On evaluation of discrete states of hidden Markov chain in condition of uncertainty

V.O. Vasilyev

Moscow Institute of Physics and Technology (State University),
Institute of Control Sciences, Russian Academy of Sciences

evil.vasy@gmail.com

VI Traditional School ”Control, Information and Science”,
22-29 June 2014

System model

- Let (ϑ_n, X_n) be a two-component process where (ϑ_n) is unobservable process, (X_n) is observable one, $n \in \overline{1, N}$, $N \in \mathbb{N}$.
- Let (ϑ_n) be a stationary Markov chain with M discrete states and transition matrix $\|p_{i,j}\|$, $p_{i,j} = \mathrm{P}\{\vartheta_n = j \mid \vartheta_{n-1} = i\}$.
- The process (X_n) is described by AR of order p:

$$X_n = \mu(\vartheta_n) + \sum_{i=1}^{p} a_i(\vartheta_n)(X_{n-i} - \mu(\vartheta_n)) + b(\vartheta_n)\xi_n, \qquad (1)$$

where $\{\xi_n\}$ are i.i.d. random variables with a $\mathcal{N}(0, 1)$, $\mu, a_i, b \in \mathbb{R}$ are coefficients controlled by the process (ϑ_n).

System model

- As a quality measure for the proposed methods we use mean risk with a simple loss function L:

$$L(\vartheta_n, \hat{\vartheta}_n) = \begin{cases} 1, & \vartheta_n \neq \hat{\vartheta}_n, \\ 0, & \vartheta_n = \hat{\vartheta}_n, \end{cases}$$

  where $\hat{\vartheta}_n$ is an estimator of ϑ_n and $X_1^n = X_1, X_2, \ldots, X_n$.
- For the mean risk with the $L(\vartheta_n, \hat{\vartheta}_n)$ the optimal estimator is

$$\hat{\vartheta}_n = \arg\max_m \mathrm{P}\{\vartheta_n = m \mid X_1^n\}, \qquad (2)$$

  where $P\{\vartheta_n = m \mid X_1^n\}$ is a posterior probability with respect to a σ-algebra, generated by r.v. X_1^n. Its realization will be denoted by

$$\mathrm{P}\{\vartheta_n = m \mid X_1^n = x_1^n\} = w_n(m \mid x_1^n) = w_n(m).$$

Optimal filtering

- **We estimate the value of ϑ_n using observable values x_1^n.**
- Probability $w_n(m)$ satisfies the recurrent Stratonovich's equation [1]

$$w_n(m) = \frac{f_m(x_n)}{f(x_n \mid x_1^{n-1})} \sum_{j=1}^{M} p_{j,m} w_{n-1}(j), \tag{3}$$

$$f_{m,n} = f_m(x_n) = f(x_n \mid x_{n-p}^{n-1}, \vartheta_n = m).$$

- Using (1) the conditional density function $f_m(x_n)$ is calculated as:

$$f_m(x_n) = \mathcal{N}(x_n, \mu(m) + \sum_{i=1}^{p} a_i(m)(x_{n-i} - \mu(m)), b^2(m))$$

- $f(x_n \mid x_1^{n-1})$ is obtained by summing of (3) over m:

$$f(x_n \mid x_1^{n-1}) = \sum_{m=1}^{M} \left(f_m(x_n) \sum_{j=1}^{M} p_{j,m} w_{n-1}(j) \right).$$

Optimal filtering

- If transition matrix $\|p_{i,j}\|$ is known, then all elements in (3) are defined and the solution of problem is found, else we need to overcome this uncertainty.

Non-parametric filtering

- In this method we overcome uncertainty in the $\|p_{i,j}\|$.
- We assume that process (ϑ_n, X_n) is α-mixing, then instead of $f(x_n \mid x_1^{n-1})$ we use "truncated" $f(x_n \mid x_{n-\tau}^{n-1})$, $\tau \in \overline{1, n-1}$:

$$w_n(m) \approx \frac{f_m(x_n) f(x_{n-\tau}^{n-1})}{f(x_{n-\tau}^{n})} \sum_{j=1}^{M} p_{j,m} w_{n-1}(j),$$

$$w_n(m) \approx \frac{f_m(x_n)}{f(x_{n-\tau}^{n})} u_n(m), \tag{4}$$

where the introduced variable $u_n(m)$ does not depend on x_n and includes unknown $p_{i,j}$.

- Then we sum of (4) over m and get:

$$f(x_{n-\tau}^{n}) \approx \sum_{m=1}^{M} f_m(x_n) u_n(m). \tag{5}$$

- To calculate $u_n(m)$ it is necessary to find yet $M-1$ equations, which we get by differentiating and integrating (5) with respect to x_n.

- Therefore obtained system of equations in matrix form is

$$\mathbf{F_n} u_n = b_n, \tag{6}$$

where

$$\mathbf{F_n} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ f_{1,n} & f_{2,n} & \cdots & f_{M,n} \\ f_{1,n}^{(1)} & f_{2,n}^{(1)} & \cdots & f_{M,n}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ f_{1,n}^{(M-2)} & f_{2,n}^{(M-2)} & \cdots & f_{M,n}^{(M-2)} \end{pmatrix}, \tag{7}$$

$$u_n = \begin{pmatrix} u_n(1) & u_n(2) & u_n(3) & \ldots & u_n(M) \end{pmatrix}^T,$$

$$b_n = \begin{pmatrix} f(x_{n-\tau}^{n-1}) & f^{(0)}(x_{n-\tau}^{n}) & \ldots & f^{(M-2)}(x_{n-\tau}^{n}) \end{pmatrix}^T.$$

- We estimate unknown elements of b_n by using multivariate kernel density estimation [2, 3].

Note

- For the process with $M = 2$:

$$\mathbf{F_n} = \begin{pmatrix} 1 & 1 \\ f_{1,n} & f_{2,n} \end{pmatrix}.$$

- If $f_{1,n} \approx f_{2,n}$ then $\mathbf{F_n}$ is almost singular.

- The latter system of equations for the process with $M = 2$ is the following

$$\begin{cases} u_n(1) + u_n(2) = f(x_{n-\tau}^{n-1}) \\ f_{1,n}u_n(1) + f_{2,n}u_n(2) = f(x_{n-\tau}^{n}) \end{cases}.$$

We add another equation

$$f'_{1,n}u_n(1) + f'_{2,n}u_n(2) = f'(x_{n-\tau}^{n}).$$

- To find u_n we can solve one of the 3 systems

$$\mathbf{F}_{n,1}u_n = b_{n,1} : \quad \begin{pmatrix} 1 & 1 \\ f_{1,n} & f_{2,n} \end{pmatrix} \cdot \begin{pmatrix} u_n(1) \\ u_n(2) \end{pmatrix} = \begin{pmatrix} f(x_{n-\tau}^{n-1}) \\ f(x_{n-\tau}^{n}) \end{pmatrix},$$

$$\mathbf{F}_{n,2}u_n = b_{n,2} : \quad \begin{pmatrix} f_{1,n} & f_{2,n} \\ f'_{1,n} & f'_{2,n} \end{pmatrix} \cdot \begin{pmatrix} u_n(1) \\ u_n(2) \end{pmatrix} = \begin{pmatrix} f(x_{n-\tau}^{n}) \\ f'(x_{n-\tau}^{n}) \end{pmatrix},$$

$$\mathbf{F}_{n,3}u_n = b_{n,3} : \quad \begin{pmatrix} 1 & 1 \\ f'_{1,n} & f'_{2,n} \end{pmatrix} \cdot \begin{pmatrix} u_n(1) \\ u_n(2) \end{pmatrix} = \begin{pmatrix} f(x_{n-\tau}^{n-1}) \\ f'(x_{n-\tau}^{n}) \end{pmatrix}.$$

- We choose the system with the best-conditioned matrix for each u_n:

$$I_n = \arg \min_{i=1,2,3} \kappa(F_{n,i}),$$

where $\kappa(A) = \|A\| \cdot \|A^{-1}\|$ is a condition number, $\|A\| = \sup\limits_{x \neq 0} \frac{\|A\bar{x}\|_2}{\|\bar{x}\|_2}$.

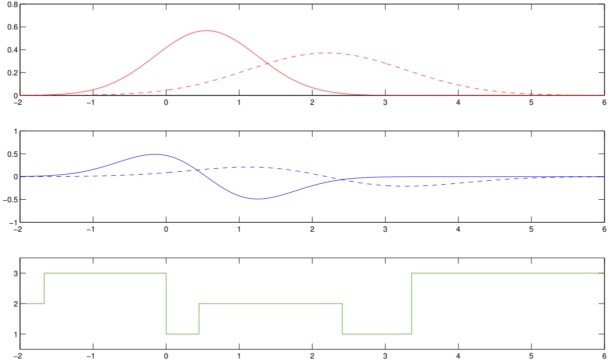

Figure 1: $f_{1,n}$(solid), $f_{2,n}$(dashed); $f'_{1,n}$(solid), $f'_{2,n}$(dashed); I_n(solid).

Optimal interpolation

- **In interpolation we estimate the value of unobservable ϑ_n using observable values x_1^N.** For posterior probability

$$\pi_n(m) = \pi_n(m \mid x_1^N) = \mathrm{P}\{\vartheta_n = m \mid X_1^N = x_1^N\}$$

we get

$$\pi_n(m) = \frac{f(x_1^{n-1}) f(x_{n+1}^N)}{f(x_1^N)} \cdot \frac{f_m(x_n) u_n(m) v_n(m)}{p_n(m)}, \qquad (8)$$

where

$$u_n(m) = \sum_{j=1}^{M} p_{j,m} w_{n-1}(j), \quad v_n(m) = \sum_{j=1}^{M} p_{j,m}^{+} \widetilde{w}_{n+1}(j),$$

$$p_{j,m}^{+} = \mathrm{P}\{\vartheta_n = m \mid \vartheta_{n+1} = j\}, \quad p_n(m) = \mathrm{P}\{\vartheta_n = m\}.$$

- Filtering probability w_n is defined in (3) and $\widetilde{w}_n$ comply with

$$\widetilde{w}_n(m) = \frac{f_m(x_n)}{f(x_n \mid x_{n+1}^N)} \sum_{j=1}^{M} p_{j,m}^{+} \widehat{w}_{n+1}(j).$$

Non-parametric interpolation

- As in non-parametric filtering we overcome the uncertainty in $\|p_{i,j}\|$ by introducing new variable $z_n(m)$ that does not depend on x_n and includes unknown $p_{i,j}$:

$$\pi_n(m) \approx \frac{f_m(x_n) z_n(m)}{f(x_{n-\tau}^{n+\tau})}.$$

- The obtained system of equations for $z_n(m)$ is

$$\mathbf{F}_n z_n = c_n,$$

where $\mathbf{F}_n$ is defined in (7) and

$$z_n = \begin{pmatrix} z_n(1) & z_n(2) & z_n(3) & \dots & z_n(M) \end{pmatrix}^T,$$
$$c_n = \begin{pmatrix} f(x_{n-\tau}^{n-1}, x_{n+1}^{n+\tau}) & f^{(0)}(x_{n-\tau}^{n+\tau}) & \dots & f^{(M-2)}(x_{n-\tau}^{n+\tau}) \end{pmatrix}^T.$$

- Using the multivariate kernel density estimation we estimate the unknown elements of c_n.

- The method is similar to the adaptive non-parametric filtering. We propose to choose the system with the best-conditioned matrix for each z_n.

Example

- Let the Markov chain (ϑ_n) have two states $(M = 2)$. The order p of AR is 1, coefficients $\mu \in \{0, 1.5\}$, $a_1 \in \{0.3, 0.2\}$, $b \in \{0.9539, 0.9797\}$ and the transition matrix

$$\|p_{i,j}\| = \begin{pmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{pmatrix}.$$

- We generate 700 random variates. Results are presented in Fig. 2. Sample mean errors after many efforts:

	Filtering	Interpolation
Optimal	0.0838	0.0738
Non-parametric	0.1510	0.1454
Adaptive non-param.	0.1338	0.1090

- **Note**: We choose coefficients a_1, μ manually; values of b were obtained automatically with Yule-Walker [4] equations so that variance for AR(p) with $\vartheta_n = 1$ is 1 and with $\vartheta_n = 2$ is 1.

Example

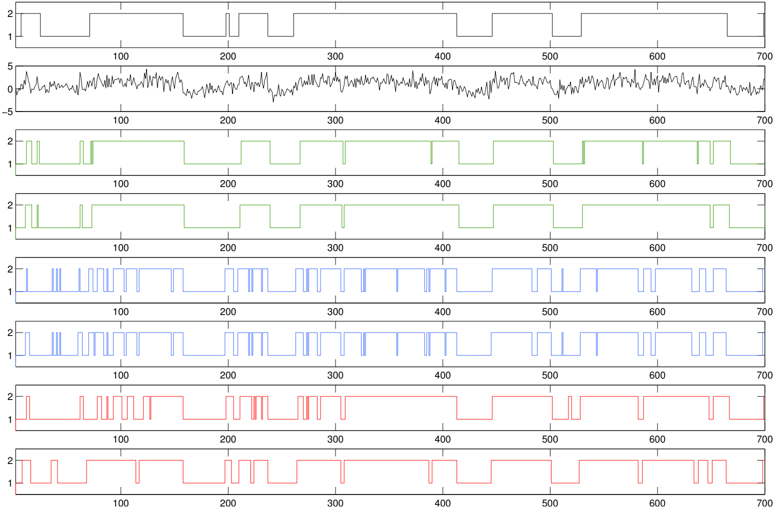

Figure 2: unobservable ϑ_n, observable X_n; optimal filter. and interp.; non-parametric filter. and interp.; adaptive non-param. filter. and interp.

Bibliography

A.V. Dobrovidov, G.M. Koshkin and V.A. Vasiliev (2012) *Non-parametric models and statistical inference from dependent observations*. USA.: Kendrick Press

Chacon, J. E.; Duon, T.; and Wand, M. P. (2009), *Asymptotics for General Multivariate Kernel Density Derivative Estimators*, Centre for Statistical and Survey Methodology, University of Wollongong, Working Paper 08-09, 2009, 30p.

Berwin A. Turlach (1993) *Bandwidth Selection in Kernel Density Estimation: A Review* In CORE and Institut de Statistique, pp. 23-493

Box, George and Jenkins, Gwilym (1970) *Time series analysis: Forecasting and control*, San Francisco: Holden-Day

Thank you for your attention :)