<table>
<thead>
<tr>
<th><strong>Statistica Sinica Preprint No:</strong> SS-2017-0276</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td><strong>Manuscript ID</strong></td>
</tr>
<tr>
<td><strong>URL</strong></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
</tr>
<tr>
<td><strong>Complete List of Authors</strong></td>
</tr>
<tr>
<td><strong>Corresponding Author</strong></td>
</tr>
<tr>
<td><strong>E-mail</strong></td>
</tr>
</tbody>
</table>

Notice: Accepted version subject to English editing.
NONPARAMETRIC ESTIMATION OF TIME–DEPENDENT QUANTILES IN A SIMULATION MODEL

Felix Heimrich, Michael Kohler and Lisa Kristl

Technische Universität Darmstadt

Abstract: The problem of estimating a time–dependent quantile at each time point \( t \in [0, 1] \), given independent samples of a stochastic process at discrete time points in \([0, 1]\), is considered. It is assumed that the quantiles depend smoothly on \( t \). Results concerning the rate of convergence of quantile estimates based on a local average estimate of the time dependent cumulative distribution functions are presented. In a simulation model importance sampling is applied to construct estimates which achieve better rates of convergences. The finite sample size performance of the estimates is illustrated by applying them to simulated data.

Key words and phrases: Conditional quantile estimation, importance sampling, rate of convergence.

1. Introduction

Let \((Y_t)_{t \in [0,1]}\) be an \( \mathbb{R} \)-valued stochastic process. For equidistant time
points \( t_1, \ldots, t_n \in [0, 1] \) we assume that we have given a data set

\[
\mathcal{D}_n = \left\{ Y_1^{(t_1)}, \ldots, Y_n^{(t_n)} \right\},
\]

(1.1)

where \( Y_1^{(t_1)}, \ldots, Y_n^{(t_n)} \) are independent and where

\[
P_{Y_k^{(t_k)}} = P_{Y_{tk}}.
\]

Let \( G_{Y_t}(y) = P(Y_t \leq y) \) be the cumulative distribution function (cdf.) of \( Y_t \), and for \( \alpha \in (0, 1) \) let

\[
q_{Y_t,\alpha} = \inf\{ y \in \mathbb{R} : G_{Y_t}(y) \geq \alpha \}
\]

be the \( \alpha \)-quantile of \( Y_t \) for \( t \in [0, 1] \). Given the data set \( \mathcal{D}_n \) we are interested in constructing estimates \( \hat{q}_{Y_t,\alpha} = \hat{q}_{Y_t,\alpha}(\mathcal{D}_n) \) of \( q_{Y_t,\alpha} \) such that we have a “small” error

\[
\sup_{t \in [0,1]} |\hat{q}_{Y_t,\alpha} - q_{Y_t,\alpha}|
\]

(1.2)

Before we explain how to construct estimates for the above estimation problem, we will next illustrate the practical relevance of this problem by an example. Here we consider a problem, which occurred in the Collaborative Research Centre 805. The German CRC 805 works on controlling uncertainty in load-carrying structures in Mechanical Engineering. To test different approaches to control uncertainty, the CRC 805 has designed a demonstrator model, which is shown in Figure 1.
Figure 1: The suspension strut demonstrator of CRC 805.

The demonstrator model is an academic example of a suspension strut, such as an aircraft landing gear. It is designed in two versions, a virtual computer experiment and a real experimental setup. Figure 2 shows a photo of the real experimental setup. In the experiments a modular spring damper system is suspended on a frame and falls down on the base of the frame. In doing so sensors measure different parameters such as acceleration, absolute position of the modular spring damper system and the force at the point of impact. Predicting this force is important to calculate the stress and its deviation in order to determine the correct load capacity for the usage phase of the product already in the development phase. If one component
of a suspension strut is time-dependent and uncertain, this has an effect on this force and should be investigated before a prototype is produced. In this context mechanical engineers need information about a whole time interval, e.g. to guarantee operational stability of the suspension strut. To support them time-dependent quantile estimation can be used to estimate \( \alpha \)-quantiles for arbitrary time points in the considered time-intervall.

We will investigate the impact of an aging spring, i.e. an over time decreasing spring stiffness \( X_t \), on the force at the point of impact \( Y_t = m(t, X_t) \). To do this, we use simulated data generated by the virtual demon-
strator. Since we want to illustrate how to support mechanical engineers in the product development process, no measured input data of an aging spring of this system is assumed to be given. In a time–invariant system the spring stiffness is assumed to be normally distributed with expectation \( \mu = 35000 \, [\text{N/m}] \) and standard deviation \( \sigma = 1166.67 \, [\text{N/m}] \) (c.f. Schüeller (2007)). It seems reasonable that the spring will weaken over time, when it is used continuously. Therefore, in this academic example, we assume that the spring constant deteriorates over time exponentially as Zill and Wright (2009) do in Chapter 3.8.1. More precisely, we assume that the spring stiffness \( X_t \) is normally distributed with expectation \( \mu_t = 35000 \times \exp(-0.5 \times t) \) [N/m] and standard deviation \( \sigma_t = 1166.67 \, [\text{N/m}] \).

For given spring stiffness, we can use physical principles to model the force at the point of impact over the time by using partial differential equations. These differential equations can be solved numerically. We do this implicitly by using the routine \textit{RecurDyn} of the software \textit{Siemens NX} and are then able to compute values of the force at the point of impact \( Y_t \) for various (randomly) chosen values for the spring stiffness (chosen according to the distribution described above) and various time points \( t \). In Subsection 4.2 we will generate \( n = 300 \) values for the force at the point of impact \( Y_t \) in this way, and will use them to estimate the time dependent
0.95-quantile $q_t$ of the force at the point of impact continuously in the time $t$.

For computation of one value of $Y_t$, we repeat here the complete computation of the force with independent data, so that the data set (1.1) is indeed independent in our example. Furthermore, since we can choose the time points which we consider in our simulation, the choice of equidistant time points is no problem in this practical application. And finally it is important to consider the estimation error (1.2), since this allows us to make some statistical inference on the force at the point of impact over the whole examined time interval.

1.1 Main results

In our first result we use plug-in estimators of $q_{Y_t,\alpha}$ based on local averaging estimators of $G_{Y_t}$ in order to define our quantile estimates. More precisely, let $K : \mathbb{R} \rightarrow \mathbb{R}$ be a nonnegative kernel function (e.g., the uniform kernel $K(z) = 1/2 \cdot 1_{[-1,1]}(z)$ or the Epanechnikov kernel $K(z) = 3/4 \cdot (1 - z^2) \cdot 1_{[-1,1]}(z)$). We estimate the cumulative distribution function of $Y_t$, i.e.,

$$G_{Y_t}(y) = P(Y_t \leq y) = E \{ 1_{(-\infty,y]}(Y_t) \}$$
by the local average estimator
\[ \hat{G}_{Y_t}(y) = \frac{\sum_{i=1}^{n} 1_{(-\infty,y]}(Y_{t_i}) \cdot K\left(\frac{t-t_i}{h_n}\right)}{\sum_{j=1}^{n} K\left(\frac{t-t_j}{h_n}\right)} \]  
(1.3)
and use the following plug-in estimator of \( q_{Y_t,\alpha} \):
\[ \hat{q}_{Y_t,\alpha} = \inf\{y \in \mathbb{R} : \hat{G}_{Y_t}(y) \geq \alpha\}. \]  
(1.4)
Under the assumptions that \( G_{Y_t}(y) \) is Hölder smooth with exponent \( p \in (0, 1] \) (as a function of \( t \in [0, 1] \)), and that a density of \( Y_t \) exists which is bounded away from zero and infinity in a neighborhood of \( q_{Y_t,\alpha} \), we show that for a suitable chosen bandwidth \( h_n \) and kernel \( K : \mathbb{R} \rightarrow \mathbb{R} \), the supremum norm error (1.2) of this estimate converges to zero in probability with rate \( (\log(n)/n)^{p/(2p+1)} \).

In our second result we show that this rate of convergence can be improved in a simulation model using importance sampling. Here we assume that \( Y_t \) is given by
\[ Y_t = m(t, X_t), \]
where \( X_t \) is an \( \mathbb{R}^d \)-valued random variable with a density \( f(t, \cdot) : \mathbb{R}^d \rightarrow \mathbb{R} \) and the function \( m : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R} \) is a costly to evaluate blackbox function. In this framework, we construct an importance sampling variant of the above plug-in quantile estimator, which is based on an initial quantile estimator and a suitably chosen estimator (surrogate) \( m_n \) of \( m \).
result here is that under suitable assumptions on \( f \) and \( m \) we can achieve in the case of Hölder smooth \( G_{Y_t}(y) \) (with exponent \( p \in (0, 1] \) and as a function of \( t \in [0, 1] \)) the rate of convergence \((\log(n)/n)^{p(4p+1)/(2p+1)^2}\). This rate of convergence is achievable for an estimate which is based on at most \( n \) evaluations of \( m \) (as the estimate \((1.4)\)). The finite sample size performance of our estimates are illustrated by applying them to simulated data.

1.2 Discussion of related results

Time-dependent quantile estimation can be regarded as conditional quantile estimation for fixed design, where we condition on the time \( t \). A short introduction into conditional quantile estimation is presented in [Yu, Lu and Stander (2003)]. Plug-in conditional quantile estimators have been considered already in quite a few papers. [? ] showed consistency in probability, [Stute (1986)] proved asymptotic normality, and [Bhattacharya and Gangopadhyay (1990)] used a Bahadur-type representation, c.f. [Bahadur (1996)], to show asymptotic normality. A double-kernel approach was presented by [Yu and Jones (1998)], who analyzed the mean squared error of their estimator.

Other conditional quantile estimation approaches are discussed for example in [Koenker and Bassett (1978)], who proposed a quantile regression
estimator, and Mehra, Rao, and Upadrasta (1991), who presented a smooth conditional quantile estimator, showed its asymptotic normality and analyzed its pointwise almost sure rate of convergence. Xiang (1996) also proposed a new kernel estimator of a conditional quantile and derived the same pointwise almost sure rate of convergence as Mehra, Rao, and Upadrasta (1991) under weaker assumptions. Additional results on quantile regression estimators can be found in Chaudhuri (1991), Fan, Yao and Tong (1996), Yu and Jones (1998), Li, Liu and Zhu (2007) and Plumlee and Tuo (2014) and the literature cited therein. In contrast to the articles cited above, we analyze the rate of convergence in probability of the supremum norm error of our quantile estimates.

As an estimate $m_n$ of $m$ any kind of nonparametric regression estimate can be chosen, cf., e.g., Györfi et al. (2002).

Importance sampling is a well-known variance reduction technique, which was originally introduced in order to improve the rate of convergence of estimates of expectations, cf., e.g., Glasserman (2004). The main idea in our setting is to consider, instead of $Y_t$, a real-valued random variable $Z_t$, where the distribution of $Z_t$ is chosen such that $Z_t$ is concentrated in a region of the sample space, which has a strong effect on the estimation of $q_{Y_t,\alpha}$. Quantile estimation based on importance sampling has been studied
by Cannamela, Garnier, and Ioos (2008), Egloff and Leippold (2010) and Morio (2012). However, only Egloff and Leippold (2010) derived theoretical properties about their estimate, such as consistency, but did not analyze the rate of convergence of their estimate. Kohler et al. (2018) studied rates of convergences of importance sampling quantile estimators based on surrogate models, but did not consider a time-dependent setting respectively conditional quantile estimation.

1.3 Outline

In Section 2 the rate of convergence of the first estimate is presented. In Section 3 a time-dependent simulation model is considered and the construction of a time-dependent importance sampling quantile estimate is described and its rate of convergence is analyzed. Section 4 illustrates the finite sample size behavior of the two presented estimates by applying them to simulated data and to the application introduced in the introduction.

2. Estimation of time-dependent quantiles

In this section we analyze the rate of convergence of our local averaging plug-in quantile estimate. As it is well-known in nonparametric curve estimation, the derivation of any non-trivial rate of convergence result requires
various regularity conditions on the data to hold.

In the sequel we will make the following assumptions on the data used for the estimation of the quantile.

(A1) Let \((Y_t)_{t \in [0,1]}\) be an \(\mathbb{R}\)-valued stochastic process, let \(t_1, \ldots, t_n \in [0, 1]\) be equidistant and set
\[
\mathcal{D}_n = \{Y_{t_1}^{(t_1)}, \ldots, Y_{t_n}^{(t_n)}\},
\]
where \(Y_{t_1}^{(t_1)}, \ldots, Y_{t_n}^{(t_n)}\) are independent and where
\[
P_{Y_{t_k}^{(t_k)}} = P_{Y_{t_k}}.
\]

(A2) Assume that \(Y_t\) has a density \(g(t, \cdot) : \mathbb{R} \to \mathbb{R}\) with respect to the Lebesgue-Borel measure, which is uniformly bounded away from zero in a neighborhood of \(q_{Y_t, \alpha}\), i.e., for some \(\epsilon > 0\) there exists a constant \(c_1 > 0\) such that
\[
\inf_{t \in [0,1]} \inf_{u \in (q_{Y_t, \alpha} - \epsilon, q_{Y_t, \alpha} + \epsilon)} g(t, u) \geq c_1. \tag{2.1}
\]

(A3) Assume that the function \(t \mapsto G_{Y_t}(y)\) is Hölder continuous with Hölder constant \(C > 0\) and Hölder exponent \(p \in (0, 1]\) for \(y \in \mathbb{R}\), i.e., assume
\[
|G_{Y_s}(y) - G_{Y_t}(y)| \leq C|s - t|^p \quad \text{for all } s, t \in [0,1] \text{ and } y \in \mathbb{R}. \tag{2.2}
\]
Here \((A_3)\) is a standard smoothness assumption, which we need to derive a non-trivial result on the rate of convergence. Assumption \((A_2)\) is needed since we want to analyze the error of our estimate in the supremum norm. The condition \((A_1)\) is used to simplify our proofs, as explained in the introduction it is possible to ensure that this condition holds by generating our data in a proper way.

Next we describe the assumptions which we need for the parameters of our estimate, i.e., the assumptions on the kernel and on the bandwidth.

(EST1) Let \(K\) a nonnegative kernel function \(K : \mathbb{R} \to \mathbb{R}\), which is left-continuous on \(\mathbb{R}_+\) and monotonically decreasing on \(\mathbb{R}_+\), and satisfies

\[
K(z) = K(-z) \quad (z \in \mathbb{R}),
\]

\[
c_2 \cdot 1_{[-\alpha,\alpha]}(z) \leq K(z) \leq c_3 \cdot 1_{[-\beta,\beta]}(z) \quad (z \in \mathbb{R})
\]

(2.3)

for some constants \(\alpha, \beta, c_2, c_3 \in \mathbb{R}_+ \setminus \{0\} \).

(EST2) Let \(h_n > 0\) be such that

\[
h_n \to 0 \quad (n \to \infty),
\]

\[
\frac{n \cdot h_n}{\log(n)} \to \infty \quad (n \to \infty).
\]

(2.5)
Theorem 1. Let $\alpha \in (0, 1)$. Let $(Y_t)_{t \in [0,1]}$ be an $\mathbb{R}$-valued stochastic process and let $G_{Y_t}$ be the cdf of $Y_t$ for $t \in [0,1]$. Let $q_{Y_t,\alpha}$ be the $\alpha$-quantile of $Y_t$. Let $n \in \mathbb{N}$ and set $t_k = k/n$ ($k = 1, \ldots, n$). Assume that (A1), (A2) and (A3) hold. Let the estimator $\hat{q}_{Y_t,\alpha}$ be defined by (1.3) and (1.4) where the kernel and the bandwidth satisfy (EST1) and (EST2). Then we have for a constant $c_4 > 0$

$$
P \left( \sup_{t \in [0,1]} |\hat{q}_{Y_t,\alpha} - q_{Y_t,\alpha}| > c_4 \cdot \left( \sqrt{\frac{\log(n)}{n \cdot h_n}} + h_n^{p} \right) \right) \rightarrow 0 \quad \text{for} \quad n \rightarrow \infty.
$$

In particular, if we set $h_n = c_5 \cdot (\log(n)/n)^{1/(2p+1)}$ for a constant $c_5 > 0$, there exists a constant $c_6 > 0$, such that

$$
P \left( \sup_{t \in [0,1]} |\hat{q}_{Y_t,\alpha} - q_{Y_t,\alpha}| > c_6 \cdot \left( \frac{\log(n)}{n} \right)^{p/(2p+1)} \right) \rightarrow 0 \quad \text{for} \quad n \rightarrow \infty.
$$

Remark 1. To derive a nontrivial statement about the rate of convergence, smoothness assumptions, such as (2.2), are required (c.f. Devroye (1982)). In Theorem 1 we have shown the same rate of convergence as the optimal minimax rate of convergence for estimation of a Hölder continuous function (with exponent $p \in (0,1]$) on a compact subset of $\mathbb{R}$ in supremum norm derived in ?.

Remark 2. To apply the time–dependent quantile estimator in practice, the bandwidth $h_n$ has to be selected in a data-driven way. We suggest to choose $h_n$ in an optimal way concerning the estimation of the time–
dependent cdf. \( G_Y \) by \( \hat{G}_Y \) using a version of the well-known splitting the sample technique (cf., e.g., Chapter 7 in Györfi et al. (2002)). Assume that for each of the equidistant time–points \( t_k \) \((k = 1, \ldots, n)\) we have given an additional random variable \( Y_{k,2}^{(t_k)} \) such that \( Y_{k,1}^{(t_k)} \) and \( Y_{k,2}^{(t_k)} \) are independent and identically distributed. Let \( y \) be the \( \alpha \)-quantile of the empirical cdf. corresponding to the data \( Y_{1,1}^{(t_1)}, \ldots, Y_{n,1}^{(t_n)} \) and define \( \hat{G}_{Y_{tk}} \) by (1.3) using \( h_n \) for \( k = 1, \ldots, n \). Then we choose the optimal bandwidth \( h_n^* \) from a finite set of possible bandwidths \( H_n \) by minimizing

\[
\Delta_{h_n} = \frac{1}{n} \sum_{k=1}^{n} \left| \mathbb{I}_{\{Y_{k,2}^{(t_k)} \leq y\}} - \hat{G}_{Y_{tk}}(y) \right|^2.
\]

**Remark 3.** It follows from the proof of Theorem 1 that the result also holds in case that \( t_k \in [0, 1] \) are chosen such that

\[
\frac{c_7}{n} \leq \left| t_k - \frac{k}{n} \right| \leq \frac{c_8}{n}
\]

for some constants \( c_7 > 0 \) and \( c_8 > 0 \).

### 3. Application of importance sampling in a simulation model

In this section we demonstrate that by using more assumptions on the given data, we can construct via importance sampling an estimate which achieves (basically for the same smoothness of the underlying cumulative distribution function) a better rate of convergence than in Theorem 1. Here
our main trick to achieve a better rate of convergence is to assume that
(as in our application in the introduction) $Y_t$ is given by some $m(t, X_t)$,
where we know the distribution of $X_t$ and can observe at the same time
$X_t$ and $m(t, X_t)$. Then we change the density of $X_t$ in such a way that it
is concentrated on a set (with small measure) which is important for the
estimation of the quantile of $m(t, X_t)$ (cf., density $h(t, x)$ defined below).

More precisely, let $(X_t)_{t \in [0,1]}$ be an $\mathbb{R}^d$–valued stochastic process and
assume that $X_t$ has a density $f(t, \cdot) : \mathbb{R}^d \to \mathbb{R}$ with respect to the Lebesgue-
Borel measure. Let $m : [0, 1] \times \mathbb{R}^d \to \mathbb{R}$ be a function, which is costly to
compute, and define $Y_t$ by

$$Y_t = m(t, X_t).$$

In the sequel we will assume that we have given independent data sets $D_{n,1}$
and $D_{n,2}$ of the form

$$D_{n,1} = \left\{ \left( t_1, X_{1,1}^{(t_1)}, Y_{1,1}^{(t_1)} \right), \ldots, \left( t_n, X_{n,1}^{(t_n)}, Y_{n,1}^{(t_n)} \right) \right\},$$

$$D_{n,2} = \left\{ \left( t_1, X_{1,2}^{(t_1)}, Y_{1,2}^{(t_1)} \right), \ldots, \left( t_n, X_{n,2}^{(t_n)}, Y_{n,2}^{(t_n)} \right) \right\},$$

(3.1)

where $t_k = k/n$ ($k = 1, \ldots, n$),

$$P(X_{k,i}^{(t_k)}, Y_{k,i}^{(t_k)}) = P(X_{k,i}, Y_{k,i})$$

for $i = 1, 2$, $k = 1, \ldots, n$ and where

$$(X_{1,1}^{(t_1)}, Y_{1,1}^{(t_1)}), \ldots, (X_{n,1}^{(t_n)}, Y_{n,1}^{(t_n)}), (X_{1,2}^{(t_1)}, Y_{1,2}^{(t_1)}), \ldots, (X_{n,2}^{(t_n)}, Y_{n,2}^{(t_n)})$$

(3.2)
are independent. I.e., we have given two independent samples of \((X_{t_k}, Y_{t_k})\) at each time point \(t_k \ (k = 1, \ldots, n)\). Here again this independence assumption can be justified in our application by generating the data properly.

Furthermore, we assume that we have given independent random variables \(X_{k,3}, X_{k,4}, \ldots\) distributed as \(X_{t_k}\) for \(k = 1, \ldots, n\) and that we are allowed to evaluate \(m\) at \(n\) additional time points. Let \(m_n\) be an estimate of \(m\) depending on the data set \(D_{n,2}\) and satisfying

\[
P\left(\sup_{t \in [0,1], \ x \in K_n} |m_n(t, x) - m(t, x)| \geq \beta_n \right) \to 0 \ (n \to \infty) \quad (3.3)
\]

for some sequence \((\beta_n)_{n \in \mathbb{N}} \in \mathbb{R}_+\) and some \(K_n \subseteq \mathbb{R}^d\). Let \(\hat{q}_{Yt,\alpha}\) be an estimate of \(q_{Yt,\alpha}\) depending on the data set \(D_{n,1}\) and satisfying

\[
P\left(\sup_{t \in [0,1]} |\hat{q}_{Yt,\alpha} - q_{Yt,\alpha}| \geq \eta_n \right) \to 0 \ (n \to \infty), \quad (3.4)
\]

for some sequence \((\eta_n)_{n \in \mathbb{N}} \in \mathbb{R}_+\), which converges to zero as \(n\) goes to infinity, e.g. the estimator \(\hat{q}_{Yt,\alpha}\) defined in \([1,4]\) and \(\eta_n = 2 \cdot c_6 \cdot (\log(n)/n)^{p/(2p+1)}\) (cf., Theorem 1). Assume that

\[
P(\exists t \in [0,1] : X_t \notin K_n) = O(\beta_n + \eta_n). \quad (3.5)
\]

Set

\[
h(t, x) = \frac{1}{c_t} \cdot \left(1_{\{x \in K_n : \hat{q}_{Yt,\alpha} - 3\beta_n - 3\eta_n \leq m_n(t, x) \leq \hat{q}_{Yt,\alpha} + 3\beta_n + 3\eta_n\}} + 1_{\{x \notin K_n\}}\right) \cdot f(t, x),
\]
where

\[ c_t = \int_{\mathbb{R}^d} \left( \mathbf{1}_{\{x \in K_n : \hat{q}_{Y_t,\alpha} - 3\beta_n - 3\eta_n \leq m_n(t,x) \leq \hat{q}_{Y_t,\alpha} + 3\beta_n + 3\eta_n \}} + \mathbf{1}_{\{x \notin K_n \}} \right) \cdot f(t, x) \, dx. \]

Set \( t_k = k/n \) for \( k = 1, \ldots, n \). Let \( Z_t \) be a random variable with density \( h(t, \cdot) \), and let \( Z_{1}^{(t_1)}, \ldots, Z_{n}^{(t_n)} \) be independent random variables such that

\[ \mathbb{P} Z_k^{(t_k)} = \mathbb{P} Z_{tk} \]

for \( k = 1, \ldots, n \). Define (for some \( h_{n,1} > 0 \))

\[ \hat{G}^{(IS)}_{Y_t}(y) = \frac{\sum_{i=1}^{n} (c_{t_i} \cdot \mathbf{1}_{\{m(t_i,Z_{ti}) \leq y\}} + b_{t_i}) \cdot K \left( \frac{t-t_i}{h_{n,1}} \right)}{\sum_{j=1}^{n} K \left( \frac{t-t_j}{h_{n,1}} \right)}, \quad (3.6) \]

where

\[ b_t = \int_{\mathbb{R}^d} \mathbf{1}_{\{x \in K_n : m_n(t,x) < \hat{q}_{Y_t,\alpha} - 3\beta_n - 3\eta_n \}} \cdot f(t, x) \, dx \]

for \( t \in [0, 1] \), and define the plug-in importance sampling estimate of \( q_{Y_t,\alpha} \) by

\[ \hat{q}^{(IS)}_{Y_t,\alpha} = \inf \{ y \in \mathbb{R} : \hat{G}^{(IS)}_{Y_t}(y) \geq \alpha \}. \quad (3.7) \]

In order to analyze the rate of convergence of this estimate, we impose besides the conditions mentioned above condition (A3) and the following two assumptions on the underlying data.

(A4) Assume that \( Y_t \) has a density \( g(t, \cdot) : \mathbb{R} \to \mathbb{R} \), which is continuous as well as uniformly bounded away from zero in a neighborhood of \( q_{Y_t,\alpha} \).
and which is uniformly bounded from above, i.e. it is assumed that (2.1) holds and that there exists a constant $c_9 > 0$ such that

$$
\sup_{t \in [0,1]} \sup_{u \in \mathbb{R}} g(t, u) \leq c_9.
$$

(3.8)

(A5) For $\alpha \in (0, 1)$ let $q_{Y_t, \alpha}$ be the $\alpha$-quantile of $Y_t$ for $t \in [0,1]$, and assume that the function $t \mapsto q_{Y_t, \alpha}$ is Hölder continuous with Hölder constant $C_1 > 0$ and Hölder exponent $q \in (0,1]$, i.e.

$$
|q_{Y_{t_1}, \alpha} - q_{Y_{t_2}, \alpha}| \leq C_1 \cdot |t_1 - t_2|^q.
$$

Here (A4) is a slightly stronger version of (A2), and (A5) is an additional smoothness assumption on the quantiles.

Concerning the parameters of our estimate, i.e., the kernel, the bandwidth, the estimate of $m$ and the original quantile estimate, we assume (EST1) and, in addition:

(EST3) The estimate $m_n$ of $m$ satisfies (3.3) for some $\beta_n > 0$, where

$$
\beta_n \to 0 \quad \text{for} \quad n \to \infty.
$$

(3.9)

(EST4) The estimate $\hat{q}_{Y_t, \alpha}$ of $q_{Y_t, \alpha}$ satisfies (3.4) for some $\eta_n \in \mathbb{R}_+$, where

$$
\eta_n \to 0 \quad \text{for} \quad n \to \infty.
$$

(3.10)
(EST5)

\[ h_{n,1} \rightarrow 0 \quad \text{for} \quad n \rightarrow \infty, \quad (3.11) \]

and

\[ \frac{n \cdot h_{n,1}}{\log(n)} \rightarrow \infty \quad \text{for} \quad n \rightarrow \infty \quad (3.12) \]

(EST6) For \( r = \min\{p, q\} \) we have

\[ \frac{h_{n,1}^r}{\beta_n + \eta_n} \rightarrow 0 \quad \text{for} \quad n \rightarrow \infty. \quad (3.13) \]

Here (EST3) and (EST4) mean that the error of the estimates of \( m \) and of our initial quantile estimate must vanish asymptotically, and (EST6) requires that the bandwidth \( h_{n,1} \) is not too small in comparison with these errors. Basically this requires rates of convergence results for the estimates of \( m \) and for our initial quantile estimate. For the latter estimate such a result is given by Theorem 1 and for the estimate of \( m \) such a result basically requires a smoothness assumption on \( m \) (see, e.g., Györfi et al. (2002) for related rate of convergence results for various estimates).

**Theorem 2.** Assume that \( (X_t)_{t \in [0,1]} \) is an \( \mathbb{R}^d \)-valued stochastic process such that \( X_t \) has a density \( f(t, \cdot) : \mathbb{R}^d \rightarrow \mathbb{R} \) with respect to the Lebesgue-Borel measure. Let \( m : [0,1] \times \mathbb{R}^d \rightarrow \mathbb{R} \) be a measurable function and assume that \( Y_t \) is given by \( Y_t = m(t, X_t) \). Let \( \alpha \in (0,1) \) and let \( q_{Y_t,\alpha} \) be the \( \alpha \)-quantile of
for $t \in [0,1]$, and assume that (A3), (A4) and (A5) hold. Let $n \in \mathbb{N}$ and set $t_k = k/n$ ($k = 1, \ldots, n$). Assume that the kernel $K$ satisfies (EST1). Let

the estimator $\hat{q}^{(IS)}_{Yt, \alpha}$ be defined by (3.6) and (3.7) with $h_{n,1} > 0$, and assume that (EST3), (EST4), (EST5) and (EST6) hold. Furthermore, assume that (3.5) is satisfied. Then there exists a constant $c_{10} > 0$ such that

$$
P \left( \sup_{t \in [0,1]} |\hat{q}^{(IS)}_{Yt, \alpha} - q_{Yt, \alpha}| > c_{10} \cdot \left( (\beta_n + \eta_n)^2 \frac{\log(n)}{n \cdot h_{n,1}} + h_{n,1}^p \right) \right) \rightarrow 0 \quad \text{for} \quad n \rightarrow \infty.$$

In particular, if we set $h_{n,1} = c_{11} \cdot (\beta_n + \eta_n)^2/(2p+1) \cdot (\log(n)/n)^{1/(2p+1)}$ for some constant $c_{11} > 0$, there exists a constant $c_{12} > 0$ such that

$$
P \left( \sup_{t \in [0,1]} |\hat{q}^{(IS)}_{Yt, \alpha} - q_{Yt, \alpha}| > c_{12} \cdot (\beta_n + \eta_n)^{2p/(2p+1)} \cdot \left( \frac{\log(n)}{n} \right)^{2p/(2p+1)} \right) \rightarrow 0 \quad \text{for} \quad n \rightarrow \infty.$$

Remark 4. If $\eta_n = 2 \cdot c_6 \cdot (\log(n)/n)^{(p/2p+1)}$ as in Theorem 1 and $\beta_n \leq \eta_n$ is satisfied, we get $(\log(n)/n)^{(4p+1)/(2p+1)}$ as rate of convergence in Theorem 2. Hence in this case the rate of convergence in Theorem 2 is better than the one in Theorem 1 but on the other hand Theorem 2 requires stronger conditions than Theorem 1. However, the basic smoothness assumption on the cumulative distribution function (in particular the value of $p$) is the same in both theorems.

Remark 5. Observations of $Z_k^{(t_k)}$ for $k = 1, \ldots, n$ can be generated by a rejection method. For this purpose one uses several observations of $(t_k, X_k^{(t_k)})$ for each $k = 1, \ldots, n$, and one selects the first observation that satisfies
either the condition

\[ X_k^{(t_k)} \in K_n \text{ and } (|m_n(t_k, X_k^{(t_k)}) - \hat{q}_{Y_k, \alpha}| \leq 3\beta_n + 3\eta_n) \text{ or } (X_k^{(t_k)} \notin K_n). \]

**Remark 6.** Since the smoothness of the system \( m \) is unknown in practice, the approximation error \( \beta_n \) of the surrogate model \( m_n \) and the estimation error \( \eta_n \) of the initial quantile estimate are unknown. A data-driven method to select \( \beta_n \) and \( \eta_n \) will be presented in Section 4.

**Remark 7.** As for the first time-dependent estimator a bandwidth \( h_{n,1} \) has to be selected in a data-driven way for any application of the importance sampling quantile estimator. To do this, we suggest to proceed as in Remark 2, but to use the importance sampling random variables. More precisely, assume that for each of the equidistant time points \( t_k \) \((k = 1, \ldots, n)\) a random variable \( Z_{k,2}^{(t_k)} \), such that \( Z_{k,1}^{(t_k)} \) and \( Z_{k,2}^{(t_k)} \) are independent and identically distributed, as well as observations \( m(t_k, Z_{k,2}^{(t_k)}) \) for \( k = 1, \ldots, n \) are available. Analogously to Remark 2 the bandwidth \( h_{n,1} \) can be selected from a set of possible bandwidths \( H_{n,1} \) by minimizing

\[ \Delta_{h_{n,1}} = \frac{1}{n} \sum_{k=1}^{n} \left| 1_{\{m(t_k, Z_{k,2}^{(t_k)}) \leq y\}} - G_{Y_k}^{(IS)}(y) \right|^2 \]

over all \( h_{n,1} \in H_{n,1} \), where \( y \) is chosen as the \( \alpha \)-quantile of the empirical cdf. corresponding to the data \( m_n(t_1, Z_{1,1}^{(t_1)}), \ldots, m_n(t_n, Z_{n,1}^{(t_n)}) \).

**Remark 8.** In Section 4 we will use Monte-Carlo simulation and addi-
tional data \((t_k, X^{(t_k)}_{k,3}), \ldots, (t_k, X^{(t_k)}_{k,N+2})\) for \(k = 1, \ldots, n\) and some \(N \in \mathbb{N}\) sufficiently large, e.g. \(N = 10,000\), in order to approximate the integrals in \(c_{t_k}\) and \(b_{t_k}\) for \(k = 1, \ldots, n\) by

\[
\hat{c}_{t_k} = \frac{1}{N} \sum_{i=3}^{N+2} \left( \mathbb{1}_{\{u \in K_n: \hat{q}_{Y_{t_k},\alpha - 3\beta_n - 3\eta_n} \leq m_n(t_k,u) \leq \hat{q}_{Y_{t_k},\alpha + 3\beta_n + 3\eta_n}\}} \left( X^{(t_k)}_{k,i} \right) + \mathbb{1}_{\{u \notin K_n\}} \left( X^{(t_k)}_{k,i} \right) \right),
\]

\[
\hat{b}_{t_k} = \frac{1}{N} \sum_{i=3}^{N+2} \mathbb{1}_{\{u \in K_n: m_n(t_k,u) < \hat{q}_{Y_{t_k},\alpha - 3\beta_n - 3\eta_n}\}} \left( X^{(t_k)}_{k,i} \right).
\]

4. Illustration of the finite sample size behaviour of the estimates

4.1 Application to simulated data

Next, we examine the finite sample size behavior of the local average based time–dependent quantile estimator \(\hat{q}_{Y_{t_k},\alpha}\) defined in (1.4) and the importance sampling time–dependent quantile estimator \(\hat{q}^{(IS)}_{Y_{t_k},\alpha}\) defined in (3.7) by applying them to simulated data. Both estimators use the same number \(3n\) of evaluations of \(m\), which we achieve by using for the local average based quantile estimator \(\hat{q}_{Y_{t_k},\alpha}\) three independent copies of \(Y_{t_k}\) for each time point \(t_k = k/n\) \((k = 1, \ldots, n)\). Here

\[
\tilde{D}_{n,1} = \{Y^{(t_1)}_{1,1}, Y^{(t_2)}_{1,2}, \ldots, Y^{(t_n)}_{n,1}, Y^{(t_n)}_{n,2}\}
\]

is used for the main quantile estimation and

\[
\tilde{D}_{n,2} = \{Y^{(t_1)}_{1,3}, \ldots, Y^{(t_n)}_{n,3}\}
\]
is used as testing data for the data–driven bandwidth selection method described in Remark 4, where for each \( k = 1, \ldots, n \) we compare \( Y_{k,1}^{(t_k)} \) and \( Y_{k,2}^{(t_k)} \) with \( Y_{k,3}^{(t_k)} \).

For the importance sampling estimator, we also use three evaluations of the function \( m \) at each time point \( t_k = k/n \) \((k = 1, \ldots, n)\) as well as additional copies \( X_{k,3}^{(t_k)}, X_{k,4}^{(t_k)}, \ldots \) of \( X_{tk} \), which are used for the generation of \( Z_{k,1}^{(t_k)} \) and \( Z_{k,2}^{(t_k)} \) for \( k = 1, \ldots, n \) and for the integral approximation by Monte-Carlo simulation in the estimation of \( c_{tk} \) and \( b_{tk} \) (cf. Remark 8) for \( k = 1, \ldots, n \) and \( N = 10,000 \). To generate observations of \( Z_{k,1}^{(t_k)} \) and \( Z_{k,2}^{(t_k)} \) for \( k = 1, \ldots, n \) by applying the rejection method presented in Remark 5, a surrogate model \( m_n \) of \( m \) as well as its approximation error \( \beta_n \) (see (3.3)) and an initial quantile estimation as well as its estimation error \( \eta_n \) (see (3.4)) are required. The local averaged based time–dependent quantile estimator \( \hat{q}_{Y_{tk},\alpha} \) of Section 2 is used for the initial quantile estimation. Here, \( \eta_n \) is unknown in reality, because the Hölder exponent \( p \) of the smoothness condition in Theorem 1 is unknown.

A data set \( D_{n,1} \), as described in (3.1), is used to generate an initial quantile estimation by the local average based time–dependent quantile estimator \( \hat{q}_{Y_{tk},\alpha} \). To determine \( \eta_n \) in a data-driven way, we suggest to use a bootstrap method and the data sets \( D_{n,1} \) and \( D_{n,2} \). For each time point \( t_k \)
$(k = 1, \ldots, n)$, we choose $(t_k, Y_{k,1}^{(t_k)})$ or $(t_k, Y_{k,2}^{(t_k)})$ randomly from $D_{n,1}$ or $D_{n,2}$ as learning or testing data sets. We repeat the procedure 30 times to obtain multiple learning and testing data sets and to estimate $q_{Y_{tk}^{\alpha}}$ by $\hat{q}_{Y_{tk}^{\alpha}}$ for $k = 1, \ldots, n$ multiple times. For each time point $t_k$ ($k = 1, \ldots, n$), we estimate the interquartile range and choose $\eta_n$ as the median of the interquartile ranges over all time points.

Next, a surrogate model $m_n$ of $m$ can be estimated by a smoothing spline estimator, such as the here applied routine $Tps()$ in the statistic package $R$, on the data set $D_{n,2}$. To estimate $\beta_n$ in a data-driven way, we suggest a cross-validation method. First, we split $D_{n,2}$ in five parts. Then for $j = 1, \ldots, 5$ we approximate $m_n^{(j)}$ of $m$ using the data $D_{n,2}$ without the $j$-th part and use the $j$-th part as testing data to compute the absolute error of $m_n^{(j)}$ for each time point $t_k$ ($k = 1, \ldots, n$). Finally, we determine the maximal absolute error of $m_n^{(j)}$ for each time point, and choose $\beta_n$ as the mean of these maximal errors.

Now, $Z_{1,1}^{(t_1)}, \ldots, Z_{n,1}^{(t_n)}$ and $Z_{1,2}^{(t_1)}, \ldots, Z_{n,2}^{(t_n)}$ can be generated according to Remark 5 for some $K_n$, where we suggest to use $K_n = [-\hat{c} \cdot \log(n), \hat{c} \cdot \log(n)]$ for some constant $\hat{c} > 0$ (c.f., Table 1).

We compare the two time–dependent quantile estimators on three different models. In all three models we consider first $n_1 = 50$, then $n_2 = 100$
and finally \( n_3 = 200 \) equidistant time-points in the time interval \([0, 1]\), i.e. overall 150, 300 or 600 evaluations of the function \( m \), and estimate the time-dependent 0.95-quantiles. Since it is not possible to compare the error in the supremum norm (1.2), we will compare the maximal absolute errors

\[
\max_{t \in \{t_1, \ldots, t_n\}} |\hat{q}_{Y_{t,\alpha}} - q_{Y_{t,\alpha}}| \to \max_{t \in \{t_1, \ldots, t_n\}} |\hat{q}_{Y_{t,\alpha}}^{(IS)} - q_{Y_{t,\alpha}}|.
\]

(4.1)

We repeat the estimation 100 times and compare the mean of these errors.

In our first model \( X_t \) is \( \mathcal{N}(0, (1/2 \cdot t - t^2 + 1/2)^2) \) distributed and

\[
m(t, x) = t \cdot \exp(x) \quad (t \in [0, 1], \ x \in \mathbb{R}).
\]

For the second model \( X_t \) is \( \mathcal{N}(0, (t^2 - t^4 + 1/2)^2) \) distributed and \( m \) is characterized by

\[
m(t, x) = \sqrt{t + x^2} \quad (t \in [0, 1], \ x \in \mathbb{R}).
\]

In our last model \( X_t \) is \( \mathcal{N}(0, (3/2 \cdot t^4 - 3/2 \cdot t^2 + 1)^2) \) distributed and \( m \) is given by

\[
m(t, x) = \begin{cases} 
0, & \text{for } x \leq 0 \\
\sin(x), & \text{for } 0 < x < \frac{\pi}{2} \\
1, & \text{for } x \geq \frac{\pi}{2}
\end{cases} \quad (t \in [0, 1], \ x \in \mathbb{R}).
\]

The results for both estimators are presented in Table 1. Moreover, Table 1
Table 1: Results for the 0.95-quantile estimation by $\hat{q}_{Y_t,\alpha}$ and $\hat{q}^{(IS)}_{Y_t,\alpha}$. Reported are the maximal absolute errors (and in brackets the relative errors).
shows the set of possible bandwidths $H$ for both estimators and the chosen constant $\hat{c}$ in the interval $K_n$. As expected the results for the importance sampling time–dependent quantile estimation $\hat{q}^{(IS)}_{Y_t,\alpha}$ are better than the results for the local average based quantile estimation $\hat{q}^{\text{LA}}_{Y_t,\alpha}$ as the sample size increases in the investigated models. Moreover, it can be seen that for both estimators the estimation becomes more accurate for higher sample sizes. If we compare the errors within the three different models it can be seen that the relative error in Model 1 is larger than the relative errors in the other models. We believe that this is due to the fact that the distributions in Model 1 have much larger tails than in the other models, which makes the estimation of quantiles more difficult.

4.2 Analysis of the effect of an aging spring on the force at the point of impact

Finally we apply our newly proposed estimation methods to the practical problem described in the introduction. As before we use 3 observations of $Y_t$ at $n = 100$ time points, i.e. 300 evaluations of the computer experiment $m$. As in the previous subsection, for the importance sampling estimator a surrogate model $m_n$ of the underlying function $m$ is estimated by a smoothing spline estimator, and the bandwidth $h$ is chosen as in Remark 7. Since the true quantiles are unknown, we only present the 0.95-quantiles
Figure 3: The 0.95-quantile of the force at the point of impact estimated by $\hat{q}_{Y_{t,0.95}}^{(IS)}$. The results are shown in Figure 3. It can be seen that less force acts on the point of impact, when the spring stiffness decreases over time.

**Supplementary Materials**

The supplementary materials contain the proofs of Theorem 1 and Theorem 2.
Acknowledgements

The authors would like to thank the German Research Foundation (DFG) for funding this project within the Collaborative Research Centre 805. Furthermore, the authors would like to thank the Associate Editor and two anonymous referees for very helpful comments on an early version of the manuscript.

References


Fachbereich Maschinenbau, Technische Universität Darmstadt, Otto-Bernd-Str. 2, 64287 Darmstadt, Germany

E-mail: heimrich@dik.tu-darmstadt.de
Fachbereich Mathematik, Technische Universität Darmstadt, Schloßgartenstr. 7,
64289 Darmstadt, Germany.
E-mail: kohler@mathematik.tu-darmstadt.de

Fachbereich Mathematik, Technische Universität Darmstadt, Schloßgartenstr. 7,
64289 Darmstadt, Germany.
E-mail: kristl@mathematik.tu-darmstadt.de