Nonparametric inference for partially observed Levy processes
Gugushvili, S.

Citation for published version (APA):
2 Decompounding

In this chapter, assuming that $X = (X_t)_{t \geq 0}$ is a compound Poisson process with known intensity $\lambda$ and unknown jump size density $f$, we consider the problem of nonparametric estimation of $f$. The estimator is constructed via Fourier inversion and kernel smoothing and the obtained results include pointwise consistency and asymptotic normality of the estimator.

2.1 Introduction

Let $N(\lambda)$ be a Poisson random variable with parameter $\lambda$ and let $Y_1, Y_2, \ldots$ be a sequence of independent and identically distributed random variables independent of $N(\lambda)$ and with a common distribution function $F$ and density $f$. Consider a Poisson sum of $Y$’s:

$$X = \sum_{j=1}^{N(\lambda)} Y_j.$$ 

Assume $\lambda$ is known. The statistical problem we consider is nonparametric estimation of the density $f$ based on observations on $X$. Since adding a Poisson number of $Y$’s is referred to as compounding, we will refer to the problem of recovering the density $f$ of $Y$’s from the observations on $X$ as decompounding. The problem of estimating the density $f$ is equivalent to the problem of estimating the jump size density $f$ of a compound Poisson process $X' = (X'_t)_{t \geq 0}$ with intensity $\lambda$, when the process is observed at equidistant time points (without loss of generality, by a rescaling argument, the observation step size can be taken to be equal to 1.). This follows from the stationary independent increments property of a compound Poisson process. Compound Poisson processes have important applications in queueing and risk theory, see e.g. Embrechts et al. (1997) and Prabhu (1980). The random variables $Y_1, Y_2, Y_3 \ldots$ may be interpreted as claims of random size that arrive at an insurance company or as numbers of customers who arrive at a service point at random times with inter-arrival time distributed exponentially. Figures 2.1 and 2.2 give a typical path of a compound Poisson process when observed continuously and discretely, respectively.

The problem of nonparametric estimation of the distribution function $F$ in case of both continuous and discrete laws was treated in Buchmann and Grübel (2003). Their estimation method is based on a suitable inversion of the compounding operation (i.e. transition from the distribution of $Y$ to the distribution of $X$) and use of an empirical estimator for the distribution of $X$, thus resulting in a plug-in type estimator for the distribution of $Y$. A further ramification of this approach in case of a discrete law is given.
in Buchmann and Gr"ubel (2004). To the best of our knowledge, the present work is the first attempt to (nonparametrically) estimate the density \( f \) instead. A very natural use of nonparametric density estimators is in informal investigation of the properties of a given set of data. They can give valuable indications about the shape of the density function, e.g. such features as multimodality. This might come in handy in applications, e.g. in insurance, where \( f \) is a claim size density.

One possible way to construct an estimator for the density \( f \) (suggested in Hansen and Pitts (2006)) is via smoothing the plug-in type estimator \( F_n \) of the distribution function \( F \) defined in Buchmann and Gr"ubel (2003) with a kernel, but at present no theoretical results for this estimator seem to be available. Moreover, as noted in Hansen and Pitts (2006), the study of the asymptotics of this estimator is a highly nontrivial task. We opt for an alternative approach based on inversion of the characteristic function \( \phi_f \), an approach which is similar in spirit to the use of kernel estimators in deconvolution problems (the latter were first introduced in Liu and Taylor (1989) and Stefanski and Caroll (1990), for a more recent overview see Wand and Jones (1995)). Before we proceed any further, we need to specify the observation scheme. Note that \( X \) can take value zero only if \( N(\lambda) \) does. Therefore, observations equal to zero contain no information on \( Y \) and hence an estimator of \( f \) should be based on the nonzero observations. In a sample of a fixed size there is a random number of nonzero observations. We want to avoid this extra technical complication and we will assume that we have observations \( X_1, \ldots, X_{T_n} \) on \( X \), where \( T_n \) is the first moment when we get precisely \( n \) nonzero observations (\( T_n \) of course is random). We denote the nonzero observations by \( Z_1, Z_2, \ldots, Z_n \).

We turn to the construction of our estimator of the density \( f \). First note that the characteristic function of \( X \) is given by

\[
\phi_X(t) = E[e^{itX}] = e^{-\lambda + \lambda \phi_f(t)},
\]

where \( \phi_f \) denotes the characteristic function of a random variable with density \( f \). Rewrite the characteristic function of \( X \) as

\[
\phi_X(t) = e^{-\lambda} + (1 - e^{-\lambda}) \frac{1}{e^{\lambda \phi_f(t)} - 1} \left( e^{\lambda \phi_f(t)} - 1 \right).
\]

Denote the density of \( X \) given \( N(\lambda) > 0 \) by \( g \). It follows that the characteristic function
of $X$ given $N(\lambda) > 0$ is equal to
\[ \phi_g(t) = \frac{1}{e^{\lambda} - 1} \left( e^{\lambda \phi_f(t)} - 1 \right). \] (2.1)
Since $\phi_f$ vanishes at plus and minus infinity, so does $\phi_g$. Inverting the above relation, we get
\[ \phi_f(t) = \frac{1}{\lambda} \log \left( (e^{\lambda} - 1)\phi_g(t) + 1 \right). \]
Here Log denotes the distinguished logarithm, called so due to its similarity to the distinguished logarithm as defined in Chung (2001, Theorem 7.6.2), Chow and Teicher (1978, Lemma 1, pp. 413–414), Finkelestein et al. (1997) and Sato (2004, Lemma 7.6). The difference lies in the fact that in our case $(e^{\lambda} - 1)\phi_g(0) + 1$ equals $e^{\lambda}$ and not 1. In our case the distinguished logarithm can be defined as
\[ \lambda + \log \left( (e^{-\lambda} - 1)\phi_g(t) + 1 \right) = \lambda + \log(\phi_X(t)), \]
where $\log(\phi_X(t))$ is the same as e.g. in Chung (2001, Theorem 7.6.2), or it can be constructed directly via the method of Chung (2001, Theorem 7.6.2). Notice that whenever $\lambda < \log 2$, the distinguished logarithm reduces to the composition of a principal branch of an ordinary logarithm with $(e^{\lambda} - 1)\phi_g(t) + 1$. This follows from the fact that the condition $\lambda < \log 2$ prevents $(e^{\lambda} - 1)\phi_g(t) + 1$ from taking values on the negative real axis, which constitutes the branch cut for the principal branch of an ordinary logarithm.

**Remark 2.1.** Notice that in general the distinguished logarithm of the non-vanishing characteristic function $\phi(t)$ cannot be reduced to the composition of a principal branch of an ordinary logarithm $\log$ with $\phi$. Consider the following trivial example: $\phi(t) = e^{it}$. This characteristic function satisfies the requirements of Chung (2001, Theorem 7.6.2), since it takes its values on the unit circle in the complex plane and hence its distinguished logarithm exists and is given by $\log(\phi(t)) = it$. On the other hand if one considers the argument of $\log(\phi(t))$, it is easy to see that it has a jump whenever $\phi$ crosses the negative real axis, see Figure 2.3 on the following page and compare to the argument of the distinguished logarithm. This fact is not surprising, given that $-1$ lies on the branch cut of the principal branch of an ordinary logarithm.

By Fourier inversion, if $\phi_f$ is integrable, we have
\[ f(x) = \frac{1}{2\pi \lambda} \int_{-\infty}^{\infty} e^{-itx} \log \left( (e^{\lambda} - 1)\phi_g(t) + 1 \right) dt. \] (2.2)
This relation suggests that if we construct an estimator of $g$ (and hence of $\phi_g$), we will automatically get an estimator for $f$ by a plug-in device. Let $w$ denote a kernel function with characteristic function $\phi_w$ and let $h$ denote a positive number, the bandwidth. The density $g$ will be estimated by the kernel density estimator
\[ g_{nh}(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h} w \left( \frac{x - Z_j}{h} \right). \] (2.3)
and Wasserman (2007). The characteristic function $\phi_{\text{g-nh}}$ serves as an estimator of $\phi_\theta$ and it is equal to $\phi_{\text{emp}}(t)\phi_w(ht)$, where $\phi_{\text{emp}}$ denotes the empirical characteristic function

$$\phi_{\text{emp}}(t) = \frac{1}{n} \sum_{j=1}^{n} e^{itZ_j}.$$ 

In view of (2.2), it is tempting to introduce the estimator

$$\frac{1}{2\pi \lambda} \int_{-\infty}^{\infty} e^{-itx} \log \left( (e^\lambda - 1)\phi_{\text{emp}}(t)\phi_w(ht) + 1 \right) dt. \tag{2.4}$$

However the problem is that the measure of the set of those $\omega$'s from the underlying sample space $\Omega$ for which the Log is undefined may be positive, since the path $(e^\lambda - 1)\phi_{\text{g-nh}}(t) + 1$ can take the value zero at some $t$ (though as $n \to \infty$ this probability tends to zero). Moreover, there is no guarantee that the integral in (2.4) is finite. Therefore we will adapt the expression in (2.4) and define our estimator as

$$\hat{f}_{nh}(x) = (M_n \land f_{nh}(x)) \lor (-M_n). \tag{2.5}$$

where $f_{nh}$ is given by

$$f_{nh}(x) = \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \log \left( (e^\lambda - 1)\phi_{\text{emp}}(t)\phi_w(ht) + 1 \right) dt, \tag{2.6}$$

for those $\omega$'s for which $(e^\lambda - 1)\phi_{\text{emp}}(t)\phi_w(ht) + 1$ vanishes nowhere in $[-1/h, 1/h]$ and $f_{nh}$ is zero otherwise. Here $M = (M_n)_{n \geq 1}$ is a sequence of positive real numbers converging to infinity at a suitable rate to be specified below. We also assume that $\phi_w$ is supported on $[-1, 1]$. Of course, for the truncation in (2.5) to make sense, $f_{nh}(x)$ must be real-valued, but this holds as is easy to check through the change of the integration variable from $t$ into $-t$. 

Figure 2.3: Arguments of the principal branch of a logarithm and the distinguished logarithm.
2.2 Asymptotic properties of the estimator

The rest of the chapter is organised as follows: Section 2.2 contains the main results of the chapter. In it we derive an order bound for the bias and an asymptotic expansion of the variance of $\hat{f}_{nh}(x)$ at a fixed point $x$ and we show that the estimator is weakly consistent and asymptotically normal. Section 2.3 provides some simulation results. All the proofs are collected in Section 2.4.

2.2 Asymptotic properties of the estimator

As usual, the nonparametric setting forces us to put some smoothness conditions on the density $f$. Let $\beta, L_1$ and $L_2$ denote some positive numbers and let $l = \lfloor \beta \rfloor$, the integer part of $\beta$. If $l = 0$, then by definition set $f^{(l)} = f$. Recall the definition of the Hölder and Nikol’ski classes of the functions, cf. Tsybakov (2004, p. 5 and 19).

**Definition 2.1.** A function $f$ is said to belong to the Hölder class $H(\beta, L_1)$, if its derivatives up to order $l$ exist and satisfy the condition

$$ |f^{(l)}(x + t) - f^{(l)}(x)| \leq L_1 |t|^{\beta-l}.$$  

for $x \in \mathbb{R}, t \in \mathbb{R}$.

**Definition 2.2.** A function $f$ is said to belong to the Nikol’ski class $N(\beta, L_2)$, if its derivatives up to order $l$ exist and verify the condition

$$ \left( \int_{-\infty}^{\infty} (f^{(l)}(x + t) - f^{(l)}(x))^2 dx \right)^{1/2} \leq L_2 |t|^{\beta-l}.$$  

for $t \in \mathbb{R}$.

We formulate our condition on the density $f$.

**Condition 2.1.** The density $f$ belongs to $H(\beta, L_1) \cap N(\beta, L_2)$ and $xf(x)$ is integrable. Moreover, $t^\beta \phi_f(t)$ is integrable and the derivatives $f', \ldots, f^{(l)}$ are integrable.

The following lemma holds true.

**Lemma 2.1.** Assume that Condition 2.1 holds. Then the density $g$ belongs to $H(\beta, L_1) \cap N(\beta, \lambda(1 - e^{-\lambda})^{-1}L_2)$. Moreover, $t^\beta \phi_g(t)$ is integrable.

We will use this fact in the proofs of Propositions 2.1, 2.2 and Theorems 2.2 and 2.3. The property $g \in N(\beta, \lambda e^{\lambda}(e^{\lambda} - 1)^{-1}L_2)$ is needed in our proofs where we will make use of an expansion of the mean integrated square error of a kernel density estimator $g_{nh}$, cf. Tsybakov (2004, p. 21). The condition $g \in H(\beta, L_1)$ is standard in ordinary kernel density estimation, see Tsybakov (2004, Proposition 1.2). The integrability of $f', \ldots, f^{(l)}$ is used in the proof of Lemma 2.1.

**Definition 2.3.** Let $l \geq 1$ be an integer. A function $w$ is called a kernel of order $l$, if the functions $w^j(u), j = 0, \ldots, l$, are integrable and satisfy the condition

$$ \int_{-\infty}^{\infty} w(u) du = 1, \int_{-\infty}^{\infty} u^j w(u) du = 0$$  

for $j = 1, \ldots, l$. 


Since it is generally recognised that the choice of a kernel is less important for the performance of an estimator (see Wand and Jones (1995, p. 31)), we feel free to impose the following condition on the kernel.

**Condition 2.2.** The kernel function \( w \) satisfies the conditions:

1. \( w \) is a bounded symmetric kernel of order \( l \);
2. The support of the Fourier transform \( \phi_w \) is contained in \([−1, 1] \);
3. \( \int_{−\infty}^{\infty} |u|^\beta |w(u)|du < \infty \);
4. \( \lim_{|u| \to \infty} |uw(u)| = 0 \).

In order to get a consistent estimator, we need to control the bandwidth and we impose the following restriction on it.

**Condition 2.3.** The bandwidth \( h \) depends on \( n \) and is of the form \( h = Cn^{-\gamma} \) for \( 0 < \gamma < 1 \), where \( C \) is some constant.

We also formulate a condition on the truncating sequence \( M = (M_n)_{n \geq 1} \) (see Section 2.1).

**Condition 2.4.** The truncating sequence \( M = (M_n)_{n \geq 1} \) is given by \( M_n = n^{\alpha} \), where \( \alpha \) is some strictly positive number, \( \alpha > \gamma \).

As the performance criterion we select the mean square error

\[
\text{MSE}[\hat{f}_{nh}(x)] = E[(\hat{f}_{nh}(x) - f(x))^2].
\]

By standard properties of mean and variance we have that

\[
\text{MSE}[\hat{f}_{nh}(x)] = (E[\hat{f}_{nh}(x)] - f(x))^2 + \text{Var}[\hat{f}_{nh}(x)].
\]

First we study the behaviour of the bias of the estimator \( \hat{f}_{nh}(x) \).

**Proposition 2.1.** Suppose that Conditions 2.1–2.4 are satisfied. Then the bias of the estimator \( \hat{f}_{nh}(x) \) admits the pointwise order bound

\[
E[\hat{f}_{nh}(x)] - f(x) = O\left(h^\beta + \frac{1}{nh}\right).
\]

In ordinary kernel density estimation, under the assumption \( g \in \mathcal{H}(\beta, L_1) \), the bias is of order \( h^\beta \), see Tsybakov (2004, Proposition 1.2). We have an additional term of order \( (nh)^{-1} \) coming from the difficulty of the decompounding problem. Under standard conditions \( h \to 0, nh \to \infty \), the bias will asymptotically vanish.

**Remark 2.2.** If \( \beta = 2 \), then it is possible to derive an exact asymptotic expansion for the bias. The leading term in the bias expansion will be

\[
-h^2 \frac{\sigma^2(e^\lambda - 1)}{4\pi \lambda} \int_{-\infty}^{\infty} e^{-itx} \frac{t^2 \phi_g(t)}{(e^\lambda - 1)\phi_g(t) + 1} dt.
\]
Now let us study the variance of the estimator $\hat{f}_{nh}(x)$.

**Proposition 2.2.** Suppose that apart of Conditions 2.1–2.4, an additional condition $nh^{1+4\beta} \to 0$ holds true. Then the variance of the estimator $\hat{f}_{nh}(x)$ admits the pointwise decomposition

$$\text{Var}[\hat{f}_{nh}(x)] = \frac{1}{nh} \frac{(e^\lambda - 1)^2}{\lambda^2} g(x) \int_{-\infty}^{\infty} (w(u))^2 du + o\left(\frac{1}{nh}\right). \quad (2.7)$$

We see that the variance of our estimator is of the same order as that of an ordinary kernel density estimator, cf. Tsybakov (2004, Proposition 1.4). Under the standard assumption $nh \to \infty$ it will vanish. From a practical point of view the restriction $nh^{1+4\beta} \to 0$ is not too restrictive, especially in view of Proposition 2.3 given below.

Combining Propositions 2.1 and 2.2, we get the following corollary.

**Corollary 2.1.** Suppose Conditions 2.1–2.4 hold. The estimator $\hat{f}_{nh}(x)$ is pointwise weakly consistent under the additional assumption $nh^{1+4\beta} \to 0$.

The bandwidth $h_{opt}$ that asymptotically minimises the mean square error of a kernel estimator is called optimal. From Propositions 2.1 and 2.2 it is now possible to determine the order of the optimal bandwidth for the estimator $\hat{f}_{nh}$.

**Proposition 2.3.** The optimal bandwidth $h_{opt}$ is of order $n^{-1/(2\beta+1)}$. Furthermore, the mean square error of the estimator $\hat{f}_{nh}$ computed for the optimal bandwidth is of order $n^{-2\beta/(2\beta+1)}$.

Note that the optimal bandwidth is of order $n^{-1/(2\beta+1)}$, just as in case of ordinary kernel density estimation.

**Remark 2.3.** When $\beta = 2$, then it is possible to derive an exact expression for $h_{opt}$.

$$h_{opt} = \left( \frac{4\pi^2 g(x) \int_{-\infty}^{\infty} (w(u))^2 du}{\sigma^4 \left( \int_{-\infty}^{\infty} e^{-itx} t^2 \phi_g(t)((e^\lambda - 1)\phi_g(t) + 1)^{-1} dt \right)^2} \right)^{1/5} n^{-1/5}.$$  

An extension of our results to the data-dependent bandwidth case is outside the scope of the present work.

It is interesting to verify whether our estimator is minimax. The theorem below and the corollary thereof demonstrate that the minimax convergence rate cannot be better than $n^{-\beta/(2\beta+1)}$, however at present we do not know whether our estimator itself is minimax. In any case, the results of the present section show that its behaviour is quite reasonable.

**Theorem 2.1.** Let $X = (X_t)_{t \geq 0}$ be a compound Poisson process with intensity $\lambda$ and jump size density $f$ and let $\mathcal{F}$ denote the class of jump size densities $f$ satisfying Condition 2.1. Then for an arbitrary $x_0 \in \mathbb{R}$ the inequality

$$\liminf_{n \to \infty} \inf_{f \in \mathcal{F}} \sup_{f_{T_n}} \mathbb{E} \left[ n^{2\beta/(2\beta+1)} |f_{T_n}(x_0) - f(x_0)|^2 \right] > 0. \quad (2.8)$$

is valid. Here the infimum is taken over all estimators $f_{T_n}$ based on continuous observations of $X = (X_t)_{t \geq 0}$ over the interval $[0, T_n]$, where $T_n$ is the first moment, when $n$ quantities among $X_1, X_2 - X_1, \ldots, X_{T_n} - X_{T_n-1}$ are nonzero.
This theorem provides the minimax convergence rate for the estimation of the density \( f \) in case the compound Poisson process is observed continuously over a (random) interval \([0, T_n]\). The proof is similar to the one used in Figueroa-Lopez and Houdré (2004), see also Ibragimov and Has’minskii (1981, Chapter 4, Theorem 5.1) and Kutoyants (1998, Theorem 6.5). Because the infimum in (2.8) over all estimators based on observations over time interval \([0, T_n]\) is obviously smaller than the same infimum taken over all estimators based on a discrete sample of size \( n \) from \( X \), the following corollary holds true.

**Corollary 2.2.** The minimax convergence rate for the decompounding problem cannot be better than \( n^{-\beta/(2\beta + 1)} \).

Concluding this section, we will derive two asymptotic normality results for \( \hat{f}_{nh} \).

**Theorem 2.2.** Assume that Conditions 2.1–2.4 hold, and that the bandwidth \( h \) satisfies the additional condition \( nh^{2\beta + 1} \to 0 \). Suppose \( x \) is such that \( g(x) \neq 0 \). Then

\[
\left( \frac{\hat{f}_{nh}(x) - f(x)}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} \right) \overset{D}{\to} N(0, 1),
\]

where \( N(0, 1) \) is the standard normal distribution.

Asymptotic normality still holds if \( nh^{2\beta + 1} \to C \), where \( C \) is some constant, but in this case the limit will not be distribution free, it will depend on the unknown function \( g \). We cannot select an optimal bandwidth to obtain (distribution free) asymptotic normality, but this is also the case in ordinary kernel density estimation. This fact comes from the trade-off between bias and variance, for details see the proof of the theorem. Now let us consider a different centring, \( \hat{f}_{nh}(x) - E[\hat{f}_{nh}(x)] \). Then the following theorem holds true.

**Theorem 2.3.** Suppose that Conditions 2.1–2.4 hold, \( g(x) \neq 0 \) and \( nh^{1+4\beta} \to 0 \). Then we have

\[
\left( \frac{\hat{f}_{nh}(x) - E[\hat{f}_{nh}(x)]}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} \right) \overset{D}{\to} N(0, 1).
\]

We see that in this case the additional condition on the bandwidth is weaker than the one from Theorem 2.2.

### 2.3 Simulation results and numerical aspects

In this section we present two simulation examples. These complement the asymptotic results of Theorems 2.1 and 2.2 and give some, albeit incomplete, indication of the finite sample properties of the estimator.

In the first example the true density \( f \) is the standard normal density and \( \lambda = 0.3 \). The kernel we used is from Fan (1992) and has a rather complicated expression

\[
w(x) = \frac{48x(x^2 - 15)\cos x - 144(2x^2 - 5)\sin x}{\pi x^7},
\] (2.9)
but its characteristic function looks much simpler and is given by

\[ \phi_w(t) = (1 - t^2)^{1/2} \text{1}_{|t|<1}. \]

The kernel \( w \) and its Fourier transform are plotted in Figures 2.4 and 2.5, respectively.

The estimator is based on 1000 observations and the bandwidth equals 0.14 (the bandwidth was selected by hand). To compute the estimator we used the Fast Fourier Transform. For more information on the Fast Fourier Transform see e.g. Brigham (1974). Our approach, in spirit close to the method for numerical evaluation of option prices proposed in Carr and Madan (1998), is as follows:

(i) As noted in Section 2.1, whenever \( \lambda < \log 2 \), the distinguished logarithm reduces to the composition of a principal branch of an ordinary logarithm with \((e^\lambda - 1)\phi_g(t) + 1\).
(ii) The main use of truncation in (2.5) is to prove asymptotic properties of the estimator and in general in practice we do not need to use it.
(iii) The computation of the empirical characteristic function can be significantly sped up by grouping the observations into bins. This idea is used to numerically evaluate ordinary kernel density estimators. However we computed the empirical characteristic function directly, without grouping the observations. Notice that we do not use the values of the empirical characteristic function in its tails.
(iv) Notice that we can rewrite (2.6) as

\[ f_{nh}(x) = f_{nh}^{(1)}(x) + f_{nh}^{(2)}(x), \]

where

\[ f_{nh}^{(1)}(x) = \frac{1}{2\pi\lambda} \int_0^\infty e^{-i\pi x} \log ((e^\lambda - 1)\phi_{\text{emp}}(t)\phi_w(ht) + 1) \, dt, \]

\[ f_{nh}^{(2)}(x) = \frac{1}{2\pi\lambda} \int_0^\infty e^{i\pi x} \log ((e^\lambda - 1)\phi_{\text{emp}}(-t)\phi_w(ht) + 1) \, dt. \]

Using the trapezoid rule and setting \( v_j = \eta(j - 1) \), \( f_{nh}^{(1)}(x) \) can be approximated by

\[ f_{nh}^{(1)}(x) \approx \frac{1}{2\pi\lambda} \sum_{j=1}^N e^{-iv_j x} \psi(v_j) \eta. \]

Here we take \( N \) to be some power of 2 and \( \psi(v_j) = \log ((e^\lambda - 1)\phi_{g_{nh}}(v_j) + 1) \). The application of the Fast Fourier Transform to this sum will give us \( N \) values of \( f_{nh}^{(1)} \) and...
we employ a regular spacing size $\delta$, so that our values of $x$ are

$$x_u = \frac{-N\lambda}{2} + \delta(u - 1),$$

where $u = 1, \ldots, N$. Thus we have

$$f^{(1)}_{nh}(x_u) \approx \frac{1}{2\pi\lambda} \sum_{j=1}^{N} e^{-i\delta\eta(j-1)(u-1)} e^{iv_j \frac{N\lambda}{2}} \psi(v_j) \eta,$$

for $u = 1, \ldots, N$. To apply the Fast Fourier Transform, we note that we must take $\delta\eta = 2\pi/N$. If we choose $\eta$ small to obtain a fine grid for integration, then we will obtain values of $f^{(1)}_{nh}$ at values of $x_u$ relatively far from each other. We would like therefore to obtain an accurate integration for larger values of $\eta$ and to this end we incorporate the Simpson rule’s weights into our summation, i.e.

$$f^{(1)}_{nh}(x_u) \approx \frac{1}{2\pi\lambda} \sum_{j=1}^{N} e^{-i\frac{2\pi}{N}(j-1)(u-1)} e^{iv_j \frac{N\lambda}{2}} \psi(v_j) \eta \frac{3 + (-1)^j - \delta_j}{3},$$

where $\delta_j$ is the Kronecker function. A similar reasoning applies to $f^{(2)}_{nh}(x)$.

The result of this procedure is given in Figure 2.6 (the estimate is represented by a bold dotted line).

![Figure 2.6: Estimation of the standard normal density.](image1)

In the second example we consider the case when $f$ is a mixture of two normal densities with means 0 and $3/2$ and variances $1$ and $1/9$ with mixing probabilities $3/4$ and $1/4$, respectively. The estimator is based on 1000 observations and the bandwidth equals 0.1; the kernel is the same as in the first example. The result is given in Figure 2.7 (the estimate is plotted by a bold dotted line). Note that the estimator captures the bimodal character of the density $f$ in a quite satisfactory manner.

### 2.4 Proofs

**Proof of Lemma 2.1.** We have

$$|\phi_f(t)| \leq \frac{\lambda e^\lambda}{e^\lambda - 1} |\phi_f(t)|,$$  \hspace{1cm} (2.10)
which follows from the relation

\[ \phi_g(t) = \frac{1}{e^\lambda - 1} \left( e^{\lambda \phi_f(t)} - 1 \right) \]

by

\[ \left| e^{\lambda \phi_f(t)} - 1 \right| = |-1 + 1 + \lambda \phi_f(t) + \ldots | \leq \lambda \phi_f(t) | e^{\lambda \phi_f(t)} | \leq \lambda e^\lambda | \phi_f(t) |. \]  

(2.11)

This implies that \( t^\beta \phi_g(t) \) is integrable. Furthermore,

\[ g(x) = \sum_{n=1}^{\infty} f^{*n}(x) P(N = n | N > 0), \]

where \( f^{*n} \) denotes the \( n \)-fold convolution of \( f \). By Parseval’s theorem

\[ \int_{-\infty}^{\infty} (g^{(l)}(x + t) - g^{(l)}(x))^2dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} | u' \phi_g(u) |^2 e^{iux} - 1 |^2du, \]  

(2.12)

where we used the fact that \( | \phi_g(u) | = | u' \phi_g(u) |, \) see Schwartz (1966, formula (V,1;25)). The latter formula is applicable, because the derivatives of \( g(x) \) up to order \( l \) are integrable, which can be verified by direct computation employing Schwartz (1966, formula (III,2;8)). From (2.12) and (2.10) it follows that

\[ \int_{-\infty}^{\infty} (g^{(l)}(x + t) - g^{(l)}(x))^2dx \leq \frac{1}{2\pi} \left( \frac{\lambda e^\lambda}{e^\lambda - 1} \right)^2 \int_{-\infty}^{\infty} | u' \phi_f(u) |^2 e^{iux} - 1 |^2du. \]

Applying Parseval’s theorem to the right-hand side and recalling that \( f \) belongs to \( \mathcal{N}(\beta, L_2) \), we conclude that \( g \) belongs to \( \mathcal{N}(\beta, \lambda e^\lambda (e^\lambda - 1)^{-1} L_2) \).

Now we will verify that \( g \in \mathcal{H}(\beta, L_1) \). We have

\[ g^{(l)}(x) = \sum_{n=1}^{\infty} f^{*(n-1)} \ast f^{(l)}(x) P(N = n | N > 0). \]

Using this expression, we get

\[ |g^{(l)}(x + t) - g^{(l)}(x)| \]

\[ = \left| \sum_{n=1}^{\infty} P(N = n | N > 0) \int_{-\infty}^{\infty} (f^{(l)}(x + t - u) - f^{(l)}(x - u)) f^{*(n-1)}(u) du \right| \]

\[ \leq L_1 |t|^{\beta - l} \sum_{n=1}^{\infty} P(N = n | N > 0) \int_{-\infty}^{\infty} f^{*(n-1)}(u) du = L_1 |t|^{\beta - l}. \]

This completes the proof of the lemma. \( \square \)

In the proof of the bias expansion we will make use of the following two lemmas. The first one is a result by Devroye, see Devroye (1991, p. 36, Remark 3). Notice that the conditions on density \( g \) are satisfied in view of Lemma 2.1.
Lemma 2.2. Let $g_{nh}$ be a kernel estimator and let us introduce

$$J_n = \int_{-\infty}^{\infty} |g_{nh}(x) - g(x)|dx,$$

the $L^1$ error of the estimator $g_{nh}$. Then for arbitrary positive $\delta$ and $\varepsilon$ there exists an integer $n_0$ such that for all $n > n_0$ the following exponential bound holds:

$$P(J_n > \delta) \leq 2 \exp\left(\frac{-n(1-\varepsilon)\delta^2}{2(e^{\int_{-\infty}^{\infty} w(x)|dx|^2})}\right).$$

Lemma 2.3. Let $z$ be a complex number and suppose $|e^z - 1| < 1/2$. Then the inequality

$$|e^z - 1 - z| \leq |e^z - 1|^2$$

holds true.

Proof. The inequality can be verified as follows:

$$|e^z - 1 - z| = \left| \int_0^1 t(e^z - 1)^2 \frac{dt}{1 + t(e^z - 1)} \right| \leq |e^z - 1|^2 \int_0^1 \left| \frac{t}{1 + t(e^z - 1)} \right| dt \leq |e^z - 1|^2,$$

where we used the fact that $|e^z - 1| < 1/2$.

Now we proceed with the proof of Proposition 2.1.

Proof of Proposition 2.1. We may write

$$b_w(n,h,x) = E[\hat{f}_{nh}(x)1_{J_n \leq \delta} + \hat{f}_{nh}(x)1_{J_n > \delta} - f(x)1_{J_n \leq \delta}] - f(x)P(J_n > \delta),$$

where $\delta$ is any positive number and $J_n$ denotes the $L_1$ error of the estimator $g_{nh}$ defined by (2.3). We have

$$|E[\hat{f}_{nh}(x)1_{J_n > \delta}]| \leq M_n P(J_n > \delta).$$

To see that this term is of lower order than $h^\beta$, recall the special form of $M_n$ and $h$ and apply the exponential bound of Lemma 2.2, valid for all $n$ sufficiently large, to $P(J_n > \delta)$. Also $f(x)P(J_n > \delta) = o(h^\beta)$.

Now we turn to $E[(\hat{f}_{nh}(x) - f(x))1_{J_n \leq \delta}]$. Notice that on the set $\{J_n \leq \delta\}$ we have

$$|\phi_{g_{nh}}(t) - \phi_g(t)| = \left| \int_{-\infty}^{\infty} e^{-itx}(g_{nh}(x) - g(x))dx \right| \leq J_n \leq \delta. \quad (2.13)$$

This in turn implies that for $\delta$ sufficiently small, e.g. $\delta = e^{-\lambda}/2$, on the set $\{J_n \leq \delta\}$ the function $(e^\lambda - 1)\phi_{g_{nh}}(t) + 1$ is bounded away from zero, because

$$|(e^\lambda - 1)\phi_g(t) + 1| = |e^{\lambda\phi(t)}| \geq e^{-\lambda}. \quad (2.14)$$
Therefore, the distinguished logarithm of \((e^\lambda - 1)\phi_{\gamma h}(t) + 1\) will be well-defined on this set. Furthermore, \(\log((e^\lambda - 1)\phi_{\gamma h}(t) + 1)\), i.e. the real part of its distinguished logarithm will be bounded on \(\{J_n \leq \delta\}\), because \((e^\lambda - 1)\phi_{\gamma h}(t) + 1\) will stay away from zero at a positive distance, and because it is bounded by \(e^\lambda\). Also the imaginary part of \(\log((e^\lambda - 1)\phi_{\gamma h}(t) + 1)\) is bounded on the same set. Indeed, let \(\psi : \mathbb{R} \to \mathbb{C}\), where

\[
\psi(t) = e^{\lambda \phi_f(t)}.
\]

By the Riemann-Lebesgue theorem \(\psi(t)\) converges to 1 as \(|t| \to \infty\) and hence there exists \(t^* > 0\), such that

\[
|\psi(t) - 1| < \frac{1}{2}, \quad |t| > t^*.
\]

Furthermore, we have

\[
|\psi(t)| \geq e^{-\lambda}, \quad t \in \mathbb{R}.
\]

Since \(f\) has the finite first moment, by Schwartz (1966, Theorem 1, p. 182) \(\phi_f\) and consequently \(\psi\) are continuously differentiable. Therefore, the path \(\psi : [-t^*, t^*] \to \mathbb{C}\) is rectifiable, i.e. has a finite length. In view of this fact and (2.16), \(\psi : [-t^*, t^*] \to \mathbb{C}\) cannot spiral infinitely many times around zero and for \(|t| > t^*\) it cannot make a turn around zero at all because of (2.15). Thus on the set \(\{J_n \leq \delta\}\), the distinguished logarithm \(\log((e^\lambda - 1)\phi_{\gamma h}(t) + 1)\) will be bounded for \(\delta\) small. Consequently

\[
\left| \int_{-1/h}^{1/h} e^{-itx} \log \left( (e^\lambda - 1)\phi_{\gamma h}(t) + 1 \right) dt \right|_{J_n \leq \delta} \leq C_1 \frac{2}{h} = C_2 n^\gamma,
\]

where \(C_1\) and \(C_2\) are some constants. Because \(\alpha > \gamma\), the expression above grows slower than the sequence \(M = (M_n)_{n \geq 1}\), and for large enough \(n\) this will imply that \(f_{nh}(x) = f_{nh}(x)\) on the set \(\{J_n \leq \delta\}\). Therefore

\[
E[(f_{nh}(x) - f(x))1_{\{J_n \leq \delta\}}] = \frac{1}{2\pi \lambda} E \left[ \left( \int_{-1/h}^{1/h} e^{-itx} \log \left( (e^\lambda - 1)\phi_{\gamma h}(t) + 1 \right) dt \right. \right.
\]

\[
- \left. \int_{-1/h}^{1/h} e^{-itx} \log((e^\lambda - 1)\phi_f(t) + 1) dt \right) 1_{\{J_n \leq \delta\}}]
\]

\[
- \frac{1}{2\pi \lambda} \int_{-\infty}^{-1/h} e^{-itx} \phi_f(t) dt P(J_n \leq \delta) - \frac{1}{2\pi \lambda} \int_{1/h}^{\infty} e^{-itx} \phi_f(t) dt P(J_n \leq \delta).
\]

The last two terms in this expression are of lower order than \(h^3\). Indeed, we have e.g.

\[
\lim_{h \to 0} \frac{1}{h^3} \int_{1/h}^{\infty} e^{-itx} \phi_f(t) dt \leq \lim_{h \to 0} \frac{1}{h^3} \int_{1/h}^{\infty} |\phi_f(t)| dt
\]

\[
\leq \lim_{h \to 0} \int_{1/h}^{\infty} t^3 |\phi_f(t)| dt = o(1). \quad (2.17)
\]
Hence we need to study
\[
\frac{1}{2\pi \lambda} E \left[ \int_{-1/h}^{1/h} e^{-itx} \log((e^\lambda - 1)\phi_{g\lambda}(t) + 1)dt - \int_{-1/h}^{1/h} e^{-itx} \log((e^\lambda - 1)\phi_g(t) + 1)dt \right]_{[J_n \leq \delta]}
\]
where \( \phi_{g\lambda}(t) = \frac{(e^\lambda - 1)(\phi_{g\lambda}(t) - \phi_g(t))}{(e^\lambda - 1)\phi_g(t) + 1} \). Note that \( z_{nh}(t) = \frac{(e^\lambda - 1)(\phi_{g\lambda}(t) - \phi_g(t))}{(e^\lambda - 1)\phi_g(t) + 1} \).

Rewrite (2.18) as
\[
\frac{1}{2\pi \lambda} E \left[ \int_{-1/h}^{1/h} e^{-itx} z_{nh}(t)dt \right]_{[J_n \leq \delta]}
\]
where
\[
R_{nh}(t) = \log(1 + z_{nh}(t)) - z_{nh}(t).
\]

We bound the first term in (2.19) as follows:
\[
\left| E \int_{-1/h}^{1/h} e^{-itx} z_{nh}(t)dt \right|_{[J_n \leq \delta]}
\leq \left| E \int_{-1/h}^{1/h} e^{-itx}\phi_{g\lambda}(t)dt \right|_{[J_n > \delta]}
\]
The second term on the right-hand side is bounded by \( Ch^{-1} P(J_n > \delta) \), where \( C \) is some constant, and this is of lower order than \( h^3 \) (recall the exponential bound of Lemma 2.2 on \( P(J_n > \delta) \)).

Using the fact that \( E[\phi_{emp}(t)] = \phi_g(t) \), we then obtain
\[
\frac{1}{2\pi \lambda} E \left[ \int_{-1/h}^{1/h} e^{-itx} \left( \frac{e^\lambda - 1)(\phi_{emp}(t)\phi_{w}(ht) - \phi_g(t))}{(e^\lambda - 1)\phi_g(t) + 1} \right)dt \right]
\]
\[
= \frac{e^\lambda - 1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \phi_g(t)\phi_{w}(ht) - \phi_g(t) dt
\]
\[
= \frac{e^\lambda - 1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} (\phi_g(t)\phi_{w}(ht) - \phi_g(t))dt
\]
\[
+ \frac{e^\lambda - 1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} (\phi_g(t)\phi_{w}(ht) - \phi_g(t))(e^{-\lambda\phi_g(t)} - 1)dt.
\]
The first summand in the latter expression differs from the bias of the kernel estimator \( g_{nh}(x) \) up to the factor \((e^\lambda - 1)/\lambda\) only by absence of the term

\[
- \int_{-\infty}^{1/h} \phi_g(t) dt - \int_{1/h}^{\infty} \phi_g(t) dt.
\]

This additional term is of lower order than \( h^{3/2} \), cf. (2.17) and Lemma 2.1. Under Conditions 2.2 and 2.1 and due to Lemma 2.1, the bias of \( g_{nh}(x) \) is of order \( h^{3/2} \), see Tsybakov (2004, Proposition 1.2). As far as the second summand in (2.20) is concerned, it is bounded by

\[
\lambda e^{\lambda} e^{\lambda} - 1 \left( \frac{1}{2 \pi \lambda} \int_{-1/h}^{1/h} |\phi_g(t)\phi_w(ht) - \phi_g(t)||\phi_f(t)| dt \right),
\]

(2.21)

since

\[
|e^{-\lambda\phi_f(t)} - 1| \leq \lambda e^{|\phi_f(t)|},
\]

(2.22)

see (2.11). Application of the Cauchy-Schwarz inequality to the integral in (2.21) yields that it is bounded from above by

\[
\sqrt{\int_{-1/h}^{1/h} |\phi_g(t)\phi_w(ht) - \phi_g(t)|^2 dt} \sqrt{\int_{-1/h}^{1/h} |\phi_f(t)|^2 dt}.
\]

The second factor in this expression is bounded uniformly in \( h \) thanks to the fact that \( \phi_f \) is integrable (\( |\phi_f(t)|^2 \) consequently is also integrable). As far as the first factor is concerned, up to the factor \( 2\pi \) by Parseval’s theorem it is bounded by the integrated square bias of the estimator \( g_{nh} \),

\[
\int_{-\infty}^{\infty} (g \ast w_h(x) - g(x))^2 dx,
\]

where \( w_h(x) = w (x/h) / h \). Since under Conditions 2.1 and 2.2 the integrated square bias of \( g_{nh} \) is of order \( h^{3/2} \) (see Tsybakov (2004, Proposition 1.8)), we conclude that (2.20) is of order \( h^{3/2} \). This gives us the order of the leading term (2.20) in the bias expansion.

Now we turn to the second term in (2.19). We have

\[
\left| \mathbb{E} \left[ \int_{-1/h}^{1/h} e^{-itx} R_{nh}(t) dt 1[J_n \leq \delta] \right] \right| \leq \mathbb{E} \left[ \int_{-1/h}^{1/h} |R_{nh}(t)| dt 1[J_n \leq \delta] \right].
\]

In order to deal with this term we will need the following inequality

\[
|\log(1 + z_{nh}(t)) - z_{nh}(t)| \leq |z_{nh}(t)|^2,
\]

(2.23)

provided that \(|z_{nh}(t)| < 1/2\). This inequality follows from Lemma 2.3 if we take \( z = \log(1 + z_{nh}(t)) \), since by choosing \( n \) large enough and \( \delta \) small, \( J_n \leq \delta \) will entail \(|z_{nh}(t)| < 1/2\).
the above expression into (2.25), we obtain that it is equal to

\[
\mathbb{E} \left[ \int_{-1/h}^{1/h} |R_{nh}(t)| dt 1_{[J_n \leq \delta]} \right] \leq \mathbb{E} \left[ \int_{-\infty}^{\infty} |z_{nh}(t)|^2 dt \right] \\
\leq K \mathbb{E} \left[ \int_{-\infty}^{\infty} |\phi_{emp}(t)\phi_w(ht) - \phi_\beta(t)|^2 dt \right] \\
= K 2\pi \mathbb{E} \left[ \int_{-\infty}^{\infty} (g_{nh}(t) - g(t))^2 dt \right] = K 2\pi \text{MISE}_n(h),
\tag{2.24}
\]

where \( K \) is a constant. Here we used the fact that \(|(e^\lambda - 1)\phi_\beta(t) + 1| = e^{\lambda|x|}t\) is bounded from below, see (2.14), and Parseval’s identity. Using the bound on \( \text{MISE}_n(h) \) of Tsybakov (2004, p. 21) and combining it with (2.20), we establish the desired result.

**Proof of Remark 2.2.** The proof is very similar to the proof of Proposition 2.1. After repeating the same steps as in the latter, we consider the middle term of (2.20),

\[
e^\lambda - 1 \\
2\pi \lambda \int_{-1/h}^{1/h} e^{-itx} \frac{\phi_\beta(t)\phi_w(ht) - \phi_\beta(t)}{(e^\lambda - 1)\phi_\beta(t) + 1} dt.
\tag{2.25}
\]

For \( \beta = 2 \) we have under Condition 2.2

\[
\phi_w(ht) = 1 - \frac{1}{2}\sigma^2 h^2 t^2 - \frac{\rho(ht)}{2} h^2 t^2,
\]

see Shiryaev (1984). Here \( \sigma^2 \) is the second moment of \( w \) and \( \rho(ht) \to 0 \). By plugging the above expression into (2.25), we obtain that it is equal to

\[
\mathbb{E}[\hat{f}_{nh}(x)] - f(x) = -h^2 \frac{\sigma^2(e^\lambda - 1)}{4\pi \lambda} \int_{-\infty}^{\infty} \frac{t^2 \phi_\beta(t)}{(e^\lambda - 1)\phi_\beta(t) + 1} dt + o(h^2).
\]

which follows from the dominated convergence theorem, because \( \rho \) is bounded. The remaining part of the proof, dealing in particular with (2.19), follows the same lines as the proof of Proposition 2.1.

**Proof of Proposition 2.2.** Throughout the proof we will frequently use the following version of the Cauchy-Schwarz inequality: if \( \xi \) and \( \eta \) are random variables, then \(|\text{Cov}[\xi, \eta]| \leq \sqrt{\text{Var}[\xi]} \sqrt{\text{Var}[\eta]}\) provided that the variances exist. Hence if the variance of \( \eta \) is negligible in comparison to that of \( \xi \), then \( \text{Cov}[\xi, \eta] \) will be also negligible in comparison to \( \text{Var}[\xi] \) and therefore \( \text{Var}[\xi + \eta] \sim \text{Var}[\xi] \), i.e. the leading term of \( \text{Var}[\xi + \eta] \) is \( \text{Var}[\xi] \).

Now we turn to the proof of the proposition itself. We have

\[
\text{Var}[\hat{f}_{nh}(x)] = \text{Var}[\hat{f}_{nh}(x)1_{[J_n \leq \delta]} + \hat{f}_{nh}(x)1_{[J_n > \delta]}].
\]

The variance of \( \hat{f}_{nh}(x)1_{[J_n > \delta]} \) is of lower order than \( (nh)^{-1} \), because of the special form of \( M_n = n^\alpha \), the exponential bound on \( P(J_n > \delta) \), see Lemma 2.2, and the inequality

\[
\text{Var}[\hat{f}_{nh}(x)1_{[J_n > \delta]}] \leq \mathbb{E}[(\hat{f}_{nh}(x))^2 1_{[J_n > \delta]}] \leq M_n^2 P(J_n > \delta).
\]
Consequently it suffices to consider $\text{Var}[\hat{f}_{nh}(x)1_{[J_n \leq \delta]}]$. We have

$$\text{Var}[\hat{f}_{nh}(x)1_{[J_n \leq \delta]}] = \text{Var}[\hat{f}_{nh}(x)1_{[J_n \leq \delta]} - f(x)],$$

and since again the variance of $f(x)1_{[J_n > \delta]}$ is of lower order than $(nh)^{-1}$, we can consider $\text{Var}[(\hat{f}_{nh}(x) - f(x))1_{[J_n \leq \delta]}]$ instead. As we have seen in the proof of Proposition 2.1, on the set $\{J_n \leq \delta\}$ for $n$ large and $\delta$ sufficiently small, $\hat{f}_{nh}(x) = f_{nh}(x)$ and the distinguished logarithm is well-defined. Write (cf. (2.18) and (2.19))

$$\text{Var}[(\hat{f}_{nh}(x) - f(x))1_{[J_n \leq \delta]}] = \text{Var} \left[ \left( \frac{1}{2\pi} \int_{-1/h}^{1/h} e^{-itx} z_{nh}(t) dt + \int_{-1/h}^{1/h} e^{-itx} R_{nh}(t) dt \right) 1_{[J_n \leq \delta]} \right].$$

The variances of the last two terms are negligible, because we have e.g.

$$\text{Var} \left[ \int_{1/h}^{\infty} e^{-itx} \phi(t) dt \right] 1_{[J_n \leq \delta]} = \left| \int_{1/h}^{\infty} e^{-itx} \phi(t) dt \right|^2 \text{Var} \left[ 1_{[J_n > \delta]} \right] \leq CP(J_n > \delta),$$

for some constant $C$.

Hence we have to deal with

$$\text{Var} \left[ \left( \frac{1}{2\pi} \int_{-1/h}^{1/h} e^{-itx} z_{nh}(t) dt + \int_{-1/h}^{1/h} e^{-itx} R_{nh}(t) dt \right) 1_{[J_n \leq \delta]} \right] = \text{Var}[II].$$

(2.26)

We will show that $II$ has a negligible variance compared to that of $I$. Using the bound (2.24) from the proof of Proposition 2.1,

$$nh \text{Var} \left[ \int_{-1/h}^{1/h} e^{-itx} R_{nh}(t) dt 1_{[J_n \leq \delta]} \right] \leq K^2 2\pi nh \text{E}[(\text{ISE}_{n}(h))^2]$$

$$= K^2 2\pi nh \text{Var}[\text{ISE}_{n}(h)] + K^2 2\pi nh (\text{MISE}_{n}(h))^2,$$

where $K$ is a constant. Due to the condition $nh \to \infty, nh^{1+4\beta} \to 0$, we see that $nh(\text{MISE}_{n}(h))^2$ tends to 0.

We deal with $nh \text{Var}[\text{ISE}_{n}(h)]$. Let us write the integrated square error as

$$\text{ISE}_{n}(h) = \frac{1}{n^2 h} \sum_{j=1}^{n} \int_{-\infty}^{\infty} (w(t))^2 dt + \frac{1}{n^2 h} \sum_{j \neq k} w \ast w \left( \frac{Z_j - Z_k}{h} \right)$$

$$- \frac{2}{nh} \sum_{j=1}^{n} \int_{-\infty}^{\infty} w \left( \frac{t - Z_j}{h} \right) g(t) dt + \int_{-\infty}^{\infty} (g(t))^2 dt,$$
since
\[
\frac{1}{h} \int_{-\infty}^{\infty} w \left( \frac{t - Z_j}{h} \right) w \left( \frac{t - Z_k}{h} \right) dt = w * w \left( \frac{Z_j - Z_k}{h} \right)
\]

because \( w \) is symmetric. Here \( w * w \) denotes the convolution of \( w \) with itself. From this it follows that

\[
\begin{align*}
\frac{n}{h} \text{Var}[\text{ISE}_n(h)] &= \frac{1}{n^3 h} \text{Var} \left[ \sum_{j \neq k} w * w \left( \frac{Z_j - Z_k}{h} \right) \right] \\
&= \frac{2}{n^3 h} \sum_{j=1}^{n} \int_{-\infty}^{\infty} w \left( \frac{t - Z_j}{h} \right) g(t) dt \\
&= 4 \frac{1}{n^3 h} \sum_{j=1}^{n} \int_{-\infty}^{\infty} w \left( \frac{t - Z_j}{h} \right) g(t) dt \\
&= 4 \sum_{j=1}^{n} \int_{-\infty}^{\infty} w \left( \frac{t - Z_j}{h} \right) g(t) dt \\
&\leq 4 \sum_{j=1}^{n} \int_{-\infty}^{\infty} w \left( \frac{t - Z_j}{h} \right) g(t) dt
\end{align*}
\]

We study the variance of each term between the brackets in (2.27) separately. For the second term we have

\[
\begin{align*}
\frac{1}{n^3 h} & \text{Var} \left[ \frac{2}{n} \sum_{j=1}^{n} \int_{-\infty}^{\infty} w \left( \frac{t - Z_j}{h} \right) g(t) dt \right] \\
&= \frac{4}{n^3 h} \sum_{j=1}^{n} \text{Var} \left[ \int_{-\infty}^{\infty} w \left( \frac{t - Z_j}{h} \right) g(t) dt \right] \\
&= \frac{4}{n^3 h} \text{Var} \left[ \int_{-\infty}^{\infty} w \left( \frac{t - Z_1}{h} \right) g(t) dt \right] \\
&\leq \frac{4}{n^3 h} \text{E} \left[ \left( \int_{-\infty}^{\infty} \left| w \left( \frac{t - Z_1}{h} \right) g(t) dt \right| \right)^2 \right]
\end{align*}
\]

Through a change of the integration variable it is easily seen that

\[
\begin{align*}
\int_{-\infty}^{\infty} w \left( \frac{t - Z_j}{h} \right) g(t) dt &= h \int_{-\infty}^{\infty} w(u) g(uh + Z_1) du \\
&\leq hA \int_{-\infty}^{\infty} |w(u)| du,
\end{align*}
\]

where we used the fact that \( g \) is bounded. Hence (2.28) vanishes as \( h \to 0 \). Now we arrive at the computation of the variance of the first term between the brackets in (2.27). We have

\[
\begin{align*}
\frac{1}{n^3 h} & \text{Var} \left[ \sum_{j \neq k} w * w \left( \frac{Z_j - Z_k}{h} \right) \right] \\
&= \frac{4}{n^3 h} \text{Var} \left[ \sum_{j \neq k} w * w \left( \frac{Z_j - Z_k}{h} \right) \right] \\
&= \frac{4}{n^3 h} \sum_{i<j} \sum_{k<l} \text{Cov} \left[ w * w \left( \frac{Z_i - Z_j}{h} \right), w * w \left( \frac{Z_k - Z_l}{h} \right) \right]
\end{align*}
\]

We have three possibilities
1. \(i, j, k, l\) are distinct. Then because of the independence, the corresponding covariances are 0.

2. \(i = k, j = l\). The number of such possibilities is of order \(n^2\) and because the covariances are bounded (the convolution \(w \ast w\) is bounded), the sum of such terms will be of order \(O(n^2)\). After the division by \(n^3h\), this will converge to zero.

3. The last possibility is that three indices out of four are distinct, e.g. \(i = k, j \neq l\).

The number of such terms is of order \(n^3\). Thus we have to study the behaviour of e.g.

\[
\frac{1}{h} \text{Cov}\left[w \ast w \left(\frac{Z_i - Z_j}{h}\right), w \ast w \left(\frac{Z_i - Z_k}{h}\right)\right].
\]

Writing out this covariance we get

\[
\frac{1}{h} \text{Cov}\left[w \ast w \left(\frac{Z_i - Z_j}{h}\right), w \ast w \left(\frac{Z_i - Z_k}{h}\right)\right] = \frac{1}{h} E \left[w \ast w \left(\frac{Z_i - Z_j}{h}\right) w \ast w \left(\frac{Z_i - Z_k}{h}\right)\right] - \frac{1}{h} E \left[w \ast w \left(\frac{Z_i - Z_j}{h}\right)\right]^2 \\
\leq \frac{1}{h} E \left[w \ast w \left(\frac{Z_i - Z_j}{h}\right) w \ast w \left(\frac{Z_i - Z_k}{h}\right)\right].
\]

(2.29)

Notice that

\[
\begin{pmatrix}
1 & -1 & 0 \\
1 & 0 & -1 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
Z_1 \\
Z_2 \\
Z_3
\end{pmatrix}
= \begin{pmatrix}
Z_1 - Z_2 \\
Z_1 - Z_3 \\
Z_2 - Z_3
\end{pmatrix}
\]

By denoting \(V_1 = Z_1 - Z_2, V_2 = Z_1 - Z_3\) and \(V_3 = Z_3\), and taking into account (2.29), we get that we have to deal with

\[
\frac{1}{h} \int \int \int w \ast w \left(\frac{V_1}{h}\right) w \ast w \left(\frac{V_2}{h}\right) g(V_1)g(V_2)g(V_3)dv_1dv_2dv_3.
\]

By introducing new integration variables \(u_1, u_2, u_3\), defined by \(v_1 = hu_1, v_2 = hu_2, v_3 = u_3\), we obtain from the above expression that we have to deal with

\[
h \int \int \int w \ast w(u_1)w \ast w(u_2)g(u_3 + hu_2)g(u_3 + hu_2 - u_1))g(u_3)du_1du_2du_3.
\]

(2.30)

By the dominated convergence theorem, the triple integral in the expression above converges to

\[
\int \int \int w \ast w(u_1)w \ast w(u_2)(g(u_3))^3du_1du_2du_3.
\]

which is finite. The dominated convergence theorem is applicable in view of the Condition 2.2 and Lemma 2.1, because \(g\) is bounded and integrable and because \(w \ast w\) is integrable.
The integrability of $w * w$ follows from Schwartz (1966, formula (III,2.8)). Consequently, \((2.30)\) converges to zero as $h \to 0$. Thus $\text{Var}[\mathcal{I}]$ is indeed negligible in comparison to $\text{Var}[\mathcal{I}]$.

Now we need to study (cf. \((2.26)\))

$$\text{Var} \left[ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-i t x} z_{nh}(t) dt 1_{[J_n \leq \delta]} \right].$$

Once again, applying the by now standard argument, we substitute $1_{[J_n \leq \delta]}$ with 1 and instead of $\int_{-1/h}^{1/h}$ we take $\int_{-\infty}^{\infty}$, because the error will be of a lower order than $(nh)^{-1}$, see the proof of Proposition 2.1. Furthermore,

$$\text{Var} \left[ \frac{1}{2\pi \lambda} \int_{-\infty}^{\infty} e^{-i t x} z_{nh}(t) dt \right] = \text{Var} [A_{nh}(x) + B_{nh}(x)],$$

where

$$A_{nh}(x) = \frac{e^{\lambda} - 1}{2\pi \lambda} \int_{-\infty}^{\infty} e^{-i t x} (\phi_{\text{emp}}(t)\phi_{w}(ht) - \phi_{g}(t)) dt = \frac{e^{\lambda} - 1}{\lambda} (\phi_{nh}(x) - \phi_{g}(x)),$$

$$B_{nh}(x) = \frac{e^{\lambda} - 1}{2\pi \lambda} \int_{-\infty}^{\infty} e^{-i t x} (\phi_{\text{emp}}(t)\phi_{w}(ht) - \phi_{g}(t))(e^{-\lambda \phi_{f}(t)} - 1) dt.$$

For the variance of $g_{nh}(x)$ we have an expansion

$$\text{Var}[g_{nh}(x)] = \frac{1}{nh} g(x) \int_{-\infty}^{\infty} (w(t))^2 dt + o \left( \frac{1}{nh} \right),$$

see Tsybakov (2004, Proposition 1.4), and this automatically provides us with an expansion for the variance of $A_{nh}(x)$,

$$\text{Var}[A_{nh}(x)] = \left( \frac{e^{\lambda} - 1}{\lambda} \right)^2 \frac{1}{nh} g(x) \int_{-\infty}^{\infty} (w(t))^2 dt + o \left( \frac{1}{nh} \right).$$

We will show that the variance of the $B_{nh}(x)$ is of lower order than $(nh)^{-1}$. Indeed,

$$nh \text{Var}[B_{nh}(x)] = \left( \frac{e^{\lambda} - 1}{(2\pi \lambda)^2} \right)^2 h \text{Var} \left[ \int_{-\infty}^{\infty} e^{-it(x-Z_1)} \phi_{w}(ht)(e^{-\lambda \phi_{f}(t)} - 1) dt \right].$$

Now note that

$$\left| \int_{-\infty}^{\infty} e^{-it(x-Z_1)} \phi_{w}(ht)(e^{-\lambda \phi_{f}(t)} - 1) dt \right| \leq \int_{-\infty}^{\infty} |e^{-\lambda \phi_{f}(t)} - 1| dt,$$

and that the right hand side is finite, thanks to the fact that the integrand on the right-hand side is bounded by an integrable function $|\phi_{f}(t)|$, which can be verified as when obtaining (2.11). Since for a random variable $\xi$ bounded in absolute value by a constant $K$ we have $\text{Var}[\xi] \leq K^2$, we conclude that $\text{Var}[B_{nh}(x)] = o \left( \frac{1}{nh} \right)$.

Combining all intermediate results, we see that the leading term of the $\text{Var}[\hat{f}_{nh}(x)]$ is

$$\frac{1}{nh} \frac{(e^{\lambda} - 1)^2}{\lambda^2} g(x) \int_{-\infty}^{\infty} (w(u))^2 du$$

and that other terms are of lower order than $(nh)^{-1}$. The theorem is proved.
Proof of Proposition 2.3. The result follows immediately from the decomposition
\[
\text{MSE}[\hat{f}_{nh}(x)] = \text{Var}[\hat{f}_{nh}(x)] + (b^w(n, h, x))^2.
\]
and Propositions 2.1 and 2.2. Here \(b^w(n, h, x)\) denotes the bias of \(\hat{f}_{nh}(x)\).

Proof of Remark 2.3. The proof is based on Remark 2.2 and Proposition 2.2. Write
\[
\text{MSE}[\hat{f}_{nh}(x)] = \text{Var}[\hat{f}_{nh}(x)] + (b^w(n, h, x))^2.
\]
Substitute in this relation the expansions of the variance and bias and neglect the terms of order \(o(h^4)\) and \(o\left(\frac{1}{nh}\right)\). We get
\[
\frac{1}{nh} \frac{(e^\lambda - 1)^2}{\lambda^2} g(x) \int_{-\infty}^{\infty} (w(u))^2 du + \frac{\sigma^4}{16\pi^2} \frac{(e^\lambda - 1)^2}{\lambda^2} h^4 \left( \int_{-\infty}^{\infty} e^{-itx} \frac{t^2 \phi_g(t)}{(e^\lambda - 1) \phi_g(t) + 1} dt \right)^2 \tag{2.31}
\]
Now minimise this with respect to \(h\), i.e. take the derivative with respect to \(h\) and set it to zero,
\[
-\frac{1}{nh^2} g(x) \int_{-\infty}^{\infty} (w(u))^2 du + \frac{\sigma^4}{4\pi^2} h^3 \left( \int_{-\infty}^{\infty} e^{-itx} \frac{t^2 \phi_g(t)}{(e^\lambda - 1) \phi_g(t) + 1} dt \right)^2 = 0.
\]
Finally determine \(h_{opt}\) from this equation. This leads to the desired result. The order of the mean square error of \(\hat{f}_{nh}(x)\) corresponding to the optimal bandwidth can be computed from (2.31) and it is easily seen that it is equal to \(n^{-4/5}\).

Proof of Theorem 2.1. We follow the same line of thought as in Figueroa-Lopez and Houdrée (2004). By relocating \(x_0\) to the origin, without loss of generality we may assume that \(x_0 = 0\). Let \(f_0\) be some density from the class \(\mathcal{H}(\beta, L_1/2) \cap \mathcal{N}(\beta, L_2/2)\), such that \(f_0(x) > 0\) for all \(x \in \mathbb{R}\) and let \(\kappa \geq 1\) be some constant. Furthermore, let \(v\) be a bounded function with compact support \(\mathbb{R} \subset [-1, 1]\) that also belongs to \(\mathcal{H}(\beta, L_1/2) \cap \mathcal{N}(\beta, L_2/2)\) and such that \(v(0) > 0\) and \(\int_{-\infty}^{\infty} v(x) dx = 0\). Notice that \(v\) is not a density and we abuse the notation. Moreover, we can select \(f_0, v\) and \(\kappa\) in such a way, that
\[
f_0(x) \pm \kappa^{-\beta} v(\kappa x) > 0, \text{ for any } x \in \mathbb{R}.
\]
Let us consider a parametric model
\[
f_\theta(x) = f_0(x) + \theta n^{-\beta/(2\beta + 1)} v(\kappa n^{1/(2\beta + 1)} x),
\]
parametrised by \(\theta \in (-\kappa^{-\beta}, \kappa^{-\beta})\). Notice that if we select \(\kappa\) large enough, then for \(n \geq 1\) the function \(f_\theta\) will be a density. Laborious, but not difficult computation demonstrates that \(f_\theta \in \mathcal{H}(\beta, L_1) \cap \mathcal{N}(\beta, L_2)\), cf. Figueroa-Lopez and Houdrée (2004).

Let \(P^{(n)}\) be the distribution of a spatial Poisson process on \([0, n] \times \mathbb{R}\) with mean measure \(\lambda f_\theta(x) dx dt\). By Kutoyants (1998, Theorem 1.3) and Jacod and Shiryaev (2003,
Theorem 3.4) we get
\[
\frac{dP_\theta(T_n)}{dP_\theta(T_0)} = \exp \left[ \int_0^{T_n} \int_{-\infty}^{\infty} \left\{ 1 + \theta n^{-\frac{2}{2+\beta}} f_0(x)^{-1} v \left( \kappa n^{1/(2+\beta)} x \right) \right\} \xi(dx, dt) \
- \lambda \theta n^{-\frac{2}{2+\beta}} \int_0^{T_n} \int_{-\infty}^{\infty} v \left( \kappa n^{1/(2+\beta)} x \right) dx dt \right],
\]
where \( \xi \) is the random measure induced by the given spatial Poisson process. Our goal is to prove LAN (local asymptotic normality) for the parametric model \( P_\theta(T_n) : \theta \in (-\kappa^{-\beta}, \kappa^{-\beta}) \) at \( \theta = 0 \), see Kutoyants (1998, Definition 2.1). Define \( R(u) = \log(1 + u) - u + \frac{u^2}{2} \). The right-hand side of the above equation can be written as follows:
\[
\frac{dP_\theta(T_n)}{dP_\theta(T_0)} = \exp \left\{ \theta \Delta_{T_n,n} - \frac{\theta^2}{2} \sigma_{T_n,n}^2 + r_{T_n}(\theta) \right\},
\]
where
\[
\Delta_{T_n,n} = n^{-\frac{2}{2+\beta}} \int_0^{T_n} \int_{-\infty}^{\infty} v \left( \kappa n^{1/(2+\beta)} x \right) \left[ \xi(dx, dt) - \lambda f_0(x)dx dt \right],
\]
\[
\sigma_{T_n,n}^2 = \lambda n^{-\frac{2}{2+\beta}} \int_0^{[cn]} \int_{-\infty}^{\infty} \left( \frac{v \left( \kappa n^{1/(2+\beta)} x \right)}{f_0(x)} \right)^2 \left[ \xi(dx, dt) - \lambda f_0(x)dx dt \right] dx dt,
\]
\[
r_{T_n}(\theta) = -\frac{\theta^2}{2} n^{-\frac{2}{2+\beta}} \int_0^{T_n} \int_{-\infty}^{\infty} \left( \frac{v \left( \kappa n^{1/(2+\beta)} x \right)}{f_0(x)} \right)^2 \left[ \xi(dx, dt) - \lambda f_0(x)dx dt \right] dx dt
- \lambda \theta n^{-\frac{2}{2+\beta}} \int_0^{T_n} \int_{-\infty}^{\infty} \left( \frac{v \left( \kappa n^{1/(2+\beta)} x \right)}{f_0(x)} \right)^2 \left[ \xi(dx, dt) - \lambda f_0(x)dx dt \right] dx dt
+ \theta^2 \int_0^{[cn]} \int_{-\infty}^{\infty} \left( \frac{v \left( \kappa n^{1/(2+\beta)} x \right)}{f_0(x)} \right)^2 \left[ \xi(dx, dt) - \lambda f_0(x)dx dt \right] dx dt
+ \int_0^{T_n} \int_{-\infty}^{\infty} R \left( \theta n^{-\frac{2}{2+\beta}} f_0(x)^{-1} v \left( \kappa n^{1/(2+\beta)} x \right) \right) \xi(dx, dt).
\]
Here \( c \) is the limit in probability of \( T_n/n \) as \( n \) tends to infinity, \( c = 1/(1 - e^{-\lambda}) \), and \([cn]\) denotes the integer part of \( cn \). In order to establish LAN, we have to prove that
\[
\mathcal{L} \left( \Delta_{T_n,n}|P_0(T_0) \right) \rightarrow N(0, I_0), \quad I_0^{-1} \sigma_{T_n,n}^2 \rightarrow 1, \quad \text{and} \quad r_{T_n}(\theta) \xrightarrow{P(T_n)} 0
\]
as \( n \to \infty \). Here \( I_0 = c\lambda^{-1} f_0(0)^{-1} \int_x (v(u))^2 du \).

Rewrite \( \Delta_{T_n,n} \) as \( \Delta_{[cn],n} + (\Delta_{T_n,n} - \Delta_{[cn],n}) \),
where
\[
\Delta_{[cn],n} = n^{-\beta/(2+\beta+1)} \int_0^{[cn]} \int_{-\infty}^{\infty} \left( f_0(x)^{-1} v \left( \kappa n^{1/(2+\beta+1)} x \right) \right) \left[ \xi(dx, dt) - \lambda f_0(x)dx dt \right] .
\]
We have for $\delta > 0$
\[
\begin{align*}
  n^{-\frac{2(2+\delta)}{2\beta+1}} &\int_0^{[cn]} \int_{-\infty}^{\infty} (f_0(x))^{-2-\delta} \left| v \left( \frac{ku^{1/(2\beta+1)} x}{k^{1/(2\beta+1)}} \right) \right|^{2+\delta} \lambda f_0(x) dx dt \\
  &= \lambda n^{-1} n^{-2-\delta} [cn] \int K \left( f_0 \left( \frac{u}{k^{1/(2\beta+1)}} \right) \right)^{-1-\delta} (v(u))^{2+\delta} du \rightarrow 0,
\end{align*}
\]
and hence $\Delta_{[cn],n}$ satisfies Lyapunov’s condition.

Furthermore, with a minor variation of the arguments in the proof of Theorem 7.3.2 in Chung (2001), one can prove that $\Delta_{T_n,n} - \Delta_{[cn],n}$ converges in probability to zero. Indeed, let $\varepsilon$ be given, $0 < \varepsilon < 1$, and put
\[
a_n = [(1 - \varepsilon^3) [cn]], \quad b_n = [(1 + \varepsilon^3) [cn]] - 1.
\]
Since $T_n/n$ converges in probability to $c$, there exists $n_0(\varepsilon)$, such that if $n \geq n_0(\varepsilon)$, then the set
\[
\Lambda = \{ \omega : a_n \leq T_n \leq b_n \}
\]
has probability at least $1 - \varepsilon$. If $\omega$ is in this set, then $\Delta_{T_n,n}$ is equal to $\Delta_{j,n}$ for some $a_n \leq j \leq b_n$. For $[cn] < j \leq b_n$ we have
\[
\Delta_{j,n} - \Delta_{[cn],n} = (\Delta_{j,n} - \Delta_{j-1,n}) + (\Delta_{j-1,n} - \Delta_{j-2,n}) + \ldots + (\Delta_{[cn]+1,n} - \Delta_{[cn],n}),
\]
and hence $\Delta_{[cn],n}$ satisfies Lyapunov’s condition.

Using Kolmogorov’s inequality, we get
\[
P \left( \max_{a_n \leq j \leq b_n} |\Delta_{j,n} - \Delta_{[cn],n}| > \varepsilon \right) \leq \frac{\text{Var} \left( \Delta_{[cn],n} \right)}{\varepsilon^2} \leq \frac{\varepsilon^3 [cn] C_1}{\varepsilon^2 n} \leq C_1 \varepsilon.
\]
A similar inequality holds for $a_n \leq j < [cn]$. Combining the two, we obtain
\[
P \left( \max_{a_n \leq j \leq b_n} |\Delta_{j,n} - \Delta_{[cn],n}| > \varepsilon \right) \leq C_2 \varepsilon.
\]
Now we have, if \( n \geq n_0(\varepsilon) \),
\[
\mathbb{P} \left( |\Delta_{T_n, n} - \Delta_{\lfloor cn \rfloor, n}| > \varepsilon \right)
\leq \sum_{j=1}^{\infty} \mathbb{P} \left( T_n = j; \max_{\lfloor an \rfloor \leq j \leq \lfloor bn \rfloor} |\Delta_{j, n} - \Delta_{\lfloor cn \rfloor, n}| > \varepsilon \right)
\leq \mathbb{P} \left( T_n = j; \max_{\lfloor an \rfloor \leq j \leq \lfloor bn \rfloor} |\Delta_{j, n} - \Delta_{\lfloor cn \rfloor, n}| > \varepsilon \right) + \sum_{j \notin [a_n, b_n]} \mathbb{P}(T_n = j)
\leq \mathbb{P}( \max_{\lfloor an \rfloor \leq j \leq \lfloor bn \rfloor} |\Delta_{j, n} - \Delta_{\lfloor cn \rfloor, n}| > \varepsilon ) + \mathbb{P}(T_n \notin [a_n, b_n])
\leq C_2 \varepsilon + 1 - \mathbb{P}(\Lambda) \leq C_3 \varepsilon.
\]
Since \( \varepsilon \) is arbitrary, this proves that \( \Delta_{T_n, n} - \Delta_{\lfloor cn \rfloor, n} \) converges to zero in probability.
Moreover, we have
\[
\text{Var} \left[ \Delta_{\lfloor cn \rfloor, n} \right] \to c \lambda \kappa^{-1}(f_0(0))^{-1} \int_{\mathbb{R}} (v(u))^2 \, du,
\]
which can be seen as follows (cf. Figueroa-Lopez and Houdré (2004))
\[
\text{Var} \left( \Delta_{\lfloor cn \rfloor, n} \right) = n^{-\frac{2\beta}{\beta + 1}} \int_{0}^{[cn]} \int_{-\infty}^{\infty} \frac{(v(kn^{1/(2\beta+1)}(x)))^2}{(f_0(x))^2} (\lambda f_0(x)) \, dx \, dt
\leq \lambda \kappa^{-1} n^{-1}[cn] \int_{\mathbb{R}} (\kappa^{-1} n^{-1/(2\beta+1)} u) (v(u))^2 \, du
\leq n^{-\infty} c \lambda \kappa^{-1} f_0^{-1}(0) \int_{\mathbb{R}} (v(u))^2 \, du.
\]
Combining these results, we conclude that \( \Delta_{T_n, n} \) is asymptotically normal with mean zero and variance \( f_0 \), and it also follows that \( \sigma_{T_n}^2 \) converges to \( f_0 \). All that remains to be proved, is the fact that \( r_{T_n}(\theta) \) converges in probability to zero.
Elementary manipulations with integrals and the fact that \( T_n/n \) converges in probability to \( c \), imply that the difference of the second and third summands in \( r_{T_n}(\theta) \) converges to zero, since this difference is equal to
\[
\frac{\theta^2}{2n} \int_{\mathbb{R}} \frac{(v(x))^2}{f_0(x/\kappa n^{1/(2\beta+1)})} dx.
\]
Now we turn to the first summand of \( r_{T_n}(\theta) \), which we denote by \( a_{T_n, n}(\theta) \). Write
\[
a_{T_n, n}(\theta) = \hat{a}_{\lfloor cn \rfloor, n}(\theta) + \hat{a}_{T_n, n}(\theta) - \hat{a}_{\lfloor cn \rfloor, n}(\theta),
\]
where
\[
\hat{a}_{j, n}(\theta) = -\frac{\theta^2}{2n} x^{-\frac{2\beta}{\beta + 1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{(v(kn^{1/(2\beta+1)}(x)))^2}{(f_0(x))^2} [\xi(dx, dt) - \lambda f_0(x) dx dt].
\]
Note that $\hat{a}_{[cn],n}(\theta)$ has expectation zero, while for its variance, given that $n$ is large enough, we have

$$\frac{\theta^4}{4} n^{-4\beta/(2\beta+1)} \int_0^{[cn]} \int_{-\infty}^{\infty} (f_0(x))^{-4} \left( v \left( \kappa n^{1/(2\beta+1)} x \right) \right)^4 \lambda f_0(x) dx dt$$

$$= \frac{\theta^4}{4} \lambda \kappa^{-1} [cn]^{-4\beta/(2\beta+1)} \int_{\mathbb{R}} \left( f_0(\kappa^{-1} n^{-1/(2\beta+1)} u) \right)^{-3} (v(u))^4 du \xrightarrow{n \to \infty} 0,$$

cf. the proof of Theorem 4.1 in Figueroa-Lopez and Houdré (2004). Consequently $\hat{a}_{[cn],n}(\theta)$ converges in probability to zero. Furthermore, notice that we can use the arguments that we used to prove $\Delta_{T_\alpha,n} - \Delta_{[cn],n} \xrightarrow{p} 0$ to prove $a_{T_\alpha,n}(\theta) - \hat{a}_{[cn],n}(\theta) \xrightarrow{p} 0$ as well.

Now we have to treat the last summand of $r_{T_n}(\theta)$, which we denote by $b_{T_n,n}(\theta)$. Write

$$b_{T_n,n}(\theta) = \hat{b}_{[cn],n}(\theta) + (b_{T_n,n}(\theta) - \hat{b}_{[cn],n}(\theta)),$$

where

$$\hat{b}_{n}(\theta) = \int_0^{[cn]} \int_{-\infty}^{\infty} R \left( \theta n^{-\beta/(2\beta+1)} (f_0(x))^{-1} v \left( \kappa n^{1/(2\beta+1)} x \right) \right) \xi(dx, dt).$$

By a computation similar to the one in the proof of Theorem 4.1 in Figueroa-Lopez and Houdré (2004) we will demonstrate that $\hat{b}_{[cn],n}(\theta)$ converges to zero in probability, because both its expectation and variance go to zero. Notice that because $u$ in our case may take negative values, instead of the inequality $|R(u)| \leq |u|^3/3$ used in Figueroa-Lopez and Houdré (2004), we will have to use the inequality

$$|R(u)| \leq \frac{|u|^3}{3(1-\varepsilon)}, \quad (2.32)$$

which is true for $|u| \leq \varepsilon < 1$. By selecting $n$ large enough, for any given $\varepsilon$, the expression

$$\theta n^{-\beta/(2\beta+1)} (f_0(x))^{-1} v \left( \kappa n^{1/(2\beta+1)} x \right)$$

can be made arbitrarily small and, hence, the aforementioned inequality is applicable. For the absolute value of the expectation we have

$$\left| \int_0^{[cn]} \int_{-\infty}^{\infty} R \left( \theta n^{-\beta/(2\beta+1)} (f_0(x))^{-1} v \left( \kappa n^{1/(2\beta+1)} x \right) \right) \lambda f_0(x) dx dt \right|$$

$$\leq \frac{\left\{ \theta^3 \lambda \right\} [cn]^{-3\beta/(2\beta+1)} \int_{-\infty}^{\infty} (f_0(x))^{-2} (v \left( \kappa n^{1/(2\beta+1)} x \right))^3 dx \xrightarrow{n \to \infty} 0,$$

which can be seen by changing the integration variable $u = \kappa n^{1/(2\beta+1)} x$ and using the fact that $v$ has a compact support $\mathbb{K} \subseteq [-1,1]$. A similar reasoning applies to the variance.
Now we have to study \( b_{T_n,n}(\theta) - \hat{b}_{[cn],n}(\theta) \). Write

\[
b_{T_n,n}(\theta) - \hat{b}_{[cn],n}(\theta) \\
= \int_{0}^{T_n} \int_{-\infty}^{\infty} R \left( \theta n^{-\beta/(\beta+1)} (f_0(x))^{-1} v \left( \kappa n^{\beta/(\beta+1)} x \right) \right) [\xi(dx, dt) - \lambda f_0(x) dt] dx dt \\
\]

\[
= \int_{0}^{[cn]} \int_{-\infty}^{\infty} R \left( \theta n^{-\beta/(\beta+1)} (f_0(x))^{-1} v \left( \kappa n^{\beta/(\beta+1)} x \right) \right) [\xi(dx, dt) - \lambda f_0(x) dt] dx dt \\
+ (T_n - [cn]) \int_{-\infty}^{\infty} R \left( \theta n^{-\beta/(\beta+1)} (f_0(x))^{-1} v \left( \kappa n^{\beta/(\beta+1)} x \right) \right) \lambda f_0(x) dx
\]

The difference of the first two terms converges to zero in probability, which can be seen by the arguments similar to those that we used to prove \( \Delta_{T_n,n} - \Delta_{[cn],n} \xrightarrow{p} 0 \). As far as the last term is concerned, an application of the inequality (2.32) and the fact that \( T_n/n \) converges to \( c \) in probability imply that it converges to zero in probability.

Summarising all these results, we conclude that \( \tau_{T_n}(\theta) \) converges to zero in probability and hence \( \{\mathbb{P}^{(T_n)}_{\theta}\}_{\theta \in (-\kappa,\kappa)} \) is LAN at \( \theta = 0 \). Consequently, we are in position of using the theory for LAN families (see Kutoyants (1998) for the case of Poisson point processes). Let \( \ell_0 \) denote a loss function, which is defined as in Kutoyants (1998). By the Hájek-Le Cam local asymptotic minimax theorem, in particular by (2.11) of Kutoyants (1998) (see also Ibragimov and Has’minskii (1981, Chapter 4, Lemma 5.1)), if \( \hat{\theta}_{T_n} \) is an arbitrary estimator of \( \theta \), based on the observations of a compound Poisson process on the time interval \([0,T_n]\), then

\[
\liminf_{n \to \infty} \inf_{\hat{\theta}_{T_n}} \sup_{|\theta| < \kappa} \mathbb{E}_{\theta} \left[ \ell_0 \left( \ell_0^{1/2} \left( \hat{\theta}_{T_n} - \theta \right) \right) \right] \geq B, \tag{2.33}
\]

where \( B = 2^{-3/2} \pi^{-1/2} \mathbb{E} \left[ \ell_0(Z)1_{\{|Z| < \kappa n^{-\beta} \}} \right] \), \( Z \sim N(0,1) \) and \( \mathbb{E}_{\theta} \) denotes the expectation taken under the parameter value \( \theta \).

Now, for each \( T_n > 0 \), let \( \hat{f}_{T_n} \) be an arbitrary estimator of \( f_0 \) based on continuous observations of a spatial Poisson process over the time interval \([0,T_n]\). Clearly, \( \hat{f}_{T_n} \) induces the estimator \( \hat{\theta}_{T_n} = n^{\beta/(2\beta+1)} (v(0))^{-1} \left( \hat{f}_{T_n}(0) - f_0(0) \right) \) of \( \theta \), and since \( v(0) (\hat{\theta}_{T_n} - \theta) = n^{\beta/(2\beta+1)} \left( \hat{f}_{T_n}(0) - f_0(0) \right) \), we can write

\[
v(0) \left( \hat{\theta}_{T_n} - \theta \right) = n^{\beta/(2\beta+1)} \left( \hat{f}_{T_n}(0) - f_0(0) \right).
\]

If we take \( \ell_0(u) := (v(0)I_0^{-1/2} u)^2 \), then (2.33) becomes

\[
B \leq \liminf_{n \to \infty} \inf_{\hat{f}_{T_n}} \sup_{|\theta| < \kappa} \mathbb{E}_{\theta} \left[ \ell_0 \left( \ell_0^{1/2} \left( \hat{\theta}_{T_n} - \theta \right) \right) \right] \\
\leq \liminf_{n \to \infty} \inf_{f_{T_n}} \sup_{|\theta| < \kappa} \mathbb{E}_{\theta} \left[ n^{2\beta/(2\beta+1)} \left( \hat{f}_{T_n}(0) - f_0(0) \right)^2 \right].
\]
Since \( \{ \theta : \theta \in (-\kappa^{-\beta}, \kappa^{-\beta}) \} \subset H(\beta, L_1) \cap \mathcal{N}(\beta, L_2) \), we have

\[
\lim \inf_{n \to \infty} \inf_{f_n \in H(\beta, L_1) \cap \mathcal{N}(\beta, L_2)} \sup_{f \in H(\beta, L_1) \cap \mathcal{N}(\beta, L_2)} E_f \left[ n^{2\beta/(2\beta+1)} \left( \hat{f}_{n \alpha}(0) - f(0) \right)^2 \right] \geq B,
\]

where

\[
B = 2^{-3/2} \pi^{-1/2} \int_{|z| < L_n^{1/2}} (v(0) I_0^{-1/2} z)^2 e^{-z^2/2} dz.
\]

This implies (2.8).

**Proof of Theorem 2.2.** The proof is based on repeated applications of Slutsky’s theorem, see van der Vaart (1998, Lemma 2.8), i.e. we will split off an expression that is asymptotically normal from our normalised and centred sum and we will show that the remainder term converges to zero in probability. Then Slutsky’s theorem will imply that the estimator itself is asymptotically normal. Write

\[
\frac{\hat{f}_{n \alpha}(x) - f(x)}{\sqrt{\text{Var}[\hat{f}_{n \alpha}(x)]}} = \frac{\hat{f}_{n \alpha}(x) - f(x)}{\sqrt{\text{Var}[\hat{f}_{n \alpha}(x)]}} 1_{[J_n \leq \delta]} + \frac{\hat{f}_{n \alpha}(x) - f(x)}{\sqrt{\text{Var}[\hat{f}_{n \alpha}(x)]}} 1_{[J_n > \delta]}.
\]

(2.34)

If we take \( n \) large and \( \delta \) small, then

\[
\frac{\hat{f}_{n \alpha}(x) - f(x)}{\sqrt{\text{Var}[\hat{f}_{n \alpha}(x)]}} 1_{[J_n \leq \delta]} = \frac{\hat{f}_{n \alpha}(x) - f(x)}{\sqrt{\text{Var}[\hat{f}_{n \alpha}(x)]}} 1_{[J_n \leq \delta]},
\]

see the proof of Proposition 2.1. We treat the first term of the right-hand side of (2.34). We have

\[
\frac{\hat{f}_{n \alpha}(x) - f(x)}{\sqrt{\text{Var}[\hat{f}_{n \alpha}(x)]}} 1_{[J_n \leq \delta]} = \frac{1}{\sqrt{\text{Var}[\hat{f}_{n \alpha}(x)]}} 1_{[J_n \leq \delta]} \frac{1}{2\pi} \int_{-1/h}^{1/h} e^{-itx} \log(1 + z_{n \alpha}(t)) dt
\]

\[- \frac{1}{\sqrt{\text{Var}[\hat{f}_{n \alpha}(x)]}} 1_{[J_n \leq \delta]} \frac{1}{2\pi} \int_{1/h}^{\infty} e^{-itx} \phi_f(t) dt
\]

\[- \frac{1}{\sqrt{\text{Var}[\hat{f}_{n \alpha}(x)]}} 1_{[J_n \leq \delta]} \frac{1}{2\pi} \int_{-1/h}^{1/h} e^{-itx} \phi_f(t) dt.
\]

(2.35)

Let us denote the second and third expressions at the right-hand side of (2.35) by \( I \) and \( II \). We can write (2.35) as

\[
\frac{1}{\sqrt{\text{Var}[\hat{f}_{n \alpha}(x)]}} 1_{[J_n \leq \delta]} \frac{1}{2\pi} \int_{-1/h}^{1/h} e^{-itx} \log(1 + z_{n \alpha}(t)) dt
\]

\[- (I - E[I]) - (II - E[II]) - E[I] - E[II].
\]
The second and third terms of this expression converge to zero in probability. This follows from an application of Chebyshev’s inequality and the facts that

\[
\text{Var} [1_{J_n \leq \delta}] = \text{Var} [1_{J_n > \delta}] \leq P (J_n > \delta) \sim e^{-Cn},
\]

\[
\text{Var}[\hat{f}_{nh}(x)] \sim \frac{1}{nh}.
\]

Application of Slutsky’s theorem shows that we may neglect these terms. Now we make a further step and rewrite the remaining terms of the right-hand side of (2.35) as

\[
\frac{1}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} 1_{[J_n \leq \delta]} \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} z_{nh}(t) dt
\]

\[
+ \frac{1}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} 1_{[J_n \leq \delta]} \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} R_{nh}(t) dt - E[I + II].
\]

Denote the second term in this expression by $III$. Rewrite the above expression as

\[
\frac{1}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} 1_{[J_n \leq \delta]} \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} z_{nh}(t) dt
\]

\[
+ (III - E[III]) - E[I + II - III].
\]

Again, $(III - E[III])$ converges to zero in probability and therefore we may neglect it. After having done so, we rewrite the above expression as

\[
\frac{1}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} 1_{[J_n \leq \delta]} \frac{1}{2\pi \lambda} \int_{-\infty}^{\infty} e^{-itx} z_{nh}(t) dt
\]

\[
- \frac{1}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} 1_{[J_n \leq \delta]} \frac{1}{2\pi \lambda} \int_{1/h}^{\infty} e^{-itx} z_{nh}(t) dt
\]

\[
- \frac{1}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} 1_{[J_n \leq \delta]} \frac{1}{2\pi \lambda} \int_{-\infty}^{-1/h} e^{-itx} z_{nh}(t) dt - E[I + II - III].
\]

Denote the second and third terms in this expression by $IV$ and $V$. Then we may write

\[
\frac{1}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} 1_{[J_n \leq \delta]} \frac{1}{2\pi \lambda} \int_{-\infty}^{\infty} e^{-itx} z_{nh}(t) dt
\]

\[
- (IV - E[IV]) - (V - E[V]) - E[I + II - III + IV + V].
\]

There is nothing random in $IV$ and $V$ except $1_{[J_n \leq \delta]}$, because $\phi_{emp}(t)\phi_{w}(ht)$ vanishes outside $[-1/h, 1/h]$. Due to Chebyshev’s inequality, $(IV - E[IV])$ and $(V - E[V])$ converge to zero in probability and therefore they may be neglected. We then have to deal with
Due to Chebyshev’s inequality, since $1_{J_n > \delta}$ may be neglected. The asymptotic normality stems from the first term in (2.36), (recall the definition of $z_{nh}$)

\[
\frac{1}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} 1_{[J_n \leq \delta]} \frac{(e^\lambda - 1)}{\lambda} (g_{nh}(x) - g(x)) + (VI - \text{E}[VI]) - \text{E}[I + II - III + IV + V - VI],
\]

where

\[
VI \equiv \frac{1}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} 1_{[J_n \leq \delta]} \frac{1}{2\pi \lambda} \int_{-\infty}^{\infty} e^{-itx}(e^\lambda - 1)\phi_{g_{nh}}(t) - \phi_g(t)(e^{-\lambda\phi_1(t)} - 1)dt.
\]

The argument from the proof of Proposition 2.2 concerning $\text{Var}[B_{nh}]$, see p. 32, demonstrates that the variance of $VI$ converges to zero and hence by Chebyshev’s inequality $VI - \text{E}[VI]$ converges to zero in probability. Therefore we may neglect it. Thus we have

\[
\frac{1}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} 1_{[J_n \leq \delta]} \frac{(e^\lambda - 1)}{\lambda} (g_{nh}(x) - g(x)) - \text{E}[I + II - III + IV + V - VI].
\]

Now rewrite this as

\[
\frac{1}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} 1_{[J_n \leq \delta]} \frac{(e^\lambda - 1)}{\lambda} (g_{nh}(x) - \text{E}[g_{nh}(x)]) + (VII - \text{E}[VII]) - \text{E}[I + II - III + IV + V - VI - VII],
\]

where

\[
VII \equiv \frac{1}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} 1_{[J_n \leq \delta]} \frac{(e^\lambda - 1)}{\lambda} (\text{E}[g_{nh}(x)] - g(x)).
\]

Due to Chebyshev’s inequality, $VII - \text{E}[VII]$ converges to zero in probability and therefore may be neglected. The asymptotic normality stems from the first term in (2.36), since $1_{[J_n \leq \delta]} \rightarrow 1$ in probability and because

\[
\left( \frac{g_{nh}(x) - \text{E}[g_{nh}(x)]}{\sqrt{\text{Var}[g_{nh}(x)]}} \right) \xrightarrow{D} N(0,1),
\]

which can be verified along the lines of Prakasa Rao (1983, pp. 61–62) by checking Lyapunov’s condition. It is easy to verify that

\[
\text{E}[I + II - III + IV + V - VI - VII] = \text{E} \left[ \frac{\hat{f}_{nh}(x) - f(x)}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} 1_{[J_n \leq \delta]} \right].
\]

Adding to this expression the second term in (2.34) results in

\[
\frac{b_\nu(n, h, x)}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} + \frac{\hat{f}_{nh}(x)1_{[J_n > \delta]} - \text{E}[\hat{f}_{nh}(x)1_{[J_n > \delta]}]}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} - \frac{f(x)1_{[J_n > \delta]} - \text{E}[f(x)1_{[J_n > \delta]}]}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}},
\]
where \( b^w(n, h, x) \) denotes the bias of the estimator \( \hat{f}_{nh}(x) \). The first term goes to zero because we assumed that \( nh^{2\beta+1} \to 0 \). The two other terms converge to zero in probability thanks to Lemma 2.2. Thus, thanks to Slutsky’s theorem, these terms may be neglected and we establish the desired result.

Proof of Theorem 2.3. Write

\[
\hat{f}_{nh}(x) - E[\hat{f}_{nh}(x)] = (\hat{f}_{nh}(x) - f(x))1_{[J_n \leq \delta]} + (\hat{f}_{nh}(x) - f(x))1_{[J_n > \delta]} + (f(x) - E[\hat{f}_{nh}(x)]).
\]

Using the same type of arguments as in Theorem 2.2 (note that we will not need \( nh^{2\beta+1} \to 0 \) since the bias divided by the root of variance will be cancelled in intermediary computations) we see that we have to deal with

\[
\frac{e^\lambda - 1}{\lambda} \frac{(g_{nh}(x) - E[g_{nh}(x)])}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} - \frac{(\hat{f}_{nh}(x) - f(x))}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}} 1_{[J_n > \delta]} - \frac{E[(\hat{f}_{nh}(x) - f(x))1_{[J_n > \delta]}]}{\sqrt{\text{Var}[\hat{f}_{nh}(x)]}}.
\]

The first term gives asymptotic normality, while the last two tend to zero in probability thanks to the exponential bound on \( P(J_n > \delta) \). The application of Slutsky’s theorem entails the desired result.