Asymptotic behavior of stochastic algorithms with statistical applications
Part 3

Bernard Bercu

University of Bordeaux, France

ETICS Annual Research School, Fréjus, 2019

Outline

1. Parametric estimation of quantiles and superquantiles.
2. Nonparametric estimation of probability density functions.
3. Semiparametric estimation in shape invariant models.

Outline

Quantiles and superquantiles.

Let X be a **continuous** random variable with **unknown** distribution function F. Assume that F is **continuous and strictly increasing.**

Definition

For any α in $]0,1[$, the quantile of order α of X is the unique solution θ_α of the equation $F(x) = \alpha$,

$$F(\theta_\alpha) = \alpha.$$

If X is **integrable**, the superquantile of order α of X is defined by

$$\vartheta_\alpha = \mathbb{E}[X | X \geqslant \theta_\alpha] = \frac{\mathbb{E}[X \, \mathrm{I}_{\{X \geqslant \theta_\alpha\}}]}{\mathbb{P}(X \geqslant \theta_\alpha)} = \frac{1}{1-\alpha} \mathbb{E}[X \, \mathrm{I}_{\{X \geqslant \theta_\alpha\}}].$$

Quantiles and superquantiles.

Example (Exponential distribution)

If X has an Exponential $\mathcal{E}(\lambda)$ distribution with $\lambda > 0$,

$$\theta_\alpha = -\frac{1}{\lambda}\log(1-\alpha) \qquad \text{and} \qquad \vartheta_\alpha = \frac{1}{\lambda}(1 - \ln(1-\alpha)).$$

Example (Pareto distribution)

If X has a Pareto $\mathcal{P}(a, b)$ distribution with $a > 1$ and $b > 0$,

$$\theta_\alpha = b(1-\alpha)^{-1/a} \qquad \text{and} \qquad \vartheta_\alpha = \frac{ab}{a-1}(1-\alpha)^{-1/a}.$$

Goal

$\longrightarrow$ Recursively estime the quantiles and superquantiles $\boldsymbol{\theta_\alpha}$ **and** $\boldsymbol{\vartheta_\alpha}$.

Recursive estimation of quantiles and superquantiles.

We already saw that we can estimate θ_α by the **slow down Robbins-Monro algorithm** given by

$$\widehat{\theta}_{n+1} = \widehat{\theta}_n - \gamma_n \left(\mathrm{I}_{\{X_{n+1} \leqslant \widehat{\theta}_n\}} - \alpha \right)$$

where

$$\gamma_n = \frac{1}{n^c} \qquad \text{with} \qquad \frac{1}{2} < c < 1$$

and its averaging version

$$\overline{\theta}_n = \frac{1}{n} \sum_{k=1}^{n} \widehat{\theta}_k.$$

Recursive estimation of quantiles and superquantiles.

We can also estimate ϑ_α by

$$\overline{\vartheta}_n = \frac{1}{n(1-\alpha)} \sum_{k=1}^{n} X_k \mathrm{I}_{\{X_k \geqslant \overline{\theta}_{k-1}\}}.$$

Another strategy is to make use of the convex version

$$\begin{aligned} \widetilde{\vartheta}_n &= \frac{1}{n} \sum_{k=1}^{n} \Big(\overline{\theta}_{k-1} + \frac{1}{1-\alpha} (X_k - \overline{\theta}_{k-1}) \mathrm{I}_{\{X_k \geqslant \overline{\theta}_{k-1}\}} \Big), \\ &= \overline{\vartheta}_n + \frac{1}{n(1-\alpha)} \sum_{k=1}^{n} \overline{\theta}_{k-1} \Big(\mathrm{I}_{\{X_k \leqslant \overline{\theta}_{k-1}\}} - \alpha \Big). \end{aligned}$$

Recursive estimation of quantiles and superquantiles.

Assume that X is square integrable and let

$$\begin{aligned} G_\alpha(\theta) &= \frac{1}{(1-\alpha)}\mathbb{E}[X\,\mathrm{I}_{\{X \geqslant \theta\}}], \\ H_\alpha(\theta) &= \frac{1}{(1-\alpha)^2}\mathbb{E}[X^2\,\mathrm{I}_{\{X \geqslant \theta\}}], \\ \sigma_\alpha^2(\theta) &= \frac{1}{(1-\alpha)^2}\mathrm{Var}(X\,\mathrm{I}_{\{X \geqslant \theta\}}). \end{aligned}$$

Denote

$$Y_n = \frac{1}{(1-\alpha)} X_n\,\mathrm{I}_{\{X_n \geqslant \overline{\theta}_{n-1}\}}.$$

We clearly have

$$\mathbb{E}[Y_n | \mathcal{F}_{n-1}] = G_\alpha(\overline{\theta}_{n-1}) \quad \text{and} \quad \mathrm{Var}(Y_n | \mathcal{F}_{n-1}) = \sigma_\alpha^2(\overline{\theta}_{n-1}).$$

The martingale decomposition.

In addition,

$$\overline{\vartheta}_n = \frac{1}{n}\sum_{k=1}^{n} Y_k = \frac{1}{n}\sum_{k=1}^{n}(Y_k - \mathbb{E}[Y_k|\mathcal{F}_{k-1}]) + \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[Y_k|\mathcal{F}_{k-1}].$$

Recalling that $\vartheta_\alpha = G_\alpha(\theta_\alpha)$, we obtain the **martingale decomposition**

$$\overline{\vartheta}_n - \vartheta_\alpha = \frac{1}{n}M_n + \frac{1}{n}\sum_{k=1}^{n} G_\alpha(\overline{\theta}_{k-1}) - G_\alpha(\theta_\alpha)$$

where

$$M_n = \sum_{k=1}^{n}(Y_k - \mathbb{E}[Y_k|\mathcal{F}_{k-1}]).$$

The martingale decomposition.

Therefore,

$$M_n = \sum_{k=1}^{n} (Y_k - \mathbb{E}[Y_k|\mathcal{F}_{k-1}]),$$

$$<M>_n = \sum_{k=1}^{n} \text{Var}(Y_k|\mathcal{F}_{k-1}) = \sum_{k=1}^{n} \sigma_\alpha^2(\overline{\theta}_{k-1}).$$

It follows from the almost sure convergence of $\overline{\theta}_n$ to θ_α that (M_n) is a **square integrable martingale** satisfying

$$\lim_{n\to\infty} \frac{<M>_n}{n} = \sigma_\alpha^2(\theta_\alpha) \qquad \text{a.s.}$$

Recursive estimation of quantiles and superquantiles.

Theorem

If X is square integrable, we have the almost sure convergence

$$\lim_{n\to\infty} \begin{pmatrix} \overline{\theta}_n \\ \overline{\vartheta}_n \end{pmatrix} = \begin{pmatrix} \theta_\alpha \\ \vartheta_\alpha \end{pmatrix} \qquad \text{a.s.}$$

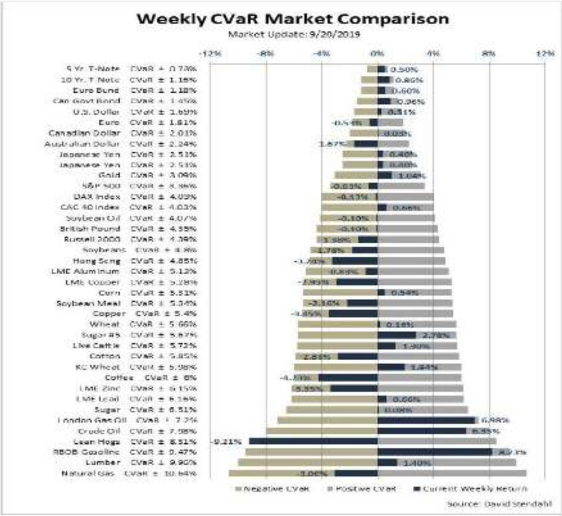
Moreover, we also have the joint asymptotic normality

$$\sqrt{n} \begin{pmatrix} \overline{\theta}_n - \theta_\alpha \\ \overline{\vartheta}_n - \vartheta_\alpha \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \Gamma_\alpha\Big)$$

where

$$\Gamma_\alpha = \begin{pmatrix} \frac{\alpha(1-\alpha)}{f^2(\theta_\alpha)} & \frac{\alpha}{f(\theta_\alpha)}(\vartheta_\alpha - \theta_\alpha) \\ \frac{\alpha}{f(\theta_\alpha)}(\vartheta_\alpha - \theta_\alpha) & \sigma_\alpha^2(\theta_\alpha) \end{pmatrix}.$$

Conditional value at risk in portfolio optimization.



Outline

1. Parametric estimation of quantiles and superquantiles.
2. Nonparametric estimation of probability density functions.
3. Semiparametric estimation in shape invariant models.

Recursive estimation of probability density functions.

Let (X_n) be a sequence of **iid** random variables with **unknown density function f**. Let K be a positive and bounded function, called **kernel**, such that

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

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

Goal

$\longrightarrow$ Recursively estimate the probability density function f.

Choice of the Kernel.

- **Uniform kernel**

$$K_a(x) = \frac{1}{2a} \mathrm{I}_{\{|x| \leqslant a\}},$$

- **Epanechnikov kernel**

$$K_b(x) = \frac{3}{4b} \left(1 - \frac{x^2}{b^2}\right) \mathrm{I}_{\{|x| \leqslant b\}},$$

- **Gaussian kernel**

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

Choice of the Kernel.

- **Uniform kernel**

$$K_a(x) = \frac{1}{2a}\mathrm{I}_{\{|x| \leqslant a\}},$$

- **Epanechnikov kernel**

$$K_b(x) = \frac{3}{4b}\left(1 - \frac{x^2}{b^2}\right)\mathrm{I}_{\{|x| \leqslant b\}},$$

- **Gaussian kernel**

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

Choice of the Kernel.

- **Uniform kernel**

$$K_a(x) = \frac{1}{2a} \mathrm{I}_{\{|x| \leqslant a\}},$$

- **Epanechnikov kernel**

$$K_b(x) = \frac{3}{4b} \left(1 - \frac{x^2}{b^2}\right) \mathrm{I}_{\{|x| \leqslant b\}},$$

- **Gaussian kernel**

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

The Wolverton-Wagner estimator.

We estimate the probability density function f by

The Wolverton-Wagner estimator

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{k=1}^{n} W_k(x)$$

where

$$W_n(x) = \frac{1}{h_n} K\Big(\frac{X_n - x}{h_n}\Big).$$

The **bandwidth** (h_n) is a sequence of positive real numbers, $h_n \searrow 0$, $nh_n \to \infty$. For $0 < \alpha < 1$, we can make use of

$$h_n = \frac{1}{n^\alpha}.$$

The martingale decomposition.

We have

$$\begin{aligned}\widehat{f}_n(x) - f(x) &= \frac{1}{n}\sum_{k=1}^{n} W_k(x) - f(x),\\ &= \frac{1}{n}\sum_{k=1}^{n}(W_k(x) - \mathbb{E}[W_k(x)]) + \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[W_k(x)] - f(x).\end{aligned}$$

Consequently,

$$\widehat{f}_n(x) - f(x) = \frac{1}{n}M_n(x) + \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[W_k(x)] - f(x)$$

where

$$M_n(x) = \sum_{k=1}^{n}(W_k(x) - \mathbb{E}[W_k(x)]).$$

The martingale decomposition.

Therefore,

$$\begin{aligned} M_n(x) &= \sum_{k=1}^{n} (W_k(x) - \mathbb{E}[W_k(x)]), \\ < M(x) >_n &= \sum_{k=1}^{n} \mathrm{Var}(W_k(x)). \end{aligned}$$

The sequence $(M_n(x))$ is a **square integrable martingale** such that

$$\lim_{n\to\infty} \frac{< M(x) >_n}{n^{1+\alpha}} = \ell \qquad \text{a.s.}$$

where

$$\ell = \frac{\xi^2 f(x)}{1+\alpha}.$$

Recursive estimation of probability density functions.

Theorem

For all $x \in \mathbb{R}$, we have the pointwise almost sure convergence

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

In addition, as soon as $1/5 < \alpha < 1$, *we have, for all $x \in \mathbb{R}$, the asymptotic normality*

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

Application to sea shores water quality.

Outline

Periodic shape invariant processes.

Consider the **shape invariant** process given, for all $n \geqslant 1$, by

$$\boldsymbol{Y_n = h(X_n) + \varepsilon_n}$$

where the function h is **periodic**

$$h(x) = m + \sum_{k=1}^{p} a_k f(x - \theta_k),$$

- The inputs (X_n) are random observation times,
- The outputs (Y_n) are the observations,
- The noises (ε_n) are unknown random errors.

Periodic shape invariant processes.

For the sake of simplicity, we focus our attention on the special case

$$Y_n = f(X_n - \theta) + \varepsilon_n$$

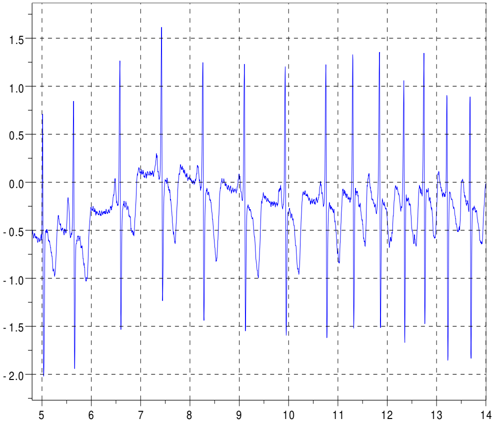
where (ε_n) is **iid** with mean zero and variance σ^2.

Goals

$\longrightarrow$ Recursively estimate the **shift parameter θ**,

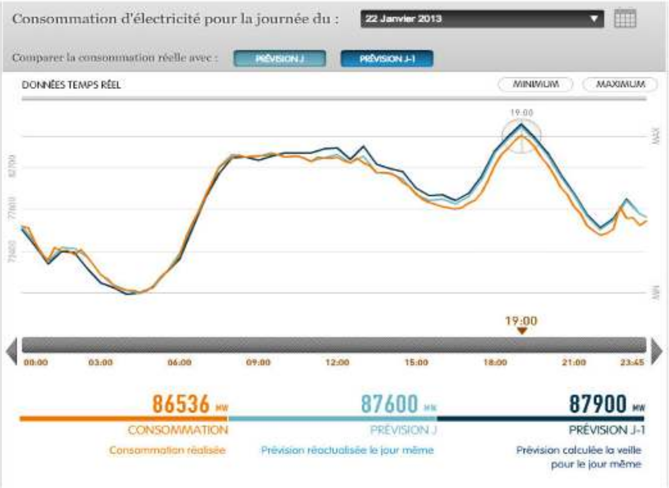
$\longrightarrow$ Recursively estimate the **shape function f**.

Detection of Atrial Fibrillation via ECG analysis.

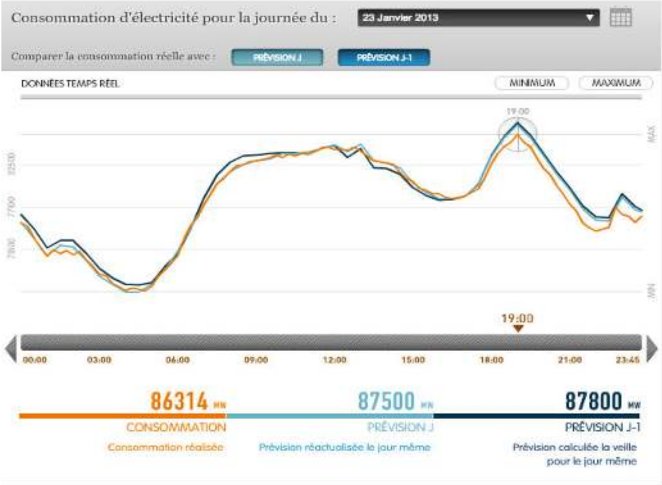


Eco2mix Forecast of electricity consumption.

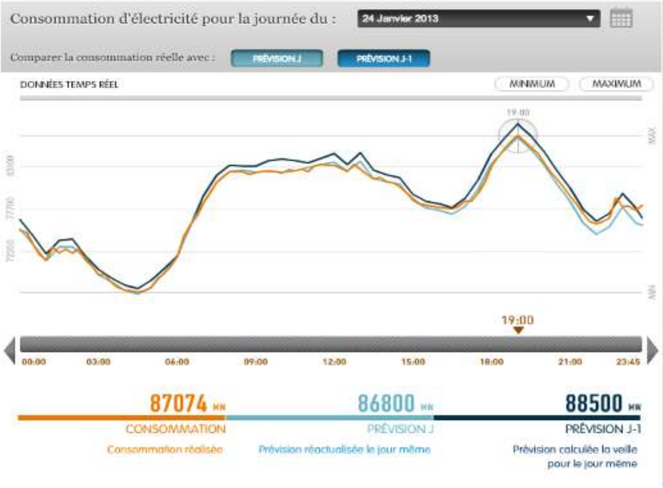
Eco2mix Forecast of electricity consumption.



Eco2mix Forecast of electricity consumption.



Eco2mix Forecast of electricity consumption.



Eco2mix Forecast of electricity consumption.

Hypothesis.

Symmetry and Periodicity

$(\mathcal{H}_1)$ The shape function f is **symmetric, bounded, periodic** with period 1.

Regularity of the density

$(\mathcal{H}_2)$ The observation times (X_n) are **iid** with density function g positive on $[-1/2, 1/2]$, continuous, twice differentiable with bounded derivatives.

A preliminary calculation.

Let X be a random variable sharing the same distribution as (X_n). We shall make use of the auxiliary function

$$\phi(t) = \mathbb{E}\Big[\frac{\sin(2\pi(X-t))}{g(X)} f(X-\theta)\Big].$$

It follows from the **periodicity and symmetry** of f that

$$\begin{aligned}\phi(t) &= \int_{-1/2}^{1/2} \sin(2\pi(x-t)) f(x-\theta)\,dx,\\ &= \int_{-1/2}^{1/2} \sin(2\pi(y+\theta-t)) f(y)\,dy,\\ &= \sin(2\pi(\theta-t)) \int_{-1/2}^{1/2} \cos(2\pi y) f(y)\,dy.\end{aligned}$$

A preliminary calculation.

Let X be a random variable sharing the same distribution as (X_n). We shall make use of the auxiliary function

$$\phi(t) = \mathbb{E}\Big[\frac{\sin(2\pi(X-t))}{g(X)} f(X-\theta)\Big].$$

It follows from the **periodicity and symmetry** of f that

$$\begin{aligned}\phi(t) &= \int_{-1/2}^{1/2} \sin(2\pi(x-t)) f(x-\theta)\, dx, \\ &= \int_{-1/2}^{1/2} \sin(2\pi(y+\theta-t)) f(y)\, dy, \\ &= \sin(2\pi(\theta-t)) \int_{-1/2}^{1/2} \cos(2\pi y) f(y)\, dy.\end{aligned}$$

A preliminary calculation.

Let X be a random variable sharing the same distribution as (X_n). We shall make use of the auxiliary function

$$\phi(t) = \mathbb{E}\Big[\frac{\sin(2\pi(X-t))}{g(X)} f(X-\theta)\Big].$$

It follows from the **periodicity and symmetry** of f that

$$\begin{aligned}
\phi(t) &= \int_{-1/2}^{1/2} \sin(2\pi(x-t)) f(x-\theta)\, dx, \\
&= \int_{-1/2}^{1/2} \sin(2\pi(y+\theta-t)) f(y)\, dy, \\
&= \sin(2\pi(\theta-t)) \int_{-1/2}^{1/2} \cos(2\pi y) f(y)\, dy.
\end{aligned}$$

A preliminary calculation.

Consequently, we obtain that

$$\phi(t) = f_1 \sin(2\pi(\theta - t))$$

where f_1 is the first Fourier coefficient of f

$$f_1 = \int_{-1/2}^{1/2} \cos(2\pi x) f(x)\, dx.$$

Obviously, ϕ is continuous and bounded function such that

$$\phi(\theta) = 0.$$

We assume in all the sequel that $f_1 > 0$. Then, for all $t \in \mathbb{R}$ such that $|t - \theta| < 1/2$, the product $(t - \theta)\phi(t) < 0$.

The Robbins-Monro procedure.

Let $K = [-1/4, 1/4]$ and denote by π_K the projection on K,

$$\pi_K(x) = \begin{cases} x & \text{if } |x| \leqslant 1/4, \\ 1/4 & \text{if } x \geqslant 1/4, \\ -1/4 & \text{if } x \leqslant -1/4. \end{cases}$$

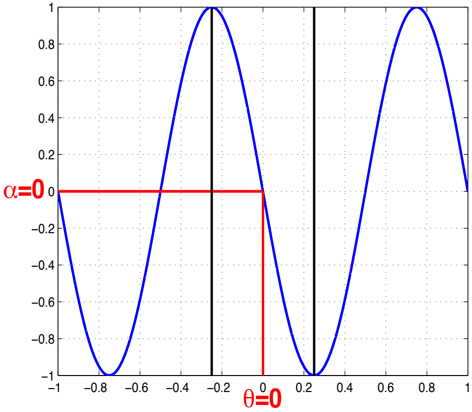
Let (γ_n) be a decreasing sequence of positive real numbers

$$\sum_{n=1}^{\infty} \gamma_n = +\infty \qquad \text{and} \qquad \sum_{n=1}^{\infty} \gamma_n^2 < +\infty.$$

For the sake of clarity, we shall make use of

$$\gamma_n = \frac{1}{n}.$$

Stochastic approximation.



The Robbins-Monro procedure.

We estimate θ by

The projected Robbins-Monro estimator

$$\widehat{\theta}_{n+1} = \pi_K\Big(\widehat{\theta}_n + \gamma_{n+1} T_{n+1}\Big),$$

where the initial value $\widehat{\theta}_0 \in K$ and

$$T_{n+1} = \frac{\sin(2\pi(X_{n+1} - \widehat{\theta}_n))}{g(X_{n+1})} Y_{n+1}.$$

$\longrightarrow$ One can observe that

$$\mathbb{E}[T_{n+1}|\mathcal{F}_n] = \phi(\widehat{\theta}_n) \qquad \text{a.s.}$$

Almost sure convergence.

Theorem

Assume that $(\mathcal{H}_1)$ *and* $(\mathcal{H}_2)$ *hold and that* $|\theta| < 1/4$*. Then,*

$$\lim_{n\to\infty} \widehat{\theta}_n = \theta \qquad a.s.$$

In addition, the **number of times** *that the random variable*

$$\widehat{\theta}_n + \gamma_{n+1} T_{n+1}$$

goes outside the compact K **is almost surely finite.**

Asymptotic normality.

In order to establish the **asymptotic normality** of $\widehat{\theta}_n$, it is necessary to introduce a second auxiliary function

$$\varphi(t) = \mathbb{E}\Big[\frac{\sin^2(2\pi(X-t))}{g^2(X)}(f^2(X-\theta)+\sigma^2)\Big],$$

$$= \int_{-1/2}^{1/2} \frac{\sin^2(2\pi(x-t))}{g(x)}(f^2(x-\theta)+\sigma^2)\,dx.$$

As soon as $\mathbf{4\pi f_1 > 1}$, denote

$$\xi^2(\theta) = \frac{\varphi(\theta)}{4\pi f_1 - 1}.$$

Asymptotic normality.

Theorem

Assume that $(\mathcal{H}_1)$ *and* $(\mathcal{H}_2)$ *hold and that* $|\theta| < 1/4$. *Moreover, suppose that* (ε_n) *has a finite moment of order* > 2 *and that* $4\pi f_1 > 1$. *Then, we have the asymptotic normality*

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \xi^2(\theta)).$$

$\longrightarrow$ If f_1 is known, we can replace γ_n by

$$\gamma_n = \frac{1}{2\pi n f_1}.$$

Then, $\widehat{\theta}_n$ is an **asymptotically efficient estimator** of θ,

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\varphi(\theta)}{4\pi^2 f_1^2}\right).$$

The symmetrized Nadaraya-Watson estimator.

We focus our attention on the estimation of the shape function f by

The symmetrized recursive Nadaraya-Watson estimator

$$\widehat{f}_n(x) = \frac{\sum_{k=1}^{n} (W_k(x) + W_k(-x)) Y_k}{\sum_{k=1}^{n} (W_k(x) + W_k(-x))},$$

where

$$W_n(x) = \frac{1}{h_n} K\Big(\frac{X_n - \widehat{\theta}_{n-1} - x}{h_n}\Big).$$

Almost sure convergence.

Lipschitzianity

$(\mathcal{H}_3)$ The shape function f is **Lipschitz.**

Theorem

Assume that $(\mathcal{H}_1)$, $(\mathcal{H}_2)$ *and* $(\mathcal{H}_3)$ *hold,* $|\theta| < 1/4$*, and that* (ε_n) *has a finite moment of order* > 2. *Then, for all* $|x| < 1/2$,

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

Asymptotic normality.

Theorem

Assume that $(\mathcal{H}_1)$, $(\mathcal{H}_2)$ *and* $(\mathcal{H}_3)$ *hold,* $|\theta| < 1/4$, *and that* (ε_n) *has a finite moment of order* > 2. *If* $1/3 < \alpha < 1$, *we have for all* $|x| < 1/2$ *with* $x \neq 0$, *the asymptotic normality*

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

In addition, for $x = 0$,

$$\sqrt{nh_n}\left(\widehat{f}_n(0) - f(0)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2\nu^2}{(1+\alpha)g(\theta)}\right).$$


!!!! Many thanks for your attention !!!!