ASYMPTOTICALLY OPTIMAL NONPARAMETRIC SIGNAL INTERPOLATION

A.V. Dobrovidov
Trapeznikov Institute of Control Sciences,
Russian Academy of Sciences, Moscow, Russia
e-mail:dobrovid@ipu.rssi.ru

Abstract

The problem of interpolation (smoothing) of a partially observable Markov random sequence is considered. For the dynamic observation models, an equation in the interpolation posterior probability density is derived. This equation has a certain form of the normalized product of the posterior probability densities in forward and backward times and differs from its counterpart for static observation models [3, 1] in an additional equation. The aim of this paper is to consider the problem of smoothing for the case of unknown distributions of the unobservable component of the random Markov sequence. For the strongly stationary Markov processes with mixing and for the conditional density of observation model belonging to the exponent family success was reached. A resultant method is based on the empirical Bayes approach and the kernel non-parametric estimation [5]. The equation of the nonlinear optimal smoothing estimate is derived in a form independent of the unknown distributions of an unobservable process. Such form of equation allows one to use the non-parametric estimates of some conditional statistics given any set of dependent observations. Modeling was carried out to compare the nonparametric estimates with optimal mean-square smoothing estimates in Kalman scheme.

1 Problem formulation

By interpolation (smoothing) of a partly observable Markov random sequence $(S_n, X_n)_{n\geqslant 1}, S_n \in \mathbb{R}^m, X_n \in \mathbb{R}^l$ is meant the problem of constructing the estimates of the unobservable vector S_k or a known one-to-one function $Q(S_k)$ from observations $x_1^n = (x_1, ..., x_n)^{\mathrm{T}}$ of a sequence $(X_n)_{n\geqslant 1}$ for all $k \leqslant n$. As is well known, the optimal mean-square smoothing estimate of $Q(S_k), k \leqslant n$, equals the conditional expectation

$$\mathsf{E}(Q(S_k)|x_1^n) = \int\limits_{\mathbb{R}^m} Q(s_k)\pi(s_k|x_1^n)ds_k, \tag{1}$$

where $\pi_k(s_k|x_1^n)$ is the conditional posterior probability density of S_k given all observable realizations x_1^n which will be called the *interpolating* posterior density. There are some ways to calculating this density. One of the interesting ways that is examined below and referred to as the two-filter smoothing lies in the recursive calculation of the *filtering* posterior density $w_k(s_k|x_1^k)$ in forward time and the *filtering* posterior density $\widetilde{w}_k(s_k|x_k^n)$ in backward time. This algorithm is as follows [3, 1]:

$$\pi_k(s_k|x_1^n) = \frac{f(x_1^k)f(x_k^n)}{f(x_1^n)} \cdot \frac{w_k(s_k|x_1^k)\widetilde{w}_k(s_k|x_k^n)}{f(s_k, x_k)}, \tag{2}$$

where the first factor is a normalizing constant, depending only of observations. The subsequent calculations in (2) may be carried out if all distributions of the composed Markov process $(S_n, X_n)_{n\geqslant 1}$ are known.

This paper is devoted to the interpolation problem with unknown distributions of the unobservable strongly stationary Markov process $(S_n)_{n\geqslant 1}$. The main idea for such problem decision is founded on the principle of empirical Bayes approach and the theory of nonparametric functional estimation from weakly dependent observations [5]. The empirical Bayes approach forces one to find such forms of estimates that are explicitly independent of the probabilistic characteristics of the unobservable random variables. This can be achieved, for instance, by using the conditional densities of observations from exponent density family [2]

$$f(x_n|s_n) = \widetilde{C}(s_n)h(x_n)\exp\left\{T^{\mathrm{T}}(x_n)Q(s_n)\right\} \tag{3}$$

where $T = (T_1, \cdots, T_m)^{\mathrm{T}}; Q = (Q^{[1]}, \cdots, Q^{[m]})^{\mathrm{T}}; h(\cdot), Q^{[j]}(\cdot)$ and $T_j(\cdot),\ j = \overline{1, m}$ are the given Borelean functions and $\widetilde{C}(s_n)$ is the normalizing factor.

2 Interpolation equation for the dynamic observation models

The equation (2) is correct both for static and for dynamic observation models, but there is some difference in calculating of joint probability density $f(s_k, x_k)$ in the denominator of this equation. Here we present a brief derivation of the expression for interpolating posterior density (2) in the case of the dynamic observation model described in terms of the conditional density $f(x_{n+1}|x_n, s_{n+1})$, because the intermediate expression will be useful in the sequel. Let $Z_k = (S_k, X_k)_{k\geqslant 1}$ be a compound Markov process. Then by definition the joint probability density

$$f(s_1^n, x_1^n) = f(s_1, x_1)\prod_{k=2}^{n} g(s_k, x_k|s_{k-1}, x_{k-1}), \tag{4}$$

where (s_k, x_k) is the value of the random variable (S_k, X_k), and $f(s_1, x_1)$ and $g(\cdot|\cdot)$ are a priori and transition densities of $(Z_n)_{n\geqslant 1}$. Taking into account (4), the interpolating posterior density $\pi_k(s_k|x_1^n)$ is representable as follows:

$$\begin{aligned}\pi_k(s_k|x_1^n) &= \frac{1}{f(x_1^n)} f(s_k, x_1^n) = \frac{1}{f(x_1^n)} f(s_k, x_1^k, x_{k+1}^n)\\ &= \frac{1}{f(x_1^n)} \cdot f(x_1^k) w_k(s_k|x_1^k) \cdot f(x_{k+1}^n|s_k, x_k)\\ &= \frac{f(x_1^k) f(x_k^n)}{f(x_1^n)} \cdot \frac{w_k(s_k|x_1^k)\widetilde{w}_k(s_k|x_k^n)}{f(s_k, x_k)},\end{aligned} \tag{5}$$

where the density $f(s_k, x_k)$ in the denominator of (5) cannot be calculated by means of the product $p(s_k)f(x_k|s_k)$ because for the dynamic model the conditional density

$f(x_k|s_k)$ is not known, whereas the conditional density $f(x_k|x_{k-1}, s_k)$ is known. Therefore, for the dynamic observation models we need add to equation (2) a new recursion equation

$$f(s_n, x_n) = \int\limits_{\mathbb{S}_{n-1}} p(s_n|s_{n-1}) \int\limits_{\mathbb{X}_{n-1}} f(x_n|x_{n-1}, s_n) f(s_{n-1}, x_{n-1}) dx_{n-1} ds_{n-1} \tag{6}$$

for $f(s_k, x_k)$ with the initial value $f(s_1, x_1)$ which is the prescribed a priori density of the Markov sequence (S_n, X_n).

3 Equation for optimal interpolation estimate under unknown distribution of unobservable signal

This section appeared owing to the practical needs to extract a useful signal with unknown distribution from noise. To construct an optimal estimate under these conditions, one has to return to equation (5). For the sequel we must choose in equation (5) any factors which depend on the observation of x_k. For that the second factor of (5) is transformed as follows:

$$f(x_1^k) w_k(s_k|x_1^k) = \int\limits_{\mathbb{R}^m} f(s_1^{k-2}, x_1^{k-2}, s_{k-1}, x_{k-1}) g(s_k, x_k|s_{k-1}, x_{k-1}) ds_{k-1}.$$

But since for dynamic models the transition density $g(s_k, x_k|s_{k-1}, x_{k-1}) = p(s_k|s_{k-1}) f(x_k|x_{k-1}, s_k)$, we get

$$\begin{aligned} f(x_1^k) w_k(s_k|x_1^k) &= f(x_k|x_{k-1}, s_k) \int\limits_{\mathbb{R}^m} p(s_k|s_{k-1}) f(s_{k-1}, x_1^{k-1}) ds_{k-1} \\ &= f(x_1^{k-1}) f(x_k|x_{k-1}, s_k) f(s_k|x_1^{k-1}). \end{aligned} \tag{7}$$

Obviously, only the second factor of (7) depends on x_k.

Let us consider the third factor in (5). Then we get

$$\begin{aligned} f(x_{k+1}^n|s_k, x_k) &= \int\limits_{\mathbb{R}^m} f(x_{k+2}^n|s_{k+1}, x_{k+1}) g(s_{k+1}, x_{k+1}|s_k, x_k) ds_{k+1} \\ &= \int\limits_{\mathbb{R}^m} f(x_{k+2}^n|s_{k+1}, x_{k+1}) p(s_{k+1}|s_k) f(x_{k+1}|x_k, s_{k+1}) ds_{k+1} \\ &= \int\limits_{\mathbb{R}^m} f(x_{k+2}^n, s_{k+1}, x_{k+1}) \frac{p(s_{k+1}|s_k)}{p(s_{k+1})} \frac{f(x_{k+1}|x_k, s_{k+1})}{f(x_{k+1}|s_{k+1})} ds_{k+1}. \end{aligned} \tag{8}$$

Here only the last ratio $\dfrac{f(x_{k+1}|x_k, s_{k+1})}{f(x_{k+1|s_{k+1}})}$ under integral depends on x_k. It is this dependence on x_k that does not permit us to construct a nonparametric version of the equation (5) for dynamic observation models. However for static observation models $X_n = \varphi(S_n, \eta_n)$, where η_n is independent noise, the conditional density of observations $f(x_{k+1}|x_k, s_{k+1}) = f(x_{k+1}|s_{n+1})$ does not depend on x_k and this makes it possible to perform cancelations in the integral (8) so in this case equation (5) takes the form

$$\pi_k(s_k|x_1^n) = \frac{1}{f(x_1^n)} \cdot f(x_1^k) w_k(s_k|x_1^k) \cdot f(x_{k+1}^n|s_k, x_k)$$
$$= \frac{1}{f(x_1^n)} f(x_k|s_k) f(s_k|x_1^{k-1}) \int\limits_{\mathbb{R}^m} f(x_{k+2}^n, s_{k+1}, x_{k+1}) \frac{p(s_{k+1}|s_k)}{p(s_{k+1})} ds_{k+1}$$
$$= \frac{\lambda_k(x_1^n \text{ without } x_k)}{f(x_k|x_1^n \text{ without } x_k)} f(x_k|s_k) f(s_k|x_1^{k-1}) f(s_k|x_{k+1}^n) p^{-1}(s_k), \tag{9}$$

where only the second factor depends on x_k. Some new definitions for the normalizing constants are introduced here:

$$\lambda_k = \lambda_k(x_1^n \text{ without } x_k) \triangleq \frac{f(x_1^{k-1}) f(x_{k+1}^n)}{f(x_1^{k-1}, x_{k+1}^n)},$$
$$f(x_k|x_1^n \text{ without } x_k) \triangleq f(x_k|x_1^{k-1}, x_{k+1}^n).$$

Such a form of the interpolation equation allows us to obtain its counterpart which is independent of the statistical characteristics of the unobserved process $(S_n)_{n\geqslant 1}$.

We denote

$$u_k(s_k) = f(s_k|x_1^{k-1}) f(s_k|x_{k+1}^n) p^{-1}(s_k) \tag{10}$$

and remark once more that u_k is independent of x_k. Then, equation (9) can be rearranged in

$$\pi_k(s_k|x_1^n) = \frac{\lambda(x_1^n \text{ without } x_k)}{f(x_k|x_1^n \text{ without } x_k)} f(x_k|s_k) u_k(s_k). \tag{11}$$

We integrate this equation with respect to s_k and carry over the normalizing factor depending only on the observations to the left-hand side. Then we get

$$\frac{f(x_k|x_1^n \text{ without } x_k)}{\lambda(x_1^n \text{ without } x_k)} = \int\limits_{\mathbb{S}_k} f(x_k|s_k) u_k(s_k) ds_k. \tag{12}$$

Assuming now that $f(x_k|s_k)$ belongs to the exponent density family (3), we differentiate (12) with respect to x_k. The possibility of differentiating under the sign of integral is justified by the assumption of existence of the second prior moment $\mathsf{E}Q^T(S_k)Q(S_k)$, that is the natural restriction of signal power. The latter is sufficiently to exist of mean-squired risk of the problem. Differentiation in x_k provides the equation

$$\frac{\nabla_{x_k} f(x_k|x_1^n \text{ without } x_k)}{\lambda(x_1^n \text{ without } x_k)} = \int\limits_{\mathbb{S}_k} \nabla_{x_k} f(x_k|s_k) u_k(s_k) ds_k. \tag{13}$$

For the exponent conditional density $f(x_k|s_k)$

$$\bigtriangledown_{x_k} f(x_k|s_k) = (\bigtriangledown_{x_k} \ln h(x_k) + \bigtriangledown_{x_k} T^{\mathrm{T}}(x_k) Q(s_k)) f(x_k|s_k). \tag{14}$$

By substituting (14) in (13) and denoting by $Q(\widehat{s}_k)$ the quantity $\int Q(s_k)\pi_k(s_k|x_1^n)ds_k$, we find the equation for the optimal mean-square estimator $Q(\widehat{s}_k)$:

$$\mathcal{T}^{\mathrm{T}}(x_k)Q(\widehat{s}_k) = \bigtriangledown_{x_k} \ln \frac{f(x_k|x_1^n \text{ without } x_k)}{h(x_k)}, \tag{15}$$

where $\mathcal{T}$ is the Jacobi matrix with elements $\partial T_i/\partial x_k^{[j]}, i = \overline{1,m}, j = \overline{1,r}$.

It is obvious that in the equation for the interpolating estimator the conditional density of observation $f(x_k|x_1^n$ without $x_k)$ is taken given all available observations on both sides of k in the future and in the past. The equation (15) is a simple linear vector equation with respect to $Q(\hat{s}_k)$, but it can be solved only under a certain density $f(x_k|x_1^n$ without $x_k)$. In the classical case, when all the distributions are known, this density can be calculated and the result will coincide with (1) and (2). But the case at hand where $f(x_k|x_1^n$ without $x_k)$ cannot be explicitly calculated and its parametric form is unknown, we may restore it from the observations using the kernel nonparametric procedures.

4 Nonparametric counterpart for interpolation estimate equation

In order to solve the problem of interpolating on the basis of one realization x_1^n of a process $(X_k)_{1\leqslant k\leqslant n}, X_k \in \mathbb{R}^l$, one may proceed to the asymptotically ε-optimal interpolating procedure [5], in which the truncated conditional density $\overline{f}(x_k|x_{k-\tau}^{k-1}, x_{k+1}^{k+\tau})$ is examined instead of the conditional density $f(x_k|x_1^n$ without $x_k) \triangleq f(x_k|x_1^{k-1}, x_{k+1}^n)$, where the parameter τ defines the connectivity of the Markov process which approximates non-Markovian process $(X_n)_{n\geqslant 1}$ with weak dependence. The criteria and methods of seeking τ which were developed in [5] with regard to filtration, can be extended in full to the interpolation problems. In doing so, the conditional density $\overline{f}(x_k|x_{k-\tau}^{k-1}, x_{k+1}^{k+\tau})$ can be written as the ratio

$$\overline{f}(x_k|x_{k-\tau}^{k-1}, x_{k+1}^{k+\tau}) = \frac{f(x_{k-\tau}^{k+\tau})}{f(x_{k-\tau}^{k-1}, x_{k+1}^{k+\tau})},$$

where the numerator is the marginal density of $l\times(2\tau+1)$-dimensional vector and the denominator is marginal density of the $l\times 2\tau$-dimensional vector of observations. By substituting the multivariate nonparametric kernel estimators

$$f_N(x_1^n) = \frac{1}{Nh_N^{nr}} \sum_{i=1}^{N} \prod_{k=1}^{n} \prod_{j=1}^{r} K\left(\frac{(x_k^{[j]} - X_k^{[j]}(i))}{h_N}\right) \tag{16}$$

for these densities, we get a nonparametric counterpart of equation (15) in the form

$$\mathcal{T}^{\mathrm{T}}(x_k)Q({}^{\tau}\widehat{s}_{k,N}) = \frac{\nabla_{x_k} f_N(x_{k-\tau}^{k+\tau})}{f_N(x_{k-\tau}^{k+\tau})} - \frac{\nabla_{x_k} h(x_k)}{h(x_k)}. \tag{17}$$

Interpretation of this equation is quite obvious. To construct the interpolating estimators at the point k, one uses the data which stand before and later of k on the distance, that do not exceed τ. For a greater τ, that is, a larger volume of data, calculations of the estimator are more difficult, but at the same time are closer to the optimal value obtained on the basis of the full data set.

The interpolation estimate from equation (17) is expressed in terms of the logarithmic gradient of the conditional density which is an unstable functional that may have an infinite value. Therefore, the nonparametric interpolation estimate will be only consistent. For a stronger convergence, one should construct a piecewise smooth approximation [5] which under some regularity conditions provides the mean-square convergence with the rate $N^{-\frac{2\nu}{2\nu+(2\tau+1)l}}$, where ν is the order of the least other than zero absolute moment of the kernel function and l is the dimension of observations vector.

5 Comparison of nonparametric interpolation estimate with optimal estimates in Kalman scheme

We consider for the purposes of illustration an example with a univariate state and observation models

$$S_{n+1} = aS_n + b\xi_{n+1}, \quad b^2 = \sigma^2(1-a^2), \tag{18}$$

$$X_n = AS_n + B\eta_n, \quad S_n,\, X_n \in \mathbb{R}. \tag{19}$$

Here, S_1, ξ_n and η_n are the mutually independent random variables with distributions $\mathsf{N}\{0,\, \sigma^2\}$ for S_1 and $\mathsf{N}\{0,\, 1\}$ for ξ_n and η_n, $n \geqslant 1$, and the coefficients a, b, A, B are given with $|a| < 1$. Such equations generate a stationary process. For them the conditional density of observations is Gaussian and, therefore, the Kalman filter and the forward and backward recursive linear interpolation equations associated with it can be obtained [4, p. 507].

The Kalman filter.

$$\begin{aligned} \widehat{S}_{k+1} &= a\widehat{S}_k + \frac{Ab^2 + a^2A\gamma_k}{B^2 + A^2b^2 + A^2a^2\gamma_k}[x_{k+1} - Aa\widehat{S}_k], \\ \gamma_{k+1} &= \frac{B^2(a^2\gamma_k + b^2)}{A^2(a^2\gamma_k + b^2) + B^2} \end{aligned} \tag{20}$$

with the initial conditions

$$\widehat{S}_1 = \frac{A\sigma^2}{A^2\sigma^2 + B^2}x_1, \quad \gamma_1 = \frac{B^2\sigma^2}{A^2\sigma^2 + B^2},$$

where $\widehat{S}_k = \mathsf{E}[S_k|x_1^k]$, $\gamma_k = \mathsf{E}[(S_k - \widehat{S}_k)^2|x_1^k]$.

Forward interpolation.

$$\begin{aligned} D_k &= A^2(a^2\gamma_k + \sigma^2(1-a^2)) + B^2 \\ \widetilde{S}_k &= \widehat{S}_k + Aa\gamma_k(X_{k+1} - Aa\widehat{S}_k)/D_k \\ \widetilde{\gamma}_k &= A^2\sigma^2(1-a^2) + B^2)\gamma_k/D_k, \quad k = 2, ..., n, \end{aligned} \tag{21}$$

where $\widehat{S}_k = \mathsf{E}[S_k|x_1^{k+1}]$, $\widetilde{\gamma}_k = \mathsf{E}[(S_k - \widetilde{S}_k)^2|x_1^{k+1}]$.

Backward interpolation.

$$\begin{aligned} \tilde{\widetilde{S}}_k &= \widetilde{S}_k + \tilde{\widetilde{\gamma}}_k a\sigma^2(1-a^2)(\tilde{\widetilde{S}}_{k+1} - \widehat{S}_k))/d_k\gamma_{k+1} \\ \tilde{\widetilde{\gamma}}_k &= \widetilde{\gamma}_k^2(\sigma^2(1-a^2))^2\tilde{\widetilde{\gamma}}_{k+1}/D_k^2\gamma_{k+1}, \quad k = 2, ..., n-1, \end{aligned} \tag{22}$$

where $\tilde{\widetilde{S}}_k = \mathsf{E}[S_k|x_k^n]$, $\tilde{\widetilde{\gamma}}_k = \mathsf{E}[(S_k - \tilde{\widetilde{S}}_k)^2|x_k^n]$.

A nonparametric interpolation can be constructed using only one observation equation (19) and a set of data of the length n. For that, we use the nonparametric density estimate (16) with Gaussian kernel function $K(\cdot)$. In this case, the nonparametric interpolation equation (15) comes to the following equation:

$$\widehat{S}_k^\tau = \frac{B^2}{A}\frac{d/dx_k f(x_{k-\tau}^{k+\tau})}{f(x_{k-\tau}^{k+\tau})} + \frac{x_k}{A}. \tag{23}$$

It should be noted that this equation does not involve the state-space equation parameters (18). The nonparametric ratio estimation of density derivative to density itself is described by the following expression:

$$\frac{(d/dx_k)f_N(x_{k-\tau}^{k+\tau})}{f_N(x_{k-\tau}^{k+\tau})} = \frac{\alpha_1}{\alpha_2} n^{\frac{4\tau}{(2\tau+5)(2\tau+15)}} \frac{\sum\limits_{i=\tau+1}^{n-\tau}(x_i - x_k)\prod\limits_{l=\tau}^{\tau}\exp\left(-\frac{(x_{k+l} - x_{i+l})^2}{2h'^2}\right)}{\sum\limits_{i=\tau+1}^{n-\tau}\prod\limits_{l=\tau}^{\tau}\exp\left(-\frac{(x_{k+l} - x_{i+l})^2}{2h^2}\right)}, \tag{24}$$

where $N = n - 2\tau$, and α_1, α_2 are the initial values of the bandwidth parameters whose optimal values depend on unknown functions. In the course of experiment these parameters are modified to get a good value of performance (risk, in case under study). The results of modeling are represented in Fig.1 given n=1000, $\sigma^2 = 2$, a=0.9, A=B=1, $\alpha_1 = 1.9, \alpha_1 = 1.2, \tau = 3$. The relative errors in percentage for each methods with regard to the risk of optimal smoothing are compiled in Table 1. The experiment shows that a quality of nonparametric estimators may be superior even to the Kalman filtering estimates, but is always inferior to the optimal backward interpolation. In spite of existence of some practical selection methods of the initial values α_1, α_2 of nonparametric estimates, the way of completely automatic selection still remains open.

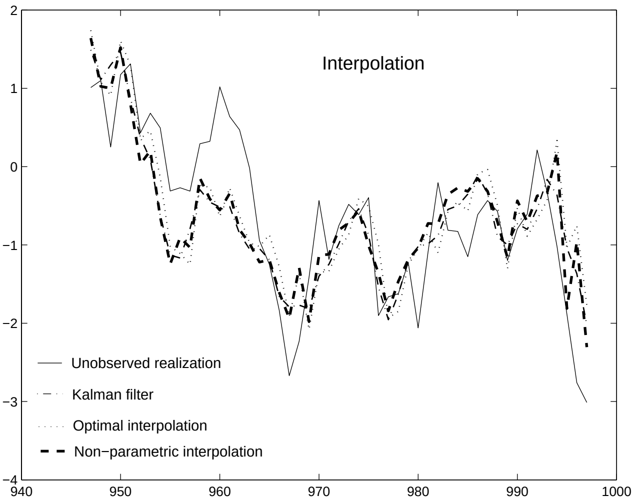

Fig. 1. Comparison of nonparametric smoothing estimate with optimal estimates.

Table 1:
Relative excess of the empirical estimate risk over the optimal smoothing risk

N	Optimal	Kalman	Nonpar
50	0%	13.8%	9.6%

References

[1] Briers M., Douset A., Maskell S.(2003). Smoothing Algorithms for State-Space Models. *IEEE Transections on signal processing*, Vol. , No. ,pp. 1–21.

[2] Chentsov N.N. (1972). *Statistical decision rules and optimal inference*. Moscow: Nauka (in Russian).

[3] Khazen E.M. (1978). Restoration of the component of a multivariate Markov process under the observations of other its components. *Probl. Control Inform. Theory*, **7 (4)** , pp. 263–275.

[4] Liptser R.Sh., Shiryaev A.N. *Statistics of random processes. I: General theory*. N. Y.: Springer-Verlag, (1977). 2: *Applications*. N. Y.: Springer-Verlag, (1978).

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