Summary

Nonparametric density and regression estimation has been the subject of intense investigation for many years and this has led to a large number of methods. The nonparametric approach is based on the idea of making little or no assumptions about the underlying distribution function. This is in contrast with the parametric approach where one assumes that the underlying distribution follows a certain parametric model, and the problem consists of estimating a finite number of parameters describing the model. Clearly, since nonparametric models require fewer assumptions, their applicability is much wider than the corresponding parametric models.

One very well–known and commonly used class of estimators consists of the so–called *kernel–type estimators*, which are frequently used to estimate densities and regression functions. A typical kernel–type estimator based upon the variables $(X_1, Y_1), \ldots, (X_n, Y_n)$ with values in $\mathbb{R}^d \times \mathbb{R}^r$ is defined as

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

where K is a kernel function, $0 < h < 1$ a bandwidth, and $\varphi : \mathbb{R}^r \to \mathbb{R}$ is a suitable measurable function. One can view $\hat{\varphi}_{n,h}(t)$ as a weighted average of the $\varphi(Y_i)$'s, where the kernel function determines the shape of the weights assigned to a neighborhood of X_i, and where the bandwidth controls the size of this neighborhood. There are basically no restrictions on the choice of the kernel, apart from being measurable and to integrate to one. Since the kernel–type estimator inherits the smoothness of the kernel, one selects a smooth or less smooth kernel function depending on how smooth $\hat{\varphi}_{n,h}(t)$ has to be.

The choice of the bandwidth, however, is more problematic, as the resulting kernel–type estimator is particularly sensitive to this choice. In the

case of the Nadaraya–Watson estimator (the most frequently applied nonparametric estimator of the regression function), a too small bandwidth results in what is essentially an interpolation of the data, while a too large bandwidth leads to a flat line. The choice of the bandwidth has also a big impact on the so–called bias, which is the difference between the expectation of the estimator one uses and the quantity one wants to estimate. In general, a small bandwidth produces an estimator with a small bias and a larger variance, whereas a larger bandwidth leads to an estimator with a small variance, but a larger bias. One thus has to find an appropriate bandwidth that produces an estimator which has a good balance between bias and variance. Typically, the bandwidth that is most appropriate will vary according to the situation and will depend on the available data. This means that one can no longer investigate the behavior of such estimators via the "classical" results for estimators based on deterministic bandwidth sequences, but one must develop methods that permit the study of estimators based on data–dependent bandwidth sequences.

Our starting point is an article of Einmahl and Mason (2005) where suprema over suitable ranges of bandwidths are added to the original consistency results. Such results are now commonly referred to as "uniform in bandwidth" results, and are typically of the form

$$\limsup_{n\to\infty} \sup_{a_n \le h \le b_0} \sup_{\varphi\in\mathcal{F}} \sup_{t\in I} \frac{\sqrt{nh^d}|\hat{\varphi}_{n,h}(t) - \mathbb{E}\hat{\varphi}_{n,h}(t)|}{\sqrt{|\log h| \vee \log\log n}} < \infty, \quad \text{a.s.}, \qquad (\star)$$

where I is a compact rectangle in $\mathbb{R}^d$, $b_0 < 1$ is a positive constant, a_n a sequence of non–random numbers going to zero at appropriate rates, and where $\mathcal{F}$ is a suitable class of functions. This extra supremum allows us to handle kernel–type estimators based upon bandwidths that are functions of the data. Indeed, if $\hat{h}_n = H(X_1, \ldots, X_n) \in [a_n, b_0]$ is such a data–driven bandwidth sequence, $(\star)$ implies immediately that $|\hat{\varphi}_{n,\hat{h}_n}(t) - \mathbb{E}\hat{\varphi}_{n,\hat{h}_n}(t)| \to 0$ with probability 1, uniformly on a compact $I \subseteq \mathbb{R}^d$ and one can also obtain convergence rates.

The main purpose of this thesis is to prove such consistency results for a wide variety of estimators. Our methodology is based on the techniques developed by Einmahl and Mason (2005) and the main steps are summarized in Section 3.3. This methodology will also be referred to as the "Standard Methodology", or simply [SM]. It relies mainly on the theory of empirical processes, and the basic tools are appropriate exponential deviation inequalities and moment inequalities for empirical processes. Throughout the different chapters of this thesis, we will apply this method several times to establish the uniform in bandwidth consistency of specific classes of kernel–type estimators.

A first application of [SM] leads to a refinement of ($\star$) for the kernel density estimator $\hat{f}_{n,h}(t)$ and consists of replacing the standard supremum–norm by a weighted supremum–norm based on suitable weight functions ψ. Such refinements have been obtained by Giné, Koltchinskii and Zinn (2004) for deterministic bandwidth sequences, and we provide "uniform in bandwidth" versions of their results. In particular, necessary and sufficient conditions are obtained for stochastic and almost sure boundedness of the process

$$\sup_{a_n \leq h \leq b_n} \sup_{t \in \mathbb{R}^d} \frac{\sqrt{nh^d}|\psi(t)(\hat{f}_{n,h}(t) - \mathbb{E}\hat{f}_{n,h}(t))|}{\sqrt{|\log h|}},$$

where ψ is an appropriate weight function (possibly unbounded) and a_n and b_n are regularly varying sequences with negative index. A detailed proof of this result is worked out in Chapter 4.

In Chapter 5 we assume that the regression function $m(t) = \mathbb{E}[Y|X = t]$, $t \in \mathbb{R}$ is $p + 1$ times differentiable at $x_0 \in \mathbb{R}$, and we consider a larger class of kernel–type estimators for $m(x_0)$. This class consists of the so–called "local polynomial regression function estimators", defined as being the solution of a weighted least–squares problem where the weights are given by $K((x_0 - X_i)/h)$. In particular, the Nadaraya–Watson estimator belongs to this class, and is a local polynomial estimator of degree $p = 0$. It turns out that the method described in [SM] which is based upon the theory of empirical processes is also applicable to establish the uniform in bandwidth consistency of local polynomial estimators. Let $\hat{m}^{(p)}_{n,h}(x_0)$ be the local polynomial regression estimator in x_0 of degree $p \geq 0$. We shall show in Section 5.2 that

$$\sup_{(\frac{c \log n}{n})^\gamma \leq h \leq b_n} \sup_{x_0 \in I} |\hat{m}^{(p)}_{n,h}(x_0) - m(x_0)| \longrightarrow 0, \quad \text{a.s.},$$

where b_n is an arbitrary sequence going to zero and where $\gamma = 1$ or $\gamma = 1 - 2/q$ according as Y is bounded or has a finite q–th moment.

In Chapter 6 we shall focus on a pointwise uniform in bandwidth result for $\hat{\varphi}_{n,h}(t)$, i.e. uniformly on a range of bandwidths, but for a fixed $t \in \mathbb{R}^d$.

Of course, this case follows directly from the general uniform result in ($\star$). However, some improvements can be achieved in several directions. We show that

$$\limsup_{n\to\infty} \sup_{a_n \le h \le b_0} \sup_{\varphi\in\mathcal{F}} \frac{\sqrt{nh^d}|\hat{\varphi}_{n,h}(t) - \mathbb{E}\hat{\varphi}_{n,h}(t)|}{\sqrt{\log\log n}} < \infty, \quad \text{a.s.,}$$

where $a_n^d = c\log\log n/n$ or $a_n^d \ge n^{-1}(\log n)^{2/(p-2)}$ depending on whether the envelope function of the class $\mathcal{F}$ has a finite moment generating function or has a finite p–th moment. Clearly, a first improvement concerns the convergence rate, which is slightly better in the pointwise case. A second improvement is achieved on the range of allowable bandwidths, which is larger than in the uniform case. Finally, the obtained result is more generally applicable compared to the uniform case ($\star$), which is only valid for bounded classes $\mathcal{F}$ or classes having an envelope function with a finite moment of order $p > 2$. From a consistency point of view, this implies that in the pointwise case, and in terms of rates and bandwidth sequences, classes with envelope functions admitting a finite moment generating function are equivalent to bounded classes. This somewhat surprising fact did not seem to be known, even in the deterministic bandwidth sequence case. It is still an open problem whether this is also case in the previous setting dealing with uniform convergence on compact sets of locations.

An important application of the aforementioned pointwise uniform in bandwidth result is the derivation of the uniform in bandwidth consistency of the kernel–based Hill estimator for the tail index of a Pareto–type distribution. If $\hat{\tau}_{n,h}$ denotes this kernel–based Hill estimator based upon n i.i.d. Pareto–distributed random variables with tail index τ, it is proved in Section 6.4 that

$$\sup_{a_n\le h\le b_n} |\hat{\tau}_{n,h} - 1/\tau| = o_{\mathbb{P}}(1),$$

where a_n and b_n are non–random sequences satisfying $b_n \to 0$ and $na_n \to \infty$ among other conditions.

As last application of [SM], we consider a much wider class of kernel–type estimators, called "conditional U–statistics". Those are denoted by $\hat{m}_{n,h,\varphi}(\mathbf{t})$ and are estimators for the "multivariate" regression function $m_\varphi(\mathbf{t}) = \mathbb{E}[\varphi(Y_1,\ldots,Y_m)|(X_1,\ldots,X_m) = \mathbf{t}]$. In Chapter 8 we establish a uniform in bandwidth consistency result for $\hat{m}_{n,h,\varphi}(\mathbf{t})$, which is a consequence of the following asymptotic result, namely

$$\limsup_{n\to\infty} \sup_{a_n^\gamma \le h < b_n} \sup_{\varphi\in\mathcal{F}} \sup_{\mathbf{t}\in I^m} \frac{\sqrt{nh^m}|\hat{m}_{n,h,\varphi}(\mathbf{t}) - \widehat{\mathbb{E}}\hat{m}_{n,h,\varphi}(\mathbf{t})|}{\sqrt{|\log h| \vee \log\log n}} < \infty, \quad \text{a.s.},$$

where $a_n = c(\log n/n)^{1/m}$, $I^m = I\times\ldots\times I$ and as usual, $\gamma = 1$ or $\gamma = 1-2/p$ depending on whether the class $\mathcal{F}$ is bounded or has an envelope function with a finite p–th moment. Note that this result also includes the uniform in bandwidth result for the Nadaraya–Watson estimator, which is a special case of $\hat{m}_{n,h,\varphi}(\mathbf{t})$ when $m = 1$.