aligned}
&\mathbb{E}_{\mu,\sigma^2}\left(U_{M,j}(Y) - L_{M,j}(Y)|\hat{M} = M\right) \\
&= \mathbb{E}_{\mu,\sigma^2}\left(\mathbb{E}_{\mu,\sigma^2}\left(U_{M,j}(Y) - L_{M,j}(Y)|\hat{M} = M, Z_{M,j}(Y)\right)|\hat{M} = M\right) \\
&< \mathbb{E}_{\mu,\sigma^2}\left(B_1\left(\mathcal{V}^-_{M,s^+_{M,j},j}(Z_{M,j}(Y))\right) + B_2\left(\mathcal{V}^+_{M,s^-_{M,j},j}(Z_{M,j}(Y))\right)|\hat{M} = M\right) + 2\sigma_{M,j}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \\
&= \frac{1}{\mathbb{P}_{\mu,\sigma^2}\left(\hat{M} = M\right)}\mathbb{E}_{\mu,\sigma^2}\left(\left(B_1\left(\mathcal{V}^-_{M,s^+_{M,j},j}(Z_{M,j}(Y))\right) + B_2\left(\mathcal{V}^+_{M,s^-_{M,j},j}(Z_{M,j}(Y))\right)\right)\mathbb{1}_{\{\hat{M}=M\}}(Y)\right) \\
&+ 2\sigma_{M,j}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \\
&< \frac{1}{\mathbb{P}_{\mu,\sigma^2}\left(\hat{M} = M\right)}\mathbb{E}_{\mu,\sigma^2}\left(B_1\left(\mathcal{V}^-_{M,s^+_{M,j},j}(Z_{M,j}(Y))\right) + B_2\left(\mathcal{V}^+_{M,s^-_{M,j},j}(Z_{M,j}(Y))\right)\right) \\
&+ 2\sigma_{M,j}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right).
\end{aligned}
\tag{3.111}
$$

Hence, to show that $\mathbb{E}_{\mu,\sigma^2}\left(U_{M,j}(Y) - L_{M,j}(Y)|\hat{M} = M\right)$ is finite, it is sufficient to show that both $\mathbb{E}_{\mu,\sigma^2}\left(B_1\left(\mathcal{V}^-_{M,s^+_{M,j},j}(Z_{M,j}(Y))\right)\right)$ and $\mathbb{E}_{\mu,\sigma^2}\left(B_2\left(\mathcal{V}^+_{M,s^-_{M,j},j}(Z_{M,j}(Y))\right)\right)$ are finite. Note that $\mathcal{V}^-_{M,s^+_{M,j},j}(Z_{M,j}(y))$ and $\mathcal{V}^+_{M,s^-_{M,j},j}(Z_{M,j}(y))$ are defined as follows:

$$
\mathcal{V}^-_{M,s^+_{M,j},j}(Z_{M,j}(y)) = \max_{i:(A_{M,s^+_{M,j}}c_{M,j})_i<0} \frac{(b_{M,s^+_{M,j}})_i - (A_{M,s^+_{M,j}}Z_{M,j}(y))_i}{(A_{M,s^+_{M,j}}c_{M,j})_i} \tag{3.112}
$$

$$
\mathcal{V}^+_{M,s^-_{M,j},j}(Z_{M,j}(y)) = \min_{i:(A_{M,s^-_{M,j}}c_{M,j})_i>0} \frac{(b_{M,s^-_{M,j}})_i - (A_{M,s^-_{M,j}}Z_{M,j}(y))_i}{(A_{M,s^-_{M,j}}c_{M,j})_i}. \tag{3.113}
$$

We define $T_{-,i}(y)$ and $T_{+,i}(y)$ as follows:

$$
T_{-,i}(y) = \frac{(b_{M,s^+_{M,j}})_i - (A_{M,s^+_{M,j}}Z_{M,j}(y))_i}{(A_{M,s^+_{M,j}}c_{M,j})_i} \quad \text{for all } i : (A_{M,s^+_{M,j}}c_{M,j})_i < 0 \tag{3.114}
$$

$$
T_{+,i}(y) = \frac{(b_{M,s^-_{M,j}})_i - (A_{M,s^-_{M,j}}Z_{M,j}(y))_i}{(A_{M,s^-_{M,j}}c_{M,j})_i} \quad \text{for all } i : (A_{M,s^-_{M,j}}c_{M,j})_i > 0. \tag{3.115}
$$

$Z_{M,j}(y)$ is a linear function in $y$, therefore $T_{-,i}(y)$ and $T_{+,i}(y)$ are also linear functions in $y$. This implies that $T_{-,i}(Y)$ and $T_{+,i}(Y)$ are normally distributed for all $i$. Since both $B_1(t)$ and $B_2(t)$ are strictly positive, we obtain

$$B_1\left(\mathcal{V}^-_{M,s^+_{M,j},j}(Z_{M,j}(y))\right) \leq \sum_{i:(A_{M,s^+_{M,j}}c_{M,j})_i<0} B_1(T_{-,i}(y)) \tag{3.116}$$

$$B_2\left(\mathcal{V}^+_{M,s^-_{M,j},j}(Z_{M,j}(y))\right) \leq \sum_{i:(A_{M,s^-_{M,j}}c_{M,j})_i>0} B_2(T_{+,i}(y)). \tag{3.117}$$

This implies that it is sufficient to show that both $\mathbb{E}\left(B_1(T)\right)$ and $\mathbb{E}\left(B_2(T)\right)$ are finite for any normally distributed random variable $T$. Let denote by $\phi_T(t)$ the p.d.f. of $T$. Consider the function $B_1(t)$ first. Recall the definition of $B_1(t)$ in (3.106). Obviously, $\mathbb{E}(|T|) < +\infty$ holds. Note that $\phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right) \to 0$ and $1 - \Phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right) \to 1$ as $t \to -\infty$. This implies that

$$\lim_{t\to-\infty} \frac{\phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right)}{1-\Phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right)} = 0. \tag{3.118}$$

Thus it follows that

$$\int\limits_{-\infty}^{t^*} \frac{\phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right)}{1-\Phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right)}\phi_T(t)dt < +\infty \quad \text{for all } t^* \in \mathbb{R}. \tag{3.119}$$

On the other hand, observe that $\frac{1}{t}\phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right) \to 0$ and $1-\Phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right) \to 0$ as $t \to +\infty$. By L'Hospital's rule it follows that

$$\begin{aligned}\lim_{t\to+\infty} \frac{\frac{1}{t}\phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right)}{1-\Phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right)} &= \lim_{t\to+\infty} \frac{-\frac{1}{t^2}\phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right) - \frac{1}{t}\frac{1}{\sigma_{M,j}}\frac{t-\beta^M_j}{\sigma_{M,j}}\phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right)}{-\frac{1}{\sigma_{M,j}}\phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right)} \\ &= \lim_{t\to+\infty} \frac{1}{t^2\sigma_{M,j}} + \frac{1}{\sigma_{M,j}} - \frac{\beta^M_j}{t\sigma_{M,j}} = \frac{1}{\sigma_{M,j}}.\end{aligned} \tag{3.120}$$

Fix an $\epsilon > 0$. Then there exists a $t^* > 0$ such that

$$\frac{\phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right)}{1-\Phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right)} < \left(\frac{1}{\sigma_{M,j}} + \epsilon\right)t \quad \text{for all } t > t^*. \tag{3.121}$$

This, in turn, implies that

$$\int\limits_{t^*}^{+\infty} \frac{\phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right)}{1-\Phi\left(\frac{t-\beta^M_j}{\sigma_{M,j}}\right)}\phi_T(t)dt < \left(\frac{1}{\sigma_{M,j}} + \epsilon\right)\int\limits_{t^*}^{+\infty} t\phi_T(t)dt < +\infty. \tag{3.122}$$

Hence, it follows that

$$\mathbb{E}\left(\frac{\phi\left(\frac{T-\beta_j^M}{\sigma_{M,j}}\right)}{1-\Phi\left(\frac{T-\beta_j^M}{\sigma_{M,j}}\right)}\right) < +\infty. \tag{3.123}$$

Thus we showed that $\mathbb{E}\left(B_1(T)\right) < +\infty$. By analogous arguments, it also follows that $\mathbb{E}\left(B_2(T)\right) < +\infty$ for any normally distributed random variable $T$. $\square$

# 4 Conclusion

In the simple setting of a normally distributed random variable, we showed that exact conditional coverage leads to unpleasant length properties of confidence intervals if the conditioning set is bounded (either from below and/or above). Thus to obtain meaningful conditional conditional confidence intervals after model selection (in the sense that the confidence intervals are reasonably narrow), the size of the conditioning set is critical. By means of the confidence intervals proposed by Lee et al. (2016), we showed that exact conditional coverage of $\beta^M$ on the set $\{\hat{M} = M, \hat{s}_{\hat{M}} = s_M\}$ leads to confidence intervals with infinite conditional expected length. On the other hand, exact conditional coverage of $\beta^M$ on the larger set $\{\hat{M} = M\}$ leads to confidence intervals with finite conditional expected length. Thus, if computationally feasible, one should always prefer the confidence interval $[L_{M,j}(Y), U_{M,j}(Y)]$ (that is the confidence interval with exact conditional coverage on $\{\hat{M} = M\}$) as this confidence interval will be significantly narrower. The findings, in general, suggest that conditioning on a larger set will lead to narrower confidence intervals on average.

# References

Bachoc, F., Leeb, H. and Pötscher, B. M. (2016). Valid confidence intervals for post-model-selection predictors. *arXiv preprint arXiv:1412.4605* .

Berk, R., Brown, L., Buja, A., Zhang, K. and Zhao, L. (2013). Valid post-selection inference. *The Annals of Statistics* **41**, 802–837.

Fithian, W., Sun, D. and Taylor, J. (2017). Optimal inference after model selection. *arXiv preprint arXiv:1410.2597* .

Frank, I. E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics* **35**, 109–135.

Kabaila, P. (1998). Valid confidence intervals in regression after variable selection. *Econometric Theory* **14**, 463–482.

Kabaila, P. and Leeb, H. (2006). On the large-sample minimal coverage probability of confidence intervals after model selection. *Journal of the American Statistical Association* **101**, 619–629.

Lee, J. D., Sun, D. L., Sun, Y. and Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics* **44**, 907–927.

Leeb, H. and Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory* **21**, 21–59.

Leeb, H. and Pötscher, B. M. (2006*a*). Can one estimate the conditional distribution of post-model-selection estimators?. *The Annals of Statistics* **34**, 2554–2591.

Leeb, H. and Pötscher, B. M. (2006*b*). Performance limits for estimators of the risk or distribution of shrinkage-type estimators, and some general lower risk-bound results. *Econometric Theory* **22**, 69–97.

Leeb, H. and Pötscher, B. M. (2008). Can one estimate the unconditional distribution of post-model-selection estimators?. *Econometric Theory* **24**, 338–376.

Leeb, H., Pötscher, B. M. and Ewald, K. (2015). On various confidence intervals post-model-selection. *Statistical Science* **30**, 216–227.

Lehmann, E. L. and Romano, J. P. (2006). *Testing statistical hypotheses.* 3rd edn. Springer. New York.

Meeks, S. L. and D'Agostino, R. B. (1983). A note on the use of confidence limits following rejection of a null hypothesis. *The American Statistician* **37**, 134–136.

Ohman Strickland, P. A. and Casella, G. (2003). Conditional inference following group sequential testing. *Biometrical Journal* **45**, 515–526.

Pötscher, B. M. (1991). Effects of model selection on inference. *Econometric Theory* **7**, 163–185.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* **58**, 267–288.

Tibshirani, R. J. (2013). The lasso problem and uniqueness. *Electronic Journal of Statistics* **7**, 1456–1490.

Tibshirani, R. J., Taylor, J., Lockhart, R. and Tibshirani, R. (2016). Exact post-selection inference for sequential regression procedures. *Journal of the American Statistical Association* **111**, 600–620.

Zeidler, E. (1984). *Nonlinear Functional Analysis and Its Applications: III: Variational Methods and Optimization.* Springer. New York.

# Appendix A English abstract

There is a growing interest in valid inference procedures after model selection as it has been demonstrated that inference procedures, which do not take the model selection into account, can be highly misleading. Lee et al. (2016) recently proposed two types of confidence intervals with exact coverage probability, conditional on a selection event defined by the lasso: the first confidence interval has exact coverage probability, conditional on the model selected by the lasso; the second has exact coverage probability, conditional on the selected model and the signs of the non-zero lasso coefficients. Lee et al. (2016) argued that the latter confidence interval is typically wider, as one conditions on a smaller set. Yet, the computation time of the former increases exponentially with the number of variables selected, while the computation time of the latter increases only linearly. This thesis shows that the first confidence interval, which is computationally intensive (even infeasible if the number of selected variables is large), has finite conditional expected length, whereas the second confidence interval, which is computationally feasible even if the number of selected variables is large, has infinite conditional expected length.

# Appendix B German abstract

Es gibt ein wachsendes Interesse an inferenzstatistischen Methoden, die nach Modellselektion valide sind, weil bereits mehrfach gezeigt wurde, dass Methoden, welche die Modellselektion nicht berücksichtigen, zu irreführenden Rückschlüssen führen können. Lee et al. (2016) stellten kürzlich zwei Arten von Konfidenzintervallen vor, die, bedingt auf ein Selektionsevent, welches durch die Lasso-Koeffizienten bestimmt wird, eine exakte Überdeckungswahrscheinlichkeit besitzen: beim ersten Konfidenzintervall wird dabei auf das von Lasso gewählte Modell bedingt; beim zweiten Konfidenzintervall wird auf das gewählte Modell und auf die Vorzeichen der Lasso-Koeffizienten, die nicht 0 sind, bedingt. Lee et al. (2016) argumentierten, dass das Letztere für gewöhnlich länger ist, weil die Menge, auf die bedingt wird, kleiner ist. Allerdings wächst die Rechenzeit des ersteren Konfidenzintervalls exponentiell mit der Zahl der gewählten Variablen, während die Rechenzeit des Letzteren bloß linear wächst. In dieser Arbeit wird gezeigt, dass die Länge des ersteren Konfidenzintervalls, welches rechenintensiv ist (sogar nicht berechenbar, wenn die Zahl der gewählten Variablen groß ist), einen endlichen bedingten Erwartungswert besitzt, während die Länge des letzteren Konfidenzintervalls, welches mit geringem Aufwand berechnet werden kann, einen unendlichen bedingten Erwartungswert besitzt.