Nonparametric inference for partially observed Levy processes
Gugushvili, S.

Citation for published version (APA):
4 Decompounding under Gaussian noise

In this chapter, assuming that the Lévy process $X = (X_t)_{t \geq 0}$ is a sum of a compound Poisson process with known intensity $\lambda$ and unknown jump size density $f$, and an independent Brownian motion $Z = (Z_t)_{t \geq 0}$, we consider the problem of nonparametric estimation of $f$. The estimator is obtained via Fourier inversion and kernel smoothing. The main result of the chapter deals with asymptotic normality of the estimator at a fixed point.

4.1 Introduction

Let $Y = (Y_t)_{t \geq 0}$ be a compound Poisson process with intensity $\lambda$ and jump size distribution $F$, which has a density $f$. Assume that $Z = (Z_t)_{t \geq 0}$ is a Brownian motion independent of $Y$ and consider the stochastic process $X_t = Y_t + Z_t$. Notice that $X = (X_t)_{t \geq 0}$ is a Lévy process, see Sato (2004, Example 8.5). Suppose that $X$ is observed at equidistant time points $\Delta, 2\Delta, \ldots, n\Delta$. By a rescaling argument, without loss of generality we may take $\Delta = 1$. Given a sample $X_1, X_2, \ldots, X_n$, the statistical problem we consider is nonparametric estimation of the density $f$.

Notice that the Lévy triplet of the process $X$ is given by $(0, 1, \nu)$, where the measure $\nu(dx)$ equals $\lambda f(x)dx$, see Sato (2004, Example 8.5). Since the Lévy triplet provides a unique means for characterisation of any Lévy process, see e.g. Sato (2004, Chapter 2), inference on the law of $X$ can be reduced to inference on $\nu$. Most of the existing literature dealing with estimation problems for Lévy processes is concerned with parametric estimation of the Lévy measure, see e.g. Akritas and Johnson (1981) and Akritas (1982), where a fairly general setting is considered. There are relatively few papers that study nonparametric inference procedures for Lévy processes, and the majority of them assume that high frequency data are available, i.e. either a Lévy process is observed continuously over a time interval $[0, T]$ with $T \to \infty$, or it is observed at equidistant time points $\Delta_n, \ldots, n\Delta_n$ and $\lim_{n \to \infty} \Delta_n = 0$, $\lim_{n \to \infty} n\Delta_n = \infty$, see e.g. Rubin and Tucker (1959), Basawa and Brockwell (1982) and Figueroa-Lopez and Houdré (2004). On the other hand, it is also interesting to study estimation problems for the case of low frequency observations. In the particular context of a compound Poisson process we mention Buchmann and Grübel (2003, 2004) and van Es et al. (2007a), see also Chapter 2 of this thesis, where given a sample $Y_1, \ldots, Y_n$ from a compound Poisson process $Y = (Y_t)_{t \geq 0}$, nonparametric estimators for the jump size distribution function $F$ (see Buchmann and Grübel (2003, 2004)) and its density $f$ (see van Es et al. (2007a)) are proposed and their asymptotics are studied as $n \to \infty$. This problem is referred to as decompounding. The process $X_t = Y_t + Z_t$ constitutes a generalisation of the compound Poisson model considered in Buchmann and Grübel (2003, 2004) and Chapter 2 and is...
related to Merton’s jump-diffusion model of an asset price, see Merton (1976). Since $Z$ is a Brownian motion, it is natural to call the estimation problem of $f$ decompounding under Gaussian noise. Figures 4.1–4.4 provide an indication of the difficulty of the problem. Figure 4.1 gives a typical path of the Brownian motion, while Figure 4.2 gives a path of the process $X$. The difference is at once clear when $X$ is observed continuously.

\begin{figure}[h]
\centering
\includegraphics[width=0.4\textwidth]{fig4_1.png}
\caption{A typical path of the Brownian motion.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.4\textwidth]{fig4_2.png}
\caption{A typical path of the process $X = (X_t)_{t \geq 0}$.}
\end{figure}

If this is the case, then one can see all the jumps in the path of $X$ and the problem of estimating $f$ is relatively easy, as no decompounding is involved. On the other hand Figures 4.3 and 4.4 provide discretised versions of the typical paths of the Brownian motion $Z$ and the process $X$. In this case both plots look similar and given the highly irregular character of Brownian paths, it is difficult to conclude at which time instances jumps occur in the process $X$. The information on $f$ is contained in the jumps and the impossibility to observe them makes the problem of estimation of $f$ much more difficult.

\begin{figure}[h]
\centering
\includegraphics[width=0.4\textwidth]{fig4_3.png}
\caption{A discretised path of the Brownian motion.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.4\textwidth]{fig4_4.png}
\caption{A discretised path of the process $X = (X_t)_{t \geq 0}$.}
\end{figure}

Nonparametric estimation of the Lévy measure of a more general Lévy process than $X$ based on low frequency observations was considered in Watteel and Kulperger (2003) and Neumann and Reiß (2007). However these authors treat the case of estimation of the Lévy measure only (or of the canonical function $K$ in case of Watteel and Kulperger (2003)) and not of its density. Moreover, they study the asymptotics of the proposed
estimators under the strong moment condition $E[|X_t|^{4+\delta}] < \infty$, where $\delta$ is some strictly positive number. We refer to those papers for additional details.

Using the stationary independent increments property of a Lévy process, see Sato (2004, Definition 1.1), we see that the problem of estimating $f$ from a discrete time sample from $X$ is equivalent to the following: let $X_1, \ldots, X_n$ be i.i.d. observations, where $X_i = Y_i + Z_i$, and $Y_i$ and $Z_i$ are independent. Assume that the unobservable $Y$’s are distributed as a random variable

$$Y = \sum_{j=1}^{N(\lambda)} W_j,$$

where $N(\lambda)$ has a Poisson distribution with parameter $\lambda$ and where the $W$’s are i.i.d. with distribution function $F$ and density $f$ and where by convention a sum over the empty set is understood to be zero. Thus we assume that $Y$ is a Poisson sum of i.i.d. $W$’s. Furthermore, let the random variables $Z_i$ have a standard normal distribution. Assume that $\lambda$ is known. The estimation problem is as follows: based on the sample $X_1, \ldots, X_n$, construct an estimator of $f$.

In this context one might also think of the $X$’s as of measurements of the realisations $Y$’s of some quantity of interest, which are corrupted by the noise $Z$. This way we are in the classical ‘signal’ plus ‘noise’ setting and the problem at hand is then related to the deconvolution problem, see e.g. Wand and Jones (1995) for an overview, and in particular to its generalisation to the case of an atomic deconvolution, see Chapter 3 of this thesis or van Es et al. (2007b).

The method that will be used to construct an estimator for $f$ is based on Fourier inversion and is similar in spirit to the use of kernel estimators in deconvolution problems, as well as our approach in Chapters 2 and 3. Let $\phi_X, \phi_Y, \phi_Z$ and $\phi_f$ denote the characteristic functions of $X, Y, Z,$ and $W$, respectively. Then by independence of $Y$ and $Z$ we have

$$\phi_X(t) = \phi_Y(t)\phi_Z(t) = e^{-\lambda + \lambda \phi_f(t)} e^{-t^2/2},$$

and therefore

$$e^{\lambda \phi_f(t)} = \frac{\phi_X(t)}{e^{-\lambda} e^{-t^2/2}}.$$

Notice that $P(Y = 0) = e^{-\lambda}$. Inverting (4.2), we get

$$\phi_f(t) = \frac{1}{\lambda} \log \left( \frac{\phi_X(t)}{e^{-\lambda} e^{-t^2/2}} \right).$$

Here Log denotes the distinguished logarithm, called so due to the similarity to the distinguished logarithm as constructed e.g. in Chow and Teicher (1978, Lemma 1, p. 413), Chung (2001, Theorem 7.6.2), Finkelestein et al. (1997) and Sato (2004, Lemma 7.6). For additional details see Section 2.1 of this thesis.

Assuming that $\phi_f$ is integrable, by Fourier inversion we obtain

$$f(x) = \frac{1}{2\pi \lambda} \int_{-\infty}^{\infty} e^{-itx} \log \left( \frac{\phi_X(t)}{e^{-\lambda} e^{-t^2/2}} \right) dt.$$
This expression will be used as the basis for construction of an estimator of $f$. Let $\phi_{\text{emp}}$ denote the empirical characteristic function of the sample $X_1, \ldots, X_n$,

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

Furthermore, let $w$ be a symmetric kernel with characteristic function $\phi_w$ supported on $[-1, 1]$ and nonzero there, and let $h > 0$ be a bandwidth. The density $q$ of $X$ can then be estimated by a kernel density estimator

$$q_{nh}(x) = \frac{1}{nh} \sum_{j=1}^{n} w \left( \frac{x - X_j}{h} \right).$$

Its characteristic function $\phi_{q_{nh}}(t) = \phi_{\text{emp}}(t)\phi_w(ht)$ will serve as an estimator of $\phi_X(t)$. For those $\omega$’s from the sample space $\Omega$, for which the distinguished logarithm in the integral below is well-defined, $f$ can be estimated by the following plug-in type estimator,

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

while for those $\omega$’s, for which the distinguished logarithm cannot be defined, we can assign an arbitrary value to $f_{nh}(x)$, e.g. zero. The distinguished logarithm in (4.3) can be defined only for those $\omega$’s for which $\phi_{\text{emp}}(t)\phi_w(ht)e^{\lambda e^{-t^2/2}}$ vanishes at every point $t \in [-1/h, 1/h]$. In fact in Section 4.2 we will prove that as $n \to \infty$, the probability of the exceptional set where the distinguished logarithm is undefined, tends to zero. For technical reasons which will become apparent in the proofs, we also need to truncate $f_{nh}(x)$, and consequently, we define the estimator of $f(x)$ not by the expression above, but by

$$\hat{f}_{nh}(x) = (M_n \wedge f_{nh}(x)) \vee (-M_n),$$

where $M = (M_n)_{n \geq 1}$ denotes a sequence of positive numbers converging to infinity at a suitable rate to be specified below.

Concluding this section, we state conditions on the density $f$, the kernel $w$, the bandwidth $h$ and the truncating sequence $M$.

**Condition 4.1.** Let the density $f$ be such that $\phi_f$ is integrable.

**Condition 4.2.** Let the kernel $w$ be the sinc kernel, $w(x) = \sin(x)/(\pi x)$.

The Fourier transform of the sinc kernel is given by $\phi_w(t) = 1_{[-1,1]}(t)$. The sinc kernel and its Fourier transform are plotted in Figures 4.5 and 4.6 on the facing page. The sinc kernel has been used successfully in kernel density estimation since a long time, see e.g. Davis (1975, 1977). It is a limit case of the so-called superkernels, i.e. integrable kernels the characteristic functions of which are identically 1 in some open neighbourhood of zero. For more information on the latter class of kernels we refer e.g. to Devroye and Győrfi (1985), Devroye (1988) or Devroye (1992). An attractive feature of the sinc kernel in ordinary kernel density estimation is that it is asymptotically optimal when one selects the mean square error or the mean integrated square error as the criterion for the performance of an estimator. Notice that the sinc kernel is not Lebesgue integrable, but its square is. Recent preprints dealing with the sinc kernel in the kernel density estimation context are Oksavik and Ushakov (2007) and Glad et al. (2007).
4.2 Main result

We first establish that with probability tending to 1 as $n \to \infty$, the distinguished logarithm in (4.3) is well-defined. Thus our goal is to find a set $B_{nh}$ such that on this set the distinguished logarithm might be undefined, while on the set $B_{nh}^c$ it is well-defined. Fix $\omega$ from the sample space $\Omega$ and consider the quantity

$$\sup_{t \in [-h^{-1}, h^{-1}]} \left| \frac{\phi_{\text{emp}}(t)}{e^{-\lambda}e^{-t^2/2}} - \frac{\phi_X(t)}{e^{-\lambda}e^{-t^2/2}} \right|. \quad (4.5)$$

Now suppose that there exists a small number $\delta$, such that

$$\sup_{t \in [-h^{-1}, h^{-1}]} e^{1/(2h^2)} \left| \frac{\phi_{\text{emp}}(t)}{e^{\lambda}} - \frac{\phi_X(t)}{e^{\lambda}} \right| \leq \delta.$$  

Obviously this implies that (4.5) is less than $\delta$. If $\delta$ is small enough, then the two functions in (4.5) are uniformly close to each other on $[-1/h, 1/h]$ and since $\phi_X(t)e^{\lambda}e^{t^2/2} = \exp[\lambda \phi_f(t)]$ is bounded away from zero, also $\phi_{\text{emp}}(t)e^{\lambda}e^{t^2/2}$ will be bounded away from
zero. From this it follows that on this interval one can define the distinguished logarithm of $\phi_{\text{emp}}(t)e^{\lambda e^{t^2/2}}$. This simple observation shows that on the set 

$$B_{nh}^c = \left\{ \omega : \sup_{t \in [-k^{-1}, k^{-1}]} e^{1/(2k^2)} \left| \frac{\phi_{\text{emp}}(t)}{e^{-\lambda}} - \frac{\phi_X(t)}{e^{-\lambda}} \right| \leq \delta \right\}$$

the distinguished logarithm will be well-defined for $\delta$ sufficiently small. Thus, what remains to be done is to prove that the probability of the complement of this set converges to zero as $n \to \infty$. To this end we will make use of the following theorem from Devroye (1994).

**Theorem 4.1.** Let $X$ be a random variable with characteristic function $\phi$ and finite first moment, and let $\phi_n$ be the empirical characteristic function of the i.i.d. sample $X_1, \ldots, X_n$ drawn from $X$. Then, for $\alpha$ and $\beta$, possibly dependent upon $n$,

$$P \left( \sup_{|t| < \alpha} |\phi_n(t) - \phi(t)| > \beta \right) \leq 4 \left( 1 + \frac{8\alpha E[|X|]}{\beta} \right) e^{-n\beta^2/2} + o(1), \quad (4.6)$$

where the $o(1)$ term is uniform over all $\alpha$ and $\beta$.

**Remark 4.1.** In our results we need additional information on the $o(1)$ term in (4.6). It follows from the proof of Theorem 4.1 that it is bounded by

$$P \left( \frac{1}{n} \sum_{j=1}^{n} X_j \geq \frac{4}{3} E[|X|] \right), \quad (4.7)$$

see Devroye (1994). Since the $X$'s are not bounded, it is not possible to apply Hoeffding’s inequality, see Hoeffding (1963), to show that this probability is exponentially small. At the same time, verification of the moment conditions needed for Bernstein’s inequality to hold is difficult in our case and might require strong conditions on $Y$. Therefore we opt for an unsophisticated application of Chebyshev’s inequality to bound this probability.

The following proposition follows from Theorem 4.1.

**Proposition 4.1.** Assume Conditions 4.2 and 4.3 and let $E[|X|] < \infty$. Then the distinguished logarithm in (4.3) is well-defined with probability tending to 1 as $n \to \infty$. Moreover, if $E[|X|^\rho] < \infty$ for $1 < \rho < 2$, then $P(B_{nh}) = O(n^{1-\rho})$, and if $E[|X|^\rho] < \infty$ for $\rho \geq 2$, then $P(B_{nh}) = O(n^{-\rho/2})$.

The main result of the chapter concerns the asymptotic normality of $\hat{f}_{nh}(x)$ at a fixed point $x$. The following theorem holds true.

**Theorem 4.2.** Suppose that $\lambda$ is known. Let the estimator $\hat{f}_{nh}(x)$ be defined as in (4.4), and assume that Conditions 4.1–4.4 hold. Furthermore, let $f$ have a finite second moment. Then

$$\sqrt{n} \left( \hat{f}_{nh}(x) - E[\hat{f}_{nh}(x)] \right) \xrightarrow{D} \mathcal{N} \left( 0, \frac{e^{2\lambda^2}}{2\pi^2 \lambda^2} \right)$$

as $n \to \infty$. 
Remark 4.2. Notice that the assumption \( x^2 f(x) \in L_1 \) implies \( E[X^2] < \infty \), see e.g. Sato (2004, Corollary 25.8).

Remark 4.3. From Theorem 4.2 it follows, that in order to get a consistent estimator, \( nh^{-2} e^{-1/k^2} \) has to diverge to infinity. This means that the bandwidth \( h \) has to be fairly large, i.e. of order \((\log n)^{-\beta}\), where \( \beta \leq 1/2 \), thus resulting in a slow, logarithmic rate of convergence of \( \hat{f}_{nh}(x) \). This is in sharp contrast with the ordinary decompounding case, where the convergence rate is polynomial, see Section 2.2. On the other hand, the convergence rate of \( \hat{f}_{nh}(x) \) is similar to that in the ordinary deconvolution, as well as the deconvolution for an atomic distribution, when the error distribution is assumed to be supersmooth, see e.g. Fan (1991a) and van Es et al. (2007b) or Chapter 3. This fact should not come as a surprise, due to the similar structure of these problems and the presence of Gaussian noise in our model. We also mention that in a recent preprint Neumann and Reiß (2007), under some conditions on the Lévy measure \( \nu \), obtained similar logarithmic lower bounds for estimation from low frequency observations of the Lévy measure \( \nu \) of a general Lévy process with a Brownian component.

Remark 4.4. We conjecture that the requirement \( E[X^2] < \infty \) that appears in Theorem 4.2 can be relaxed to the requirement \( E[|X|^\rho] < \infty \) with \( \rho > 3/2 \). This observation is based on the results in van Es and Uh (2004).

Remark 4.5. Using the estimator \( p_{ng} \), cf. (3.8) from Section 3.1 of Chapter 3, an estimator of \( \lambda \) can be defined as \( \lambda_{ng} = -\log p_{ng} \), of course provided that \( p_{ng} \) is strictly positive. However the proof of Theorem 4.2 for the case of unknown \( \lambda \) is a highly nontrivial task.

Apart of Theorem 4.2, it is also interesting to study the asymptotic distribution of

\[
\frac{\sqrt{n}}{he^{1/(2h^2)}} (\hat{f}_{nh}(x) - f(x)),
\]

i.e. of the estimator \( \hat{f}_{nh}(x) \) centred at the true density \( f(x) \). After rewriting the above expression as

\[
\frac{\sqrt{n}}{he^{1/(2h^2)}} (\hat{f}_{nh}(x) - f(x)) = \frac{\sqrt{n}}{he^{1/(2h^2)}} (\hat{f}_{nh}(x) - E[\hat{f}_{nh}(x)])
+ \frac{\sqrt{n}}{he^{1/(2h^2)}} (E[\hat{f}_{nh}(x)] - f(x)),
\]

we see that we have to study the behaviour of the bias of the estimator \( \hat{f}_{nh}(x) \), which is given by \( E[\hat{f}_{nh}(x)] - f(x) \). It will turn out that the behaviour of the bias depends on the tail behaviour of the characteristic function of \( f \). We will distinguish two cases: in the first case we will assume that \( \phi_f(t) = O(e^{-|t|^\alpha}) \) with \( 1 < \alpha \leq 2 \), and in the second case we will assume that \( \phi_f(t) = O(|t|^{-\gamma}) \) as \( t \to \infty \) with \( \gamma > 1 \). These two cases find a parallel in deconvolution problems, where a distinction is made between the use of supersmooth or ordinary smooth distributions to model the error distribution, see e.g. Fan (1991a).
Proposition 4.2. Suppose that $\lambda$ is known. Let the estimator $\hat{f}_{nh}(x)$ be defined as in (4.4), and assume that Conditions 4.1–4.4 hold. Furthermore, let $f$ have a finite $\rho$th moment, $\rho > 1$.

(i) If $\phi_f(t) = O(e^{-|t|^\alpha})$ as $|t| \to \infty$ for $1 < \alpha \leq 2$, then we have
\[
E[\hat{f}_{nh}(x)] - f(x) = O(h^{\alpha - 1}e^{-1/h^\alpha})
\]
as $n \to \infty$.

(ii) If $\phi_f(t) = O(|t|^{-\gamma})$ as $|t| \to \infty$ for $\gamma > 1$, then
\[
E[\hat{f}_{nh}(x)] - f(x) = O(h^{\gamma - 1})
\]
as $n \to \infty$.

Remark 4.6. Despite the fact that the bias of $\hat{f}_{nh}$ asymptotically vanishes under Condition 4.3, the consequence of Proposition 4.2 is that the asymptotic normality of (4.8) cannot be established for symmetric stable densities. Of course, it cannot be established for other densities either, the characteristic functions of which decay algebraically. Examination of the proof of Proposition 4.2 demonstrates that in order to have that (4.8) is asymptotically normal, one has to assume that, e.g., $\phi_f(t) = e^{-|t|^\alpha}$ with $\alpha > 2$. However, if this is the case, then $\phi'_f(0) = \phi''_f(0) = 0$ and consequently the first two moments of $f$ have to vanish. There does not exist a density with such properties.

Remark 4.7. It appears that in our case the square bias always dominates the variance of the estimator. This is not surprising in view of similar results obtained in Butucea and Tsybakov (2004) for the ordinary deconvolution problem: suppose
\[
X = Y + Z,
\]
where $Y$ and $Z$ are such that
\[
\int_{-\infty}^{\infty} |\phi_Y(t)|^2 \exp(2\alpha |t|^\gamma) \leq 2\pi L,
\]
\[
b_{\min}|t|^\gamma \exp(-\beta |t|^s) \leq \phi_Z(t) \leq b_{\max}|t|^\gamma' \exp(-\beta' |t|^s).
\]
Here $\alpha, r, L, b_{\min}, r, b_{\max}$, are strictly positive constants, $\gamma$ are $\gamma'$ are real numbers, and it is assumed that $r < s$. Then the square bias of the deconvolution kernel density estimator, which is based on observations on $X$ and is evaluated for the sinc kernel, dominates the variance. The similarity to the model in Butucea and Tsybakov (2004) holds true when comparing $\phi_f$ and $\phi_Z$, as $\phi_Z$ in a certain sense represents an extreme case among characteristic functions of supersmooth distributions.

4.3 Simulation example

Practical implementation of the estimator (4.4) is not a straightforward task. The idea we use is similar to that of Section 2.3. Notice that we can rewrite (4.3) as $f_{nh}(x) =$
\[ f^{(1)}_{nh}(x) + f^{(2)}_{nh}(x), \]

where

\[ f^{(1)}_{nh}(x) = \frac{1}{2\pi\lambda} \int_{0}^{\infty} e^{-itx} \log \left( \frac{\phi_{emp}(t)}{e^{-\lambda e^{-t^2/2}}} \right) dt, \]

\[ f^{(2)}_{nh}(x) = \frac{1}{2\pi\lambda} \int_{0}^{\infty} e^{itx} \log \left( \frac{\phi_{emp}(-t)}{e^{-\lambda e^{-t^2/2}}} \right) dt. \]

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

\[ f^{(1)}_{nh}(x) \approx \frac{1}{2\pi\lambda} \sum_{j=1}^{N} e^{-iv_jx} \psi(v_j) \eta. \quad (4.10) \]

Here we take \( N \) to be some power of 2 and \( \psi \) is defined by

\[ \psi(v_j) = \log \left( \frac{\phi_{emp}(v_j)}{e^{-\lambda e^{-v_j^2/2}}} \right). \]

From this point on one can proceed as in Section 2.3 of Chapter 2. A similar reasoning applies to \( f^{(2)}_{nh}(x) \).

The general difficulty with implementing the estimator is the computation of the distinguished logarithm, i.e. of function \( \psi \). A way to do this is to take a fine grid of points, evaluate the argument of the ordinary logarithm there and if one sees large jumps of size comparable to \( 2\pi \) between two consecutive points, make appropriate changes to the argument, thus obtaining an approximation to the argument of the distinguished logarithm, e.g. the way it was done in Figure 2.3 from Chapter 2. Of course this approach works only when \( \phi_{emp}(t) \) does not vanish on \([-1/h, 1/h]\]. The latter fact can be verified in theory only, while in practice this can be done only for a grid of points \( t_1, t_2, \ldots, t_k \), which thus has to be taken rather fine, so that one does not possibly miss the value zero.

Though our emphasis is more on theoretical aspects of decompounding under Gaussian noise, we nevertheless will consider one simulation example in this section. We took \( \lambda = 1 \) and \( f \) the standard normal density and simulated a sample of size \( n = 5000 \). The bandwidth \( h = 0.5 \) was selected by hand. The resulting estimate \( f_{nh} \) (bold dotted line) together with the true density \( f \) (dashed line) is plotted in Figure 4.7 on the next page. We notice that the fit is quite good. Furthermore, notice that

\[ P(N(\lambda) \geq 2) = 1 - \frac{2}{e} \approx 0.264. \]

It turns out that we considered a nontrivial example, since a considerable number among the \( Y_i \)'s are sums of the \( W_j \)'s.

We should stress the fact that this simulation example serves as an illustration only and an extensive simulation study is needed to investigate the finite sample performance of our estimator and its behaviour in practice. We have to be very careful when generalising our conclusions concerning this simulation example because of the fact that the empirical characteristic function is oscillatory in its tails. If the integration step size \( \eta \) is not small enough, we might miss instances when \( \phi_{emp} \) crosses the negative real axis. This will have direct consequences for the argument of the distinguished logarithm. This
is especially true for relatively small sample sizes, for which the empirical characteristic function \( \phi_{emp} \) might not approximate the true characteristic function \( \phi_X \) well enough. The issue of selection of \( \eta \) in practice remains open and a thorough simulation study is needed to obtain some practical recommendations how this can be done. Additionally, a data-dependent method of the bandwidth selection has to be created.

\[
\text{Figure 4.7: Estimation of the standard normal density, } n = 5000.
\]

4.4 Proofs

*Proof of Proposition 4.1.* Note that we have

\[
P(B_{nh}) \leq 4 \left( 1 + \frac{8E[|X|] e^{1/(2\delta^2) h}}{e^\lambda \delta} \right) \exp \left( \frac{-e^{-2\lambda \delta^2 n} e^{-1/(2h^2)}}{72} \right) + o(1),
\]

where we assume that \( \delta \) is small enough. This bound follows from Theorem 4.1 with \( \alpha = 1/h \) and \( \beta = e^{-\lambda \delta} e^{-1/(2h^2)} \). The right-hand side converges to zero as \( n \to \infty \) due to Condition 4.3, cf. (4.12). To prove the proposition, the only additional fact that we need to verify is that the \( o(1) \) term from Theorem 4.1, which in the proof of Theorem 4.1 from Devroye (1994) is bounded by (4.7), is of order \( n^{1-\rho} \), if \( 1 < \rho < 2 \), and is of order \( n^{-\rho/2} \), if \( \rho \geq 2 \). In fact, if the inequality

\[
\left| \frac{1}{n} \sum_{j=1}^{n} X_j \right| \geq \frac{4}{3} E[|X_1|]
\]

holds, then we have

\[
\left| \frac{1}{n} \sum_{j=1}^{n} X_j - E[X_1] \right| \geq \left| \frac{1}{n} \sum_{j=1}^{n} X_j - E[X_1] \right| \geq \frac{4}{3} E[|X_1|] - E[|X_1|] = \frac{1}{3} E[|X_1|].
\]
By Chebyshev’s inequality this implies

\[ P\left(\left|\frac{1}{n} \sum_{j=1}^{n} X_j \right| \geq \frac{4}{3} \mathbb{E}[|X_1|]\right) \leq P\left(\left|\frac{1}{n} \sum_{j=1}^{n} X_j - \mathbb{E}[X_1]\right| \geq \frac{1}{3} \mathbb{E}[|X_1|]\right) \]

\[ \leq 3^\rho (\mathbb{E}[|X_1|])^{-\rho} \mathbb{E}\left[\left|\frac{1}{n} \sum_{j=1}^{n} X_j - \mathbb{E}[X_1]\right|^\rho\right]. \tag{4.11} \]

Suppose first that \(1 < \rho < 2\). Then it follows from Theorem 4 of von Bahr and Esseen (1965) that the expectation in the rightmost term in (4.11) is of order \(n^{1-\rho}\). Now suppose \(\rho \geq 2\). Then Theorem 2 of Dharmadhikari and Jogdeo (1969) implies that the expectation in the rightmost term of (4.11) is of order \(n^{-\rho/2}\). For explicit constants we refer to the same papers.

Assume again that \(1 < \rho < 2\). To complete the proof of the proposition, we have to verify that

\[ \frac{e^{1/(2h^2)}}{h} \exp\left(-\frac{e^{-2\lambda^2 n e^{-1/h^2}}}{72}\right) n^{\rho-1} \to 0. \tag{4.12} \]

To this end we take the logarithm of the left-hand side to obtain

\[ 1/(2h^2) - \log h - \frac{e^{-2\lambda^2}}{72} n e^{-1/h^2} + (\rho - 1) \log n. \tag{4.13} \]

The first term here is of order \((\log n)^{2/3}\) and is negligible compared to \(\log n\). The second term is of order \(\log \log n\) and is thus negligible, while the third term dominates \(\log n\). Therefore (4.13) diverges to minus infinity and consequently (4.12) holds. The proof for the case \(\rho \geq 2\) is virtually identical and therefore it is omitted.

\[ \square \]

**Proof of Theorem 4.2.** Write \(\zeta_n(h) = \sqrt{n}h^{-1}e^{-1/(2h^2)}\). We have

\[ \zeta_n(h)(\hat{f}_{\text{nh}}(x) - \mathbb{E}[\hat{f}_{\text{nh}}(x)]) = \zeta_n(h)(\hat{f}_{\text{nh}}(x) - f(x)) + \zeta_n(h)(f(x) - \mathbb{E}[\hat{f}_{\text{nh}}(x)]) \]

\[ = \zeta_n(h)((\hat{f}_{\text{nh}}(x) - f(x))1_{B_{\text{nh}}} - \mathbb{E}[(\hat{f}_{\text{nh}}(x) - f(x))1_{B_{\text{nh}})]) \]

\[ + \zeta_n(h)((\hat{f}_{\text{nh}}(x) - f(x))1_{B_{\text{nh}}} - \mathbb{E}[(\hat{f}_{\text{nh}}(x) - f(x))1_{B_{\text{nh}}})], \]

where the set \(B_{\text{nh}}\) is defined as in Section 4.2. Now notice that, for an arbitrary constant \(\eta > 0\), by Chebyshev’s inequality we have

\[ P(\zeta_n(h)(\hat{f}_{\text{nh}}(x) - f(x))1_{B_{\text{nh}}} - \mathbb{E}[(\hat{f}_{\text{nh}}(x) - f(x))1_{B_{\text{nh}}}) > \eta \]

\[ \leq \frac{2}{\eta} \zeta_n(h) \mathbb{E}[(\hat{f}_{\text{nh}}(x) - f(x))1_{B_{\text{nh}}}). \tag{4.14} \]

Since \(\phi_f\) is integrable, it follows that \(|f(x)| \leq C\), where \(C\) is some constant. It then follows that the probability at the left-hand side of (4.14) is bounded by

\[ \frac{2}{\eta} \zeta_n(h)(M_n + C) P(B_{\text{nh}}). \tag{4.15} \]
Now we apply the bound of Proposition 4.1 to \( P(B_{nh}) \). To prove that (4.15) converges to zero, it is sufficient to verify that

\[
\frac{\sqrt{n}}{he^{1/(2h^2)}} (M_n + C) \frac{1}{n} \rightarrow 0. \tag{4.16}
\]

This is obviously true due to Condition 4.3 and 4.4. Therefore

\[
\zeta_n(h) \{ (\hat{f}_{nh}(x) - f(x))1_{B_{nh}} - E[(\hat{f}_{nh}(x) - f(x))1_{B_{nh}}] \} \overset{P}{\rightarrow} 0.
\]

Hence by Slutsky’s theorem, see van der Vaart (1998, Lemma 2.8), this term can be neglected and it suffices to consider

\[
\zeta_n(h) \{ (\hat{f}_{nh}(x) - f(x))1_{B_{nh}^c} - E[(\hat{f}_{nh}(x) - f(x))1_{B_{nh}^c}] \}.
\]

By the arguments similar to those that we used on p. 25, on the set \( B_{nh}^c \), for \( n \) large enough, the truncation in (4.4) becomes unimportant,

\[
\hat{f}_{nh}(x)1_{B_{nh}^c} = f_{nh}(x)1_{B_{nh}^c}. \tag{4.17}
\]

Thus we have to consider

\[
\zeta_n(h) \{ (f_{nh}(x) - f(x))1_{B_{nh}^c} - E[(f_{nh}(x) - f(x))1_{B_{nh}^c}] \}.
\]

Plugging in the expressions for \( f_{nh}(x) \) and \( f(x) \), we obtain that the above expression is equal to

\[
\zeta_n(h) \left\{ \frac{1}{2\pi} \int_{-1/h}^{1/h} e^{-ixt} \log \left( \frac{\phi_{emp}(t)}{\phi_X(t)} \right) dt 1_{B_{nh}^c} 
- E \left[ \frac{1}{2\pi} \int_{-1/h}^{1/h} e^{-ixt} \log \left( \frac{\phi_{emp}(t)}{\phi_X(t)} \right) dt 1_{B_{nh}^c} \right] 
- \frac{1}{2\pi} \left\{ \int_{-\infty}^{-1/h} e^{-ixt} \phi_f(t) dt + \int_{1/h}^{\infty} e^{-ixt} \phi_f(t) dt \right\} (1_{B_{nh}^c} - E[1_{B_{nh}^c}]) \right\}. \tag{4.18}
\]

First notice that

\[
\left| \int_{-\infty}^{-1/h} e^{-ixt} \phi_f(t) dt + \int_{1/h}^{\infty} e^{-ixt} \phi_f(t) dt \right| \leq \int_{\infty}^{\infty} |\phi_f(t)| dt < \infty.
\]

Consequently, the last term in (4.18) converges to zero in probability if

\[
\zeta_n(h) (1_{B_{nh}^c} - E[1_{B_{nh}^c}]) \overset{P}{\rightarrow} 0.
\]

This in turn is equivalent to

\[
\zeta_n(h) (1_{B_{nh}} - E[1_{B_{nh}}]) \overset{P}{\rightarrow} 0.
\]
because \(1_{B_{nh}^c} = 1 - 1_{B_{nh}}\). By Chebyshev’s inequality it is sufficient to prove that 
\(\zeta_n(h) P(B_{nh}) \to 0\). However, this follows from (4.15) and (4.16).

Hence, by Slutsky’s theorem we have to consider the first term of (4.18),

\[
\zeta_n(h) \left\{ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \log \left( \frac{\phi_{emp}(t)}{\phi_X(t)} \right) dt 1_{B_{nh}^c} \right. 
- \left. \mathbb{E} \left[ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \log \left( \frac{\phi_{emp}(t)}{\phi_X(t)} \right) dt 1_{B_{nh}^c} \right] \right\}. 
\]

Rewrite this as

\[
\zeta_n(h) \left\{ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \left( \phi_{emp}(t) \phi_X(t) - 1 \right) dt 1_{B_{nh}^c} \right. 
- \left. \mathbb{E} \left[ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \left( \phi_{emp}(t) \phi_X(t) - 1 \right) dt 1_{B_{nh}^c} \right] \right\} 
+ \zeta_n(h) \left\{ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} R_{nh}(t) dt 1_{B_{nh}^c} \right. 
- \left. \mathbb{E} \left[ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} R_{nh}(t) dt 1_{B_{nh}^c} \right] \right\}, 
\]

where

\[
R_{nh}(t) = \log \left( 1 + \left\{ \frac{\phi_{emp}(t)}{\phi_X(t)} - 1 \right\} - \left\{ \frac{\phi_{emp}(t)}{\phi_X(t)} - 1 \right\} \right). 
\]

Notice that on the set \(B_{nh}^c\) we have

\[
\left| \frac{\phi_{emp}(t)}{\phi_X(t)} - 1 \right| < \frac{1}{2}, 
\]

if \(\delta\) is small enough. It follows from Lemma 2.3 of Section 2.4 that

\[
|R_{nh}(t)| \leq \left( \frac{\phi_{emp}(t)}{\phi_X(t)} - 1 \right)^2. 
\]

Consequently, to prove that the second term in (4.19) asymptotically vanishes, it is sufficient to prove that

\[
\zeta_n(h) \frac{1}{2\pi \lambda} \mathbb{E} \left[ \int_{-1/h}^{1/h} |R_{nh}(t)| dt 1_{B_{nh}^c} \right] 
\leq \zeta_n(h) \frac{1}{2\pi \lambda} \mathbb{E} \left[ \int_{-1/h}^{1/h} \left| \frac{\phi_{emp}(t)}{\phi_X(t)} - 1 \right|^2 dt \right] \to 0. 
\]

Since \(|\phi_Y(t)|^{-1} \leq e^{2\lambda}\), we have

\[
\mathbb{E} \left[ \int_{-1/h}^{1/h} \left| \frac{\phi_{emp}(t)}{\phi_X(t)} - 1 \right|^2 dt \right] \leq C h^{1/2} \mathbb{E} \left[ \int_{-\infty}^{\infty} |\phi_{emp}(t)\phi_u(ht) - \phi_X(t)\phi_u(ht)|^2 dt \right]. 
\]
where $\phi_w$ is the characteristic function of the sinc kernel and $C$ is a constant. By Parseval’s identity the expectation on the right-hand side equals

$$\frac{1}{2\pi} E \left[ \int_{-\infty}^{\infty} (q_{nh}(x) - q * w_h(x))^2 dx \right].$$

The expectation here is the integrated variance of a kernel estimator $q_{nh}$, which is of order $(nh)^{-1}$, see Tsybakov (2004, Proposition 1.7). Thus we have to show that

$$\sqrt{n} he 1/(2h^2) e^{1/(2h^2)} h^2 \frac{1}{\sqrt{n}} \to 0.$$  \hspace{1cm} (4.22)

The result follows from Condition 4.3 and can be verified by taking the logarithm of the left-hand side of the above expression and concluding that it diverges to minus infinity. We obtain

$$\frac{1}{2h^2} = \log h^2 - \frac{1}{2} \log n \to -\infty,$$

because $h^{-2} = (\log n)^{2\beta}$ and $2\beta < 1$, and hence the dominating term on the left-hand side in the above expression is the last one.

We deal with the first summand in (4.19). Rewrite it as

$$\zeta_n(h) \left\{ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \frac{\phi_{emp}(t)}{\phi_X(t)} dt 1_{B_{nh}^c} - \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \frac{\phi_{emp}(t)}{\phi_X(t)} dt 1_{B_{nh}^c} - \zeta_n(h) \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} dt (1_{B_{nh}^c} - E[1_{B_{nh}^c}]) \right\}.$$  \hspace{1cm} (4.22)

We want to show that the second summand in this expression converges to zero in probability. Notice, that it is bounded by

$$C \zeta_n(h) \frac{1}{h} |1_{B_{nh}^c} - E[1_{B_{nh}^c}]|,$$

because $1_{B_{nh}^c} = 1 - 1_{B_{nh}}$. Here $C$ is some constant. By Chebyshev’s inequality it is sufficient to prove that

$$\zeta_n(h) \frac{1}{h} P(B_{nh}) \to 0.$$  \hspace{1cm} (4.22)

This is obviously true thanks to (4.15) and (4.16).

Thus, by Slutsky’s theorem, instead of (4.22) we may consider

$$\zeta_n(h) \left\{ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \frac{\phi_{emp}(t)}{\phi_X(t)} dt 1_{B_{nh}^c} - \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \frac{\phi_{emp}(t)}{\phi_X(t)} dt 1_{B_{nh}^c} \right\}.$$  \hspace{1cm} (4.22)

Note that for $|t| \leq 1/h$ the inequality

$$|\phi_X(t)| = \left| e^{-t^2/2} e^{-\lambda + \lambda \phi_f(t)} \right| \geq e^{-1/(2h^2)} e^{-2\lambda}$$
holds. Consequently, we have
\[
\zeta_n(h) E \left[ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \frac{\phi_{\text{emp}}(t)}{\phi_X(t)} dt \right] = C_n(h) \frac{2}{2\pi \lambda h} e^{1/(2h^2)} e^{2\lambda} P(B_{nh}), \quad (4.23)
\]
which converges to zero thanks to the fact that \( P(B_{nh}) = O(n^{-1}) \), see Proposition 4.1.

Hence by Slutsky’s theorem we may consider
\[
\zeta_n(h) \left\{ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \phi_{\text{emp}}(t) \frac{1}{\phi_X(t)} dt - \mathbb{E} \left[ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \phi_{\text{emp}}(t) \frac{1}{\phi_X(t)} dt \right] \right\}.
\]

By (4.1) the expression above can be rewritten as
\[
\zeta_n(h) \left\{ \frac{e^\lambda}{2\pi} \int_{-1/h}^{1/h} e^{-itx} \left( \frac{\phi_{\text{emp}}(t)}{e^{-t^2/2}} - \frac{\phi_X(t)}{e^{-t^2/2}} \right) dt 
+ \zeta_n(h) \frac{e^\lambda}{2\pi} \int_{-1/h}^{1/h} e^{-itx} \left( \frac{\phi_{\text{emp}}(t)}{e^{-t^2/2}} - \frac{\phi_X(t)}{e^{-t^2/2}} \right) \left( e^{-\lambda \phi_f(t)} - 1 \right) dt. \quad (4.24)
\]

Let us consider the first summand in this expression. By Corollary 3.1 it is asymptotically normal with zero mean and variance given by
\[
\sigma^2 = \frac{e^{2\lambda}}{2\pi^2 \lambda^2}.
\]

Now we will show that the second term in (4.24) asymptotically vanishes in probability. By Chebyshev’s inequality it suffices to show
\[
(\zeta_n(h))^2 \mathbb{E} \left[ \left| \frac{e^\lambda}{2\pi} \int_{-1/h}^{1/h} e^{-itx} \left( \frac{\phi_{\text{emp}}(t)}{e^{-t^2/2}} - \frac{\phi_X(t)}{e^{-t^2/2}} \right) \left( e^{-\lambda \phi_f(t)} - 1 \right) dt \right|^2 \right] 
= (\zeta_n(h))^2 \text{Var} \left[ \int_{-1/h}^{1/h} e^{-itx} \phi_{\text{emp}}(t) \frac{1}{e^{-t^2/2}} \left( e^{-\lambda \phi_f(t)} - 1 \right) dt \right] \to 0.
\]

Using the independence of the \( X_i \)'s, after further simplification we obtain that we have to prove that
\[
\frac{1}{h^2} e^{1/h^2} \left( \int_{-1/h}^{1/h} e^{t^2/2} |e^{-\lambda \phi_f(t)} - 1| dt \right)^2 \to 0.
\]

Thus we have to prove that
\[
\frac{1}{he^{1/(2h^2)}} \int_{-1/h}^{1/h} e^{t^2/2} |e^{-\lambda \phi_f(t)} - 1| dt \to 0.
\]

By the same argument which was used to prove formula (2.11), we have
\[
|e^{-\lambda \phi_f(t)} - 1| \leq C_\lambda |\phi_f(t)|,
\]
where the constant $C_\lambda$ depends on $\lambda$ only. Therefore it suffices to prove
\[
\frac{1}{he^{1/(2h^2)}} \int_{-1/h}^{1/h} e^{t^2/2} |\phi_f(t)| dt \to 0. \tag{4.25}
\]
This can be done either via an application of L'Hôpital's rule or via the method similar to the one used in the proof of Lemma 5 in van Es and Uh (2005). We follow the latter path. It is enough to consider the integral over $[0, 1/h]$ as the integral over $[-1/h, 0]$ can be dealt with in a similar fashion. After the change of integration variable $v = (1 - ht)/h^2$, we obtain from (4.25)
\[
h \int_0^{1/h^2} e^{(1-h^2v^2)/(2h^2)} \phi_f \left( \frac{1 - vh^2}{h} \right) dv = he^{1/(2h^2)} \int_0^{1/h^2} e^{v + v^2h^2/2} \phi_f \left( \frac{1 - vh^2}{h} \right) dv.
\]
By the Riemann-Lebesgue theorem $\lim_{|u| \to \infty} \phi_f(u) = 0$, and therefore by the dominated convergence theorem the above expression is of lower order than $he^{1/(2h^2)}$. The dominated convergence theorem is applicable, because
\[
(e^{-v^2/2} - e^{-v+v^2h^2/2})1_{[0, 1/h^2]} \geq 0,
\]
and hence $e^{-v^2/2}$ can be taken as the dominating function. Consequently (4.25) vanishes as $h \to 0$ and this argument concludes the proof of the theorem.

\[
\square
\]

**Proof of Proposition 4.2.** We will prove both parts of the proposition simultaneously. Write
\[
E[\hat{f}_{nh}(x)] - f(x) = E[(\hat{f}_{nh}(x) - f(x))1_{B_{nh}}] + E[(\hat{f}_{nh}(x) - f(x))1_{B^c_{nh}}]. \tag{4.26}
\]
Notice that for some $C > 0$,
\[
\left| E[(\hat{f}_{nh}(x) - f(x))1_{B_{nh}}] \right| \leq (M_n + C) P(B_{nh}). \tag{4.27}
\]
Here we used the fact that $f$ is bounded, because $\phi_f$ is integrable. Due to Proposition 4.1 and Condition 4.4, we see that (4.27) converges to zero as $n \to \infty$. Moreover, this term is negligible compared to $h^{\alpha-1} e^{-1/h^2}$ (case (i)) or $h^{\gamma-1}$ (case (ii)), since $P(B_{nh}) = O(n^{1-\rho})$ or $P(B_{nh}) = O(n^{-\rho/2})$, depending whether $1 < \rho < 2$ or $\rho \geq 2$, see Proposition 4.1.

Now we turn to the second summand in (4.26). By selecting $\delta$ small enough, on the set $B^c_{nh}$ truncation in the definition of $\hat{f}_{nh}(x)$ becomes unimportant, see the arguments that led to (4.17). Hence we have to deal with $E[(\hat{f}_{nh}(x) - f(x))1_{B^c_{nh}}]$. Using expressions for $f_{nh}(x)$ and $f(x)$, we see that this term equals
\[
E \left[ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \log \left( \frac{\phi_{emp}(t)}{\phi_x(t)} \right) dt 1_{B^c_{nh}} \right] = \frac{1}{2\pi} \int_{-1/h}^{1/h} e^{-itx} \phi_f(t) dt P(B^c_{nh}) - \frac{1}{2\pi} \int_{1/h}^{\infty} e^{-itx} \phi_f(t) dt P(B^c_{nh}). \tag{4.28}
\]
The last two terms in this expression can be treated similarly and therefore we consider only the second one. It will turn out that these are the leading terms in the bias expansion. Notice that
\[
\frac{1}{2\pi} \int_{1/h}^{\infty} e^{-itx} \phi_f(t) dt P(B_{nh}^c) \to 0
\]
as \( n \to 0 \), because \( \phi_f \) is integrable. Moreover, if \( \phi_f(t) = O \left( e^{-|t|^{\alpha}} \right), \alpha > 1 \), then
\[
\int_{1/h}^{\infty} |\phi_f(t)| dt = O \left( h^{\alpha-1} e^{-1/h^\alpha} \right).
\]
This fact can be proved using the same type of arguments as in Casella and Berger (2002, Example 3.6.3, p. 123). Furthermore, if \( \phi_f(t) = O \left( |t|^{-\gamma} \right), \) then
\[
\int_{1/h}^{\infty} |\phi_f(t)| dt \leq C \int_{1/h}^{\infty} t^{-\gamma} dt = O(h^{\gamma-1}).
\]
Now we turn to the first term in (4.28). We have
\[
E \left[ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \left( \frac{\phi_{emp}(t)}{\phi_X(t)} - 1 \right) dt 1_{B_{nh}^c} \right] = E \left[ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} R_{nh}(t) dt 1_{B_{nh}^c} \right], \tag{4.29}
\]
where \( R_{nh}(t) \) is defined as in (4.20). Consider the first term in this expression. Rewrite it as
\[
E \left[ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \left( \frac{\phi_{emp}(t)}{\phi_X(t)} - 1 \right) dt 1_{B_{nh}^c} \right] = E \left[ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \left( \frac{\phi_{emp}(t)}{\phi_X(t)} - 1 \right) dt \right] = E \left[ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \left( \frac{\phi_{emp}(t)}{\phi_X(t)} - 1 \right) dt 1_{B_{nh}^c} \right].
\]
The expectation of the first summand here is equal to zero. As far as the second summand is concerned, notice that for \( n \) large enough
\[
\left| \int_{-1/h}^{1/h} e^{-itx} \left( \frac{\phi_{emp}(t)}{\phi_X(t)} - 1 \right) dt \right| \leq C \frac{1}{h} e^{1/(2h^2)},
\]
where \( C \) is some constant. This inequality follows from the facts that for \( t \in [-1/h, 1/h] \),
\[
\left| \frac{\phi_{emp}(t)}{\phi_X(t)} - 1 \right| \leq \left| \frac{\phi_{emp}(t)}{\phi_X(t)} \right| + 1,
\]
\[
\left| \frac{\phi_{emp}(t)}{\phi_X(t)} \right| \leq e^{2\lambda} e^{1/(2h^2)},
\]
because \( \phi_X(t) = \phi_Y(t)e^{-t^2/2} \) and \( |\phi_Y(t)| \geq e^{-2\lambda} \). Consequently
\[
\left| \mathbb{E} \left[ \frac{1}{2\pi \lambda} \int_{-1/h}^{1/h} e^{-itx} \left( \frac{\hat{\phi}_{\text{emp}}(t)}{\hat{\phi}_X(t)} - 1 \right) dt 1_{B_{nh}} \right] \right| \leq C \frac{1}{h} e^{1/(2h^2)} P(B_{nh}).
\]
The right-hand side converges to zero as \( n \to \infty \) due to Proposition 4.1 and Condition 4.3. Moreover, due to the same facts it is negligible compared to \( h^{-1} \) or to \( h^{\alpha-1}e^{-1/h^\alpha} \).

Now we consider the second term in (4.29). Notice that this term is of order \((nh)^{-1}\), which was shown in the proof of Theorem 4.2 in the arguments concerning (4.21). Consequently, it is be negligible compared to \( h^{-1} \) or to \( h^{\alpha-1}e^{-1/h^\alpha} \). This completes the proof of the proposition.

**Proof of Remark 4.6.** We have to study the behaviour of
\[
\zeta_n(h)h^{\alpha-1}e^{-1/h^\alpha}.
\]
After taking the logarithm, we obtain
\[
\frac{1}{2} \log n - \log h - \frac{1}{2h^2} + (\alpha - 1) \log h - \frac{1}{h^\alpha}.
\]

Dominating terms here are the first, the third and the last one. Now note that the third and the last terms equal \(-(\log n)^{2\beta}/2\) and \(-(\log n)^{\alpha\beta}\), respectively. In view of \( 2\beta < 1 \) and \( \alpha\beta < 1 \), these terms are dominated by \( \log n \) and hence (4.31) diverges to plus infinity. It follows that so does (4.30). The case of \( \zeta_n(h)h^{-1} \to \infty \) is trivial given Condition 4.3.