Lower bounds on bandwidth selection in hazard estimation

Jean-David Fermanian

To cite this article: Jean-David Fermanian (2001) Lower bounds on bandwidth selection in hazard estimation, Journal of Nonparametric Statistics, 13:4, 515-567, DOI: 10.1080/10485250108832864

To link to this article: https://doi.org/10.1080/10485250108832864

Published online: 10 Jan 2011.

Article views: 22

View related articles
LOWER BOUNDS ON BANDWIDTH SELECTION IN HAZARD ESTIMATION

JEAN-DAVID FERMANIAN*

CREST-ENSAE, 3 Av. Pierre Larousse, 92245 Malakoff Cedex, France

(Received 13 June 2000; In final form 20 July 2000)

In the setting of nonparametric hazard estimation under right random censorship by the kernel method, asymptotic lower bounds for bandwidth selection are provided. If the error criterion is the Integrated Squared Error (ISE), and if the distribution function of the underlying lifetime is sufficiently regular, then it is shown that the relative error of any data-driven bandwidth selector cannot be reduced below order $n^{-1/10}$ asymptotically. On the other hand, if the error criterion is the Mean Integrated Squared Error (MISE), the relative error of bandwidth selection can be reduced to order $n^{-1/2}$, when the hazard function is sufficiently smooth. Possible extensions to the multivariate setting are pointed out. These results are similar with those obtained by Hall and Marron (1991) in univariate density estimation without censoring.

Keywords: Hazard functions; Bandwidth selection; Asymptotic lower bounds; Integrated squared error; Right-censoring

AMS 1991 Subject Classifications: Primary: 62G07; Secondary: 62G20

1. INTRODUCTION

In univariate survival analysis, hazard rates or hazard functions are commonly used. They estimate instantaneous probabilities of occurrence of an event (such as death, breakdown or recovery) at each date $t$. If the random variable $T$ denotes the underlying lifetime, the hazard rate $\lambda$ of $T$ at time $t$ is by definition

$$\lambda(t) = \lim_{dt \to 0} \frac{1}{dt} P(T \in [t, t+dt] | T > t).$$

*e-mail: fermania@ensae.fr
Assume that $T$ is a continuous random variable with survival function $F$ and density $f$. Then, obviously, we have $\lambda(t) = f(t)/F(t)$.

In practice, we do not often observe an i.i.d. sample from $T$. For some part of the sample, we know only that the event will occur after some date (in general depending on each individual). In this case, observations are said to be right censored. Right censoring is the most usual disturbance. Suppose moreover that the censoring mechanism is independent from the failure time. Formally, there exists a r.v. $C$ independent from $T$. For each individual, we observe $X$, the infimum of $T$ and $C$ and $\delta$, an indicator r.v. whose value is one if $T \leq C$ and zero otherwise. $G$ and $H$ will be denote the survival functions of $C$ and $X$ respectively. The nonparametric estimation of univariate hazard functions under right random censoring has generated a large amount of literature: Tanner and Wong (1983); Yandell (1983); Lo et al. (1989); Xiang (1994); Zhang (1996)… among others.

It is well known that the choice of a bandwidth is crucial in kernel estimation. To define a data-driven bandwidth selector, we need to choose a criterion, viz a function of $h$. The minimizer of this criterion, evaluated with the observed data, will be selected. In the literature, there are mainly two criteria with a selection (see Devroye and Györfi, 1984 for another viewpoint): the integrated squared error (or $ISE$) and the mean integrated squared error (or $MISE$), defined, for every hazard function $\lambda$ and every bandwidth $h$ by

$$ISE(h, \lambda) = \int (\hat{\lambda}_h - \lambda)^2 w, \quad MISE(h, \lambda) = E \left[ \int (\hat{\lambda}_h - \lambda)^2 w \right],$$

where $w$ is a positive weight function and $\hat{\lambda}_h$ is the $K$-kernel estimate of $\lambda$, viz

$$\hat{\lambda}_h(x) = \frac{1}{h} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h} \right) \frac{\delta_i}{\sum_{j=1}^{n} I(X_j \geq x_i)}.$$

Denote $h_\lambda$ (resp. $\hat{h}_\lambda$) a minimizer of $MISE(h, \lambda)$ (resp. $ISE(h, \lambda)$) with respects to its first argument. For technical reasons, we impose that these minimizers belong to the real interval $[0, \nu^{-1}]$, where $\nu$ is a (small) positive number.

Since $\lambda$ is unknown, an important issue is to find a bandwidth selector, that goes respectively to $h_\lambda$ or $\hat{h}_\lambda$ (depending on which
BANDWIDTH SELECTION IN HAZARD SELECTION

517

criterion has being chosen). In the literature, the case of the hazard function has not received as much attention as the density case. Nonetheless, some cross-validation techniques have been proposed. The major contributions are probably those of Sarda and Vieu (1991) for non-censored data and Patil (1993a, 1993b) for randomly right-censored data. The latter author shows that the data-driven bandwidth selector converges to the minimizer of the integrated squared error (ISE) at the relative rate \( n^{-1/10} \), in density and hazard rate estimation and in the censored and uncensored settings. Sarda et al. (1994) and Hassani and Youndjé (1995) defined some other estimates of the hazard function as the ratio of two kernel-type estimates. They obtain some results similar to those of Patil with some cross-validation methods. See Gregoire (1993) who provides some theoretical results about intensities of counting processes.

By different bootstrap resampling methods, Gonzalez-Manteiga, Cao and Marron (1996) have proposed a bandwidth selector based on the bootstrap and the MISE criterion (its asymptotic equivalent AMISE more precisely). The previous authors conjecture that the relative rate of convergence towards \( h_\lambda \) is here \( n^{-5/14} \). Moreover, concerning the MISE criterion, Fermanian (1999) has found recently a bandwidth selector by adapting the method of Jones et al. (1991). The relative rate of convergence is \( n^{-1/2} \) in this case.

The performances of these two previous selectors are convincingly better than cross-validation methods, although the implementation is heavier. There exist other studies concerning local choices of the bandwidth based on MISE (see Müller and Wang, 1990, 1994).

In the literature about kernel density estimation, some classical arguments can be found in favour of or against each criterion (see e.g., Hall and Marron, 1991; Turlach, 1993 or Jones et al., 1996). An important point has been stated by Hall and Marron (1987b and 1991): with the ISE criterion, the rate of convergence to the theoretical ISE's minimizer of any bandwidth selector (any measurable function of the data) cannot be smaller than \( n^{-1/10} \). With the MISE criterion, this rate is \( n^{-1/2} \). In our mind, this much faster relative rate turns the bandwidth choice to MISE's advantage even if we estimate a function in an average sense over all possible data sets, and not for our own data set. Nonetheless, others argue that practical, data-driven bandwidths can produce an expected integrated squared error that is
strictly less than the minimum of the MISE. A complete view of this controversy can be found in Grund et al. (1994). By simulations for small or moderate samples, they show particularly that bandwidth rules for each criterion provide close values in all cases. They conclude that "the viewpoints reflect different decision-theoretic problems, and none of the two is priori superior to the other". Thus, the choice between ISE or MISE seems to be a matter of personal taste.

This paper seeks to adapt some previous results obtained by Hall and Marron (1987a, 1987b, 1991) in the density case to hazard functions and possibly right censored data. Most of their comments (particularly the introduction of Hall and Marron, 1991) are available in our setting. In fact, some of them appear in the beginning of the paper. The main conclusions are the same in both frameworks. Thus, it is proved it is impossible to find a bandwidth selector that tends to \( h_\lambda \) (respectively \( \hat{h}_\lambda \)) at relative rate faster than \( n^{-1/2} \) (resp. \( n^{-1/10} \)) under suitable regularity conditions. Contrary to Hall and Marron (1991), our paper is (almost) self-contained. It could be surely be generalized to include in the same way the density and the hazard case (censored or not), following the framework of Patil (1993b).

2. BOUNDS INVOLVING TWO ALTERNATIVES

2.1. Introduction and Summary

To obtain the lower bounds in the current section, it is enough to consider only two alternative hazard functions for each \( n \). Following Hall and Marron (1991) a way to construct these ones is to start with a fixed hazard function \( \lambda_0 \) and a function \( \alpha > 0 \). Consider the alternative hazard function

\[
\lambda_1(x) \equiv \{1 + n^{-1/2} \alpha(x)\} \lambda_0(x).
\]

The fact that \( \lambda_0 \) and \( \lambda_1 \) are distant only \( n^{-1/2} \) apart means that our bounds will apply even in a parametric setting, not solely to nonparametric classes of hazard functions. To distinguish between the different distributions associated with different hazard functions, denote \( T_\lambda \) the r.v. whose hazard function is \( \lambda \). Moreover set \( X_\lambda = C \wedge T_\lambda \). Let \( F_\lambda \) and \( H_\lambda \) be the survival functions of \( T_\lambda \) and \( X_\lambda \),
Particularly, set \( F_0 = F_{\lambda_0} \) and \( H_0 = H_{\lambda_0} \). Denote \( C^k(D) \) the set of \( k \)-times continuously differentiable functions on the domain \( D \). Assume

i. \( 0 < \sigma^2 = \int \alpha^2 \lambda_0(t) \exp(-\int_0^t \lambda_0) dt < \infty \).

ii. \( \lambda_0, \alpha \) and \( G \) belong to \( C^2(\mathbb{R}) \), \( \lambda(0) \neq 0 \), and

\[
\sup_{x,y \geq 0} \left| \frac{\lambda_0(x) - \lambda_0(y)}{x - y} \right| + |\alpha(x) - \alpha(y)| + |G(x) - G(y)|/|x - y|''
\]

is bounded for some positive constant \( \nu \).

To simplify the proofs, we have supposed that \( \lambda_0, \alpha \) and \( G \) belong to \( C^2(\mathbb{R}) \). The reader will verify it is not restrictive. Hall and Marron (1991) have supposed that these functions are fifth times continuously differentiable. Our assumption (ii) is then weaker than theirs. Convenient technical assumptions concerning the estimates are

iii. the kernel \( K \), whose support is included into \([-A, A]\), is even and belongs to \( C^2(\mathbb{R}) \).

iv. the support of \( w \) is bounded and included in the interior of \( \tau = [0, \tau_0] \).

v. \( G(\tau_0) \neq 0 \) and \( F_0(\tau_0) \neq 0 \).

Further useful notation is

\[
p = 1 - \Phi(\sigma/2),
\]

where \( \Phi \) denotes the standard normal cumulative distribution function.

For technical reasons and to simplify the proofs, we impose that every considered bandwidth \( h \) (particularly \( h_\lambda \) and \( \hat{h}_\lambda \)) belongs to \( I_n = [n^{-1}, n^{-\nu}] \), \( 1/5 > \nu > 0 \). This condition seems to be restrictive, but is common in kernel estimation. Indeed, the condition \( h \to 0 \) when \( n \) tends to infinity is necessary to obtain asymptotically unbiased estimates of density, regression or hazard functions. Moreover, the variance of each kernel estimate tends to zero only if \( nh \to \infty \). Finally, the behaviour of hazard functions is very different from density functions when arguments are large. Hazard functions are often unbounded when density functions are integrable. That is why various versions of this condition are commonly used in most of the papers in
the literature about hazard estimation (see e.g., Patil, 1991a; Sarda and Vieu, 1991).

Nonetheless, it is possible to weaken the condition $h \in I_n$, but under some additional and unsatisfying regularity assumptions. Such elements can be found in Fermanian (1998). Particularly, Lemma 5.1 is true even if $h < n^{-1}$. And without assumption on $h$, we can prove that there exists some constant $h_0 > 0$ such that, for every sequence $(\lambda_n)_n$ in $\mathcal{H}_n$ (defined below), $P_{\lambda_n}(\hat{h}_n < h_0) \to 1$.

2.2. Bounds in the Case of the ISE

In addition to the technical assumptions made previously, also assume that the alternative hazard functions $\lambda_0$ and $\lambda_1$ are distinct in the sense that

$$\int d^2(\omega,\lambda_0)\lambda_0 w \neq 0. \quad (2.1)$$

The implications of this condition will be made clear in the proof of the following theorem, which prove the fact that it is impossible to find a data-based bandwidth which is closer to $\hat{h}_\lambda$ than $n^{-1/10}$ in a relative error sense.

**Theorem 2.1** Under the assumptions i–v and (2.1), for $\hat{h}$ some measurable function of the data,

$$\lim_{\epsilon \to 0} \lim_{n \to +\infty} \max_{\lambda \in \{\lambda_0, \lambda_1\}} P_{\lambda}(\hat{h} - \hat{h}_\lambda)/\hat{h}_\lambda > \epsilon n^{-1/10} \geq p, \quad (2.2)$$

$$\lim_{\epsilon \to 0} \lim_{n \to +\infty} \max_{\lambda \in \{\lambda_0, \lambda_1\}} P_{\lambda}(\text{ISE}(\hat{h}, \lambda) - \text{ISE}(\hat{h}_\lambda, \lambda))/\text{ISE}(\hat{h}_\lambda, \lambda) > \epsilon n^{-1/5} \geq p. \quad (2.3)$$

If $\hat{h}$ is chosen by cross-validation then the convergence rates in (2.2) and (2.3) are achievable (see Patil, 1993b). Therefore, the convergence rates described in the previous theorem are the best possible for two alternatives and the ISE criterion.

**Remark 2.1** As pointed out by Hall and Marron (1991), the probability $p$ may be increased to 1 if more than just the two alternatives $\lambda_0$ and $\lambda_1$ are considered. To do this, the simplest way is to
replace \( \{\lambda_0, \lambda_1\} \) (or the set of the densities associated with this set) by
the set of all linear combinations of \( \lambda_0 \) and \( \lambda_1 \).

2.3. Bounds in the Case of the MISE

In addition to the technical assumptions made in Section 2.1, also
assume that \( w \) is three times continuously differentiable on \( \tau \) and that
the alternative hazard functions \( \lambda_0 \) and \( \lambda_1 \) are distinct in the sense that

\[
\int \alpha \lambda_0 \left[ \lambda_0^{(4)} w + \lambda_0^{(2)} w^{(2)} + 2 \lambda_0^{(3)} w^{(1)} \right] \neq 0.
\]  (2.4)

Note that this condition is simpler if \( w \) is constant on the support of \( \lambda_0 \).
In this case, we find the same condition as Hall and Marron (1991).
The following theorem shows that it is impossible to use a data-based
bandwidth which is closer to \( h_A \), the minimizer of the mean integrated
squared error \( MISE(h, \lambda) \), than \( n^{-1/2} \) in a relative error sense.

**Theorem 2.2** Under the assumptions i-\( \nu \) and (2.4), for \( \hat{h} \) some
measurable function of the data,

\[
\lim_{\epsilon \to 0} \lim_{n \to \infty} \inf_{\lambda \in \{\lambda_0, \lambda_1\}} \max_{\lambda \in \{\lambda_0, \lambda_1\}} P_\lambda(|\hat{h} - h_\lambda|/h_\lambda > \epsilon n^{-1/2}) \geq p, \quad (2.5)
\]

\[
\lim_{\epsilon \to 0} \lim_{n \to \infty} \inf_{\lambda \in \{\lambda_0, \lambda_1\}} \max_{\lambda \in \{\lambda_0, \lambda_1\}} P_\lambda(|MISE(\hat{h}, \lambda) - MISE(h_\lambda, \lambda)|/MISE(h_\lambda, \lambda) > \epsilon n^{-1}) \geq p. \quad (2.6)
\]

The rate of convergence \( n^{-1/2} \) has been reached in a multi-
dimensional framework by Fermanian (1999). The same comment as
Remark 2.1 can be made in the case of the MISE.

3. BOUNDS INVOLVING MULTIPLE ALTERNATIVES

The rates of convergence depend on the amount of smoothness
assumed about the underlying hazard function. To quantify this in a
form convenient for minimax lower bound results, consider smooth-
ness classes indexed by a parameter \( \nu \geq 0 \). In particular, let \( l \) be the
largest integer strictly less than \( \alpha = 2 + \nu \) and define \( G_\nu(\tau) \) or simpler \( G_\nu \)
to be the set of all hazard functions which have \( l \) derivatives on \( \tau = [0, \tau_0] \) and satisfy, for some constant \( C(G_v) \),

\[
\sup_{(x,y) \in \mathbb{R}^2} \frac{|\lambda^{(l)}(x) - \lambda^{(l)}(y)|}{|x - y|^{d-l}} \leq C(G_v). \tag{3.1}
\]

Moreover, to identify the underlying distribution beyond \( \tau_0 \), assume that

\[
\liminf_{n} \inf_{\lambda \in G_v} H_\lambda(\tau_0) > 0. \tag{3.2}
\]

A minimax lower bound for the relative rate of convergence of \( \hat{h} \) to \( h_\lambda \), in terms of the smoothness index \( \nu \) is now stated.

**Theorem 3.1** Under the assumptions \( i-\nu \) and (3.2), for \( \hat{h} \) any measurable function of the data,

\[
\lim_{\epsilon \to 0} \liminf_{n \to \infty} \sup_{\lambda \in G_v} P_\lambda(\frac{\hat{h} - h_\lambda}{h_\lambda} > \epsilon n^{-\nu/5}) = 1, \tag{3.3}
\]

\[
\lim_{\epsilon \to 0} \liminf_{n \to \infty} \sup_{\lambda \in G_v} P_\lambda(\frac{\text{MISE}(\hat{h}, \lambda)}{\text{MISE}(h_\lambda, \lambda)} > \epsilon n^{-2\nu/5}) = 1. \tag{3.4}
\]

Note that Theorems 2.2 and 3.1 each provide useful information for different values of \( \nu \). Indeed, if \( \nu \geq 2.5 \), then the lower bound of Theorem 2.2 is more informative. On the other hand, if \( \nu < 2.5 \), the bound of Theorem 3.1 is more useful.

**Remark 3.1** If \( \nu \geq 2.5 \), the optimal rate of \( n^{-1/2} \) is achieved by the bandwidth selector proposed in Fermanian (1999). An open question is to find a bandwidth selector that achieves the optimal rate when \( \nu < 2.5 \).

**Remark 3.2** The bounds obtained in the previous theorem are sharper than those obtained by Hall and Marron (1991, Theorem 3.2) in the density case. It is due to the fact that we have proved equality (5.54) directly, without using an inequality between the variational distance and the Kullback-Leibler information. Thus, using the latter tool and invoking Bickel and Ritov (1988), Hall and Marron need an
additional condition – the sequence \((n^2m^{-4a-1})_n\) is bounded – which induces the rate \(4\nu/(4\nu+9)\). This rate is larger than ours rate \((\nu/5)\) when \(\nu \leq 2.75\), viz for every interesting values of \(\nu\). Moreover, our bound \(\nu/5\) is surely available in the density case too, because the techniques are very similar and the calculations are simpler in the latter case.

**Remark 3.3** The class of hazard functions \(G_\nu\) is bigger than is required to obtain the bound stated in the theorem. Indeed, the proof uses some finite subsets of \(G_\nu\) only. The more general result is left to the reader, because it makes cumbersome the presentation of the main result.

4. EXTENSION TO THE MULTIVARIATE CASE

Multivariate survival analysis is much more difficult than univariate survival analysis. This is partly due to the fact that identifiability of the multivariate distributions is no more easily ensured (see Pruitt, 1993). Moreover, we have to deal with messier notations and concepts than in dimension one (see Gill, 1992a, 1992b, to get a general view of these topics). Multivariate hazard rates have been received more attention, particularly since the papers of Dabrowska (1988, 1989); Pons (1989) and Fermanian (1997).

Our results have been stated in dimension one, where there is a one-to-one correspondence between the hazard rate and the density function. This is no more the case in larger dimensions, because there exist several hazard functions for each distribution. More precisely, consider \(T=(T_1, \ldots, T_d)\) a \(d\)-vector of positive variables (failure times or lifetimes) with survival function \(F(x) = P(T > x) = P(T_1 > x_1, \ldots, T_d > x_d)\) and density function \(f_\). The orders in \(\mathbb{R}^d\) are defined coordinate by coordinate. For example, \(u \leq x\) if for all \(i \in \{1, \ldots, d\}, u_i \leq x_i\). Here \(1\{\}\) will be the indicator function of an event. For each subset \(A\) of \(\{1, \ldots, d\}\) and each vector \(v \in \mathbb{R}^d\), the vector of \(\mathbb{R}^{|A|}\) whose coordinates are \(v_k, k \in A\) is denoted by \(v_A\). Let \(I\) be a subset of \(\{1, \ldots, d\}\) and \(J\) its complementary. The indices in \(I\) point out the at hand failure times, whose number is denoted by \(r\), while the other indices indicate conditional lifetimes.
Define the hazard function (or hazard rate) at a point $x$ such that $F(x) > 0$ by

$$
\lambda_I(x) = \lim_{h_i \to 0, \forall i \in I} \left( \prod_{i \in I} h_i \right)^{-1} \frac{1}{F(x_i, \ldots, x_j) \int f(x_i, u_j) 1\{u_j > x_j\} du_j} 
$$

(4.1)

We know that in the univariate case ($d = 1$), there is a single hazard function, and this function characterizes completely the distribution. In the multivariate case ($d > 1$), $2^d - 1$ hazard functions are necessary to characterize the distribution function of $T$. These hazard functions provide some tools to describe the probability of occurrence of one or a subset of failures, knowing that $d$ failures can happen potentially. Moreover, we can adopt the straightforward generalization of the univariate censoring mechanism, where each failure time $T_k$ is possibly censored by a specific r.v.

These concepts are some straightforward extensions of those used in Dabrowska (1988, 1989), which were restricted to the bivariate case ($d = 2$). An explicite formula that would provide the density function according to the set of functions $\{\lambda_I, I \subseteq \{1, \ldots, d\}\}$, viz something like the reverse of (4.1), is relatively cumbersome. See the method in Dabrowska (1988).

Our goal is to discuss the performances of bandwidth selectors to estimate $\lambda_I$, the subset $I$ being at hand. It is not so easy to generalize the results of the previous sections, because the proofs need to work with the underlying distribution functions, viz with the density functions. For each $I \subseteq \{1, \ldots, d\}$, let $\phi_I$ be the application which sends every density function (in $\mathbb{R}^d$) to its related $I$-hazard function $\lambda_I$.

To define $MISE(h, \lambda_I)$, we have to consider an underlying density function $f$ which must satisfy $\phi_I(f) = \lambda_I$. But a lot of functions $f$ satisfy this identity and to describe $\phi_I^{-1}(\lambda_I)$ is not so simple. Thus, the set $\mathcal{G}_\lambda$ (see the proof of Theorem 3.1) will be much more tedious than in the univariate case.

For example, in dimension 2, the set $\{\lambda_0, \lambda_1\}$ can be defined like in Subsection 2.1, with here $I = \{1, 2\}$, but it is no more sufficient for our purpose. We have to consider $\{f_0, f_1\}$ where $\phi(f_0) = \lambda_0$ and $\phi(f_1) = \lambda_1$. 
in the theorems and proofs, because we state our results with respects to the distributions summarized by $f_0$ or $f_1$.

That is why we have preferred to state the results in the univariate framework. Nonetheless, with the previous alterations, all our results can be extended in dimension larger than 1. More precisely, in a multivariate setting, we conjecture that the rate of convergence obtained in the ISE case becomes $n^{-r/(2q+r)}$ in Eq. (2.2) and $n^{-r/(4+r)}$ in Eq. (2.3). In the MISE case, multivariate kernel functions are integrated. That is why the bounds obtained in Theorem (2.2) should be unchanged. That is partly the case too in Theorem 3.1, changing the bound $\nu/5$ into $\nu/(4+r)$ only.

5. PROOFS

Denote $\psi^{[p]}(h, \cdot)$ the $p$-th derivative of any function $\psi$ with respects to $h$. It will be useful to state some preliminary lemmas in a relatively general framework, viz for a wide class of hazard functions. Thus, let $(\mathcal{H}_n)_{n \geq 1}$ be a sequence of sets of unspecified hazard functions. We need the following technical assumptions.

vi. $\inf_{\lambda} \inf_{x} H_{\lambda}(t) > 0$.

vii. $\limsup_{n, \lambda} \sup_{x} (\lambda H_{\lambda}^{-1}(x))$ is bounded.

viii. for some constant $\nu > 0$, $\limsup_{n, \lambda} \sup_{x} (\lambda^{(2)}(x) - \lambda^{(2)})/|x-y|^\nu$ is bounded.

ix. $\liminf_{n, \inf_{\lambda}} \int \lambda w/H_{\lambda} > 0$.

x. $\liminf_{n, \inf_{\lambda}} \int \lambda^{(2)} w > 0$.

xi. $\liminf_{n, \inf_{\lambda}} \int \lambda^{2} w > 0$.

xii. There exists $\bar{w} < \tau_0$ such that $\liminf_{n, \inf_{\lambda}} \int F_{\lambda}(dt) > 0$.

To simplify the proofs, we have assumed a bit more than viii, viz that each function in $\mathcal{H}_n$ is three times continuously differentiable, and that the third derivative of the hazard functions are bounded uniformly on $x \in \tau$ and $\lambda$. The reader will prove that this additional assumption is not necessary. Particularly, condition xii holds if $\liminf_{n, \inf_{\lambda}} P_{\lambda}(T < \tau_0) > 0$.

Suppose these conditions hold for any sequence $(\mathcal{H}_n)_{n > 0}$. For convenience, we will not write the subscript $\lambda$ of $T_\lambda$, $H_\lambda$... when there is no ambiguity. Thus, the arguments will be available for the current
distribution, denoted often by $\lambda$, belonging to $\mathcal{H}_n$. Denote as usual $K_\lambda(x) = h^{-1}K(x/h)$. Moreover, let $h^{-1}\tilde{K}_\lambda$ and $h^{-1}\tilde{K}_\lambda$ be the derivatives of $K_\lambda$ and $\tilde{K}_\lambda$ respectively with respects to $h$. For every integer $p$, denote $\int K^{(p)} = \int K^{(p)} dt$.

5.1. Preliminary Lemmas

For convenience, we need to prove first some results that have been stated for the density function by Hall and Marron (1987a and 1987b).

**Lemma 5.1** Under vi–xii and some $n_0 > 0$,

$$0 < \inf_{n \geq n_0, \lambda \in \mathcal{H}_n} n^{1/3} h_\lambda \leq \sup_{n \geq n_0, \lambda \in \mathcal{H}_n} n^{1/3} h_\lambda < \infty. \quad (5.1)$$

For any $\varepsilon > 0$, there exist $\eta = \eta(\varepsilon)$ and $n_1$ such that

$$\inf_{|h-h_\lambda| > cn^{-1/3}} MISE(h, \lambda) \geq (1 + \eta)MISE(h_\lambda, \lambda) \quad (5.2)$$

for all $\lambda \in \mathcal{H}_n$ and all $n$ larger than $n_1$. For some $n_2 > 0$,

$$0 < \inf_{n \geq n_0, \lambda \in \mathcal{H}_n} n^{2/5} MISE^{(2)}(h_\lambda, \lambda) \leq \sup_{n \geq n_0, \lambda \in \mathcal{H}_n} n^{2/5} MISE^{(2)}(h_\lambda, \lambda) < \infty. \quad (5.3)$$

For any $\varepsilon > 0$, there exist $\eta = \eta(\varepsilon)$ and $n_3$ such that

$$\sup_{|h-h_\lambda| \leq cn^{-1/3}} |MISE^{(2)}(h, \lambda) - MISE^{(2)}(h_\lambda, \lambda)| \leq \eta(\varepsilon)n^{-2/3} \quad (5.4)$$

for all $\lambda \in \mathcal{H}_n$ and all $n$ larger than $n_3$, and $\eta(\varepsilon) \to 0$ as $\varepsilon \to 0$.

Denote $\|D_h * \lambda\|^2_2 = \int (K_\lambda * \lambda - \lambda)^2 w$ and $J_{nh} = \|D_h * \lambda\|^2_2 + (nh)^{-1}$.

**Lemma 5.2** Let $\varepsilon$ be a strictly positive constant and $(\lambda_n)_{n \geq 0}$ be a sequence of hazard functions, $\lambda_n$ belonging to $\mathcal{H}_n$ for each $n$. For any $\tilde{h}$, a random bandwidth such that $P_{\lambda_n}(\tilde{h} \in I_n) \to 1$, we have

$$P_{\lambda_n}(\hat{\lambda}(\tilde{h}, \lambda_n) - MISE(\tilde{h}, \lambda_n)|/MISE(\tilde{h}, \lambda_n) > \varepsilon) \to 0.$$
Remark 5.1 The previous result is still true replacing ISE and MISE by their derivative with respects to \( h \). Indeed, the only modifications to the proof is to replace kernel \( K \) by its derivatives with respect to \( h \), which are higher order functions.

Lemma 5.2 is an obvious consequence of Lemma 5.3.

**Lemma 5.3** For all \( \varepsilon > 0 \) and every sequence of hazard functions \((\lambda_n)_{n \geq 0}, \lambda_n \in \mathcal{H}_n\) for each \( n \), we have

\[
\liminf_{n} \inf_{h \in I_{n}} \frac{MISE(h, \lambda_n)}{J_{nh}} > 0,
\]

(5.5)

\[
P_{\lambda_n}\left( \sup_{h \in I_{n}} \frac{|ISE(h, \lambda_n) - MISE(h, \lambda_n)|}{J_{nh}} > \varepsilon \right) \to 0.
\]

(5.6)

The following inequality will be useful to prove the results.

**Lemma 5.4** If \( \lambda \) is \( C^2(\tau) \) and if \( \int (\lambda^{(2)})^2 w \neq 0 \), then there exists a positive constant \( c_0 \) such that, for all \( h \in [0, h_0] \),

\[
\int (K_h * \lambda - \lambda)^2 w \geq c_0 h^4.
\]

To use the latter result "uniformly" over \( \mathcal{H}_n \) and \( n \), we need stronger assumptions.

**Corollary 5.1** Under \( x \), if for any \( \varepsilon > 0 \) there exists \( \eta > 0 \) such that, for all \( x \in \tau, n \) and \( \lambda \in \mathcal{H}_n \),

\[
|t| < \eta \Rightarrow |\lambda^{(2)}(x + t) - \lambda^{(2)}(x)| < \varepsilon,
\]

(5.7)

then there exists a positive constant \( c_0 \) such that, for all \( h \in I_{n} \),

\[
\liminf_{n} \inf_{\lambda \in \mathcal{H}_n} \int (K_h * \lambda - \lambda)^2 w \geq c_0 h^4.
\]

Particularly, the first assumption (5.7) holds under viii.

**Lemma 5.5** For all \( \varepsilon > 0 \),

\[
\sup_{\lambda \in \mathcal{H}_n} P_{\lambda}(|\hat{\lambda}_n - \lambda_n| > \varepsilon n^{-1/5}) \to 0
\]

(5.8)
Proof of Lemma 5.1  Some limited expansions allow us to prove these results. With the notations of Fermanian (1997), note that

\[ MISE(h, \lambda) = \int E[(\hat{\lambda}_h - \lambda)^2 w] = \int E[(\tilde{\xi}_h + B_h + R_{n,h})^2] w \]

\[ = \int \text{Var}\tilde{\xi}_h w + \int B_h^2 w + \epsilon_n(h) \]

where \( S(t) = P(X > t, \delta = 1) \), \( \hat{S}(t) = n^{-1} \sum_{t=1}^{n} 1\{X_i > t, \delta_i = 1\} \), and

\[ \tilde{\xi}_h(x) = \int K_h(x-u) \frac{(H-H)(u-)}{H^2(u-)} \hat{S}(du) \]

\[ - \int K_h(x-u) \frac{\hat{S}-S}{H(u-)}(du) , \]

\[ B_h(x) = D_h \ast \lambda(x) = K_h \ast \lambda(x) - \lambda(x) , \]

\[ R_{n,h}(x) = \int K_h(x-u) \frac{(H-H)(u-)}{H^2(u-)} (\hat{S}-S)(du) \]

\[ - \int K_h(x-u) \frac{(H-H)(u-)}{H^2(u-)} \hat{S}(du) . \]

Under our assumptions, \( \sup_{x \in \mathbb{R}} |R_{n,h}(x)| = O(n^{-1}h^{-1/2}\ln n) \) a.s. (see Fermanian, 1997). First,

\[ \int \text{Var}\tilde{\xi}_h w = \frac{1}{n} \int K_h^2(x-u) \frac{\lambda}{H} (u) w(x) du dx \]

\[ + \frac{1}{n} \int K_h(x-u) K_h(x-v) \frac{\lambda}{H} (u) \frac{\lambda}{H} (v) [H(u \vee v) \]

\[ - 1\{u \geq v\}H(u) - 1\{v \geq u\}H(v)] w(x) du dv dx \]

\[ = \frac{1}{n} \int K_h^2(x-u) \frac{\lambda}{H} (u) w(x) du dx + 0. \]

We deduce

\[ \int \text{Var}\tilde{\xi}_h^{[1]} w = \frac{2}{nh} \int K_h(x-u) \tilde{K}_h(x-u) \frac{\lambda}{H} (u) w(x) du dx \]

\[ = \frac{2}{nh^2} \int K\tilde{K} \cdot \int \frac{\lambda w}{H} + \frac{\phi_1}{nh}, \]

(5.9)
where $\phi_1$ is uniformly bounded on $\mathcal{H}_n$ when $\sup_{x \in T, \lambda \in \mathcal{H}_n} |(\lambda H^{-1})(x)|$ is bounded. Similarly,

$$\int \text{Var} \varepsilon_h^3 \, w = \frac{2}{nh^2} \int [K_h(x - u) \tilde{K}(x - u)$$

$$+ \tilde{K}_h(x - u) - K_h \tilde{K}(x - u)] \frac{\lambda}{H}(u) w(x) du \, dx$$

$$= \frac{2}{nh^3} \int [\tilde{K} + \tilde{K}_h - K \tilde{K}] \cdot \left( \int \frac{\lambda w}{H} + \frac{\phi_2}{nh^2} \right)$$

$$= \frac{2}{nh^3} \int K^2 \cdot \left( \int \frac{\lambda w}{H} + \frac{\phi_2}{nh^2} \right),$$

where $\phi_2$ is uniformly bounded on $n$ and $\mathcal{H}_n$ under vii. Second, it is easy to prove that

$$\int (B_h^2)^{[1]} w = \frac{2}{h} \int (D_h * \lambda)(\tilde{K} * \lambda) w$$

$$= \frac{h^3}{2} \int \tilde{K}^2 \int K^{(2)} \cdot \int (\lambda^{(2)})^2 w + h^4 \phi_3,$$

where $\phi_3$ is uniformly bounded on $n$ and $\mathcal{H}_n$ under vii. Moreover,

$$\int (B_h^2)^{[2]} w = \frac{2}{h^2} \int (\tilde{K} * \lambda)^2 w + \frac{2}{h^2} \int (D_h * \lambda)(\tilde{K} * \lambda - \tilde{K}_h * \lambda) w$$

$$= \frac{h^2}{2} \left[ \int (\tilde{K}^{(2)})^2 + \int K^{(2)} \left( \int \tilde{K}^{(2)} - \int \tilde{K}^{(2)} \right) \right]$$

$$\cdot \int (\lambda^{(2)})^2 w + h^3 \phi_4$$

$$= 3h^2 \int (\tilde{K}^{(2)})^2 \cdot \int (\lambda^{(2)})^2 w + h^3 \phi_4,$$

where $\phi_4$ is uniformly bounded on $n$ and $\mathcal{H}_n$ under viii.

Third, the same calculations as in Lemma A-3 of Fermanian (1999) shows that

$$\varepsilon_n(h) = \left( \frac{1}{n^2 h^2 + \frac{h}{n}} \right) \phi_5 \quad \varepsilon_n^{[1]}(h) = \left( \frac{1}{n^2 h^2 + \frac{h}{n}} \right) \phi_6$$

$$\varepsilon_n^{[2]}(h) = \left( \frac{1}{n^2 h^2 + \frac{1}{n}} \right) \phi_7$$

where $\phi_5$, $\phi_6$ and $\phi_7$ are bounded uniformly on $n$ and $\mathcal{H}_n$ under viii.
Remark 5.2 Note that $\limsup_{n \to \infty} \sup_{\lambda \in \mathcal{H}_n} \sup_{h>1} MISE(h, \lambda)$ is bounded from above by $\limsup_{n \to \infty} \sup_{\lambda \in \mathcal{H}_n} \int \lambda^2 w$. Moreover

$$MISE(h) \geq \frac{1}{nh} \int K^2(i) \frac{\lambda}{H} (x - h t)w(x) dx dt + \left( \frac{1}{n^2 h} + \frac{h^2}{n} \right) \phi_5$$

$$\geq \frac{1}{nh} \int K^2 \cdot \int \frac{\lambda w}{H}$$

$$- \frac{1}{n} \sup_{\lambda \in \mathcal{H}_n} \sup_{x \in T} \left| \int K^2 \cdot \frac{\lambda}{H} \left( \frac{1}{n^2 h} + \frac{h^2}{n} \right) \phi_5.\right.$$ 

Then, under the assumptions vii and ix, there exists a constant $C$ sufficiently small such that, for $n$ sufficiently large and any $\lambda \in \mathcal{H}_n$, $h, \in [Cn^{-1}, n^{-1}]$.

To prove (5.1), note that

$$MISE[1](h) = 0 \iff 2 \frac{\lambda w}{H} + \phi_1 + h^3 \left( \int K^2 \right)^2 \left( \int (\lambda(2))^2 w \right)$$

$$+ h^4 \phi_3 + \left( \frac{1}{n^2 h^2} + \frac{h^2}{n} \right) \phi_5 = 0.$$

Then, under vii and viii,

$$\sup_{\lambda \in \mathcal{H}_n} \int 2 \int K^2 \cdot \frac{\lambda w}{H} + nh^5 \left( \int K^2 \right)^2 \left( \int (\lambda(2))^2 w \right) \rightarrow 0.$$

We deduce the left hand of (5.1) under viii and ix, and the right hand of (5.1) under x.

Note that we have found the asymptotic equivalent of $h\lambda$.

$$nh^5 \sim \frac{\int K^2 \cdot \int \lambda w / H}{\left( \int K^2 \right)^2 \cdot \int (\lambda(2))^2 w}$$

Denote $C(\lambda)$ the constant such that $h\lambda \sim n^{-1/5} C(\lambda)$. The proof of (5.3) is similar. Thus,

$$MISE[2](h) = \frac{1}{nh^3} \left\{ \int K^2 \cdot \frac{\lambda w}{H} + h \phi_2 + 3nh^5 \left( \int K^2 \right)^2 \left( \int (\lambda(2))^2 w \right)$$

$$+ nh^6 \phi_4 + \left( \frac{1}{n} + h^2 \right) \phi_7 \right\}.$$
Deduce

\[ n^{2/5} MISE^{[2]}(h, \lambda) \sim C st \left( \int \frac{\lambda w}{H} \right)^{2/5} \cdot \left( \int (\lambda^{(2)})^2 w \right)^{3/5}, \]

where \(C st\) is a constant strictly positive depending only on \(K\). Under the previous assumptions, deduce (5.3).

We prove easily (5.4) using the previous expansion of \(MISE^{[2]}\) to bound each term. Particularly, if \(\varepsilon > 0\) is small enough and if |\(h - h_\lambda| \leq \varepsilon n^{-1/5}\) then, for \(n\) sufficiently large,

\[ \frac{\phi_2}{nh^2} \leq \frac{4\phi_3}{nh_\lambda^2} < 8\phi_2 C(\lambda)^{-2} n^{-3/5} = o(n^{-2/5}) \]

Moreover, under vii and viii, we can choose \(\eta(\varepsilon)\) such that

\[ \left| 3(h^2 - h_\lambda^2) \left( \int K^{(2)} \right)^2 \cdot \int (\lambda^{(2)})^2 w \right| \]
\[ \leq 9|h - h_\lambda| \left( \int K^{(2)} \right)^2 \cdot \int (\lambda^{(2)})^2 w \cdot h_\lambda \]
\[ \leq 9\varepsilon n^{-2/5} \left[ \left( \int K^{(2)} \right)^2 \cdot \int (\lambda^{(2)})^2 w \right]^{4/5} \cdot \left[ \int K^2 \cdot \int \frac{\lambda w}{H} \right]^{1/5} \]
\[ \leq \eta(\varepsilon)n^{-2/5}/10. \]

It remains to prove (5.2). If \(|h - h_\lambda| > \varepsilon n^{-1/5}\), a Taylor Laplace expansion provides

\[ MISE(h) = MISE(h_\lambda) + 0 \]
\[ + (h - h_\lambda)^2 \int_0^1 (1 - t) MISE^{[2]}(t(h - h_\lambda) + h_\lambda) dt \]

or

\[ \frac{MISE(h) - MISE(h_\lambda)}{MISE(h_\lambda)} \geq \frac{\varepsilon^2 n^{-2/5}}{MISE(h_\lambda)} \int_0^1 (1 - t) MISE^{[2]}(t(h - h_\lambda) + h_\lambda) dt. \]

Note that the first part of the current lemma implies that \(\lim sup_n \sup_{\lambda \in \mathcal{H}} n^{4/5} MISE(h_\lambda)\) is bounded. Set \(h_t = t(h - h_\lambda) + h_\lambda\). The
previous expansion of $MISE^{[2]}$ shows that, if $h < h_\lambda$, then for each $t \in [0, 1]$ and $n$ large enough,

$$\inf_{\lambda \in \mathcal{N}_n} MISE^{[2]}(h_t) \geq \frac{Cst}{nh^3} > \frac{Cst}{nh^3} \geq Cst \cdot n^{-2/5} > 0.$$ 

Moreover, if $n^{-1}h > h_\lambda$, then for each $t \in [0, 1]$ and $n$ large enough,

$$\inf_{\lambda \in \mathcal{N}_n} MISE^{[2]}(h_t) \geq L^2 Cst > h^3 Cst > n^{-2/5} Cst > 0.$$ 

Deduce

$$\frac{MISE(h) - MISE(h_\lambda)}{MISE(h_\lambda)} \geq \epsilon^2 Cst > 0 \quad (5.10)$$

and the lemma follows.

**Proof of Lemma 5.2** As previously, the index $n$ in the sequence of hazard functions $(\lambda_n)_{n>0}$ will be forgotten. For $n$ large enough,

$$P_{\lambda} \left( \sup_{h \in I_n} |\text{ISE}(h, \lambda) - MISE(h, \lambda)|/MISE(h, \lambda) > \epsilon \right) \leq P_{\lambda} \left( \sup_{h \in I_n} |\text{ISE}(h, \lambda) - MISE(h, \lambda)|/J_{nh} > \epsilon \alpha \right) \xrightarrow{n \to \infty} 0,$$

where $\alpha = 2^{-1} \lim \inf_{n} \inf_{h \in I_n} MISE(h, \lambda)/J_{nh}$, and Lemma 5.2 follows from Lemma 5.3.

**Proof of Lemma 5.3** Note that, for any $h \in I_n$, we have

$$MISE(h, \lambda) = \int (D_h * \lambda)^2 w + \int \text{Var}(\tilde{\xi}_h)w + \int E[R_{nh}^2]w$$

$$+ 2 \int E[\tilde{\xi}_h R_{nh}]w + 2 \int E[R_{nh}]D_h * \lambda w. \quad (5.11)$$

We have found previously that

$$\int \text{Var}(\tilde{\xi}_h)w = \frac{1}{nh^2} \int K^2 \left( \frac{x - u}{h} \right) \frac{\lambda}{H}(u)w(x)du dx. \quad (5.12)$$
When $h$ is "small", more precisely $h < \eta$, a Taylor expansion provides under $\xi$ where $\xi_0$ is uniformly bounded on $\mathcal{H}$ and $\mathcal{H}_n$. Then,

$$\int \text{Var}(\xi_h)w = \int K^2 \cdot \int \frac{\lambda w}{H} \cdot \frac{1}{nh} + \phi_0 \cdot \frac{n}{n},$$

where $\phi_0$ is uniformly bounded on $n$ and $\mathcal{H}_n$. Then,

$$\int \text{Var}(\xi_h)w \geq \frac{\text{Cst}}{nh}$$

for all $n$, $\lambda \in \mathcal{H}_n$ and $h < \eta$. Hence, for $n$ large enough, $h < \eta$ and

$$\int (K_h \cdot \lambda - \lambda)^2w + \int \text{Var}(\xi_h)w \geq \text{Cst} \cdot J_{nh}, \quad (5.13)$$

where $\text{Cst}$ is a constant independent from $n$ and $\lambda \in \mathcal{H}_n$. Thus the result (5.5) is proved if we state that

$$\lim \inf \max_{n} J_{nh}^{-1} \left| \int E[R_{nh}^2]w + 2 \int E[\xi_h R_{nk}]w \right| \geq 2 \int E[R_{nk}](K_h \cdot \lambda_n - \lambda_n)w = 0. \quad (5.14)$$

This can be made by the same arguments as Lemma A-3 of Fermanian (1999), with no derivations with respects to $h$. This provides

$$\int E[R_{nh}^2]w + 2 \int E[\xi_h R_{nk}]w + 2 \int E[R_{nk}](K_h \cdot \lambda_n - \lambda_n)w = O(n^{-2}h^{-1} + h^2n^{-1}).$$

The previous upper bound is true uniformly with respects to $h < h_0$, $n$ and $\lambda \in \mathcal{H}_n$. Then (5.14) holds because $nhJ_{nh} > 1$ and because

$$\sup_{h \in \mathcal{H}, \lambda \in \mathcal{H}_n} \frac{h^2}{nh} \to 0.$$ 

This last point is obvious when $h < n^{-\nu}$, which we have supposed. In fact, it is still true if $h$ is bounded by a constant (see Fermanian, 1998).
We have to prove now (5.6). Since $K$ belongs to $C^1(\mathbb{R})$, it is sufficient to prove the result for $h \in J_n$, where $J_n$ is a finite subset of $I_n$, $n = n', d$ being large enough. In the remainder of the proof, we will suppose that $h \in J_n$. With the previous notations, we have

$$(ISE - MISE)(h, \lambda) = \int \{\xi_h^2 - E[\xi_h^2]\} w + 2 \int (K_h * \lambda - \lambda) \xi_h w$$

$$+ \int \{R_{nh}^2 - E[R_{nh}^2]\} w$$

$$+ 2 \int (K_h * \lambda - \lambda) \cdot \{R_{nh} - E[R_{nh}]\} w$$

$$+ 2 \int (\xi_h R_{nh} - E[\xi_h R_{nh}]) w.$$  \hspace{1cm} (5.15)

We will discuss each term of the previous expansion. Recall that

$$\xi_h(x) = \int K_h(x - u) \left[ -\frac{(\check{S} - S)(du)}{H(u)} + S(du) \cdot \frac{(\check{H} - H)}{H^2} (u) \right].$$  \hspace{1cm} (5.16)

Note that

$$\int (\xi_h^2 - E[\xi_h^2]) w = T_1 - E[T_1] + T_2 - E[T_2] + T_3 - E[T_3]$$

$$T_1 = \int K_h(x - u)K_h(x - v)$$

$$\cdot \frac{(\check{S} - S)(du)}{H(u)} \cdot \frac{(\check{S} - S)(dv)}{H(v)} \cdot w(x) dx$$

$$T_2 = 2 \int K_h(x - u)K_h(x - v) \cdot \frac{(S - S)(du)}{H(u)}$$

$$\cdot S(du) \cdot \frac{(\check{H} - H)}{H^2} (v) \cdot w(x) dx$$

$$T_3 = \int K_h(x - u)K_h(x - v)$$

$$\cdot \frac{(\check{H} - H)(du)}{H^2} (u) \cdot \frac{(\check{H} - H)}{H^2} (v) \cdot S(du) S(dv) w(x) dx$$
We will discuss $T_1$. The two other terms can be dealt similarly.

$$T_1 - E[T_1] = \frac{1}{n} \int K_h^2(x - u) \frac{(\hat{S} - S)(du)}{H^2(u)} w(x) du dx$$

$$+ \int_{\text{w.p}0} K_h(x - u) K_h(x - v) \frac{(\hat{S} - S)(du)}{H(u)} \cdot \frac{(\hat{S} - S)(dv)}{H(v)} w(x) dx.$$ 

First, from Lemma B.3 of Fermanian (1999),

$$\sup_{\lambda \in \mathcal{H}_n} \left( \sup_{h \in J_n \cap J_{nh}} \frac{1}{n} \left| \int K_h^2(x - u) \frac{(\hat{S} - S)(du)}{H^2(u)} w(x) du dx \right| > \varepsilon \right)$$

$$\leq \sup_{\lambda \in \mathcal{H}_n} n^d \cdot \frac{1}{(n \lambda e J_{nh})^{2p}} E_{\lambda} \left[ \left( \int (K^2h(x - u) \frac{(\hat{S} - S)(du)}{H^2(u)} w(x) du dx \right)^{2p} \right]$$

$$\leq n^d \cdot \left( \frac{C_{s0}}{n} \right)^p \cdot \frac{1}{(n \lambda e J_{nh})^{2p}}.$$ 

where $C_{s0}$ is independent from $n$ and $\lambda \in \mathcal{H}_n$. Note that, $w$ and $K$ being compactly supported, we can suppose that the variable $u$ belongs to $\tau$, and that $H$ is uniformly bounded from below on $\tau$. Choosing $p > d + 1$, we have obtained that

$$\sup_{\lambda \in \mathcal{H}_n} \left( \sup_{h \in J_n \cap J_{nh}} \frac{1}{n} \left| \int K_h^2(x - u) \frac{(\hat{S} - S)(du)}{H^2(u)} w(x) du dx \right| > \varepsilon \right) n \rightarrow 0. \tag{5.17}$$

Second, using Lemma B.3 of Fermanian (1999),

$$\sup_{\lambda \in \mathcal{H}_n} \left( \sup_{h \in J_n \cap J_{nh}} \int_{\text{w.p}0} K_h(x - u) K_h(x - v) \frac{(\hat{S} - S)(du)}{H(u)} \cdot \frac{(\hat{S} - S)(dv)}{H(v)} w(x) dx \right) > \varepsilon$$
where $C_{st}$ is independent from $n$ and $\lambda \in \mathcal{H}_a$. Setting $p$ sufficiently large, we have obtained that

$$
sup_{\lambda \in \mathcal{H}_a} \left( \sup_{k \in J_{A_{\lambda}}} \int_{A_{\lambda}} K_h(x-u)K_h(x-v) \frac{(\bar{S} - S)(du)}{H(u)} \cdot \frac{(\bar{S} - S)(dv)}{H(v)} w(x)dx \right)^{2p} \leq C_{st}^{2p} n^d \cdot \frac{1}{(e/n_{nh})^{2p}} \left[ \left( \frac{p}{n} \right)^{4p} h^{-p+1/2} + \left( \frac{p}{n} \right) h^{-p} \right].
$$

Also

$$
\lim_{n \to \infty} \max_{k \in J_{A_{\lambda}}} \left| \int_{A_{\lambda}} (\bar{S} - S) w(x)dx \right| \to 0.
$$

The same result is true for $T_2$ and $T_3$. Then

$$
\lim_{n \to \infty} \max_{k \in J_{A_{\lambda}}} \frac{1}{n} \left| \int_{A_{\lambda}} (\bar{S}_h^2 - E[\bar{S}_h^2]) w(x)dx \right| = 0.
$$

Hence the first part of the result (5.6). To deal with the second term of (5.15), note that

$$
2 \int (K_h \ast \lambda - \lambda) \bar{S}_h w = \frac{2}{n} \sum_{i=1}^{n} Z_{ih},
$$

where

$$
Z_{ih} = \int D_h \ast \lambda(x) K_h(x - X_i) \frac{\delta_{i}}{H(X_i)} w(x)dx
$$

$$
- \int D_h \ast \lambda(x) K_h(x - u) 1{\{X_i \geq u\}} \frac{\lambda}{H(u)} w(x)du dx
$$
Evidently, there exists a constant $M_0$ such that
\[
\sup_n \sup_{h \in H_n} \sup_{\lambda \in \mathcal{H}_n, x \in \mathcal{E}} |D_h \ast \lambda(x)| \leq M_0 h^2. \tag{5.20}
\]
Thus, under vi and vii and for $n$ large enough \(i.e. h\) small enough,
\[
|Z_h| \leq M_0 h^2 \int |K|(t) \frac{w(X_t + ht)}{H(X_t)} dt
+ M_0 h^2 \int |K|(t) \frac{\lambda}{H}(u)w(u + ht)du dt
\leq Cst \cdot h^2 \left[ \frac{1}{H(X_t)} + \int \frac{\lambda}{H} \right] \leq Cst \cdot h^2.
\]
Moreover,
\[
\text{Var } Z_h = \int D_h \ast \lambda(x)D_h \ast \lambda(y)K_h(x - t)K_h(y - t)
- 2 \int D_h \ast \lambda(x)D_h \ast \lambda(y)K_h(x - t)K_h(y - u)
+ \int D_h \ast \lambda(x)D_h \ast \lambda(y)K_h(x - u)K_h(y - v)H(u \lor v)
\]
\[
= V_1 + V_2 + V_3.
\]
Each of these terms can be bounded by some constant times $\|D_h \ast \lambda\|_2^2$ (denoted by $u^2_2$). For example, we precise this point for $V_1$. For $n$ sufficiently large, under viii, we have
\[
|V_1| \leq Cst \int D_h \ast \lambda(t + hx)D_h \ast \lambda(t + hy)
\]
\[
\int ( \frac{\lambda}{H} (t) 1\{|x| \leq A, |y| \leq A, t \in \tau\} w(t + hx)w(t + hy) dt dx dy
\leq Cst \int \left[ \int (D_h \ast \lambda)^2(t_1 + hx)w(t_1 + hx) \frac{\lambda}{H} (t_1) dt_1 \right]^{1/2}
\]
Thus, we can make the same development as Stone (1984, Lemma 2). It follows from the Bernstein's inequality that

\[
P_\lambda \left( 2 \int D_h \ast \lambda \xi_h w \, \left| \frac{\lambda}{h} \right| > t \right) \leq 2 \exp \left( - \frac{\kappa \rho}{1 + \rho/3} \right),
\]

(5.21)

where \( 0 \leq \rho \leq t/u_h^2 \) and \( \kappa = Cst \cdot n^{- 2 \nu} \). If \( u_h \geq n^{- 1/2} \), set \( t = n^{- 1/2} u_h \) and \( \rho = n^{- 1/2} u_h^{- 1} \). Thus, if \( h \in J_n \), then \( \kappa \rho \geq Cst n^{2 \nu} \) (the constant is a function of \( M_0 \) only, independent from \( n \) and \( \lambda \in \mathcal{H}_n \)). Otherwise, if \( u_h < n^{- 1/2} \), set \( t = n^{- 1} \) and \( \rho = 1 \). Thus, if \( h \in J_n \), then \( \kappa \rho \geq Cst n^{2 \nu} \).

In either case, it follows

\[
P_{\lambda, n} \left( \sup_{h \in J_n} 2 \int D_h \ast \lambda \xi_n w \, \left| \frac{\lambda}{h} \right| > n^{- 1/2} u_h + n^{- 1} \right) \rightarrow 0
\]

Thus, to verify the second part of (5.15), it is sufficient to show that

\[
\liminf_{n} \inf_{\lambda \in \mathcal{H}_n} \inf_{h \in J_n} \frac{J_{nh}}{n^{- 1/2} u_h + n^{- 1}} \geq \varepsilon^{- 1}.
\]

If \( n^{- 1/2} u_h \leq n^{- 1} \), then the previous ratio is less than \( (nh)^{- 1/2} (2n^{- 1}) \) or \( n^\varepsilon/2 \) which goes to infinity. Else, this ratio is less than \( u_h^2/(2n^{- 1/2} u_h) \) or \( 1/2 \). Hence, for all \( \varepsilon > 0 \), we have

\[
P_{\lambda, n} \left( \max_{h \in J_n} 2 \int D_h \ast \lambda \xi_n w \, \left| \frac{\lambda}{h} \right| > \varepsilon \right) \rightarrow 0.
\]

(5.22)

To obtain the result (5.6), it is now sufficient to prove

\[
P_{\lambda, n} \left( \max_{h \in J_n} \left| \int R_{nh}^2 w + 2 \int (K_h \ast \lambda_h - \lambda_n) R_{nh} w + 2 \int \xi_h R_{nh} w \right| > \varepsilon J_{nh} \right) \rightarrow 0.
\]

(5.23)
Indeed, recall that the argument about the expectation of the terms inside the brackets has been done in (5.14). To deal with $\int R_{nh}^2 w$, it is sufficient to use Lemma A.2 of Fermanian (1997). After some manipulations, this amounts to deal with some quantities like

$$P_\lambda \left( \max_{h \in J_n} \left| \int K_h(x-u)K_h(x-v) \cdot \frac{(\hat{H} - H)}{H^2} (u) \cdot \frac{(\hat{H} - H)}{H^2} (v) \cdot (\hat{S} - S) (du) (\hat{S} - S) (dv) w(x) dx \right| > \varepsilon \right)$$

$$\leq n^d \max_{h \in J_n} \varepsilon (eJ_n)^{-2p} E \left[ \left( \int K_h(x-u)K_h(x-v)(\hat{H} - H) (u) (\hat{H} - H) (v) H^{-2}(u) H^{-2}(v) \cdot (\hat{S} - S) (du) (\hat{S} - S) (dv) w(x) dx \right)^{2p} \right]$$

$$\leq n^d \max_{h \in J_n} \frac{1}{(eJ_n)^{2p}} C_{st}^d (4p)^{4p} n^{-4p} h^{-2p} \leq n^d \left( \frac{C_{st}^d}{n} \right)^{2p}.$$ 

Under vi and vii, the constant is independent from $h$, $n$ and $\lambda \in \mathcal{H}_n$. Thus, choose e.g. $p > d$ to obtain the result. The three other terms of $\int R_{nh}^2 w$ are dealt similarly.

Let us discuss $\int D_h \ast \hat{\lambda} \cdot R_{nh} w$, which is the sum of two terms. To deal with the first part, $\int D_h \ast \hat{\lambda} \cdot \hat{R}_{nh} w$ (notations of Fermanian, 1997), use Lemma B.3 of Fermanian (1999). Thus

$$P_\lambda \left( \max_{h \in J_n} \left| \int D_h \ast \lambda(x) \hat{R}_{nh} (x) w(x) dx \right| > \varepsilon \right)$$

$$\leq n^d \max_{h \in J_n} \varepsilon (eJ_n)^{-2p} E \left[ \left( \int D_h (x-u) \lambda(u) \cdot K_h (x-v) \frac{(\hat{H} - H)}{H^2} (v) \cdot (\hat{S} - S) (dv) w(x) dx \right)^{2p} \right]$$

$$\leq n^d \max_{h \in J_n} C_{st}^d \left( \frac{p}{n \varepsilon J_n} \right)^{2p} \cdot \left( \frac{ph}{n} \right)^p \leq n^d C_{st}^d \cdot \left( \frac{p^3}{n} \right)^p \rightarrow 0,$$

where, as previously, $C_{st}$ denotes a constant independent from $n$, $h$, $\lambda \in \mathcal{H}_n$ and $p > d$. The same technique can be used to discuss the
This technique provides the corresponding results for $\int \tilde{\xi}_h R_{nh} w$. In this case, the upper bound is now $n^p \sup_{h \in [\hat{h}, h_0]} \int_{-h_0 \hat{h}}^{h_0} n^{-3p} h^{-p}$ which tends to zero in every case, proving the result.

**Proof of Lemma 5.4** This result is similar to Lemma 1 of Stone (1984). We cannot use directly his methodology since the hazard function does not always belong in $L^1(\mathbb{R})$. Thus the Fourier transform does not exist necessarily. For convenience, we have supposed that the hazard functions are $C^2(\mathbb{R})$. Hence our result is weaker than Stone's one.

A Taylor expansion provides

$$
\int (K_h * \lambda - \lambda)^2 w = \int \left[ \int K(t) \lambda(x - ht) dt - \lambda(x) \right]^2 w(x) dx
$$

$$= h^4 \int \left[ \int 1 \{u \in [0, 1]\} K(t) \lambda^{(2)}(x - hu(t))^2 (1 - u) du dt \right]^2 w(x) dx.
$$

Since $\lambda^{(2)}$ is uniformly continuous on $[-h_0 \hat{h}, \tau_0 + h_0 \hat{h}]$, for all $\varepsilon > 0$, there exists $\eta > 0$ such that, for all $t \in [-\hat{h}, \hat{h}], x \in \tau, u \in [0, 1]$ and $h < \eta$,

$$|\lambda^{(2)}(x - hu(t)) - \lambda^{(2)}(x)| < \varepsilon.
$$

Hence

$$
\int (K_h * \lambda - \lambda)^2 w \geq \frac{h^4}{4} \left( \int K^{(2)} \right)^2 \left[ \int (\lambda^{(2)})^2 w - 2 \varepsilon \int |\lambda^{(2)}(x)| w(x) dx \right]
$$

$$\geq \frac{h^4}{4} \left( \int K^{(2)} \right)^2 \left[ (1 - \varepsilon) \int (\lambda^{(2)})^2 w - \varepsilon \right]
$$

$$\geq \frac{h^4}{8} \left( \int K^{(2)} \right)^2 \int (\lambda^{(2)})^2 w,
$$

for $\varepsilon$ small enough. We have use the fact that $2 \int |\lambda^{(2)}| w \leq 1 + \int (\lambda^{(2)})^2 w$ because $w \leq 1$. Thus, the result is proved when $h < \eta$. If $h \in [\eta, h_0]$, then
this is true due to the continuity of the considering function with respects to $h$.

**Proof of Lemma 5.5** It is sufficient to show that for any sequence of choices $\lambda = \lambda_1, \ldots, \lambda_n \in \mathcal{H}_n$ and for each $\varepsilon > 0$,

$$P_n(\hat{h}_n - h_n \geq \varepsilon n^{-1/5}) \to 0.$$  \hfill (5.25)

Invoking Lemma 5.1, the event $|\hat{h}_n - h_n| > \varepsilon n^{-1/5}$ implies that there exists $\eta(\varepsilon)$ (independent from each $\lambda \in \mathcal{H}_n$) such that $MISE(h_n, \lambda) \geq (1 + \eta)MISE(h_n, \lambda)$. Moreover, we have anyway

$$MISE(h_n, \lambda) \leq MISE(h_n, \lambda)$$
$$\leq MISE(h_n, \lambda) + (MISE - ISE)(\hat{h}_n, \lambda)$$
$$+ ISE(\hat{h}_n, \lambda) - ISE(h_n, \lambda) + (ISE - MISE)(h_n, \lambda).$$

Thus in this case,

$$1 + \eta \leq \frac{MISE(\hat{h}_n, \lambda)}{MISE(h_n, \lambda)} \leq 1 + \left(\frac{MISE - ISE}{MISE}\right)(\hat{h}_n, \lambda)$$
$$\cdot \left[\frac{MISE(\hat{h}_n, \lambda)}{MISE(h_n, \lambda)} + \left(\frac{ISE - MISE}{MISE}\right)(h_n, \lambda)\right],$$

and then

$$\frac{MISE(\hat{h}_n, \lambda)}{MISE(h_n, \lambda)} \leq \left[1 + \left(\frac{MISE - ISE}{MISE}\right)(h_n, \lambda)\right]$$
$$\cdot \left[1 - \left(\frac{ISE - MISE}{MISE}\right)(\hat{h}_n, \lambda)\right]^{-1}.$$  \hfill (5.27)

Due to Lemma 5.2, if $\alpha$ is any positive number $< 1$,

$$P_n(|\hat{h}_n - h_n| > \varepsilon n^{-1/5})$$
$$\leq P_n\left[1 + \left(\frac{MISE - ISE}{MISE}\right)(h_n, \lambda)\right]$$
$$\cdot \left[1 - \left(\frac{ISE - MISE}{MISE}\right)(\hat{h}_n, \lambda)\right]^{-1} \geq 1 + \eta$$
$$\leq P_n\left(1 + \eta \leq \frac{1 + \alpha}{1 - \alpha}, \left(\frac{ISE - MISE}{MISE}\right)(h_n, \lambda)\right)$$
5.2. Proofs of Main Theorem: The ISE Case

Supposing (2.3), equality (2.2) may be easily proved as in Hall and Marron (1987, p. 171). Use the identity

\[ \text{ISE}(\hat{h}, \lambda) - \text{ISE}(\hat{h}_{\lambda}, \lambda) = \frac{1}{2} (\hat{h} - \hat{h}_{\lambda})^2 \text{ISE}^{\text{II}}(h^*, \lambda), \]

where \( h^* \) lies between \( \hat{h} \) and \( \hat{h}_{\lambda} \). Now apply (5.3) and (5.4) of Lemma 5.1, Lemma 5.5 and Remark 5.1.

To prove (2.3), use the previous lemmas with \( \mathcal{H}_n = (\lambda_0, \lambda_1) \). Note that \( \lambda_i \) is implicitly \( n \)-dependent and that \( \mathcal{H}_n \) satisfies conditions vi-xii.

Consider \( \tilde{h} = \tilde{h}_{\lambda} \) where

\[ \lambda \in \arg \min_{\lambda \in \mathcal{H}_n} |\tilde{h} - \tilde{h}_{\lambda}|. \]

Thus, if \( \lambda \in \mathcal{H}_n \),

\[ |\hat{h} - \hat{h}_{\lambda}| \leq |\hat{h} - \tilde{h}| + |\tilde{h} - \hat{h}_{\lambda}| \leq 2|\hat{h} - \hat{h}_{\lambda}|. \]

Therefore result (2.2) will follow if we prove that

\[ \lim_{\varepsilon \to 0} \inf_{n \to \infty} \lambda \in \mathcal{H}_n \max \text{P} \lambda (|\hat{h} - \hat{h}_{\lambda}| \geq \varepsilon n^{-1/10}) \geq \rho. \quad (5.28) \]

Indeed, for \( n \) large enough and every \( \varepsilon > 0 \), we have

\[ P \lambda \left( \frac{|\hat{h} - \hat{h}_{\lambda}|}{\hat{h}_{\lambda}} > \varepsilon n^{-1/10} \right) \leq P \lambda (|\hat{h} - \hat{h}_{\lambda}| > \varepsilon n^{-1/10} (\hat{h}_{\lambda} - c n^{-1/3} / 3)), \]

\[ |\hat{h}_{\lambda} - h_{\lambda}| \leq c n^{-1/3} / 3 + P \lambda (|\hat{h}_{\lambda} - h_{\lambda}| > c n^{-1/3} / 3), \]

\[ \leq P \lambda (|\hat{h} - \hat{h}_{\lambda}| > \varepsilon n^{-1/10} \cdot c n^{-1/3} / 3) + P \lambda (|\hat{h}_{\lambda} - h_{\lambda}| > c n^{-1/3} / 3), \]

where \( c = \lim \inf_{\lambda \in \mathcal{H}_n} n^{1/3} h_{\lambda} \) (invoke Lemmas 5.1 and 5.5).
It is easy to show that, for each \( \lambda \in \mathcal{H}_n \), there exists \( h^*_\lambda \) such that
\[
|h^*_\lambda - \hat{h}| \leq |\hat{h} - \hat{h}_\lambda| \quad \text{and}
\]
\[
\hat{h}_\lambda - \hat{h} = \frac{2}{\hat{h} \cdot ISE^{[2]}(\hat{h}_\lambda, \lambda)} \int (\lambda - \lambda) \hat{\mu}_h w, \tag{5.29}
\]
where \( \hat{\mu}_h(x) = \hat{h}_\lambda^{[1]}(x) = n^{-1} \sum_i \hat{k}_h(x - X_i) \delta_i \hat{h}^{-1}(X_i -) \).

If the three following lemmas are proven, under the assumptions of Theorem 2.1, the proof is completed.

**Lemma 5.6** We have
\[
\lim_{a \to 0, b \to \infty} \lim_{n \to \infty} \inf_{\lambda \in \mathcal{H}_n} P_{\lambda}(an^{-1/5} \leq \hat{h}_\lambda \leq bn^{-1/5}) = 1.
\]

**Lemma 5.7** For all \( 0 < a < b < \infty \), we have
\[
\lim_{A \to \infty} \limsup_{n \to \infty} \max_{\lambda \in \mathcal{H}_n} P_{\lambda} \left( \inf_{n^{-1/5} h \in [a,b]} |ISE^{[2]}(h, \lambda)| > An^{-2/5} \right) = 0.
\]

**Lemma 5.8** For all \( 0 < a < b < \infty \), we have
\[
\lim_{\varepsilon \to 0} \liminf_{n \to \infty} \max_{\lambda \in \mathcal{H}_n} P_{\lambda} \left( \inf_{n^{-1/5} h \in [a,b]} \left| \int (\lambda - \hat{\lambda}) \hat{\mu}_h w \right| > \varepsilon n^{-3/10} \right) \geq p.
\]

Indeed, (5.28) follows from
\[
P_{\lambda}(h - \hat{h}_\lambda \geq \varepsilon n^{-3/10})
\]
\[
= P_{\lambda} \left( \left| \int (\lambda - \hat{\lambda}) \hat{\mu}_h w \right| > \varepsilon n^{-1/10} \right)
\]
\[
\geq P_{\lambda} \left( \left| \int (\lambda - \hat{\lambda}) \hat{\mu}_h w \right| > \varepsilon n^{-1/10} bn^{-1/5} An^{-2/5} \right)
\]
\[
- P_{\lambda}(n^{-1/5} \hat{h} \notin [a,b])
\]
\[
- P_{\lambda} \left( \inf_{n^{-1/5} h \in [a,b]} ISE^{[2]}(h, \lambda) > An^{-2/5} \right).
\]
Proof of Lemma 5.6  For all positive numbers $a$ and $b$, we have

$$P_x(a < n^{1/5}h < b) \leq P_x(a/2 < n^{1/5}h < 2b) + \sup_{\lambda \in \mathcal{F}_{4\lambda}} P_x(|\bar{\lambda} - \lambda| > n^{-1/5}a/2)$$

and symmetrically

$$P_x(a < n^{1/5}h < b) \leq P_x(a/2 < n^{1/5}h < 2b) + \sup_{\lambda \in \mathcal{F}_{4\lambda}} P_x(|\bar{\lambda} - \lambda| > n^{-1/5}a/2)$$

By Lemma 5.5, deduce

$$\liminf_{n} \min_{\lambda \in \mathcal{F}_{4\lambda}} P_x(a < n^{1/5}h < b) \leq \liminf_{n} \min_{\lambda \in \mathcal{F}_{4\lambda}} P_x(a/2 < n^{1/5}h < 2b) \leq \liminf_{n} \sup_{\lambda \in \mathcal{F}_{4\lambda}} P_x(a/4 < n^{1/5}h < 4b)$$

The result follows by Lemma 5.1.

Proof of Lemma 5.7  By the proof of Lemma 5.1, there exists a constant $C_0$ such that, for $n$ sufficiently large,

$$\sup_{\lambda \in \mathcal{F}_{4\lambda}} \sup_{n^{1/5}h \in [a,b]} |MISE^{[2]}(h, \lambda)| < C_0n^{-2/5}.$$  

Thus, it is sufficient to show that for every $\varepsilon > 0$

$$\sup_{\lambda \in \mathcal{F}_{4\lambda}} \left( \sup_{n^{1/5}h \in [a,b]} |(ISE^{[2]} - MISE^{[2]})(h, \lambda)| > \varepsilon n^{-2/5} \right) \to 0.$$

The proof of the latter result can be lead exactly like in Lemma 5.2. The preliminary results are independent from the kernel functions $K$, $\tilde{K}$ or $\hat{K}$. Here we can choose $c = 4$ and $J_{nh} \sim n^{-4/5}$. The power $n^{-2/5}$ on the right side of the inequality is of the order of $J_{nh}^2$, where the extra factor $h^2$ comes from the two first derivatives of the objective functions.

Proof of Lemma 5.8  If $\lambda \neq \lambda, \lambda \in \{\lambda_0, \lambda_1\}$, then

$$\left| \int (\lambda - \lambda)\hat{\mu}_n w \right| = n^{-1/2} \left| \int \alpha\lambda_0\hat{\mu}_n w \right|.$$
By the Neyman-Pearson lemma,

\[
\max_{\lambda \in \{\lambda_0, \lambda_1\}} P_\lambda(\bar{\lambda} \neq \lambda) \geq \frac{1}{2} \left( P_{\lambda_0}(\bar{\lambda} = \lambda_1) + P_{\lambda_1}(\bar{\lambda} = \lambda_0) \right) \\
\geq \frac{1}{2} \left( P_{\lambda_0}(\bar{f} = f_1) + P_{\lambda_1}(\bar{f} = f_0) \right),
\]

where \(\bar{f}\) is the likelihood ratio rule for deciding between \(f_0\) and \(f_1\) (the density functions associated respectively with \(\lambda_0\) and \(\lambda_1\)). Note that we need at this step a one-to-one correspondence between hazard and density functions. Now

\[
P_{\lambda_0}(\bar{f} = f_1) = P_{\lambda_0} \left[ \sum_i \ln \left\{ 1 + n^{-1/2} \alpha(X_i) \right\} - n^{-1/2} \int_0^{X_i} \alpha \lambda_0 > 0 \right]
\]

\[
= P_{\lambda_0} \left[ n^{-1/2} \sum_i \alpha(X_i) - \frac{1}{2} n^{-1} \sum_i \alpha(X_i)^2 \right.
\]

\[\left. - n^{-1/2} \int_0^{X_i} \alpha \lambda_0 + o(1) > 0 \right]
\]

\[
\rightarrow 1 - \phi(\sigma/2) = p,
\]

and similarly \(P_{\lambda_1}(\bar{f} = f_0) \rightarrow p\). Therefore

\[
\lim \inf_{n \to \infty} \max_{\lambda \in \{\lambda_0, \lambda_1\}} P_\lambda(\bar{\lambda} \neq \lambda) \geq p,
\]

(5.30)

and Lemma 5.8 will follow if we prove that

\[
\lim_{\epsilon \to 0} \lim \inf_{n \to \infty} \max_{\lambda \in \{\lambda_0, \lambda_1\}} P_\lambda \left( \inf_{\alpha \in [0, \alpha]} \right| \int \alpha \lambda_0 \hat{\mu}_n \right| > \epsilon n^{-2/3} \right) = 1. \quad (5.31)
\]

This last result can be proved easily with some minor modifications of Hall and Marron (1991). Note that, since \(\hat{H}^{-1}(X_i) \geq n^{-1},\)

\[
\hat{H}^{-1}(X_i) = H^{-1}(X_i) + H^{-2}(H - \hat{H})(X_i) + \cdots + H^{-4}(H - \hat{H})^4(X_i)
\]

\[
\geq H^{-1}(X_i) \left( 1 + \frac{(H - \hat{H})}{H} + \frac{(H - \hat{H})^2}{H^2} + \frac{(H - \hat{H})^3}{H^3} \right)(X_i)
\]

\[
- n \frac{(H - \hat{H})^4}{H^4}(X_i) \geq H^{-1}(X_i)(1 + \hat{R}(X_i)) - n \frac{(H - \hat{H})^4}{H^4}(X_i),
\]
where \( \sup_{t \in \mathbb{R}} |\dot{R}(t)| \) is \( o(1) \) almost everywhere. Recall that
\[
\hat{\mu}_h(x) = \frac{1}{n} \sum_{i=1}^{n} \vec{K}_h(x - X_i) \delta_i \dot{H}^{-1}(X_i -),
\]
and denote
\[
\mu'_h(x) = \frac{1}{n} \sum_{i=1}^{n} \vec{K}_h(x - X_i) \delta_i H^{-1}(X_i -) \cdot (1 + \dot{R}(X_i -)).
\]
Then,
\[
P_\lambda \left( \inf_{n^{1/4} h \in [a, b]} \int |\alpha \lambda_0 \hat{\mu}_h| > \varepsilon n^{-2/5} \right)
\geq P_\lambda \left( \inf_{n^{1/4} h \in [a, b]} \int |\alpha \lambda_0 \mu'_h| > 2\varepsilon n^{-2/5} \right)
- P_\lambda \left( \sup_{n^{1/4} h \in [a, b]} \int \alpha \lambda_0 w(x) \sum_i |\vec{K}_h|(x - X_i) \delta_i \frac{(H - \dot{E})^4}{H^4} (X_i -) > \varepsilon n^{-2/5} \right)
\geq p_1 - p_2.
\]
It is possible to deal with \( p_1 \) like Hall and Marron (1991, p. 163) because, for all \( \lambda \in \{\lambda_0, \lambda_1\} \), we have
\[
E_\lambda \left[ \int \alpha \lambda_0 w \mu'_h \right] = \int \alpha \lambda_0 w(x) \vec{K}_h(x - t) \lambda(t) (1 + n^{-1} \phi(t)) dt dx
= \int \alpha_0 \lambda_0 w(hy + t) \hat{K}(y) \lambda(t) (1 + n^{-1} \phi(t)) dy dt
= \frac{h^2}{2} \int \frac{\dot{K}^{(2)}}{2} \int d^2(\alpha_0 \lambda_0 w)_\lambda + O(n^{-1} h^2)
= C_{st} \cdot h^2 + O(n^{-1} h^{-2}),
\]
where \( C_{st} \) is nonzero and \( \phi \) denotes a bounded function. To deal with \( p_2 \), note that there exists a constant \( C^* \) such that, a.s.
\[
\left| \int \alpha \lambda_0 w(x) \delta_i \frac{(H - \dot{E})^4}{H^4} (X_i -) dx \right|
\leq \frac{C^* \ln^2 n}{n^2 h} = o(n^{-2/5}).
\]
Hence \( \limsup_n \max_\lambda p_2 \to 0 \), proving Lemma 5.8.
5.3. Proof of the Main Theorem: the MISE Case with Multiple Alternatives

First, we prove (3.3). We may suppose that \( \nu \leq 2.5 \), for otherwise \( \nu/5 \geq 1/2 \) and then Theorem 3.1 follows from Theorem 2.2. Following some ideas of Bickel and Ritov (1988) and Hall and Marron (1991), the first step is to construct a sufficiently rich class of hazard functions in \( G_\nu(\tau) \) (we have supposed \( \tau_0 > 2 \) for convenience).

Let \( \psi_0 \) be a symmetrical, odd, six times differentiable function on \( \mathbb{R} \) vanishing outside \([-1/4, 1/4]\) and satisfying \( \sup_j |\psi_0^{(j)}| \leq C_0 \) for \( 0 \leq j \leq 6 \). Put \( m = \eta_0 \theta \) where \( \eta_0 > 0 \) and \( \theta \) is a positive function of \( \nu \) which will be specified below. Moreover, assume that \( n^{-1/3} m \) is bounded, \( \nu \leq \theta \geq 1/5 \). Let \( \psi = \delta \psi_0 \) where \( 0 < \delta \leq 1 \). Let \( \lambda_0 \) be a hazard function in \( G_\nu(\tau) \) which is constant on \([0, \tau_0/2]\) and vanishes outside \([0, \tau_0]\), so that \( \lambda_0 \) is constant on a compact real subset strictly larger than \([0, 1]\). To fix the ideas, \( \lambda_0 = 1 \) on \([0, \tau_0/2]\). Let \( r \) be an arbitrary nonnegative integer and \( \tau \) arbitrary in \((0, \ldots, r-1)\). Let \( \tau = (\tau_1, \ldots, \tau_m) \) be some real vector. Recall that \( \alpha = 2 + \nu \) and define

\[
\gamma_\nu(x) \equiv \beta m^{-\alpha} \psi(mx - \nu),
\]

\[
\lambda(x) \equiv \lambda_0(x) \left( 1 + \sum_{i=1}^{m} \tau_i \gamma_\nu(x) \right).
\]  

(5.32)

Let \( F_n \) or simpler \( F \) be the set of hazard functions \( \lambda \) which can be written as in (5.32), where \( \beta \) is chosen in \( 0, \ldots, (r-1) \) and the \( \tau_i, i = 1, \ldots, m \) are chosen independently into \( \{1, -1\} \).

To ensure that \( \lambda \) is a hazard function in \( F \), a sufficient condition is \( rm^{-\alpha}C_0 < 1 \). For every positive constants \( C_0 \) and \( \eta_0 \), it holds if \( n \) is large enough. Furthermore, \( F \subset G_\nu(\tau) \). Verify that \( F \) satisfies conditions vi-xii.

Consider \( \tilde{h} = h_{\tilde{h}, \tau} \) which is the minimizer of \( |\hat{h} - h_\lambda| \) on all \( \lambda \in F \). Then \( |\tilde{h} - h_\lambda| \leq 2|\hat{h} - h_\lambda| \) for any \( \lambda \in F \). To state the result, it is sufficient to prove that

\[
\lim_{\epsilon \to 0, \delta \to 0} \liminf_{n \to \infty} \sup_{\lambda \in F} P_\lambda( |\tilde{h} - h_\lambda| > \epsilon n^{-1/3} - \gamma ) = 1,
\]

(5.33)

where the power \( \gamma > 0 \) will be chosen as small as possible. The first step in establishing (5.33) is to develop an analog of (5.29). The same
considerations as in Hall and Marron (1991, p. 164) provide
\[ h_\lambda - \bar{h} = \frac{2\eta(\bar{h}, \lambda, \bar{\lambda})}{\bar{h} \cdot \text{MISE}^{[2]}(h^*, \lambda)}, \]  
\[ (5.34) \]

where \( h^* \) lies between \( h_\lambda \) and \( \bar{h} \), and
\[ \eta(n, \lambda, \lambda_1) = h \int \left( [(E_{h_\lambda} - E_{\lambda}) \{ (\lambda_0 - \lambda) \lambda^{[1]}_{\lambda_0} \} + E_{\lambda_0} \{ \lambda^{[1]}_{\lambda_0} (\lambda_1 - \lambda) \} \right) w. \]
\[ (5.35) \]

Through to the same limited expansions as in Lemma 5.1, we may show that given \( \xi > 0 \), we may choose \( \delta \) small enough, in the definition of \( \psi \), and \( n_0 \geq 1 \) large enough so that, if \( C(\lambda) > 0 \) is the constant such that \( h_\lambda \sim n^{-1/5} C(\lambda) \) and if \( a = C(\lambda_0) \), then
\[ a - \xi < \inf_{n \geq n_0, \lambda \in \mathcal{F}} n^{1/5} h_\lambda \leq \sup_{n \geq n_0, \lambda \in \mathcal{F}} n^{1/5} h_\lambda < a + \xi, \]
and, for any \( 0 < b < c < \infty \) and some \( \zeta = \zeta(b, c) > 0 \), we have
\[ \sup_{n^{1/5} h \in [b, c], \lambda \in \mathcal{F}} |\text{MISE}^{[2]}(h, \lambda)| \leq \zeta n^{-2/5}. \]

In view of these results and (5.34) we see that (5.33) will follow if we prove that for each sufficiently small constants \( \xi, \eta_0 > 0 \) (not depending on \( \delta \)), and any nonparametric rule \( \bar{\lambda} \) for selecting an element of \( \mathcal{F} \),
\[ \lim_{\varepsilon \to 0} \lim_{\delta \to 0} \lim_{n \to \infty} \sup_{\lambda \in \mathcal{F}} P_\lambda \left( \min_{h \in H_n} \| \eta(h, \lambda, \bar{\lambda}) \| > \varepsilon n^{-(4/5) - \gamma} \right) = 1, \]
\[ (5.36) \]
where \( H_n = [(a - \xi)n^{-1/5}, (a + \xi)n^{-1/5}] \).

The next step is to simplify \( \eta(h, \lambda, \lambda_1) \). Write
\[ \lambda = \lambda_0 \left( 1 + \sum_{\sigma} \tau_\sigma \gamma_\sigma \right) \quad \text{and} \quad \lambda_1 = \lambda_0 \left( 1 + \sum_{\sigma} \tau_\sigma \gamma_\sigma \right). \]

We have to deal with \( E_\lambda[\phi(X_t) \delta \hat{H}^{-1}(X_t -)] \) for some borelian functions \( \phi \). Note that we can use the expectation of \( \hat{H} \) (with respects to the underlying distribution associated with \( \lambda \)) instead of \( \bar{H} \) (see below). This expectation is denoted by \( H_\lambda \) or simpler \( \bar{H} \) (the context will leave
out any ambiguity). Thus

\[(E_\lambda - E_{\lambda_1}) \left[ \phi(X_1) \frac{\delta}{H(X_1 -)} \right] = \sum \tau_0 \gamma_0 \int \phi \lambda_0 \gamma_0.\]

Simple calculations provide

\[
\eta(h, \lambda, \lambda_1) = \int \left[ n^{-1}(E_\lambda - E_{\lambda_1}) \left[ K_h \tilde{K}_h(x - X_1) \frac{\delta}{\tilde{H}^2(X_1 -)} \right] \right. \\
+ (1 - n^{-1}) \left( E_\lambda \left[ K_h(x - X_1) \frac{\delta}{\tilde{H}(X_1 -)} \right] \right. \\
\left. \cdot E_{\lambda_1} \left[ \tilde{K}_h(x - X_1) \frac{\delta}{\tilde{H}(X_1 -)} \right] \right) \\
- E_{\lambda_1} \left[ K_h(x - X_1) \frac{\delta}{\tilde{H}(X_1 -)} \right] \cdot E_{\lambda_1} \left[ \tilde{K}_h(x - X_1) \frac{\delta}{\tilde{H}(X_1 -)} \right] \\
- \lambda(x)(E_\lambda - E_{\lambda_1})[h\tilde{\lambda}^{(1)}_h(x)] \\
+ \left( \lambda - \lambda_1 \right)(x)E_{\lambda_1} [h\tilde{\lambda}^{(1)}_h(x)] w(x)dx.
\]

Hence

\[
\eta(h, \lambda, \lambda_1) = \sum \tau_0 \gamma_0 I_0 + n^{-1} \int (E_\lambda - E_{\lambda_1}) \left[ K_h \tilde{K}_h(x - X) \frac{\delta}{\tilde{H}^2(X -)} \right] w(x)dx + R(\eta),
\]

where

\[
I_0 = \int \lambda_0 \gamma_0(y) \left[ (1 - n^{-1}) \left( K_h(x - y) \cdot E_\lambda \left[ \tilde{K}_h(x - X) \frac{\delta}{\tilde{H}(X -)} \right] \right. \right. \\
+ \tilde{K}_h(x - y) \cdot E_{\lambda_1} \left[ K_h(x - X) \frac{\delta}{\tilde{H}(X -)} \right] \right) \\
- \lambda(x)\tilde{K}_h(x - y) \right] w(x)dx dy \\
+ \int E_{\lambda_1} \left[ \tilde{K}_h(x - X) \frac{\delta}{\tilde{H}(X -)} \right] \lambda_0(x)\gamma_0(x)w(x)dx.
\]

The remainder term \(R(\eta)\) stems from replacing the empirical survival function \(\tilde{H}\) by its expectation \(\bar{H}\). More precisely, using the expansion

\[
\tilde{H}^{-1}(X_1 -) = H^{-1}(X_1 -) + (H - \tilde{H})(X_1 -)/H^2(X_1 -) \\
+ \cdots (H - \tilde{H})^p(X_1 -)/H^2(X_1 -),
\]

(5.37)
and since \( \dot{H}(X_1-) \geq 1/n \), we deduce easily that

\[
E_\lambda \left[ K_h \bar{K}_h (x - X_1) \frac{\delta}{\dot{H}(X_1-)} \right] \\
= E_\lambda \left[ K_h \bar{K}_h (x - X_1) \frac{\delta}{H^2(X_1-)} \left( 1 + \frac{(H - \dot{H})}{H} (X_1-) + \ldots \right) \right] \\
= \int K_h \bar{K}_h (x - t) \frac{\lambda}{H} (t) \left( 1 + 2n^{-1} \frac{(1 - H)}{H} (t) + \ldots \right) dt \\
= \int K_h \bar{K}_h (x - t) \frac{\lambda}{H} (t) (1 + n^{-1} R_1(t) + O(n \cdot n^{-p})) dt,
\]

where \( R_1(\cdot) \) is a non random function independent from \( \lambda \), bounded on \( \tau \) uniformly with respect to \( \lambda \in \mathcal{F} \). Similarly,

\[
E_\lambda \left[ K_h (x - X_1) \frac{\delta}{\dot{H}(X_1-)} \right] \\
= \int K_h (x - t) \lambda(t) \left( 1 + n^{-1} R_2(t) + O(n \cdot n^{-p}) \right) dt. \quad (5.38)
\]

Similar formulas are available for each term involving a denominator \( \dot{H}(X_1) \). We will not detail all these non random functions \( R_i \) subsequently. They could be considered as being included into the integrated function \( \lambda \) e.g. They do not modify any convergence or inequality rate. Thus they will be forgotten. The remainder terms could be taken as small as desired, choosing \( p \) large enough. Thus, from now on, we consider only \( H \) in each denominator (\( R(\eta) \) being considered as small as necessary).

Let us discuss directly the second term of the previous expansion of \( \eta \). Denote \( H \) and \( H_1 \) (resp. \( F \) and \( F_1 \)) the survival functions of \( X \wedge T \) (resp. \( T \)) associated with \( \lambda \) and \( \lambda_1 \). Then

\[
n^{-1} \int (E_\lambda - E_{\lambda_1}) \left[ K_h \bar{K}_h (x - X_1) \frac{\delta}{H^2(X_1-)} \right] w(x) dx \\
= n^{-1} \int K_h \bar{K}_h (x - t) \left( \frac{\lambda}{H} - \frac{\lambda_1}{H_1} \right) (t) w(x) dx dt \\
= \frac{1}{nh} \int K \bar{K}(u) \left( \frac{\lambda}{F} - \frac{\lambda_1}{F_1} \right) (t) w(t + hu) du dt.
\]
But we have easily
\[
\frac{\lambda}{F} - \frac{\lambda_1}{F_1} = \sum \limits_{\nu} (\tau_\nu - \tau_{1\nu}) \frac{\lambda_0 \gamma_\nu}{F} + \frac{\lambda_1}{FF_1} \left( \exp \left\{ \int_0^1 \sum \limits_{\nu} \lambda_0 (\tau_\nu - \tau_{1\nu}) \gamma_\nu \right\} - 1 \right). \tag{5.39}
\]

Since \( \int \sum (\tau_\nu - \tau_{1\nu}) \lambda_0 \gamma_\nu \) is \( O(m^{-\alpha}) \), or even simpler \( o(1) \), uniformly on \( t \in \tau \), deduce

\[
n^{-1} \int (E_\lambda - E_{\lambda_0}) \left[ K_\lambda K_h(x - X_1) \frac{\delta}{HF(x - \cdot)} \right] w(x) \, dx
\]
\[
= \frac{1}{nh} \sum \limits_{\nu} (\tau_\nu - \tau_{1\nu}) \int K \hat{K}(u) \left[ \frac{\lambda_0 \gamma_\nu}{H} (t) + \frac{\lambda_1}{FH_1} (t) \int_0^t \lambda_0 \gamma_\nu \right] w(t + hu) \, dt \, du + O \left( \frac{1}{nh} m^{-2\alpha} \right)
\]
\[
= \frac{1}{nhm^{\alpha+1}} \sum \limits_{\nu} (\tau_\nu - \tau_{1\nu}) \int K \hat{K}(u) \psi(y) \frac{\lambda_0}{H} \left( \frac{y + v}{m} \right)
\]
\[
\times w \left( \frac{y + v}{m} + hu \right) \, du \, dy + \frac{\beta}{nhm^{\alpha+1}} \sum \limits_{\nu} (\tau_\nu - \tau_{1\nu})
\]
\[
\int K \hat{K}(u) \frac{\lambda_1}{FH_1} (t) \int_{-v}^{m-\nu} \lambda_0 \left( \frac{s + v}{m} \right) \psi(s) w(t + hu) \, du \, dt \, ds + O \left( \frac{1}{nh} m^{-2\alpha} \right).
\]

Since \( \psi \) is odd, a limited expansion with respects to \( y \) provides that this last term is \( O((nhm^{\alpha+1})^{-1} m m^{-1}) \). Thus, we need to verify that, for the considered values,

\[
\frac{1}{nhm^{\alpha+1}} = o(n^{-4/5} m^{-\gamma}) \tag{5.40}
\]

\[
\frac{1}{nhm^{2\alpha}} = o(n^{-4/5} m^{-\gamma}). \tag{5.41}
\]
Let us deal with the main part of \( \eta \) viz the sum over \( \nu \) of the previous quantities denoted by \( I_\nu \). Set \( y = z/m \) and \( x - y = ht \). Then

\[
I_\nu = \int \lambda_0 \gamma_\nu(z/m) \\
\left[ (1 - n^{-1}) \left( K(t) E_h \left[ K_h(z/m + ht - X) \frac{\delta}{H(X)} \right] \right) \\
+ \bar{K}(t) E_h \left[ K_h(z/m + ht - X) \frac{\delta}{H(X)} \right] - \lambda(z/m + ht) \bar{K}(t) \right] \\
w(z/m + ht) dt dz/m \\
+ \int E_h \left[ \bar{K}_h(l/m - X) \frac{\delta}{H(X)} \right] \lambda_0(l/m) \gamma_\nu(l/m) w(l/m) dl/m.
\]

In the first integral, suppose that \( z \in \nu + \text{supp}(\psi_0) = [v - 1/4, v + 1/4] \) and that \( t \in \text{supp}(K) = [-A, A] \). Note that

\[
E_h \left[ \bar{K}_h(z/m + ht - X) \frac{\delta}{H(X)} \right] \\
= \int \bar{K}_h(z/m + ht - u) \lambda_0(u) du \\
+ \sum_{\nu'} \gamma_{\nu'} \int \bar{K}_h(z/m + ht - s) \lambda_0(s) \gamma_{\nu'}(s) ds \\
= \int \bar{K}(u) \lambda_0(z/m + h(t - u)) du + \beta \sum_{\nu'} \gamma_{\nu'} \int \bar{K}(u) \lambda_0(z/m + h(t - u)) \\
\psi(z - \nu' + mh(t - u)) du/m^\alpha.
\]

Recall that \( h \leq (a + \xi)n^{-1/3} \). For all considered \( (z, t, u) \), we have

\[
|z/m + h(t - u)| \leq (v + 1/4)/m + 2Ah \leq 1 + 1/(4m) + 2A(a + \xi)n^{-1/3}.
\]

For \( n \) large enough (more precisely, \( n \geq n^* \), where \( n^* \) is independent from \( \lambda \in \mathcal{F} \)), we obtain

\[
\int \bar{K}_h(z/m + ht - u) \lambda_0(u) du = \int \bar{K} = 0
\]

Moreover, for each \( \nu' \neq \nu \), \( |v - \nu'| \geq 1 - 1/4 \). Thus

\[
|z - \nu' + mh(t - u)| \geq (1 - 1/4) - 2Ahm \geq 3/4 - 2A(a + \xi) \gamma_0 > 1/4,
\]
for \(\eta_0\) sufficiently small, and, if \(\theta' \neq \theta\), we have
\[
\int \tilde{K}(u) \lambda_0(z/m + h(t - u)) \psi(z - \theta' + mh(t - u)) du = 0.
\]

Hence, if \(n\) is large enough (viz \(n \geq n^*\)) then for any \(\lambda \in \mathcal{F}\), we have
\[
E_{\lambda} \left[ \tilde{K}_h(z/m + ht - X) \frac{\delta}{H(X)} \right] = \beta \tau_0 \int \tilde{K}(u) \lambda_0(z/m + h(t - u)) \psi(z - \theta + mh(t - u)) du/m^\alpha
\]
\[
= \beta \tau_0 \int \tilde{K}(u) \psi(z - \theta + mh(t - u)) du/m^\alpha.
\]

Similarly, if \(n \geq n^*\),
\[
E_{\lambda} \left[ \tilde{K}_h(l/m - X) \frac{\delta}{H(X)} \right] = \beta \tau_0 \int \tilde{K}(u) \psi(l - \theta - hmu) du/m^\alpha, \quad (5.42)
\]
and since the integral of \(K\) is nonzero,
\[
E_{\lambda} \left[ K_h(z/m + ht - X) \frac{\delta}{H(X)} \right] = 1 + \beta \tau_0 \int K(u) \psi(z - \theta + hm(t - u)) du/m^\alpha. \quad (5.43)
\]

Moreover, note that with the considered values, if \(n \geq n^*\),
\[
\lambda(z/m + ht) = 1 + \beta \tau_0 \psi(z + mht - \theta)/m^\alpha,
\]
\[
\lambda_0(z/m + h(t - u)) = 1.
\]

We deduce, if \(n \geq n^*\) and \(\beta \neq 0\),
\[
I_\theta/\beta = \frac{1}{m^\alpha+1} \int \lambda_0(z/m) \psi(z - \theta)
\]
\[
\cdot \left[ (1 - n^{-1}) \left( \beta K(t) \tau_0 \int \tilde{K}(u) \lambda_0(z/m + h(t - u)) \psi(z - \theta + hm(t - u)) du/m^\alpha \right) \right]
\]
\[ + \bar{\psi}(t) \left\{ 1 + \beta \tau_v \int \bar{K}(u) \lambda_0 ((z - v)/m + h(t - u)) \psi(z - v + hm(t - u)) du/m^a \right\} \]

\[ - \bar{\psi}(t) \left\{ 1 + \beta \tau_v \lambda_0 (z/m + ht) \psi(z - v + mht)/m^a \right\} \]

\[ w(z/m + ht) dt dz + m^{-2a-1} \beta \tau_v \int \bar{K}(u) \lambda_0 (z/m - hu) \psi(z - v - hm) \psi(z - v) \lambda_0 (z/m) w(z/m) dz du \]

\[ = \frac{\beta}{m^{2a+1} \tau_v} \int \psi(z - v) \]

\[ \cdot [\bar{K}(u) \bar{K}(t) \psi(z - v + hm(t - u)) du - \bar{\psi}(z - v + mht) w(z/m) \]

\[ + \frac{\beta}{m^{2a+1} \tau_v} \int \psi(z - v) \bar{K}(t) \left[ \bar{K}(u) \psi(z - v + hm(t - u)) du - \psi(z - v + mht) w(z/m) \right] dz dt \]

\[ - \frac{1}{nm^{2a+1}} \int \psi(z - v) [\tau_v \bar{K}(t) \bar{K}(u) + \tau_v \bar{K}(t) \bar{K}(u)] \psi(z - v + hm(t - u)) du \]

\[ w(z/m + ht) \int \psi(z - v) \bar{K}(t) w(z/m + ht) dt dz \]

\[ = \frac{\beta}{m^{2a+1} \tau_v} \int \psi(y) \cdot \left[ -\bar{K}(u) \bar{K}(t) \psi(y + hm(t + u)) du \right] \]

\[ - \bar{K}(t) \psi(y + mht) w\left( \frac{y + v}{m} + ht \right) dy dt \]

\[ - \frac{1}{nm^{2a+1}} \int \psi(y) \left[ \bar{K}(u) \bar{K}(t) \psi(y + hm(t + u)) \right] \]

\[ w\left( \frac{y + v}{m} + hu \right) du - \bar{K}(t) \psi(y + mht) w\left( \frac{y + v}{m} \right) \] \[ dy dt \]

\[ + \frac{\beta}{nm^{2a+1}} \int \psi(y) \tau_v \bar{K}(t) \bar{K}(u) \]

\[ \psi(y + hm(t + u)) w\left( \frac{y + v}{m} + ht \right) dy dt du \]
Note that we have used the symmetry of the kernel functions. Then,

\[ m^{2\alpha+1}\frac{\beta}{\beta^2} = (\tau_0 + \tau_{10}) \int \psi(y) \]

\[ \cdot \left[ \kappa(u)K(t)\psi(y + hm(t + u))w\left(\frac{y + v}{m} + ht\right) dt \\
- \kappa(u)\psi(y + hmu)w\left(\frac{y + v}{m}\right) dy du \\
+ \frac{1}{n}(\tau_0 + \tau_{10}) \int \psi(y)K(t)\kappa(u)\psi(y + hm(t + u))w\left(\frac{y + v}{m} + ht\right) dy du \\
+ \tau_0 \int \psi(y)K(t)\psi(y + hm(t + u)) \right] \]

\[ \left[ w\left(\frac{y + v}{m} + hu\right) - w\left(\frac{y + v}{m} + ht\right) \right] du dy dt \\
- \tau_0 \int \psi(y)K(t)\psi(y + hmt) \]

\[ \left[ w\left(\frac{y + v}{m} + ht\right) - w\left(\frac{y + v}{m}\right) \right] dy dt \\
+ \frac{1}{n} \tau_{10} \int \psi(y)K(t)\psi(y + hm(t + u)) \]

\[ \left[ w\left(\frac{y + v}{m} + hu\right) - w\left(\frac{y + v}{m} + ht\right) \right] dy dt du \\
- \frac{m^\alpha}{n\beta} \int \psi(y)K(t)w\left(\frac{y + v}{m} + ht\right) dt dy \\
= (\tau_0 + \tau_{10})J_1 + \frac{1}{n}(\tau_0 + \tau_{10})J_2 \\
+ \frac{1}{m^{2\alpha+1}} \left[ \tau_{10}J_3 \left(1 + \frac{1}{n}\right) - \tau_0 J_4 \right] - \frac{m^\alpha}{n\beta}J_5 \]
Deduce

\[
\sum_v (\tau_v - \tau_{v}) I_v = \frac{\beta^2}{m^{2\alpha+1}} \sum_v (|\tau_v| - |\tau_{v}|)(J_1 - J_2/n)
\]
\[
+ \frac{\beta^2(1 + 1/n)}{m^{2\alpha+1}} \sum_v (\tau_v \tau_{v} - |\tau_{v}|) J_2
\]
\[
- \frac{\beta^2}{m^{2\alpha+1}} \sum_v (|\tau_v| - \tau_{v}) J_4 - \frac{\beta}{nm^{\alpha+1}} \sum_v (\tau_v - \tau_{v}) J_5.
\]

(5.44)

Note that all the quantities \(J_k, k = 1, \ldots, 5\), depend on \(v\), even if we have not indexed by \(v\). We have denoted

\[
J_1 = \int \psi(y) \tilde{K}(u) \left[ \psi(y + hm(t + u)) K(t) \left( \frac{(y + v)}{m} + ht \right) \right] dt
\]
\[
- \psi(y + hmu) \left( \frac{(y + v)}{m} \right) du dy,
\]
\[
J_2 = \int \psi(y) \psi(y + hm(t + u)) \tilde{K}(u) \left( \frac{(y + v)}{m} + ht \right) dudy dt,
\]
\[
J_3 = \int \psi(y) \psi(y + hm(t + u)) \tilde{K}(t) \tilde{K}(u)
\]
\[
\left[ w \left( \frac{(y + v)}{m} + hu \right) - w \left( \frac{(y + v)}{m} + ht \right) \right] dt du dy,
\]
\[
J_4 = \int \psi(y) \psi(y + hmt) \tilde{K}(t) \left[ w \left( \frac{(y + v)}{m} + ht \right) - w \left( \frac{(y + v)}{m} \right) \right] dt dy,
\]
\[
J_5 = \int \psi(y) \tilde{K}(t) \left( \frac{y + v}{m} + ht \right) dt dy.
\]

Set \(\zeta = hm\). By assumption, \(\zeta \to 0\) or \(\zeta \sim \text{Cst.}\)

**Study of \(J_1\)** If \(\zeta \to 0\), then, since \(\int \tilde{K} = 0\), some limited expansions of \(\psi\) and \(w\) provide, with obvious notations,

\[
J_{1/\beta} = \int K(t) \tilde{K}(u) \psi(y)
\]
\[
\left( \sum_{k=0}^{4} \frac{d^k \psi(y) \zeta^k [(t + u)^k - u^k]}{k!} + \frac{d^4 \psi(y) \zeta^5 (t + u)^5}{5!} \right)
\]
\[
- \frac{d^3 \psi(y) \zeta^4 u^4}{5!} \right) dt du dy + \int K(t) \tilde{K}(u) \psi(y).
\]
BANDWIDTH SELECTION IN HAZARD SELECTION

\[
\left( \sum_{k=0}^{4} \frac{d^k \psi(y) \zeta^k(t + u)^k}{k!} + \frac{d^3 \psi(y^*) \zeta^3(t + u)^3}{5!} \right) \\
\cdot dw \left( \frac{y + v}{m} \right) ht(1 + O(h)) dt du dy \\
= \zeta^4 \int K(t) \tilde{K}(u) \psi(y) (d^4 \psi(y) [(t + u)^4 - u^4]/4! \\
+ \frac{d^3 \psi(y^*) \zeta(t + u)^3}{5!} - \frac{d^2 \psi(y^*) \zeta u^2}{5!} w \left( \frac{y + v}{m} \right) dt du dy \\
+ \zeta^2 \int K(t) \tilde{K}(u) \psi(y) d^2 \psi(y) (t + u)^2/2! \\
\cdot dw \left( \frac{y + v}{m} \right) ht \cdot (1 + O(h)) \cdot (1 + O(\zeta)) dt du dy.
\]

Also, deduce, when \( n \to \infty \),

\[
J_1(v) \sim t_4(v) \zeta^4,
\]

where

\[
t_4(v) = \beta^2/4 \int K^{(2)} \int \tilde{K}^{(2)} \int \psi(y) d^4 \psi(y) w \left( \frac{y + v}{m} \right) dy.
\]

We will suppose that \( \int \phi^{(4)} \psi \neq 0 \). If \( \zeta \sim Cst \), we can choose \( \eta_0 \) so small that

\[
\liminf_{n} \inf_{v} |J_{1v}| > 0
\]

\( J_2(v) / n \) will be negligible with respect to \( J_1(v) \) under the assumption \( \zeta^2/n = o(\zeta^4) \) or

\[
n(hm)^2 \to \infty.
\]

This condition is always satisfied since \( nh^2 \to +\infty \).
Study of $J_3$ and $J_4$. The two terms $J_3$ and $J_4$ can be viewed as negligible. The first solution is to suppose that $w$ is constant on $[0, 2]$, or, more generally on $[0, 2c_m]$, $c_m$ being a strictly positive constant (by means of some minor modifications, like changing $m$ into $[mc_m]$ in the summations over $v$). In this case, $J_k, k = 3, 4$ is zero for $n$ large enough.

The second solution, that we adopt here, is to take no assumptions about the values of $w$ but about its regularity only. More precisely, suppose that $w \in C^3(r)$. Then, by some Taylor expansions of $w, J_3$ and $J_4$ can be taken small enough. More precisely, recall that

$$J_3 = \int \psi(y)w(y + \zeta(t + u))K(t)K(u)$$
$$\left[ w\left(\frac{y + u}{m} + hu\right) - w\left(\frac{y + u}{m} + ht\right) \right] du \, dy \, dt, \tag{5.48}$$

$$J_4 = \int \psi(y)w(y + \zeta u)K(u)$$
$$\left[ w\left(\frac{y + u}{m} + ht\right) - w\left(\frac{y + u}{m}\right) \right] du \, dy. \tag{5.49}$$

For these two terms, we can make some limited expansions of $w$ and $\psi$. For simplicity, we do not precise the order of the remainder terms.

$$J_4 = \int \sum_{p \geq 0, q \geq 1} \psi(y)\frac{d^pw}{p!} w\left(\frac{y + u}{m}\right) \frac{t^p}{q!} \frac{\zeta^p h^q}{2q!} du \, dy$$
$$= \int \sum_{p \geq 0, q \geq 1, q \geq 0} \psi(y)\frac{d^pw}{p!} w\left(\frac{y + u}{m}\right) \frac{t^p}{q!} \frac{\zeta^p h^q}{2q!} du \, dy,$$

$$J_3 = \int \sum_{p \geq 0, q \geq 1} \psi(y)\frac{d^pw}{p!} w\left(\frac{y + u}{m}\right) \frac{t^p}{q!} \frac{\zeta^p h^q}{2q!} du \, dy \, dt$$
$$= \int \sum_{p \geq 0, q \geq 1, q \geq 0} \psi(y)\frac{d^pw}{p!} w\left(\frac{y + u}{m}\right) \frac{t^p}{q!} \frac{\zeta^p h^q}{2q!} du \, dy \, dt.$$
Note that \( p + q \) can be supposed even \( \geq 2 \) since \( \bar{K} \) has an even order. Moreover, \( \int \psi \) and \( \int \psi \psi' \) are zero, hence the main part of \( J_3 \) and \( J_4 \) is obtained (at worse) with \((p, q, r) \in \{(0, 1, 2), (1, 1, 1), (2, 1, 0)\})

To state (5.36), the two previous terms can be viewed as residuals when we have
\[
\frac{1}{m^{2a+1}} \left[(1 + 1/n) \sum_v (\tau_v \tau_0 - |\tau_v|) J_3 + \sum_v (\tau_v \tau_0 - |\tau_v|) J_4\right]
= O \left( m^{-2a} \left[ \frac{h}{m^2} + \frac{ch}{m} + \frac{\zeta h}{m}\left(1 + \frac{\zeta h h}{m}\right) \right] \right) = o(n^{-4/5}n^{-\gamma}).
\] (5.50)

Thus, it is necessary that
\[
\frac{h}{m^{2a}} \left[ \frac{1}{m^2} + h + h^2 m^2 \right] = o(n^{-4/5}n^{-\gamma}).
\] (5.51)

To dealing with \( J_5 \), note that
\[
\frac{1}{nm^{2a+1}} \sum_v (\tau_v - \tau_0) J_5(v) = O(n^{-1}m^{-a-1}m[m^{-1} + h]).
\]

Hence, \( J_5 \) is negligible if
\[
n^{-1}m^{-a-1} = o(n^{-4/5}n^{-\gamma}).
\] (5.52)

Hence the main part of \( \eta(h, \lambda, A, \lambda) \) is \( \sum_v (\tau_v - \tau_0) I_v + \sum_v (|\tau_v| - |\tau_0|) J_1(v) \). Then Eq. (5.36) follows from
\[
\lim_{\varepsilon \to 0, \delta \to 0} \liminf_{n \to \infty} \max_{\lambda \in \mathcal{F}} P_{\lambda} \left\{ m^{-2a-1} \beta \left| \sum_v (|\tau_v| - |\bar{\tau}_0|) J_1(v) \right| > \varepsilon n^{-4/5}n^{-\gamma} \right\} = 1,
\] (5.53)

denoting \( \bar{\tau} \) the quantities concerning \( \bar{\lambda} \). Recall that \( J_1(v) \sim t_4(v) \zeta^4 \). All the quantities \( t_4(v), v = 1, \ldots, m \) have the same sign, say positive. They can be seen like weights. Note that \( \sum_v t_4(v) \) is proportional to \( m \). For convenience, we impose \( \sum_v t_4(v) = m \). Note that, when \( w \) is constant on \([0, \tau_0/2]\), then the sequence \( (t_4(v))_v \) is constant too for \( n \) sufficiently large. Then, if we prove that
\[
\lim_{\varepsilon \to 0, \delta \to 0} \liminf_{n \to 0} \max_{\lambda \in \mathcal{F}} P_{\lambda} \left\{ \left| \sum_v (|\tau_v| - |\bar{\tau}_0|) t_4(v) \right| > \varepsilon m \right\} = 1,
\] (5.54)
then the result follows, choosing \( m = n^{-1/2} \) of the same order as \( n^{-4/3} \), or

\[
m = \tau_0 [\eta^n] = \tau_0 [\eta^n (2\alpha - 4)].
\]  

(5.55)

Note that this statement implies that \( \gamma / (2\alpha - 4) \leq 1/5 \) or

\[
\gamma \leq 2\nu/5.
\]  

(5.56)

To prove (5.54), we follow Bickel and Ritov (1988). We prove the assertion by presenting a sequence of Bayes problems. In the \( n \)th problem we observe \( X_1, \ldots, X_n \) iid, whose hazard function is \( \lambda \). The loss function is \( L_n(\theta, d) = 1\{|\theta - d| > \varepsilon m\} \).

\( \lambda \) has been picked according to the following measure \( \pi_m \). Conditional on the data, let \( \tau_1, \ldots, \tau_m \) be a sequence of independent, symmetric r.v. taking only values 1 and \(-1\), and let \( \beta \) be a r.v., independent from the previous ones, equal to \( 0, 1, \ldots, r - 1 \) with probability \( 1/r \). The random measure is governed by the hazard function

\[
\lambda^* = \lambda_0 \left( 1 + \sum_{i=1}^{m} \tau_i^* \gamma \right),
\]

\[
\gamma(x) = \beta m^{-\alpha} \psi(mx - v).
\]

We will show that, if \( \lambda^* \) is governed by the distribution \( \pi_m \), then the variational distance between the probability measures of \( X_1, \ldots, X_n \) under \( \beta = i \) and \( \beta = j \) tends to 0 (for any choice of \( i \) and \( j \)). Assume it is the case.

We want to estimate \( \beta \) or \( m\beta \) rather. Note that, if \( \beta = j \), then \( \beta \sum_{i=1}^{m} |\tau_i|_4(v) = mj \).

A random estimate of \( m\beta \) is provided by \( T_n = \beta \sum_{i=1}^{m} |\tau_i|_4(v) \) (the "true" \( \beta \)). If

\[
A_{\beta} = \{ |T_n - jm| \leq \varepsilon m \}
\]

then the \( A_{\beta} \) are disjoint when \( \varepsilon < 1 \). The Bayes risk for estimating \( mj \) using our loss function is

\[
R_n = \frac{1}{r} \sum_{j=0}^{r-1} P_{\lambda^*}(A_{\beta} \mid \beta = j)
\]

\[
= 1 - \frac{1}{r} \sum_{j=0}^{r-1} P_{\lambda^*}(A_{\beta} \mid \beta = j)
\]
But, by the equivalence of $P_{\lambda}(\cdot| \beta = i)$ and $P_{\lambda}(\cdot| \beta = j)$, we have observed that

$$P_{\lambda}(A_{nj}| \beta = i) - P_{\lambda}(A_{nj}| \beta = 0) \to 0$$

for each $j$. Thus,

$$\lim \inf_n R_n \geq 1 - \frac{1}{r} \lim \sup_n \sum_{j=0}^{r-1} P_{\lambda}(A_{nj}| \beta = 0)$$

$$\geq 1 - \frac{1}{r} \lim \sup_n P_{\lambda} \left( \bigcup_j A_{nj}| \beta = 0 \right) \geq 1 - \frac{1}{r}$$

Since

$$\sup_{\lambda \in \mathcal{F}} P_{\lambda} \left( \left| \beta \sum_v (|r_v| - |\bar{r}_v|)t_4(v) \right| > \epsilon m \right)$$

$$\geq \frac{1}{r} \sum_{j=0}^{r-1} P_{\lambda} \left( \left| \beta \sum_v (|r_v| - |\bar{r}_v|)t_4(v) \right| > \epsilon m | \beta = j \right)$$

$$\geq \frac{1}{r} \sum_{j=0}^{r-1} P_{\lambda}(|T_n - jm| > \epsilon m | \beta = j) \geq R_n,$$

and since $r$ is arbitrary, we deduce (5.54), i.e.,

$$\lim_{\epsilon \to 0} \lim \inf_n \sup_{\lambda \in \mathcal{F}} P_{\lambda} \left\{ \left| \beta \sum_v (|r_v| - |\bar{r}_v|)t_4(v) \right| > \epsilon m \right\} = 1.$$

Thus, the proof is completed if the variational distance between the probability measures of $X_1, \ldots, X_n$ under $\beta = i$ and $\beta = j$ tends to 0, where $i, j = 1, \ldots, r - 1$. Without loss of generality, consider $\beta = 0$ and $\beta = 1$.

To prove this result, Bickel and Ritov (1988) use the inequality between the variational distance and the Kullback-Leibler information (Shorack and Wellner, p. 159), and some calculations derived from the likelihood ratio. For technical reasons, this method is too cumbersome in our case. That is why we prefer proving the assertion by a direct argument. Indeed, for any measurable subset $A$ and denoting by $\pi$ the
probability distribution induced by the r.v. \((\tau_1^*, \ldots, \tau_m^*)\), we have

\[
P_{\lambda^*}(A|\beta = 1) = \int P_{\lambda}(A|\beta = 1) d\pi(\lambda)
\]

\[
= \sum_{\tilde{e}=(e_1, \ldots, e_m)} \frac{1}{2^m} \int_A \lambda_{\tilde{e}}(x) \exp \left( - \int_0^x \lambda_{\tilde{e}}(u) du \right) dx
\]

\[
= \sum_{\tilde{e}=(e_1, \ldots, e_m)} \frac{1}{2^m} \int_A \lambda_0(x) \left( 1 + \sum_{i=1}^m e_i \gamma_i(x) \right)
\]

\[
\exp \left( - \int_0^x \lambda_0 \right) \exp \left( - \sum_{i=1}^m e_i \int_0^x \lambda_0 \gamma_i(u) du \right) dx
\]

\[
= P_{\lambda_0}(A|\beta = 0) + \sum_{\tilde{e}=(e_1, \ldots, e_m)} \frac{1}{2^m} \int_A \lambda_0(x) \exp \left( - \int_0^x \lambda_0 \right)
\]

\[
\left( \sum_{i=1}^m e_i \left[ \gamma_i(x) - \int_0^x \lambda_0 \gamma_i(u) du + O(m^{-2\alpha}) \right] \right) dx.
\]

Note that the support of the functions \(\gamma_i, i=1, \ldots, m\), are disjoints. Then, \(\sum_i e_i \gamma_i(x)\) is \(O(m^{-\alpha})\) uniformly with respects to \(x \in \tau\). Because \(\lambda_0\) is compactly supported, this provides

\[
\sup_A |P_{\lambda^*}(A|\beta = 1) - P_{\lambda}(A|\beta = 0)|
\]

\[
\leq \sum_{\tilde{e}} \frac{Cst}{2^m m^{\alpha}} \int \lambda_0(x) \exp \left( - \int_0^x \lambda_0 \right) dx = O(m^{-\alpha}) \rightarrow 0,
\]

(5.57)

and result (3.3) follows if conditions (5.40), (5.41), (5.47), (5.51), (5.52) and (5.56) are satisfied. Recalling that \(m\) is of order \(n^{\gamma/2\nu}\), it is easy to show that (5.40) is equivalent to \(\nu < 3\) and that (5.41), (5.47) and (5.52) are always satisfied. Moreover, condition (5.51) can be rewritten \(\gamma > \nu/5\). Thus, every constant \(\gamma\) in \([\nu/5, 2\nu/5]\), provide the result.

In fact, it is possible to obtain a little more stronger result, viz. to reach the rate \(\rho = \nu/5\). Indeed, in this case, the sum (5.50) involving \(J_3(\nu)\) and \(J_4(\nu)\) is a priori of the same order as the dominant part of the sum involving \(J_1(\nu)\) (see 5.53). But, noting that the dominant part of \(J_3(\nu)\) and \(J_4(\nu)\) are the same, the result is still true since the main part
of the Eq. (5.44) is now

$$\frac{\beta^2}{m^{2\alpha+1}} \sum_p (|\tau_p| - |\tau_{1p}|)(J_1 + J_2). \tag{5.58}$$

Thus, the technical details are the same as previously, replacing an equivalent of $$J_1(v)$$ by an equivalent of $$J_1(v) + J_3(v)$$.

To prove (3.4), note that

$$MISE(\hat{h}, \lambda) - MISE(h, \lambda) = (\hat{h} - h)MISE^2(h, \lambda)/2,$$

where $$|\hat{h} - h| \leq |h - h_\lambda|$$. It is sufficient to restrict the set $$G_{\nu, \theta}$$ to its previously defined subset $$F$$, which satisfies conditions vi-xii. We have then, through to the proof of Lemmas 5.1 and 5.5,

$$\sup_{n, \lambda \in F} n^{4/5}MISE(h, \lambda) < \infty,$$

$$0 < n^{2/5} \inf_{n, \lambda \in F} \{h, |h-h_\lambda| \leq |h-h_\lambda|\} MISE^2(h, \lambda).$$

Thus, for any $$\epsilon > 0$$, there exists $$\epsilon^*$$ such that

$$\sup_{\lambda \in F} P_{\lambda} \left( \frac{|MISE(\hat{h}, \lambda) - MISE(h, \lambda)|}{MISE(h, \lambda)} > \epsilon n^{-2p} \right)$$

$$= \sup_{\lambda \in F} P_{\lambda} \left( (\hat{h} - h)MISE^2(h, \lambda)/2 \epsilon n^{-2p} \right)$$

$$\geq \sup_{\lambda \in F} P_{\lambda} ((\hat{h} - h)^2/h^2_\lambda > \epsilon^* n^{-2p})$$

and (3.4) is an obvious consequence of (3.3).

\[5.4. \text{Proof of the Main Theorem: The MISE Case with Two Alternatives}\]

The context here is related to that in Theorem 3.1, if we consider $$m = 1, \nu = \infty, m^{-\nu} = m^{-1/2}$$ and $$\psi = \alpha \lambda_0$$. The two hazard functions are $$\lambda_0$$ and $$\lambda_1 = \lambda_0 + n^{-1/2} \psi$$ instead of $$\lambda_0$$ and $$\lambda_0 + \gamma_1$$. With these modifications, the argument is a hybrid of those for Theorems 2.1 and 2.2. Particularly, it suffices to prove, instead of (5.36), that

$$\lim_{\epsilon \to 0, \delta \to 0} \lim_{n \to \infty} \sup_{\lambda \in \{\lambda_0, \lambda_1\}} P_{\lambda} \left( \min_{h \in F_n} |\eta(h, \lambda, \hat{\lambda})| > \epsilon n^{-13/10} \right) \geq \rho, \tag{5.59}$$
where, by an analogue of \( I_n \),

\[
|\eta(h, \lambda, \bar{\lambda})| = \begin{cases} 
0 & \text{if } \lambda = \bar{\lambda} \\
n^{-1/2}|I| & \text{if } \lambda \neq \bar{\lambda}
\end{cases}
\]

and \( I \) is given by

\[
I = \int \psi(y) \left[ (1 - 1/n) \left( K_h(x - y) E_{\lambda_0} \left[ \tilde{K}_h(x - X) \frac{\delta}{H(X)} \right] 
+ \tilde{K}_h(x - y) E_{\lambda_1} \left[ K_h(x - X) \frac{\delta}{H(X)} \right] \right) 
- \lambda_0(x) \tilde{K}_h(x - y) \right] w(x) dx \, dy
\]

\[
+ \int E_{\lambda_0} \left[ \tilde{K}_h(x - X) \frac{\delta}{H(X)} \right] \psi(x) w(x) dx
\]

\[
+ \frac{1}{nh} \int K\bar{K}(x) \psi(y) \left[ \frac{\lambda_0}{H_0} - \frac{\lambda_1}{H_1} \right] (y) w(y + hx) dy \, dx.
\]

Note that

\[
E_{\lambda_0} \left[ \tilde{K}_h(x - X) \frac{\delta}{H(X)} \right] = \int \tilde{K}(t) \lambda_0(x - ht) dt,
\]

\[
E_{\lambda_1} \left[ K_h(x - X) \frac{\delta}{H(X)} \right] = \int K(t) \lambda_1(x - ht) dt.
\]

Thus

\[
I = \int \psi(y) \left[ (1 - 1/n) \left( K(x) \int \tilde{K}(t) \lambda_0(y + hx - ht) dt 
+ \tilde{K}(x) \int K(t) \lambda_1(y + h(x - t)) dt \right) 
- \lambda_0(y + hx) \tilde{K}(x) \right] w(y + hx) dx \, dy
\]

\[
+ \int \psi(y) \tilde{K}(t) \lambda_1(y - ht) w(y) dy \, dt
\]

\[
+ \frac{1}{nh} \int K\bar{K}(x) \psi(y) \left[ \frac{\lambda_0}{H_0} - \frac{\lambda_1}{H_1} \right] (y) w(y + hx) dy \, dx
\]

\[
= I_1 + I_2 + I_3 + I_4,
\]
where

\[ I_1 = \int \psi(y) \left[ (1 - 1/n) \int K(x) \tilde{K}(t) \right. \]
\[ \left. \{\lambda_0(y + h x - h t) + \lambda_1(y + h (x - t))\} dt \right] w(y) dy, \]
\[ I_2 = (1 - 1/n) \int \psi(y) K(x) \tilde{K}(t) \lambda_0(y + h(x + t)) \]
\[ \{w(y + h x) - w(y)\} dy \ dt \ dx, \]
\[ I_3 = (1 - 1/n) \int \psi(y) K(x) \tilde{K}(t) \lambda_1(y + h(x + t)) \]
\[ \{w(y + h t) - w(y)\} dy \ dt \ dx \]
\[ - \int \psi(y) \tilde{K}(t) \lambda_0(x + h t) \{w(y + h t) - w(y)\} dy \ dt, \]
\[ I_4 = \frac{1}{nh} \int K(x) \psi(y) \left[ \frac{\lambda_0}{H_0} - \frac{\lambda_1}{H_1} \right] (y) w(y + h x) dy \ dx. \]

Some limited expansions provide

\[ I_1 = \frac{h^4}{12} \int \psi(y) \lambda_0^{(4)}(y) w(y) dy \cdot \int [K(x)(x - \delta)^4 - \delta^4] \tilde{K}(t) dt \ dx + o(h^4) \]
\[ = h^4 t_1 \int \psi(y) \lambda_0^{(4)}(y) w(y) dy + o(h^4), \]

where \( t_1 = \int K^{(2)}(x) \). Moreover,

\[ I_2 = h^4 t_1 \int \psi(y) [\lambda_0^{(2)} w^{(2)} / 2 + w^{(1)} \lambda_0^{(3)}(y)] dy + o(h^4), \]

and we have

\[ I_3 = \int \psi(y) \tilde{K}(t) \left[ \int K(x) \lambda_0(y + h(x + t)) dx - \lambda_0(y + h t) \right] \]
\[ \{w(y + h t) - w(y)\} dy \ dt + o(h^4) \]
\[ = \frac{h^2}{2} \int K^{(2)} \int \psi(y) \tilde{K}(t) \lambda_0^{(2)}(y + h(x + t)) \]
\[ \{w(y + h t) - w(y)\} dy \ dt + o(h^4) \]
\[ = h^4 t_1 \int \psi(y) [\lambda_0^{(2)} w^{(2)}(y) / 2 + w^{(1)} \lambda_0^{(3)}(y)] dy + o(h^4). \]
Finally, we have obviously
\[ I_4 = O(n^{-3/2}h^{-1}). \]

Hence,
\[ I = h^4 I_1 \int \psi(y)[\lambda_0(y)w + \lambda_2 w' + 2w'(\lambda_0)(\lambda_0y + o(h^4)). \tag{5.60} \]

The result follows from this formula, (5.59), the definition of \( \eta \) and
\[ \liminf \max_{\lambda} P_{\lambda}(\lambda \neq \lambda_0) \geq p. \tag{5.61} \]

The last point is stated like in (5.30), proving the result.

\section*{References}


