Distribution function estimation by constrained polynomial spline regression

Lan Xue *; Jing Wang b
*a Department of Statistics, Oregon State University, Corvallis, OR, USA *b Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL, USA

First published on: 30 November 2009
A smooth monotone polynomial spline (PS) estimator is proposed for the cumulative distribution function. The proposed method applies a constrained PS regression to smooth the empirical distribution function, while simultaneously ensures monotonicity by imposing a set of linear constraints on the coefficients of the PS functions. This feature is not shared by its kernel counterpart in [Cheng, M.Y., and Peng, L. (2002), 'Regression Modeling for Nonparametric Estimation of Distribution and Quantile Functions', *Statistica Sinica*, 12, 1043–1060], as the kernel estimator is not necessarily monotone. Under mild assumptions, both $L_2$ and uniform convergence rates are obtained. Our simulation studies show that the proposed estimator has better finite sample performance than the simple empirical distribution function. We also illustrate the use of the proposed method by analysing two real data examples.

**Keywords:** empirical process; knot; monotone spline; strong approximation; uniform consistency

**AMS Subject Classification:** 62G08; 62G10; 62G15

1. Introduction

The estimation of the probability density function and cumulative distribution function (CDF) is fundamental in applied statistical data analysis. Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d) random variables with an unknown CDF $F(x)$. The empirical cumulative distribution function (ECDF) $F_n(x) = \sum_{i=1}^{n} I(X_i \leq x)/n$ provides a convenient nonparametric estimator for $F(\cdot)$ with desirable theoretical properties. However, the ECDF is discontinuous as it jumps with size $1/n$ at each data point. It is undesirable since the CDF itself is a continuous function in most cases.

In the literature, the main method to obtain a smooth nonparametric CDF estimator is to integrate a kernel density function. This idea dates back to Nadaraya (1964), Yamato (1973), Azzalini (1981), and Reiss (1981) studied its asymptotics for univariate i.i.d. data. Altman and Léger (1995) proposed a plug-in bandwidth selection method. Liu and Yang (2008) systematically studied the asymptotics of this integrated kernel estimator for multivariate and weakly dependent data. Recently, Cheng and Peng (2002) proposed a novel approach by smoothing the ECDF.
directly. The idea is to treat \( \{F_n(X_i)\}_{i=1}^n \) as responses from a regression model with the value of the \( F_n(\cdot) \) equal to the value of the sum of the true distribution and an error term, and then the local linear regression is applied to the \( n \) pairs \( \{(X_i, F_n(X_i))\}_{i=1}^n \) to obtain a smooth CDF estimator. Cheng and Peng (2002) showed that, for most commonly used kernel functions, their local linear estimator has a smaller asymptotic mean integrated squared error than the traditional integrated kernel distribution estimator. However, Cheng and Peng’s estimator is not necessarily monotone nondecreasing, which is undesirable for estimating distribution functions.

In this paper, we propose to estimate the CDF by applying the constrained polynomial spline (CPS) regression to the \( n \) pairs \( \{(X_i, F_n(X_i))\}_{i=1}^n \). The proposed estimator is not only smooth but also monotone nondecreasing. As an alternative to kernel smoothing, the polynomial spline (PS) smoothing has been successfully used in estimating various non- and semi-parametric models (He and Shi 1998; Huang 1998, 2003; Qu and Li 2006; Stone 1985; Wang 2009; Xue 2009; Zhou 2009; Zhou and He 2008). However, the use of PS in estimating monotone functions, such as the distribution function, is less understood. Ramsay (1988) proposed to use I-splines, which use integrated B-splines as basis and impose non-negative constraints on the coefficients to ensure monotonicity. But as He and Shi (1998) pointed out, the class of I-splines is relatively small compared with the class of all monotone splines, therefore, there is always a possibility that the fit to the data could be improved by allowing more general monotone splines. The important work of He and Shi (1998) proposed a more general approach that used quadratic splines and ensured monotonicity by constraining its derivatives to be non-negative. Since the derivative of a quadratic spline is a linear spline, it essentially imposes a set of linear constraints on the coefficients, which can be efficiently solved by linear programming. However, extension of He and Shi’s idea to higher order of PS is computationally difficult as its derivative is no longer linear. In this paper, we propose a class of linearly CPSs, with a specific set of linear constraints imposed to guarantee monotonicity and can be used with any order of PSs. If the quadratic spline is used, it is also equivalent to He and Shi’s method. We also show that the proposed estimator is both \( L_2 \) and uniformly consistent asymptotically. Although we focus on the estimation of the distribution function, the idea can be generally applied to estimate any monotone function.

The rest of this paper is organised as follows. Section 2 proposes a smooth monotone spline estimator for the CDF and discusses a knot number selection procedure based on Akaike information criterion (AIC). Section 3 presents the asymptotic properties of the proposed estimator. In Section 4, we present some simulation studies and apply the proposed method to analyse the old faithful data and the blood fat concentration data. The proofs are left in the appendices.

2. Spline estimation

One can express the empirical distribution \( F_n(\cdot) \) as the following ‘regression model’

\[
F_n(X_i) = F(X_i) + \varepsilon_i, \quad i = 1, \ldots, n,
\]

where \( \varepsilon_i \)'s are zero mean error terms. This representation motivates us to produce a smooth estimator of the unknown distribution function \( F(\cdot) \) by PS regression of the pseudo observations \( \{(X_i, F_n(X_i))\}_{i=1}^n \).

PSs are piece-wise polynomials connected smoothly over a set of interior knots. The space of the PSs is determined by the order of polynomials and the location of knots. Suppose that we are estimating the distribution function \( F \) on the interval \([0, 1]\). Let \( k_n = \{0 = s_0 < s_1 < \cdots < s_{N_n} < s_{N_n+1} = 1\} \) be a partition of the interval. Then using \( k_n \) as knots, PSs with degree \( p \) forms a linear space with dimension \( J_n = N_n + p + 1 \), denoted as \( G^{(p)}_n = G(p, k_n) \). Let \( s_{-p} = \cdots = s_{-1} = 0 \) and \( s_{N_n+2} = \cdots = s_{N_n+p+1} = 1 \) be auxiliary knots, and \( (x - s)^p_+ = (x - s)^p I(x \geq s) \). For any
function \( g \) on \([0, 1]\), the divided difference of \( g \) on a grid of \( v \) points \( 0 \leq s_1^* \leq s_2^* \leq \cdots \leq s_v^* \leq 1 \) is defined by

\[
[s_i^*]g = g(s_i^*),
\]

\[
[s_1^*, \ldots, s_i^*]g = \frac{[s_2^*, \ldots, s_i^*]g - [s_1^*, \ldots, s_{i-1}^*]g}{s_i^* - s_1^*}
\]

for \( s_v^* \neq s_1^* \),

\[
[s_1^*, \ldots, s_i^*]g = \frac{(d^{v-1}/dv-1)g(s_1^*)}{(v - 1)!}
\]

for \( s_1^* = s_v^* \in (0, 1) \),

\[
[s_1^*, \ldots, s_i^*]g = \frac{(d^{v-1}/dv-1)g(s_1^*)}{(v - 1)!}
\]

for \( s_v^* = s_i^* = 0 \),

\[
[s_1^*, \ldots, s_i^*]g = \frac{(d^{v-1}/dv-1)g(s_1^*)}{(v - 1)!}
\]

for \( s_v^* = s_i^* = 1 \),

where \( d_j^l/dr^l \) and \( d_j^l/rd^r^j \) denote the \( j \)th left and right derivatives of a given function. Then the normalised B-spline basis of \( G_{n}^{(p)} \) is defined as,

\[
B_l(x) = (-1)^{p+1}(s_l - s_{l-p-1})[s_{l-p-1}, \ldots, s_l](x - s)^p, \quad l = 1, \ldots, J_n.
\]

Then one can approximate the unknown distribution function \( F \) by a linear combination of the normalised B-spline basis, i.e. \( F(x) \approx \sum_{l=1}^{J_n} \beta_l B_l(x) \), where the unknown coefficients \( \beta = (\beta_1, \ldots, \beta_{J_n})^T \) can be estimated by minimising the following sum of squares

\[
\tilde{\beta} = \arg \min_{\beta \in \mathbb{R}^{J_n}} \sum_{i=1}^{n} \left( F_n(X_i) - \sum_{l=1}^{J_n} \beta_l B_l(X_i) \right)^2.
\]

As a result, the PS estimator of the distribution function is given as

\[
\tilde{F}(x) = \sum_{l=1}^{J_n} \tilde{\beta}_l B_l(x), \quad (2)
\]

which is continuous as long as the order of polynomial \( p \geq 1 \). However, it is not guaranteed to be monotone nondecreasing, although the ‘data points’ \( \{X_i, F_n(X_i)\}_{i=1}^{n} \) are monotone. According to Theorem 5.9 of Schumaker (1981), a sufficient condition for a PS \( g(x) = \sum_{l=1}^{J_n} \beta_l B_l \in G_{n}^{(p)} \) to be monotone nondecreasing is that its basis coefficients satisfy \( \beta_l \geq \beta_{l-1} \) for \( l = 2, \ldots, J_n \). Therefore, to produce a smooth and monotone distribution estimator, we consider the following CPS estimation. Denote \( S_n = \{ (\beta_1, \ldots, \beta_{J_n}) \in \mathbb{R}^{J_n} | \beta_l \geq \beta_{l-1}, l = 2, \ldots, J_n \} \). Let \( \hat{\beta} \) be the minimiser of the sum of squares constrained on \( S_n \), i.e.

\[
\hat{\beta} = \arg \min_{\beta \in S_n} \sum_{i=1}^{n} \left( F_n(X_i) - \sum_{l=1}^{J_n} \beta_l B_l(X_i) \right)^2. \quad (3)
\]

Then the resulting monotone PS estimation of the distribution is given as

\[
\hat{F}(x) = \sum_{l=1}^{J_n} \hat{\beta}_l B_l(x). \quad (4)
\]

Since the minimisation problem being solved in Equation (3) is convex, solutions to Equation (3) exist and form a closed convex set if not unique. We can take any solution to be our estimate.
When \( p = 2 \), the condition that \( \beta_l \geq \beta_{l-1} \) for \( l = 2, \ldots, J_n \) is necessary and sufficient for a PS \( g(x) = \sum_{l=1}^{J_n} \beta_l B_l \in \mathcal{G}_n^{(p)} \) to be monotone nondecreasing. Therefore, Equation (3) with \( p = 2 \) is also equivalent to He and Shi (1998)'s approach to constrain on the derivative of the quadratic spline functions. However, the formulation of Equation (3) is more general and applicable with any order of PSs. In the following, we denote \( \tilde{F}(\cdot) \) in Equation (2) and \( \hat{F}(\cdot) \) in Equation (4) as PS estimator and CPS estimator, respectively.

### 2.1. Knot number selection

The selection of the knot sequence is critical to the accuracy of the proposed PS and CPS estimators. Following Xue and Yang (2006) and Xue (2009), we propose to use a set of knots equally spaced in the percentile ranks by taking \( s_j = x_{\lfloor n j/(N_n + 1) \rfloor} \), the \( j/(N_n + 1) \)th quantile of the distinct values of \( x_i \) for \( j = 1, \ldots, N_n \). The number of interior knots \( N_n \) can be selected by the AIC criteria. To be specific, we denote a distribution estimator \( \hat{F}(\cdot) \) with number of knots \( N_n \) by \( \hat{F}(\cdot, N_n) \). Then the optimal knot number \( N_n^{opt} \) minimises the AIC values, i.e.

\[
N_n^{opt} = \arg \min_{N_n} \left\{ \log \left( \frac{1}{n} \sum_{i=1}^{n} \{ F_n(X_i) - \hat{F}(X_i, N_n) \}^2 \right) + 2J_n/n \right\}.
\]

One can conduct the knot number selection for the PS and CPS separately. For simplicity, we only conduct the knot selection for the PS method, and use the same set of knots for the CPS. It is justified by the asymptotic property developed in the following section.

### 3. Asymptotic properties

Throughout this paper, we denote by \( L^2[0, 1] \) the space of square integrable real-valued functions on \([0, 1] \). For any \( \phi \in L^2[0, 1] \), define \( \| \phi \|^2 = E\{ \phi^2(X) \} \) and \( \| \phi \|^2_n = (1/n) \sum_{i=1}^{n} \phi^2(X_i) \). In the following sections, we use \( a_n \sim b_n \) to denote that there are constants \( 0 < A < B < \infty \) such that \( A \leq a_n/b_n \leq B \) for all \( n \).

We need the following technical assumptions for our main theorems.

(A1) The distribution function \( F \) is \( p + 1 \) times continuously differentiable for some \( p \geq 1 \).

(A2) Let \( f \) be the first-order derivative of \( F \), also is the density function. Assume that \( f \) is compactly supported. Without loss of generality, let the support be \( \chi = [0, 1] \).

(A3) The density function \( f \) is uniformly bounded below from 0 and above from infinity on \( \chi \).

(A4) The knot sequence \( k_n = \{ 0 = s_0 < s_1 < \cdots < s_{N_n} < s_{N_n+1} = 1 \} \) has a bounded mesh ratio. That is, there exists a constant \( c \) such that \( \max_{j=0, \ldots, N_n} \{ s_{j+1} - s_j \} / \min_{j=0, \ldots, N_n} \{ s_{j+1} - s_j \} \leq c \).

(A5) The number of knots \( N_n \to \infty \), and \( N_n/\sqrt{n} \to 0 \) as \( n \to \infty \).

These assumptions are common in the nonparametric estimation literature. Assumptions similar to (A1)–(A3) are used in Bowman, Hall and Prvan (1998). Assumptions (A4) and (A5) are considered in Huang (1998, 2003) and Xue (2009).

Let \( \tilde{f}(x) \) be the first-order derivative of \( \tilde{F}(x) \), and it provides an estimator for \( f(x) \). In the following, we first provide the rates of both \( L_2 \) and uniform convergence of the unconstrained PS estimators \( \tilde{F}(x) \) and \( \tilde{f}(x) \) in Theorems 3.1 and 3.2, respectively. Then in Theorem 3.3, we show that the rate of convergence of the CPS estimators \( \hat{F}(x) \) is the same as the unconstrained \( \tilde{F}(x) \). Under additional assumption on the smoothing parameter \( N_n \), Theorem 3.4 further shows that
both $\tilde{F}(x)$ and $\hat{F}(x)$ follow the same asymptotic normal distribution as the empirical distribution $F_n(x)$.

**Theorem 3.1** Under regularity conditions (A1)–(A5), one has

$$\| \tilde{F} - F \| = O_p(N_n^{-p} + n^{-1/2})$$

$$\| \tilde{f} - f \| = O_p(N_n^{-p} + n^{-1/2}).$$

**Theorem 3.2** Under regularity conditions (A1)–(A5), one has

$$\sup_{x \in X} | \tilde{F}(x) - F(x) | = O_p \left( N_n^{-p} + \left( \frac{\log(n)}{n} \right)^{1/2} \right),$$

$$\sup_{x \in X} | \tilde{f}(x) - f(x) | = O_p \left( N_n^{-p} + \left( \frac{\log(n)N_n^2}{n} \right)^{1/2} \right).$$

**Theorem 3.3** Under regularity conditions (A1)–(A5), one has

$$\| \hat{F} - F \| = O_p(N_n^{-p} + n^{-1/2}).$$

$$\sup_{x \in X} | \hat{F}(x) - F(x) | = O_p \left( N_n^{-p} + \left( \frac{\log(n)}{n} \right)^{1/2} \right).$$

**Theorem 3.4** Under regularity conditions (A1)–(A5) and if $n^{1/2}N_n^{-(p+1)} \to 0$, one has, for any $x \in (0, 1)$,

$$\sqrt{n}(\tilde{F}(x) - F(x)) \to_d N(0, \sigma^2(x)),$$

$$\sqrt{n}(\hat{F}(x) - F(x)) \to_d N(0, \sigma^2(x)),$$

where $\sigma^2(x) = F(x)(1 - F(x))$.

One referee pointed out that kernel smoothers for CDF improves the second order only and its mean-squared error expansion shows how the bandwidth affects, like those cases in Azzalini (1981) and Reiss (1981). We agree that the choice of the size of the smoothing window can greatly affect the bias and variance of the smoothed values for all smoothers. The authors suspect that under regular conditions, the asymptotic bias and variance of the constrained polynomial estimator can be derived explicitly as a function of knot number and sample size. Those important issues deserve further investigation, but are beyond the scope of this paper.

### 4. Examples

In this section, we conduct simulation studies to evaluate the finite sample performance of the proposed methods. We also illustrate the proposed method by analysing the old faithful data and the blood fat concentration data.

#### 4.1. Simulation study

The data are generated from five mixture normal distributions selected from Marron and Wand (1992). They are standard normal distribution, strongly skewed distribution #3, outlier
distribution #5, separated bimodal distribution #7, and smooth comb #14. The distribution functions are given in Table 1. These distributions cover a broad range of shapes and characteristics.

From each of the five models, 1000 samples of sizes $n = 20, 50$ and 200 are generated respectively. We consider four estimators of the distribution function: linear spline (PS1), cubic spline (PS3), constrained linear spline (CPS1), and constrained cubic spline (CPS3). Unlike the empirical distribution function, all four spline distribution estimators are smooth. Furthermore, the two constrained splines (CPS1 and CPS3) are also monotone nondecreasing. We compare the estimation accuracy of the four spline estimators with the empirical estimators in terms of their averaged squared errors (ASE). For a given $\hat{F}$, the ASE is defined as $\text{ASE}(\hat{F}) = \frac{1}{n} \sum_{i=1}^{n} [\hat{F}(X_i) - F(X_i)]^2$, which is a discrete approximation of $\| \hat{F} - F \|^2$.

Table 2 reports the estimation results. It clearly shows that the ASE decreases as the sample size $n$ increases, supporting asymptotic results in Section 3. Furthermore, the proposed four spline estimators always have smaller ASEs than the empirical distribution estimator for all models and sample size combinations. Overall, the two spline estimators (PS1 and PS3) have smaller ASEs.

**Table 1. Distribution functions used in the simulation study.**

<table>
<thead>
<tr>
<th>Name</th>
<th>Distribution function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard normal distribution</td>
<td>$N(0, 1)$</td>
</tr>
<tr>
<td>Strongly skewed distribution #3</td>
<td>$\sum_{i=0}^{7} \frac{1}{8} N \left( \left( \frac{2}{3} \right)^{i} - 1, \left( \frac{2}{3} \right)^{2i} \right)$</td>
</tr>
<tr>
<td>Outlier distribution #5</td>
<td>$\frac{1}{10} N(0, 1) + \frac{9}{10} N\left(0, \left(\frac{1}{10}\right)^2\right)$</td>
</tr>
<tr>
<td>Separated bimodal distribution #7</td>
<td>$\frac{1}{2} N\left(-\frac{3}{2}, \left(\frac{1}{2}\right)^2\right) + \frac{1}{2} N\left(\frac{3}{2}, \left(\frac{1}{2}\right)^2\right)$</td>
</tr>
<tr>
<td>Smooth comb #14</td>
<td>$\sum_{i=0}^{5} \left(\frac{25-i}{63}\right) N\left(\frac{65-96/2^{i}}{21}, \left(\frac{32/63}{2^{i}}\right)^2\right)$</td>
</tr>
</tbody>
</table>

**Table 2. Simulation results: ASE of five distribution estimators.**

<table>
<thead>
<tr>
<th>Model</th>
<th>$n$</th>
<th>PS1 ($\times 10^3$)</th>
<th>PS3 ($\times 10^3$)</th>
<th>CPS1 ($\times 10^3$)</th>
<th>CPS3 ($\times 10^3$)</th>
<th>$F_n$ ($\times 10^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guassian</td>
<td>20</td>
<td>6.97</td>
<td>6.86</td>
<td>8.08</td>
<td>7.17</td>
<td>8.53</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>2.87</td>
<td>2.76</td>
<td>3.13</td>
<td>2.97</td>
<td>3.29</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.82</td>
<td>0.75</td>
<td>0.83</td>
<td>0.78</td>
<td>0.84</td>
</tr>
<tr>
<td>Strongly skewed</td>
<td>20</td>
<td>8.07</td>
<td>7.36</td>
<td>8.70</td>
<td>7.89</td>
<td>9.02</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>3.48</td>
<td>2.98</td>
<td>3.33</td>
<td>3.25</td>
<td>3.51</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.84</td>
<td>0.79</td>
<td>0.83</td>
<td>0.82</td>
<td>0.84</td>
</tr>
<tr>
<td>Outlier</td>
<td>20</td>
<td>8.25</td>
<td>8.10</td>
<td>8.47</td>
<td>9.20</td>
<td>8.71</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>3.38</td>
<td>3.33</td>
<td>3.41</td>
<td>3.38</td>
<td>3.46</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.78</td>
<td>0.77</td>
<td>0.79</td>
<td>0.82</td>
<td>0.81</td>
</tr>
<tr>
<td>Separated bimodal</td>
<td>20</td>
<td>8.07</td>
<td>7.86</td>
<td>8.39</td>
<td>8.04</td>
<td>8.56</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>3.19</td>
<td>3.12</td>
<td>3.20</td>
<td>3.14</td>
<td>3.32</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.79</td>
<td>0.77</td>
<td>0.79</td>
<td>0.78</td>
<td>0.81</td>
</tr>
<tr>
<td>Smooth comb</td>
<td>20</td>
<td>8.20</td>
<td>7.80</td>
<td>8.28</td>
<td>7.98</td>
<td>8.56</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>3.23</td>
<td>3.19</td>
<td>3.27</td>
<td>3.25</td>
<td>3.36</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.81</td>
<td>0.79</td>
<td>0.81</td>
<td>0.83</td>
<td>0.82</td>
</tr>
</tbody>
</table>
than the constrained ones (CPS1 and CPS3). It is not surprising since the feasible sets of the CPSs are smaller than the unconstrained ones. When the sample size is large \((n = 200)\), the performance of all the four spline estimators are comparable with the empirical distribution estimators.

4.2. Old faithful data

In this section, we illustrate the proposed method by analysing the old faithful data, which is available in the R package. The old faithful data contain the waiting time between eruptions and the duration of eruption for the old faithful geyser in Yellowstone National Park, Wyoming, USA. There are 272 observations in the data. This analysis focuses on the duration of eruption (duration). We used the proposed constrained quadratic spline (CPS2) to estimate the distribution function of duration. The estimation using constrained spline with other orders gave similar results. We also generated 500 bootstrap samples and obtained 95% pointwise bootstrap confidence intervals of the estimated distribution function. In Figure 1, we plotted the estimated distribution function with its 95% pointwise bootstrap confidence intervals. It clearly shows that the duration data are bimodal. Therefore, we also fitted a parametric normal mixture model to the duration data with distribution function \( F(x) = pN(x; \mu_1, \sigma_1^2) + (1 - p)N(x; \mu_2, \sigma_2^2) \), where the parameters \((\mu_1, \mu_2, \sigma_1, \sigma_2, p)\) are estimated using the maximum likelihood estimation (MLE).

Figure 1 shows that the estimated parametric distribution function is very close to the nonparametric estimate CPS2, also lies within the 95% pointwise bootstrap confidence intervals. Both support the use of a normal mixture model for the duration data.

4.3. Blood fat concentration

In a chest pain study, concentrations of plasma cholesterol and plasma triglycerides (mg/dl) were measured on 371 males with chest pain. Among those patients, 51 patients had no evidence of heart

![Figure 1. Old faithful data: plots of the estimated distribution function using constrained quadratic spline (solid) with its 95% pointwise bootstrap confidence intervals (dot-dash), and the estimated distribution from normal mixture model (dashed), along with the empirical distribution at data points (point).](image-url)
disease; for the remaining 320 there was evidence of narrowing of arteries. Please see Chapter 11 in Everitt (2002) for more details about the data set.

One research interest is on whether the concentrations of cholesterol and triglycerides are significantly different between men with and men without heart disease. To this end, we applied the proposed constrained quadratic spline to estimate the distribution functions of cholesterol and triglycerides concentrations for both groups, with and without heart disease respectively. Figure 2 plots the estimated distribution functions along with their 95% bootstrap confidence intervals. For both cholesterol and triglycerides concentrations, the estimated distribution function for group with heart disease is uniformly smaller than those without heart disease group on majority of the interval under consideration. One also notes that two 95% bootstrap confidence intervals of with and without heart disease groups do not overlap completely. It suggests evidence that men

![Figure 2](image-url)

**Figure 2.** Blood fat data: plots of the estimated distribution functions of the cholesterol and triglycerides concentrations for with (dashed) and without heart disease (solid) groups, and their 95% point-wise bootstrap confidence intervals (dotted). In each plot, the empirical distributions at data points (point) are also plotted.
with heart disease tend to be more likely to have higher levels of plasma cholesterol and plasma triglycerides.

Furthermore, for each of the bootstrap sample used to construct confidence intervals, we also estimated the 25th, 50th and 75th percentiles by inverting the estimated distribution function. Using the estimated percentiles, we then constructed 95% confidence intervals for the 25th, 50th and 75th percentiles which were plotted in Figure 3. It also supports that men with heart disease tend to be more likely to have higher levels of plasma cholesterol and plasma triglycerides.

![Cholesterol and Triglycerides](image-url)

Figure 3. Blood fat data: the 95% bootstrap confidence intervals of the 25th, 50th and 75th percentiles for the cholesterol and triglycerides concentrations. In each plot, dashed and solid lines represent the confidence intervals for with and without heart disease groups, respectively. The circles represent the point estimates of corresponding percentiles.
Acknowledgements

The authors are grateful to Dr. Javier Rojo, an editorial board member, and two referees for their helpful comments and suggestions. The work is supported in part by National Science Foundation grant DMS-0906739.

References


Appendix 1. Notation and auxiliary lemmas

Let $\mathcal{B} = \{B_1, \ldots, B_{J_n}\}$ be the set of normalised B-spline basis for $\mathcal{G}_n^{(p)}$. Let $\pi = \{\pi_1, \ldots, \pi_{J_n}\}$ be scaled B-spline basis with $\pi_j = \sqrt{n/a_n} B_j$. For any $g_1, g_2 \in \mathcal{G}(p, k_n)$, define $(g_1, g_2) = E(\langle g_1(X)g_2(X) \rangle)$ and $(g_1, g_2)_n = \langle g_1, g_2 \rangle_n = (1/n) \sum_{i=1}^n \langle g_1(X_i)g_2(X_i) \rangle$. The induced norms are $\|g\|_2 = E(\|g^2(X)\|)$ and $\|g\|_2^2 = (1/n) \sum_{i=1}^n \|g^2(X_i)\|$, respectively. In what follows, $\| \cdot \|_\infty$ denotes the supremum norm of a function on $[0, 1]$. We first present the following two lemmas regarding the properties of the scaled B-spline basis, which were proved in Xue and Yang (2006).

**Lemma A.1** The scaled spline basis $\{\pi_j\}_{j=1}^{J_n}$ satisfies

(i) $E(\pi_j^2(X)) \sim \frac{1}{n^{3/2}-1}$.

(ii) There exists a constant $C > 0$, such that for any vector $a = (a_1, \ldots, a_{J_n})^T$, as $n \to \infty$, $\| \sum_{j=1}^{J_n} a_j \pi_j \|^2 \geq C \sum_{j=1}^{J_n} a_j^2$. 

Lemma A.2 Under assumptions (A1)–(A5), one has
\[
\sup_{g_1, g_2 \in \mathcal{G}(p, k_n)} \left\{ \frac{\langle g_1, g_2 \rangle_n - \langle g_1, g_2 \rangle}{\|g_1\| \|g_2\|} \right\} = O_p \left( \frac{\log^2(n)}{nh} \right).
\]
In particular, there exist constants 0 < c < 1 < C, such that except on an event whose probability tends to zero, as \( n \to \infty \), \( c\|g\| \leq \|g\|_n \leq C\|g\| \), \( \forall g \in \mathcal{G}(p, k_n) \).

Here we cite Theorem 2.2. of Csörgö, Horváth, and Mason (1986) which gives the strong approximation theory for the uniform empirical process.

Lemma A.3 Let \( U_1, \ldots, U_n \) be \( n \) independent uniform \([0, 1]\) random variables, and \( G_n(x) = \sum_{i=1}^n I(U_i \leq x) \) be the uniform empirical distribution function. Let \( \alpha_n(x) = n^{1/2} \{ G_n(x) - x \} \) for \( 0 \leq x \leq 1 \). Then there exists a sequence of Brownian bridges \( \{ B_n(x) : 0 \leq x \leq 1 \} \) defined on the same probability space as \( \{ \alpha_n(x) : 0 \leq x \leq 1 \} \), such that \( \limsup_{n \to \infty} n^{1/2} \| \alpha_n(x) - B_n(x) \|_\infty / (\log n)^{1/2} (\log \log n)^{1/4} = 2^{-1/4}, \) a.s.

Note that \( \tilde{F}(\cdot) \) defined in Equation (2) can be viewed as the orthogonal projection of \( F_n(\cdot) \) onto \( \mathcal{G}_n^{(p)} \) with respect to the empirical inner product \( \langle \cdot, \cdot \rangle_n \). With a slightly abuse of notation, let \( \varepsilon = \varepsilon(\cdot) \) be a random function which interpolates the values \( \varepsilon_1, \ldots, \varepsilon_n \) at \( X_1, \ldots, X_n \). Together with Equation (1), one has the following error decomposition,
\[
\tilde{F} - F = (\tilde{F}^* - F) + \tilde{\varepsilon},
\]
where \( \tilde{F}^* \) and \( \tilde{\varepsilon} \) are orthogonal projection of \( F \) and \( \varepsilon \) onto \( \mathcal{G}_n^{(p)} \), respectively. Note that the first term of the right-hand side of Equation (A1) stands for the approximation error while the second term stands for estimation error. We will treat those two terms separately.

Lemma A.4 Under assumptions (A1)–(A5), one has \( \| \tilde{F}^* - F \| = Op(N_n^{-(p+1)}) \), and \( \| \tilde{F}^* - F \|_\infty = Op(N_n^{-(p+1)}) \).

Proof For \( F \in C^{p+1}([0, 1]) \), PS approximation theorem (de Boor (2001), p. 145) entails that there exist a \( g \) \( \in \mathcal{G}(p, k_n) \) and a constant \( c > 0 \) such that
\[
\| g - F \|_\infty \leq cN_n^{-(p+1)}.
\]
On the other hand, following Theorem 5.1 in Huang (2003), one has
\[
\| \tilde{F}^* - F \|_\infty \leq \inf_{g \in \mathcal{G}(p, k_n)} \| g - F \|_\infty.
\]
Therefore Equations (A2) and (A3) gives that \( \| \tilde{F}^* - F \|_\infty = Op(N_n^{-(p+1)}) \). The result on \( L_2 \) norm follows directly from the result on supremum norm and assumptions (A1)–(A3).

Lemma A.5 Under assumptions (A1)–(A5), the estimation error term \( \tilde{\varepsilon}(x) \) for any \( x \in [0, 1] \) follows an asymptotically normal distribution \( \sqrt{n}\tilde{\varepsilon}(x) \to_d N(0, F(x)(1 - F(x))) \).

Proof For notation simplicity, we consider the constant spline with \( p = 0 \) where \( \{ \pi_j \}_{j=1}^{J_n} \) are orthogonal. For general order \( p \), the proof follows similarly with a choice of the orthogonal basis, which can be constructed from \( \{ \pi_j \}_{j=1}^{J_n} \) using the Gram–Schmidt process. Note that the error term can be written as
\[
\tilde{\varepsilon}(x) = \sum_{j=1}^{J_n} \langle \pi_j, \pi_j \rangle_n \pi_j \| \pi_j \|^2_{2, n} \pi_j(x) = n^{-1} \sum_{j=1}^{J_n} \pi_j(x) \| \pi_j \|^2_{2, n} \pi_j(X_j) s_i
\]
\[
= n^{-1} \pi_{j(x)}(x) \| \pi_{j(x)} \|^2_{2, n} \pi_{j(x)}(X_j) s_i,
\]
where the location index \( j(x) \) is defined such that \( s_{j(x)}(x) \leq x < s_{j(x)}(x+1) \). Lemma A3 entails that there exist a sequence of Brownian bridge \( B_n(\cdot) \) such that, \( \sup_x |\sqrt{n}[F_n(x) - F(x)] - B_n[F(x)]| = Op(n^{-1/4} \log^{1/2} n (\log \log n)^{1/4}) = Op(a_n) \).
Therefore, one has

\[ \sum_{i=1}^{n} \pi_{j(x)}(X_i)x_i = \sum_{i=1}^{n} \pi_{j(x)}(X_i)(F_n(X_i) - F(X_i)) = n \int \pi_{j(x)}(y)(F_n(y) - F(y)) \, dF_n(y) \]

\[ = n^{1/2} \int \pi_{j(x)}(y)B_n[F(y)] \, dF(y) + n^{1/2} \int \pi_{j(x)}(y)\sigma_p(a_n) \, dF(y) \]

\[ + n^{1/2} \int \pi_{j(x)}(y)[B_n[F(y)] + O(a_n)] \, d[F_n(y) - F(y)] \]

\[ = I + II + III. \]  

(A5)

It is known that Brownian bridge can be constructed as \( B_n(t) = W_n(t) - tW_n(1) \), where \( W_n(t) \) is a Gaussian process with mean 0 and variance \( t \). Write \( F_j(x) = F(s_j(x)) \) and \( F_j(x)+1 = F(s_j(x)+1) \). For simplicity, substitute \( F(y) \) with \( t \), then one has

\[ I = n^{1/2} \int \pi_{j(x)}(y)B_n[F(y)] \, dF(y) = n^{1/2}N_n^{1/2} \int_{F_{j(x)}}^{F_{j(x)+1}} [W_n(t) - tW_n(1)] \, dt, \]

which can be shown to follow Gaussian distribution with mean zero and variance

\[ \sigma^2(x) = nN_n E \left[ \int_{F_{j(x)}}^{F_{j(x)+1}} [W_n(t) - tW_n(1)] \, dt \right]^2 \]

\[ = nN_n E \left[ \int_{F_{j(x)}}^{F_{j(x)+1}} [W_n(t) - tW_n(1)] \, dt \int_{F_{j(x)}}^{F_{j(x)+1}} [W_n(s) - sW_n(1)] \, ds \right] \]

\[ = nN_n \int_{F_{j(x)}}^{F_{j(x)+1}} \int_{F_{j(x)}}^{F_{j(x)+1}} E[(W_n(t) - tW_n(1))(W_n(s) - sW_n(1))] \, dt \, ds \]

\[ = nN_n \int_{F_{j(x)}}^{F_{j(x)+1}} \int_{F_{j(x)}}^{F_{j(x)+1}} (\min(t, s) - ts) \, dt \, ds \]

\[ = nN_n^{1/2} F_j(x)(1 - F(x)) = O(nN_n^{1/2}). \]  

(A6)

For term II in Equation (A5), one has

\[ n^{1/2} \int \pi_{j(x)}(y) \cdot O(a_n) \, dF(y) = n^{1/2}N_n^{1/2} \int_{F_{j(x)}}^{F_{j(x)+1}} O_p(a_n) \, dt = O(a_n^{1/2}N_n^{-1/2}). \]  

(A7)

For term III in Equation (A5), applying Lemma A3 and following similar arguments as before, term III is dominated by

\[ \int \pi_{j(x)}(y)B_n[F(y)] \, dF_n[F(y)] = N_n^{1/2} \int_{F_{j(x)}}^{F_{j(x)+1}} B_n(t) \, dB_n(t) \]

\[ = \frac{N_n^{1/2}}{2} [B_n^2(F_{j(x)+1}) - B_n^2(F_{j(x)}) - (F_{j(x)+1} - F_{j(x)})] = V. \]

One can show that

\[ E(V) = \frac{\sqrt{N_n}}{2} [E(B_n^2(F_{j(x)+1}) - B_n^2(F_{j(x)})) - (F_{j(x)+1} - F_{j(x)})] \]

\[ = \frac{\sqrt{N_n}}{2} [F_{j(x)+1}(1 - (F_{j(x)+1}) - F_{j(x)}(1 - (F_{j(x)})) - (F_{j(x)+1} - F_{j(x)})] \]

\[ = -\frac{\sqrt{N_n}}{2}(F_{j(x)+1} - F_{j(x)})(F_{j(x)+1} + F_{j(x)}) = O(N_n^{-1/2}) \]
Appendix 2. Proof of Theorem 3.1

From the error decomposition in Equation (A1), one has that
\[ \| \tilde{F} - F \| \leq \| \tilde{F}^* - F \| + \| \tilde{\varepsilon} \|. \tag{A9} \]
where \( \| \tilde{F}^* - F \| = O_p(N_n^{-\rho+1}) \) as shown in Lemma A4. In the following, we show that \( \| \tilde{\varepsilon} \| = O_p(1/\sqrt{n}) \). Note that one can write \( \tilde{\varepsilon} = \sum_{j=1}^{J_n} a_j \pi_j \), for a set of coefficient \( a = (a_1, \ldots, a_{J_n})^T \). Let \( N = (N(\pi_1), \ldots, N(\pi_{J_n}))^T \) with \( N(\pi_j) = (\pi_j, \varepsilon_n) \).

By the projection theory, one has \( \langle (\pi_j, \pi_j')a_j \rangle_{\pi_j'=\pi_j} = \mathbf{a}^T \mathbf{N} \). Hence \( \mathbf{a}^T (\langle \pi_j, \pi_j' \rangle a_{j'})_{\pi_j'=\pi_j} \mathbf{a} = \mathbf{a}^T \mathbf{N} \), where the LHS is \( \| \sum_{j=1}^{J_n} a_j \pi_j \|^2 \geq c(1 - Q_n) \sum_{j=1}^{J_n} a_j^2 \), while the RHS is \( \mathbf{a}^T \mathbf{N} \leq (\sum_{j=1}^{J_n} a_j^2)^{1/2} (\sum_{j=1}^{J_n} (\pi_j, \varepsilon_n)^2)^{1/2} \). Therefore
\[ \| \tilde{\varepsilon} \|^2 \leq c \left( \sum_{j=1}^{J_n} a_j^2 \right)^{1/2} \leq c \left( \sum_{j=1}^{J_n} (\pi_j, \varepsilon_n)^2 \right)^{1/2}. \tag{A10} \]

Note that \( \sum_{j=1}^{J_n} (\pi_j, \varepsilon_n)^2 = \sum_{j=1}^{J_n} \left( \int \pi_j(x)[F_n(x) - F(x)] \, dF_n(x) \right)^2 = I + II + III \), where
\[ I = \sum_{j=1}^{J_n} \left( \int \pi_j(x)[F_n(x) - F(x)] \, dF(x) \right)^2, \]
\[ II = \sum_{j=1}^{J_n} \left( \int \pi_j(x)[F_n(x) - F(x)] \, d(F_n(x) - F(x)) \right)^2, \]
\[ III = 2 \sum_{j=1}^{J_n} \left[ \int \pi_j(x)[F_n(x) - F(x)] \, dF(x) \right] \times \left[ \int \pi_j(x)[F_n(x) - F(x)] \, d(F_n(x) - F(x)) \right]. \]

One can show that \( I \) is the dominated term. Applying Lemma A3, one has
\[ I = \sum_{j=1}^{J_n} \left( \int \pi_j(x)[F_n(x) - F(x)] \, dF(x) \right)^2 \]
\[ = \frac{1}{n} \sum_{j=1}^{J_n} \left( \int \pi_j(x)B_n(F(x)) \, dF(x) \right)^2 + \frac{1}{n} \sum_{j=1}^{J_n} \left( \int \pi_j(x)[\alpha_n(F(x)) - B_n(F(x))] \, dF(x) \right)^2 \]
\[ + \frac{2}{n} \sum_{j=1}^{J_n} \left( \int \pi_j(x)B_n(F(x)) \, dF(x) \right) \left( \int \pi_j(x)[\alpha_n(F(x)) - B_n(F(x))] \, dF(x) \right) \]
\[ = I_1 + I_2 + I_3. \]
For the first term above, note that its expectation

\[
E(I_1) = \frac{1}{n} \sum_{j=1}^{J_n} \int \pi_j(x) F_n(x) \, dF(x) \int \pi_j(y) F_n(y) \, dF(y)
\]

\[
= \frac{1}{n} \sum_{j=1}^{J_n} \int \pi_j(x) \pi_j(y) E(F_n(x) F_n(y)) \, dF(x) \, dF(y)
\]

\[
\leq \frac{1}{n} \sum_{j=1}^{J_n} E^2(\pi_j(X)) = \mathcal{O}_p\left(\frac{1}{n}\right).
\]

In addition, one has \(I_2 \leq \left(\sup_{0 \leq t \leq 1} |a_n(u) - B_n(u)|^2\right)^{(1/n)} \sum_{j=1}^{J_n} E^2(\pi_j(X)) = \mathcal{O}_p(n^{-3/2} \log^2(n))\) and \(I_3 = \mathcal{O}_p(n^{-5/4} \log(n) N_n^{-1/2})\). Therefore \(I = \mathcal{O}_p(1/n)\). Hence \(\|\tilde{e}\| = \mathcal{O}_p(1/\sqrt{n})\). Together with Equation (A9) and Lemma A4, one has \(\|\tilde{F} - F\| = \mathcal{O}_p(N_n^{-1/2} + n^{-1/2})\). The rate of convergence of \(\tilde{f}\) is obtained by applying Lemma 8 of Stone (1985).

**Appendix 3. Proof of Theorem 3.2**

The order of approximation error follows from Lemma A4. To obtain the supremum order of the estimation error term, we use the following discretisation idea. Divide the interval \([0, 1]\) into \(M_n\) equally spaced intervals with disjoint endpoints \(0 = x_0 < x_1 < \cdots < x_{M_n} = 1\). One has the decomposition as follows:

\[
\sup_{x \in [0, 1]} |\hat{\varepsilon}(x)| = \sup_{0 \leq k \leq M_n} |\hat{\varepsilon}(x_k)| + \sup_{1 \leq k \leq M_n} \sup_{x \in [x_{k-1}, x_k]} |\hat{\varepsilon}(x) - \hat{\varepsilon}(x_k)|. \tag{A11}
\]

Let the standardised error be \(\tilde{\varepsilon}^*(x) = \hat{\varepsilon}(x) \cdot (\text{var}(\tilde{\varepsilon}(x)))^{-1/2}\). From Lemma A5, the distribution of \(\tilde{\varepsilon}^*(x)\) is standard normal asymptotically.

Based on the tail property of the normal distribution, there exists a \(c' > 0\), such that \(1 - \Phi(x) \leq c' \phi(x)\) for large \(x\), where \(\Phi(x)\) and \(\phi(x)\) are CDF and density function of a standard normal random variable. Take \(t_n = \sqrt{C_1 \log n}\) for a large enough \(C_1\), then there exists a constant \(c\) such that

\[
\sum_{n=1}^{\infty} P \left( \sup_{0 \leq k \leq M_n} |\tilde{\varepsilon}^*(x_k)| \geq t_n \right) = \sum_{n=1}^{\infty} P \left( \sup_{0 \leq k \leq M_n} |Z_k| \geq t_n \right)
\]

\[
\leq \sum_{n=1}^{\infty} M_n \cdot P(|Z_0| \geq t_n) \leq c \sum_{n=1}^{\infty} M_n \cdot \exp\{-t_n^2/2\} < \infty,
\]

where \(Z_0, \ldots, Z_{M_n} \sim N(0, 1)\). Consequently for a large value \(\delta > 0\), we have

\[
\sum_{n=1}^{\infty} P \left( \sup_{0 \leq k \leq M_n} |\tilde{\varepsilon}^*(x_k)| \geq \delta \sqrt{\log n} \right) < \infty.
\]

Therefore, the Borel–Cantelli lemma implies that \(\sup_{0 \leq k \leq M_n} |\varepsilon^*(x_k)| = O_p(\sqrt{\log n})\). Note that for any \(x \in [0, 1]\), \(\text{var}(\tilde{\varepsilon}(x)) = E(\tilde{\varepsilon}^2(x))\). Theorem 3.1 entails that \(\sup_{x \in [0, 1]} E(\tilde{\varepsilon}^2(x)) = O_p(n^{-1})\). It immediately implies that \(\sup_{0 \leq k \leq M_n} \text{var}(\tilde{\varepsilon}(x_k)) = O_p(n^{-1})\). Therefore, the first term at RHS of Equation (A11)

\[
\sup_{0 \leq k \leq M_n} |\tilde{\varepsilon}(x_k)| \leq \sup_{0 \leq k \leq M_n} |\varepsilon^*(x_k)| \cdot \sup_{0 \leq k \leq M_n} \sqrt{\text{var}(\tilde{\varepsilon}(x_k))} = O_p\left(\sqrt{\frac{\log(n)}{n}}\right).
\]
In the proof of Theorem 3.1, one expresses the error as \( \tilde{\varepsilon}(x) = \sum_{j=1}^{J_n} a_j \pi_j(x) \). The latter term in Equation (A11) can be explicitly expressed as

\[
\sup_{1 \leq k \leq M_n} \sup_{x \in [x_{k-1}, x_k]} |\tilde{\varepsilon}(x) - \tilde{\varepsilon}(x_k)| = \sup_{1 \leq k \leq M_n} \sup_{x \in [x_{k-1}, x_k]} \left| \sum_{j=1}^{J_n} a_j \pi_j(x) - \sum_{j=1}^{J_n} a_j \pi_j(x_k) \right|
\]

\[
\leq \sup_{1 \leq k \leq M_n} \sup_{x \in [x_{k-1}, x_k]} \max_{j=1,...,N_n} |\pi_j(x) - \pi_j(x_k)| \cdot \sum_{j=1}^{J_n} a_j
\]

\[
\leq \left[ C \sqrt{N_n} \cdot M^{-1} \right] \cdot \frac{J^{1/2}}{\sqrt{\sum_{j=1}^{J_n} a_j^2}}.
\]

The last step is based on special polynomial structure of the B-spline basis and the Hölder inequality. Note that \( \sum_{j=1}^{J_n} a_j^2 = O_p(n^{-1}) \) as shown in the proof of Theorem 3.1. Taking \( M_n \sim n^2 \), one has \( \sup_{1 \leq k \leq M_n} \sup_{x \in [x_{k-1}, x_k]} |\tilde{f}(x) - \tilde{f}(x_k)| = o_p(n^{-1}) \). Based on Equation (A11), the first part of the theorem follows, i.e. \( \sup_{x \in [0,1]} |\tilde{\varepsilon}(x)| = O_p(\sqrt{\log(n)/n}) \). The uniform convergence of \( \tilde{f} \) follows by applying a uniform norm version of Lemma 8 of Stone (1985).

Appendix 4. Proof of Theorem 3.3

According to condition (A2), there exists a \( \varepsilon_0 > 0 \) such that \( f(x) \geq \varepsilon_0 \) for all \( x \in [0,1] \). On the other hand, Theorem 3.2 entails that there exists an integer \( n(\varepsilon_0) > 0 \) such that, when \( n \geq n(\varepsilon_0) \), \( \sup_{x \in [0,1]} |\tilde{f}(x) - f(x)| = o_p(n^{-1}) \). Therefore, when \( n \geq n(\varepsilon_0) \), \( \tilde{f}(x) \geq f(x) - \varepsilon_0/2 \geq \varepsilon_0/2 > 0 \) for all \( x \in [0,1] \). It implies that when the sample size is large enough, the unconstrained \( \tilde{F}(x) \) is actually monotone increasing. Therefore, the constrained polynomial estimator \( F(x) \) is identical to the unconstrained \( \tilde{F}(x) \) when \( n \geq n(\varepsilon_0) \). Therefore, \( F(x) \) enjoys the same asymptotic properties as \( \tilde{F}(x) \).

Appendix 5. Proof of Theorem 3.4

Under \( n^{1/2}N^{-1-\alpha} \rightarrow 0 \), the bias term \( N^{-1-\alpha} \) in Theorems 3.1 and 3.3 are of negligible order. Therefore, the asymptotic normality of \( \tilde{F} \) follows directly from Lemma A5, and asymptotic normality of \( F \) follows from a similar argument as the proof of Theorem 3.3.