Monotone least squares and isotonic quantiles

Alexandre Mösching and Lutz Dümbgen

University of Bern
e-mail: alexandre.moesching@stat.unibe.ch; duembgen@stat.unibe.ch

Abstract: We consider bivariate observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ such that, conditional on the $X_i$, the $Y_i$ are independent random variables. Precisely, the conditional distribution function of $Y_i$ equals $F_{X_i}$, where $(F_x)_x$ is an unknown family of distribution functions. Under the sole assumption that $x \mapsto F_x$ is isotonic with respect to stochastic order, one can estimate $(F_x)_x$ in two ways:
(i) For any fixed $y$ one estimates the antitonic function $x \mapsto F_x(y)$ via nonparametric monotone least squares, replacing the responses $Y_i$ with the indicators $1_{[Y_i \leq y]}$.
(ii) For any fixed $\beta \in (0, 1)$ one estimates the isotonic quantile function $x \mapsto F_x^{-1}(\beta)$ via a nonparametric version of regression quantiles.

We show that these two approaches are closely related, with (i) being more flexible than (ii). Then, under mild regularity conditions, we establish rates of convergence for the resulting estimators $\hat{F}_x(y)$ and $\hat{F}_x^{-1}(\beta)$, uniformly over $(x, y)$ and $(x, \beta)$ in certain rectangles as well as uniformly in $y$ or $\beta$ for a fixed $x$.

MSC 2010 subject classifications: 62G08, 62G20, 62G30.
Keywords and phrases: Regression quantiles, stochastic order, uniform consistency.

Received August 2019.

Contents

1 Introduction .................................. 25
2 Estimation of the conditional distributions . . . . . . . . . . . . . . 27
  2.1 Estimation of $F_x$ via monotone least squares ................. 27
  2.2 Plug-in estimation of $Q_x$ ................................ 28
  2.3 Estimation of $Q_x$ via monotone regression quantiles ...... 28
  2.4 Connection between the two estimation paradigms ......... 30
  2.5 A data example ........................................... 30
3 Asymptotic considerations .................................. 30
  3.1 Uniform consistency in both arguments ...................... 31
  3.2 Uniform consistency at a single point ....................... 34
4 Proofs and technical details ................................ 34
  4.1 Monotone regression ..................................... 34
  4.2 Proofs of Lemma 2.1 and 2.3 ............................. 39

*This work was supported by Swiss National Science Foundation.
1. Introduction

Suppose we observe \( n \geq 1 \) pairs

\[(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n) \in \mathcal{X} \times \mathbb{R}\]

with random or fixed covariate values \( X_1, \ldots, X_n \) in a set \( \mathcal{X} \subset \mathbb{R} \) such that, conditional on \( X = (X_i)_{i=1}^n \), the response values \( Y_1, \ldots, Y_n \) are independent with

\[
P(Y_i \leq y \mid X) = F_{X_i}(y),
\]

for \( 1 \leq i \leq n \) and \( y \in \mathbb{R} \). Here \( (F_x)_{x \in \mathcal{X}} \) is an unknown family of distribution functions on \( \mathbb{R} \). Note that some values \( X_i \) could be identical, so the corresponding random variables \( Y_i \) have the same conditional distribution, given \( X \).

Our goal is to estimate the whole family \( (F_x)_{x \in \mathcal{X}} \) under the sole assumption that \( x \mapsto F_x \) is isotonic (non-decreasing) with respect to stochastic order. This can be expressed in three equivalent ways:

1. (SO.1) For arbitrary fixed \( y \in \mathbb{R} \), \( F_x(y) \) is antitonic (non-increasing) in \( x \in \mathcal{X} \).
2. (SO.2) For any fixed \( \beta \in (0, 1) \), the minimal \( \beta \)-quantile \( F^{-1}_x(\beta) := \min\{y \in \mathbb{R} : F_x(y) \geq \beta\} \) is isotonic in \( x \in \mathcal{X} \).
3. (SO.3) For any fixed \( \beta \in (0, 1) \), the maximal \( \beta \)-quantile \( F^{-1}_x(\beta^+) := \inf\{y \in \mathbb{R} : F_x(y) > \beta\} \) is isotonic in \( x \in \mathcal{X} \).

In what follows, we denote with \( Q_x(\beta) \) any \( \beta \)-quantile of \( F_x \) and assume that it is isotonic in \( x \).

Such a constraint appears natural in several settings. For instance, an employee’s income \( Y \) tends to increase with his or her age \( X \). Other examples in which such a stochastic order is plausible are: The expenditures \( Y \) of a household for certain goods in relation to its monthly income \( X \); the body height or weight \( Y \) of a child in relation to its age \( X \). Stochastic ordering constraints also have applications in forecasting. For example, \( X_1, \ldots, X_n \) and \( Y_1, \ldots, Y_n \) could be the predicted and actual cumulative precipitation amounts on \( n \) different days, respectively, with the predictions being obtained from a numerical weather prediction model, see Henzi [7].

With condition (SO.1) in mind, one could think about estimating the antitonic function \( x \mapsto F_x(y) \) by means of monotone least squares regression, replacing the response values \( Y_i \) with the indicator variables \( 1_{Y_i \leq y} \). Precisely, we would set \( \hat{F}_x(y) = h(x) \) with an antitonic function \( h : \mathcal{X} \to [0, 1] \) such that

\[
\sum_{i=1}^n (1_{Y_i \leq y} - h(X_i))^2
\]
is minimal. The solution $h$ is unique on the set $\mathcal{X}_n := \{X_1, \ldots, X_n\}$, and on $\mathcal{X} \setminus \mathcal{X}_n$ one could extrapolate it in some reasonable way. In the special case of a finite $\mathcal{X}$, this approach has been proposed by Hogg [8] and analyzed by El Barmi and Mukerjee [6]. For the 2-sample problem, the nonparametric maximum likelihood estimator was found by Brunk et al. [2], but it has no known extension to the $k$-sample case, for $k \geq 3$.

Conditions (SO.2-3) suggest to imitate the regression quantiles of Koenker and Bassett [10]. That means, we estimate the conditional $\beta$-quantiles $Q_x(\beta)$ by $\hat{Q}_x(\beta) = h(x)$ with an isotonic function $h : \mathcal{X} \to \mathbb{R}$ minimizing the empirical risk

$$\sum_{i=1}^n \rho_\beta(Y_i - h(X_i)),$$

where $\rho_\beta$ denotes the loss function

$$\rho_\beta(z) := (\beta - 1[z<0])z.$$

This estimator has been considered, for instance, by Poiraud-Casanova and Thomas-Agnan [14] who showed that it coincides with an estimator of Casady and Cryer [3] which is given by a certain minimax formula involving sample $\beta$-quantiles. The characterization of isotonic estimators in terms of minimax formulae has also been derived by Robertson and Wright [15] in a rather general framework including arbitrary partial orders on $\mathcal{X}$ and general loss functions $R_i(\cdot)$ in place of $\rho_\beta(Y_i - \cdot)$, see also Section 4.1.

The goals of the present paper are to clarify the connection between these two estimation paradigms and to provide new consistency results in a suitable asymptotic framework.

In Section 2, we give a detailed description of the estimator $(\hat{F}_x)_{x \in \mathcal{X}}$ based on monotone least squares and estimators $(\hat{Q}_x)_{x \in \mathcal{X}}$ based on monotone regression quantiles. Then we show that the estimators $\hat{Q}_x$ are essentially quantiles of the estimators $\hat{F}_x$, but the latter allow for smoother estimated quantile curves.

In Section 3, we analyze the estimators in a suitable asymptotic framework with a triangular scheme of observations and $\mathcal{X}$ being a real interval. It turns out that under certain regularity conditions on the design points and the true distribution functions $F_x$, one can prove rates of convergence for quantities such as

$$\sup_{x \in I, y \in J} |\hat{F}_x(y) - F_x(y)|$$

and

$$\sup_{x \in I, \beta \in B} |\hat{Q}_x(\beta) - Q_x(\beta)|$$

with intervals $I \subset \mathcal{X}$, $J \subset \mathbb{R}$ and $B \subset (0, 1)$. These results generalize and improve the findings of Casady and Cryer [3], see also Mukerjee [13] who analyzed a slightly different estimator. In addition, we investigate

$$\sup_{y \in J} |\hat{F}_{x_o}(y) - F_{x_o}(y)|$$

and

$$\sup_{\beta \in B} |\hat{Q}_{x_o}(\beta) - Q_{x_o}(\beta)|$$

for a fixed interior point $x_o$ of $\mathcal{X}$. These results complement the analysis of a single quantile curve by Wright [17].

Proofs and technical details are deferred to Section 4. We also provide some general results about isotonic regression which are of independent interest.
2. Estimation of the conditional distributions

Throughout this section, we view the observations \((X_i, Y_i), 1 \leq i \leq n\), as fixed and focus mainly on computational aspects. Let \(x_1 < \cdots < x_m\) be the different elements of the set \(X_n\) of observed values \(X_i\), implying \(m \leq n\). For \(1 \leq j \leq m\), we set 
\[
   w_j := \#\{i : X_i = x_j\}.
\]
Then 
\[
   \mathbb{P}(Y_i \leq y) = F_{x_j}(y) \quad \text{whenever } X_i = x_j,
\]
and the unconstrained maximum likelihood estimator of \(F_{x_j}(y)\) is given by 
\[
   \hat{F}_j(y) := w_j^{-1} \sum_{i : X_i = x_j} 1_{Y_i \leq y}.
\]

### 2.1. Estimation of \(F_x\) via monotone least squares

The estimator \(\hat{F}_j(y)\) in (2.1) is rather poor by itself, unless the corresponding subsample size \(w_j\) is large. But in connection with our stochastic order constraint, it becomes a useful tool. Note first that, for any function \(h : \mathcal{X} \to \mathbb{R}\), 
\[
   \sum_{i=1}^{n} (1_{Y_i \leq y} - h(X_i))^2 = \sum_{j=1}^{m} w_j (\hat{F}_j(y) - h(x_j))^2 + \sum_{j=1}^{m} w_j \hat{F}_j(y) (1 - \hat{F}_j(y)),
\]
and the stochastic order assumption implies that the vector \(F(y) = (F_{x_j}(y))_{j=1}^{m}\) belongs to the cone 
\[
   \mathbb{R}_m^m := \{f \in \mathbb{R}^m : f_1 \geq f_2 \geq \cdots \geq f_m\}.
\]
Hence one can estimate \(F(y)\) by the unique least squares estimator 
\[
   \hat{F}(y) = (\hat{F}_{x_j}(y))_{j=1}^{m} := \arg\min_{f \in \mathbb{R}_m^m} \sum_{j=1}^{m} w_j (\hat{F}_j(y) - f_j)^2.
\]
It is known that \(\hat{F}(y)\) may also be represented by the following minimax and maximin formulae, see Robertson, Wright and Dykstra [16]: For \(1 \leq j \leq m\), 
\[
   \hat{F}_{x_j}(y) = \min_{r \leq j} \max_{s \geq j} \hat{F}_{rs}(y) = \max_{s \geq j} \min_{r \leq j} \hat{F}_{rs}(y),
\]
where 
\[
   \hat{F}_{rs}(y) := w_{rs}^{-1} \sum_{j=r}^{s} w_j \hat{F}_j(y) = \arg\min_{f \in \mathbb{R}} \sum_{j=r}^{s} w_j (\hat{F}_j(y) - f)^2,
\]
\[
   w_{rs} := \sum_{j=r}^{s} w_j = \#\{i : x_r \leq X_i \leq x_s\},
\]
and \( r, s \) stand for indices in \( \{1, 2, \ldots, m\} \) such that \( r \leq s \). These formulae are useful for theoretical considerations. In particular, since the pointwise maximum or minimum of finitely many distribution functions is a distribution function, too, we may conclude that for \( 1 \leq j \leq m \),

\[
\hat{F}_{x_j}(\cdot) \text{ is a distribution function.}
\]

The computation of \( \hat{F}(y) \) is easily achieved via the pool-adjacent-violators algorithm (PAVA), see Robertson, Wright and Dykstra [16]. Note also that it suffices to compute \( \hat{F}(y) \) for at most \( n - 1 \) different values of \( y \). Precisely, if \( y_1 < y_2 < \cdots < y_\ell \) are the elements of \( \{Y_1, Y_2, \ldots, Y_n\} \), then \( \hat{F}(y) = 0 \) for \( y < y_1 \), \( \hat{F}(y) = 1 \) for \( y \geq y_\ell \), and \( \hat{F}(y) = \hat{F}(y_k) \) for \( 1 \leq k < \ell \) and \( y \in [y_k, y_{k+1}) \). Consequently, since the PAVA is known to have linear complexity, the computation of all estimators \( \hat{F}_{x_j}(\cdot), \ 1 \leq j \leq m \), requires \( O(n \log n + m \ell) = O(n^2) \) steps.

Finally, we extrapolate \( \hat{F}(y) \) to an antitonic function \( x \mapsto \hat{F}_x(y) \) on \( X \). We set \( \hat{F}_x(y) := \hat{F}_{x_1}(y) \) for \( x < x_1 \) and \( \hat{F}_x(y) := \hat{F}_{x_m}(y) \) for \( x > x_m \). For \( x_{j-1} \leq x \leq x_j \), \( 1 < j \leq m \), one could define \( \hat{F}_x(y) \) by linear interpolation, but other antitonic interpolations are possible without affecting our asymptotic results.

### 2.2. Plug-in estimation of \( Q_x \)

Once we have estimated \( (F_x)_{x \in X} \) by \( (\hat{F}_x)_{x \in X} \) as in Section 2.1, we can easily determine the corresponding quantile functions. For any fixed \( \beta \in (0, 1) \) and \( x_j \), \( 1 \leq j \leq m \), we could determine the minimal and maximal \( \beta \)-quantiles,

\[
\begin{align*}
\hat{F}_x^{-1}(\beta) & := \min \{ y \in \mathbb{R} : \hat{F}_x(y) \geq \beta \} \\
\hat{F}_x^{-1}(\beta+) & := \inf \{ y \in \mathbb{R} : \hat{F}_x(y) > \beta \}.
\end{align*}
\]

Both vectors \( (\hat{F}_x^{-1}(\beta))^m \) and \( (\hat{F}_x^{-1}(\beta+))^m \) are isotonic, and any choice of an isotonic function \( X \ni x \mapsto \tilde{Q}_x(\beta) \) such that \( \hat{F}_x^{-1}(\beta) \leq \tilde{Q}_x(\beta) \leq \hat{F}_x^{-1}(\beta+) \), \( 1 \leq j \leq m \), is a plausible estimator of a \( \beta \)-quantile curve.

### 2.3. Estimation of \( Q_x \) via monotone regression quantiles

Similarly as in Section 2.1, we focus on the vector \( Q(\beta) = (Q_{x_j}(\beta))_{j=1}^m \). Writing

\[
\sum_{i=1}^n \rho_\beta(Y_i - h(X_i)) = \sum_{j=1}^m \sum_{i : X_i = x_j} \rho_\beta(Y_i - h(x_j)),
\]

one can estimate \( Q(\beta) \) by some vector in the set

\[
\tilde{Q}(\beta) := \arg \min_{q \in \mathbb{R}_+^m} T_\beta(q),
\]
where \(\mathbb{R}_+^m := -\mathbb{R}_+^m = \{ \mathbf{q} \in \mathbb{R}^m : q_1 \leq q_2 \leq \cdots \leq q_m \} \) and

\[
T_\beta(\mathbf{q}) := \sum_{j=1}^m \sum_{i : X_i = x_j} \rho_\beta(Y_i - q_j).
\]

Note that the function \(T_\beta(\cdot)\) is convex but not strictly convex on \(\mathbb{R}^m\). Hence it need not have a unique minimizer. The next result provides more precise information in terms of the minimal and maximal sample \(\beta\)-quantiles

\[
\hat{F}^{-1}_{rs}(\beta) := \min\{ y \in \mathbb{R} : \hat{F}_{rs}(y) \geq \beta \},
\]

\[
\hat{F}^{-1}_{rs}(\beta+) := \inf\{ y \in \mathbb{R} : \hat{F}_{rs}(y) > \beta \}.
\]

Lemma 2.1. The set \(\hat{Q}(\beta)\) is a compact and convex subset of \(\mathbb{R}_+^m\).

Two particular elements of \(\hat{Q}(\beta)\) are the vectors \(\ell = (\ell_j)_{j=1}^m\) and \(\mathbf{u} = (u_j)_{j=1}^m\) with components

\[
\ell_j := \max \min_{r \leq j} \hat{F}^{-1}_{rs}(\beta) = \min \max_{s \geq j} \hat{F}^{-1}_{rs}(\beta),
\]

\[
u_j := \min \max_{r \leq j} \hat{F}^{-1}_{rs}(\beta+) = \max \min_{s \geq j} \hat{F}^{-1}_{rs}(\beta+).
\]

Any vector \(\mathbf{q} \in \hat{Q}(\beta)\) satisfies \(\ell \leq \mathbf{q} \leq \mathbf{u}\) componentwise.

On the other hand, suppose that \(\mathbf{q} \in \mathbb{R}_+^m\) satisfies \(\ell \leq \mathbf{q} \leq \mathbf{u}\) and that \(\{ j < m : q_j < q_{j+1} \}\) is a subset of \(\{ j < m : \ell_j < \ell_{j+1} \text{ or } u_j < u_{j+1} \}\). Then \(\mathbf{q} \in \hat{Q}(\beta)\).

Finally, for any \(j \in \{1, \ldots, m\}\), the set \(\{ x_j \times (\ell_j, u_j) \}\) contains no data point \((X_i, Y_i)\).

Remark 2.2. At first glance, one might suspect that any isotonic \(\mathbf{q} \in \mathbb{R}_+^m\) satisfying \(\ell \leq \mathbf{q} \leq \mathbf{u}\) minimizes \(T_\beta\). But this conjecture is wrong. As a counterexample, consider the case of \(n = 2\) observations with \(X_1 < X_2\) but \(Y_1 > Y_2\). Here \(m = 2\), and \(\hat{F}_{11}(y) = 1_{[y \geq Y_1]}, \hat{F}_{12}(y) = 1_{[y \geq Y_2]}\) and

\[
\hat{F}_{12}(y) = \begin{cases} 
0 & \text{if } y < Y_2, \\
0.5 & \text{if } Y_2 \leq y < Y_1, \\
1 & \text{if } y \geq Y_1.
\end{cases}
\]

Hence

\[
\ell = (Y_2, Y_2)^\top \quad \text{and} \quad \mathbf{u} = (Y_1, Y_1)^\top,
\]

because \(\hat{F}^{-1}_{11}(0.5) = \hat{F}^{-1}_{12}(0.5+) = Y_1, \hat{F}^{-1}_{21}(0.5) = \hat{F}^{-1}_{22}(0.5) = Y_2\) and

\[
\hat{F}^{-1}_{12}(0.5) = Y_2, \quad \hat{F}^{-1}_{12}(0.5+) = Y_1,
\]

But

\[
\hat{Q}(0.5) = \{(q, q)^\top : q \in [Y_2, Y_1]\},
\]

because for \(\mathbf{q} \in [Y_2, Y_1]^2\) with \(q_1 \leq q_2\),

\[
\rho_{0.5}(Y_1 - q_1) + \rho_{0.5}(Y_2 - q_2) = 0.5(Y_1 - q_1 + q_2 - Y_2) \geq 0.5(Y_1 - Y_2)
\]

with equality if, and only if, \(q_1 = q_2\).
2.4. Connection between the two estimation paradigms

Restricting the plug-in quantile estimators of Section 2.2 to the set $\mathcal{X}_n$ of observed $X$-values leads to the set

$$\hat{Q}_{\text{plug-in}}(\beta) := \{q \in \mathbb{R}^m_+: \hat{F}^{-1}_{x_j}(\beta) \leq q_j \leq \hat{F}^{-1}_{x_j}(\beta+) \text{ for } 1 \leq j \leq m\}.$$ 

This set is closely related to the set $\hat{Q}(\beta)$:

**Lemma 2.3.** The vectors $\ell$ and $u$ in Lemma 2.1 are given by

$$\ell_j = \hat{F}^{-1}_{x_j}(0.5) \quad \text{and} \quad u_j = \hat{F}^{-1}_{x_j}(0.5+) \quad \text{for } 1 \leq j \leq m.$$ 

In particular, $\hat{Q}(\beta) \subset \hat{Q}_{\text{plug-in}}(\beta)$.

**Example 2.4.** The simple example in Remark 2.2 shows that generally $\hat{Q}(\beta) \neq \hat{Q}_{\text{plug-in}}(\beta)$. Let us illustrate this point with a more interesting numerical example. Figure 1 shows a simulated sample of size $n = 100$. In addition, it shows the minimal and maximal median curves $x \mapsto \hat{F}^{-1}_{f_x}(0.5)$, $\hat{F}^{-1}_{f_x}(0.5+) \text{ obtained by linear interpolation of the points } \ell_j = \hat{F}^{-1}_{x_j}(0.5) \text{ and } u_j = \hat{F}^{-1}_{x_j}(0.5+)$, respectively, as well as a piecewise linear median curve $x \mapsto \hat{Q}(0.5)$ minimizing $\int q'(x)^2 \, dx$ among all isotonic functions $q : \mathbb{R} \to \mathbb{R}$ such that $\ell_j \leq q(x_j) \leq u_j$, $1 \leq j \leq m$. Although $\hat{Q}(0.5)$ is a natural candidate and smoother in $x$ than $\hat{F}^{-1}_{f_x}(0.5)$ or $\hat{F}^{-1}_{f_x}(0.5+)$, the corresponding values of $T_{0.5}(\cdot)$ are (rounded to three digits)

$$T_{0.5}(\hat{Q}(0.5)) = 45.343 > T_{0.5}(\ell) = T_{0.5}(u) = 44.112.$$ 

The true medians $F^{-1}_{x}(0.5) = F^{-1}_{x}(0.5+)$ are depicted as well.

2.5. A data example

Figure 2 shows weight for age quantile curves $\hat{Q}_x(\beta)$ of girls between 2 and 20 years old for different values of $\beta$. The dataset used to compute these curves comprises $n = 19,459$ individuals and was publicly released as part of the National Health and Nutrition Examination Survey conducted in the US between 1963 and 1991 (data available from www.cdc.gov). These data were analyzed by Kuczmarski et al. [11] with parametric models to produce smooth quantile curves. However, the details provided in their paper were not sufficient to reproduce their work, so a direct comparison with our nonparametric approach is not possible.

3. Asymptotic considerations

We provide some asymptotic properties of the estimators just introduced in case of a real interval $\mathcal{X}$ and a triangular scheme of observations: For each sample size
Fig 1. $n = 100$ data pairs, together with the true medians $F^{-1}_x(0.5)$ (green, dashed) and the estimated medians $\hat{F}^{-1}_x(0.5)$ (lower red), $\hat{F}^{-1}_x(0.5 +)$ (upper blue) and $\hat{Q}_x(0.5)$ (middle black).

$n \geq 2$, consider observations $(X_{n1}, Y_{n1}), \ldots, (X_{nn}, Y_{nn})$ with $X_{n1}, \ldots, X_{nn} \in \mathcal{X}$ such that conditional on $X_n := (X_{ni})_{i=1}^n$, the random variables $Y_{n1}, \ldots, Y_{nn}$ are independent with

$$\mathbb{P}(Y_{ni} \leq y \mid X_n) = F_{X_n}(y),$$

for $1 \leq i \leq n$ and $y \in \mathbb{R}$. The resulting constrained estimators of $F_x(y)$ and $Q_x(\beta)$ are denoted by $\hat{F}_x(y)$ and $\hat{Q}_x(\beta)$, respectively. In what follows, we derive asymptotic properties of these estimators under moderate assumptions, where asymptotic statements refer to $n \to \infty$.

El Barmi and Mukerjee [6] have derived asymptotic properties, including asymptotic distributions, in case of a fixed finite set $\mathcal{X}$, which is easier to handle than the present setting. We are focusing on settings with a growing number of different design points $X_{ni}$ and rates of convergence. Asymptotic distributions or functional limit theorems are beyond the scope of the present paper, but an interesting topic for future research.

3.1. Uniform consistency in both arguments

First of all, we assume that the distribution functions $F_x$ are Hölder-continuous in $x$, at least on some subinterval of $\mathcal{X}$:
Fig 2. Weight for age quantile curves $\hat{Q}_x(\beta)$ of girls between 2 and 20 years old for different values of $\beta$.

(A.1) For given intervals $I \subset X$ and $J \subset \mathbb{R}$, there exist constants $\alpha \in (0, 1]$ and $C_1 > 0$ such that

$$\sup_{y \in J} |F_w(y) - F_x(y)| \leq C_1 |w - x|^\alpha$$

for arbitrary $w, x \in I$.

Secondly, we assume that the design points are ‘asymptotically dense’ within this interval $I$. To state this precisely, we need some notation. We write

$$\rho_n := \log \frac{n}{n},$$

and $\lambda(\cdot)$ stands for Lebesgue measure. Moreover, the absolute frequency of the design points $X_{ni}$ is denoted by $w_n(\cdot)$, that means,

$$w_n(B) := \#\{i \leq n : X_{ni} \in B\} \quad \text{for} \ B \subset X.$$

(A.2) For given constants $C_2, C_3 > 0$, let $A_n$ be the event that for arbitrary intervals $I_n \subset I$,

$$\frac{w_n(I_n)}{n \lambda(I_n)} \geq C_2 \quad \text{whenever} \quad \lambda(I_n) \geq \delta_n := C_3 \rho_n^{1/(2\alpha+1)}.$$

Then,

$$\mathbb{P}(A_n) \to 1.$$
Remark 3.1 (Fixed design points). Suppose that $I = \mathcal{X} = [a, b]$ with real numbers $a < b$, and let $X_{ni} = a + (i/n)(b - a)$ for $1 \leq i \leq n$. Then Assumption (A.2) is satisfied for any fixed $C_2 < 1$ and $C_3 > 0$.

Remark 3.2 (Random design points). Suppose that $X_{n1}, X_{n2}, \ldots, X_{nn}$ are independent random variables with density $g$ on $\mathcal{X}$ such that $\inf_{x \in I} g(x) > 0$ on $I$. Then for any choice of $\alpha \in (0, 1]$, $0 < C_2 < \inf_{x \in I} g(x)$ and $C_3 > 0$, 

$$\inf \left\{ \frac{w_n(I_n)}{n \lambda(I_n)} : \text{intervals } I_n \subset I \text{ with } \lambda(I_n) \geq \delta_n \right\} \geq C_2$$

with asymptotic probability one as $n \to \infty$. This follows directly from a more general basic fact about empirical distributions in Section 4.3. Hence Assumption (A.2) is satisfied.

Under the two assumptions above, the estimator $\hat{F}_{nx}$ satisfies a uniform consistency property.

Theorem 3.3. Suppose that Assumptions (A.1–2) are satisfied. Then there exists a $C = C(C_1, C_2, C_3) > 0$ such that

$$\lim_{n \to \infty} \mathbb{P} \left( \sup_{x \in I_n, y \in \mathcal{J}} \left| \hat{F}_{nx}(y) - F_x(y) \right| \geq C \rho_n^{\alpha/(2\alpha + 1)} \right) = 0,$$

where $I_n := \{ x \in \mathbb{R} : [x \pm \delta_n] \subset I \}$.

Concerning estimated quantiles, we combine Assumptions (A.1–2) with a growth condition on the conditional distribution functions $F_x$:

(A.3) For some numbers $0 \leq \beta_1 < \beta_2 \leq 1$ and $\kappa > 0$,

$$F_x(y_2) - F_x(y_1) \geq \kappa (y_2 - y_1),$$

for all $x \in I$ and $y_1, y_2 \in \mathbb{R}$ such that $y_1 < y_2$ and $F_x(y_1), F_x(y_2) \in (\beta_1, \beta_2)$.

For instance, if each $F_x$, $x \in I$, has a density $f_x$ such that

$$\kappa := \inf_{x \in I} \inf_{y : \beta_1 < F_x(y) < \beta_2} f_x(y) > 0,$$

then (A.3) is satisfied with the latter parameter $\kappa$.

Theorem 3.4. Suppose that Assumptions (A.1–3) are satisfied with $J = \mathbb{R}$ in (A.1). Then, for any plug-in estimator $(\hat{Q}_{nx})_{x \in \mathcal{X}}$ of $(Q_x)_{x \in \mathcal{X}}$,

$$\lim_{n \to \infty} \mathbb{P} \left( \sup_{x \in I_n, \beta \in B_n} \left| \hat{Q}_{nx}(\beta) - Q_x(\beta) \right| \geq \kappa^{-1} C \rho_n^{\alpha/(2\alpha + 1)} \right) = 0,$$

where $I_n \subset I$ and $C = C(C_1, C_2, C_3)$ are defined as in Theorem 3.3, and $B_n$ denotes the interval $(\beta_1 + C \rho_n^{\alpha/(2\alpha + 1)}, \beta_2 - C \rho_n^{\alpha/(2\alpha + 1)})$. 

3.2. Uniform consistency at a single point

In addition to the previous uniform convergence results, one may verify uniform consistency of $\hat{F}_{nx_o}$ and $\hat{Q}_{nx_o}$ for a fixed interior point $x_o$ of $X$. These results require similar but weaker assumptions.

(A’1$x_o$) For a neighbourhood $U \subset X$ of $x_o$ and an interval $J \subset \mathbb{R}$, there exist constants $\alpha \in (0,1]$ and $C_1 > 0$ such that
\[
\sup_{y \in J} \left| F_x(y) - F_{x_o}(y) \right| \leq C_1 |x - x_o|^\alpha \text{ for arbitrary } x \in U.
\]

(A’2$x_o$) For given constants $C_2, C_3 > 0$, let $A_n$ be the event that
\[
\frac{w_n([x_o - \delta_n, x_o])}{\delta_n}, \frac{w_n([x_o, x_o + \delta_n])}{\delta_n} \geq C_2 \text{ where } \delta_n := C_3 n^{-1/(2\alpha + 1)}.
\]

Then,
\[
P(A_n) \rightarrow 1.
\]

Under these two assumptions, the following consistency property holds.

Theorem 3.5. Suppose that Assumptions (A’1–2$x_o$) are satisfied. Then
\[
\sup_{y \in J} \left| \hat{F}_{nx_o}(y) - F_{x_o}(y) \right| = O_p\left(n^{-\alpha/(2\alpha + 1)}\right).
\]

(A’3$x_o$) For some numbers $0 \leq \beta_1 < \beta_2 \leq 1$ and $\kappa > 0$,
\[
F_{x_o}(y_2) - F_{x_o}(y_1) \geq \kappa(y_2 - y_1),
\]
for all $y_1, y_2 \in \mathbb{R}$ such that $y_1 < y_2$ and $F_{x_o}(y_1), F_{x_o}(y_2) \in (\beta_1, \beta_2)$.

Theorem 3.6. Suppose that Assumptions (A’1–3$x_o$) are satisfied with $J = \mathbb{R}$ in (A’1$x_o$). Then, for any plug-in estimator $(\hat{Q}_{nx})_{x \in X}$ of $(Q_x)_{x \in X}$,
\[
\sup_{\beta \in B_n} \left| \hat{Q}_{nx_o}(\beta) - Q_{x_o}(\beta) \right| = O_p\left(n^{-\alpha/(2\alpha + 1)}\right),
\]
where $B_n := (\beta_1 + \Delta_n, \beta_2 - \Delta_n)$ and $\Delta_n = O(n^{-\alpha/(2\alpha + 1)})$.

4. Proofs and technical details

4.1. Monotone regression

In this section we review isotonic regression on a totally ordered set in a rather general setting, summarizing and extending results of numerous authors. Our main goal is a thorough understanding of isotonic regression in situations with potentially non-unique solutions. For extensions to partially ordered sets we refer to Mühlemann, Jordan and Ziegel [12].
The starting point are \( m \geq 2 \) loss functions \( R_1, \ldots, R_m : \mathbb{R} \to \mathbb{R} \) with the following property: For arbitrary indices \( 1 \leq a \leq b \leq m \), the function

\[
R_{ab} := \sum_{j=a}^{b} R_j
\]

is minimal on a compact interval \( [L_{ab}, U_{ab}] \subset \mathbb{R} \), strictly antitonic on \(( -\infty, L_{ab} \] \) and strictly isotonic on \([ U_{ab}, \infty \)).

This property is satisfied if all functions \( R_j \) are convex with \( R_j(x) \to \infty \) as \( |x| \to \infty \). It implies a refined version of the so-called Cauchy-mean-value property.

**Proposition 4.1.** Let \( \{a, \ldots, b\} \subset \{1, \ldots, m\} \) be partitioned into \( k \geq 2 \) index intervals \( \{a_1, \ldots, b_1\}, \ldots, \{a_k, \ldots, b_k\} \). Then

\[
\min_{1 \leq i \leq k} L_{a_i b_i} \leq L_{ab} \leq \max_{1 \leq i \leq k} L_{a_i b_i}
\]

and

\[
\min_{1 \leq i \leq k} U_{a_i b_i} \leq U_{ab} \leq \max_{1 \leq i \leq k} U_{a_i b_i}.
\]

**Proof.** The smallest minimizer \( L_{ab} \) of \( R_{ab} \) is the largest real number \( r \) such that \( R_{ab} \) is strictly antitonic on \(( -\infty, r \] \) and the smallest real number \( s \) such that \( R_{ab} \) is isotonic on \([ s, \infty \)). Since \( R_{ab} = \sum_{i=1}^{k} R_{a_i b_i} \), this function is strictly antitonic on the interval \( \bigcap_{1 \leq i \leq k} (-\infty, L_{a_i b_i}] = (-\infty, \min_{1 \leq i \leq k} L_{a_i b_i}] \) and isotonic on the interval \( \bigcap_{1 \leq i \leq k} [L_{a_i b_i}, \infty) = [\max_{1 \leq i \leq k} L_{a_i b_i}, \infty) \). This yields the desired inequalities for \( L_{ab} \). The largest minimizer \( U_{ab} \) can be handled analogously. \( \square \)

Now we consider the function \( T : \mathbb{R}^m \to \mathbb{R} \),

\[
T(x) := \sum_{j=1}^{m} R_j(x_j)
\]

and the set

\[
Q := \arg\min_{q \in \mathbb{R}^m} T(q).
\]

The elements of \( Q \) can be characterized completely in terms of the minimizers of the functions \( R_{ab} \). Throughout the sequel, we set \( x_0 := -\infty \) and \( x_{m+1} := \infty \) for a vector \( x \in \mathbb{R}^m \). Moreover, the componentwise minimum and maximum of vectors \( x, y \in \mathbb{R}^m \) are denoted by \( \min(x, y) \) and \( \max(x, y) \), respectively.

**Proposition 4.2.** For a vector \( x \in \mathbb{R}^m \), the following two properties are equivalent:

(i) \( x \in Q \).

(ii) For arbitrary indices \( 1 \leq a \leq b \leq m \),

\[
x_a \leq U_{ab} \quad \text{if} \quad x_{a-1} < x_a,
\]

\[
x_b \geq L_{ab} \quad \text{if} \quad x_b < x_{b+1}.
\]
Corollary 4.3. If $\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in \mathcal{Q}$, then $\min(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ and $\max(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ belong to $\mathcal{Q}$ as well.
Proof. For symmetry reasons it suffices to verify that \( x := \min(x^{(1)}, x^{(2)}) \in Q \), and this is equivalent to \( x \) satisfying property (iii) in Proposition 4.2. Let \( 1 \leq a \leq b \leq m \), and suppose that \( x_{a-1} < x_a \). Then for some \( k \in \{1, 2\} \),

\[
    x_{a-1} = x^{(k)}_a < x_a \leq x^{(k)}_a,
\]

so property (iii) of \( x^{(k)} \) implies that \( x_a \leq x^{(k)}_a \leq U_{ab} \). In case of \( x_b < x_{b+1} \), we choose \( k \in \{1, 2\} \) such that

\[
    x_b = x^{(k)}_b < x_{b+1} \leq x^{(k)}_{b+1},
\]

and then property (iii) of \( x^{(k)} \) implies that \( x_b = x^{(k)}_b \geq L_{ab} \).

Now we provide the main result involving min-max and max-min formulae for the set \( Q \).

**Theorem 4.4.** For any index \( 1 \leq j \leq m \),

\[
    \ell^{(1)}_j := \max_{a \leq j} \min_{b \geq j} L_{ab} = \ell^{(2)}_j := \min_{a \leq j} \max_{b \geq j} L_{ab}
\]

and

\[
    u^{(1)}_j := \min_{b \geq j} \max_{a \leq j} U_{ab} = u^{(2)}_j := \max_{a \leq j} \min_{b \geq j} U_{ab}.
\]

This defines vectors \( \ell = (\ell^{(1)}_j)_{j=1}^m \) and \( u = (u^{(1)}_j)_{j=1}^m \) in \( Q \), and any vector \( x \in Q \) satisfies \( \ell \leq x \leq u \) componentwise.

**Proof of Theorem 4.4.** For symmetry reasons, if suffices to verify the claims about \( \ell \). Precisely, with \( \ell^{(k)} := (\ell^{(k)}_j)_{j=1}^m \), we show subsequently that

\[
    \ell^{(1)} \leq \ell^{(2)}, \quad (4.1) \\
    \ell^{(2)} \leq x \quad \text{for any } x \in Q, \quad (4.2) \\
    \ell^{(1)} \in Q. \quad (4.3)
\]

Inequality (4.1) follows from

\[
    \ell^{(1)}_j \leq \max_{a \leq j} \min_{b \geq j} L_{ab} = \max_{a \leq j} \ell^{(2)}_j = \ell^{(2)}_j
\]

for \( 1 \leq j \leq m \).

As to (4.2), for \( x \in Q \) and \( 1 \leq j \leq m \) let \( \tilde{b} \) be the largest index such that \( x_{\tilde{b}} = x_j \). Then \( x_{\tilde{b}} < x_{\tilde{b}+1} \), so property (ii) of \( x \) in Proposition 4.2 implies that

\[
    \ell^{(2)}_j \leq \max_{a \leq j} L_{a, \tilde{b}} \leq x_{\tilde{b}} = x_j.
\]

It remains to verify (4.3). For indices \( 1 \leq j < k \leq m \),

\[
    \ell^{(1)}_j = \max_{a \leq j} \min_{b \geq j} L_{ab} \leq \max_{a \leq j \leq k} \min_{b \geq j} L_{ab} \leq \max_{a \leq k} \min_{b \geq k} L_{ab} = \ell^{(1)}_k,
\]
whence $\ell^{(1)} \in \mathbb{R}^m_+$. To show that $\ell^{(1)} \in Q$, it suffices to show that it has property (iii) in Proposition 4.2, and this is an immediate consequence of the following two claims: For $1 \leq j \leq m$,

$$
\ell^{(1)}_{j-1} < \ell^{(1)}_j \quad \text{implies that} \quad \ell^{(1)}_j = \min_{b \geq j} L_{jb},
$$

and

$$
\ell^{(1)}_j < \ell^{(1)}_{j+1} \quad \text{implies that} \quad \ell^{(1)}_j = \max_{a \leq j} L_{aj}.
$$

As to (4.4), suppose that the conclusion is wrong, i.e. $\ell^{(1)}_j > \min_{b \geq j} L_{jb}$. Then $j > 1$, and for some index $\tilde{a} \leq j - 1,$

$$
\ell^{(1)}_j = \min_{b \geq j} L_{\tilde{a}b} \leq \min_{b \geq j} \max(L_{\tilde{a},j-1}, L_{jb}) = \max(L_{\tilde{a},j-1}, \min_{b \geq j} L_{jb}) = L_{\tilde{a},j-1},
$$

where we used Proposition 4.1. But then

$$
\ell^{(1)}_{j-1} \geq \min_{b \geq j-1} L_{\tilde{a}b} = \min(L_{\tilde{a},j-1}, \min_{b \geq j} L_{\tilde{a}b}) = \ell^{(1)}_j,
$$

i.e. the assumption of (4.4) is wrong as well.

Concerning (4.5), suppose that that the conclusion is wrong, i.e. $\ell^{(1)}_j < L_{\tilde{a}j}$ for some $\tilde{a} \leq j$. Then $j < m$, and

$$
\ell^{(1)}_j \geq \min_{b \geq j} L_{\tilde{a}b} = \min(L_{\tilde{a}j}, \min_{b \geq j+1} L_{\tilde{a}b}) = \min_{b \geq j+1} L_{\tilde{a}b}
$$

$$
\geq \min_{b \geq j+1} \min(L_{\tilde{a}j}, L_{j+1,b}) = \min(L_{\tilde{a}j}, \min_{b \geq j+1} L_{j+1,b}) = \min_{b \geq j+1} L_{j+1,b}.
$$

Consequently,

$$
\min_{b \geq j+1} L_{j+1,b} \leq \ell^{(1)}_j \quad \text{and} \quad \min_{b \geq j+1} L_{\tilde{a}b} \leq \ell^{(1)}_j.
$$

This is true for any index $\tilde{a} \leq j$ with $L_{\tilde{a}j} > \ell^{(1)}_j$. If $a \leq j$ is such that $L_{aj} \leq \ell^{(1)}_j$, then

$$
\min_{b \geq j+1} L_{ab} \leq \min_{b \geq j+1} \max(L_{aj}, L_{j+1,b}) = \max(L_{aj}, \min_{b \geq j+1} L_{j+1,b}) \leq \ell^{(1)}_j.
$$

Thus $\min_{b \geq j+1} L_{aj} \leq \ell^{(1)}_j$ for any $a \leq j + 1$. Consequently, $\ell^{(1)}_{j+1} \leq \ell^{(1)}_j$, i.e. the assumption of (4.5) is wrong as well.

We end this subsection with two additional conclusions for the special case of convex functions $R_j$.

**Theorem 4.5.** Suppose in addition that all loss functions $R_j$ are convex. Then the set $Q$ is compact and convex. If $x \in \mathbb{R}^m_+$ is such that $\ell \leq x \leq u$ and \{\(j < m : x_j < x_{j+1}\)\} $\subset$ \{\(j < m : \ell_j < \ell_{j+1} \text{ or } u_j < u_{j+1}\)\}, then $x \in Q$. Moreover, each function $R_j$ is linear on the interval $[\ell_j, u_j]$. 


Proof. The general assumptions imply that each function \( R_j = R_{jj} \) has a compact set of minimizers. Together with convexity, this implies that \( R_j \) is continuous with \( R_j(x) \to \infty \) as \( |x| \to \infty \). But then, \( T : \mathbb{R}^m \to \mathbb{R} \) is a continuous and convex function such that \( T(x) \to \infty \) as \( \|x\| \to \infty \). Moreover, \( \mathbb{R}_+^m \) is a closed convex cone in \( \mathbb{R}^m \). This implies that \( Q \) is a compact and convex set.

To verify the remaining statements, consider the vectors \( \mathbf{x}(\lambda) := (1-\lambda)\ell + \lambda \mathbf{u} \), \( \lambda \in [0,1] \). Since \( Q \) is a convex set, all these vectors belong to \( Q \). But for \( 0 < \lambda < 1 \),

\[
\{ j < m : x_j(\lambda) < x_{j+1}(\lambda) \} = \{ j < m : \ell_j < \ell_{j+1} \text{ or } u_j < u_{j+1} \}.
\]

Exploiting property (ii) of \( \mathbf{x}(\lambda) \) in Proposition 4.2 for all \( \lambda \in (0,1) \), we may conclude that for arbitrary indices \( 1 \leq a \leq b \leq m \),

\[
u_a \leq U_{ab} \text{ if } \ell_{a-1} < \ell_a \text{ or } u_{a-1} < u_a,\]

\[
\ell_b \geq L_{ab} \text{ if } \ell_b < \ell_{b+1} \text{ or } u_b < u_{b+1}.
\]

In particular, any vector \( \mathbf{x} \in \mathbb{R}_+^m \) such that \( \ell \leq \mathbf{x} \leq \mathbf{u} \) and \( \{ j < m : x_j < x_{j+1} \} \) is a subset of \( \{ j < m : \ell_j < \ell_{j+1} \text{ or } u_j < u_{j+1} \} \) satisfies property (iii) in Proposition 4.2. Hence \( \mathbf{x} \in Q \).

Finally, since

\[
T_{\beta}(q(\lambda)) = \sum_{j=1}^{m} R_j((1 - \lambda)\ell_j + \lambda u_j)
\]

is constant in \( \lambda \in [0,1] \), each summand \( R_j((1 - \lambda)\ell_j + \lambda u_j) \) has to be linear in \( \lambda \in [0,1] \), which is equivalent to \( R_j \) being linear on \( [\ell_j, u_j] \).

\[\square\]

4.2. Proofs of Lemma 2.1 and 2.3

Proof of Lemma 2.1. For \( 1 \leq j \leq m \), set

\[
R_j(q) := \sum_{i : X_i = x_j} \rho_\beta(Y_i - q).
\]

This is a convex function of \( q \in \mathbb{R} \) with \( R_j(q) \to \infty \) as \( |q| \to \infty \). To apply the results of the previous subsection, we need to determine the sets \( [L_{ab}, U_{ab}] \) for \( 1 \leq a \leq b \leq m \). Note that \( R_j'(q+) = \sum_{i : X_i = x_j} (1_{[Y_i \leq q]} - \beta) \), whence

\[
R_{ab}'(q+) = w_{ab}(\hat{F}_{ab}(q) - \beta).
\]

Consequently,

\[
L_{ab} = \min \{ q \in \mathbb{R} : R_{ab}'(q+) \geq 0 \} = \hat{F}_{ab}^{-1}(\beta),
\]

\[
U_{ab} = \inf \{ q \in \mathbb{R} : R_{ab}'(q+) > 0 \} = \hat{F}_{ab}^{-1}(\beta+).
\]

Now all but the last statement of Lemma 2.1 follow from Theorems 4.4 and 4.5. As to the last statement, note that each \( R_j \) is a convex and piecewise linear
function with strict changes of slope at each \( Y_i \) such that \( X_i = x_j \). Consequently, since \( R_j \) is linear on \([\ell_j, u_j]\), there is no data point \((X_i, Y_i)\) such that \( X_i = x_j \) and \( Y_i \in (\ell_j, u_j)\). \(\square\)

**Proof of Lemma 2.3.** For arbitrary \( y \in \mathbb{R} \),

\[
y \geq \hat{F}_{x_j}^{-1}(\beta) \quad \text{if and only if} \quad \hat{F}_{x_j}(y) \geq \beta.
\]

But the min-max formula (2.2) for \( \hat{F}_{x_j}(y) \) implies that the inequality on the right hand side is equivalent to the following statements:

\[
\min_{r \leq j} \max_{s \geq j} \hat{F}_{rs}(y) \geq \beta,
\]

for all \( r \leq j \), \( \hat{F}_{rs}(y) \geq \beta \) for some \( s = s(r) \geq j \),

for all \( r \leq j \), \( y \geq \hat{F}_{rs}^{-1}(\beta) \) for some \( s = s(r) \geq j \),

\[
y \geq \max_{r \leq j} \min_{s \geq j} \hat{F}_{rs}^{-1}(\beta) = \ell_j.
\]

Hence \( \hat{F}_{x_j}^{-1}(\beta) = \ell_j \). Analogously, for any \( y \in \mathbb{R} \),

\[
y \geq \hat{F}_{x_j}^{-1}(\beta+) \quad \text{if and only if} \quad \hat{F}_{x_j}(y-) \leq \beta.
\]

But (2.2) remains valid if we replace ‘(y)’ with ‘(y-)’, so the inequality on the right hand side is equivalent to the following statements:

\[
\max_{s \geq j} \min_{r \leq j} \hat{F}_{rs}(y-) \leq \beta,
\]

for all \( s \geq j \), \( \hat{F}_{rs}(y-) \leq \beta \) for some \( r = r(s) \leq j \),

for all \( s \geq j \), \( y \leq \hat{F}_{rs}^{-1}(\beta+) \) for some \( r = r(s) \geq j \),

\[
y \leq \min_{s \geq j} \max_{r \leq j} \hat{F}_{rs}^{-1}(\beta+) = u_j.
\]

Hence \( \hat{F}_{x_j}^{-1}(\beta+) = u_j \). \(\square\)

### 4.3. Ratios of empirical and true probabilities

Let \( \hat{P}_n \) be the empirical distribution of independent r.v.s \( X_1, \ldots, X_n \) with distribution \( P \) on the real line. Further let \( \delta_n > 0 \) such that \( \delta_n \to 0 \) while \( n \delta_n / \log n \to \infty \) (as \( n \to \infty \)). Then there exist numbers \( \epsilon_n > 0 \) such that \( \epsilon_n \to 0 \) and

\[
\inf \left\{ \frac{\hat{P}_n(I)}{P(I)} : \text{intervals } I \subset \mathbb{R} \text{ with } P(I) \geq \delta_n \right\} \geq 1 - \epsilon_n \quad (4.6)
\]

with asymptotic probability one.
Proof. By means of the quantile transformation, this claim can be reduced to the special case of $P$ being the uniform distribution on $[0, 1]$. Then let $0 =: X_{(0)} < X_{(1)} < X_{(2)} < \cdots < X_{(n)} < X_{(n+1)} := 1$ be the augmented order statistics of $X_1, \ldots, X_n$. With $U_{jk} := X_{(k)} - X_{(j)} = P((X_{(j)}, X_{(k)}))$ for $0 \le j < k \le n$, an elementary consideration shows that the left-hand side of (4.6) equals
\[
\min\left\{ \frac{k - j - 1}{nU_{jk}} : 0 \le j < k \le n + 1, U_{jk} \ge \delta_n \right\}.
\] (4.7)

But $U_{jk} \sim \text{Beta}(k - j, n + 1 - k + j)$ (where $\text{Beta}(n + 1, 0) := \delta_1$), so
\[
\mathbb{E}(U_{jk}) = p_{jk} := \frac{k - j}{n + 1},
\]
and Proposition 2.1 of D{"u}mbgen [4] implies that for any $c > 0$,
\[
\mathbb{P}(U_{jk} > p_{jk} + \sqrt{2p_{jk}(1 - p_{jk})c + c}) \le \exp(- (n + 1)c).
\]
Consequently, setting $c_n := \gamma \log(n + 2)/(n + 1)$ for some fixed $\gamma > 2$,
\[
\mathbb{P}\left( U_{jk} \le p_{jk} + \sqrt{2p_{jk}(1 - p_{jk})c_n + c_n} \text{ for } 0 \le j < k \le n + 1 \right)
\]
\[
> 1 - (n + 2)^{-\gamma/2} \to 1.
\]

But elementary calculations show that $U_{jk} \le p_{jk} + \sqrt{2p_{jk}(1 - p_{jk})c_n + c_n}$ implies that
\[
p_{jk} \ge U_{jk} - \max(c_n, \sqrt{2c_nU_{jk}})
\]
and thus
\[
\frac{k - j - 1}{nU_{jk}} = \frac{(n + 1)p_{jk} - 1}{nU_{jk}} \ge 1 - \max(c_n/\sqrt{2c_nU_{jk}}, \sqrt{2c_n/\delta_n}) - (nU_{jk})^{-1}.
\]

Consequently, with asymptotic probability one, (4.7) is not smaller than $1 - \epsilon_n$, where
\[
\epsilon_n := \max(c_n/\delta_n, \sqrt{2c_n/\delta_n}) + (n\delta_n)^{-1} \to 0.
\]

### 4.4. Asymptotics

In what follows, we always work with the conditional distribution of $(Y_{ni})_{n=1}^\infty$ given $X_n$. Moreover, we tacitly assume that $X_n$ is a “good” vector in the sense that the event $A_n$ in Assumption (A.2) or (A’2$_{\infty}$) occurs.

To lighten the notation, we do not introduce an extra subscript $n$ for the weights $w_{rs}$ or the empirical distribution functions $\hat{F}_{rs}$. Furthermore, we define
\[
\overline{F}_{rs}(\cdot) := w_{rs}^{-1} \sum_{j=r}^s w_j F_{x_j}(\cdot).
\]

The norm $\| \cdot \|_\infty$ denotes the usual supremum norm of functions on the real line.

The proofs make use of the following exponential inequality which follows from Bretagnolle [1] and Hu [9].
Theorem 4.6. Let \( Y_1, Y_2, Y_3, \ldots \) be independent random variables with respective distribution functions \( F_1, F_2, F_3, \ldots \). For \( k \in \mathbb{N} \), let
\[
\hat{F}(\cdot) := \frac{1}{k} \sum_{i=1}^{k} 1_{[Y_i \leq \cdot]} \quad \text{and} \quad \bar{F}(\cdot) := \frac{1}{k} \sum_{i=1}^{k} F_i(\cdot).
\]
Then there exists a universal constant \( C_4 \leq \frac{25}{2}e \) such that for all \( \eta \geq 0 \),
\[
\mathbb{P}\left( \sqrt{k} \| \hat{F} - \bar{F} \|_\infty \geq \eta \right) \leq C_4 \exp(-2\eta^2).
\]

Corollary 4.7. Let
\[
M_n := \max_{1 \leq r \leq s \leq m} w_{rs}^{1/2} \| \hat{F}_{rs} - \bar{F}_{rs} \|_\infty.
\]
Then for any constant \( D > 1 \),
\[
\lim_{n \to \infty} \mathbb{P}(M_n \leq (D \log n)^{1/2}) = 1.
\]

Proof of Corollary 4.7. Note that the number \( M_n \) is the maximum of the \( \binom{m}{2} + m < (n+1)^2/2 \) quantities
\[
w_{rs}^{1/2} \| \hat{F}_{rs} - \bar{F}_{rs} \|_\infty,
\]
and we may apply Theorem 4.6 to each of them. Consequently,
\[
\mathbb{P}(M_n \geq \eta_n) \leq \sum_{1 \leq r \leq s \leq m} \mathbb{P}(w_{rs}^{1/2} \| \hat{F}_{rs} - \bar{F}_{rs} \|_\infty \geq \eta_n) \leq (C_4/2) \exp(2\log(n+1) - 2\eta_n^2)
\]
for arbitrary \( \eta_n \geq 0 \). But the right hand side converges to zero as \( n \to \infty \) if \( \eta_n = (D \log n)^{1/2} \) for fixed \( D > 1 \).

Proof of Theorem 3.3. Recall that \( \rho_n = \log(n)/n, \delta_n = C_3 \rho_n^{1/(2\alpha+1)} \) and \( I_n = \{ x \in I : [x \pm \delta_n] \subset I \} \). Recall also that we treat \( X_n \) as fixed and assume that the event \( A_n \) in Assumption (A.2) occurs. Let \( n \) be sufficiently large so that \( I_n \neq \emptyset \). For \( x \in I_n \) the indices
\[
r(x) := \min\{ j \in \{1, \ldots, m\} : x_j \geq x - \delta_n \},
\]
\[
j(x) := \max\{ j \in \{1, \ldots, m\} : x_j \leq x \}
\]
are well-defined, because \( [x - \delta_n, x] \) is a subinterval of \( I \) of length \( \delta_n \), so Assumption (A.2) guarantees that this interval contains at least one observation \( x_j \). Moreover, we have
\[
r(x) \leq j(x),
\]
\[
x - \delta_n \leq x_{r(x)} \leq x_{j(x)} \leq x,
\]
\[
w_{r(x)j(x)} = w_n([x - \delta_n, x]) \geq C_2 n \delta_n.
\]
Consequently, with $M_n$ as in Corollary 4.7, for any $y \in J$ we obtain the inequalities
\[
\hat{F}_{nx}(y) - F_x(y) \leq \hat{F}_{nx_j(x)}(y) - F_x(y) \\
= \min_{r \leq j(x)} \max_{s \geq j(x)} \hat{F}_{rs}(y) - F_x(y) \\
\leq \max_{s \geq j(x)} \hat{F}_{r(s)}(y) - F_x(y) \\
\leq w_{r(x)} M_n + \max_{s \geq j(x)} \hat{F}_{r(s)}(y) - F_x(y) \\
\leq (C_2 n \delta_n)^{-1/2} M_n + F_{x_{r(x)}}(y) - F_x(y) \\
\leq (C_2 n \delta_n)^{-1/2} M_n + C_1 \delta_n^\alpha.
\]

In the first step we used antitonicity of $\hat{x} \mapsto \hat{F}_{n\hat{x}}(y)$, in the second last step we used antitonicity of $x \mapsto F_x(y)$, and the last step utilizes Assumption (A.1).

But $\mathbb{P}(M_n \leq (D \log n)^{1/2}) \to 1$ for any fixed $D > 1$, and on the event $\{M_n \leq (D \log n)^{1/2}\}$, the previous considerations imply that
\[
\sup_{x \in I_n, y \in J} (\hat{F}_{nx}(y) - F_x(y)) \leq (C_2 n \delta_n)^{-1/2} (D \log n)^{1/2} + C_1 \delta_n^\alpha = C \rho_n^{\alpha/(2\alpha + 1)}
\]
with $C := (C_2 D / C_3)^{1/2} + C_1 C_3^\alpha$.

Analogously one can show that on $\{M_n \leq (D \log n)^{1/2}\}$,
\[
\sup_{x \in I_n, y \in J} (F_x(y) - \hat{F}_{nx}(y)) \leq (n \delta_n)^{-1/2} (D \log n)^{1/2} + C_1 \delta_n^\alpha = C \rho_n^{\alpha/(2\alpha + 1)}
\]
with the same constant $C$.

The proof of Theorem 3.4 is based on Theorem 3.3 and two elementary inequalities for distribution functions:

**Lemma 4.8.** Suppose that $F, G$ are distribution functions such that
\[
\|F - G\|_\infty \leq \Delta < 1.
\]

Then
\[
G^{-1}(\beta) \geq F^{-1}(\beta - \Delta), \quad \text{for } \Delta < \beta < 1,
\]
\[
G^{-1}(\beta^+) \geq F^{-1}((\beta + \Delta)^+), \quad \text{for } 0 < \beta < 1 - \Delta.
\]

**Lemma 4.9.** Suppose that $F$ is a distribution function so that, for given $0 \leq \beta_1 < \beta_2 \leq 1$ and $\kappa > 0$,
\[
F(y_2) - F(y_1) \geq \kappa (y_2 - y_1)
\]
for arbitrary $y_1 < y_2$ such that $F(y_1), F(y_2) \in (\beta_1, \beta_2)$. Then $F^{-1}(\beta) = F^{-1}(\beta^+)$ and
\[
|F^{-1}(\beta) - F^{-1}(\beta')| \leq \kappa^{-1} |\beta - \beta'|,
\]
for arbitrary $\beta, \beta' \in (\beta_1, \beta_2)$. 


Proof of Lemma 4.8. Let $\Delta < \beta < 1$ and $y < F^{-1}(\beta - \Delta)$. Then $F(y) < \beta - \Delta$ and thus
$$G(y) \leq F(y) + \Delta < \beta - \Delta + \Delta = \beta.$$ Therefore, we have $y < G^{-1}(\beta)$ and letting $y \to F^{-1}(\beta - \Delta)$ yields the first inequality.

Next, let $0 < \beta < 1 - \Delta$ and $y > F^{-1}((\beta + \Delta)+)$. Then $F(y-) > \beta + \Delta$ and thus
$$G(y-) \geq F(y-) - \Delta > \beta + \Delta - \Delta = \beta.$$ This gives $y > G^{-1}(\beta+)$, and letting $y \to F^{-1}((\beta - \Delta)+)$ proves the second claim. \hfill \Box

Proof of Lemma 4.9. Let $\beta, \beta' \in (\beta_1, \beta_2)$ be such that $\beta < \beta'$. Define $y_1 := F^{-1}(\beta)$ and $y_2 := F^{-1}(\beta')$, so that $y_1 \leq y_2$. If $y_1 = y_2$, then (4.8) is trivial. In case $y_1 < y_2$, we have, for all $h \in (0, y_2 - y_1]$, that
$$\beta_1 < \beta \leq F(y_1) \leq F(y_2 - h) \leq \beta' < \beta_2,$$ so that $F(y_1), F(y_2 - h) \in (\beta_1, \beta_2)$. Therefore, we get
$$\beta' - \beta \geq \lim_{h \downarrow 0} F(y_2 - h) - F(y_1) \geq \lim_{h \downarrow 0} \kappa(y_2 - h - y_1) = \kappa(F^{-1}(\beta') - F^{-1}(\beta)).$$ \hfill \Box

Proof of Theorem 3.4. With $\Delta_n := C \rho_n^{\alpha/(2\alpha+1)}$, we may write $B_n = (\beta_1 + \Delta_n, \beta_2 - \Delta_n)$. Let $n$ be large enough so that $I_n$ and $B_n$ are nondegenerate intervals; in particular, $\Delta_n < 1/2$. The proof of Theorem 3.3 reveals that $\mathbb{P}(A_n^*) \to 1$, where $A_n^*$ is the event that
$$\sup_{x \in I_n} \|\tilde{F}_{nx,k} - F_x\|_\infty \leq \Delta_n \quad \text{for } k = 1, 2.$$ Here $\tilde{F}_{nx,1}$ and $\tilde{F}_{nx,2}$ denote two extremal ways to extrapolate $\tilde{F}_{nx}$ from $x \in \{x_1, \ldots, x_m\}$ to arbitrary $x \in X$: With $x_0 := -\infty$ and $x_{m+1} := \infty$, we define
$$\tilde{F}_{nx,1} := \begin{cases} \tilde{F}_{nx,j} & \text{if } x_{j-1} < x \leq x_j, \ 1 \leq j \leq m, \\ 0 & \text{if } x > x_m, \end{cases}$$
$$\tilde{F}_{nx,2} := \begin{cases} 1 & \text{if } x < x_1, \\ \tilde{F}_{nx,j} & \text{if } x_j \leq x < x_{j+1}, \ 1 \leq j \leq m. \end{cases}$$ Then $\tilde{F}_{nx,1} \geq \tilde{F}_{nx} \geq \tilde{F}_{nx,2}$ for any choice of $(\tilde{F}_x)_{x \in X}$. The event $A_n^*$ implies that $\tilde{F}_{nx,k}$ is a proper distribution function for $k = 1, 2$ and all $x \in I_n$. Moreover, for $x \in I_n$ and $\beta \in B_n$, it follows from Lemmas 4.8 and 4.9 that
$$\tilde{Q}_x(\beta) \geq \tilde{F}^{-1}_{nx,1}(\beta) \geq F^{-1}_x(\beta - \Delta_n) \geq F_x^{-1}(\beta) - \kappa^{-1}\Delta_n,$$$$
\tilde{Q}_x(\beta) \leq \tilde{F}^{-1}_{nx,2}(\beta+) \leq F^{-1}_x((\beta + \Delta_n)+) \leq F_x^{-1}(\beta) + \kappa^{-1}\Delta_n.$$
Consequently,
\[
\mathbb{P}\left( \sup_{x \in I_n, \beta \in B_n} |\tilde{Q}_x(\beta) - Q_x(\beta)| > \kappa^{-1} \Delta_n \right) \geq \mathbb{P}(A_n^*) \to 1
\]
as \(n \to \infty\).

We now proceed to the proof of Theorem 3.5. Theorem 4.6 and Lemma 4.11 in the next subsection imply the following exponential inequality:

**Corollary 4.10.** With the same notation as in Theorem 4.6, for any \(D' \in (0, 2)\) there exists a universal constant \(D'' = D''(D')\) such that
\[
\mathbb{P}\left( \sup_{k \geq k_o} \|\hat{F}_k - \bar{F}_k\|_\infty \geq \eta \right) \leq D'' \exp(-D'k_o\eta^2)
\]
for all \(k_o \in \mathbb{N}\) and \(\eta \geq 0\).

**Proof of Theorem 3.5.** Let us define the indices
\[
r_n := \min\{j \in \{1, \ldots, m\} : x_j \geq x_o - \delta_n\}
\]
and
\[
j_n := \max\{j \in \{1, \ldots, m\} : x_j \leq x_o\}.
\]
Since we assume the event \(A_n^*\) in \((A', 2x_o)\) to occur, we know that
\[
x_o - \delta_n \leq x_{r_n} \leq x_{j_n} \leq x_o,
\]
\[
w_{r_n, j_n} = w_n([x_o - \delta_n, x_o]) \geq C_2n\delta_n > 0.
\]
One can easily deduce from Corollary 4.10 that
\[
M_n := \max_{j \geq j_n} w_{r_n, j}^{1/2} \|\hat{F}_{r_n, j} - \bar{F}_{r_n, j}\|_\infty = O_p(1).
\]
Consequently, for \(y \in J\),
\[
\hat{F}_{nx_o}(y) - F_{x_o}(y) \leq \hat{F}_{njn}(y) - F_{x_o}(y) \leq \hat{F}_{r_n, s}(y) - F_{x_o}(y) \leq w_{r_n, j_n}^{-1/2} M_n + \max_{s \geq j_n} \bar{F}_{r_n, s}(y) - F_{x_o}(y) \leq (C_2n\delta_n)^{-1/2} M_n + F_{x_o - \delta_n}(y) - F_{x_o}(y) \leq (C_2n\delta_n)^{-1/2} M_n + C_1\delta_n^\alpha.
\]
But the right hand side does not depend on \(y\) and is of order \(O_p((n\delta_n)^{-1/2} + \delta_n^\alpha) = O_p(n^{-\alpha/(2\alpha+1)})\). Consequently,
\[
\sup_{y \in J} (\hat{F}_{x_o}(y) - F_{x_o}(y)) = O_p(n^{-\alpha/(2\alpha+1)}).
\]
With analogous arguments one shows that \( \sup_{y \in J} \left( F_{x_0}(y) - \hat{F}_{x_0}(y) \right) \) is of order \( O_p(n^{-\alpha/(2\alpha+1)}) \) as well.

Proof of Theorem 3.6. The proof uses essentially the same arguments as the proof of Theorem 3.4. The main differences are that we replace \( I_n \) with \( \{x_o\} \) and \( \rho_n \) with \( n^{-1} \).

4.5. An exponential inequality for the LLN

We consider stochastically independent random elements \( Z_1, Z_2, Z_3, \ldots \) with values in a normed vector space \( (\mathcal{Z}, \| \cdot \|) \). Defining the partial sums \( S_0 := 0 \) and \( S_n := \sum_{i=1}^{n} Z_i \) for \( n \in \mathbb{N} \), we assume that \( \|S_b - S_a\| \) is measurable for arbitrary integers \( 0 \leq a < b \).

Lemma 4.11. Suppose that there are constants \( c > 0 \) and \( C \geq 1 \) such that for arbitrary integers \( 0 \leq a < b \) and real numbers \( \eta > 0 \),

\[
P(\|S_b - S_a\| > \eta) \leq C \exp(-c\eta^2/(b-a)). \tag{4.9}
\]

Then for arbitrary \( c' \in (0,c) \) there exists a constant \( C' \) such that

\[
P(\sup_{n \geq n_o} \|S_n/n\| \geq \eta) \leq C' \exp(-c'n_o \eta^2) \tag{4.10}
\]

for arbitrary numbers \( n_o, \eta > 0 \).

Corollary 4.10 is a consequence of this result, where \( Z_i := 1_{[Y_i \leq \cdot]} - F_i \) is a random bounded function on the real line, and \( c = 2 \).

Proof of Lemma 4.11. Note that the right hand side of (4.10) is continuous in \( \eta \geq 0 \) and \( n_o \geq 0 \), and it is not smaller than 1 in case of \( \eta = 0 \) or \( n_o = 0 \). Hence it suffices to verify that

\[
P(\sup_{n \geq n_o} \|S_n/n\| > \eta) \leq C' \exp(-c'n_o \eta^2) \tag{4.11}
\]

for arbitrary numbers \( n_o, \eta > 0 \).

The essential ingredient will be the following inequality: For arbitrary real numbers \( 0 \leq a < b \) and \( \eta > 0 \),

\[
P(\max_{a \leq n \leq b} \|S_n\| > \eta) \leq 2C \exp(-c\eta^2/(\sqrt{b} + \sqrt{b-a})) \tag{4.12}
\]

(with the maximum over the empty set interpreted as 0). To verify this, it suffices to consider the case of \( a \) and \( b \) being integers; otherwise one could replace \( a \) with \( \lfloor a \rfloor \) and \( b \) with \( \lfloor b \rfloor \), and this would even decrease the term \( \sqrt{b} + \sqrt{b-a} \) in (4.12). Define the stopping time

\[
\tau := \min(\{n \in \{a, \ldots, b\} : \|S_n\| > \eta\} \cup \{\infty\}).
\]
Then, for $0 < \lambda < 1$,
\[
\mathbb{P}(\max_{a \leq n \leq b} \|S_n\| > \eta) = \mathbb{P}(\tau \leq b)
\]
\[
\leq \mathbb{P}(\|S_b\| > \lambda \eta) + \mathbb{P}(\tau \leq b, \|S_b\| \leq \lambda \eta)
\]
\[
= \mathbb{P}(\|S_b\| > \lambda \eta) + \sum_{n=a}^{b-1} \mathbb{P}(\tau = n, \|S_b\| \leq \lambda \eta)
\]
\[
\leq \mathbb{P}(\|S_b\| > \lambda \eta) + \sum_{n=a}^{b-1} \mathbb{P}(\tau = n, \|S_n - S_b\| > (1 - \lambda) \eta)
\]
\[
= \mathbb{P}(\|S_b\| > \lambda \eta) + \sum_{n=a}^{b-1} \mathbb{P}(\tau = n) \mathbb{P}(\|S_n - S_b\| > (1 - \lambda) \eta)
\]
\[
\leq C \exp\left(-\frac{c \lambda^2 \eta^2}{b}\right) + \sum_{n=a}^{b-1} \mathbb{P}(\tau = n) C \exp\left(-\frac{c(1 - \lambda)^2 \eta^2}{b - a}\right)
\]
\[
\leq C \exp\left(-\frac{c \lambda^2 \eta^2}{b}\right) + C \exp\left(-\frac{c(1 - \lambda)^2 \eta^2}{b - a}\right).
\]

Here the fourth last step follows from the triangle inequality for $\|\cdot\|$: $\|S_n - S_b\| \geq \|S_n\| - \|S_b\| > \eta - \lambda \eta$ in case of $\tau = n$ and $\|S_b\| \leq \lambda \eta$. The third last step follows from independence of the $Z_i$ and the fact that the event $\{\tau = n\}$ depends on $Z_n, \ldots, Z_b$, whereas $\|S_n - S_b\|$ is a function of $Z_{n+1}, \ldots, Z_b$. If we take
\[
\lambda := \frac{\sqrt{b}}{\sqrt{b + b - a}},
\]
then the two exponents in our inequality are identical, and we obtain (4.12).

Since $c' < c$, the constant
\[
\beta := \frac{(c/c' + 1)^2}{4c/c'}
\]
satisfies $\beta > 1$ and
\[
c' = \frac{c}{(\sqrt{\beta} + \sqrt{\beta - 1})^2}.
\]

With (4.12) at hand, we may argue that for arbitrary numbers $n_o > 0$,
\[
\mathbb{P}\left(\sup_{n \geq n_o} \|S_n/n\| > \eta\right) \leq \sum_{k=0}^{\infty} \mathbb{P}\left(\max_{\beta^k n_o \leq n \leq \beta^{k+1} n_o} \|S_n\| > \beta^k n_o \eta\right)
\]
\[
\leq 2C \sum_{k=0}^{\infty} \exp\left(-\frac{c \beta^{2k} n_o^2 \eta^2}{(\sqrt{\beta^{k+1} n_o + \sqrt{\beta^{k+1} n_o - \beta^k n_o})^2}}\right)
\]
\[
= 2C \sum_{k=0}^{\infty} \exp\left(-\frac{c \beta^k n_o \eta^2}{(\sqrt{\beta + \sqrt{\beta - 1})^2}}\right).
\]
\[
= 2C \sum_{k=0}^{\infty} \exp(-p(\eta)\beta^k),
\]
where \(p(\eta) := c' n_0 \eta^2 > 0\). Since \(\beta^x\) is increasing in \(x \geq 0\), we find the upper bound
\[
\sum_{k=1}^{\infty} \exp(-p(\eta)\beta^k) \leq \int_{0}^{\infty} \exp(-p(\eta)\beta^x) \, dx
\]
\[
= (\log \beta)^{-1} \int_{0}^{\infty} \exp(-p(\eta)e^y) \, dy
\]
\[
\leq (\log \beta)^{-1} \int_{0}^{\infty} \exp(-p(\eta)(1+y)) \, dy
\]
\[
= \frac{1}{p(\eta) \log \beta} \exp(-p(\eta)),
\]
which yields
\[
\mathbb{P}\left( \sup_{n \geq n_0} \|S_n/n\| > \eta \right) \leq 2C\left( 1 + \frac{1}{p(\eta) \log \beta} \right) \exp(-p(\eta)).
\]
For a number \(p_o > 0\) to be specified later, the bound above is not greater than
\[
2C\left( 1 + \frac{1}{p_o \log \beta} \right) \exp(-p(\eta)) = 2C\left( 1 + \frac{1}{p_o \log \beta} \right) \exp(-c'n_0\eta^2)
\]
whenever \(p(\eta) \geq p_o\). But in case of \(p(\eta) \leq p_o\), the latter bound is at least
\[
2C\left( 1 + \frac{1}{p_o \log \beta} \right) \exp(-p_o) \geq 1
\]
if we set \(p_o := \min\{ (\log \beta)^{-1}, \log(4C) \}\). Consequently, with this choice of \(p_o\), (4.11) is true with \(C' := 2C\left( 1 + (p_o \log \beta)^{-1} \right)\).

Acknowledgements

The authors are grateful to Geurt Jongbloed for drawing their attention to El Barmi and Mukerjee [6] and to Johanna Ziegel for stimulating discussions. We also thank two reviewers for constructive comments.

References

[1] Bretagnolle, J. (1981). Statistique de Kolmogorov-Smirnov pour un échantillon non équiréparti. In Statistical and physical aspects of Gaussian processes (Saint-Flour, 1980). Colloq. Internat. CNRS 307 39–44. CNRS, Paris. MR0716526
[2] Brunk, H. D., Franck, W. E., Hanson, D. L. and Hogg, R. V. (1966). Maximum likelihood estimation of the distributions of two stochastically ordered random variables. *J. Amer. Statist. Assoc.* **61** 1067–1080. MR0205367

[3] Casady, R. J. and Cryer, J. D. (1976). Monotone percentile regression. *Ann. Statist.* **4** 532–541. MR0403050

[4] Dumbgen, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. *Ann. Statist.* **26** 288–314. MR1611768

[5] Dumbgen, L. and Kovac, A. (2009). Extensions of smoothing via taut strings. *Electron. J. Stat.* **3** 41–75. MR2471586

[6] El Barmi, H. and Mukerjee, H. (2005). Inferences under a stochastic ordering constraint: the k-sample case. *J. Amer. Statist. Assoc.* **100** 252–261. MR2156835

[7] Henzi, A. (2018). Isotonic Distributional Regression (IDR): A powerful nonparametric calibration technique, Master’s thesis, University of Bern.

[8] Hogg, R. V. (1965). On models and hypotheses with restricted alternatives. *J. Amer. Statist. Assoc.* **60** 1153–1162. MR0191035

[9] Hu, I. (1985). A uniform bound for the tail probability of Kolmogorov-Smirnov statistics. *Ann. Statist.* **13** 821–826. MR0790579

[10] Koenker, R. and Bassett, G. Jr. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644

[11] Kuczmaszski, R. J., Ogden, C. L., Guo, S. S., Grummer-Strawn, L. M., Flegal, K. M., Mei, Z., Wei, R., Curtin, L. R., Roche, A. F. and Johnson, C. L. (2002). CDC Growth Charts for the United States: Methods and development. *Vital Health Stat.* **246**.

[12] Mühlemann, A., Jordan, A. I. and Ziegel, J. F. (2019). Optimal solutions to the isotonic regression problem. Preprint, arXiv:1904.04761.

[13] Mukerjee, H. (1993). An improved monotone conditional quantile estimator. *Ann. Statist.* **21** 924–942. MR1232526

[14] Poiraud-Casanova, S. and Thomas-Agnan, C. (2000). About monotone regression quantiles. *Statist. Probab. Lett.* **48** 101–104. MR1767613

[15] Robertson, T. and Wright, F. T. (1980). Algorithms in order restricted statistical inference and the Cauchy mean value property. *Ann. Statist.* **8** 645–651. MR0568276

[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. MR0961262

[17] Wright, F. T. (1984). The asymptotic behavior of monotone percentile regression estimates. *Canad. J. Statist.* **12** 229–236. MR0776113