Statistica Sinica: Supplement

OPTIMAL DESIGNS FOR SERIES ESTIMATION IN NONPARAMETRIC REGRESSION WITH CORRELATED DATA

Fakultät für Statistik, Technische Universität Dortmund

Fakultät für Mathematik, Ruhr-Universität Bochum

Sustainable Energy Laboratory, Cyprus University of Technology

Supplementary Material

S1 Further numerical results

S1.1 Different choices of the truncation parameter

As pointed out by a referee it is of interest to study the impact of the choice of the truncation parameter J on the optimal design. For the sake of brevity we concentrate on processes with exponential covariance kernel with $L = 1$. We now assume that either four, five or six basis functions (i.e. $J = 4, 5,$ or 6) are used in the series estimator. For the truncation parameter $J = 6$ at least $n = 7$ observations are necessary to get a well-defined estimator such that we further assume that the sample size is $n = 7$. The optimal time points minimizing the criterion (4.22) are given by

$$0.00, 0.17, 0.33, 0.50, 0.67, 0.83, 1.00$$

$$0.00, 0.22, 0.37, 0.51, 0.72, 0.87, 1.00$$

$$0.00, 0.22, 0.28, 0.49, 0.61, 0.89, 1.00$$

for the truncation parameter $J = 4$, $J = 5$ and $J = 6$, respectively. These designs differ substantially indicating some sensitivity of the optimal design with respect to the number of basis functions in the series estimator. An interesting problem for future research is the construction of optimal designs addressing the problem of uncertainty in the truncation parameter.

In Table 1 we display the simulated mean integrated squared errors of the estimators $\hat{f}^{(J),n}$ and $\check{f}^{(J),n}$ for $J = 4, 5, 6$, where we consider the optimal design and the design in (5.4). We obtain similar results as in Section 5.1, where three basis functions are used and the sample size is also given by $n = 7$. For each choice of the truncation parameter the new estimator $\hat{f}^{(J),n}$ outperforms $\check{f}^{(J),n}$ in all cases under consideration, and using the optimal time points even improves its performance.

		design ($J = 4$)		design ($J = 5$)		design ($J = 6$)	
f	estimator	optimal	(5.4)	optimal	(5.4)	optimal	(5.4)
(5.5)	$\hat{f}^{(J),n}$	1.65	1.70	1.69	1.83	1.96	2.08
	$\check{f}^{(J),n}$	1.80	1.83	1.79	1.93	2.03	2.14
(5.6)	$\hat{f}^{(J),n}$	1.64	1.77	1.92	1.99	1.74	2.07
	$\check{f}^{(J),n}$	1.80	1.93	2.04	2.10	1.81	2.14

Table 1: *Simulated mean integrated squared error of the estimators* $\hat{f}^{(J),n}$ *and* $\check{f}^{(J),n}$ *defined in (5.1) and (5.2) for different seven-point designs, for different regression functions and for different truncation parameters* J *of the series expansion. The covariance kernel is given by* $\exp(-|s-t|)$.

S2 Technical details

Proof of Theorem 1 We restrict ourselves to the proof of the result in case (A), the other cases can be proved in a similar way. Note that the function Ψ_j is convex on the space of all signed measures and therefore, a signed measure ξ_j^* minimizes Ψ_j if and only if the directional derivative from ξ_j^* in any direction is nonnegative, that is

$$\frac{\partial}{\partial \alpha} \Psi_j((1-\alpha)\xi_j^* + \alpha\eta)\Big|_{\alpha=0} \geq 0,$$

for all signed measures η on the interval $[0,1]$. A straightforward calculation gives

$$\frac{\partial}{\partial\alpha}\Psi_j((1-\alpha)\xi_j^*+\alpha\eta)\Big|_{\alpha=0} = \int_0^1\int_0^1 f(s)f(t)+K(s,t)\big(\xi_j^*(ds)\xi_j^*(dt)-\xi_j^*(dt)\eta(dt)\big) + \theta_j\int_0^1 f(t)\big(\eta(dt)-\xi_j^*(dt)\big). \tag{S.1}$$

Consequently, the signed measure ξ_j^* minimizes Ψ_j if and only the inequality

$$\int_0^1\int_0^1 f(s)f(t)+K(s,t)\big(\xi_j^*(ds)\xi_j(dt)-\xi_j^*(dt)\eta(dt)\big) + \theta_j\int_0^1 f(t)\big(\eta(dt)-\xi_j^*(dt)\big)\geq 0,$$

is satisfied for all signed measures η on the interval $[0,1]$.

In order to check (S.2) for the signed measure ξ_j^* we calculate each term in (S.2) separately, where we use the following representation of the quantities in (3.13) - (3.16)

$$\begin{aligned}
c &= \int_0^1 \frac{1}{v^2(t)}\frac{(\dot f(t)v(t)-f(t)\dot v(t))^2}{\dot u(t)v(t)-u(t)\dot v(t)}dt+\frac{f^2(0)}{u(0)v(0)},\\
P_0 &= \frac{1}{u(0)}\frac{f(0)\dot u(0)-\dot f(0)u(0)}{v(0)\dot u(0)-\dot v(0)u(0)},\\
P_1 &= \frac{1}{v(1)}\frac{\dot f(1)v(1)-f(1)\dot v(1)}{\dot u(1)v(1)-u(1)\dot v(1)},\\
p(t) &= -\frac{1}{v(t)}\frac{d}{dt}\left[\frac{\dot f(t)v(t)-f(t)\dot v(t)}{\dot u(t)v(t)-u(t)\dot v(t)}\right].
\end{aligned}$$

To simplify (S.2) we note that integration by parts yields

$$\begin{aligned}
\int_0^1 f(s)\xi_j^*(ds) &= \frac{\theta_j}{1+c}\Big(\frac{f(0)(\dot{u}(0)f(0)-u(0)\dot{f}(0))}{u(0)(\dot{u}(0)v(0)-u(0)\dot{v}(0))} \\
&\quad + \frac{f(1)(v(1)\dot{f}(1)-\dot{v}(1)f(1))}{v(1)(\dot{u}(1)v(1)-u(1)\dot{v}(1))} \\
&\quad - \int_0^1 \frac{f(s)}{v(s)}\frac{d}{ds}\left[\frac{v(s)\dot{f}(s)-\dot{v}(s)f(s)}{\dot{u}(s)v(s)-u(s)\dot{v}(s)}\right]ds\Big) \\
&= \frac{\theta_j}{1+c}\Big(\frac{f(0)(\dot{u}(0)f(0)-u(0)\dot{f}(0))}{u(0)(\dot{u}(0)v(0)-u(0)\dot{v}(0))} \\
&\quad + \frac{f(1)(v(1)\dot{f}(1)-\dot{v}(1)f(1))}{v(1)(\dot{u}(1)v(1)-u(1)\dot{v}(1))} \\
&\quad - \Big[\frac{f(s)}{v(s)}\frac{v(s)\dot{f}(s)-\dot{v}(s)f(s)}{\dot{u}(s)v(s)-u(s)\dot{v}(s)}\Big]_0^1 \\
&\quad + \int_0^1 \frac{1}{v^2(s)}\frac{\big(\dot{v}(s)f(s)-v(s)\dot{f}(s)\big)^2}{\dot{u}(s)v(s)-u(s)\dot{v}(s)}ds\Big) \\
&= \frac{\theta_j}{1+c}\Big(\frac{f^2(0)}{u(0)v(0)} + \int_0^1 \frac{1}{v^2(s)}\frac{\big(\dot{v}(s)f(s)-v(s)\dot{f}(s)\big)^2}{\dot{u}(s)v(s)-u(s)\dot{v}(s)}ds\Big) \\
&= \frac{\theta_j}{1+c}c\,.
\end{aligned}$$

Similarly, we obtain

$$\begin{aligned}
\int_0^1 K(s,t)\xi_j^*(ds) &= \int_0^t u(s)v(t)\xi_j^*(ds) + \int_t^1 u(t)v(s)\xi_j^*(dt) \\
&= \frac{\theta_j}{1+c}\Big(v(t)\frac{\dot u(0)f(0)-u(0)\dot f(0)}{\dot u(0)v(0)-u(0)\dot v(0)} \\
&\quad -v(t)\int_0^t \frac{u(s)}{v(s)}\frac{d}{ds}\left[\frac{v(s)\dot f(s)-\dot v(s)f(s)}{\dot u(s)v(s)-u(s)\dot v(s)}\right]ds \\
&\quad +u(t)\frac{v(1)\dot f(1)-\dot v(1)f(1)}{\dot u(1)v(1)-u(1)\dot v(1)} \\
&\quad -u(t)\int_t^1 \frac{d}{ds}\left[\frac{v(s)\dot f(s)-\dot v(s)f(s)}{\dot u(s)v(s)-u(s)\dot v(s)}\right]ds\Big) \\
&= \frac{\theta_j}{1+c}\Big(v(t)\frac{f(0)}{v(0)} + f(t) - v(t)\frac{f(0)}{v(0)}\Big) \\
&= \frac{\theta_j}{1+c}f(t),
\end{aligned}$$

where we have used again integration by parts for the third equality. Consequently, we get

$$\int_0^1\int_0^1 f(s)f(t)\xi_j^*(ds)\xi_j^*(dt) = \Big(\int_0^1 f(s)\xi_j^*(ds)\Big)^2 = \frac{\theta_j^2}{(1+c)^2}c^2,$$

$$\int_0^1 K(s,t)\xi_j^*(ds)\xi_j^*(dt) = \frac{\theta_j}{1+c}\int_0^1 f(t)\xi_j^*(dt) = \frac{\theta_j^2}{(1+c)^2}c,$$

and thus the left hand side of (S.2) reduces to

$$\frac{\theta_j^2c^2}{(1+c)^2} + \frac{\theta_j^2 c}{(1+c)^2} - \frac{\theta_j^2 c}{1+c} - \int_0^1 f(t)\eta(dt)\big(\frac{\theta_j}{1+c}c + \frac{\theta_j}{1+c} - \theta_j\big) = 0,$$

for an arbitrary signed measure η. This proves that (S.2) holds and the signed measure ξ_j^* defined in Theorem 1 minimizes the function Ψ_j.

Proof of Proposition 1 For the term on the left hand side of equation (4.11) we obtain

$$\begin{aligned}
\mathbb{E}\big[\|\hat{\theta}^{(J),*} - \hat{\theta}^{(J),n}\|^2\big] &= \mathrm{tr}\Big\{\mathbb{E}\Big[\big(\hat{\theta}^{(J),*} - \hat{\theta}^{(J),n}\big)\big(\hat{\theta}^{(J),*} - \hat{\theta}^{(J),n}\big)^T\Big]\Big\} \\
&= \frac{\|\theta^{(J)}\|^4}{(1+c^{(J)})^2}\mathrm{tr}\Big\{\mathbb{E}\Big[\Big(\sum_{i=2}^{n}\int_{t_{i-1}}^{t_i}\Big(\frac{d}{dt}\left[\frac{\Phi^{(J)}(t)}{v(t)}\right]\big(\frac{d}{dt}q(t)\big)^{-1} - \mu_i\Big)d\Big(\frac{Y_t}{v(t)}\Big)\Big) \\
&\times\Big(\sum_{i=2}^{n}\int_{t_{i-1}}^{t_i}\Big(\frac{d}{dt}\left[\frac{\Phi^{(J)}(t)}{v(t)}\right]\big(\frac{d}{dt}q(t)\big)^{-1} - \mu_i\Big)d\Big(\frac{Y_t}{v(t)}\Big)\Big)^T\Big]\Big\}.
\end{aligned}$$

For the determination of the expected value inside the trace

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sum_{i=2}^{n}\int_{t_{i-1}}^{t_i}\Big(\frac{d}{dt}\frac{\Phi^{(J)}(t)}{v(t)}\big(\frac{d}{dt}q(t)\big)^{-1} - \mu_i\Big)d\Big(\frac{Y_t}{v(t)}\Big)\Big) \\
&\quad\times\Big(\sum_{i=2}^{n}\int_{t_{i-1}}^{t_i}\Big(\frac{d}{dt}\frac{\Phi^{(J)}(t)}{v(t)}\big(\frac{d}{dt}q(t)\big)^{-1} - \mu_i\Big)d\Big(\frac{Y_t}{v(t)}\Big)\Big)^T\Big],
\end{aligned} \tag{S.2}$$

we use a transformation of the Gaussian process $\{Y_t : t \in [0,1]\}$ to a Brownian motion, as it was introduced by Doob (1949). This result shows that the error process $\{\varepsilon_t : t \in [0,1]\}$ with covariance kernel (3.3) can be represented by

$$\varepsilon_t = \varepsilon(t) = v(t)W(q(t)),$$

where $W = \{W(s) : s \in [q(0), q(1)]\}$ is a Brownian motion on the interval $[q(0), q(1)]$. We use this relationship to represent the process $\{Y_t : t \in [0,1]\}$ as

$$Y_t \;=\; f(t) + \varepsilon_t = f(t) + v(t)W(q(t)), \quad t \in [0,1].$$

Dividing by $v(t)$ and using the transformation $s = q(t)$, we get the transformed model

$$Z_s = g(s) + W(s), \quad s \in [q(0), q(1)],$$

where

$$Z_s = \frac{Y_{q^{-1}(s)}}{v(q^{-1}(s))} \quad \text{and} \quad g(s) = \frac{f(q^{-1}(s))}{v(q^{-1}(s))}.$$

Consequently, we obtain for arbitrary $0 \leq t_{i-1} < t_i \leq 1$

$$\begin{aligned}
&\int_{t_{i-1}}^{t_i} \Big(\frac{d}{dt}\left[\frac{\Phi^{(J)}(t)}{v(t)}\right]\big(\frac{d}{dt}q(t)\big)^{-1} - \mu_i\Big) d\Big(\frac{Y_t}{v(t)}\Big) \\
&\qquad = \int_{q(t_{i-1})}^{q(t_i)} \Big(\frac{d}{ds}\left[\frac{\Phi^{(J)}(q^{-1}(s))}{v(q^{-1}(s))}\right] - \mu_i\Big) d\Big(\frac{Y_{q^{-1}(s)}}{v(q^{-1}(s))}\Big) \\
&\qquad \int_{q(t_{i-1})}^{q(t_i)} \Big(\frac{d}{ds}\tilde{\Phi}^{(J)}(s) - \mu_i\Big) dZ_s \\
&\qquad \int_{q(t_{i-1})}^{q(t_i)} \Big(\frac{d}{ds}\tilde{\Phi}^{(J)}(s) - \mu_i\Big)\big(dg(s) + dW(s)\big),
\end{aligned}$$

where the function $\tilde{\Phi}^{(J)}(s)$ is given by $\tilde{\Phi}^{(J)}(s) = \frac{\Phi^{(J)}(q^{-1}(s))}{v(q^{-1}(s))}$ and we set $t = q^{-1}(s)$. This gives for the transformed derivatives

$$\begin{aligned}
\frac{d}{dt}\left[\frac{\Phi^{(J)}(t)}{v(t)}\right] &= \frac{d}{ds}\left[\frac{\Phi^{(J)}(q^{-1}(s))}{v(q^{-1}(s))}\right]\frac{ds}{dt} = \frac{d}{ds}\left[\frac{\Phi^{(J)}(q^{-1}(s))}{v(q^{-1}(s))}\right]\Big(\frac{d}{ds}q^{-1}(s)\Big)^{-1}, \\
\frac{d}{dt}q(t) &= \frac{d}{ds}q(q^{-1}(s))\,\frac{ds}{dt} = \big(\frac{d}{ds}q^{-1}(s)\big)^{-1}.
\end{aligned}$$

We now introduce the notation

$$X_i = \int_{q(t_{i-1})}^{q(t_i)} \Big(\frac{d}{ds}\tilde{\Phi}^{(J)}(s) - \mu_i\Big)\big(dg(s) + dW(s)\big) \quad i = 2, \ldots, n.$$

As W is a Brownian motion, the random variables $X_2, \ldots, X_n$ are independent and the expected value in (S.2) can be rewritten as

$$\mathbb{E}\Big[\sum_{i=2}^{n} X_i \sum_{i=2}^{n} X_i^T\Big] = \sum_{i=2}^{n} \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_i - \mathbb{E}[X_i])^T] + \sum_{i=2}^{n} \mathbb{E}[X_i] \sum_{i=2}^{n} \mathbb{E}[X_i^T]. \tag{S.3}$$

Obviously

$$\begin{aligned}
\mathbb{E}\big[X_i\big] &= \int_{q(t_{i-1})}^{q(t_i)} \Big(\frac{d}{ds}\tilde{\Phi}^{(J)}(s) - \mu_i\Big)\frac{d}{ds}g(s)ds \\
&= \int_{t_{i-1}}^{t_i} \Big(\frac{d}{dt}\Big[\frac{\Phi^{(J)}(t)}{v(t)}\Big]\Big(\frac{d}{dt}q(t)\Big)^{-1} - \mu_i\Big)\Big(\frac{d}{dt}\left[\frac{f(t)}{v(t)}\right]\Big)dt,
\end{aligned}$$

and Itô's isometry gives

$$\begin{aligned}
&\mathbb{E}[(X_i - \mathbb{E}[X_i])(X_i - \mathbb{E}[X_i])^T] \\
= &\int_{q(t_{i-1})}^{q(t_i)} \Big(\frac{d}{ds}\tilde{\Phi}^{(J)}(s) - \mu_i\Big)\Big(\frac{d}{ds}\tilde{\Phi}^{(J)}(s) - \mu_i\Big)^T ds \\
= &\int_{t_{i-1}}^{t_i} \Big(\frac{d}{dt}\Big[\frac{\Phi^{(J)}(t)}{v(t)}\Big]\Big(\frac{d}{dt}q(t)\Big)^{-1} - \mu_i\Big)\Big(\frac{d}{dt}\Big[\frac{\Phi^{(J)}(t)}{v(t)}\Big]\Big(\frac{d}{dt}q(t)\Big)^{-1} - \mu_i\Big)^T \frac{d}{dt}q(t)dt.
\end{aligned}$$

Inserting these representations in (S.3) results in (4.11), which proves Proposition 1.

Bibliography

Doob, J. L. (1949). Heuristic Approach to the Kolmogorov-Smirnov Theorems. *Ann. Math. Statist. 20*(3), 393–403.