AUSTRIAN JOURNAL OF STATISTICS
Volume 40 (2011), Number 1 & 2, 15–24

Data-Based Nonparametric Signal Filtration

Alexander V. Dobrovidov[1] and Gennadij M. Koshkin[2]
[1]Russian Academy of Sciences, Moscow, Russia
[2]Tomsk State University, Tomsk, Russia

Abstract: The problem of stochastic signal filtration under nonparametric uncertainties is considered. A probabilistic description of the signal process is assumed to be completely unknown. The Bayes estimator can not be constructed in this case. However if the conditional density of the observation process given signal process belongs to conditionally exponential family, the optimal Bayes estimator is a solution to some non-recurrent equation which is explicitly independent upon the signal process distribution. In this case, the Bayes estimator is expressed in terms of conditional distribution of the observation process, which can be approximated by using of the stable nonparametric procedures, adapted to dependent samples. These stable approximations provide the mean square convergence to Bayes estimator. In the stable kernel nonparametric procedures, a crucial step is to select a proper smoothing parameter (bandwidth) and a regularized parameter, which have a considerable influence on the quality of signal filtration. The optimal procedures for selecting of these parameters are proposed. These procedures allow to construct the automatic (data-based) signal filtration algorithm.

Keywords: Signal Filtration, Nonparametric Uncertainty, Kernel Estimates, Bandwidth Selection, Regularization.

1 Introduction and Problem Setting

In the sixties and seventies of the last century much efforts have been made to develop the supervised and nonsupervised methods of machine learning, based on the strict methods of mathematical statistics. One of these efforts, directed to the solution of the problems of nonsupervised stochastic signal processing with unknown probabilistic characteristics, led to the theory of nonparametric signal estimation, introduced first in Dobrovidov (1983). In this theory stochastic state models of the useful signals are assumed to be completely unknown but the observation models and distribution of noise in measurement devices assumed to be known. Such situation is typical for applications, e.g., in astronomy or hydroacoustics, when mathematical models of telescope or hydroacoustic radar and their characteristics are well-known, but useful signals are generated by the unknown sources. If useful signals are not observable, then it is impossible to restore theirs distributions. Consequently, one can't construct the optimal Bayes estimator of unobservable random signal by using the realization of observed process. The solution of the signal filtration in this case was proposed in Dobrovidov and Koshkin (1997) and Vasiliev, Dobrovidov, and Koshkin (2004). The principal result of this approach is the obtaining of the optimal filtration equation in the form explicitly independent of unknown distribution of useful stochastic signal. Such form is possible when the conditional observation density under fixed useful signal belongs to a *conditionally-exponential family* Dobrovidov and Koshkin (1997). This family, particularly, contains Gaussian density

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

which will be considered in this paper as an example. It follows from expression (1) that the observation equation has the form

$$X_n = S_n + \sigma\eta_n\,, \tag{2}$$

where σ is a known constant, X_n is an observation, S_n is an unobservable signal, and η_n is an independent noise at the instant n. This example clearly demonstrates an opportunity of proposed approach.

The problem is to obtain an optimal in mean square sense estimator $\hat{S}_n$ of an useful signal S_n at the moment n from given observations $X_1^n = (X_1, \dots, X_n)$. As it is well-known, the optimal estimator of S_n is a conditional mean $\hat{S}_n = \mathrm{E}[S_n|X_1^n = x_1^n]$. This conditional mean can be calculated by the 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 an independent Gaussian noise, Kalman filter is optimal, and for its construction it is necessary to know (3) exactly. Very often such information is not available for users. Are there any ways to circumvent the necessity to know signal state equation? One of these ways is the empirical Bayes approach, by following which one can obtain an equation for conditional mean $\hat{S}_n$ without information about the state equation (3). In this case it is necessary only to have the information about the observation equation of the type (1). Then it follows from Dobrovidov and Koshkin (1997) that the equation for optimal estimator $\hat{S}_n$ takes on a form

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

where $f(x_n|x_1^{n-1})$ is a conditional density of the 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 (3) is unknown. However it can be restored from observations x_1^n with the prescribed precision, using the nonparametric kernel method of estimation from dependent data (Vasiliev et al., 2004). According to this method we must replace the unknown density $f(x_n|x_1^{n-1})$ by the truncated density $\bar{f}(x_n|x_{n-\tau}^{n-1})$, where τ is degree of dependence of observable process (X_n). Here τ represents an order of connectivity of Markov process approximating the non-Markov 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

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

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

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

The nonparametric estimate for the numerator of (5) has the form

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

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

Using (6), (7) we obtain the nonparametric estimate for the logarithmic density derivative $\psi(x_{n-\tau}^n)$:

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

Such estimate is referred as *the plug-in estimate*. To calculate (8) it remains only to select bandwidths h_n for the density in (6) and h_{1n} for the derivative in (7).

2 Bandwidth Selection for Densities and their Derivatives

Several data-based selection methods of the kernel function bandwidth are well-known. The methods of cross-validation CV (Bowman, 1984; Rudemo, 1982), smoothed cross-validation SCV (Hall, Marron, and Park, 1992), and plug-in (Park and Marron, 1990) seem to be the basic ones as the most clear and rapidly converging procedures. In Dobrovidov and Rudko (2010) the method SCV developed in Duong and Hazelton (2005) for density estimation was extended to the kernel estimates of the density derivatives. Both of these methods generate data-based bandwidth estimates with a higher rate of convergence and substantially smaller scatter than in CV methods.

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

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

This criterion depends on the bandwidth h, and it is natural to select such h, that minimizes $MISE_r(h)$. Unfortunately it cannot be done directly because the 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 above-mentioned SCV method for criterion $MISE(h)$. Applying the Gaussian kernels $K(\cdot)$ in (6) gives the expression (Hall et al., 1992)

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

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 leads to the following expression:

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

where

$$\nu_k = \int f^{(k)}(x) f(x) dx , \qquad k = \overline{0,8}. \tag{12}$$

An estimate for the derivative $MISE_1$ has the form (Dobrovidov and Rudko, 2010)

$$\begin{aligned} SCV_1(h_1) = & \frac{1}{4\sqrt{\pi} n h_1^3} + \frac{1}{n} \left(\frac{1}{4\sqrt{\pi} g^3} - \frac{2}{\sqrt{2\pi}(h_1^2+2g^2)^{3/2}} + \frac{(n-1)/n}{\sqrt{2\pi}(2h_1^2+2g^2)^{3/2}} \right) \\ & + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{2g^2 - (x_i - x_j)^2}{(2g^2)^2} \varphi_{\sqrt{2g}}(x_i - x_j) \\ & - 2\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{h_1^2 + 2g^2 - (x_i - x_j)^2}{(h_1^2+2g^2)^2} \varphi_{(h_1^2+2g^2)^{1/2}}(x_i - x_j) \\ & + \frac{n-1}{n} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{2h_1^2 + 2g^2 - (x_i - x_j)^2}{(2h_1^2+2g^2)^2} \varphi_{(2h_1^2+2g^2)^{1/2}}(x_i - x_j) , \end{aligned} \tag{13}$$

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

$$\hat{g}_1 = \left(\frac{105}{32\sqrt{\pi}\nu_8} \right)^{1/9} n^{-1/9} . \tag{14}$$

As it is shown in Dobrovidov and Rudko (2010), the expression (14) allows to calculate the mean square rate of convergence of $\hat{h}_{SCV_1}$ to h_{AMISE_1}, which is equal to $n^{-32/63} = n^{-0.507}$. The value

$$h_{AMISE_1} = C_0 n^{-1/7} , \qquad C_0 = \left(\frac{3R(K')}{\mu_2^2(K) R(f''')} \right)^{1/7} \tag{15}$$

is the explicit expression for the bandwidth, that minimizes the asymptotic $MISE_1$ for the derivative.

Both formulae (11) and (14) contain parameters ν_6 and ν_8, which depend upon the unknown density $f(x)$ and its derivatives. They also can be estimated using the cross-validation method for the density and the rule of thumb (see below) for higher derivative.

According to the law of large numbers, the integral (12) is approximated by the sum

$$\frac{1}{n} \sum_{i=1}^{n} f_{h,i}^{(k)}(X_i) ,$$

where the function $f_{h,i}^{(k)})(X_i)$ can be estimated by the cross-validation method:

$$\hat{f}_{h,i}^{(k)}(X_i) = \frac{1}{n-1} \sum_{j \neq i} K_h^{(k)}(X_i - X_j) . \tag{16}$$

Such estimates unlike to (14) are referred to as the estimates of the second level, where less precision is admissible. For Gaussian kernels $K(u) = \varphi_1(x)$, where $\varphi_1(x)$ is the standard normal density, calculation of derivatives in (16) is accomplished by using the well-known formula

$$\varphi_1^{(k)}(x) = (-1)^k H_k(x)\varphi_1(x)\,, \tag{17}$$

where $H_k(x)$ is the Hermitian polynomial

$$H_{k+1}(x) = xH_k(x) - nH_{k-1}(x)\,, \qquad H_0(x) = 1\,, \quad k = 1, 2, \ldots.$$

At last, the kernel bandwidth h of the second level is estimated roughly from observations by using the *rule of thumb* (Duong and Hazelton, 2005):

$$\tilde{h} = 1.06\hat{\sigma}n^{-1/5}\,.$$

The quantity $\tilde{h}$ gives the optimal value of kernel bandwidth under the assumption that unknown density is Gaussian, where $\hat{\sigma}$ is the sample standard deviation, calculated from x_1^n.

As a result we obtain the following data-based expressions:

$$\nu_6 = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{1}{\tilde{h}^6} \left(\frac{b_{ij}^6}{\tilde{h}^6} - 15\frac{b_{ij}^4}{\tilde{h}^4} + 45\frac{b_{ij}^2}{\tilde{h}^2} - 15 \right) \varphi_{\tilde{h}}(b_{ij})\,,$$

$$\nu_8 = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{1}{\tilde{h}^8} \left(\frac{b_{ij}^8}{\tilde{h}^8} - 28\frac{b_{ij}^6}{\tilde{h}^6} + 210\frac{b_{ij}^4}{\tilde{h}^4} - 420\frac{b_{ij}^2}{\tilde{h}^2} + 105 \right) \varphi_{\tilde{h}}(b_{ij})\,,$$

where $b_{ij} = (X_i - X_j)$.

3 Mean Square Convergence and Stability

Logarithmic density derivative estimate $\hat{\psi}_n(x_{n-\tau}^n)$, described by (8), is a special case of general plug-in estimating composite function $G(t_n(x))$, where $x \in \mathbb{R}^{\tau+1}$, $t_n : \mathbb{R}^{\tau+1} \to \mathbb{R}^m$, $G : \mathbb{R}^m \to \mathbb{R}$. In the case under consideration $m = 2$, $t_n = (t_{1n}, t_{2n})$, $t_{1n}(x) = f_n(x)$, $t_{2n} = f_n'(x)$, and $G(t_n) = t_{2n}/t_{1n}$. If the statistics t_n converges to a function t as $n \to \infty$, than under some regularity conditions $G(t_n) \to G(t)$. The main part of these regularity conditions is

1. the existence and boundedness of several derivatives of $G(t)$;

2. sequence $\{|G(t_n)|\}$ is dominated by the number sequence $(C_0 d_n^\gamma)$, where C_0 is a constant, $d_n \to \infty$ as $n \to \infty$ and $0 \leqslant \gamma < \infty$.

Roughly speaking, the function $G(t_n)$ must grow slower, than power function of n. These conditions provide a mean square convergence of $G(t_n) \to G(t)$ (see Theorem 1.8.1 in Vasiliev et al., 2004).

If the mean Euclidean distance $\mathrm{E}\|t_n - t\|$ satisfies the inequality $\mathrm{E}\|t_n - t\| < \varepsilon$, $\varepsilon > 0$, then for a small ε the equality

$$G(t_n) - G(t) = \nabla G(\vartheta_n)(t_n - t)^{\mathrm{T}}\,, \qquad \vartheta_n \in (t, t_n)\,,$$

is valid, where ∇ is the gradient. This yields a following result (see Theorem 1.9.1 in Vasiliev et al., 2004):

$$\left| \mathrm{E}[G(t_n) - G(t)]^2 - \mathrm{E}\left[\nabla G(t)(t_n - t)^{\mathrm{T}}\right]^2 \right| = O(d_n^{-3/2}) \,, \tag{18}$$

i.e., the mean square closeness of the composite functions $G(t_n)$ and $G(t)$ is replaced by the mean square closeness of more simple statistics' t_n and t.

There are a number of cases, when conditions 1. and 2. do not hold. For example, function $G = 1/t$ does not satisfy both conditions and its estimator $G(t_n)$ becomes unstable, because it may be unbounded. In our case, $G = t_2/t_1$, where $t_1 = f(x)$ and $t_2 = f'(x)$ and for the Gaussian density $f(x)$ we have $G = -x$. This function is unbounded on $\mathbb{R}$. Because of the proposition (18) is valid only for bounded functions G, it is proposed here some regularized procedure, called a *piecewise smooth approximation* (Vasiliev et al., 2004). In the special case this procedure coincides with the Tychonoff regularization method. Using this procedure we may obtain a stable approximation of G in the form

$$\Phi(G(t), \delta_n) = \widetilde{\Phi}(t, \delta_n) = \frac{G(t)}{(1 + \delta_n |G(t)|^4)} \,, \tag{19}$$

where $\delta_n > 0$ is a regularization parameter. As it is proved in Vasiliev et al. (2004), $\widetilde{\Phi}(t_n, \delta_n)$ satisfies both above mentioned conditions and therefore is dominated by the power function of n. Moreover $\widetilde{\Phi}(t_n, \delta_n)$ converges to $G(t)$ in the mean square sense, i.e., as $\mathrm{E}\|t_n - t\| \to 0$ and $\delta_n \to 0$, then

$$\lim_{n\to\infty} \mathrm{E}(\widetilde{\Phi}(t_n, \delta_n) - G(t))^2 = 0 \,. \tag{20}$$

The statistic $\hat{\psi}_n(x^n_{n-\tau})$ in the expression (8) is unstable as well when the denominator of (8) is close to zero. Therefore it is reasonable to apply to it the procedure of regularization

$$\breve{\psi}(x^n_{n-\tau}) = \frac{\hat{\psi}_n(x^n_{n-\tau})}{1 + \delta |\hat{\psi}_n(x^n_{n-\tau})|^4} \,, \tag{21}$$

where the regularization parameter δ has to be find. In Vasiliev et al. (2004) it was found an optimal value of this parameter, which minimizes the mean square deviation of $\widetilde{\Phi}(t_n, \delta_n)$ from $G(t)$ at each point x. This performance is not so good for practice, because a minimization procedure has to be repeated in each signal processing. It is proposed here to make an optimization procedure only once before signal processing, using the criterion of mean integrated square error ($MISE$) for estimating the logarithmic density derivative with the weight density $f^2(\cdot)$

$$\begin{aligned} MISE(\delta) \; &= \int u^2\left(\breve{\psi}_n(x^n_{n-\tau})\right) f^2(x^n_{n-\tau}) dx^n_{n-\tau} \\ &:= \int \mathrm{E}\left(\breve{\psi}_n(x^n_{n-\tau}) - \psi(x^n_{n-\tau})\right)^2 f^2(x^n_{n-\tau}) dx^n_{n-\tau} \,. \end{aligned} \tag{22}$$

The weight function is selected in such a form in order the criterion would exist.

Calculating of the expectation of ratio in (22) is laborious. According to (20) for the mean square convergence of the regularized estimate $\breve{\psi}_n(\cdot)$ to logarithmic density derivative $\psi(\cdot)$ it is necessary that $\delta = \delta_n \to 0$ as $n \to \infty$. Therefore under the assumption of small δ we expand (21) over parameter δ and approximately obtain

$$\breve{\psi}(x_{n-\tau}^n) \approx \hat{\psi}_n(x_{n-\tau}^n) - \delta\hat{\psi}_n^5(x_{n-\tau}^n)\,. \tag{23}$$

Putting (23) into the mean integrated square error (22) yields

$$\int \mathrm{E}\left(\hat{\psi}_n(x_{n-\tau}^n) - \delta\hat{\psi}_n^5(x_{n-\tau}^n) - \psi(x_{n-\tau}^n)\right)^2 f^2(x_{n-\tau}^n)dx_{n-\tau}^n\,. \tag{24}$$

The equality (18), valid for plug-in estimates, helps to calculate this expression, if we slightly touch up the criterion, replacing $\psi(x_{n-\tau}^n)$ by $\psi(x_{n-\tau}^n) - \delta\psi^5(x_{n-\tau}^n)$, i.e., $G = \psi - \delta\psi^5$. Then

$$\begin{aligned}
&\int u^2\left(\hat{\psi}_n(x_{n-\tau}^n) - \delta\hat{\psi}_n^5(x_{n-\tau}^n)\right) f^2(x_{n-\tau}^n)dx_{n-\tau}^n \\
&:= \int \mathrm{E}\left[\left(\hat{\psi}_n(x_{n-\tau}^n) - \delta\hat{\psi}_n^5(x_{n-\tau}^n\right) - \left(\psi(x_{n-\tau}^n) - \delta\psi^5(x_{n-\tau}^n)\right)\right]^2 f^2(x_{n-\tau}^n)dx_{n-\tau}^n \\
&\approx \int G_1^2 u^2(\hat{f}'(x_{n-\tau}^n))f^2(x_{n-\tau}^n)dx_{n-\tau}^n \\
&\quad +2\int G_1G_2\mathrm{cov}(\hat{f}'(x_{n-\tau}^n), \hat{f}(x_{n-\tau}^n))f^2(x_{n-\tau}^n)dx_{n-\tau}^n \\
&\quad +\int G_2^2 u^2(\hat{f}(x_{n-\tau}^n))f^2(x_{n-\tau}^n)dx_{n-\tau}^n\,,
\end{aligned} \tag{25}$$

where $G_1 = (1 - 5\delta\psi^4)/f$ and $G_2 = (-\psi + 5\delta\psi^5)/f$.

By minimizing (25) with respect to δ we find the optimal value of δ:

$$\delta_{opt} = \frac{\int \psi^4 u^2(\hat{f}') - 2\int \psi^5 \mathrm{cov}(\hat{f}', \hat{f}) + \int \psi^6 u^2(\hat{f})}{5\int \psi^8 u^2(\hat{f}') - 10\int \psi^9 \mathrm{cov}(\hat{f}', \hat{f}) + 5\int \psi^{10} u^2(\hat{f})}\,. \tag{26}$$

Integrals in numerator and denominator of δ_{opt} depend on unknown densities and can not be calculated directly. So they must be replaced by estimates.

The main parts of expansions of $u^2(\cdot)$ and $\mathrm{cov}(\cdot,\cdot)$ as $n \to \infty$ are

$$\begin{aligned}
u^2(\hat{f}') &\approx \frac{f}{nh_n^3}\int (K^{(1)}(u))^2du + \frac{h_n^4}{4}(f^{(3)})^2\left(\int u^2K(u)du\right)^2, \\
\mathrm{cov}(\hat{f}', \hat{f}) &\approx \frac{f}{nh_n^2}\int K^{(1)}(u)K(u)du + \frac{h_n^4}{4}f^{(3)}f^{(2)}\left(\int u^2K(u)du\right)^2, \\
u^2(\hat{f}) &\approx \frac{f}{nh_n}\int K^2(u))du + \frac{h_n^4}{4}(f^{(2)})^2\left(\int u^2K(u)du\right)^2.
\end{aligned}$$

Substituting these formulae into (26) we find the intermediate expression for δ_{opt}, in which it is necessary to estimate slightly more complicated integrals than (12):

$$J_k = \int (f^{(k)}(u))^q f(u)du\,, \qquad \nu = 0, \ldots, 4\,, \quad q = 1, 2, \ldots.$$

It can be done by cross-validation method, described above in Section 3.

4 Comparison with Kalman Filter

Computer modeling is started by generation 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 Kalman filter equation is well known and isn't represented here.

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

$$\tilde{S}_n = \sigma^2 \hat{\psi}_n(x_{n-\tau}^n) + x_n \,, \tag{27}$$

where

$$\hat{\psi}_n(x_{n-\tau}^n) = \frac{h_{1n}^{-(\tau+3)} \sum_{i=1}^{n-\tau-1} (x_{n-j-i+1} - x_{n-j+1}) \prod_{j=1}^{\tau} \exp\left(-\frac{(x_{n-j+1} - x_{n-j-i+1})^2}{2h_{1n}^2}\right)}{h_n^{-(\tau+1)} \sum_{i=1}^{n-\tau-1} \prod_{j=1}^{\tau+1} \exp\left(-\frac{(x_{n-j+1} - x_{n-j-i+1})^2}{2h_n^2}\right)} \tag{28}$$

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

Unfortunately PE is unstable when the denominator of (28) vanishes. In this case the estimate may have spikes, which can be seen in Figure 1 (left). This spikes are sharply impaired the performance of PE (look at table). To eliminate the spikes we use the regularization method, introduced in (21). In our modeling example this method is reduced to replacement the expression $\hat{\psi}_n(x_{n-\tau}^n)$ in (27) by the approximation (21), where δ is defined by expression (26). The direct calculation of (26) is impossible in view of lack of knowledge about the true density and only the estimate is designed using the cross-validation method. For the different samples, generated by the model (2), we receive the following range of values $[0.01 - 0.05]$ for the regularized parameter δ. This fact implies that the assumption about smallness of the parameter δ is confirmed.

The equation for regularized estimation takes the form

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

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

$$\varepsilon = \frac{u_{non} - u_{kal}}{u_{kal}} 100 \,, \tag{29}$$

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

It is easy to note that the discrepancy ε between both estimates is very small when the spikes are 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 given in Table 1.

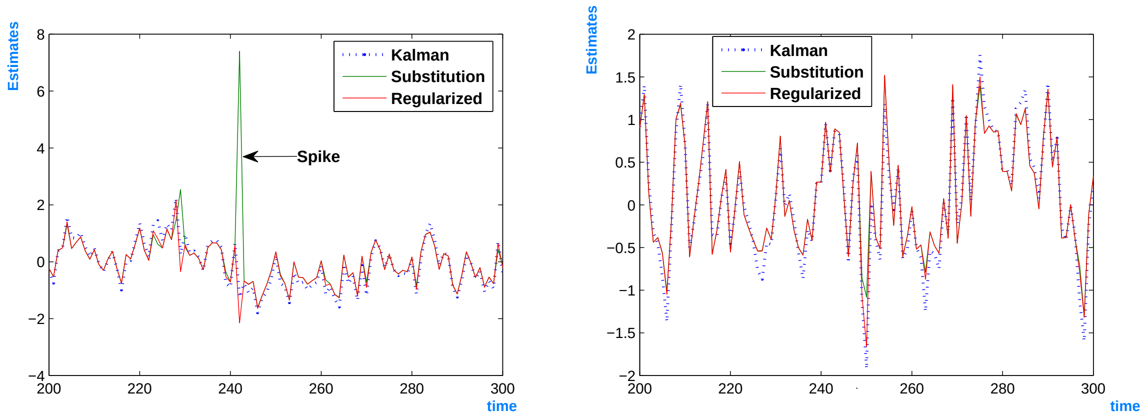

Figure 1: Comparison of nonparametric and optimal Kalman filtration with spikes (left) and without spikes (right).

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

Plug-in $\hat{\varepsilon}$	Regularized $\breve{\varepsilon}$	Spikes
83.13%	1.42%	yes
1.13%	1.31%	no

5 Conclusion

The nonparametric filtering algorithm of a random signal with unknown distribution is presented. The stable counterpart of the filtering procedure is proposed. It is proved that the nonparametric stable estimate converges in the mean square sense to the optimal Bayes estimate, constructed by using full information about stochastic models under consideration. Applying the smoothed cross-validation method to the bandwidth and to the regularization parameter selection we have found the optimal estimates of these parameters for fixed sample size, making the algorithm to be automatic. For linear stochastic models, for which Kalman filter exists, it is shown by computer modeling that the stable nonparametric filtration approximates the optimal filter with sufficiently high precision.

Acknowledgements

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

References

Bowman, A. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrica*, *71*, 353-360.

Dobrovidov, A. (1983). Nonparametric methods of nonlinear filtering of stationary random sequenses. *Automat. and Remote Control*, *44*, 757-768.

Dobrovidov, A., and Koshkin, G. (1997). *Nonparametric Signal Estimation*. Moscow: Phizmatlit.

Dobrovidov, A., and Rudko, I. (2010). Bandwidth selection in nonparametric estimator of density derivative by smoothed cross-validation method. *Automation and Remote Control*, *71*, 42-58.

Duong, T., and Hazelton, M. L. (2005). Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scandinavian Journal of Statistics*, *32*, 485-506.

Hall, P., Marron, J., and Park, B. (1992). Smoothed cross-validation. *Probability Theory and Related Fields*, *92*, 1-20.

Park, B., and Marron, J. (1990). Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, *85*, 66-72.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, *9*, 65-78.

Vasiliev, V., Dobrovidov, A., and Koshkin, G. (2004). *Nonparametric Estimation of Functionals of Stationary Sequences Distributions (in Russian)*. Moscow: Nauka.

Authors' addresses:

Alexander V. Dobrovidov
Trapeznikov Institute of Control Sciences
Russian Academy of Sciences
6-3-174 Sumskaya, Moscow, 113587
Russia

E-mail: dobrovid@ipu.ru

Gennadiy M. Koshkin
Tomsk State University
Department of Applied Mathematics and Cybernetics
Tomsk State University
43-9 Lenina, Tomsk
Russia

E-mail: kgm@mail.tsu.ru