CREDIBLE INTERVALS AND BOOTSTRAP CONFIDENCE INTERVALS IN MONOTONE REGRESSION

BY PIET GROENEBOOM AND GEURT JONGBLOED

Delft Institute of Applied Mathematics, Delft University of Technology
P.Groeneboom@tudelft.nl, G.Jongbloed@tudelft.nl

In the recent paper [5], a Bayesian approach for constructing confidence intervals in monotone regression problems is proposed, based on credible intervals. We view this method from a frequentist point of view, and show that it corresponds to a percentile bootstrap method of which we give two versions. It is shown that a (non-percentile) smoothed bootstrap method has better behavior and does not need correction for over- or undercoverage. The proofs use martingale methods.

1. Introduction. In many fields of application, inference on a monotone function is both natural and needed. Many examples of applications can be found in the books on shape constrained statistical inference [1], [16], [19] and [8]. Natural estimators of monotone functions exist. In order to assess the accuracy of these, there is a need for uncertainty quantification. In the frequentist setting, this can be done by (pointwise) confidence intervals for the monotone function of interest. From a Bayesian perspective, credible intervals are a common method to quantify uncertainty.

We consider the following model for our observations $D_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$, also considered in [5].

$$Y_i = f_0(X_i) + \epsilon_i, \qquad i = 1, \dots, n.$$

Here $f_0 : [0,1] \mapsto \mathbb{R}$ is monotone nondecreasing, the ε_i are i.i.d. sub-Gaussian with expectation 0 and variance σ_0^2, independent of the X_i's, and the X_i are i.i.d. with non-vanishing density g on $[0,1]$. The classical least squares estimator (LSE) $\hat f_n$ of f_0, under the condition that $\hat f_n$ is nondecreasing minimizes

$$\sum_{i=1}^n \left(Y_i - f(X_i)\right)^2 \tag{1.1}$$

over all nondecreasing functions $f : [0,1] \mapsto \mathbb{R}$. This estimator, which is also the maximum likelihood estimator if we assume the ϵ_i to be i.i.d. centered normally distributed, can be explicitly constructed based on the data. Denoting by $0 < x_1 < x_2 < \dots < x_n < 1$ the observed ordered X_i's and by y_i the corresponding Y values (so $(x_1, y_1), \dots, (x_n, y_n)$ represent the original data pairs, but with first coordinate sorted in increasing order), Lemma 2.1 in [8] shows that $\hat f_n$ can be taken piecewise constant on the intervals $(x_{i-1}, x_i]$, $1 \le i \le n$ and that $\hat f_n(x_i)$ can be obtained as the left derivative of the greatest convex minorant of the diagram consisting of the points $\{P_i : 0 \le i \le n\}$ with

$$P_0 = (0,0) \text{ and for } 1 \le i \le n,\;\; P_i = \left(\frac{i}{n}, \frac{1}{n}\sum_{j=1}^i y_j\right),$$

AMS 2000 subject classifications: Primary 62G05, 62N01; secondary 62-04.

Keywords and phrases: credible intervals, confidence intervals, Chernoff's distribution, smoothed bootstrap, monotone regression.

evaluated at the point P_i.

As is clear from this characterization, $\hat{f}_n$ will be a nondecreasing step function with its jumps concentrated on a data-dependent subset of the observed points x_i. Fixing a number of points in $[0,1]$, say $0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_m = 1$, one can also minimize (1.1) over all nondecreasing functions, piecewise constant on the intervals $I_j = (\tau_{j-1}, \tau_j]$. Then, writing $n_j = |\{i : i \in I_j\}|$ and $\bar{y}_j = (\sum_{x_i \in I_j} y_i)/n_j$, we have

$$\sum_{i=1}^{n} (y_i - f(x_i))^2 = \sum_{j=1}^{m} \sum_{x_i \in I_j} (y_i - \bar{y}_j + \bar{y}_j - f(\tau_j))^2 =$$

$$= \sum_{j=1}^{m} \sum_{x_i \in I_j} (y_i - \bar{y}_j)^2 + \sum_{j=1}^{m} (\bar{y}_j - f(\tau_j))^2 n_j, \tag{1.2}$$

where we use that for the current function class, $f(x_i) = f(\tau_j)$ for $x_i \in I_j$. As the first term in (1.2) does not involve f, minimizing it boils down to a weighted isotonic regression. The solution to this minimization problem also allows for a graphical construction. The optimal value of $f(\tau_j)$ is the left derivative, taken at the point P_j, of the greatest convex minorant of the diagram of points consisting of

$$P_0 = (0,0),\ P_j = \left(\frac{1}{n}\sum_{k=1}^{j} n_k, \frac{1}{n}\sum_{k=1}^{j} n_k \bar{y}_k\right),\ 1 \le j \le m.$$

From decomposition (1.2) it also follows that the piecewise constant function f defined by $f(x_i) = \bar{y}_j$ for $x_i \in I_j$ is the least squares estimator over the class of piecewise constant functions without imposing the restriction of monotonicity.

In [5], a Bayesian method is proposed for constructing pointwise confidence intervals for a monotone regression function, based on credible intervals. The method is proved to give overcoverage for large sample sizes, but a correction table is given in [5] to correct for the overcoverage. Purely based on the algorithm that results in the credible intervals, the approach can be seen as a particular percentile bootstrap method.

In Section 2 we describe the approach in [5] to construct confidence intervals via credible intervals. In Section 3 we give the interpretation of the credible intervals as percentile bootstrap intervals and in particular Theorem 3.1 for the bootstrap procedure, corresponding to the key Theorem 3.3 in [5] for the construction of the credible intervals. In proving Theorem 3.1 we use a martingale method.

In analogy with the Bayesian procedure, we construct the bootstrap intervals by generating normal noise variables (following [5]), using the empirical Bayes method for estimating the variance of these variables, defined in Section 3. In subsection 3.2 we define a classical bootstrap procedure, where we resample with replacement from the original data, and do not have to estimate the variance. These two methods correspond, respectively, to the "regression method" (holding the regressors X_i in the regression model fixed), and the "correlation model" (where we consider the X_i as random) in the terminology of [12]. The results of the three methods are highly similar.

It has been proved by several authors that the straightforward bootstrap is inconsistent in this situation (see, e.g., [13], [18] and [17] for results related to this phenomenon).This straightforward bootstrap uses resampling with replacement from the pairs (X_i, Y_i) and computes the monotone least squares estimator $\hat{f}_n^*$ based bootstrap samples and approximates the distribution of $n^{1/3}\left(\hat{f}_n(t_0) - f_0(t_0)\right)$ by that of the analogous 'bootstrap quantity' $n^{1/3}\left(\hat{f}_n^*(t_0) - \hat{f}_n(t_0)\right)$. The Bayesian approach and the percentile bootstrap approach circumvent this difficulty by using the convergence in distribution of the random variable (as a function of $D_n = \{(X_1, Y_n), \dots, (X_n, Y_n)\}$)

$$\mathbb{P}\left(n^{1/3}\{\hat{f}_n^*(t_0) - f_0(t_0)\} \le x \,\middle|\, D_n\right) \tag{1.3}$$

to

$$\mathbb{P}\left[\left(\frac{4\sigma_0^2 f_0'(t_0)}{g(t_0)}\right)^{1/3} \text{argmin}_{t\in\mathbb{R}}\left[W_1(t) + W_2(t) + t^2\right] \le x \,\middle|\, W_1\right] \tag{1.4}$$

see Theorem 3.3 in [5] and Theorem 3.1, where W_1 and W_2 are two independent standard two-sided Brownian motions, originating from zero. Here $\hat{f}_n^*$ is either the projection-posterior Bayes estimate (to be described in Section 2), in which case we would write

$$\Pi\left(n^{1/3}\{\hat{f}_n^*(t_0) - f_0(t_0)\} \le x \,\middle|\, D_n\right)$$

instead of (1.3), or the percentile bootstrap estimate $\hat{f}_n^*$. The limit (1.4) leads to credible intervals which asymptotically give overcoverage, which can be corrected for as described in [5].

In Section 4 an altogether different method for constructing the confidence intervals is given, where we use the smoothed (non-percentile) bootstrap. Here we keep the regressors fixed again, and resample with replacement residuals w.r.t. a smooth estimate of the regression function: the Smoothed Least Squares Estimator (SLSE). In this way, using theory from [9], consistent confidence intervals are constructed.

In fact, instead of $n^{1/3}\{\hat{f}_n^*(t_0) - f_0(t_0)\} = n^{1/3}\{\hat{f}_n^*(t_0) - \hat{f}_n(t_0) + \hat{f}_n(t_0) - f_0(t_0)\}$ we can now consider

$$n^{1/3}\{\hat{f}_n^*(t_0) - \tilde{f}_{nh}(t_0) + \hat{f}_n(t_0) - f_0(t_0)\},$$

where $\hat{f}_n^*$ is based on sampling with replacement from the residuals w.r.t. the SLSE $\tilde{f}_{nh}$ with bandwidth h of order $n^{-1/5}$. In contrast with Theorem 3.3 in [5] and Theorem 3.1 in the present paper, we now have convergence of

$$\mathbb{P}\left(n^{1/3}\{\hat{f}_n^*(t_0) - \tilde{f}_{nh}(t_0) + \hat{f}_n(t_0) - f_0(t_0)\} \le 0 \,\middle|\, D_n\right)$$

to the uniform distribution on $[0,1]$ (using the symmetry of the limit distribution of $n^{1/3}\{\hat{f}_n(t_0) - f_0(t_0)\}$), see Theorem 4.2.

For the non-percentile bootstrap, however, it is more natural to consider

$$n^{1/3}\{\hat{f}_n(t_0) - \{\hat{f}_n^*(t_0) - \tilde{f}_{nh}(t_0)\} - f_0(t_0)\}$$

(avoiding "looking up the wrong tables, backwards", see the discussion on p. 938 of [11]), for which we also get convergence to the the uniform distribution of

$$\mathbb{P}\left(n^{1/3}\{\hat{f}_n(t_0) - \{\hat{f}_n^*(t_0) - \tilde{f}_{nh}(t_0)\} - f_0(t_0)\} \le 0 \,\middle|\, D_n\right),$$

implying the consistency of the smoothed bootstrap method. This method of constructing confidence intervals seems superior in comparison to the Bayesian method and the percentile bootstrap intervals, as is suggested by our simulations of the coverage of the different methods.

2. Credible intervals. In [5], a Bayesian approach to construct confidence intervals for a monotone regression function is proposed. A prior distribution is defined on the class of functions on $[0,1]$, supported on a sieve of piecewise constant functions. More specifically, the interval $[0,1]$ is partitioned into J intervals $I_j = ((j-1)/J, j/J]$, $1 \le j \le J$. In the notation of the previous section, $\tau_j = j/n$. A draw from the prior distribution is then represented by

$$f_{\boldsymbol{\theta}} = \sum_{j=1}^{J} \theta_j 1_{I_j}, \qquad \boldsymbol{\theta} = (\theta_1, \dots, \theta_J), \tag{2.1}$$

where the θ_j are independent normal random variables with expectation ζ_j and variance $\sigma_0^2\lambda_j^2$, where $0 < \lambda_j < \infty$ (including noise variance σ_0^2 as a factor is only done for convenience in formulas to follow). Note that function (2.1) will not automatically be monotone, a requirement that would seem natural in this setting. The main reason not to impose this, is that with this prior distribution, the posterior distribution can be conveniently analytically computed. Indeed, as seen in the Appendix, the posterior distribution of $\boldsymbol{\theta}$ has independent coordinates, where θ_j has distribution

$$\theta_j \sim N\left(\frac{n_j\bar{y}_j + \zeta_j/\lambda_j^2}{n_j + 1/\lambda_j^2}, \frac{\sigma_0^2}{n_j + 1/\lambda_j^2}\right), \qquad n_j = \sum_{i=1}^{n} 1_{I_j}(x_i). \tag{2.2}$$

Here, as before, $\bar{y}_j$ is the mean of the y_i for the x_i belonging to the j-th interval I_j. As mentioned in the previous section, this corresponds to the MLE of f over the (nonrestricted) class of functions which are constant on the intervals I_j. A draw from the posterior on the set of piecewise constant functions on $[0,1]$ proceeds via (2.1), based on a draw from the posterior of $\boldsymbol{\theta}$. The resulting function will in general not be monotone, so the support of the posterior extends outside the set of monotone functions on $[0,1]$.

Following ideas of [14], [2] and [3], in [5] a draw $f_{\boldsymbol{\theta}}$ from the 'raw posterior' is subsequently projected on the set of nondecreasing functions on $[0,1]$, piecewise constant on the intervals I_j, $1 \le j \le J$, via weighted isotonic regression. This projection f_θ^* is computed using Lemma 2.1 of [8]. This boils down to computing the left derivative of the greatest convex minorant of the cusum diagram consisting of the points P_j, for $0 \le j \le J$ with $P_0 = (0,0)$ and

$$P_j = \left(\frac{1}{n}\sum_{k=1}^{j} n_k, \frac{1}{n}\sum_{k=1}^{j} n_k\theta_k\right), \; 1 \le j \le J$$

if all $n_j > 0$, see (3.2) in [5] (note that Lemma 2.1 in [8] has the condition that all weights are *strictly* positive). It is clear that computing the isotonic regression can be restricted to those j with $n_j > 0$ and that for those j with $n_j = 0$, f_θ^* can be given any value such that monotonicity is not violated.

In this procedure, various choices need to be made. One is the number of intervals J. In [5], the asymptotic bounds

$$n^{1/3} \ll J \ll n^{2/3} \tag{2.3}$$

are given and the closest integer to $n^{1/3}\log n$ is chosen in the simulations. Here the symbol "$\ll$" means "is of lower order than", as $n \to \infty$. Also the noise variance σ_0^2 needs to be dealt with. For this, [5] choose the natural empirical Bayes estimate (fixing ζ and Λ), given by

$$\hat{\sigma}_n^2 = n^{-1}(\boldsymbol{y} - \boldsymbol{B\zeta})^T(\boldsymbol{B\Lambda B}^T + \boldsymbol{I})^{-1}(\boldsymbol{y} - \boldsymbol{B\zeta}) \qquad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1^2, \dots, \lambda_J^2), \tag{2.4}$$

where $\boldsymbol{B} = (b_{ij})$ the $n \times J$ 'design matrix' with entries $b_{ij} = 1_{I_j}(x_i)$, corresponding to the regression model $\boldsymbol{y} = \boldsymbol{B\theta} + \boldsymbol{\epsilon}$ following from the representation $f_{\boldsymbol{\theta}}(x_i) = (\boldsymbol{B\theta})_i$; see the Appendix. As also shown in the Appendix, this estimate can be rewritten as

$$\begin{aligned}\hat{\sigma}_n^2 &= \frac{1}{n}\sum_{j=1}^{J}\sum_{x_i \in I_j}(y_i - \bar{y}_j)^2 + \frac{1}{n}\sum_{j=1}^{J}\frac{n_j(\bar{y}_j - \zeta_j)^2}{1 + n_j\lambda_j^2} \\ &= \frac{1}{n}\sum_{i=1}^{n}\big(y_i - \bar{f}(x_i)\big)^2 + \frac{1}{n}\sum_{j=1}^{J}\frac{n_j(\bar{y}_j - \zeta_j)^2}{1 + n_j\lambda_j^2}\end{aligned} \tag{2.5}$$

where $\bar{f}(x) = \sum_{j=1}^{J}\bar{y}_j 1_{I_j}(x)$ is the aforementioned maximum likelihood estimate of f_0 over all piecewise constant (not necessarily monotone) functions on the intervals I_j. The first term in this expression is the mean of the squared residuals of the observations with respect to $\bar{f}$. This is a quite natural estimator of the variance. The influence of the hyper parameters $\boldsymbol{\zeta}$ and $\boldsymbol{\Lambda}$ on the estimate of σ_0^2 can be inferred from the second term in the expression.

As shown in the Appendix, the empirical Bayes estimate for ζ, not taking into account monotonicity is given by

$$\bar{\zeta} = (\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_J)^T. \tag{2.6}$$

Substituting this in (2.5) makes the second term vanish. Using the empirical Bayes estimate over the monotone vectors ζ, being the isotonic regression of $\bar{\zeta}$ with weights $n_j/(1 + n_j\lambda_j^2)$ increases the empirical Bayes estimate for σ_0^2.

With the choices $\boldsymbol{\zeta} = 0$ and $\lambda_j \equiv \lambda$ made in [5], the empirical Bayes estimate for σ_0^2 becomes

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\big(y_i - \bar{f}(x_i)\big)^2 + \frac{1}{n}\sum_{j=1}^{J}\frac{n_j\bar{y}_j^2}{1 + n_j\lambda^2} \tag{2.7}$$

For relatively large values of $\lambda^2 n_j$, the second term becomes negligible to the first.

Because the density g generating the X_i's is nonvanishing on $[0, 1]$, the (random) number N_j of points in intervals of length of the order $1/J = 1/J_n$ is (in the setting of [5]) of the order n/J_n. With the restriction $n^{1/3} << J_n << n^{2/3}$, this means that N_j will be of bigger order than $n^{1/3}$; taking $J_n \approx n^{1/3}\log n$, N_j will be of order $n^{2/3}/\log n$. Therefore, for reasonable choice of λ, $\lambda^2 N_j >> 1$ with high probability when n is large.

Considering (2.2) with $\zeta_j \equiv 0$ and $\lambda_i \equiv \lambda$, fixed, as chosen in [5], a draw from the raw posterior of θ_j can be viewed as

$$\theta_j = \bar{f}(j/J) + \tilde{\epsilon}_j$$

where

$$\bar{f}(j/J) = \frac{\bar{y}_j}{1 + 1/(n_j\lambda^2)} \text{ and } \tilde{\epsilon}_j \sim^{\text{indep}} N\left(0, \frac{\sigma_0^2/n_j}{1 + 1/(n_j\lambda^2)}\right), \tag{2.8}$$

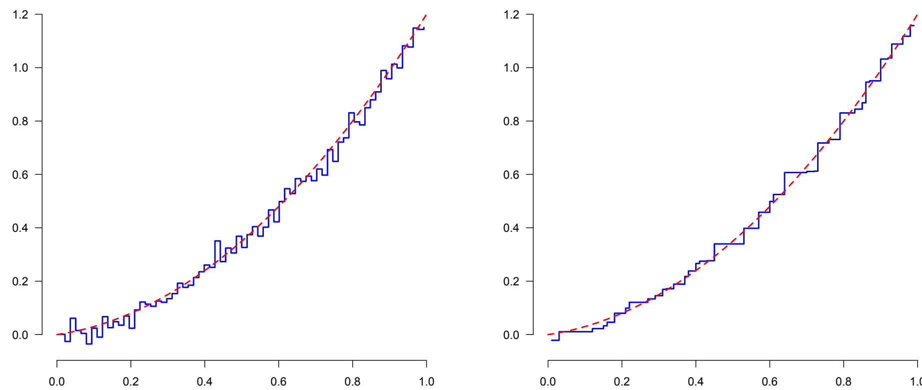
where σ_0^2 is estimated by its Empirical Bayes estimate (2.7). Again due to the restriction $n^{1/3} << J_n << n^{2/3}$, $\bar{f}$ in (2.8) is a (generally non-monotone) local average estimator of f_0. The added noise $(\tilde{\epsilon}_j)$ is normal and reflects the variance of the original $\bar{Y}_j$. This means that the draw from the projected posterior is computed as left derivative of the cumulative sum diagram consisting of the points P_j, $0 \le j \le J$ with $P_0 = (0, 0)$ and

$$P_j = \left(\sum_{k=1}^{j} n_k, \sum_{k=1}^{j} n_k\left(\frac{\bar{y}_k}{1 + 1/(n_k\lambda^2)} + \tilde{\epsilon}_k\right)\right) \tag{2.9}$$

In [5], the following example is considered:

$$f_0(x) = x^2 + x/5, \qquad x \in [0,1].$$

Here the X_i are independently uniformly distributed on $[0,1]$ and the ε_i have a normal $N(0, 0.01)$ distribution. The choices for J, ζ_j and λ_j are $J = \lfloor n^{1/3} \log n \rfloor$, $\zeta_j = 0$ and $\lambda_j = 10$, following the parametrization in the `R` code, kindly sent to us by Moumita Chakraborty. A picture of a single draw $f_{\boldsymbol{\theta}}$ from the raw posterior and its isotonic projection $f^*_{\boldsymbol{\theta}}$, for a sample of size $n = 500$ is shown in Figure 1.



(a) (b)

Fig 1: (a) A single draw $f_{\boldsymbol{\theta}}$ (blue) from the raw posterior and (b) the corresponding projection $f^*_{\boldsymbol{\theta}}$ (blue), for a sample of size $n = 1000$ and $f_0(x) = x^2 + x/5$ (red, dashed).

Now, one can generate 1000 posterior samples of $\boldsymbol{\theta} = (\theta_1, \dots, \theta_J)$ from the posterior normal distribution, specified in (2.2), and consider the $\frac{1}{2}\alpha$th and $(1 - \frac{1}{2}\alpha)$th percentiles of the isotonic projections $\hat{f}^*_{\boldsymbol{\theta}}(t)$ at a fixed point t. Would this give us, at least asymptotically, valid 95% confidence intervals for $f_0(t)$?

The question is answered in [5] by Theorem 3.3 on p. 1017: $\Pi\left(n^{1/3}\{\hat{f}^*_{\boldsymbol{\theta}}(t) - f_0(t)\} \le z | D_n\right)$ converges to a limit distribution, leading to wider intervals than in the situation in which we have the Chernoff distribution as limit. The fraction by which they become wider is given in [5].

3. Credible intervals as bootstrap percentile confidence intervals.

3.1. *The percentile bootstrap for the regression model.* In the Bayes approach, we considered random parameters θ_j, with (posterior) distribution given in (2.2). In the simulations, accompanying the paper [5], the prior parameter $\boldsymbol{\zeta}$ was taken $\boldsymbol{\zeta} = \mathbf{0}$ and $\lambda_j \equiv \lambda > 0$. Moreover, the empirical Bayes estimator $\hat{\sigma}^2_n$ was taken as estimator for σ^2_0.

With these choices, we get:

$$\theta_j \sim N\left(\frac{\bar{y}_j}{1 + 1/(n_j\lambda^2)}, \frac{\hat{\sigma}^2_n}{n_j\{1 + 1/(n_j\lambda^2)\}}\right), 1 \le j \le J, \text{ independently.}$$

This means asymptotically, in first order:

$$\theta_j \sim N\left(\bar{y}_j, \frac{\hat{\sigma}_n^2}{n_j}\right),\ 1 \le j \le J \tag{3.1}$$

if we keep λ bounded away from zero ($\lambda^2 = 100$ was taken in the simulations with paper [5]). Next the confidence intervals were determined by taking the percentiles of simulated values of the (weighted) monotonic projections of the θ_j's with distribution given by (3.1).

Algorithmically, this can be viewed as a percentile bootstrap method, where a bootstrap sample is generated by adding noise to an estimate of the regression function. The regression estimate in this setting is the (weighted) least squares estimate of f_0, piecewise constant on intervals I_j and not taking into account the monotonicity constraint (so: $\bar{y}_j$ on I_j). The noise is sampled from a centered normal distribution with estimated variance. Then, $\hat{f}_n^*$ is determined by computing the (weighted) isotonic regression based on the bootstrap dataset. Adopting the "bootstrap notation" rather than the "Bayesian notation" θ_j, define where

$$Y_j^* \sim N(\bar{Y}_j, \hat{\sigma}_n^2/N_j), \qquad j = 1, \dots, J \tag{3.2}$$

and note that given the original data, $(Y_1^*, \dots, Y_J^*) =^D (\theta_1, \dots, \theta_J)$, in view of (3.1) and (3.2). Using this notation, $\hat{f}_n^*$ is found by taking the left derivative of the convex minorant of the cusum diagram, running through the points

$$P_j^* = \left(\sum_{i=1}^j N_i, \sum_{i=1}^j N_i Y_i^*\right), \qquad j = 1, \dots, J. \tag{3.3}$$

To study the asymptotic behavior of $\hat{f}_n^*$, we define the (local) “bootstrap” process

$$\widetilde{W}_n^*(t) = n^{-1/3}\left\{\sum_{j: j/J \in [0, t_0 + n^{-1/3}t]} N_j\left\{Y_j^* - \bar{Y}_j\right\} - \sum_{j: j/J \in [0, t_0]} N_j\left\{Y_j^* - \bar{Y}_j\right\}\right\} \tag{3.4}$$

and the (local) “sample” process $\widetilde{W}_n$ by:

(3.5)

$$\widetilde{W}_n(t) = n^{-1/3}\left\{\sum_{j: j/J \in [0, t_0 + n^{-1/3}t]} 1_{\{N_j > 0\}} N_j \left\{\bar{Y}_j - \frac{n}{N_j}\int_{I_j} f_0(u)\, d\mathbb{G}_n(u)\right\}\right.$$

$$\left. - \sum_{j: j/J \in [0, t_0]} 1_{\{N_j > 0\}} N_j \left\{\bar{Y}_j - \frac{n}{N_j}\int_{I_j} f_0(u)\, d\mathbb{G}_n(u)\right\}\right\}. \tag{3.6}$$

With these definitions we have the following theorem, similar to Theorem 3.3 in [5].

THEOREM 3.1. *Let $D_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ and let $\hat{f}_n^*$ be a draw generated according to the bootstrap procedure described above. Then, for each fixed $x \in \mathbb{R}$, as $n \to \infty$,*

$$\mathbb{P}\left(n^{1/3}\{\hat{f}_n^*(t_0) - f_0(t_0)\} \le x \,\middle|\, D_n\right)$$

$$\xrightarrow{\mathscr{D}} \mathbb{P}\left[\left(\frac{4\sigma_0^2 f_0'(t_0)}{g(t_0)}\right)^{1/3} \operatorname{argmin}_{t \in \mathbb{R}}\left[W_1(t) + W_2(t) + t^2\right] \le x \,\middle|\, W_1\right] \tag{3.7}$$

where W_1 and W_2 are independent standard two-sided Brownian motions.

Note that, for $t > 0$,

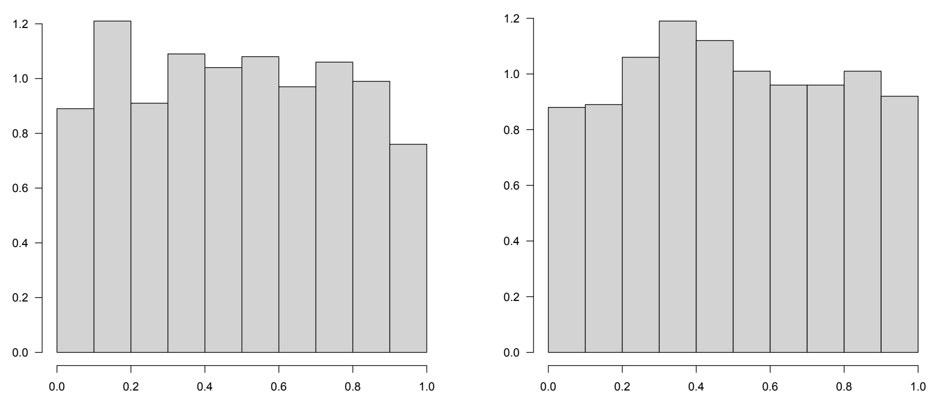
$$n^{-1/3}\left\{\sum_{j:j/J\in[0,t_0+n^{-1/3}t]} N_j\{Y_j^* - f_0(t_0)\} - \sum_{j:j/J\in[0,t_0]} N_j\{Y_j^* - f_0(t_0)\}\right\}$$

$$= \widetilde{W}_n^*(t) + \widetilde{W}_n(t) + n^{-1/3} \sum_{j:j/J\in[t_0,t_0+n^{-1/3}t]} n\int_{I_j}\{f_0(u) - f_0(t_0)\}\,d\mathbb{G}_n(u)$$

$$\sim \widetilde{W}_n^*(t) + \widetilde{W}_n(t) + \tfrac{1}{2}f_0'(t_0)g(t_0)t^2, \tag{3.8}$$

with a similar expansion for $t \le 0$.

So the percentile bootstrap estimates have the same behavior as the Bayes estimates in [5]. Histograms of estimates of the posterior probabilities for the Bayesian procedure in [5] and the corresponding conditional probabilities of the percentile bootstrap in Lemma 3.1 for varying D_n of size $n = 2000$ and $t_0 = 0.5$ are shown in Figure 2. The estimates are the relative frequencies in 1000 posterior, resp. percentile bootstrap samples for each of the original (1000) samples.



(a) (b)

Fig 2: (a) Histogram of estimate of $\Pi\{n^{1/3}\{f_n^*(t_0) - f_0(t_0)\} \le 0 \mid D_n\}$ for the projection-posterior samples in [5], (b) Histogram of estimate of $\mathbb{P}\{n^{1/3}\{\hat{f}_n^*(t_0) - f_0(t_0)\} \le 0 \mid D_n\}$ for the percentile bootstrap. Both histograms are based on 1000 samples D_n of size $n = 2000$ and $t_0 = 0.5$.

REMARK 3.1. Let $\Delta_n = \Pi\{n^{1/3}\{f_n^*(t_0) - \hat{f}_n(t_0)\} \le 0 | D_n\}$. In Figure 1 on p. 1017 of [5] three pictures of Δ_n are shown for three different sets of simulated data, where $\hat{f}_n$ is the LSE. Is is not completely clear to us how Δ_n is sampled here. Since we do not have an explict expression for Δ_n, it seems that an estimate of Δ_n has to be based on a sample of posterior draws $f^*(t_0)$. If we use such a procedure and consider the fluctuation of Δ_n as a function of D_n, we get a histogram similar to the histograms in Figure 1 of [5]. The estimates are relative frequencies in 1000 samples of size 2000. See Figure 3.

Theorem 3.1 is the consequence of the following two lemmas.

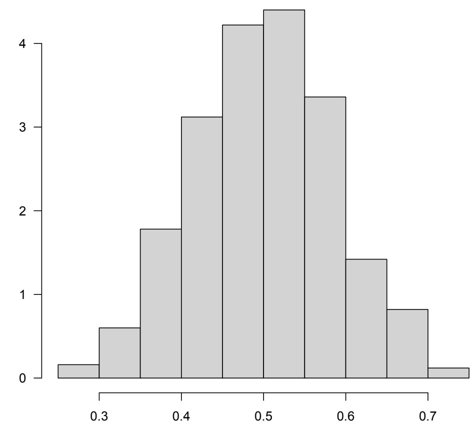


Fig 3: Histogram of estimate of $\Pi\{n^{1/3}\{f_n^*(t_0)-\hat f_n(t_0)\}\le 0 \mid D_n\}$, for varying D_n of size $n=2000$ and $t_0=0.5$. The estimates are based on relative frequencies in 1000 draws.

LEMMA 3.1. *Let W be standard two-sided Brownian motion on $\mathbb{R}$, originating from zero. Let $D(\mathbb{R})$ be the space of right continuous functions, with left limits (cadlag functions) on $\mathbb{R}$, equipped with the metric of uniform convergence on compact sets, and let $t_0\in(0,1)$. Let $\widetilde W_n^*$ be defined by (3.4). Then, along almost all sequences $(X_1,Y_1),(X_2,Y_2),\dots$, the process $\widetilde W_n^*$ defined by (3.4) converges in $D(\mathbb{R})$ in distribution conditionally to the process V, defined by*

$$V(t)=\sigma_0\sqrt{g(t_0)}\,W(t),\ t\in\mathbb{R}. \tag{3.9}$$

Here W is standard two-sided Brownian motion, originating from zero.

PROOF. We consider the case $t\ge 0$. It is clear that $t\mapsto \widetilde W_n(t)$ is a martingale with respect to the family of σ-algebras $\mathscr{F}^*_{n,t}$, $t\ge0$, defined by:

$$\mathscr{F}^*_{n,t}=\sigma\left\{(j/J,\bar y_j): j/J\in(t_0,t_0+n^{-1/3}t]\right\}. \qquad t\ge 0.$$

The quadratic variation process is, for $t\ge0$ given by:

$$\left[\widetilde W_n^*\right](t)=n^{-2/3}\sum_{j:j/J\in(t_0,t_0+n^{-1/3}t]} N_j^2\left\{\theta_j^*-\bar y_j\right\}^2\sim n^{-2/3}\sum_{j:j/J\in(t_0,t_0+n^{-1/3}t]} N_j\sigma_0^2\,.$$

If, for example as in ([5]), $J\sim n^{1/3}\log n$, we get:

$$N_j\sim n^{2/3}g(t_0)/\log n,$$

and

$$n^{-2/3}\sum_{j:j/J\in(t_0,t_0+n^{-1/3}t]} N_j\sigma_0^2\sim(\log n)^{-1}\sum_{j:j/J\in(t_0,t_0+n^{-1/3}t]} g(t_0)\sigma_0^2\sim\sigma_0^2 g(t_0)\,t.$$

The case $t<0$ is treated similarly. The result now follows from Rebolledo's theorem, see Theorem 3.6, p. 68 of [8] and [15]. $\square$

LEMMA 3.2. *Let W, t_0 and V be as defined in Lemma 3.1 and $\widetilde{W}_n$ by (3.5). Then the process $\widetilde{W}_n$ converges in $D(\mathbb{R})$ in distribution, conditionally on the sequence $X_1, X_2, \ldots$, to the process V.*

PROOF OF LEMMA 3.2. This is proved in the same way as Lemma 3.1, using that (3.5) is a martingale. □

PROOF OF THEOREM 3.1. We use the "switch relation" (see, e.g., Section 3.8 of [8] and section 5.1 of [10]; the terminology is due to Iain Johnstone to denote a construction introduced in a course given by the first author in Stanford, 1990). The bootstrap estimate f_n^* is computed as left derivative of the greatest convex minorant of cumulative sum diagram (3.3). Let the processes G_n and V_n^* be defined by

$$G_n(t) = \sum_{j/J \le t} N_j/n, \qquad V_n^*(t) = \sum_{j/J \le t} N_j Y_j^*/n, \qquad t \in [0,1].$$

Moreover, let U_n^* be defined by

$$U_n^*(a) = \text{argmin}\{t \in [0,1] : V_n^*(t) - aG_n(t)\},$$

for a in the range of f_0. Then we have the "switch relation":

$$\hat{f}_n^*(t) \ge a \iff G_n(t) \ge G_n(U_n^*(a)) \iff t \ge U_n^*(a),$$

(compare with (3.35), p. 69 of [8]). So we get if $a = f_0(t_0)$,

$$\begin{aligned}
&\mathbb{P}\left\{n^{1/3}\{\hat{f}_n^*(t_0) - f_0(t_0)\} \ge x | D_n\right\} = \mathbb{P}\left\{\hat{f}_n^*(t_0) \ge a + n^{1/3}x | D_n\right\}\\
&= \mathbb{P}\left\{U_n^*(a + n^{-1/3}x) \le t_0 | D_n\right\} = \mathbb{P}\left\{n^{1/3}\{U_n^*(a + n^{-1/3}x) - t_0\} \le 0 | D_n\right\}\\
&= \mathbb{P}\left\{\text{argmin}\left[t \in [0,1] : V_n^*(t) - (a + n^{-1/3}x)\, G_n(t)\right] \le 0 | D_n\right\}\\
&= \mathbb{P}\left\{\text{argmin}\left[t \in [0,1] : V_n^*(t) - V_n^*(t_0) - (a + n^{-1/3}x)\{G_n(t) - G_n(t_0)\}\right] \le 0 | D_n\right\},
\end{aligned}$$

where the last equality holds since the values of the argmin function do not change if we add constants to the function for which we determine the argmin.

By Lemmas 3.1 and 3.2 we get the local expansion:

$$\begin{aligned}
&\mathbb{P}\left\{n^{1/3}\{\hat{f}_n^*(t_0) - f_0(t_0)\} \ge x \;\middle|\; D_n\right\}\\
&\sim \mathbb{P}\left[\text{argmin}_t\left[\tilde{W}_n^*(t) + \tilde{W}_n(t) + \tfrac{1}{2}f_0'(t_0)g(t_0)t^2 - xg(t_0)\right] \le 0 \;\middle|\; D_n\right],
\end{aligned}$$

which (using Brownian scaling) converges in distribution to

$$\begin{aligned}
&\mathbb{P}\left[\left(\frac{4\sigma_0^2 f_0'(t_0)}{g(t_0)}\right)^{1/3} \text{argmin}_{t\in\mathbb{R}}\left[W_1(t) + W_2(t) + t^2\right] \le -x \;\middle|\; W_1\right]\\
&= \mathbb{P}\left[\left(\frac{4\sigma_0^2 f_0'(t_0)}{g(t_0)}\right)^{1/3} \text{argmin}_{t\in\mathbb{R}}\left[W_1(t) + W_2(t) + t^2\right] \ge x \;\middle|\; W_1\right].
\end{aligned}$$

The last line uses the type of symmetry used in the proof of Theorem 5.2 of [10]. □

In the proof of Theorem 3.1 we use the tightness of $n^{1/3}\{U_n^*(a + n^{-1/3}x) - t_0\}$, which can be proved along entirely similar lines as the proof of Lemma 3.5 in [8].

3.2. *Convergence of a classical percentile bootstrap.* It is of interest to investigate what happens if we perform a classical empirical bootstrap, where we resample with replacement from the pairs (X_i, Y_i). This situation, where we also treat the X_i as random from the start instead of keeping them fixed, is called the "correlation model" in [12]. In this case we compute the local means

$$\bar{Y}_j^* = (N_j^*)^{-1} \sum_{X_i^* \in I_j} Y_i^*, \qquad I_j = ((j-1)/J, j/J], \qquad N_j^* = \#\{i : X_i^* \in I_j\}, \tag{3.10}$$

where the (X_i^*, Y_i^*) are (discretely) uniformly (re-)sampled with replacement from the set $D_n = \{(X_1, Y_i), \ldots, (X_n, Y_n)\}$. If $N_j^* = 0$ we define $\bar{Y}_j^* = 0$, these values play no role in the isotonization step.

Note that we can write alternatively, if $N_j^* > 0$:

$$\bar{Y}_j^* = (N_j^*)^{-1} \sum_{X_i \in I_j} M_{in}^* Y_i, \qquad N_j^* = \sum_{i : X_i \in I_j} M_{in}^*, \tag{3.11}$$

where

$$(M_{1n}^*, \ldots, M_{nn}^*) \sim \text{Multinomial}\left(n; n^{-1}, \ldots, n^{-1}\right).$$

This means that

$$\mathbb{E}\left\{N_j^* \bar{Y}_j^* \mid (X_1, Y_1), \ldots, (X_1, Y_n)\right\} = N_j \bar{Y}_j, \qquad j = 1, \ldots, J.$$

The points of the cusum diagram needed to compute the bootstrap realization of the LSE are given by

$$P_j^* = \left(\sum_{i=1}^{j} N_i^*, \sum_{i=1}^{j} N_i^* \bar{Y}_i^*\right), \qquad j = 1, \ldots, J.$$

In order to study the local asymptotics of the greatest convex minorant of this diagram, we consider the process

$$\begin{aligned} \widetilde{W}_n^*(t) = n^{-1/3} \Bigg\{ & \sum_{j : j/J \in [0, t_0 + n^{-1/3} t]} \sum_{X_i \in I_j} (M_{in}^* - 1)(Y_i - a_0) \\ & - \sum_{j : j/J \in [0, t_0]} \sum_{X_i \in I_j} (M_{in}^* - 1)(Y_i - a_0) \Bigg\}, \end{aligned} \tag{3.12}$$

where $a_0 = f_0(t_0)$, and

$$\widetilde{W}_n(t) = n^{-1/3} \Bigg\{ \sum_{j : j/J \in [0, t_0 + n^{-1/3} t]} \sum_{X_i \in I_j} (Y_i - a_0) - \sum_{j : j/J \in [0, t_0]} \sum_{X_i \in I_j} (Y_i - a_0) \Bigg\}.$$

Defining

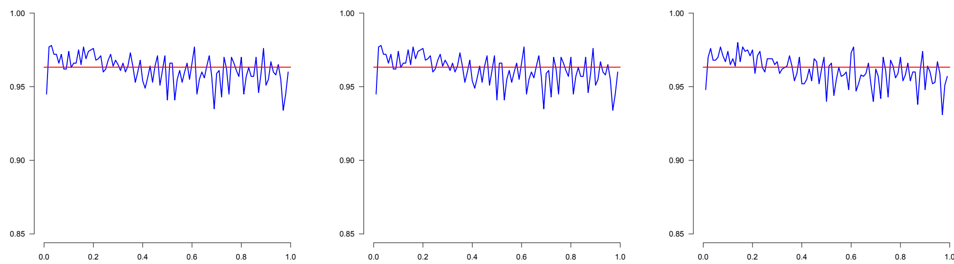
$$U_n(a) = \operatorname{argmin}\Bigg\{ t \in [0,1] : n^{-1} \sum_{j : j/J \le t} \left\{ N_j^* \bar{Y}_j^* - a\, N_j^* \right\} \Bigg\},$$

results analogous to Lemmas 3.1 and 3.2 hold. For example we get Lemma 3.3 (the analogue to Lemma 3.1) which is proved in the Appendix.

LEMMA 3.3. *Let W, t_0 and V be as defined in Lemma 3.1 and $\widetilde{W}_n^*$ be defined by (3.12). Then, along almost all sequences $(X_1, Y_1), (X_2, Y_2), \ldots$, the process $\widetilde{W}_n^*$ converges in $D(\mathbb{R})$ in distribution conditionally to the process V*

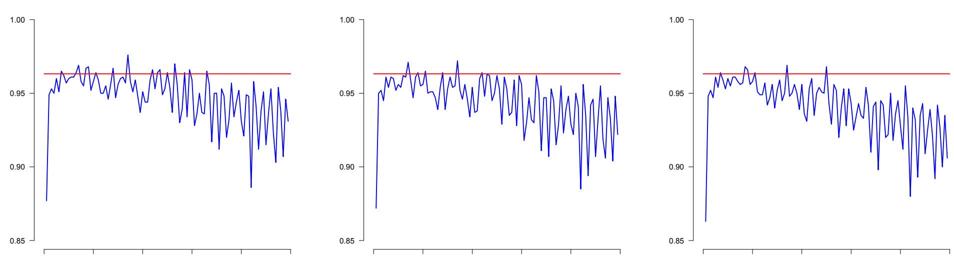
So we get the same behavior as in subsection 3.1, but the present approach has the interesting feature that we do not have to estimate the variance of the errors separately. We can just resample with replacement from the original sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ and compute the estimator $\hat{f}_n^*$ in the bootstrap samples.

The simulations, based on the regression function $f(x) = x^2 + x/5$ with normal noise with expectation 0 and variance 0.01 show almost no difference beween the three methods if $n = 20,000$, see Figure 4. At smaller sample size, like, e.g., $n = 1000$, the overcoverage is still not reached, as can be seen in Figure 5. So the phenomenon of overcoverage also here only seems to occur with very large sample sizes.



(a) (b) (c)

Fig 4: Coverage percentages for credible intervals and percentile confidence intervals, for $n = 20,000$. (a) credible intervals, (b) percentile confidence intervals of section 3.1, (c) percentile confidence intervals of section 3.2. The red line is at level 0.96324, but the intervals are based on the 0.025 and 0.975 quantiles of the credible, respectively percentile bootstrap simulations. The level 0.96324 was determined from the values of the function A in [5].



(a) (b) (c)

Fig 5: Coverage percentages for credible intervals and percentile confidence intervals, for $n = 1000$. (a) credible intervals, (b) percentile confidence intervals of section 3.1, (c) percentile confidence intervals of section 3.2. The red line is at level 0.96324, but the intervals are based on the 0.025 and 0.975 quantiles of the credible, respectively percentile bootstrap simulations. The level 0.96324 was determined from the values of the function A in [5].

4. Cube root n consistent smoothed bootstrap confidence intervals. In [18] it was shown for interval censoring models that cube root n convergent bootstrap confidence intervals can be computed for the distribution function at a fixed point with the right asymptotic coverage. Key in this, is the convergence of the nonparametric maximum likelihood estimator to Chernoff's distribution. We show that a similar approach is possible in the present context.

In the regression context this means that we use, as in [9], the smoothed least squares estimator (the SLSE) $\tilde f_{nh}$, for a bandwidth $h = cn^{-1/5}$. To define $\tilde f_{nh}$, let K be a symmetric twice continuously differentiable nonnegative kernel with support $[-1,1]$ such that $\int K(u)\,du = 1$. Let $h > 0$ be a bandwidth and define the scaled kernel K_h by

$$K_h(u) = \frac{1}{h}K\left(\frac{u}{h}\right), \;\; u \in \mathbb{R}. \tag{4.1}$$

The SLSE $\tilde f_{nh}$ is then for $t \in [h, 1-h]$ defined by

$$\tilde f_{nh}(t) = \int K_h(t-x)\,\hat f_n(x)\,dx. \tag{4.2}$$

For $t \notin [h, 1-h]$ we use the boundary correction, defined in [9] (see (2.6) and (2.7) in [9]).

We now generate residuals E_i with respect to (4.2), defined by

$$E_i = Y_i - \tilde f_{nh}(X_i), \qquad i = 1, \dots, n,$$

and compute the centered residuals $\tilde E_i$,

$$\tilde E_i = E_i - n^{-1}\sum_{j=1}^n E_j, \qquad i = 1, \dots, n. \tag{4.3}$$

From these residuals, we generate bootstrap samples

$$(X_i, Y_i^*), \qquad Y_i^* = \tilde f_{nh}(X_i) + \tilde E_i^*, \qquad i = 1, \dots, n, \tag{4.4}$$

where the $\tilde E_i^*$ are drawn uniformly with replacement from the residuals $\tilde E_i$ defined by (4.3). For the bootstrap samples (4.4), we compute the monotone (non-smoothed) LSE $\hat f_n^*$ and consider the differences

$$\hat f_n^*(t) - \tilde f_{nh}(t), \tag{4.5}$$

and the 95% bootstrap confidence intervals, given by

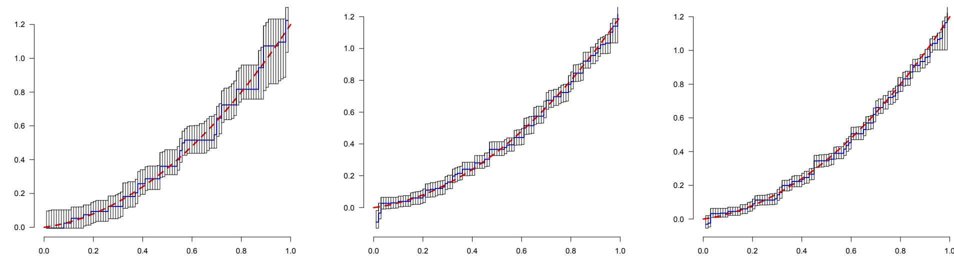
$$\left(\hat f_n(t) - Q^*_{0.975}, \hat f_n(t) - Q^*_{0.025}\right), \tag{4.6}$$

where $Q^*_{0.025}$ and $Q^*_{0.975}$ are the 2.5th and 97.5th percentiles of 1000 bootstrap samples of (4.5) and $\hat f_n$ is the LSE in the original sample. Note that this is the more conventional bootstrap approach, rather than the percentile method.

The SLSE $\tilde f_{nh}$. which plays a central role in the construction of these confidence intervals, has the limit behavior specified in the following theorem, which is Theorem 1 in [9].

THEOREM 4.1. *Let f_0 be a nondecreasing continuous function on $[0,1]$. Let $X_1, X_2, \dots$ be i.i.d random variables with continuous density g, staying away from zero on $[0,1]$, where the derivative g' is continuous and bounded on $(0,1)$. Furthermore, let $\varepsilon_1, \varepsilon_2 \dots$ be i.i.d. random variables distributed according to a sub-Gaussian distribution with expectation zero and variance $0 < \sigma_0^2 < \infty$, independent of the X_i's. Then consider Y_i, defined by*

$$Y_i = f_0(X_i) + \varepsilon_i, \;\; i = 1, 2, \dots$$



(a) (b) (c)

Fig 6: Confidence intervals, based on (4.6), for $n = 100, 500$ and 1000. The blue curve is the nonparametric monotone LSE, and the red dashed curve the real regression function $f_0(x) = x^2 + x/5$. (a) $n = 100$, (b) $n = 500$, (c) $n = 1000$.

Suppose $t_0 \in (0,1)$ such that f_0 has a strictly positive derivative and a continuous second derivative $f_0''(t_0) \neq 0$ at t_0. Then, for the SLSE $\tilde f_{nh}$ defined by (4.2) based on the pairs $(X_1,Y_1),\dots,(X_n,Y_n)$, and $h \sim cn^{-1/5}$ for $c > 0$,

$$n^{2/5}\left\{\tilde f_{nh}(t_0) - f_0(t_0)\right\} \xrightarrow{\mathscr{D}} N(\beta,\sigma^2).$$

Here

$$\beta = \tfrac12 c^2 f_0''(t_0)\int u^2 K(u)\,du \quad \text{and} \quad \sigma^2 = \frac{\sigma_0^2}{cg(t)}\int K(u)^2\,du. \tag{4.7}$$

The asymptotically Mean Squared Error optimal constant c is given by:

$$c = \left\{\frac{\sigma_0^2}{g(t_0)}\int K(u)^2\,du \Big/ \left\{f_0''(t_0)\int u^2K(u)\,du\right\}^2\right\}^{1/5}.$$

We have the following lemma of which the proof is given in the Appendix.

LEMMA 4.1. *Let W, t_0 and V be as defined in Lemma 3.1 and $\widetilde W_n^*$ be defined by:*

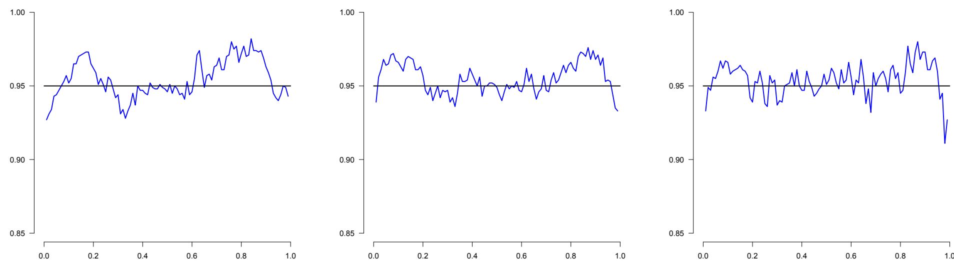
$$\widetilde W_n^*(t) = n^{-1/3}\left\{\sum_{i:X_i\in[0,t_0+n^{-1/3}t]}\tilde E_i^* - \sum_{i:X_i\in[0,t_0]}\tilde E_i^*\right\}, \qquad t \in \mathbb{R},$$

where $\tilde E_i^*$ is drawn with replacement from the residuals $\tilde E_i$, defined by (4.3). Then, along almost all sequences $(X_1,Y_1),(X_2,Y_2),\dots$, the process $\widetilde W_n^*$ converges in $D(\mathbb{R})$ in distribution conditionally to the process V

This leads to the following corollary.

COROLLARY 4.1. *Let the bootstrap LSE f_n^* be constructed as in (4.5) and let t_0 and the SLSE $\tilde f_{nh}$ be defined as in Theorem 4.1. Then, along almost all sequences $(X_1,Y_1),\dots$, we have, under the condtions of Theorem 4.1,*

$$n^{1/3}\{\hat f_n^*(t_0) - \tilde f_{nh}(t_0)\} \xrightarrow{\mathscr{D}} Z,$$



(a) (b) (c)

Fig 7: Coverage percentages for confidence intervals, based on (4.6), for $n = 100, 500$ and 1000. (a) $n = 100$, (b) $n = 500$, (c) $n = 1000$.

where

$$Z = \left(\frac{4\sigma_0^2 f_0'(t_0)}{g(t_0)}\right)^{1/3} \text{argmin}_{t\in\mathbb{R}}\left[W(t) + t^2\right],$$

and W is standard two-sided Brownian motion, originating from zero.

PROOF. We use the "switch relation" again (see the proof of Theorem 3.1). Let G_n be the empirical distribution function of the X_i and let V_n^* be defined by

$$V_n^*(t) = \sum_{X_i \le t} Y_i^*,$$

where Y_i^* is defined by (4.4). Let U_n^* be defined by

$$U_n^*(a) = \text{argmin}\{t \in [0,1] : V_n^*(t) - aG_n(t)\},$$

for a in the range of f_0. Then we have the "switch relation":

$$\hat f_n^*(t) \ge a \iff G_n(t) \ge G_n(U_n^*(a)) \iff t \ge U_n^*(a).$$

Hence we get if $a_n = \tilde f_{nh}(t_0)$ and $D_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$,

$$\begin{aligned}
&\mathbb{P}\left\{n^{1/3}\{\hat f_n^*(t_0) - \tilde f_{nh}(t_0)\} \ge x | D_n\right\} = \mathbb{P}\left\{\hat f_n^*(t_0) \ge a_n + n^{-1/3}x | D_n\right\}\\
&= \mathbb{P}\left\{U_n^*(a_n + n^{-1/3}x) \le t_0 | D_n\right\} = \mathbb{P}\left\{n^{1/3}\{U_n^*(a_n + n^{-1/3}x) - t_0\} \le 0 | D_n\right\}\\
&= \mathbb{P}\left\{\text{argmin}\left[t \in [0,1] : V_n^*(t) - (a_n + n^{-1/3}x)\, G_n(t)\right] \le 0 | D_n\right\}\\
&= \mathbb{P}\left\{\text{argmin}\left[t \in [0,1] : V_n^*(t) - V_n^*(t_0) - (a_n + n^{-1/3}x)\{G_n(t) - G_n(t_0)\}\right] \le 0 | D_n\right\},
\end{aligned}$$

where the last equality holds since the values of the argmin function do not change if we add constants to the function for which we determine the argmin.

We have:

$$\begin{aligned}
&\mathbb{P}\left\{\text{argmin}\left[t \in [0,1] : V_n^*(t) - V_n^*(t_0) - (a_n + n^{-1/3}x)\{G_n(t) - G_n(t_0)\}\right] \le 0 | D_n\right\}\\
&= \mathbb{P}\left\{\text{argmin}_t\left[W_n^*(t) + n^{1/3}\int_{(t_0, t_0 + n^{-1/3}t]} \tilde f_{nh}(u)\, dG_n(u)\right.\right.
\end{aligned}$$

$$-n^{1/3}(\tilde{f}_{nh}(t_0)+n^{-1/3}x)\{G_n(t_0+n^{-1/3})-G_n(t_0)\}\Big] \le 0|D_n\Big\}$$
$$\sim \mathbb{P}\left\{\text{argmin}_t\left[W_n^*(t)+\tfrac{1}{2}f_{nh}'(t_0)g_0(t_0)t^2-xg(t_0)t\right]\le 0|D_n\right\}$$
$$\sim \mathbb{P}\left\{\text{argmin}_t\left[W_n^*(t)+\tfrac{1}{2}f_0'(t_0)g_0(t_0)t^2-xg(t_0)t\right]\le 0|D_n\right\}$$
$$\longrightarrow \mathbb{P}\left\{\text{argmin}_t\left[\sigma_0\sqrt{g(t_0)}\,W(t)+\tfrac{1}{2}f_0'(t_0)g(t_0)t^2-xg(t_0)t\right]\le 0\right\},$$

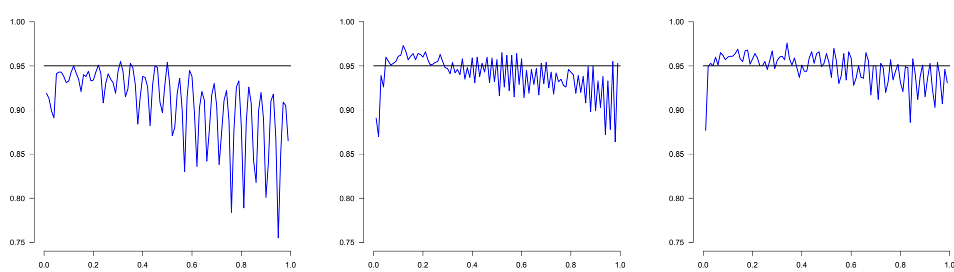
where we use Lemma 4.1 in the last step. By Brownian scaling, we can write

$$\mathbb{P}\left\{\text{argmin}_t\left[\sigma_0\sqrt{g(t_0)}\,W(t)+\tfrac{1}{2}f_0'(t_0)g_0(t_0)t^2-xg(t_0)t\right]\le 0\right\}$$
$$=\mathbb{P}\left\{\left(\frac{4\sigma_0^2 f_0'(t_0)}{g(t_0)}\right)^{1/3}\text{argmin}_t\left[W(t)+t^2\right]\le -x\right\}$$
$$=\mathbb{P}\left\{\left(\frac{4\sigma_0^2 f_0'(t_0)}{g(t_0)}\right)^{1/3}\text{argmin}_t\left[W(t)+t^2\right]\ge x\right\}.$$

□

REMARK 4.1. Corollary 4.1 shows that the limit distribution of $n^{1/3}\{\hat{f}_n^*(t_0)-\tilde{f}_{nh}(t_0)\}$, along almost all sequences D_n, is the same as the limit distributionof $n^{1/3}\{\hat{f}_n(t_0)-\tilde{f}_0(t_0)\}$. In the latter case the limit distribution was first derived in [4].

Using this corollary it is clear that in using this method the (ordinary, not percentile) smoothed bootstrap simulations recreate the actual asymptotic distribution correctly, and that we do not have to use a correction for over- or undercoverage. It is also clear from Figure 7 that its behavior is much better than the behavior of the confidence intervals in the preceding section. Even for sample size $n = 100$ the confidence intervals are more or less "on target". In comparison, the credible intervals are still far off the target for these sample sizes, and the overcoverage has still not set in for these sample sizes, see Figure 8.



(a) (b) (c)

Fig 8: Coverage percentages for the confidence intervals, for (a) $n = 100$, (b) $n = 500$, (c) $n = 1000$.

We can also prove the result below, illustrating the fact that there is no need for correction for over-or under coverage.

THEOREM 4.2. *Let $D_n = \{(X_1,Y_1),\dots,(X_n,Y_n)\}$ and $h \asymp n^{-1/5}$, and let t_0 and $\tilde f_{nh}$ be defined as in Corollary 4.1. Let the conditions of Theorem 4.1 be satisfied. Then, for $z \in (0,1)$, as $n \to \infty$,*

$$\mathbb{P}\left\{\mathbb{P}\left(n^{1/3}\{\hat f_n^*(t_0) - \tilde f_{nh}(t_0) + \hat f_n(t_0) - f_0(t_0)\} \le 0 \,\middle|\, D_n\right) \le z\right\} \longrightarrow z. \tag{4.8}$$

PROOF. Let

$$U_n = n^{1/3}\{\hat f_n^*(t_0) - \tilde f_{nh}(t_0)\}, \qquad V_n = n^{1/3}\{\hat f_n(t_0) - f_0(t_0)\},$$

and let Φ be the distribution function of Z, defined in Corollary 4.1. We have

$$\begin{aligned}
&\mathbb{P}\left\{\mathbb{P}\left(n^{1/3}\{\hat f_n^*(t_0) - \tilde f_{nh}(t_0) + \hat f_n(t_0) - f_0(t_0)\} \le 0 \,\middle|\, D_n\right) \le z\right\}\\
&= \mathbb{P}\left\{\mathbb{P}\{V_n \le -U_n | D_n\} \le z\right\}\\
&\sim \mathbb{P}\left\{\Phi(-U_n)\} \le z\right\} = \mathbb{P}\left\{-U_n \le \Phi^{-1}(z)\right\} \longrightarrow \Phi\circ\Phi^{-1}(z) = z,
\end{aligned}$$

using the symmetry around zero of the distribution of Z (the limit in distribution of U_n). □

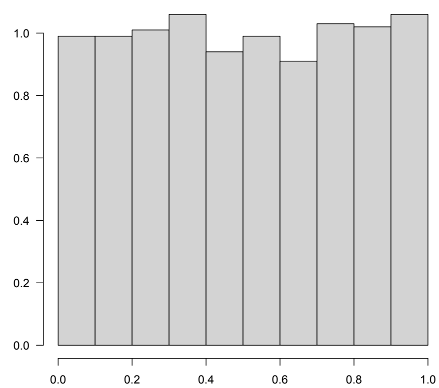


Fig 9: Histogram of 1000 estimates of $\mathbb{P}\left(n^{1/3}\{\hat f_n^*(t_0) - \tilde f_{nh}(t_0) + \hat f_n(t_0) - f_0(t_0)\} \le 0 \mid D_n\right)$ for the smoothed bootstrap, for 1000 samples D_n of size $n = 2000$ and $t_0 = 0.5$.

Note that $\hat f_n^*(t_0)$ is now centered by $\tilde f_{nh}(t_0)$ instead of $\hat f_n$, and that $n^{1/3}\{\hat f_n^*(t_0) - \tilde f_{n,h}(t_0)\}$ tends to the right limit distribution, in contrast with $n^{1/3}\{\hat f_n^*(t_0) - \hat f_n(t_0)\}$. The histogram of estimates of $\mathbb{P}\big(n^{1/3}\{\hat f_n^*(t_0) - \tilde f_{nh}(t_0) + \hat f_n(t_0) - f_0(t_0)\} \le 0 \mid D_n\big)$, based on relative frequencies, is shown in Figure 9.

For the ordinary (non-percentile) bootstrap we get by an entirely similar proof, in which we do not need the symmetry of the limit distribution:

THEOREM 4.3. *Let $D_n = \{((X_1,Y_1),\dots,(X_n,Y_n)\}$ and $h \asymp n^{-1/5}$. Let the conditions of Theorem 4.1 be satisfied. Then, for $z \in (0,1)$, as $n \to \infty$,*

$$\mathbb{P}\left\{\mathbb{P}\left(\hat f_n^*(t_0) - \tilde f_{nh}(t_0) \le \hat f_n(t_0) - f_0(t_0) \,\middle|\, D_n\right) \le z\right\} \longrightarrow z. \tag{4.9}$$

The result gives an interesting consequence of what it means to say that the "bootstrap works". This phenomenon also occurs in the simple bootstrap setting where one resamples with replacement from samples $U_1,\dots,U_n$ from a normal $N(\mu,\sigma^2)$ distribution with the aim to construct a confidence set for the mean. Then also, for all $z\in(0,1)$,

$$\mathbb{P}\left\{\mathbb{P}\left(\bar U_n^*-\bar U_n\le\bar U_n-\mu\,\middle|\,U_1,\dots U_n\right)\le z\right\}\longrightarrow z. \tag{4.10}$$

5. Concluding remarks. We showed that the construction of pointwise credible intervals for the monotone regression function, as proposed in [5], has an interpretation as the construction of percentile bootstrap intervals. The overcoverage, as explained by Theorem 3.1, only sets in for very large sample sizes, like $n=20,000$; for smaller sample sizes we have observed undercoverage.

Because the confidence intervals are based on piecewise constant estimates of the regression function, on intervals of equal length, the effect of bias is very pronounced, which does not hold for the confidence intervals, based on the smoothed bootstrap in Section 4. The latter confidence intervals have the further advantages of being on target also for smaller sample sizes, and not needing correction for overcoverage or undercoverage, since the estimates are consistent.

The consistency is also borne out by Theorem 4.3, showing convergence to the uniform distribution of $\mathbb{P}(\hat f_n^*(t_0)-\tilde f_{nh}(t_0)\le\hat f_n(t_0)-f_0(t_0)|D_n)$, $D_n=\{(X_1,Y_1),\dots,(X_n,Y_n)\}$, for the smoothed bootstrap estimates $\hat f_n^*$, in contrast with the situation for the credible intervals and the percentile bootstrap intervals, where we need a correction for convergence of $\mathbb{P}(\hat f_n^*(t_0)-f_0(t_0)\le 0\,\big|\,D_n)$ to a distribution different from the uniform distribution.

As shown in [9], it is also possible to use the smoothed least squares estimator (SLSE) directly as the basis for confidence intervals. In this case, resampling is done from residuals w.r.t. an oversmoothed estimate of the regression function to treat the bias in the right way. The bias is in this case much more of a problem because the variance and squared bias of the SLSE are of the same order if the bandwidth is of order $n^{-1/5}$. A picture of confidence intervals of this type is given in Figure 10. For more details, see [9].

All simulations in our paper can be recreated using the `R` scripts in [7].

Appendix. DERIVATION OF RAW POSTERIOR DISTRIBUTION (2.2)
Considering $\boldsymbol{X}$ fixed, we have $\boldsymbol{Y}=\boldsymbol{B\theta}+\boldsymbol{\epsilon}$, where $\boldsymbol{B}=(b_{ij})$ with $b_{ij}=1_{I_j}(X_i)$ giving $\boldsymbol{Y}|\boldsymbol{\theta}\sim N_n(\boldsymbol{B\theta},\sigma_0^2I_n)$. Writing $\boldsymbol{\Lambda}=\mathrm{Diag}(\lambda_j^2)$, this can be combined with the prior $\boldsymbol{\theta}\sim N_J(\boldsymbol{\zeta},\sigma_0^2\Lambda)$ yielding

$$f(\boldsymbol{\theta}|\boldsymbol{Y})\propto f(\boldsymbol{Y}|\boldsymbol{\theta})f(\boldsymbol{\theta})\propto\exp\left(-\frac{1}{2\sigma_0^2}\left[\boldsymbol{\theta}^T(\boldsymbol{\Lambda^{-1}}+\mathbf{B^TB})\boldsymbol{\theta}-\mathbf{2}\boldsymbol{\theta}^\mathbf{T}(\mathbf{B^TY}+\boldsymbol{\Lambda}^{\mathbf{-1}}\boldsymbol{\zeta})\right]\right)$$

By 'completing the square', this function can be seen to be proportional to the normal density with covariance matrix

$$\sigma_0^2\left(\boldsymbol{\Lambda}^{-1}+\boldsymbol{B}^T\boldsymbol{B}\right)^{-1}=\sigma_0^2\mathrm{Diag}(1/\lambda_j^2+n_j)^{-1}=\mathrm{Diag}\left(\frac{\sigma_0^2}{1/\lambda_j^2+n_j}\right)$$

where we use that both $\boldsymbol{\Lambda}$ and $\boldsymbol{B}^T\boldsymbol{B}$ are diagonal matrices with diagonal entries λ_j^2 and $n_j=\sum_i b_{ij}=\#\{i:x_i\in I_j\}$ respectively. The expectation is given by

$$\left(\boldsymbol{\Lambda}^{-1}+\boldsymbol{B}^T\boldsymbol{B}\right)^{-1}(\boldsymbol{B}^T\boldsymbol{Y}+\boldsymbol{\Lambda}^{-1}\boldsymbol{\zeta})=\mathrm{Diag}\left(\frac{1}{1/\lambda_j^2+n_j}\right)\begin{pmatrix}\zeta_1/\lambda_1^2+\sum_{\{i:x_i\in I_1\}}y_i\\ \vdots\\ \zeta_J/\lambda_J^2+\sum_{\{i:x_i\in I_J\}}y_i\end{pmatrix},$$

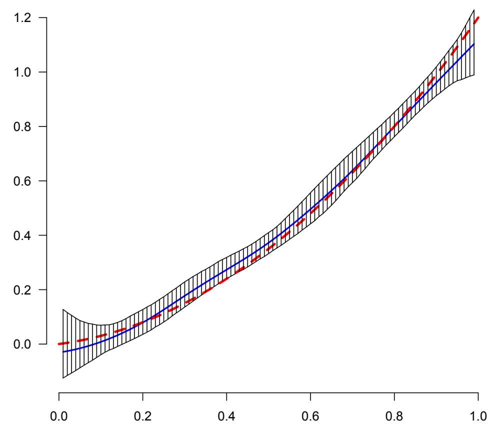


Fig 10: The SLSE (blue, solid) and 95% confidence intervals, using the theory in [9], for sample size $n = 100$ and $f_0(x) = x^2 + x/5$ (red dashed curve).

boiling down to the expression in (2.2).

DERIVATION OF (2.5)
First note that $I + B\Lambda B^T$ is a block diagonal matrix, where block j, has size $n_j \times n_j$, diagonal elements $1 + \lambda_j^2$ and off-diagonal elements λ_j^2 for $1 \le j \le J$. The j-th block can be written as

$$A_j = I_{n_j \times n_j} + \lambda_j^2 \mathbf{1}_{n_j} \mathbf{1}_{n_j}^T,$$

where $I_{k\times k}$ is the $k \times k$ identity matrix and $\mathbf{1}_k$ is the column vector of length k with all elements equal to one.

This means that $(I + B\Lambda B^T)^{-1}$ is also a block diagonal matrix, with j-th block

$$A_j^{-1} = I_{n_j \times n_j} - \frac{\lambda_j^2}{1 + n_j \lambda_j^2} \mathbf{1}_{n_j} \mathbf{1}_{n_j}^T,$$

by the Sherman Morrison formula. For convenience, write $z = y - B\zeta$ and denote by subscript $[j]$ the part of a vector corresponding to the j-th block (so i for which $x_i \in I_j$; $z_{[j]}$ has length n_j). Then

$$\left[(I + B\Lambda B^T)^{-1} z\right]_{[j]} = A_j^{-1} z_{[j]} = z_{[j]} - \frac{\lambda_j^2}{1 + n_j \lambda_j^2} \mathbf{1}_{n_j} \mathbf{1}_{n_j}^T z_{[j]} = z_{[j]} - \frac{\lambda_j^2 n_j \bar{z}_{[j]}}{1 + n_j \lambda_j^2} \mathbf{1}_{n_j},$$

where $\bar{z}_{[j]}$ denotes the average of the entries of $z_{[j]}$. Therefore

$$z_{[j]}^T \left[(I + B\Lambda B^T)^{-1} z\right]_{[j]} = z_{[j]}^T z_{[j]} - \frac{\lambda_j^2 n_j^2 \bar{z}_{[j]}^2}{1 + n_j \lambda_j^2}.$$

Hence,

$$z^T (I + B\Lambda B^T)^{-1} z = \sum_{j=1}^{J} z_{[j]}^T \left[(I + B\Lambda B^T)^{-1} z\right]_{[j]} = z^T z - \sum_{j=1}^{J} \frac{n_j \bar{z}_{[j]}^2}{1 + 1/(n_j \lambda_j^2)}. \tag{5.1}$$

Now write $(1+1/(n_j\lambda_j^2))^{-1} = 1-\delta_j$, so $\delta_j = (1+n_j\lambda_j^2)^{-1}$. Then (5.1) can be further rewritten as

(5.2)

$$
\begin{aligned}
z^T(I+B\Lambda B^T)^{-1}z &= z^Tz - \sum_{j=1}^J (1-\delta_j)n_j\bar z_{[j]}^2 = z^Tz - \sum_{j=1}^J n_j \bar z_{[j]}^2 + \sum_{j=1}^J \delta_j n_j \bar z_{[j]}^2 \\
&= \sum_{j=1}^J \sum_{x_i\in I_j} \left((y_i-\zeta_j)^2 - (\bar y_j - \zeta_j)^2\right) + \sum_{j=1}^J \delta_j n_j (\bar y_j - \zeta_j)^2 \\
&= \sum_{j=1}^J \sum_{x_i\in I_j} \left(y_i^2 - \bar y_j^2\right) + \sum_{j=1}^J \delta_j n_j (\bar y_j - \zeta_j)^2 = \sum_{j=1}^J \sum_{x_i\in I_j} (y_i-\bar y_j)^2 + \sum_{j=1}^J \delta_j n_j (\bar y_j - \zeta_j)^2.
\end{aligned}
$$

Substituting $\delta_j = (1+n_j\lambda_j^2)^{-1}$ and writing the first term as one sum, yields (2.5).

DERIVATION OF EMPIRICAL BAYES ESTIMATORS (2.6) AND (2.4) The distribution of the observed $\boldsymbol{Y}$ can be expressed in terms of the parameters σ_0^2, Λ and ζ,

$$\boldsymbol{Y} = \boldsymbol{B\theta} + \epsilon = \boldsymbol{B}(\boldsymbol{\zeta} + \sigma_0\tilde\epsilon) + \sigma_0\epsilon = \boldsymbol{B}\zeta + \sigma_0(\boldsymbol{B}\tilde\epsilon + \epsilon),$$

where $\epsilon \sim N_n(0, I_{n\times n})$ and $\tilde\epsilon \sim N_J(0,\Lambda)$ are independent. Therefore,

$$\boldsymbol{Y} \sim N_n\left(\boldsymbol{B}\zeta, \sigma_0^2\left(I_{n\times n} + \boldsymbol{B\Lambda B}^T\right)\right).$$

Maximizing the likelihood in ζ, for fixed values of σ_0^2 and Λ entails minimizing

$$(y-B\zeta)^T\left(I_{n\times n} + \boldsymbol{B\Lambda B}^T\right)^{-1}(y-B\zeta).$$

Recognizing (5.2) in this expression, it is clear that the empirical Bayes estimate of ζ is either given by the vector $(\bar y_1,\ldots,\bar y_J)^T$ if the likelihood is maximized over $\mathbb{R}^J$ or its isotonic regression with weights $n_j\delta_j = n_j/(1+n_j\lambda_j^2)$ if monotonicity is taken into account.

For any fixed value of ζ, maximizing the the log likelihood of σ_0 corresponds to minimizing

$$\frac{n}{2}\log\sigma_0^2 + (y-B\zeta)^T\left(I_{n\times n} + \boldsymbol{B\Lambda B}^T\right)^{-1}(y-B\zeta)/(2\sigma_0^2),$$

yielding (2.4).

PROOF OF LEMMA 3.3. We employ a construction, used in the proof of Lemma 2.2 in [6]. Let A_n be the interval $[t_0 - n^{-1/3}\log n, t_0 + n^{-1/3}\log n]$ and let $(U_1^*, V_1^*), (U_2^*, V_2^*), \ldots$ be an i.i.d sequence of points, (discretely) uniformly distributed on the set of points (X_i, Y_i) such that $X_i \in A_n$.

Let M_n be the number of points $X_i \in A_n$. The number of bootstrap draws such that the first component belongs to A_n has distribution

$$M_n^* \sim \text{Binom}(n, M_n/n), \tag{5.3}$$

so, taking the random variable M_n^* defined by (5.3), independent of the sequence (U_1^*, V_1^*), $(U_2^*, V_2^*), \ldots$, we can represent the bootstrap variables such that the first component belongs to A_n by

$$\sum_{i=1}^{M_n^*} \delta_{\{(U_i^*, V_i^*)\}},$$

where δ_x denotes Dirac measure.

We can couple this process with a Poisson process

$$\sum_{i=1}^{N_n} \delta_{\{(U_i^*, V_i^*)\}},$$

where

$$N_n \sim \text{Poisson}\,(M_n),$$

independent of the (U_i^*, V_i^*), using the construction with a Uniform(0,1) random variable U as in the construction in the proof of Lemma 2.2 in [6]. We find in this way:

$$\mathbb{P}\left\{\sum_{i=1}^{M_n^*} \delta_{\{(U_i^*, V_i^*)\}} \neq \sum_{i=1}^{N_n} \delta_{\{(U_i^*, V_i^*)\}}\right\} \le 2M_n/n,$$

where M_n/n tends to zero almost surely, using an inequality from [20].

This means that we can replace M_{in}^* by N_{in} in (3.12), where the N_{in} are independent Poisson(1) random variables and can replace $\widetilde{W}_n^*$ by its Poissonized version

(5.4)
$$W_n^{(P)}(t) = n^{-1/3} \sum_{j:j/J \in [0, t_0 + n^{-1/3}t]} \sum_{X_i \in I_j} (N_{in} - 1)(Y_i - a_0 - n^{-1/3}x)$$

(5.5)
$$- n^{-1/3} \sum_{j:j/J \in [0, t_0]} \sum_{X_i \in I_j} (N_{in} - 1)(Y_i - a_0 - n^{-1/3}x).$$

For the latter process we have the martingale structure again, and the quadratic variation process $[W_n^{(P)}](t)$, $t \ge 0$ satisfies

$$\left[W_n^{(P)}\right](t) = n^{-2/3} \sum_{j:j/J \in (t_0, t_0 + n^{-1/3}t]} \left\{\sum_{X_i \in I_j} (N_{in} - 1)(Y_i - a_0 - n^{-1/3}x)\right\}^2$$
$$\longrightarrow \sigma_0^2 g(t_0)t.$$

for almost all sequences $(X_1, Y_1), \ldots$. A similar reation holds for $t < 0$. So the result follows in the same way as in the proof of Lemma 3.1. □

PROOF OF LEMMA 4.1. We consider the case $t \ge 0$. It is clear that, conditionally on $(X_1, Y_1), \ldots, (X_n, Y_n)$, $t \mapsto \widetilde{W}_n^*(t)$ is a martingale with respect to the family of σ-algebras $\mathscr{F}_{n,t}^*$, $t \ge 0$, defined by:

$$\mathscr{F}_{n,t}^* = \sigma\left\{(X_i, E_i^*) : X_i \in [t_0 + n^{-1/3}t]\right\}. \qquad t \ge 0.$$

The quadratic variation process is given by:

$$\left[\widetilde{W}_n^*\right](t) = n^{-2/3} \sum_{i:X_i \in [t_0, t_0 + n^{-1/3}t]} (E_i^*)^2$$

We have:

$$\begin{aligned}&\mathbb{E}\left\{\left[\widetilde{W}_n^*\right](t)\,\middle|\,(X_1,Y_1),\dots,(X_n,Y_n)\right\}\\&=n^{-2/3}\sum_{i:X_i\in[t_0,t_0+n^{-1/3}t]}\mathbb{E}\left\{(E_i^*)^2\,\middle|\,(X_1,Y_1),\dots,(X_n,Y_n)\right\}\\&=n^{-2/3}\sum_{i:X_i\in[t_0,t_0+n^{-1/3}t]}n^{-1}\sum_{j=1}^n\tilde E_j^2\xrightarrow{a.s.}\sigma_0^2g(t_0)t.\end{aligned}$$

Note that

$$\begin{aligned}&\{Y_j-\tilde f_{nh}(X_j)\}^2\\&=\{Y_j-f_0(X_j)\}^2+\{f_0(X_j)-\tilde f_{nh}(X_j)\}^2+2\{Y_j-f_0(X_j)\}\{f_0(X_j)-\tilde f_{nh}(X_j)\},\end{aligned}$$

and that therefore

$$n^{-1}\sum_{j=1}^n\tilde E_j^2\sim n^{-1}\sum_{j=1}^n\{Y_j-f_0(X_j)\}^2$$

almost surely, as $n\to\infty$, by the properties of $\tilde f_{nh}$.

We can treat the case $t<0$ in a similar way. This means that we can apply Rebolledo's theorem, see Theorem 3.6, p. 68 of [8] and [15]. The conclusion of the lemma now follows. □

REFERENCES

[1] BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical inference under order restrictions. The theory and application of isotonic regression*. John Wiley & Sons, London-New York-Sydney Wiley Series in Probability and Mathematical Statistics. MR0326887 (48 ##5229)

[2] BHAUMIK, P. and GHOSAL, S. (2015). Bayesian two-step estimation in differential equation models.

[3] BHAUMIK, P. and GHOSAL, S. (2017). Efficient Bayesian estimation and uncertainty quantification in ordinary differential equation models.

[4] BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969)* 177–197. Cambridge Univ. Press, London. MR0277070

[5] CHAKRABORTY, M. and GHOSAL, S. (2021). Coverage of credible intervals in nonparametric monotone regression. *Ann. Statist.* **49** 1011–1028. https://doi.org/10.1214/20-aos1989 MR4255117

[6] GROENEBOOM, P. (1988). Limit theorems for convex hulls. *Probab. Theory Related Fields* **79** 327–368. https://doi.org/10.1007/BF00342231 MR959514

[7] GROENEBOOM, P. (2021). Monotone Regression. https://github.com/pietg/monotone-regression.

[8] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation under Shape Constraints*. Cambridge Univ. Press, Cambridge.

[9] GROENEBOOM, P. and JONGBLOED, G. (2023). Confidence intervals in monotone regression. Submitted.

[10] GROENEBOOM, P. and WELLNER, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*. *DMV Seminar* **19**. Birkhäuser Verlag, Basel. MR1180321 (94k:62056)

[11] HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* **16** 927–985. With a discussion and a reply by the author. https://doi.org/10.1214/aos/1176350933 MR959185

[12] HALL, P. (1992). *The bootstrap and Edgeworth expansion*. *Springer Series in Statistics*. Springer.

[13] KOSOROK, M. R. (2008). Bootstrapping the Grenander estimator. In *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*. *Inst. Math. Stat. Collect.* **1** 282–292. Inst. Math. Statist., Beachwood, OH.

[14] LIN, L. and DUNSON, D. B. (2014). Bayesian monotone regression using Gaussian process projection. *Biometrika* **101** 303–317.

[15] REBOLLEDO, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. Verw. Gebiete* **51** 269–286. https://doi.org/10.1007/BF00587353 MR566321 (81g:60023)

[16] ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order restricted statistical inference*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. John Wiley & Sons Ltd., Chichester. MR961262 (90b:62001)

[17] SEN, B., BANERJEE, M. and WOODROOFE, M. B. (2010). Inconsistency of bootstrap: the Grenander estimator. *Ann. Statist.* **38** 1953–1977. https://doi.org/10.1214/09-AOS777 MR2676880 (2011f:62046)

[18] SEN, B. and XU, G. (2015). Model based bootstrap methods for interval censored data. *Comput. Statist. Data Anal.* **81** 121–129. https://doi.org/10.1016/j.csda.2014.07.007 MR3257405

[19] SILVAPULLE, M. J. and SEN, P. K. (2005). *Constrained statistical inference: Inequality, order and shape restrictions*. John Wiley & Sons.

[20] VERVAAT, W. (1969). Upper bounds for the distance in total variation between the binomial or negative binomial and the Poisson distribution. *Statistica Neerlandica* **23** 79–86. https://doi.org/10.1111/j.1467-9574.1969.tb00075.x MR242235