Kernel density estimation in adaptive tracking

Bernard Bercu and Bruno Portier

Universities of Bordeaux and Rouen, France

47th IEEE Conference on Decision and Control

Cancun, Mexico, December 10, 2008

Outline

Outline

Goal

Let (X_n) be a sequence of **iid** random variables with **unknown density function f**.

Goal

Estimate the density function f by a **kernel density estimator**.

Let K be a **nonnegative, bounded, Lipschitz** function called **Kernel**, such that

$$\int_{\mathbb{R}} K(x)\,dx = 1, \qquad \int_{\mathbb{R}} xK(x)\,dx = 0,$$

$$\int_{\mathbb{R}} K^2(x)\,dx = \tau^2.$$

Choices for the kernel

- **Uniform kernel**

$$K_a(x) = \begin{cases} \dfrac{1}{2a} & \text{if } |x| \leqslant a, \\ 0 & \text{otherwise.} \end{cases}$$

- **Epanechnikov kernel**

$$K_b(x) = \begin{cases} \dfrac{3}{4b}\left(1 - \dfrac{x^2}{b^2}\right) & \text{if } |x| \leqslant b, \\ 0 & \text{otherwise.} \end{cases}$$

- **Gaussian kernel**

$$K_c(x) = \frac{1}{c\sqrt{2\pi}} \exp\left(-\frac{x^2}{2c^2}\right).$$

Choices for the kernel

- **Uniform kernel**

$$K_a(x) = \begin{cases} \dfrac{1}{2a} & \text{if } |x| \leqslant a, \\ 0 & \text{otherwise.} \end{cases}$$

- **Epanechnikov kernel**

$$K_b(x) = \begin{cases} \dfrac{3}{4b}\left(1 - \dfrac{x^2}{b^2}\right) & \text{if } |x| \leqslant b, \\ 0 & \text{otherwise.} \end{cases}$$

- **Gaussian kernel**

$$K_c(x) = \frac{1}{c\sqrt{2\pi}} \exp\left(-\frac{x^2}{2c^2}\right).$$

Choices for the kernel

- **Uniform kernel**

$$K_a(x) = \begin{cases} \dfrac{1}{2a} & \text{if } |x| \leqslant a, \\ 0 & \text{otherwise.} \end{cases}$$

- **Epanechnikov kernel**

$$K_b(x) = \begin{cases} \dfrac{3}{4b}\left(1 - \dfrac{x^2}{b^2}\right) & \text{if } |x| \leqslant b, \\ 0 & \text{otherwise.} \end{cases}$$

- **Gaussian kernel**

$$K_c(x) = \frac{1}{c\sqrt{2\pi}} \exp\left(-\frac{x^2}{2c^2}\right).$$

Parzen-Rosenblatt or Wolverton-Wagner

Let (h_n) be a sequence of positive real numbers decreasing to zero called **bandwidth**. We can estimate the density f by the **Parzen-Rosenblatt** estimator given for all $x \in \mathbb{R}$ by

$$\widetilde{f}_n(x) = \frac{1}{nh_h}\sum_{i=1}^{n} K\Big(\frac{X_i - x}{h_n}\Big).$$

We can also estimate the density f by the **Wolverton-Wagner** estimator given for all $x \in \mathbb{R}$ by

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_i} K\left(\frac{X_i - x}{h_i}\right).$$

On Wolverton-Wagner

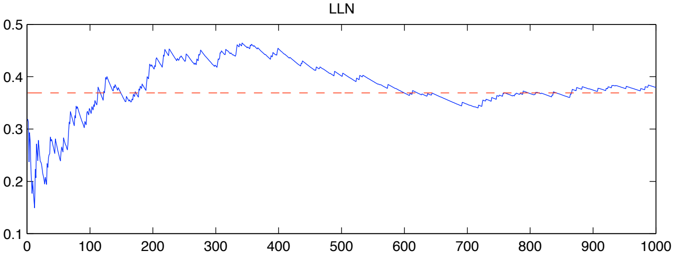
Theorem

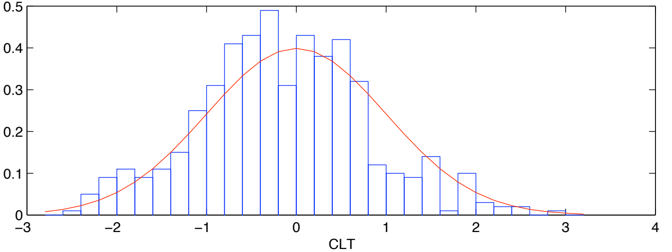
Assume that f is derivable with bounded derivative. If the bandwidth $h_n = 1/n^{\alpha}$ *with* $0 < \alpha < 1$*, we have*

(LLN) $$\lim_{n \to \infty} \widehat{f}_n(x) = f(x) \quad a.s.$$

In addition, if $1/5 < \alpha < 1$*, we also have*

(CLT) $$\sqrt{nh_n}(\widehat{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\tau^2 f(x)}{1+\alpha}\right).$$


LLN
0.5
0.4
0.3
0.2
0.1
0
100
200
300
400
500
600
700
800
900
1000


0.5
0.4
0.3
0.2
0.1
0
−3
−2
−1
0
1
2
3
4
CLT

Outline

Consider the d-dimensional **ARMAX(p,q,r)** model given by

$$A(R)X_n = B(R)U_n + C(R)\varepsilon_n$$

where R is the shift-back operator, X_n is the **system output**, U_n is the **system input** and ε_n is the **driven noise**,

- $A(R) = I_d - A_1 R - \cdots - A_p R^p$,
- $B(R) = B_1 R + B_2 R^2 + \cdots + B_q R^q$,
- $C(R) = I_d - C_1 R - \cdots - C_r R^r$

where A_i, B_j, and C_k are unknown matrices. We assume that the **high frequency gain** B_1 is known with $B_1 = I_d$.

Consider the d-dimensional **ARMAX(p,q,r)** model given by

$$A(R)X_n = B(R)U_n + C(R)\varepsilon_n$$

where R is the shift-back operator, X_n is the **system output**, U_n is the **system input** and ε_n is the **driven noise**,

- $A(R) = I_d - A_1 R - \cdots - A_p R^p$,
- $B(R) = B_1 R + B_2 R^2 + \cdots + B_q R^q$,
- $C(R) = I_d - C_1 R - \cdots - C_r R^r$

where A_i, B_j, and C_k are unknown matrices. We assume that the **high frequency gain** B_1 is known with $B_1 = I_d$.

Consider the d-dimensional **ARMAX(p,q,r)** model given by

$$A(R)X_n = B(R)U_n + C(R)\varepsilon_n$$

where R is the shift-back operator, X_n is the **system output**, U_n is the **system input** and ε_n is the **driven noise**,

- $A(R) = I_d - A_1 R - \cdots - A_p R^p$,
- $B(R) = B_1 R + B_2 R^2 + \cdots + B_q R^q$,
- $C(R) = I_d - C_1 R - \cdots - C_r R^r$

where A_i, B_j, and C_k are unknown matrices. We assume that the **high frequency gain** B_1 is known with $B_1 = I_d$.

The unknown parameter of the model is given by

$$\theta^t = (A_1, \dots, A_p, B_2, \dots, B_q, C_1, \dots, C_r).$$

The **ARMAX(p,q,r)** model can be rewritten as

$$\boldsymbol{X_{n+1} = \theta^t \Psi_n + U_n + \varepsilon_{n+1},}$$

where $\Psi_n = \left(X_n^p, U_n^q, \varepsilon_n^r\right)^t$ with

$$\begin{aligned} X_n^p &= (X_n^t, \dots, X_{n-p+1}^t), \\ U_n^q &= (U_{n-1}^t, \dots, U_{n-q+1}^t), \\ \varepsilon_n^r &= (\varepsilon_n^t, \dots, \varepsilon_{n-r+1}^t). \end{aligned}$$

Causality and Passivity

Definition

The matrix polynomial B is **causal** if for all $z \in \mathbb{C}$ with $|z| \leqslant 1$

$$\det(z^{-1}B(z)) \neq 0.$$

Definition

The matrix polynomial C is **passif** if for all $z \in \mathbb{C}$ with $|z| = 1$

$$\det(C(z)) \neq 0 \quad \text{and} \quad C^{-1}(z) > \frac{1}{2} I_d$$

Extended least squares

We estimate θ by the **extended least squares** estimator

$$\widehat{\theta}_{n+1} = \widehat{\theta}_n + S_n^{-1}\Phi_n(X_{n+1} - U_n - \widehat{\theta}_n^t\Phi_n)^t,$$

$$\widehat{\varepsilon}_{n+1} = X_{n+1} - U_n - \widehat{\theta}_n^t\Phi_n,$$

where the vector $\Phi_n = \left(X_n^p, U_n^q, \widehat{\varepsilon}_n^r\right)^t$ with $\widehat{\varepsilon}_n^r = (\widehat{\varepsilon}_n^t, \ldots, \widehat{\varepsilon}_{n-r+1}^t)$,

$$S_n = \sum_{i=0}^{n} \Phi_i \Phi_i^t + S,$$

where S is a positive definite and deterministic matrix.

Adaptive Control

The role played by U_n is to force X_n to track step by step a given trajectory (x_n). We make use of the **adaptive tracking control**

$$U_n = x_{n+1} - \widehat{\theta}_n^t \Phi_n.$$

Then, the closed-loop system is given by

$$X_{n+1} - x_{n+1} = \pi_n + \varepsilon_{n+1}$$

where the prediction error

$$\pi_n = (\theta - \widehat{\theta}_n)^t \Phi_n.$$

We assume that (ε_n) is a sequence of **iid** random vectors with **unknown density *f***. If (ε_n) were observable, we could estimate f by

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i^d} K\left(\frac{\varepsilon_i - x}{h_i}\right).$$

However, ε_{n+1} is unobservable but it can estimated by

$$\widehat{\varepsilon}_{n+1} = X_{n+1} - U_n - \widehat{\theta}_n^t \Phi_n = X_{n+1} - x_{n+1}.$$

Consequently, we can use the **Wolverton-Wagner** estimator

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i^d} K\left(\frac{X_i - x_i - x}{h_i}\right).$$

Outline

Uniform law of large numbers

Theorem

Assume that f is positive and differentiable with **bounded gradient** *and that* (ε_n) *has finite moment of order* > 2*. If the bandwidth* $\boldsymbol{h_n = 1/n^{\alpha}}$ *with* $\alpha \in]0, 1/d[$*, then*

(LLN) $$\lim_{n \to \infty} \sup_{x \in \mathbb{R}^d} |\widehat{f}_n(x) - f(x)| = 0 \quad a.s.$$

Central limit theorem

Theorem

Assume that f is positive and differentiable with **bounded gradient** *and that* (ε_n) *has finite moment of order* > 2. *If the bandwidth* $\boldsymbol{h_n = 1/n^{\alpha}}$ *with* $\alpha \in]1/(d+2), 1/d[$, *then*

$$G_n(x) = \sqrt{nh_n^d}(\widehat{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\tau^2 f(x)}{1+\alpha d}\right) = G(x).$$

In addition, for N distinct points $x_1, \ldots, x_N$ *of* $\mathbb{R}^d$, *we also have*

$$\text{(MCLT)} \quad \Big(G_n(x_1), \cdots, G_n(x_N)\Big) \xrightarrow{\mathcal{L}} \Big(G(x_1), \cdots, G(x_N)\Big)$$

where $G(x_1), \ldots, G(x_N)$ *are independent.*

Outline

goodness of fit test

We wish to test

$$\mathcal{H}_0 \ : \ \langle\langle\, f = f_0 \,\rangle\rangle \qquad \textbf{versus} \qquad \mathcal{H}_1 \ : \ \langle\langle\, f \neq f_0 \,\rangle\rangle$$

where f_0 is a given density function. Our statistical test is

$$T_n(N) = \frac{1}{\tau^2 \ell_h} \sum_{i=1}^{N} \frac{(\widehat{f}_n(x_i) - f_0(x_i))^2}{\widehat{f}_n(x_i)}$$

where $x_1, \ldots, x_N$ are N distinct points of $\mathbb{R}^d$ and

$$\ell_h = \frac{1}{1 + \alpha d}.$$

Theorem

Assume that f is positive and differentiable with **bounded gradient** *and that (ε_n) has finite moment of order > 2. If the bandwidth* $\boldsymbol{h_n = 1/n^{\alpha}}$ *with $\alpha \in]1/(d+2), 1/d[$, then under $\mathcal{H}_0$*

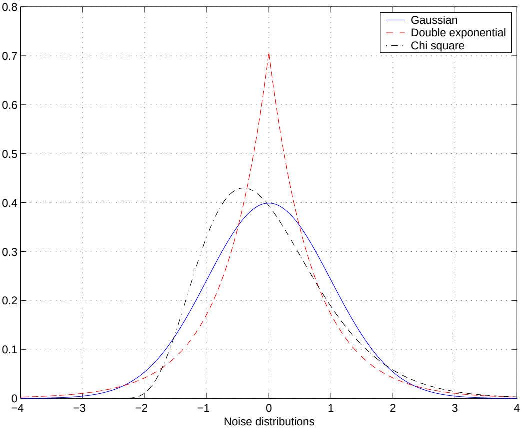
$$\boldsymbol{n h_n^d T_n(N) \xrightarrow{\mathcal{L}} \chi^2(N).}$$

In addition, under $\mathcal{H}_1$ and if one can find some point x of $\mathbb{R}^d$ in $\{x_1, x_2, \ldots, x_N\}$ such that $f(x) \neq f_0(x)$, then $T_n(N) \to \sigma^2$ a.s.

$$\boldsymbol{\sqrt{n h_n^d}(T_n(N) - \sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \lambda^2)}$$

$$\sigma^2 = \frac{1}{\tau^2 \ell_h} \sum_{i=1}^{N} \frac{(f(x_i) - f_0(x_i))^2}{f(x_i)}, \qquad \lambda^2 = \frac{1}{\tau^2 \ell_h} \sum_{i=1}^{N} \frac{(f^2(x_i) - f_0^2(x_i))^2}{f^3(x_i)}.$$

Simulations



Noise distributions

- **Gaussian**

$$f_0(x) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big),$$

- **Double exponential**

$$f_1(x) = \frac{1}{\sqrt{2}} \exp\Big(-\sqrt{2}|x|\Big),$$

- **Chi square**

$$f_2(x) = \begin{cases} \frac{9}{5}(x+\sqrt{6})^5 \exp\Big(-\sqrt{6}(x+\sqrt{6})\Big) & \text{if } |x| \geqslant -\sqrt{6}, \\ 0 & \text{otherwise.} \end{cases}$$

Noise distributions

- **Gaussian**

$$f_0(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right),$$

- **Double exponential**

$$f_1(x) = \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}|x|\right),$$

- **Chi square**

$$f_2(x) = \begin{cases} \frac{9}{5}(x+\sqrt{6})^5 \exp\left(-\sqrt{6}(x+\sqrt{6})\right) & \text{if } |x| \geqslant -\sqrt{6}, \\ 0 & \text{otherwise.} \end{cases}$$

Noise distributions

- **Gaussian**

$$f_0(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right),$$

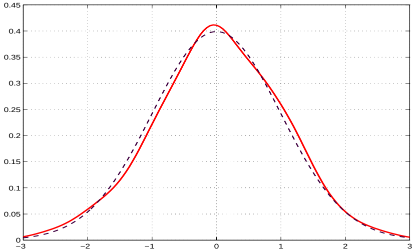
- **Double exponential**

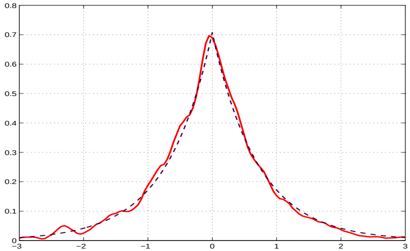
$$f_1(x) = \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}|x|\right),$$

- **Chi square**

$$f_2(x) = \begin{cases} \frac{9}{5}(x+\sqrt{6})^5 \exp\left(-\sqrt{6}(x+\sqrt{6})\right) & \text{if } |x| \geqslant -\sqrt{6}, \\ 0 & \text{otherwise.} \end{cases}$$

Law of large numbers





ARX Goodness of fit test

$$X_{n+1} = \theta X_n + U_n + \varepsilon_{n+1}$$

Table: Results under $\mathcal{H}_0$ and $\mathcal{H}_1$ with test level 5%.

	$n = 200, N = 8$				$n = 1000, N = 22$	
	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$
$\mathcal{G}f_0$	**3.8%**	35.7%	28%	**3.7%**	99.7%	98.2%
$\mathcal{G}f_1$	45.8%	**5.5%**	71.5%	100%	**5%**	100%
$\mathcal{G}f_2$	21.2%	54.5%	**3.2%**	96.7%	100%	**5.1%**

NARX Goodness of fit test

$$X_{n+1} = \theta X_n^2 + U_n + \varepsilon_{n+1}$$

Table: Results under $\mathcal{H}_0$ and $\mathcal{H}_1$ with test level 5%.

	$n = 200,\ N = 8$				$n = 1000,\ N = 22$	
	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$	$\mathcal{H}f_0$	$\mathcal{H}f_1$	$\mathcal{H}f_2$
$\mathcal{G}f_0$	**3%**	37.1%	28.5%	**4.3%**	99.5%	98.6%
$\mathcal{G}f_1$	44.6%	**5.2%**	72%	100%	**5.1%**	100%
$\mathcal{G}f_2$	19.8%	58.3%	**3.7%**	97.2%	100%	**5%**