Nonparametric regression estimation :

an empirical process approach to uniform in bandwidth consistency of kernel–type estimators and conditional U–statistics

Julia Dony

Free University of Brussels (VUB)

Vrije Universiteit Brussel

PhD thesis written under the supervision of

Uwe Einmahl (VUB) and David Mason (Univ. Delaware)

Public Defense

May 19, 2008

I. A short introduction

Nonparametric regression estimation

What is regression ?

regression = finding a relation between two "things" (variables)

Nonparametric regression estimation

What is regression ?

regression = finding a relation between two "things" (variables)

Examples :

- weight $\longleftrightarrow$ length

Nonparametric regression estimation

What is regression ?

regression = finding a relation between two "things" (variables)

Examples :

- weight $\longleftrightarrow$ length
- length $\longleftrightarrow$ nationality

Nonparametric regression estimation

What is regression ?

regression = finding a relation between two “things” (variables)

Examples :

- weight $\longleftrightarrow$ length
- length $\longleftrightarrow$ nationality
- life expectancy $\longleftrightarrow$ age of grandparents

Nonparametric regression estimation

What is regression ?

regression = finding a relation between two "things" (variables)

Examples :

- weight $\longleftrightarrow$ length
- length $\longleftrightarrow$ nationality
- life expectancy $\longleftrightarrow$ age of grandparents
- type of car $\longleftrightarrow$ city in which you live

Nonparametric regression estimation

What is regression ?

regression = finding a relation between two "things" (variables)

Examples :

- weight $\longleftrightarrow$ length
- length $\longleftrightarrow$ nationality
- life expectancy $\longleftrightarrow$ age of grandparents
- type of car $\longleftrightarrow$ city in which you live
- salary $\longleftrightarrow$ education level
- etc...

Nonparametric regression estimation

In mathematical words :

Let (X, Y) be a random vector where Y is assumed to be related to X. Then we denote the regression function of Y w.r.t. X as

$$m(t) = \mathbb{E}[Y|X = t].$$

read it as : *"the expected value of* Y *given that* $X = t$*".*

Nonparametric regression estimation

In mathematical words :

Let (X, Y) be a random vector where Y is assumed to be related to X. Then we denote the regression function of Y w.r.t. X as

$$m(t) = \mathbb{E}[Y|X = t].$$

read it as : "the expected value of Y given that $X = t$".

Example : expected salary in function of the study time

Nonparametric regression estimation

In mathematical words :

Let (X, Y) be a random vector where Y is assumed to be related to X. Then we denote the regression function of Y w.r.t. X as

$$m(t) = \mathbb{E}[Y|X = t].$$

read it as : "the expected value of Y given that $X = t$".

Example : expected salary in function of the study time

- let Y = salary
- let X = number of years of study

Nonparametric regression estimation

In mathematical words :

Let (X, Y) be a random vector where Y is assumed to be related to X. Then we denote the regression function of Y w.r.t. X as

$$m(t) = \mathbb{E}[Y|X = t].$$

read it as : *"the expected value of Y given that $X = t$".*

Example : expected salary in function of the study time

- let $Y =$ salary
- let $X =$ number of years of study

$$\Longrightarrow \quad m(4) = \mathbb{E}[Y|X = 4] = \text{expected salary if you have studied for 4 years}$$

Nonparametric regression estimation

Estimation :

an estimation $=$ a smart way to approximate $m(t)$

Nonparametric regression estimation

Estimation :

an estimation = a smart way to approximate $m(t)$

- essentially a clever calculation based upon the data
- many different ways to make an estimation

Nonparametric regression estimation

Estimation :

an estimation = a smart way to approximate $m(t)$

- essentially a clever calculation based upon the data
- many different ways to make an estimation

Ex. : estimate the mean length of coffee beans

Nonparametric regression estimation

Estimation :

an estimation = a smart way to approximate $m(t)$

- essentially a clever calculation based upon the data
- many different ways to make an estimation

Ex. : estimate the mean length of coffee beans

- average
- median
- something more sophisticated

Nonparametric regression estimation

Estimation :

an estimation = a smart way to approximate $m(t)$

- essentially a clever calculation based upon the data
- many different ways to make an estimation

  Ex. : estimate the mean length of coffee beans

  - average
  - median
  - something more sophisticated
- we will focus on the so–called "kernel–type estimators".

Nonparametric regression estimation

Nonparametric estimation :

nonparametric = the idea of making no assumption about underlying distribution of the variables

Nonparametric regression estimation

Nonparametric estimation :

nonparametric = the idea of making no assumption about underlying distribution of the variables

- in contrast with parametric models described only by parameters

Nonparametric regression estimation

Nonparametric estimation :

nonparametric = the idea of making no assumption about underlying distribution of the variables

- in contrast with parametric models described only by parameters
- much more generally applicable than parametric estimation

Nonparametric regression estimation

Nonparametric estimation :

nonparametric = the idea of making no assumption about underlying distribution of the variables

- in contrast with parametric models described only by parameters
- much more generally applicable than parametric estimation
- many many methods available, e.g. : histogram, averages, splines, kernel–type estimators, ...

II. Consistency

Consistency

What is consistency ?

consistency = the property of improving the approximation when more and more data are available

Consistency

What is consistency ?

consistency = the property of improving the approximation when more and more data are available

$\hookrightarrow$ the more observations, the better the estimation

Consistency

What is consistency ?

consistency = the property of improving the approximation when more and more data are available

$\hookrightarrow$ the more observations, the better the estimation

Example : estimators for the mean length of coffee beans

Consistency

What is consistency ?

consistency = the property of improving the approximation when more and more data are available

$\hookrightarrow$ the more observations, the better the estimation

Example : estimators for the mean length of coffee beans

- take an average of the lengths of the beans in the sample

Consistency

What is consistency ?

consistency = the property of improving the approximation when more and more data are available

$\hookrightarrow$ the more observations, the better the estimation

Example : estimators for the mean length of coffee beans

- take an average of the lengths of the beans in the sample
  $\hookrightarrow$ consistent

Consistency

What is consistency ?

consistency = the property of improving the approximation when more and more data are available

$\hookrightarrow$ the more observations, the better the estimation

Example : estimators for the mean length of coffee beans

- take an average of the lengths of the beans in the sample
  $\hookrightarrow$ consistent
- take the longest bean in the sample and divide it by 2

Consistency

What is consistency ?

consistency = the property of improving the approximation when more and more data are available

$\hookrightarrow$ the more observations, the better the estimation

Example : estimators for the mean length of coffee beans

- take an average of the lengths of the beans in the sample
  $\hookrightarrow$ consistent
- take the longest bean in the sample and divide it by 2
  $\hookrightarrow$ (generally) not consistent

Consistency

In mathematical words :

a consistent estimator = an estimator that converges to the estimated quantity as the sample size $n \to \infty$.

More precisely, let $\hat{m}_n(t)$ be an estimator for $m(t)$ based upon i.i.d. random variables $(X_1, Y_1), \ldots, (X_n, Y_n)$. Then $\hat{m}_n(t)$ is consistent if

$$\hat{m}_n(t) \longrightarrow m(t), \quad \text{when } n \to \infty.$$

III. Kernel–type estimators

Kernel–type estimators

Kernel–type estimators are based upon **two quantities** :

Kernel–type estimators

Kernel–type estimators are based upon **two quantities** :

(1) a kernel function K

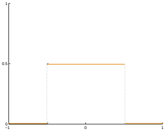
- bounded,
- $\int K(u)du = 1$,
- usually (but not necessarily) compact support.

Kernel–type estimators

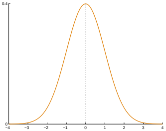
Kernel–type estimators are based upon **two quantities** :

(1) a kernel function K

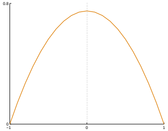
- bounded,
- $\int K(u)du = 1$,
- usually (but not necessarily) compact support.



Uniform kernel



Gaussian kernel



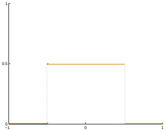
Epanechnikov kernel

Kernel–type estimators

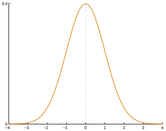
Kernel–type estimators are based upon **two quantities** :

(1) a kernel function K

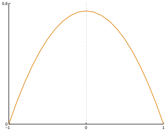
- bounded,
- $\int K(u)du = 1$,
- usually (but not necessarily) compact support.



Uniform kernel



Gaussian kernel



Epanechnikov kernel

(2) a bandwidth :

- a number $0 < h < 1$,
- usually (and necessarily for consistency) $h_n \to 0$.

Kernel–type estimators

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. random variables in $\mathbb{R}^d \times \mathbb{R}^r$, and $\varphi : \mathbb{R}^r \to \mathbb{R}$ a measurable function. A typical kernel–type estimator looks like

$$\hat{\varphi}_{n,h}(t) = \frac{1}{nh^d} \sum_{i=1}^{n} \varphi(Y_i) K\Big(\frac{t - X_i}{h}\Big), \quad t \in \mathbb{R}^d.$$

Kernel–type estimators

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. random variables in $\mathbb{R}^d \times \mathbb{R}^r$, and $\varphi : \mathbb{R}^r \to \mathbb{R}$ a measurable function. A typical kernel–type estimator looks like

$$\hat{\varphi}_{n,h}(t) = \frac{1}{nh^d} \sum_{i=1}^{n} \varphi(Y_i) K\Big(\frac{t - X_i}{h}\Big), \quad t \in \mathbb{R}^d.$$

- it is a weighted sum of the $\varphi(Y_i)$'s

Kernel–type estimators

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. random variables in $\mathbb{R}^d \times \mathbb{R}^r$, and $\varphi : \mathbb{R}^r \to \mathbb{R}$ a measurable function. A typical kernel–type estimator looks like

$$\hat{\varphi}_{n,h}(t) = \frac{1}{nh^d} \sum_{i=1}^{n} \varphi(Y_i) K\left(\frac{t - X_i}{h}\right), \quad t \in \mathbb{R}^d.$$

- it is a weighted sum of the $\varphi(Y_i)$'s
- **role of the kernel** : determines the shape of the weights that are assigned in a neighborhood of the observation X_i

Kernel–type estimators

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. random variables in $\mathbb{R}^d \times \mathbb{R}^r$, and $\varphi : \mathbb{R}^r \to \mathbb{R}$ a measurable function. A typical kernel–type estimator looks like

$$\hat{\varphi}_{n,h}(t) = \frac{1}{nh^d} \sum_{i=1}^{n} \varphi(Y_i) K\left(\frac{t - X_i}{h}\right), \quad t \in \mathbb{R}^d.$$

- it is a weighted sum of the $\varphi(Y_i)$'s
- **role of the kernel** : determines the shape of the weights that are assigned in a neighborhood of the observation X_i
- **role of the bandwith** : controls the size of the neighborhood

Examples of particular interest

1. the kernel density estimator

$$\hat{f}_{n,h}(t) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h}\right), \quad t \in \mathbb{R}^d.$$

Examples of particular interest

1. the kernel density estimator

$$\hat{f}_{n,h}(t) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h}\right), \quad t \in \mathbb{R}^d.$$

- for estimating the density of X

Examples of particular interest

1. the kernel density estimator

$$\hat{f}_{n,h}(t) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h}\right), \quad t \in \mathbb{R}^d.$$

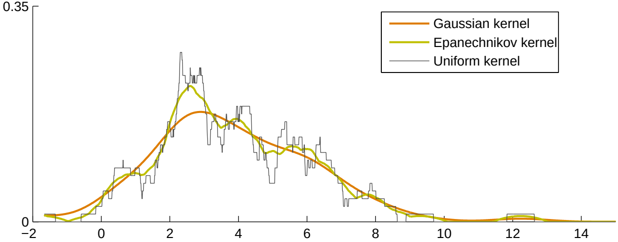
- for estimating the density of X
- improved, smoothed histogram

Examples of particular interest

1. the kernel density estimator

$$\hat{f}_{n,h}(t) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h}\right), \quad t \in \mathbb{R}^d.$$

- for estimating the density of X
- improved, smoothed histogram



Examples of particular interest

2. the Nadaraya–Watson–type estimator

$$\hat{m}_{n,h,\varphi}(t) = \frac{\sum_{i=1}^{n} \varphi(Y_i) K\left(\frac{t-X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{t-X_i}{h}\right)}.$$

Examples of particular interest

2. the Nadaraya–Watson–type estimator

$$\hat{m}_{n,h,\varphi}(t) = \frac{\sum_{i=1}^{n} \varphi(Y_i) K\left(\frac{t-X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{t-X_i}{h}\right)}.$$

- to estimate the regression function $m_\varphi(t) = \mathbb{E}[\varphi(Y)|X = t]$

Examples of particular interest

2. the Nadaraya–Watson–type estimator

$$\hat{m}_{n,h,\varphi}(t) = \frac{\sum_{i=1}^{n} \varphi(Y_i) K\left(\frac{t-X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{t-X_i}{h}\right)}.$$

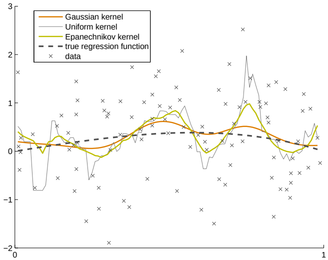
- to estimate the regression function $m_\varphi(t) = \mathbb{E}[\varphi(Y)|X = t]$
- weighted average of the $\varphi(Y_i)$'s

Examples of particular interest

2. the Nadaraya–Watson–type estimator

$$\hat{m}_{n,h,\varphi}(t) = \frac{\sum_{i=1}^{n} \varphi(Y_i) K\left(\frac{t-X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{t-X_i}{h}\right)}.$$

- to estimate the regression function $m_\varphi(t) = \mathbb{E}[\varphi(Y)|X = t]$
- weighted average of the $\varphi(Y_i)$'s



Three NW estimators for

$$m_\varphi(t) = t(1 - t^2)$$

based upon three different kernel functions.

Examples of particular interest

3. the conditional U–statistic of order $m \leq n$

$$\hat{m}_{n,h,\varphi}(\mathbf{t}) := \frac{\sum_{\mathbf{i}\in I_n^m} \varphi(Y_{i_1},\ldots,Y_{i_m}) K\left(\frac{t_1 - X_{i_1}}{h}\right)\cdots K\left(\frac{t_m - X_{i_m}}{h}\right)}{\sum_{\mathbf{i}\in I_n^m} K\left(\frac{t_1 - X_{i_1}}{h}\right)\cdots K\left(\frac{t_m - X_{i_m}}{h}\right)},$$

where $I_n^m = \{(i_1,\ldots,i_m) : 1 \leq i_j \leq n, i_j \neq i_k \text{ if } j \neq k\}$.

Examples of particular interest

3. the conditional U–statistic of order $m \leq n$

$$\hat{m}_{n,h,\varphi}(\mathbf{t}) := \frac{\sum_{\mathbf{i}\in I_n^m} \varphi(Y_{i_1},\ldots,Y_{i_m}) K\left(\frac{t_1 - X_{i_1}}{h}\right)\cdots K\left(\frac{t_m - X_{i_m}}{h}\right)}{\sum_{\mathbf{i}\in I_n^m} K\left(\frac{t_1 - X_{i_1}}{h}\right)\cdots K\left(\frac{t_m - X_{i_m}}{h}\right)},$$

where $I_n^m = \{(i_1,\ldots,i_m) : 1 \leq i_j \leq n, i_j \neq i_k \text{ if } j \neq k\}$.

- generalization of the Nadaraya–Watson–type estimators $(m = 1)$
- useful to estimate "multivariate" regression functions

$$m_\varphi(\mathbf{t}) = \mathbb{E}[\varphi(Y_1,\ldots,Y_m)|(X_1,\ldots,X_m) = \mathbf{t}], \mathbf{t} \in \mathbb{R}^m.$$

Examples of particular interest

Examples :

- conditional moments : take $\varphi(y_1) = y_1^k$, then for $\mathbf{t} = t$,

$$m_\varphi(\mathbf{t}) = \mathbb{E}[\varphi(Y_1)|X_1 = t] = \mathbb{E}[Y^k|X = t].$$

Examples of particular interest

Examples :

- conditional moments : take $\varphi(y_1) = y_1^k$, then for $\mathbf{t} = t$,

$$m_\varphi(\mathbf{t}) = \mathbb{E}[\varphi(Y_1)|X_1 = t] = \mathbb{E}[Y^k|X = t].$$

- conditional variance : take $\varphi(y_1, y_2) = (y_1 - y_2)^2/2$, then for $\mathbf{t} = (t, t)$,

$$m_\varphi(\mathbf{t}) = \mathbb{E}[\varphi(Y_1, Y_2)|(X_1, X_2) = (t, t)] = \mathrm{Var}(Y|X = t).$$

Examples of particular interest

Examples :

- conditional moments : take $\varphi(y_1) = y_1^k$, then for $\mathbf{t} = t$,

$$m_\varphi(\mathbf{t}) = \mathbb{E}[\varphi(Y_1)|X_1 = t] = \mathbb{E}[Y^k|X = t].$$

- conditional variance : take $\varphi(y_1, y_2) = (y_1 - y_2)^2/2$, then for $\mathbf{t} = (t, t)$,

$$m_\varphi(\mathbf{t}) = \mathbb{E}[\varphi(Y_1, Y_2)|(X_1, X_2) = (t, t)] = \text{Var}(Y|X = t).$$

- conditional probabilities : take $\varphi(y_1, y_2) = \mathbb{I}\{y_1 > y_2\}$, then for $\mathbf{t} = (t_1, t_2)$,

$$m_\varphi(\mathbf{t}) = \mathbb{P}\{Y_1 > Y_2|X_1 = t_1, X_2 = t_2\}.$$

Examples of particular interest

Known results when $h_n \to 0$:

- uniform consistency : under some appropriate assumptions, $\hat{f}_{n,,h_n}(t)$, $\hat{m}_{n,,h_n,\varphi}(t)$ and $\hat{m}_{n,,h_n,\varphi}(\mathbf{t})$ are uniformly consistent estimators, i.e.,

$$\sup_{t\in\mathbb{R}^d} |\hat{f}_{n,h_n}(t) - f_X(t)| \longrightarrow 0, \quad \text{a.s.}$$

$$\sup_{t\in I}\sup_{\varphi\in\mathcal{F}} |\hat{m}_{n,h_n,\varphi}(t) - m_\varphi(t)| \longrightarrow 0, \quad \text{a.s.}$$

$$\sup_{\mathbf{t}\in I^m} |\hat{m}_{n,h_n,\varphi}(\mathbf{t}) - m_\varphi(\mathbf{t})| \longrightarrow 0, \quad \text{a.s.}$$

Examples of particular interest

Known results when $h_n \to 0$:

- convergence rates : under some appropriate assumptions, it holds when in particular $h_n \to 0$ and $nh_n^d / \log n \to \infty$ that

$$\sup_{t \in \mathbb{R}^d} |\hat{f}_{n,h_n}(t) - \mathbb{E}\hat{f}_{n,h_n}(t)| = O\left(\sqrt{\frac{|\log h_n|}{nh_n^d}}\right), \quad \text{a.s.}$$

$$\limsup_{n \to \infty} \sup_{\varphi \in \mathcal{F}} \sup_{t \in I} \frac{\sqrt{nh_n^d}|\hat{m}_{n,h_n,\varphi}(t) - \widehat{\mathbb{E}}\hat{m}_{n,h_n,\varphi}(t)|}{\sqrt{|\log h_n|}} < \infty, \quad \text{a.s.}$$

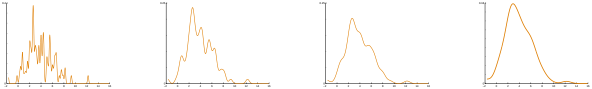
$$\sup_{\mathbf{t} \in I^m} |\hat{m}_{n,h_n,\varphi}(\mathbf{t}) - m_\varphi(\mathbf{t})| = O\left(\sqrt{\frac{\log n}{nh_n^{2m}}}\right), \quad \text{a.s.}$$

IV. A brief motivation for UiB results

Motivation of UiB

Issue : choice of the bandwidth

- kernel density estimator

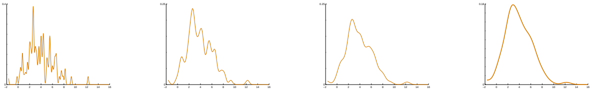


$h = 0.1$ $h = 0.3$ $h = 0.5$ $h = 0.8$

Motivation of UiB

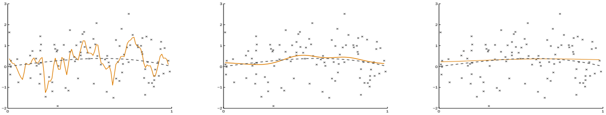
Issue : choice of the bandwidth

- kernel density estimator



$h = 0.1$ $h = 0.3$ $h = 0.5$ $h = 0.8$

- Nadaraya–Watson estimator

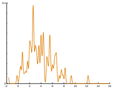


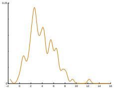
$h = 0.01$ $h = 0.1$ $h = 0.3$

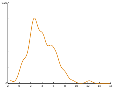
Motivation of UiB

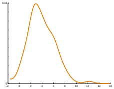
Issue : choice of the bandwidth

- kernel density estimator

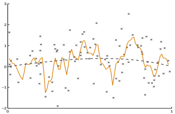

$h = 0.1$

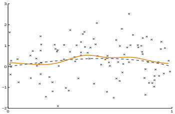

$h = 0.3$

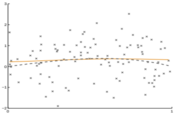

$h = 0.5$


$h = 0.8$

- Nadaraya–Watson estimator


$h = 0.01$


$h = 0.1$


$h = 0.3$

$\Rightarrow$ **bias–variance trade–off**

Motivation of UiB

Question : what is the optimal bandwidth ?

- a huge amount of literature exists about how to find the "optimal bandwidth" for $\hat{m}_{n,h,\varphi}(t)$, e.g.,
  - cross validation : Härdle and Marron (1985)
  - plug–in methods : Tsybakov (1987)
  - many more ...

Motivation of UiB

Question : what is the optimal bandwidth ?

- a huge amount of literature exists about how to find the "optimal bandwidth" for $\hat{m}_{n,h,\varphi}(t)$, e.g.,
  - cross validation : Härdle and Marron (1985)
  - plug–in methods : Tsybakov (1987)
  - many more ...
- most of those bandwidth selectors are **data–dependent**

$\hookrightarrow$ **we need to consider estimators with random bandwidths** $h = H(X_1, \ldots, X_n)$

Motivation of UiB

Question : what is the optimal bandwidth ?

- a huge amount of literature exists about how to find the "optimal bandwidth" for $\hat{m}_{n,h,\varphi}(t)$, e.g.,
  - cross validation : Härdle and Marron (1985)
  - plug–in methods : Tsybakov (1987)
  - many more ...
- most of those bandwidth selectors are **data–dependent**

  $\hookrightarrow$ **we need to consider estimators with random bandwidths** $h = H(X_1, \dots, X_n)$

$\Rightarrow$ classical consistency results with fixed $h_n \to 0$ do not apply

Motivation of UiB

Solution : Uniform in bandwidth (UiB) consistency

Motivation of UiB

Solution : Uniform in bandwidth (UiB) consistency

- Typical uniform–in–h result : add an extra $\sup_{a_n \le h \le b_n}$

Motivation of UiB

Solution : Uniform in bandwidth (UiB) consistency

- Typical uniform–in–h result : add an extra $\sup_{a_n \leq h \leq b_n}$
- Einmahl and Mason (2005) : for $b_n \to 0, c_n \to \infty$,

$$\sup_{\frac{c_n \log n}{n} \leq h \leq b_n} \sup_{\varphi \in \mathcal{F}} \sup_{t \in I} \left| \hat{m}_{n,h,\varphi}(t) - m_\varphi(t) \right| \longrightarrow 0, \quad a.s.$$

Motivation of UiB

Solution : Uniform in bandwidth (UiB) consistency

- Typical uniform–in–h result : add an extra $\sup_{a_n \leq h \leq b_n}$
- Einmahl and Mason (2005) : for $b_n \to 0, c_n \to \infty$,

$$\sup_{\frac{c_n \log n}{n} \leq h \leq b_n} \sup_{\varphi \in \mathcal{F}} \sup_{t \in I} |\hat{m}_{n,h,\varphi}(t) - m_\varphi(t)| \longrightarrow 0, \quad a.s.$$

- **Application to random bandwidths :**

  Let $\hat{h}_n = H(X_1, \ldots, X_n)$ be a solution of any data–driven bandwidth selector such that $\hat{h}_n \in [a_n, b_n]$,

$$\Rightarrow \quad \hat{m}_{n,\hat{h}_n,\varphi}(t) \longrightarrow m_\varphi(t) \quad \text{uniformly on compact sets.}$$

Motivation of UiB

Solution : Uniform in bandwidth (UiB) consistency

- Typical uniform–in–h result : add an extra $\sup_{a_n \le h \le b_n}$
- Einmahl and Mason (2005) : for $b_n \to 0, c_n \to \infty$,

$$\sup_{\frac{c_n \log n}{n} \le h \le b_n} \sup_{\varphi \in \mathcal{F}} \sup_{t \in I} \left| \hat{m}_{n,h,\varphi}(t) - m_\varphi(t) \right| \longrightarrow 0, \quad a.s.$$

- **Application to random bandwidths :**

  Let $\hat{h}_n = H(X_1, \ldots, X_n)$ be a solution of any data–driven bandwidth selector such that $\hat{h}_n \in [a_n, b_n]$,

$$\Rightarrow \quad \hat{m}_{n,\hat{h}_n,\varphi}(t) \longrightarrow m_\varphi(t) \quad \text{uniformly on compact sets.}$$

$\hookrightarrow$ we do not lose the consistency as long as $\hat{h}_n \in [\frac{c_n \log n}{n}, b_n]$.

Motivation of UiB

Solution : concretely...

Motivation of UiB

Solution : concretely...

Aim : making coffee. Therefore one needs :

Motivation of UiB

Solution : concretely...

Aim : making coffee. Therefore one needs :

(1) a coffee machine




Motivation of UiB

Solution : concretely...

Aim : making coffee. Therefore one needs :

(1) a coffee machine




(2) appropriate coffee



Motivation of UiB

Solution : concretely...

Suppose that : you go for a coffee break, so

Motivation of UiB

Solution : concretely...

Suppose that : you go for a coffee break, so

- you bring your favorite coffee ,

Motivation of UiB

Solution : concretely...

Suppose that : you go for a coffee break, so

- you bring your favorite coffee ,
- but your friend only has a fancy  machine.

Motivation of UiB

Solution : concretely...

Suppose that : you go for a coffee break, so

- you bring your favorite coffee ,
- but your friend only has a fancy  machine.

$\Longrightarrow$  is useless, and thus no coffee.

Motivation of UiB

Solution : concretely...

Imagine : a universal wondermachine able to make coffee with any sort of coffee system :

Motivation of UiB

Solution : concretely...

Imagine : a universal wondermachine able to make coffee with any sort of coffee system :

accepts all sorts of coffee systems

⟵

Motivation of UiB

Solution : concretely...

Imagine : a universal wondermachine able to make coffee with any sort of coffee system :

accepts all sorts of coffee systems

⟵

Then this wondermachine would be called a

"uniform–in–coffee–system" coffee machine...

Motivation of UiB

UiCS coffee machines $\longleftrightarrow$ UiB results :

Aim :	make your favorite coffee 	analyze estimators based upon optimal bandwidths
existing tool : (useless)	 for 	consistency results for $h_n \rightarrow 0$
solution :	 , , , 	UiB consistency results $a_n \leq h \leq b_n$

V. A general UiB result

General Theorem [GT]

The first UiB result for general processes of the form

$$W_{n,h,\varphi}(t) := \sum_{i=1}^{n} \{c_\varphi(t)\varphi(Y_i) + d_\varphi(t)\} K\left(\frac{t - X_i}{h}\right)$$
$$- n\mathbb{E}\{c_\varphi(t)\varphi(Y) + d_\varphi(t)\} K\left(\frac{t - X}{h}\right)$$

was proved by Einmahl and Mason (2005).

General Theorem [GT]

Under the "Usual Assumptions [UA]", it holds that with probability 1,

$$\limsup_{n\to\infty} \sup_{c(\log n/n)^{\gamma} \le h \le b_0} \sup_{\varphi\in\mathcal{F}} \sup_{t\in I} \frac{|W_{n,h,\varphi}(t)|}{\sqrt{nh^d(|\log h| \vee \log\log n)}} = O(1).$$

General Theorem [GT]

The General Theorem implies the UiB consistency of

- kernel density estimators $\hat{f}_{n,h}(t)$,
- Nadaraya–Watson–type estimators $\hat{m}_{n,h,\varphi}(t)$.

General Theorem [GT]

The General Theorem implies the UiB consistency of

- kernel density estimators $\hat{f}_{n,h}(t)$,
- Nadaraya–Watson–type estimators $\hat{m}_{n,h,\varphi}(t)$.

We shall establish UiB (consistency) results to

General Theorem [GT]

The General Theorem implies the UiB consistency of

- kernel density estimators $\hat{f}_{n,h}(t)$,
- Nadaraya–Watson–type estimators $\hat{m}_{n,h,\varphi}(t)$.

We shall establish UiB (consistency) results to

- kernel density estimators w.r.t. weighted sup–norms $\psi(t)\hat{f}_{n,h}(t)$,

General Theorem [GT]

The General Theorem implies the UiB consistency of

- kernel density estimators $\hat{f}_{n,h}(t)$,
- Nadaraya–Watson–type estimators $\hat{m}_{n,h,\varphi}(t)$.

We shall establish UiB (consistency) results to

- kernel density estimators w.r.t. weighted sup–norms $\psi(t)\hat{f}_{n,h}(t)$,
- local polynomial estimators $\hat{m}^{(p)}_{n,h}(t)$,

General Theorem [GT]

The General Theorem implies the UiB consistency of

- kernel density estimators $\hat{f}_{n,h}(t)$,
- Nadaraya–Watson–type estimators $\hat{m}_{n,h,\varphi}(t)$.

We shall establish UiB (consistency) results to

- kernel density estimators w.r.t. weighted sup–norms $\psi(t)\hat{f}_{n,h}(t)$,
- local polynomial estimators $\hat{m}^{(p)}_{n,h}(t)$,
- conditional U–statistics $\hat{m}_{n,h,\varphi}(\mathbf{t})$,

General Theorem [GT]

The General Theorem implies the UiB consistency of

- kernel density estimators $\hat{f}_{n,h}(t)$,
- Nadaraya–Watson–type estimators $\hat{m}_{n,h,\varphi}(t)$.

We shall establish UiB (consistency) results to

- kernel density estimators w.r.t. weighted sup–norms $\psi(t)\hat{f}_{n,h}(t)$,
- local polynomial estimators $\hat{m}_{n,h}^{(p)}(t)$,
- conditional U–statistics $\hat{m}_{n,h,\varphi}(\mathbf{t})$,
- kernel–type estimators at a fixed point $\hat{\varphi}_{n,h}(t)$,

General Theorem [GT]

The General Theorem implies the UiB consistency of

- kernel density estimators $\hat{f}_{n,h}(t)$,
- Nadaraya–Watson–type estimators $\hat{m}_{n,h,\varphi}(t)$.

We shall establish UiB (consistency) results to

- kernel density estimators w.r.t. weighted sup–norms $\psi(t)\hat{f}_{n,h}(t)$,
- local polynomial estimators $\hat{m}^{(p)}_{n,h}(t)$,
- conditional U–statistics $\hat{m}_{n,h,\varphi}(\mathbf{t})$,
- kernel–type estimators at a fixed point $\hat{\varphi}_{n,h}(t)$,
- the kernel–based Hill estimator $\hat{\tau}_{n,h}$.

Main steps of the proof :

Standard Methodology – [SM]

Standard Methodology [SM]

STEP 1 : empirical process representation

$$\begin{aligned} W_{n,h,\varphi}(t) &= \sum_{i=1}^{n}\{c_\varphi(t)\varphi(Y_i)+d_\varphi(t)\}K\left(\frac{t-X_i}{h}\right) \\ &\qquad -n\mathbb{E}\{c_\varphi(t)\varphi(Y)+d_\varphi(t)\}K\left(\frac{t-X}{h}\right) \\ &= \sqrt{n}\alpha_n(g), \end{aligned}$$

where α_n is the empirical process indexed by a suitable subclass $\mathcal{G}_n$ of the class of functions

$$\mathcal{G}=\{(x,y)\mapsto c_\varphi(t)\varphi(y)K\left(\frac{t-x}{h}\right):\varphi\in\mathcal{F},t\in I,0<h<1\}.$$

Standard Methodology [SM]

STEP 2 : blocking argument and moment bound

- blocking : set $n_k = 2^k, k \geq 0$.

Standard Methodology [SM]

STEP 2 : blocking argument and moment bound

- blocking : set $n_k = 2^k, k \geq 0$.
- $\mathcal{G}_{n_k}$ is decomposed into subclasses $\mathcal{G}_{n_k,j}$ in which

$$2^{j-1} a_{n_k} \leq h \leq 2^j a_{n_k}, \quad (a_n = c \log n / n).$$

Standard Methodology [SM]

STEP 2 : blocking argument and moment bound

- blocking : set $n_k = 2^k, k \geq 0$.
- $\mathcal{G}_{n_k}$ is decomposed into subclasses $\mathcal{G}_{n_k,j}$ in which

$$2^{j-1} a_{n_k} \leq h \leq 2^j a_{n_k}, \quad (a_n = c \log n/n).$$

- crucial bounds :

$$\mathbb{E} G_n^2(X,Y) \equiv \beta_{k,j}^2 \quad \text{and} \quad \sup_{g \in \mathcal{G}_{n_k,j}} \mathbb{E} g^2(X,Y) \leq 2^j C_1 \beta_{k,j}^2 a_{n_k}^d =: \sigma_{k,j}^2.$$

Standard Methodology [SM]

STEP 2 : blocking argument and moment bound

- blocking : set $n_k = 2^k, k \geq 0$.
- $\mathcal{G}_{n_k}$ is decomposed into subclasses $\mathcal{G}_{n_k,j}$ in which

$$2^{j-1} a_{n_k} \leq h \leq 2^j a_{n_k}, \quad (a_n = c \log n/n).$$

- crucial bounds :

$$\mathbb{E} G_n^2(X,Y) \equiv \beta_{k,j}^2 \quad \text{and} \quad \sup_{g \in \mathcal{G}_{n_k,j}} \mathbb{E} g^2(X,Y) \leq 2^j C_1 \beta_{k,j}^2 a_{n_k}^d =: \sigma_{k,j}^2.$$

- Moment bound of Einmahl and Mason (2005) :

$$\mathbb{E} \| \sum_{i=1}^{n_k} \varepsilon_i g(X_i, Y_i) \|_{\mathcal{G}_{n_k,j}}^2 \leq n_k a_{n_k}^d (|\log a_{n_k}| \vee \log\log n_k) =: \lambda_{k,j}^2.$$

Standard Methodology [SM]

STEP 3 : exponential inequality

- Talagrand's exponential inequality (1994) with $t = \rho\lambda_{k,j}, \rho > 1$:

$$\mathbb{P}\left\{ \max_{n_{k-1}<n\leq n_k} \|\sqrt{n}\alpha_n(g)\|_{\mathcal{G}_{n_k,j}} \geq A_1(1+\rho)\lambda_{k,j} \right\} \leq 4(\log n_k)^{-A_2\rho}.$$

Standard Methodology [SM]

STEP 3 : exponential inequality

- Talagrand's exponential inequality (1994) with $t = \rho\lambda_{k,j}, \rho > 1$:

$$\mathbb{P}\left\{\max_{n_{k-1}<n\leq n_k} \|\sqrt{n}\alpha_n(g)\|_{\mathcal{G}_{n_k,j}} \geq A_1(1+\rho)\lambda_{k,j}\right\} \leq 4(\log n_k)^{-A_2\rho}.$$

- Set $L(k) = \max\{j \geq 0 : 2^j a_{n_k} \leq 2b_0\}$, then

$$[a_{n_k}, b_0] \subseteq [a_{n_k}, 2^{L(k)}a_{n_k}] \quad \text{and} \quad \mathcal{G}_{n_k} \subseteq \bigcup_{j=0}^{L(k)} \mathcal{G}_{n_k,j},$$

and

$$L(k) \leq 2\log n_k.$$

Standard Methodology [SM]

STEP 4 : Borel–Cantelli

- a little calculation yields that

$$\mathbb{P}\left\{ \max_{n_{k-1}<n\le n_k} \sup_{a_n\le h\le b_0} \frac{\|\sqrt{n}\alpha_n(g)\|_{\mathcal{G}_{n_k}}}{\sqrt{nh^d(|\log h| \vee \log\log n)}} \ge A_1(1+\rho) \right\}$$
$$\le \sum_{j=0}^{L(k)} 4(\log n_k)^{-A_2\rho} \le 8(\log n_k)^{1-A_2\rho}.$$

Standard Methodology [SM]

STEP 4 : Borel–Cantelli

- a little calculation yields that

$$\mathbb{P}\left\{\max_{n_{k-1}<n\le n_k}\sup_{a_n\le h\le b_0}\frac{\|\sqrt{n}\alpha_n(g)\|_{\mathcal{G}_{n_k}}}{\sqrt{nh^d(|\log h|\vee\log\log n)}}\ge A_1(1+\rho)\right\}$$

$$\le\sum_{j=0}^{L(k)}4(\log n_k)^{-A_2\rho}\ \le\ 8(\log n_k)^{1-A_2\rho}.$$

- Borel–Cantelli implies for a large enough $\rho>1$ that wp.1,

$$\limsup_{k\to\infty}\max_{n_{k-1}<n\le n_k}\sup_{a_n\le h\le b_0}\sup_{\varphi\in\mathcal{F}}\sup_{t\in I}\frac{|W_{n,h,\varphi}(t)|}{\sqrt{nh^d(|\log h|\vee\log\log n)}}<\infty.$$