AUTOMATIC NONPARAMETRIC SIGNAL FILTRATION[1]

Dobrovidov A.V.
Trapeznikov Institute of Control Sciences,
Russian Academy of Sciences, Moscow, Russia
e-mail:dobrovid@ipu.ru
Koshkin G.M.
Tomsk State University; Department of Informatization Problem,
Tomsk Scientific Center SB RAS, Tomsk, Russia
e-mail:kgm@mail.tsu.ru

Abstract

Recent results in nonparametric bandwidth selection allow us to create data-based algorithms of automatic nonparametric signal filtration. Such algorithms are based on the optimal filtering equation and its nonparametric counterpart from the theory of nonparametric signal processing [1, 2]. This approach was developed for the case when state equation and probability distribution of unobservable useful signal are unknown, but the observation equation and perturbation distribution are known completely. Term "automatic filtration" means that the output data of the observation equation is only used to derive a nonparametric signal filtration equation. The estimation equation contains a term that is a non-parametric estimator of logarithmic derivative of density, which depends on bandwidths for probability and its derivative estimates. Using the results of [3, 4] for bandwidth selection by Smoothed Cross-Validation method, we give an automatic filtration method. To obtain a stable non-parametric estimator of logarithmic density derivative some regularization procedure is used that is named piecewise smooth approximation [5]. Modeling was carried out to compare the behavior of nonparametric estimates with the optimal Kalman ones.

1 Introduction

More than twenty years ago there were developed nonsupervised methods of extraction of useful stochastic signal with unknown distribution from mixture with noise disturbances [1]. In this approach it is assumed that noise distribution is known because often there is an opportunity to observe noise without signal and one can restore the sufficient noise distribution approximation given noise observation. Inverse situation – signal observation without noise – is very sparse (unreal case). So, in this situation the restoration of pure signal distribution is impossible, and signal distribution is assumed to be unknown for this approach.

The methods mentioned above were gathered in the theory of nonparametric signal estimation published in [2] (1997) and in [5](2004) in Russia. The principal result of this

[1]This work was supported by the Russian Foundation for Basic Research, project no. 09-08-00595-a and Program no. 29 of the RAS Presidium.

theory concerning the problem of filtration is the optimal filtration equation of Markov processes, represented in the form explicitly independent of unknown distribution of useful stochastic signal. Such form is possible when noise distribution and observation equation make up a so called *exponential pair*, i.e. the family of conditional observation densities under fixed useful signal constitutes a *conditionally-exponential family* [2]. This family, particularly, contains gaussian density

$$f(x_n|s_n) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_n - s_n)^2}{2\sigma^2}\right\}, \; x_n \in \mathbf{R}, \; s_n \in \mathbf{R}, \tag{1}$$

which as an example will be considered in this paper for simplicity. It follows from expression (1) that observation equation represents an additive model

$$X_n = AS_n + B\eta_n, \tag{2}$$

where A and $B = \sigma$ are known constants, X_n is an observation, S_n is an unobservable signal, and η_n is a noise perturbation at the time n. In this approach an important assumption is made that (X_n, S_n) is Markovian process. This example clearly demonstrates an opportunity of proposal approach.

The problem is to design an optimal in mean square sense estimator $\hat{S}_n$ of a useful signal S_n at moment n from given observations $X_1^n = (X_1, \ldots, X_n)$. As is well-known the optimal estimator of S_n is a conditional mean $\hat{S}_n = \mathsf{E}[S_n|X_1^n = x_1^n]$. This conditional mean can be calculated by method of transformation for posterior probabilities if the state equation in S_n is specified. In the case of linear state equation, for instance,

$$S_{n+1} = aS_n + b\xi_n, \tag{3}$$

where ξ_n is gaussian noise, Kalman filter is optimal, and for its construction it is necessary to know (3) exactly. Such information frequently is not available for users. Are there any ways to circumvent the necessity to know signal state equation? One of this ways is the empirical Bayes approach, by following which one can design an equation for conditional mean $\hat{S}_n$ without information about state equation (3). For this design in our example it is only necessary the information about observation equation of the type (1). Then the equation for optimal estimator $\hat{S}_n$ takes on a form

$$\hat{S}_n = \frac{B^2}{A} \frac{\partial}{\partial x_n} \ln f(x_n|x_1^{n-1}) + \frac{x_n}{A}, \tag{4}$$

where $f(x_n|x_1^{n-1})$ is a conditional density of observation x_n at given previous observations x_1^{n-1}. Unlike Kalman filter equation (4) for $\hat{S}_n$ is not recurrent. The conditional density $f(x_n|x_1^{n-1})$ can not be exactly calculated if the equation (4) is unknown. However it can be restored from observations x_1^n with the demanded degree of precision, using nonparametric kernel method of estimation from dependent data [5]. According to this method we must replace the unknown density $f(x_n|x_1^{n-1})$ by truncated density $\bar{f}(x_n|x_{n-\tau}^{n-1})$, where τ is degree of dependence of observable process (X_n). Per se τ represents an order of connectivity of Markov process approximating the non-Markovian process (X_n). By definition $\bar{f}(x_n|x_{n-\tau}^{n-1}) = f(x_{n-\tau}^{n})/f(x_{n-\tau}^{n-1})$. Then

$$\frac{\partial}{\partial x_n} \ln \bar{f}(x_n|x_{n-\tau}^{n-1}) = \frac{\partial/\partial x_n f(x_{n-\tau}^{n})}{f(x_{n-\tau}^{n})} \triangleq \psi(x_{n-\tau}^{n}). \tag{5}$$

The denominator in the latest formula represents a $(\tau+1)$-dimensional marginal density, the nonparametric kernel estimate for this density can be written as following:

$$\hat{f}(x_{n-\tau}^{n}) = n^{-1} h_n^{-(\tau+1)} \sum_{i=1}^{n-\tau-1} \prod_{j=1}^{\tau+1} K\left(\frac{x_{n-j+1} - x_{n-j-i+1}}{h_n}\right). \tag{6}$$

The nonparametric estimate for the numerator of (5) can be represented as

$$\hat{f}'(x_{n-\tau}^{n}) = n^{-1} h_{1n}^{-(\tau+2)} \sum_{i=1}^{n-\tau-1} K'\left(\frac{x_{n-j-i+1} - x_{n-j+1}}{h_{1n}}\right) \prod_{j=1}^{\tau} K\left(\frac{x_{n-j+1} - x_{n-j-i+1}}{h_{1n}}\right), \tag{7}$$

where f', K' denote the partial derivatives with respect to x_n.

So, the nonparametric estimate for logarithmic density derivative $\psi(x_{n-\tau}^{n})$ can be written as

$$\hat{\psi}_n(x_{n-\tau}^{n}) = \frac{\hat{f}'(x_{n-\tau}^{n})}{\hat{f}(x_{n-\tau}^{n})}. \tag{8}$$

For calculating (8) it remains only to select bandwidths h_n in (6) and h_{1n} in (7).

2 Bandwidth selection for densities and their derivatives

For the time being, several data-based selection methods of the kernel function bandwidth are known of which the methods of cross-validation CV [6, 7], smoothed cross-validation SCV [8], and plug-in [9] seem to be the basic ones as the most clear and rapidly converging procedures. In [4] the method SCV developed in [3] for density estimation was extended to the kernel estimates of the density derivatives. Both of this methods generate data-based bandwidth estimates with a higher rate of convergence and substantially smaller scatter than in CV methods.

Here a measure of distance between the true object $f^{(r)}(\cdot)$ and its estimator $f_n^{(r)}(\cdot)$ is selected as a mean integrated square error (*MISE*)

$$MISE_r(h) = \mathsf{E} \int \left(\hat{f}_h^{(r)}(x) - f^{(r)}(x)\right)^2 dx,\ r = 0, 1,\ f^{(0)}(x) = f(x). \tag{9}$$

This criterion depends on bandwidth h, and it would be natural to select such h, that will minimize $MISE_r(h)$. Unfortunately it can't be done directly because the true object $f^{(r)}(\cdot)$ is unknown. Therefore we will try to construct an estimate of $MISE_r(h)$, which will be minimized over h. This will be done by using the aforementioned SCV method for criterion $MISE(h)$. Applying gaussian kernels $K(\cdot)$ in (6), it provides for the expression [8]

$$SCV(h)=\frac{1}{2\sqrt{\pi}nh}+\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left\{\varphi_{\sqrt{2h^2+2g^2}}-2\varphi_{\sqrt{h^2+2g^2}}+\varphi_{\sqrt{2}g}\right\}(x_i-x_j), \quad (10)$$

where a new constant g is responsible for data presmoothing. Selection of g in turn is performed by minimization of mean square error of bandwidth estimate $\hat{h}_n(g)$, which minimizes (10). It brings to the following expression:

$$\hat{g}=\left(\frac{15}{16\sqrt{\pi}\nu_6}\right)^{1/7}n^{-1/7}, \quad (11)$$

where $\nu_{2k}=\int f^{(2k)}(x)f(x)dx,\ k=\overline{0,4}$.

Analogous technique provides an estimate for derivative $MISE_1$ in the more complex form [4]

$$\begin{aligned} SCV_1(h_1) = {} & \frac{1}{4\sqrt{\pi}nh_1^3}+\frac{1}{n}\left(\frac{1}{4\sqrt{\pi}g^3}-\frac{2}{\sqrt{2\pi}(h_1^2+2g^2)^{3/2}}+\frac{(n-1)/n}{\sqrt{2\pi}(2h_1^2+2g^2)^{3/2}}\right)+ \\ & +\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{2g^2-(x_i-x_j)^2}{(2g^2)^2}\varphi_{\sqrt{2}g}(x_i-x_j)- \\ & -2\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{h_1^2+2g^2-(x_i-x_j)^2}{(h_1^2+2g^2)^2}\varphi_{(h_1^2+2g^2)^{1/2}}(x_i-x_j)+ \\ & +\frac{n-1}{n}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{2h_1^2+2g^2-(x_i-x_j)^2}{(2h_1^2+2g^2)^2}\varphi_{(2h_1^2+2g^2)^{1/2}}(x_i-x_j), \end{aligned} \quad (12)$$

where g minimizing the mean square error of $\hat{h}_1(g)$ is defined as

$$\hat{g}_1=\left(\frac{105}{32\sqrt{\pi}\nu_8}\right)^{1/9}n^{-1/9}. \quad (13)$$

Both formulae (11) and (13) contain parameters ν_6 and ν_8, which are dependent from unknown density $f(x)$ and its derivatives. They are also can be estimated using cross-validation method for density and rule of thumb for higher derivative. In the end we get the following data-based expressions:

$$\nu_6=\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{\hat{\sigma}^6}\left(\frac{b_{ij}^6}{\hat{\sigma}^6}-15\frac{b_{ij}^4}{\hat{\sigma}^4}+45\frac{b_{ij}^2}{\hat{\sigma}^2}-15\right)\varphi_1(b_{ij}), \quad (14)$$

$$\nu_8=\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{\hat{\sigma}^8}\left(\frac{b_{ij}^8}{\hat{\sigma}^8}-28\frac{b_{ij}^6}{\hat{\sigma}^6}+210\frac{b_{ij}^4}{\hat{\sigma}^4}-420\frac{b_{ij}^2}{\hat{\sigma}^2}+105\right)\varphi_1(b_{ij}), \quad (15)$$

where $\varphi_1(\cdot)$ is the standard gaussian density, $b_{ij} = (X_i - X_j)$, and $\hat{\sigma}$ is the standard rms deviation calculated from the sample $X_1, \ldots, X_n$.

Now everything is ready for modeling the algorithm of automatic nonparametric filtration of the unknown signal from additive mixer with noise according the equation (2).

3 Regularized estimate

The statistics $\hat{\psi}_n(x_{n-\tau}^n)$ in expression (8), representing the ratio of density derivative estimate to density estimate, is unstable when denominator is near zero. To circumvent this drawback it is proposed some regularized procedure, called piecewise smooth approximation [5]. In special case this procedure coincides with the Tychonoff regularization method. Using this procedure we may design a stable statistics

$$\breve{\psi}(x_{n-\tau}^n) = \frac{\hat{\psi}_n(x_{n-\tau}^n)}{1 + \delta|\hat{\psi}_n(x_{n-\tau}^n)|^4}, \tag{16}$$

where optimal value for δ is defined by expression

$$\delta_{opt} = \frac{\int u^2\left(\hat{\psi}_n(x_{n-\tau}^n)\right)\omega(x_{n-\tau}^n)dx_{n-\tau}^n + \int b\left(\hat{\psi}_n(x_{n-\tau}^n)\right)\psi_n(x_{n-\tau}^n)\omega(x_{n-\tau}^n)dx_{n-\tau}^n}{\int \left(\psi_n(x_{n-\tau}^n)\right)^6\omega(x_{n-\tau}^n)dx_{n-\tau}^n + \int b\left(\hat{\psi}_n(x_{n-\tau}^n)\right)\left(\psi_n(x_{n-\tau}^n)\right)^5\omega(x_{n-\tau}^n)dx_{n-\tau}^n}. \tag{17}$$

Here

$$u^2\left(\hat{\psi}_n(x_{n-\tau}^n)\right) = \mathsf{E}\left(\hat{\psi}_n(x_{n-\tau}^n) - \psi(x_{n-\tau}^n)\right)^2$$

and

$$b\left(\hat{\psi}_n(x_{n-\tau}^n)\right) = \mathsf{E}\hat{\psi}_n(x_{n-\tau}^n) - \psi(x_{n-\tau}^n)$$

are the mean square error and the bias of the estimate $\hat{\psi}_n(x_{n-\tau}^n)$ accordingly.

To facilitate finding of the integrals in (17) we take the weight function $\omega(x) = f^2(x)$. Notice that all functions in the integrals of (17) must be estimated preliminary to calculate the parameter δ_{opt}.

4 Modeling results

First of all in modeling we must generate a sequence of dependent observations, using the state equation (3) for S_n and observation equation (2) for X_n. The exact information about both mentioned equations gives us the opportunity to design Kalman filter with respect to optimal estimate $\hat{S}_n$. The filter equation is well known and isn't represent here.

When the state equation is unknown, we make use of nonparametric counterpart of the optimal equation (4), which, taking into account expressions (6), (7), can be represented as

$$\tilde{S}_n = \frac{B^2}{A}\hat{\psi}_n(x_{n-\tau}^n) + \frac{x_n}{A}, \tag{18}$$

where

$$\hat{\psi}_n(x_{n-\tau}^n) = \frac{h_{1n}^{-(\tau+3)}\sum_{i=1}^{n-\tau-1}(x_{n-j-i+1} - x_{n-j+1})\prod_{j=1}^{\tau}\exp\left(-\frac{(x_{n-j+1} - x_{n-j-i+1})^2}{2h_{1n}^2}\right)}{h_n^{-(\tau+1)}\sum_{i=1}^{n-\tau-1}\prod_{j=1}^{\tau+1}\exp\left(-\frac{(x_{n-j+1} - x_{n-j-i+1})^2}{2h_n^2}\right)} \tag{19}$$

is a nonparametric estimate of $\psi(x_{n-\tau}^n)$.

This nonparametric estimate is a substitution estimate (SE). Unfortunately SE is unstable when denominator of (19) possesses a small values. In this case the estimate may have spikes, which can be seen in Fig.1. This spikes are sharply impaired the performance of SE (look at table). To eliminate the spikes we use the regularization method, introduced in (16). In our modeling example this method is reduced to replacement the expression $\hat{\psi}_n(x_{n-\tau}^n)$ in (18) by the approximation (16), where δ is defined by expression (17). Unfortunately the direct calculation of (17) is impossible in view of lack of knowledge about true density and only some estimate is possible, that will be calculated in the next paper. Now in modeling we let $\delta = 0.05$, bearing in mind that we can't make worse the substitution estimate more then 5% in the absence of spikes. The equation for regularized estimation takes the form

$$\breve{S}_n = \frac{B^2}{A}\breve{\psi}_n(x_{n-\tau}^n) + \frac{x_n}{A}.$$

Comparison of nonparametric estimates $\tilde{S}_n$ and $\breve{S}_n$ with optimal Kalman estimate $\hat{S}_n$ is carried out by calculating the relative error ε in percentage

$$\varepsilon = \frac{u_{non} - u_{kal}}{u_{kal}}100, \tag{20}$$

where $u_{non} = (\tilde{u}_{non}$ or $\breve{u}_{non})$, $\tilde{u}_{non} = (1/n\sum_k (S_k - \tilde{S}_k)^2)^{1/2}$, $\breve{u}_{non} = (1/n\sum_k (S_k - \breve{S}_k)^2)^{1/2}$, and $u_{kal} = (1/n\sum_k (S_k - \hat{S}_k)^2)^{1/2}$. Nonparametric estimates $\tilde{S}_n$ and $\breve{S}_n$ together with optimal Kalman estimate $\hat{S}_n$ are represented in Fig.1 and Fig.2.

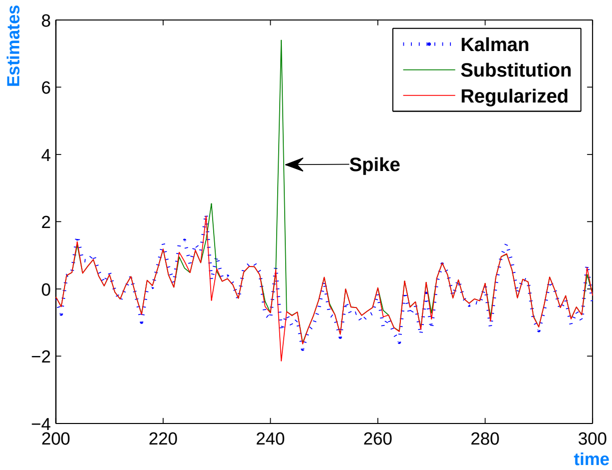

Fig. 1. Comparison of nonparametric and optimal Kalman filtration with spikes

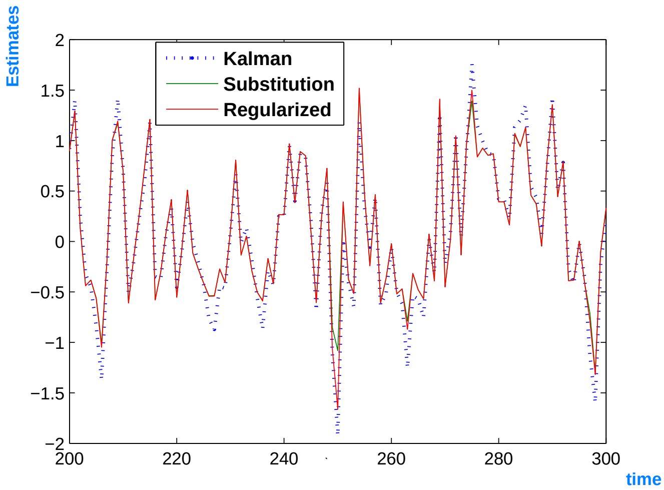

Fig. 2. Comparison of nonparametric and optimal Kalman filtration without spikes

It is easy to note that discrepancy ε between both estimates is very little when the spikes is out. But when the spikes are present the advantage of the regularization procedure becomes obvious.

The distances between nonparametric estimates $\tilde{S}_n$ and S_n and optimal Kalman estimate $\hat{S}_n$ in ε-units are reflected in Table.

Table 1:
Measure of closeness of estimates $\tilde{S}_n$ and $\breve{S}_n$
to Kalman estimate $\hat{S}_n$

Substitution $\hat{\varepsilon}$	Regularized $\breve{\varepsilon}$	Spikes
83,13%	1,42%	yes
1,13%	1,31%	no

References

[1] Dobrovidov A.V. (1983). Nonparametric methods of nonlinear filtering of stationary random sequenses. *Automat. and Remote Control*, **44(6)**, pp. 757–768.

[2] Dobrovidov A.V.and Koshkin G.M. (1997). *Nonparametric signal estimation*. Moscow: Phizmatlit (in Russian).

[3] Duong, T. and Hazelton M.L. (2005). Cross-validation Bandwidth Matrices for Multivariate Kernel Density Estimation, *Scand. J. Statist.*, 2005, vol. 32, pp. 485–506.

[4] Dobrovidov A.V. and Rudko I.M. (2010). Bandwidth Selection in Nonparametric Estimator of Density Derivative by Smoothed Cross-validation Method. *Automation and Remote Control*, **71(2)**, pp. 42-58.

[5] Vasiliev V.A., Dobrovidov A.V. and Koshkin G.M.(2004). *Nonparametric estimation of functionals of stationary sequences distributions*. Moscow: Nauka (in Russian).

[6] Bowman, A., An Alternative Method of Cross-validation for the Smoothing of Density Estimates, *Biometrica*, 1984, vol. 71, pp. 353–360.

[7] Rudemo, M., Empirical Choice of Histograms and Kernel Density Estimators, *Scand. J. Stat.*, 1982, vol. 9, pp. 65–78.

[8] Hall, P., Marron, J., and Park, B., Smoothed Cross-validation, *Prob. Theory Related Fields*, 1992, vol. 90, pp. 1–20.

[9] Park, B. and Marron, J., Comparison of Data-driven Bandwidth Selectors, *J. Amer. Statist. Assoc.*, 1990, vol. 85, pp. 66–72.