Estimation of Stopping Times for Some Stopped Random Processes

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

Der Fakultät für Mathematik
der Technischen Universität Dortmund
vorgelegt von

M.Sc.-Wirt.-Math. Viktor Schulmann

im Februar 2019

Dissertation

Estimation of Stopping Times for Some Stopped Random Processes

Fakultät für Mathematik

Technische Universität Dortmund

Erstgutachter: Prof. Dr. Michael Voit

Zweitgutachterin: Prof. Dr. Jeannette Woerner

Tag der mündlichen Prüfung: 09.05.2019

"Math is useful because it's difficult."

— A. ALEKSANDROV, Soviet/Russian mathematician, physicist, philosopher and mountaineer

Contents

Acknowledgments v

Chapter 1. Introduction 1

Chapter 2. Preliminaries 5
2.1. Complex Random Variable: Expected Value, Covariance, Variance 5
2.2. Mellin Transform 9
2.3. Notation for Asymptotic Analysis 15
2.4. Gamma Function 15

Chapter 3. Estimation for Self-Similar Processes 19
3.1. The Setting 19
3.2. Construction of an Estimator for Self-Similar Processes 21
3.3. Convergence Analysis 23
3.4. Classes $\mathcal{C}$ and $\mathcal{D}$ 24
3.5. Upper Bounds on the Bias 28
3.6. Upper Bounds on the Variance 31
3.7. Application to Bessel Processes 32
3.8. Asymptotic Normality, Unbiasedness and Variance in Context of Bessel Processes 38
3.9. Simulation Study 50
3.10. Some Other Self-Similar Processes 53

Chapter 4. Processes Associated with a Convolution Semi-group 61
4.1. Introduction to Hypergroups 61
4.2. Lévy Processes on Commutative Hypergroups 64
4.3. Noncompact Sturm-Liouville Hypergroups 65
4.4. Lévy Processes with Spherical Symmetry 69

Chapter 5. Estimation for Lévy Processes on Sturm-Liouville Hypergroups 71
5.1. Construction of the Estimator 71
5.2. Convergence 74
5.3. Application to a Bessel Process 81

Chapter 6. Optimality 85
6.1. Minimax Risk 85

6.2. Reduction to two Hypotheses 86
6.3. Optimality for the Class $\mathcal{C}$ 88
6.4. Optimality for the Class $\mathcal{D}$ 104

Appendix. R Source Code 107

Appendix. List of Symbols 111

Bibliography 113

Acknowledgments

I would like to thank my advisor Prof. Dr. Michael Voit for suggesting the topic of this thesis, for his helpful ideas and our insightful discussions along the way, and also for constantly encouraging me to complete my work.

I am also deeply grateful to all of my colleagues from the chair “Analysis und Stochastik” at TU Dortmund for maintaining a productive and amicable working atmosphere.

Additionally, I wish to thank the Research Training Group “High-dimensional Phenomena in Probability - Fluctuations and Discontinuity” for their financial support.

CHAPTER 1

Introduction

Consider the following problem from physics: A radiation source is placed at the center of a screen (see Figure 1). At certain time intervals the source releases particles. These move around the screen following a path of some known random process $(Y_t)_{t\geq 0}$ without interacting with each other and without us being able to observe their movement until they die after some random time T. During its death a particle leaves a mark such that we can measure the distance $X = ||Y_T||_2$ it traveled from the source during its lifetime. Based on these observed distances we wish to infer the life span T of a particle or, in particular, the density f_T of T.

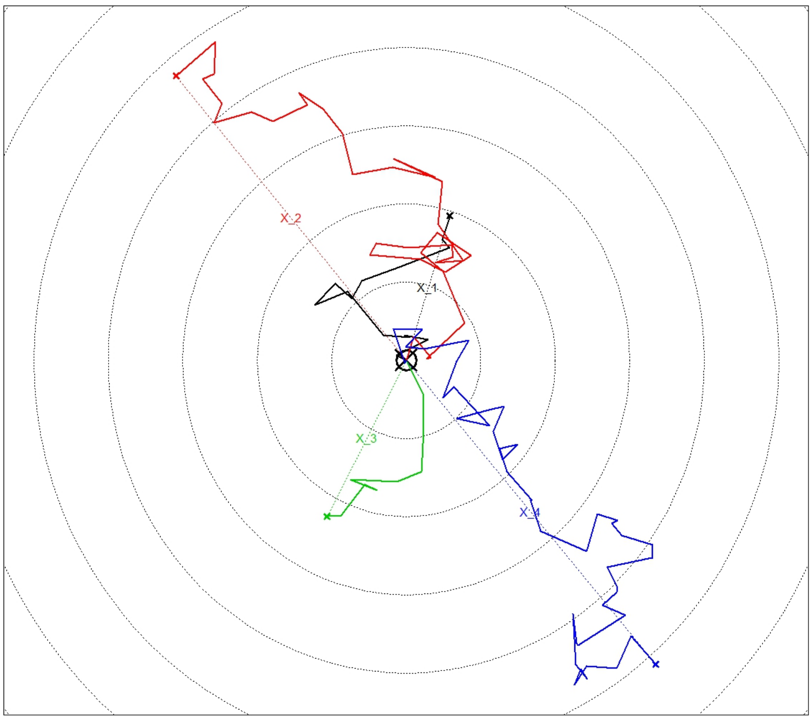

FIGURE 1. Four particles are released from the source and travel distances $X_1,\ X_2,\ X_3,\ X_4$ before expiring.

This statistical problem is inspired by the so-called *Skorokhod embedding problem* or *Skorokhod stopping problem* named after the Ukrainian mathematician who first stated and solved it in [**34**] (English translation [**35**]). Originally it consists of representing a given centered distribution with finite second moment as the law of Brownian motion at a chosen integrable stopping time. A number of reformulations and different solutions was presented by several authors since Skorokhod's publication in 1961 (see [**28**] for an overview). In [**6**] Belomestny and Schoenmakers considered "the so-called statistical Skorokhod embedding problem of recovering the distribution on an independent random time T based on i.i.d. sample from [a one-dimensional Brownian motion at time T]". In [**7**] Belomestny and Schoenmakers extended this problem by replacing Brownian motion with a one-dimensional Lévy process. We profit from their work and firstly generalize their results concerning Brownian motion to self-similar processes. We particularly focus on Bessel processes. As a consequence, we extend results from [**6**] to multi-dimensional Brownian motion. This is accomplished by considering the two-norm of the multi-dimensional Brownian motion, thus reducing the problem to the case of a Bessel process which is a one-dimensional process and can be treated similarly to the case of one-dimensional Brownian motion. Secondly, we consider the class of so-called *Sturm-Liouville processes* which will be introduced properly in Chapter 4. These processes generalize the concept of Bessel processes and often arise as a norm of other multi-dimensional processes. This makes them perfect candidates for the role of the process $(||Y_t||_2)_{t\geq 0}$ from our physics example above which describes the distance of a particle to the source at time t. Moreover, they exhibit a similar structure to Lévy processes, so we can build on the results available for this case.
We classify the inference on T as an inverse problem or, more specifically, a parameter identification problem, which is described by Engl, Hanke and Neubauer in [**16**] as "the identification of physical parameters from observations of the evolution of the system". The authors of [**16**] warn us that inverse problems often "do not fulfill Hadamard's definition of well-posedness, i.e., for [them] one of the following properties does not hold:

(H1) For all admissible data, a solution exists.
(H2) For all admissible data, the solution is unique.
(H3) The solution depends continuously on the data."

This definition goes back to [**21**]. It is not precise and has to be adjusted to a given situation. While (H1) proves unproblematic for us, (H2) is a major concern and will require the additional assumption of stochastic independence of Y and T. Property (H3) corresponds to the boundedness of an estimator we construct for f_T. For all admissible data boundedness is not guaranteed but it can be installed by a regularization procedure. This regularization creates a trade-off between accuracy and stability in our estimation method. More specifically, we can either construct an estimator with a small bias or with a small variance but not both. A sensible

compromise will be introduced.
More specifically, inference on f_T can be classified as a problem of non-parametric estimation, which according to Tsybakov (see [**37**]) "consists in estimation, from observations, of an unknown function belonging to a sufficiently large class of functions." Tsybakov sees the focus of non-parametric estimation in the following four topics:

(1) methods of construction of the estimators,
(2) statistical properties of the estimators (convergence, rates of convergence),
(3) study of optimality of the estimators,
(4) adaptive estimation.

In the preliminary Chapter 2 we introduce the so-called *Mellin transform* which is our main tool throughout this thesis. Using this transform we cover topic (1). See Section 3.2 for the case of self-similar processes and Section 5.1 for the case of Sturm-Liouville processes. Topic (2) for these two cases is discussed in Sections 3.3-3.6 and Section 5.2 respectively. See Sections 3.7 and 5.3 for the particular case of Bessel processes. Section 3.10 provides two other brief examples for the self-similar case. Furthermore, we show asymptotic normality of the estimator constructed in Section 3.7 for the Bessel case. Topic (3) is content of Chapter 6. Topic (4) is addressed only briefly within Sections 3.9 and 5.3 when we implement the estimators for the Bessel case based on different approaches with the free software environment for statistics R.

CHAPTER 2

Preliminaries

2.1. Complex Random Variable: Expected Value, Covariance, Variance

We expect the reader to be familiar with the notion of expected value, covariance and variance of real random variables on the level of [**5**, §3]. Here we present a way to generalize these notions to variables assuming complex values. In this endeavor we roughly follow the outline in [**2**] until Lemmas 2.1.8 and 2.1.10, where two more intricate but well-known (see for instance [**6**]) facts are shown. The complex generalization will be needed later, when we construct estimators that are complex random variables and investigate their properties. Basic concepts regarding the set of complex numbers are assumed to be known.

Definition 2.1.1. *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space. We call* $X : \Omega \to \mathbb{C}$ *a complex random variable, if* $X = \mathrm{Re}(X) + i\,\mathrm{Im}(X)$, *where* $\mathrm{Re}(X)$ *and* $\mathrm{Im}(X)$ *are real random variables.*

We will define the expected value of a complex variable X with an operator on the complex vector space given by

$$L^2_{\mathbb{C}} := L^2_{\mathbb{C}}(\Omega, \mathcal{A}, P) := \{X | X \text{ is a complex random variable with } \mathrm{E}[|X|^2] < \infty\},$$

where E denotes the expectation operator of a real random variable. Since

$$\mathrm{E}[|X|^2] = \mathrm{E}[\mathrm{Re}(X)^2 + \mathrm{Im}(X)^2],$$

$L^2_{\mathbb{C}}$ can be characterized as the vector space of complex random variables with square-integrable real and imaginary parts.

Definition 2.1.2. *The expectation of a complex random variable* X *is an operator* $\mathrm{E} : L^2_{\mathbb{C}} \to \mathbb{C}$ *defined by*

$$\mathrm{E}[X] := \mathrm{E}[\mathrm{Re}(X)] + i\,\mathrm{E}[\mathrm{Im}(X)].$$

The value of $\mathrm{E}[X]$ *is called expected value (or mean) of* X.

Note that we use E both as the symbol for the expected value operator of a real and complex random variable. Next we define the covariance of two complex random variables.

Definition 2.1.3. *Let X and Y be complex random variables. The covariance operator of X and Y, $\operatorname{Cov} : L^2_{\mathbb{C}} \times L^2_{\mathbb{C}} \to \mathbb{C}$, is defined by*

$$\operatorname{Cov}[X, Y] = \operatorname{E}\left[(X - \operatorname{E}[X])\overline{(Y - \operatorname{E}[Y])}\right].$$

The value of $\operatorname{Cov}[X, Y]$ is called covariance of X and Y (order matters here).

In the special case where $X = Y$ the covariance operator is called the *variance operator*. This leads to the following definition.

Definition 2.1.4. *Let X be a complex random variable. The variance operator of X, $\operatorname{Var} : L^2_{\mathbb{C}} \to \mathbb{R}_+$, is defined by*

$$\begin{aligned}\operatorname{Var}[X] &= \operatorname{Cov}[X, X] \\ &= \operatorname{E}\left[(X - \operatorname{E}[X])\overline{(X - \operatorname{E}[X])}\right].\end{aligned}$$

The value of $\operatorname{Var}[X]$ is called variance of X.

Theorems 2.1.5, 2.1.6 and 2.1.7 about expectation, covariance and variance are immediate consequences of the definitions above or each other.

Theorem 2.1.5. *Let $X, Y \in L^2_{\mathbb{C}}$ and $c, d \in \mathbb{C}$. Then we have*

(i) $\operatorname{E}[cX + d] = c\operatorname{E}[X] + d$;
(ii) $\operatorname{E}[X + Y] = \operatorname{E}[X] + \operatorname{E}[Y]$;
(iii) $\operatorname{E}[\overline{X}] = \overline{\operatorname{E}[X]}$.

Theorem 2.1.5 means that the expectation operator is a linear operator on $L^2_{\mathbb{C}}$.

Theorem 2.1.6. *Let $X, Y, Z \in L^2_{\mathbb{C}}$ and $c_1, c_2, d_1, d_2 \in \mathbb{C}$. Then we have*

(i) $\operatorname{Cov}[X, Y] = \operatorname{E}[X\overline{Y}] - \operatorname{E}[X]\overline{\operatorname{E}[Y]}$;
(ii) $\operatorname{Cov}[X, Y] = \overline{\operatorname{Cov}[Y, X]}$;
(iii) $\operatorname{Cov}[X, Y + Z] = \operatorname{Cov}[X, Y] + \operatorname{Cov}[X, Z]$;
(iv) $\operatorname{Cov}[c_1X + d_1, c_2Y + d_2] = c_1\overline{c_2}\operatorname{Cov}[X, Y]$.

According to 2.1.6 the covariance operator is a conjugate bilinear operator on $L^2_{\mathbb{C}} \times L^2_{\mathbb{C}}$.

Theorem 2.1.7. *Let $X, Y \in L^2_{\mathbb{C}}$ and $c, d \in \mathbb{C}$. Then we have*

(i) $\operatorname{Var}[X] = \operatorname{E}[X\overline{X}] - \operatorname{E}[X]\overline{\operatorname{E}[X]} = \operatorname{Var}[\operatorname{Re}(X)] + \operatorname{Var}[\operatorname{Im}(X)]$;
(ii) $\operatorname{Var}[cX + d] = c\overline{c}\operatorname{Var}[X]$;
(iii) $\operatorname{Var}[X + Y] = \operatorname{Var}[X] + \operatorname{Var}[Y] + 2\operatorname{Re}(\operatorname{Cov}[X, Y])$.

Later we will deal with estimators that are complex random variables. In this context the following lemma will be useful as we will derive error bounds for those estimators.

LEMMA 2.1.8. *Let $a, b \in \mathbb{R}$ with $a < b$ and let $(X_t)_{t\in(a,b)} \subset L^2_{\mathbb{C}}$ be a family of complex random variables on a probability space $(\Omega, \mathcal{A}, P)$ such that the mappings $X(\omega) : (0, \infty) \to \mathbb{C},\ t \mapsto X_t(\omega)$ are continuous for all $\omega \in \Omega$. If, in addition, the mapping*

$$X : \Omega \times (a, b) \to \mathbb{C},\ (\omega, t) \mapsto X_t(\omega)$$

is bounded, then

$$\mathrm{Var}\left[\int_a^b X_t dt\right] \leq \left[\int_a^b \sqrt{\mathrm{Var}[X_t]} dt\right]^2.$$

REMARK 2.1.9. *In this thesis, by dt we always mean integration with respect to the Lebesgue measure on $\mathbb{R}$, where complex integrands are treated by splitting them in their real and imaginary parts as we have done for the expectation operator. This means*

$$\int_a^b X_t dt := \int_a^b \mathrm{Re}(X_t)dt + i \int_a^b \mathrm{Im}(X_t)dt.$$

Note further that $X(\omega)$ is continuous and bounded for all $\omega \in \Omega$ by assumption. This implies that $X(\omega)$ is Riemann integrable for all $\omega \in \Omega$ with respect to t. By [**14**, Satz 6.1] *we may then interpret $\int_a^b X_t dt$ as a complex Riemann integral and compute it by using the Riemann sum*

$$\int_a^b X_t dt = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{b-a}{n} X_{a+(b-a)k/n}. \tag{2.1.1}$$

This representation as a sum will be useful in the following proof of Lemma 2.1.8 as it allows an estimation with the Minkowski inequality.

PROOF OF LEMMA 2.1.8. At first let $\mathrm{E}[X_t] = 0$ for all $t \in (a, b)$. Then by Fubini's theorem (which is applicable because X_t is bounded) we get

$$\begin{aligned}
\mathrm{E}\left[\int_a^b X_t dt\right] &= \mathrm{E}\left[\int_a^b \mathrm{Re}(X_t)dt\right] + i\, \mathrm{E}\left[\int_a^b \mathrm{Im}(X_t)dt\right] \\
&= \int_a^b \mathrm{E}[\mathrm{Re}(X_t)]dt + i \int_a^b \mathrm{E}[\mathrm{Im}(X_t)]dt \\
&= 0.
\end{aligned}$$

In combination with Theorem 2.1.7(i) this yields

$$\mathrm{Var}\left[\int_a^b X_t dt\right] = \mathrm{E}\left[\left|\int_a^b X_t dt\right|^2\right]. \tag{2.1.2}$$

Next we write the integral with respect to t in (2.1.2) as the Riemann sum (2.1.1) as announced in the remark before this proof. After doing so we use Fatou's lemma to obtain

$$\operatorname{Var}\left[\int_a^b X_t dt\right] = \mathrm{E}\left[\left|\lim_{n\to\infty}\sum_{k=1}^n \frac{b-a}{n} X_{a+(b-a)k/n}\right|^2\right]$$
$$\leq \lim_{n\to\infty} \mathrm{E}\left[\left|\sum_{k=1}^n \frac{b-a}{n} X_{a+(b-a)k/n}\right|^2\right].$$

Note that $\sqrt{\mathrm{E}[|\cdot|^2]}$ is a norm on $L^2_{\mathbb{C}}$. So Minkowski inequality leads to

$$\operatorname{Var}\left[\int_a^b X_t dt\right] \leq \lim_{n\to\infty}\left(\sum_{k=1}^n \frac{b-a}{n}\sqrt{\mathrm{E}\left[|X_{a+(b-a)k/n}|^2\right]}\right)^2 = \left(\int_a^b \sqrt{\operatorname{Var}[X_t]}dt\right)^2.$$

Now let $\mathrm{E}[X_t] \in \mathbb{C}$ for all $t \in (a,b)$. Then it follows from the already shown case, that

$$\operatorname{Var}\left[\int_a^b X_t - \mathrm{E}[X_t]\ dt\right] \leq \left[\int_a^b \sqrt{\operatorname{Var}[X_t - \mathrm{E}[X_t]]}dt\right]^2.$$

As a consequence, the claim is shown considering the linearity of the integral and the fact $\operatorname{Var}[Y-c] = \operatorname{Var}[Y]$ for all $Y \in \mathcal{L}^2_{\mathbb{C}}, c \in \mathbb{C}$ (see Theorem 2.1.7(ii)). □

The next lemma will reappear in Section 3.8 as we will compute the variance of a complex random variable (that is itself an integral) by Definition 2.1.4.

LEMMA 2.1.10. *Let $a, b, c, d \in \mathbb{R}$ with $a < b$ and $c < d$ and let $(X_t)_{t\in(a,b)}$, $(Y_s)_{s\in(c,d)} \subset L^2_{\mathbb{C}}$ be families of complex random variables on a probability space $(\Omega, \mathcal{A}, P)$ such that the mappings $X(\omega) : (a,b) \to \mathbb{C},\ t \mapsto X_t(\omega)$ and $Y(\omega) : (c,d) \to \mathbb{C},\ s \mapsto Y_s(\omega)$ are continuous for all $\omega \in \Omega$. If, in addition, the mappings*

$$X : \Omega \times (a,b) \to \mathbb{C},\ (\omega,t) \mapsto X_t(\omega) \quad \text{and} \quad Y : \Omega \times (c,d) \to \mathbb{C},\ (\omega,s) \mapsto Y_s(\omega)$$

are bounded, then

$$\operatorname{Cov}\left[\int_a^b X_t dt, \int_c^d Y_s ds\right] = \int_a^b \int_c^d \operatorname{Cov}\left[X_t, Y_s\right] dt ds.$$

PROOF. Denote $I(Y) := \int_c^d Y_s ds$ and consider the special case where $\mathrm{E}[X_t] = 0$ and $\mathrm{E}[Y_s] = 0$ for all $t \in (a,b)$ and $s \in (c,d)$. Firstly, extract the integral with respect to t from the covariance. This is possible by Definition 2.1.3, linearity of the expectation operator and Fubini's theorem. In fact,

$$\begin{aligned}
\operatorname{Cov}\left[\int_a^b X_t dt, \int_c^d Y_s ds\right] &= \mathrm{E}\left[\left(\int_a^b X_t dt\right)\overline{I(Y)}\right] \\
&= \mathrm{E}\left[\left(\int_a^b \operatorname{Re}(X_t)dt + i\int_a^b \operatorname{Im}(X_t)dt\right)\overline{I(Y)}\right] \\
&= \mathrm{E}\left[\int_a^b \operatorname{Re}(X_t)\overline{I(Y)}dt\right] + i\,\mathrm{E}\left[\int_a^b \operatorname{Im}(X_t)\overline{I(Y)}dt\right] \\
&= \int_a^b \mathrm{E}\left[X_t\overline{I(Y)}\right] dt \\
&= \int_a^b \operatorname{Cov}\left[X_t, \overline{I(Y)}\right] dt.
\end{aligned}$$

Secondly, the integral with respect to s is extracted from the covariance in the analogous way (taking into consideration Theorem 2.1.5(iii)).
Now turn to the general case where we may have $\mathrm{E}[X_t] \neq 0$ or $\mathrm{E}[Y_s] \neq 0$ for some $t \in (a, b)$ or $s \in (c, d)$. Here, according to the already shown special case and Theorem 2.1.6(iv), we have

$$\begin{aligned}
\operatorname{Cov}\left[\int_a^b X_t dt, \int_c^d Y_s ds\right] &= \operatorname{Cov}\left[\int_a^b X_t dt - \mathrm{E}[\int_a^b X_t dt], \int_c^d Y_s ds - \mathrm{E}[\int_c^d Y_s ds]\right] \\
&= \operatorname{Cov}\left[\int_a^b (X_t - \mathrm{E}[X_t])dt, \int_c^d (Y_s - \mathrm{E}[Y_s]ds)\right] \\
&= \int_a^b\int_c^d \operatorname{Cov}\left[X_t - \mathrm{E}[X_t], Y_s - \mathrm{E}[Y_s]\right] dtds \\
&= \int_a^b\int_c^d \operatorname{Cov}\left[X_t, Y_s\right] dtds,
\end{aligned}$$

which concludes the proof. □

2.2. Mellin Transform

In this section we introduce the Mellin transform and list some of its properties. This integral transform will be our main tool in estimation procedures of the next chapters. We roughly follow the outline in [**12**], which is our general reference for Mellin transforms.

Definition 2.2.1. *For $c \in \mathbb{R}$ define the space*

$$\mathfrak{M}_c := \left\{ f : \mathbb{R}_+ \to \mathbb{C} \,\middle|\, \int_0^\infty |f(x)|x^{c-1}dx < \infty \right\}.$$

The associated norm on $\mathfrak{M}_c$ *is given by*

$$||f||_{\mathfrak{M}_c} := \int_0^\infty |f(u)|u^{c-1}du.$$

Moreover, for $a, b \in \mathbb{R}$*,* $a < b$ *define the spaces*

$$\mathfrak{M}_{(a,b)} := \bigcap_{c\in(a,b)} \mathfrak{M}_c, \quad \mathfrak{M}_{[a,b]} := \bigcap_{c\in[a,b]} \mathfrak{M}_c.$$

If f is the density function of an $\mathbb{R}_+$-valued random variable, then we have at least $f \in \mathfrak{M}_1$. If X is an $\mathbb{R}_+$-valued random variable with density $f \in \mathfrak{M}_{(a,b)}$, then we shall also write $X \in \mathfrak{M}_{(a,b)}$.
The following lemma gives an easy to check sufficient condition for $f \in \mathfrak{M}_{(a,b)}$.

Lemma 2.2.2. *Let* $a, b \in \mathbb{R}$ *with* $a < b$*. If* $f : \mathbb{R}_+ \to \mathbb{R}$ *is locally integrable on* $\mathbb{R}_+$ *with*

$$f(x) = \begin{cases} \mathcal{O}(x^{-a}), & \text{for } x \to 0 \\ \mathcal{O}(x^{-b}), & \text{for } x \to \infty \end{cases},$$

then $f \in \mathfrak{M}_{(a,b)}$ *holds.*

Proof. Elementary calculus; see [**40**, page 203]. □

Definition 2.2.3. *For* $f \in \mathfrak{M}_{(a,b)}$*,* $s \in \mathbb{C}$ *with* $\mathrm{Re}(s) \in (a, b)$ *define*

$$\mathcal{M}[f](s) := \int_0^\infty f(x)x^{s-1}dx$$

the Mellin transform of f *in* s*.*
If f *is the density function of a random variable* X*, then we call*

$$\mathcal{M}[X](s) := \mathrm{E}[X^{s-1}] = \mathcal{M}[f](s)$$

the Mellin transform of X*.*

If $f \in \mathfrak{M}_{(a,b)}$ holds, then $\mathcal{M}[f](s)$ is well defined and holomorphic on the strip $\{s \in \mathbb{C}|\, \mathrm{Re}(s) \in (a, b)\}$ according to [**12**]. The following facts hold for the Mellin transform. See for example [**12**] again for the proof.

Theorem 2.2.4. *Let* $a, b \in \mathbb{R}$ *with* $a < b$ *and* $f, g \in \mathfrak{M}_{(a,b)}$*. Then we have*

(i) $f + g \in \mathfrak{M}_{(a,b)}$ *and*

$$\mathcal{M}[f + g](s) = \mathcal{M}[f](s) + \mathcal{M}[g](s)$$

for all $s \in \mathbb{C}$ *with* $\mathrm{Re}(s) \in (a, b)$*;*

(ii) $\lambda f \in \mathfrak{M}_{(a,b)}$ *and*

$$\mathcal{M}[\lambda f](s) = \lambda\mathcal{M}[f](s)$$

for all $\lambda \in \mathbb{C}$*,* $s \in \mathbb{C}$ *with* $\mathrm{Re}(s) \in (a, b)$*;*

(iii) *Define* $f_\lambda(x) := f(\lambda x)$ *for* $x, \lambda \geq 0$. *Then* $f_\lambda \in \mathfrak{M}_{(a,b)}$ *and*

$$\mathcal{M}[f_\lambda](s) = \lambda^{-s}\mathcal{M}[f](s)$$

for all $s \in \mathbb{C}$ *with* $\mathrm{Re}(s) \in (a, b)$*;*

(iv) *Define* $f_\lambda(x) := x^\lambda f(x)$ *for* $\lambda \in \mathbb{C}$, $x \geq 0$. *Then* $f_\lambda \in \mathfrak{M}_{(a-\mathrm{Re}(\lambda),b-\mathrm{Re}(\lambda))}$ *and*

$$\mathcal{M}[f_\lambda](s) = \mathcal{M}[f](s + \lambda)$$

for all $s \in \mathbb{C}$ *with* $\mathrm{Re}(s) \in (a - \mathrm{Re}(\lambda), b - \mathrm{Re}(\lambda))$*;*

(v) *Define* $f_\lambda(x) := f(x^\lambda)$ *for* $\lambda, x \geq 0$. *Then* $f_\lambda \in \mathfrak{M}_{(\lambda a,\lambda b)}$ *and*

$$\mathcal{M}[f_\lambda](s) = \lambda^{-1}\mathcal{M}[f](\lambda^{-1}s)$$

for all $s \in \mathbb{C}$ *with* $\mathrm{Re}(s) \in (\lambda a, \lambda b)$*;*

(vi) *Define* $f_\lambda(x) := f(x^{-\lambda})$ *for* $\lambda, x \geq 0$. *Then* $f_\lambda \in \mathfrak{M}_{(-\lambda b,-\lambda a)}$ *and*

$$\mathcal{M}[f_\lambda](s) = \lambda^{-1}\mathcal{M}[f](-\lambda^{-1}s)$$

for all $s \in \mathbb{C}$ *with* $\mathrm{Re}(s) \in (-\lambda a, -\lambda b)$.

Let us now look at some examples of Mellin transforms.

Example 2.2.5. (i) *Consider* $f : \mathbb{R}_+ \to \mathbb{R}$, $f(x) = e^{-x}$. *The Mellin transform* $\mathcal{M}[f]$ *of* f *is given by the Gamma function:*

$$\Gamma(s) = \int_0^\infty e^{-x}x^{s-1}dx \quad \text{for } s \in \mathbb{C} \text{ with } \mathrm{Re}(s) > 0.$$

See Section 2.4 for further discussion of the Gamma function.

(ii) *Consider gamma densities*

$$f(x) = \frac{r^\sigma}{\Gamma(\sigma)}x^{\sigma-1}e^{-rx}$$

for $x, \sigma, r > 0$. *We have*

$$\mathcal{M}[f](s) = \frac{r^{1-s}}{\Gamma(\sigma)}\Gamma(s + \sigma - 1).$$

for all $s \in \mathbb{C}$ *with* $\mathrm{Re}(s - \sigma + 1) > 0$.

Proof. The change of variables $y = rx$ gives

$$\begin{aligned}\mathcal{M}[f](s) &= \int_0^\infty x^{s-1}\frac{r^\sigma}{\Gamma(\sigma)}x^{\sigma-1}e^{-rx}dx \\ &= \frac{r^{-s+1}}{\Gamma(\sigma)}\int_0^\infty y^{s+\sigma-2}e^{-y}dy \\ &= \frac{r^{1-s}}{\Gamma(\sigma)}\Gamma(s + \sigma - 1)\end{aligned}$$

for all $s \in \mathbb{C}$ with $\mathrm{Re}(s - \sigma + 1) > 0$. □

(iii) *Consider for $t > 0$, $d \geq 1$ the densities*

$$f_t(x) = \frac{2^{1-\frac{d}{2}} t^{-\frac{d}{2}}}{\Gamma(d/2)} x^{d-1} e^{-\frac{x^2}{2t}}, \quad x > 0.$$

*The associated distributions will reappear in the next chapter as marginal densities of a Bessel process. For $t = d^{-1}$ [**23**] refers to these distributions as the d-dimensional Rayleigh distributions. For $\mathrm{Re}(s) > 1 - d$ we have*

$$\mathcal{M}[f_1](s) = \frac{1}{\Gamma(d/2)} \Gamma\left(\frac{s+d-1}{2}\right) 2^{\frac{s-1}{2}}.$$

PROOF. This follows easily from Theorem 2.2.4. In fact,

$$\begin{aligned}
\mathcal{M}[f_1](s) &= \mathcal{M}\left[\frac{1}{\Gamma(d/2)} 2^{1-\frac{d}{2}} x^{d-1} e^{-\frac{x^2}{2}}\right](s) \\
&= \frac{1}{\Gamma(d/2)} 2^{1-\frac{d}{2}} \mathcal{M}\left[e^{-\frac{x^2}{2}}\right](s+d-1) \\
&= \frac{1}{2\Gamma(d/2)} 2^{1-\frac{d}{2}} \mathcal{M}\left[e^{-\frac{x}{2}}\right]\left(\frac{s+d-1}{2}\right) \\
&= \frac{1}{2\Gamma(d/2)} 2^{1-\frac{d}{2}} \left(\frac{1}{2}\right)^{-\frac{s+d-1}{2}} \mathcal{M}\left[e^{-x}\right]\left(\frac{s+d-1}{2}\right) \\
&= \frac{1}{\Gamma(d/2)} \Gamma\left(\frac{s+d-1}{2}\right) 2^{\frac{s-1}{2}}
\end{aligned}$$

for $\mathrm{Re}(s) > 1 - d$. □

REMARK 2.2.6. *There is a close connection between the Mellin transform and the Fourier transform*

$$\mathcal{F}[f](u) := \int_{-\infty}^{\infty} f(y) e^{iuy} dy, \quad u \in \mathbb{R}$$

of an integrable function $f : \mathbb{R} \to \mathbb{C}$. Setting $s = \sigma + it$ and applying the change of variables $x := e^y$ yield

$$\begin{aligned}
\mathcal{M}[f](s) &= \int_{-\infty}^{\infty} f(e^y) e^{\sigma y} e^{ity} dy = \mathcal{F}[g](t), \\
\mathcal{F}[f](u) &= \int_0^{\infty} f(\log(x)) x^{iu-1} dx = \mathcal{M}[h](iu),
\end{aligned}$$

where $g(x) = e^{\sigma x} f(e^x)$ for $x \in \mathbb{R}$ and $h(x) = f(\log(x))$ for $x > 0$. Hence, the Mellin transform allows a representation as a Fourier transform, and vice versa.

Similar to the multiplicative behavior of the classical Fourier transform with respect to sums of independent random variables, the Mellin transform behaves multiplicatively with respect to products of independent random variables. More precisely we have:

THEOREM 2.2.7. *Let X and Y be independent $\mathbb{R}_+$-valued random variables with densities $f_X \in \mathfrak{M}_{(a,b)}$ and $f_Y \in \mathfrak{M}_{(c,d)}$, and Mellin transforms $\mathcal{M}[X]$ and $\mathcal{M}[Y]$ for $a < b,\ c < d,\ (a,b) \cap (c,d) \neq \emptyset$.*
Then the density f_{XY} of XY satisfies $f_{XY} \in \mathfrak{M}_{(a,b)\cap(c,d)}$, and

$$\mathcal{M}[XY](s) = \mathcal{M}[X](s)\mathcal{M}[Y](s)$$

for all $s \in \mathbb{C}$ with $\mathrm{Re}(s) \in (a,b) \cap (c,d)$.

PROOF. This follows from independence of X and Y, and the definition of Mellin transform. □

In the setting of Theorem 2.2.7 it is easy to see that f_{XY} is identical to

$$(f_X \odot f_Y)(s) := \int_0^\infty f_X\left(\frac{s}{x}\right) f_Y(x) \frac{1}{x} dx \tag{2.2.1}$$

for all $s \in \mathbb{C}$ with $\mathrm{Re}(s) \in (a,b) \cap (c,d)$. The function $f_X \odot f_Y$ is called *Mellin convolution* of f_X and f_Y. Theorem 2.2.7 implies

$$\mathcal{M}[f_X \odot f_Y](s) = \mathcal{M}[f_X](s)\mathcal{M}[f_Y](s)$$

for all $s \in \mathbb{C}$ with $\mathrm{Re}(s) \in (a,b) \cap (c,d)$.
For $a,b \in \mathbb{R}$ with $a < b$ denote by $\mathcal{H}(a,b)$ the space of holomorphic functions on $\{s \in \mathbb{C} | \mathrm{Re}(s) \in (a,b)\}$. The mapping $\mathcal{M} : \mathfrak{M}_{(a,b)} \to \mathcal{H}(a,b),\ f \mapsto \mathcal{M}[f]$ is injective (see [**12**, Corollary 5]). Given the Mellin transform of a function f we can reconstruct f via the following theorem.

THEOREM 2.2.8. *For $a,b \in \mathbb{R}$ with $a < b$ let $f \in \mathfrak{M}_{(a,b)}$. If*

$$\int_{-\infty}^\infty |\mathcal{M}[f](\gamma + iv)| dv < \infty$$

for all $\gamma \in (a,b)$, then the inversion formula

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{M}[f](\gamma + iv) x^{-\gamma - iv} dv$$

holds almost everywhere for $x \in \mathbb{R}_+$ and all $\gamma \in (a,b)$. If f is continuous on $\mathbb{R}_+$, then this equation is true for all $x \in \mathbb{R}_+$.

PROOF. See [**12**]. □

Another important result in the theory of Mellin transforms is the *Parseval formula for Mellin tranforms.* It goes as follows:

THEOREM 2.2.9. *Suppose that $f : \mathbb{R}_+ \to \mathbb{R}$ and $g : \mathbb{R}_+ \to \mathbb{R}$ are measurable functions such that*

$$\int_0^\infty f(x) g(x) dx$$

exists. Suppose further that $\mathcal{M}[f](1-\cdot)$ and $\mathcal{M}[g](\cdot)$ are holomorphic on some vertical strip $\mathcal{S} := \{z \in \mathbb{C} | a < \mathrm{Re}(z) < b\}$ for $a, b \in \mathbb{R}$. If there is a $\gamma \in (a, b)$ such that

$$\int_{-\infty}^{\infty} |\mathcal{M}[f](1-\gamma-is)|ds < \infty \quad \text{and} \quad \int_{0}^{\infty} x^{\gamma-1}|g(x)|dx < \infty,$$

then this γ satisfies

$$\int_0^\infty f(x)g(x)dx = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \mathcal{M}[f](1-s)\mathcal{M}[g](s)ds.$$

Proof. See [**9**, page 108]. □

Let us now discuss the Mellin transform from the statistical point of view. Given i.i.d. samples of a random variable, its Mellin transform can be approximated by its so-called *empirical Mellin transform.*

Definition 2.2.10. *For $a, b \in \mathbb{R}$ with $a < b$ let $Y \in \mathfrak{M}_{(a,b)}$ and $X_1, ..., X_n$ be i.i.d. samples of Y. We call*

$$\mathcal{M}_n[Y](s) := \frac{1}{n}\sum_{k=1}^{n} X_k^{s-1},$$

where $a < \mathrm{Re}(s) < b$, the empirical Mellin transform of Y.

We have the following fact about the complex conjugate of an (empirical) Mellin transform.

Lemma 2.2.11. *For a, b with $a < b$ let $Y \in \mathfrak{M}_{(a,b)}$ a random variable and $X_1, ..., X_n$ be i.i.d. samples of Y. We have*

(i) $\overline{\mathcal{M}[Y](s)} = \mathcal{M}[Y](\bar{s})$ *and*
(ii) $\overline{\mathcal{M}_n[Y](s)} = \mathcal{M}_n[Y](\bar{s})$

for all $s \in \mathbb{C}$ with $a < \mathrm{Re}(s) < b$.

Proof. Claim (i) follows immediately from Definition 2.2.3 and Theorem 2.1.5(iii). Claim (ii) is obvious with Definition 2.2.10. □

The facts in Lemma 2.2.12 about the empirical Mellin transform follow from linearity of the expected value and calculation rules for the variance (Theorems 2.1.5 and 2.1.7).

Lemma 2.2.12. *For $a, b \in \mathbb{R}$ with $a < b$ let $Y \in \mathfrak{M}_{(a,b)}$ and $X_1, ..., X_n$ be i.i.d. samples of Y. Then the empirical Mellin transform $\mathcal{M}_n[Y]$ of Y satisfies*

(i) $\mathrm{E}\left[\mathcal{M}_n[Y](s)\right] = \mathcal{M}[Y](s)$ *and*
(ii) $\mathrm{Var}\left[\mathcal{M}_n[Y](s)\right] = \frac{1}{n}\mathrm{Var}\left[Y^{s-1}\right]$

for all $n \in \mathbb{N}$ and all $s \in \mathbb{C}$ with $a < \mathrm{Re}(s) < b$.

Lemma 2.2.12 shows that $\mathcal{M}_n[Y](s)$ is an unbiased estimator for $\mathcal{M}[Y](s)$ and its variance converges to 0 for $n \to \infty$, if $\mathrm{Var}\left[Y^{s-1}\right] < \infty$ is assumed. So $\mathcal{M}_n[Y](s)$ is also weakly consistent for $\mathcal{M}[Y](s)$ in this case.

2.3. Notation for Asymptotic Analysis

Let f and g be either sequences of real numbers or complex-valued functions of real numbers, and $a \in \mathbb{R} \cup \{\infty\}$. For the sake of brevity we introduce the notation

$$f(x) \lesssim g(x) \text{ for } x \to a, \tag{2.3.1}$$

if $f = \mathcal{O}(g)$ in the Landau notation, that is

$$\limsup_{x \to a} \frac{|f(x)|}{|g(x)|} < \infty. \tag{2.3.2}$$

If f and g are functions of real numbers and $a \in \mathbb{R}$, then (2.3.2) is equivalent to the statement

$$\exists C > 0 \ \exists \varepsilon > 0 \text{ such that } \forall x \in \mathbb{R} \text{ with } |x - a| < \varepsilon \ : \ |f(x)| \leq C|g(x)|.$$

And if $a = \infty$, then (2.3.2) is equivalent to

$$\exists C > 0 \ \exists x_0 \in \mathbb{R} \text{ such that } \forall x > x_0 \ : \ |f(x)| \leq C|g(x)|.$$

We write

$$f(x) \sim g(x) \text{ for } x \to a,$$

if $f(x) \lesssim g(x)$ and $g(x) \lesssim f(x)$ for $x \to a$.
Using the notation from (2.3.1) allows us to omit constants and lower order terms that are often irrelevant when we investigate the asymptotic behavior of a function.

2.4. Gamma Function

As we have seen from Example 2.2.5 the Gamma function often appears in the Mellin transform of a function that has an exponential expression in it. Thus, it deserves closer attention. We will see in Lemma 3.5.2 from Section 3.3 that the asymptotic behavior of a Mellin transform determines the quality of the estimators we consider there. Hence, the asymptotic behavior of the Gamma function is of a special interest to us. We refer the reader to [**3**] for a thorough study on this subject. Here we only state some facts that are relevant to us.
We define the Gamma function by

$$\Gamma(s) := \int_0^\infty t^{s-1} e^{-t} dt \tag{2.4.1}$$

for all $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 0$. According to the remark below Definition 2.2.3, as the Mellin transform of the exponential function, Gamma function is holomorphic for all $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 0$.
With Lemmas 2.4.1 and 2.4.2 we present versions of the well-known Stirling formula for the Gamma function of real arguments,

$$\Gamma(x) \sim \sqrt{2\pi} x^{x-1/2} e^{-x} \text{ for } x \in \mathbb{R}, \ x \to \infty.$$

Lemma 2.4.1 describes the behavior of the Gamma function when the absolute value of its complex argument is large.

LEMMA 2.4.1. *For $\delta > 0$ and $|\arg(s)| \leq \pi - \delta$,*

$$\Gamma(s) \sim \sqrt{2\pi} s^{s-1/2} e^{-s}$$

for $|s| \to \infty$. Complex power is defined by $s^{s-1/2} := e^{(s-1/2)\log(s)}$ and the value of $\log(s)$ is the branch with $\log(s)$ real when s is real and positive.

PROOF. See [**3**, Corollary 1.4.3]. □

Lemma 2.4.2 describes the behavior of the Gamma function when the real part of its argument is fixed and the imaginary part is large.

LEMMA 2.4.2. *Let $a, b \in \mathbb{R}$ with $0 \leq a < b$. For all $\alpha \in [a, b]$ we have*

$$|\Gamma(\alpha + i\beta)| = \sqrt{2\pi}|\beta|^{\alpha-1/2}e^{-|\beta|\pi/2}(1 + \mathcal{O}(1/|\beta|))$$

for $|\beta| \to \infty$. The constant implied by $\mathcal{O}$ depends only on a and b.

PROOF. See [**3**, Corollary 1.4.4]. □

The following lemma is an immediate consequence of Lemma 2.4.2.

LEMMA 2.4.3. *For all $\alpha \in \mathbb{R}$ there are $C_1, C_2 \geq 0$ such that*

$$C_1|\beta|^{\alpha-1/2}e^{-|\beta|\pi/2} \leq |\Gamma(\alpha + i\beta)| \leq C_2|\beta|^{\alpha-1/2}e^{-|\beta|\pi/2}$$

for all $|\beta| \geq 2$.

From Lemma 2.4.3 we can derive the behavior of integrals of the Gamma function with respect to its imaginary part. In fact, we have

LEMMA 2.4.4. *(i) For all $\alpha \in (0, 1/2), \delta > 0$ and $U > 2$ there is a $C(\alpha, \delta) > 0$ such that*

$$\int_{-U}^{U} \frac{1}{|\Gamma(\alpha + iv)|^\delta} dv \leq C(\alpha, \delta) U^{(1/2-\alpha)\delta} e^{U\pi\delta/2}$$

(ii) For all $\alpha \geq 1/2$, $\delta > 0$ and $U > 2$ there are $C_1(\alpha, \delta)$ and $C_2(\alpha, \delta) > 0$ such that

$$\int_{-U}^{U} \frac{1}{|\Gamma(\alpha + iv)|^\delta} dv \leq C_1(\alpha, \delta) + C_2(\alpha, \delta) e^{U\pi\delta/2}$$

PROOF. (i) By Lemma 2.4.3 there is a $C_\alpha > 0$ such that for all $|v| \geq 2$ we have

$$C_\alpha |v|^{(\alpha-1/2)\delta} e^{-|v|\pi\delta/2} \leq |\Gamma(\alpha + iv)|^\delta.$$

Define $C_1 := \int_{-2}^{2} \frac{1}{|\Gamma(\alpha+iv)|^\delta} dv$ (which is finite, because $\Gamma(\alpha+iv)$ is nonzero and continuous in v). Hence,

$$\begin{aligned}\int_{-U}^{U} \frac{1}{|\Gamma(\alpha+iv)|^\delta} dv &= C_1 + \int_{\{2<|v|<U\}} \frac{1}{|\Gamma(\alpha+iv)|^\delta} dv \\ &\le C_1 + C_\alpha \int_{\{2<|v|<U\}} |v|^{(1/2-\alpha)\delta} e^{|v|\pi\delta/2} dv \\ &\le C_1 + 2C_\alpha U^{(1/2-\alpha)\delta} \int_2^U e^{v\pi\delta/2} dv \\ &= C_1 + 4C_\alpha(\pi\delta)^{-1} U^{(1/2-\alpha)\delta}(e^{U\pi\delta/2} - e^{\pi\delta}) \\ &\le C(\alpha,\delta) U^{(1/2-\alpha)\delta} e^{U\pi\delta/2}\end{aligned}$$

for $C(\alpha,\delta) := \max\{2C_1, 8C_\alpha(\pi\delta)^{-1}\}$.

(ii) By Lemma 2.4.3 there is a $C_\alpha > 0$ such that for all $|v| \ge 2$ we have

$$C_\alpha |v|^{(\alpha-1/2)\delta} e^{-|v|\pi\delta/2} \le |\Gamma(\alpha+iv)|^\delta.$$

Define $C_1(\alpha,\delta) := \int_{-2}^{2} \frac{1}{|\Gamma(\alpha+iv)|^\delta} dv$ (which is finite). Hence,

$$\begin{aligned}\int_{-U}^{U} \frac{1}{|\Gamma(\alpha+iv)|^\delta} dv &= C_1(\alpha,\delta) + \int_{\{2<|v|<U\}} \frac{1}{|\Gamma(\alpha+iv)|^\delta} dv \\ &\le C_1(\alpha,\delta) + C_\alpha \int_{\{2<|v|<U\}} |v|^{(1/2-\alpha)\delta} e^{|v|\pi\delta/2} dv \\ &\le C_1(\alpha,\delta) + 2C_\alpha \int_2^U e^{v\pi\delta/2} dv \\ &= C_1(\alpha,\delta) + 4C_\alpha(\pi\delta)^{-1}(e^{U\pi\delta/2} - e^{\pi\delta}) \\ &\le C_1(\alpha,\delta) + C_2(\alpha,\delta) e^{U\pi\delta/2}\end{aligned}$$

with $C_2(\alpha,\delta) := 4C_\alpha(\pi\delta)^{-1}$. □

CHAPTER 3

Estimation for Self-Similar Processes

Self-similar processes are stochastic processes that are invariant in distribution under suitable scaling of time and space. These processes also arise naturally in the analysis of random phenomena (in time) exhibiting certain forms of long-range dependence (see [**15**]). Applications are found in statistical physics (see [**33**]). See [**24**] for a study of general self-similar processes from a probabilistic point of view. An important example of such processes is the fractional Brownian motion which is a Gaussian self-similar process. See for instance [**8**] for a thorough study of these processes. Non-Gaussian self-similar processes include Bessel processes (see Section 3.7, one-dimensional Dunkl processes (see [**13**]), the Rosenblatt process, $S\alpha S$ Lévy motion, linear fractional stable motion, log-fractional stable motion and the Telecom prosess (see [**29**] and [**36**]).

DEFINITION 3.0.1. *A real-valued stochastic process* $(Y_t)_{t\geq 0}$ *is called self-similar with scaling parameter* H*, if*

$$(Y_{at})_{t\geq 0} \stackrel{d}{=} (a^H Y_t)_{t\geq 0} \quad \text{for all } a > 0. \tag{3.0.1}$$

Here, $\stackrel{d}{=}$ *denotes identity of all finite dimensional distributions.*

The following properties of self-similar processes are well-known.

THEOREM 3.0.2. *Let* $(Y_t)_{t\geq 0}$ *be self-similar with scaling parameter* H*. Then we have*

(i) $Y_0 = 0$ *almost surely;*
(ii) H *is unique. Moreover, if* $(Y_t)_{t\geq 0}$ *is a stochastic process such that for all* $a > 0$ *there is* $b_a > 0$ *with* $(Y_{at})_{t\geq 0} \stackrel{d}{=} (b_a Y_t)_{t\geq 0}$*, then necessarily* $b_a = a^H$ *for a unique* $H > 0$*;*
(iii) For all $a > 0, t \geq 0$*,*

$$Y_{at} \stackrel{d}{=} a^H Y_t. \tag{3.0.2}$$

Here, $\stackrel{d}{=}$ *denotes identity in law of two random variables.*

See for example [**24**] or [**15**] for proofs of these facts.

3.1. The Setting

Let $Y = (Y_t)_{t\geq 0}$ be a known self-similar process with a scaling parameter H and càdlàg paths. Let $T \geq 0$ be a stopping time with density f_T independent

of Y. When we talk about stopping times in this thesis, we refer to a random variable. The name "stopping time" is motivated by our interpretation of T as the time where Y "stops", it is not related to any particular filtration. The aim of this chapter is to estimate f_T non-parametrically based on i.i.d. samples $X_1, \dots, X_n$ of Y_T.
This was done by Belomestny and Schoenmakers in [**6**] for the case where Y is a one-dimensional Brownian motion. We modify their approach in order to present a consistent estimate for the case where Y is a general self-similar process.
Before we begin to tackle this problem, let us specify our setting with a remark about the underlying probability space. It will remain valid throughout the thesis, not only for the self-similar case.

REMARK 3.1.1. *Let Y and T be realized on a canonical product space, meaning that $Y_t : \Omega_1 \to \mathbb{R}$ for $t \in \mathbb{R}_+$ and $T : \Omega_2 \to \mathbb{R}_+$ are random variables living on probability spaces $(\Omega_1, \mathcal{A}_1, P_1)$ and $(\Omega_2, \mathcal{A}_2, P_2)$, respectively. Then Y_t and T can also be interpreted as variables on the product of the two probability spaces, denoted by $(\Omega, \mathcal{A}, P) := (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P_1 \otimes P_2)$, via $(\omega_1, \omega_2) \mapsto Y_t(\omega_1)$ and $(\omega_1, \omega_2) \mapsto T(\omega_2)$, respectively. The mapping $Y_T : \Omega \to \mathbb{R}$, $(\omega_1, \omega_2) \mapsto Y_{T(\omega_2)}(\omega_1)$ is measurable, because Y has càdlàg paths (cf.* [**5**, Lemma 49.11]*).*

The estimation procedure is based on the fact that the property (3.0.2) still holds for a self-similar process, if we replace numbers $t \geq 0$ with stopping times $T \geq 0$ independent of Y. In fact:

LEMMA 3.1.2. *Let $Y = (Y_t)_{t \geq 0}$ be a real valued self-similar process with scaling parameter H and càdlàg paths. If $T \geq 0$ is a stopping time independent of Y, then*

$$T^H Y_1 \stackrel{d}{=} Y_T.$$

PROOF. We show the identity in law by comparing the Fourier transforms of the two variables. Using Fubini's theorem and self-similarity of Y we get

$$\begin{aligned} \mathrm{E}[e^{isY_T}] &= \int \left(\int e^{isY_{T(\omega_2)}(\omega_1)} dP_1(\omega_1) \right) dP_2(\omega_2) \\ &= \int \left(\int e^{isT(\omega_2)^H Y_1(\omega_1)} dP_1(\omega_1) \right) dP_2(\omega_2) \\ &= \mathrm{E}[e^{isT^H Y_1}] \end{aligned}$$

for all $s \in \mathbb{R}$ and thereby the claim. $\square$

The density of the random variable Y_T can be calculated via the following lemma.

LEMMA 3.1.3. *Let $(Y_t)_{t \geq 0}$ be a real-valued random process with càdlàg paths and marginal densities g_t, $t \geq 0$. Suppose that $t \mapsto g_t(x)$ is continuous for all*

$x \in \mathbb{R}$. *If* $T \geq 0$ *is a stopping time with density* f_T *and independent of* $(Y_t)_{t\geq 0}$, *then the density of* Y_T *is given by*

$$p_T(y) = \int_0^\infty g_z(y) f_T(z) dz, \quad y \in \mathbb{R}.$$

PROOF. Using the notation introduced in Remark 3.1.1 and Fubini's theorem we can write

$$\begin{aligned} P(Y_T \in A) &= \int \mathbb{1}_A(\omega) dP_{Y_T}(\omega) \\ &= \iint \mathbb{1}_A(Y_{T(\omega_2)}(\omega_1)) dP_1(\omega_1) dP_2(\omega_2) \\ &= \int_0^\infty \int \mathbb{1}_A(y) g_{T(\omega_2)}(y) dy dP_2(\omega_2) \\ &= \int_A \int_0^\infty g_z(y) f_T(z) dz dy \end{aligned}$$

for any Borel set A, which implies the claim. □

3.2. Construction of an Estimator for Self-Similar Processes

Given the setting presented in Section 3.1 we can now proceed to construct an estimator for the density of a stopping time $T \geq 0$ based on samples $X_1, ..., X_n$ of Y_T, where $Y = (Y_t)_{t\geq 0}$ is a known self-similar process with scaling parameter $H > 0$ and càdlàg paths.
Due to Lemma 3.1.2 we have

$$T^H Y_1 \overset{d}{=} Y_T. \tag{3.2.1}$$

We take the absolute value on both sides and assume that $T \in \mathfrak{M}_{(a,b)}$ with $0 \leq a < b$ and $|Y_1| \in \mathfrak{M}_{(0,\infty)}$, so we can apply the Mellin transform on both sides of (3.2.1). These assumptions yield

$$\begin{aligned} \mathcal{M}[|Y_T|](s) &= \mathcal{M}[T^H](s)\mathcal{M}[|Y_1|](s) \\ &= \mathcal{M}[T]\,(Hs - H + 1)\,\mathcal{M}[|Y_1|](s) \end{aligned}$$

for $\max\{0, \frac{a+H-1}{H}\} < \mathrm{Re}(s) < \frac{b+H-1}{H}$. Setting $z := Hs - H + 1$ we conclude that for $\max\{1 - H, a\} < \mathrm{Re}(z) < b$ we have

$$\mathcal{M}[T](z) = \frac{\mathcal{M}[|Y_T|](\frac{z+H-1}{H})}{\mathcal{M}[|Y_1|](\frac{z+H-1}{H})}. \tag{3.2.2}$$

Provided that the Mellin inversion formula (Lemma 2.2.8) is applicable to T, we may write

$$f_T(x) = \frac{1}{2\pi i} \int\limits_{\gamma - i\infty}^{\gamma + i\infty} \mathcal{M}[T](z) x^{-z} dz = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \mathcal{M}[T](\gamma + iv) x^{-\gamma - iv} dv \tag{3.2.3}$$

for $a < \gamma < b$. Combining (3.2.2) and (3.2.3) we obtain the representation

$$f_T(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\mathcal{M}[|Y_T|](\frac{\gamma+H-1+iv}{H})}{\mathcal{M}[|Y_1|](\frac{\gamma+H-1+iv}{H})} x^{-\gamma-iv} dv \tag{3.2.4}$$

for $\max\{1-H, a\} < \gamma < b$. In order to obtain an estimator of f_T based on (3.2.4) we would like to substitute $\mathcal{M}[|Y_T|]$ by its empirical counterpart

$$\mathcal{M}_n[|Y_T|](s) = \frac{1}{n} \sum_{k=1}^{n} |X_k|^{s-1}, \tag{3.2.5}$$

where $a < \mathrm{Re}(s) < b$ and $X_1, ..., X_n$ are i.i.d. samples of Y_T (see Definition 2.2.10). However, this substitution may prevent the integral in (3.2.4) from converging. Thus, we introduce a sequence $(h_n)_{n\in\mathbb{N}}$ with $h_n \in \mathbb{R}_+$ for all $n \in \mathbb{N}$ and $h_n \to 0$ in order to regularize our estimator. In view of (3.2.4) and (3.2.5) define

$$\hat{f}_n(x) := \frac{1}{2\pi} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\mathcal{M}_n[|Y_T|](\frac{\gamma+H-1+iv}{H})}{\mathcal{M}[|Y_1|](\frac{\gamma+H-1+iv}{H})} x^{-\gamma-iv} \, dv \tag{3.2.6}$$

for $x > 0$ and $\max\{1-H, a\} < \gamma < b$ as an estimator for f_T. In contrast to (3.2.4), the definition in (3.2.6) depends on γ. That is, for each admissible γ we obtain a different estimator. The appropriate choice of γ will become apparent at the end of Section 3.7.

The right hand side of (3.2.6) can be rewritten as

$$\hat{f}_n(x) := \frac{1}{\pi} \int K(h_n v) \frac{\mathcal{M}_n[|Y_T|](\frac{\gamma+H-1+iv}{H})}{\mathcal{M}[|Y_1|](\frac{\gamma+H-1+iv}{H})} x^{-\gamma-iv} \, dv,$$

where $K(v) := \frac{1}{2}\mathbb{1}_{[-1,1]}(v)$ is called *rectangular kernel* and h_n is called *bandwidth* of $\hat{f}_n$. A simple generalization of (3.2.6) can be obtained by considering a general kernel, that is any integrable function $K : \mathbb{R} \to \mathbb{R}$ satisfying $\int K(u)du = 1$. Some classical examples of kernels are given in [**37**, page 3]. In this thesis we will only consider the rectangular kernel, because it already leads to a rate-optimal estimator (see Chapter 6) among other properties discussed below.

Remark 3.2.1. *Despite the presence of a complex integral in* (3.2.6) $\hat{f}_n(x)$ *can only assume real values. That is because the integration interval is symmetrical and so Lemma 2.2.11 implies*

$$\overline{\hat{f}_n(x)} = \frac{1}{2\pi} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\mathcal{M}_n[|Y_T|](\frac{\gamma+H-1-iv}{H})}{\mathcal{M}[|Y_1|](\frac{\gamma+H-1-iv}{H})} x^{-\gamma+iv} \, dv = \hat{f}_n(x).$$

Consistency and other desirable properties of the estimator (3.2.6) are not obvious and will be the subject of our study for the following sections in this chapter.

3.3. Convergence Analysis

According to [**37**, page 4] the *mean squared risk* is a basic measure of the accuracy of an estimator $\hat{f}_n(x)$ for a function $f_T : \mathbb{R}_+ \to \mathbb{R}$. It (the risk) is defined for a fixed point $x \in \mathbb{R}$ by

$$\text{MSE} := \text{MSE}(x) := \text{E}[|\hat{f}_n(x) - f_T(x)|^2]. \tag{3.3.1}$$

For estimators considered in this thesis we will obtain error bounds of the form

$$\text{MSE}(x) \lesssim x^{-2\gamma}\psi_n \tag{3.3.2}$$

for some sequences $(\psi_n)_{n\in\mathbb{N}}$ with $\psi_n \to 0$ for $n \to \infty$ and where $\gamma \in \mathbb{R}_+$ is a constant appearing in (3.2.6) and may be chosen freely from some interval depending on the setting. We define the *(weighted) mean squared risk* via

$$\text{MSE}_\gamma := \text{MSE}_\gamma(x) := x^{2\gamma}\,\text{MSE}(x). \tag{3.3.3}$$

Using this quantity (3.3.2) is equivalent to

$$\text{MSE}_\gamma(x) \lesssim \psi_n.$$

Upper bounds on $\text{MSE}_\gamma(x)$ that we may find do not provide upper bounds on $\text{MSE}(x)$ for $x \in (0,1)$. Nevertheless, since $\text{MSE}(x) \leq \text{MSE}_\gamma(x)$ for $x \geq 1$, an upper bound on MSE_γ is also an upper bound on MSE in this case. In fact, simulations in Section 3.9 indicate that our estimation method performs poorly for small x. This issue could be a subject for further studies, but we will not address it further in this thesis.

DEFINITION 3.3.1. *We call an estimator $\hat{f}_n$ of a function $f : \mathbb{R}_+ \to \mathbb{R}$ consistent, if its* $\text{MSE}(x)$ *converges to* 0 *for $n \to \infty$ and all $x > 0$.*

The mean squared error of an estimator $\hat{f}_n$ can be decomposed in the following way. By Theorem 2.1.7(i),

$$\text{MSE}_\gamma = |b_\gamma(x)|^2 + \sigma_\gamma^2(x) \tag{3.3.4}$$

for all $x \in \mathbb{R}_+$, where

$$b_\gamma(x) := x^\gamma(f_T(x) - \text{E}[\hat{f}_n(x)]),$$

which we call *(weighted) bias* of $\hat{f}_n$ and

$$\sigma_\gamma^2(x) := x^{2\gamma}\,\text{Var}[\hat{f}_n(x)],$$

which we call *(weighted) variance* of $\hat{f}_n$. During a convergence analysis it is often useful to consider the bias and the variance of an estimator separately.

3.4. Classes $\mathcal{C}$ and $\mathcal{D}$

In non-parametric estimation it is usual to assume that f_T belongs to some "massive" class of densities (see [**37**, page 1]), for example the class of all continuous functions or all square integrable functions. Here, we consider the following two classes suggested in [**6**].

DEFINITION 3.4.1. *For $\beta, L > 0$ and $\gamma \in \mathbb{R}$ with $a < \gamma < b$ define classes of functions*

(i) $\mathcal{C}(\beta, L, \gamma) := \left\{ f \in \mathfrak{M}_\gamma \middle| \int e^{\beta|v|} \left|\mathcal{M}[f](\gamma + iv)\right| dv \le L \right\}$ *and*

(ii) $\mathcal{D}(\beta, L, \gamma) := \left\{ f \in \mathfrak{M}_\gamma \middle| \int (1 + |v|^\beta) \left|\mathcal{M}[f](\gamma + iv)\right| dv \le L \right\}$.

For a random variable X with density $f \in \mathcal{C}(\beta, L, \gamma)$ or $f \in \mathcal{D}(\beta, L, \gamma)$ we will also write $X \in \mathcal{C}(\beta, L, \gamma)$ or $X \in \mathcal{D}(\beta, L, \gamma)$, respectively.

REMARK 3.4.2. *Since $x + 1 \le e^x$ for all $x \in \mathbb{R}$, we have $\mathcal{C}(\beta, L, \gamma) \subseteq \mathcal{D}(\beta, L, \gamma)$ for all $\beta, L > 0$ and $\gamma \in \mathbb{R}$.*

It turns out that classes $\mathcal{C}(\beta, L, \gamma)$ and $\mathcal{D}(\beta, L, \gamma)$ are quite large and contain functions with different degrees of smoothness (see [**7**]). We will make this statement precise in Corollary 3.4.6. But first let us describe the connection between the degree of smoothness of a function and the rate of decay of its Mellin transform.

Recall the notation

$$\mathfrak{M}_{(a,b)} := \left\{ f : \mathbb{R}_+ \to \mathbb{R} \middle| \int_0^\infty |f(t)t^{z-1}| dt < \infty \text{ for all } \operatorname{Re}(z) \in (a, b) \right\}$$

from Definition 2.2.1.

LEMMA 3.4.3. *Let $a, b \in \mathbb{R}$ with $a < b$.*

(i) If $f \in \mathfrak{M}_{(a,b)}$ is locally integrable, then

$$\mathcal{M}[f](\gamma + iv) = o(1) \quad \text{for } v \to \pm\infty$$

uniformly with respect to γ in every closed subinterval of (a, b).

(ii) If $f \in \mathfrak{M}_{(a,b)}$ is β times continuously differentiable ($\beta \in \mathbb{N}$) with β-th derivative $f^{(\beta)} \in \mathfrak{M}_{(a,b)}$, then

$$\mathcal{M}[f](\gamma + iv) = o(|v|^{-\beta}) \quad \text{for } v \to \pm\infty$$

uniformly with respect to γ in every closed subinterval of (a, b).

(iii) Let $f : \mathbb{C} \to \mathbb{C}$ be analytical on the cone

$$S_\beta := \{z \in \mathbb{C} : 0 < |z| < \infty,\ |\arg(z)| \le \beta\} \text{ with } 0 < \beta < \pi$$

and further suppose that the restriction $f|_{\mathbb{R}_+}$ of f to $\mathbb{R}_+$ is real-valued. If $f(x) = \mathcal{O}(x^{-a})$ for $x \to 0$ in S_β and $f(x) = \mathcal{O}(x^{-b})$ for $|x| \to \infty$ in S_β, then

$$\mathcal{M}[f|_{\mathbb{R}_+}](\gamma + iv) = \mathcal{O}(e^{-\beta|v|}) \quad \text{for } v \to \pm\infty$$

uniformly with respect to γ in every closed subinterval of (a, b).

PROOF. For claims (i) and (ii) see [**19**, Proposition 3 and 5]. The proof is based on the representation of the Mellin transform as a Fourier transform (see Remark 2.2.6) and the Lemma of Riemann-Lebesgue. Additionally, for (ii) the rule

$$\mathcal{M}[f^{(\beta)}](s) = (-1)^\beta s^\beta \mathcal{M}[f](s)$$

is used.
For (iii) we roughly follow the sketch of the proof in [**19**, Proposition 5]. Note that $f|_{\mathbb{R}_+} \in \mathfrak{M}_{(a,b)}$ by Theorem 2.2.2. Since f is analytical on S_β, we can compute the Mellin trasform of $f|_{\mathbb{R}_+}$ using Cauchy's integral theorem in the following way:

$$\mathcal{M}[f|_{\mathbb{R}_+}](z) = \int_0^\infty f(t)t^{z-1}dt = \lim_{R\to\infty} \int_0^{Re^{i\beta}} f(t)t^{z-1}dt + \int_{K_R} f(t)t^{z-1}dt, \tag{3.4.1}$$

where K_R denotes the arc between $Re^{i\beta}$ and R parametrized by $t \mapsto Re^{-it}$, $t \in (-\beta, 0)$. By denoting $z = \gamma + iv$ we obtain

$$|t^{z-1}| = |t|^{-1}|t|^\gamma e^{-v\arg(t)} \le e^{\pi|\gamma|/2}|t|^{\gamma-1} =: C_z|t|^{\gamma-1}$$

for $\mathrm{Re}(t) > 0$. Let R be so large that $|f(Re^{i\theta})| \le R^{-b}$ almost everywhere. It then holds at K_R that

$$|I_R| := \left|\int_{K_R} f(t)t^{z-1}dt\right| = R\left|\int_{-\beta}^0 f(Re^{i\theta})(Re^{i\theta})^{z-1}dt\right| \le C_z\beta R^{\gamma-b}. \tag{3.4.2}$$

So I_R converges to 0 as $R\to\infty$ because $\gamma < b$. Thus, (3.4.1) and (3.4.2) yield

$$\mathcal{M}[f|_{\mathbb{R}_+}](z) = \int_0^{e^{i\beta}\infty} f(t)t^{z-1}dt.$$

By applying the change of variables $t = \rho e^{i\beta}$ we obtain

$$\mathcal{M}[f|_{\mathbb{R}_+}](z) = e^{i\beta z}\int_0^\infty f(\rho e^{i\beta})\rho^{z-1}d\rho,$$

which yields

$$|\mathcal{M}[f|_{\mathbb{R}_+}](\gamma+iv)| \le e^{-\beta v}\int_0^\infty \left|f(\rho e^{i\beta})\rho^{\gamma-1}\right| d\rho,$$

where the integral converges by $f(x) = \mathcal{O}(x^{-a})$ for $x\to 0$ in S_β and $f(x) = \mathcal{O}(x^{-b})$ for $|x|\to\infty$ in S_β. □

Next, let us establish a connection between elements of $\mathcal{C}(\beta, L, \gamma)$ and the rate of decay of their Mellin transform.

LEMMA 3.4.4. *For $a, b \in \mathbb{R}$ with $a < b$ let $f \in \mathfrak{M}_{(a,b)}$ and $\gamma \in (a, b)$.*
(i) If $f \in \mathcal{C}(\beta, L, \gamma)$ for some $L > 0$, then

$$\mathcal{M}[f](\gamma+iv) = \mathcal{O}(e^{-\beta|v|})$$

almost everywhere for $v \to \pm\infty$ for all $\beta > 0$;

(ii) If

$$\mathcal{M}[f](\gamma+iv)=\mathcal{O}(e^{-\beta|v|})$$

almost everywhere for $v\to\pm\infty$ and some $\beta>0$, then $f\in\mathcal{C}(\beta',L,\gamma)$ for all $\beta'<\beta$ and some $L>0$.

Moreover, if $f\in\mathcal{C}(\beta,L,\gamma)$ or $\mathcal{M}[f](\gamma+iv)=\mathcal{O}(e^{-\beta|v|})$ almost everywhere for $v\to\pm\infty$, then we also have

$$\int_{-\infty}^{\infty}|\mathcal{M}[f](\gamma+iv)|dv<\infty, \tag{3.4.3}$$

i.e. the Mellin inversion formula (Theorem 2.2.8) is valid for f.

Proof. Let $\beta,L>0$ with $\int e^{\beta|v|}\,|\mathcal{M}[f](\gamma+iv)|\,dv\le L$. Then

$$e^{\beta|v|}\,|\mathcal{M}[f](\gamma+iv)|\to 0 \text{ almost everywhere for } v\to\pm\infty,$$

and thus (i) is shown.
Now let $\mathcal{M}[f](\gamma+iv)=\mathcal{O}(e^{-\beta|v|})$ almost everywhere for $v\to\pm\infty$. This means, there are $C,v_0>0$ such that for all $|v|\ge v_0$ we have

$$|\mathcal{M}[f](\gamma+iv)|\le Ce^{-\beta|v|}.$$

Choose any β' with $0<\beta'<\beta$ and define $I:=\int\limits_{-v_0}^{v_0}e^{\beta|v|}\,|\mathcal{M}[f](\gamma+iv)|\,dv$. Note that $0<I<\infty$, since $\mathcal{M}[f](\gamma+iv)$ is continuous in v for all $\gamma\in(a,b)$. So we can conclude

$$\begin{aligned}\int e^{\beta'|v|}\,|\mathcal{M}[f](\gamma+iv)|\,dv &= I+\int\limits_{\{|v|\ge v_0\}} e^{\beta'|v|}\,|\mathcal{M}[f](\gamma+iv)|\,dv\\ &\le I+C\int\limits_{\{|v|\ge v_0\}} e^{(\beta'-\beta)|v|}dv\\ &= I+\frac{C}{\beta-\beta'}(e^{(\beta'-\beta)v_0}+e^{(\beta-\beta')v_0})=:L.\end{aligned}$$

Finally we have (3.4.3), because by defining
$I':=\int\limits_{-v_0}^{v_0}|\mathcal{M}[f](\gamma+iv)|\,dv<\infty$ we obtain

$$\int|\mathcal{M}[f](\gamma+iv)|\,dv\le I'+C\int\limits_{\{|v|\ge v_0\}} e^{-\beta'|v|}dv,$$

which is finite and thus yields the desired result. □

The analogous statement to Lemma 3.4.4 for the class $\mathcal{D}(\beta,L,\gamma)$ goes as follows.

Lemma 3.4.5. *For $a,b\in\mathbb{R}$ with $a<b$ let $f\in\mathfrak{M}_{(a,b)}$ and $\gamma\in(a,b)$.*

(i) *If* $f \in \mathcal{D}(\beta, L, \gamma)$ *for some* $L > 0$*, then*

$$\mathcal{M}[f](\gamma + iv) = \mathcal{O}(|v|^{-\beta})$$

almost everywhere for $v \to \pm\infty$ *and* $\int_{-\infty}^{\infty} |\mathcal{M}[f](\gamma + iv)| dv < \infty$*;*

(ii) *If there is a* $\beta > 1$ *and a* $\gamma \in (a, b)$ *with*

$$\mathcal{M}[f](\gamma + iv) = \mathcal{O}(|v|^{-\beta})$$

almost everywhere for $v \to \pm\infty$ *and* $\int_{-\infty}^{\infty} |\mathcal{M}[f](\gamma + iv)| dv < \infty$*, then* $f \in \mathcal{D}(\beta', \gamma, L)$ *for all* $\beta' \in (0, \beta - 1)$ *and some* $L > 0$*.*

Proof. Let $\beta, L > 0$ with $\int (1 + |v|^{\beta}) |\mathcal{M}[f](\gamma + iv)| \, dv \leq L$. Then

$$(1 + |v|^{\beta}) |\mathcal{M}[f](\gamma + iv)| \to 0 \text{ for } v \to \pm\infty,$$

and thus (i) is shown.
Now let $\mathcal{M}[f](\gamma + iv) = \mathcal{O}(|v|^{-\beta})$ almost everywhere for $v \to \pm\infty$ and $J := \int |\mathcal{M}[f](\gamma + iv)| \, dv$. This means, there are $C, v_0 > 0$ such that for all $|v| \geq v_0$ we have

$$|\mathcal{M}[f](\gamma + iv)| \leq C|v|^{-\beta}.$$

Choose any β' with $0 < \beta' < \beta - 1$ and define $I := \int\limits_{-v_0}^{v_0} (1 + |v|^{\beta}) |\mathcal{M}[f](\gamma + iv)| \, dv$.
Note that $0 < I < \infty$, since $\mathcal{M}[f](\gamma + iv)$ is continuous in v for all $\gamma \in (a, b)$. Hence,

$$\begin{aligned}
\int (1 + |v|^{\beta'}) |\mathcal{M}[f](\gamma + iv)| \, dv &\leq \int |\mathcal{M}[f](\gamma + iv)| \, dv + I + C \int\limits_{\{|v| \geq v_0\}} |v|^{\beta' - \beta} dv \\
&= J + I + 2C \int\limits_{v_0}^{\infty} v^{\beta' - \beta} dv \\
&= J + I + \frac{2C v_0^{\beta' - \beta + 1}}{\beta - \beta' + 1} =: L,
\end{aligned}$$

which concludes our proof. $\square$

Summing up all lemmas of this section we obtain simple sufficient criteria which imply that a density belongs to the class $\mathcal{C}(\beta, L, \gamma)$ or $\mathcal{D}(\beta, L, \gamma)$ without computing the Mellin transform.

Corollary 3.4.6. *Let* $a, b, \gamma \in \mathbb{R}$ *with* $a < \gamma < b$*.*

(i) *Let* $f \in \mathfrak{M}_{(a,b)}$*. If* f *is* β *times continuously differentiable (*$\beta \in \mathbb{N} \setminus \{1\}$*) with* $f^{(\beta)} \in \mathfrak{M}_{(a,b)}$*, then*

$$f \in \mathcal{D}(\beta', \gamma, L)$$

for all $\beta' < \beta - 1$ *and some* $L > 0$*.*

(ii) Let $f : \mathbb{C} \to \mathbb{C}$ be analytical on the sector

$$S_\beta := \{z \in \mathbb{C} : 0 < |z| < \infty,\ |\arg(z)| \leq \beta\} \text{ with } 0 < \beta < \pi$$

and further suppose that the restriction $f|_{\mathbb{R}_+}$ of f to $\mathbb{R}_+$ is real-valued. If $f(x) = \mathcal{O}(x^{-a})$ as $x \to 0$ in S_β and $f(x) = \mathcal{O}(x^{-b})$ as $|x| \to \infty$ in S_β for some $a < b$, then

$$f|_{\mathbb{R}_+} \in \mathcal{C}(\beta', L, \gamma)$$

for all $\beta' < \beta$, all $\gamma \in (a, b)$ and some $L > 0$.

PROOF. Claim (i) follows from Lemmas 3.4.3(ii) and 3.4.5(ii). To show (ii) apply Lemmas 3.4.3(iii) and 3.4.4(ii). □

With the help of Corollary 3.4.6 it is easy to see that the class $\mathcal{C}(\beta, L, \gamma)$ is fairly large and includes such well-known families of distributions as Gamma, Weibull, Beta, log-normal, inverse Gaussian and F for suitable β, L, γ. By Remark 3.4.2, the class $\mathcal{D}(\beta, L, \gamma)$ also contains these distributions.

EXAMPLE 3.4.7. *Consider the Gamma density*

$$f_\sigma(x) = \frac{x^{\sigma-1}e^{-x}}{\Gamma(\sigma)}, \quad x \geq 0$$

with parameter $\sigma \geq 0$ and define $\tilde{f}_\sigma : \{z \in \mathbb{C} | \operatorname{Re}(z) > 0\} \to \mathbb{C}$, $z \mapsto \frac{z^{\sigma-1}e^{-z}}{\Gamma(\sigma)}$. Then $\tilde{f}_\sigma$ is analytical in its domain. We have $\tilde{f}_\sigma(z) = \mathcal{O}(z^{-a})$ as $z \to 0$ for all $a \geq \max\{0, 1-\sigma\}$ and further $f(z) = \mathcal{O}(z^{-b})$ as $|z| \to \infty$ in $\{z \in \mathbb{C} | \operatorname{Re}(z) \geq 0\}$ for all $b > 0$. Moreover, $\tilde{f}_\sigma|_{\mathbb{R}_+} = f_\sigma$. Corollary 3.4.6 implies $f_\sigma \in \mathcal{C}(\beta, L, \gamma)$ for all $\beta \in (0, \pi/2)$, all $\gamma > \max\{0, 1-\sigma\}$ and some $L > 0$.

3.5. Upper Bounds on the Bias

In this section we show upper bounds on the bias of the previously constructed estimator (3.2.6). Together with a bound on the variance we use it in Sections 3.7 and 3.10 to show consistency of the estimator (3.2.6) for the classes discussed in the previous section.

LEMMA 3.5.1. *Let $Y = (Y_t)_{t\geq 0}$ be a self-similar process with scaling parameter H and càdlàg paths. Let $T \geq 0$ be a stopping time with density f_T independent of Y. Suppose $T \in \mathfrak{M}_{(a,b)}$ with $0 \leq a < b$ and $|Y_1| \in \mathfrak{M}_{(0,\infty)}$ and consider*

$$\hat{f}_n(x) = \frac{1}{2\pi} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\mathcal{M}_n[|Y_T|](\frac{\gamma+H-1+iv}{H})}{\mathcal{M}[|Y_1|](\frac{\gamma+H-1+iv}{H})} x^{-\gamma-iv}\, dv$$

from (3.2.6) *for $x > 0$ and $\max\{1-H, a\} < \gamma < b$ as an estimator for f_T. For all $n \in \mathbb{N}$ we have*

$$x^{\gamma}(f_T(x) - \mathrm{E}[\hat{f}_n(x)]) \leq \frac{1}{2\pi}\int_{\{|v|>\frac{1}{h_n}\}} |\mathcal{M}[T](\gamma + iv)|\, dv. \tag{3.5.1}$$

for all $x > 0$.

Proof. Let $x > 0$. By definition (3.2.6),

$$\mathrm{E}[\hat{f}_n(x)] = \mathrm{E}\left[\frac{1}{2\pi}\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\mathcal{M}_n[|Y_T|](\frac{\gamma+H-1+iv}{H})}{\mathcal{M}[|Y_1|](\frac{\gamma+H-1+iv}{H})} x^{-\gamma-iv}\, dv\right].$$

The integrand here is a continuous function in v, so its absolute value is bounded on the integration interval. This allows us to interchange the order of integration with Fubini's theorem to get

$$\mathrm{E}[\hat{f}_n(x)] = \frac{1}{2\pi}\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\mathrm{E}\left[\mathcal{M}_n[|Y_T|](\frac{\gamma+H-1+iv}{H})\right]}{\mathcal{M}[|Y_1|](\frac{\gamma+H-1+iv}{H})} x^{-\gamma-iv}\, dv.$$

Considering Lemma 2.2.12 and (3.2.2) we see that

$$\begin{aligned}\mathrm{E}[\hat{f}_n(x)] &= \frac{1}{2\pi}\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\mathcal{M}[|Y_T|](\frac{\gamma+H-1+iv}{H})}{\mathcal{M}[|Y_1|](\frac{\gamma+H-1+iv}{H})} x^{-\gamma-iv}\, dv \\ &= \frac{1}{2\pi}\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \mathcal{M}[T](\gamma+iv) x^{-\gamma-iv} dv. \end{aligned} \tag{3.5.2}$$

We combine Theorem 2.2.8 with (3.5.2) to get

$$f_T(x) - \mathrm{E}[\hat{f}_n(x)] = \frac{1}{2\pi}\int\limits_{\{|v|>\frac{1}{h_n}\}} \mathcal{M}[T](\gamma+iv) x^{-\gamma-iv} dv \tag{3.5.3}$$

and after taking the absolute value on the right hand side of (3.5.3) and applying the triangle inequality we see

$$f_T(x) - \mathrm{E}[\hat{f}_n(x)] \leq \frac{x^{-\gamma}}{2\pi}\int\limits_{\{|v|>\frac{1}{h_n}\}} |\mathcal{M}[T](\gamma+iv)|\, dv,$$

which is our claim. □

Assuming now that f_T belongs to one of the two classes introduced in Section 3.4 we can give a further bound on the bias of $\hat{f}_n$.

LEMMA 3.5.2. *(i) If $T \in \mathcal{C}(\beta, L, \gamma)$ for some $\beta, L > 0$ and $\gamma \in (a, b)$ with $\gamma > 1 - H$, then*

$$x^\gamma \,\mathrm{E}[f_T(x) - \hat{f}_n(x)] \le \frac{L}{2\pi} e^{-\beta/h_n} \tag{3.5.4}$$

for all $x > 0$.

(ii) If $T \in \mathcal{D}(\beta, L, \gamma)$ for some $\beta, L > 0$ and $\gamma \in (a, b)$ with $\gamma > 1 - H$, then

$$x^\gamma \,\mathrm{E}[f_T(x) - \hat{f}_n(x)] \le \frac{L}{2\pi} h_n^\beta \tag{3.5.5}$$

for all $x > 0$.

PROOF. (i) Lemma 3.5.1 implies

$$\begin{aligned} x^\gamma(f_T(x) - \mathrm{E}[\hat{f}_n(x)]) &\le \frac{1}{2\pi} \int\limits_{\{|v| > \frac{1}{h_n}\}} |\mathcal{M}[T](\gamma + iv)| \, dv \\ &= \frac{e^{-\beta/h_n}}{2\pi} \int\limits_{\{|v| > \frac{1}{h_n}\}} e^{\beta/h_n} \, |\mathcal{M}[T](\gamma + iv)| \, dv \\ &\le \frac{e^{-\beta/h_n}}{2\pi} \int\limits_{\{|v| > \frac{1}{h_n}\}} e^{\beta|v|} \, |\mathcal{M}[T](\gamma + iv)| \, dv \\ &\le L \frac{e^{-\beta/h_n}}{2\pi} \end{aligned}$$

for all $x > 0$.

(ii) We have

$$\begin{aligned} x^\gamma(f_T(x) - \mathrm{E}[\hat{f}_n(x)]) &\le \frac{1}{2\pi} \int\limits_{\{|v| > \frac{1}{h_n}\}} |\mathcal{M}[T](\gamma + iv)| \, dv \\ &= \frac{h_n^\beta}{2\pi} \int\limits_{\{|v| > \frac{1}{h_n}\}} h_n^{-\beta} \, |\mathcal{M}[T](\gamma + iv)| \, dv \\ &\le \frac{h_n^\beta}{2\pi} \int\limits_{\{|v| > \frac{1}{h_n}\}} |v|^\beta \, |\mathcal{M}[T](\gamma + iv)| \, dv \\ &\le L \frac{h_n^\beta}{2\pi} \end{aligned}$$

for all $x > 0$, which is our claim. □

It turns out that the bound in Lemma 3.5.2 depends on the rate of decay of the Mellin transform of T, which in turn is tied to the smoothness of f_T (cf. Section 3.4): The more smoothness we impose on f_T, the faster $\mathcal{M}[T]$ decays and the better the performance of the estimator (3.2.6).

3.6. Upper Bounds on the Variance

Having established an upper bound on the bias of $\hat{f}_n$, we now shall do the same for the variance of our estimator.

LEMMA 3.6.1. *Let $Y = (Y_t)_{t\geq 0}$ be a self-similar process with scaling parameter H and càdlàg paths. Let $T \geq 0$ be a stopping time independent of Y with density f_T. Suppose $T \in \mathfrak{M}_{(a,b)}$ with $0 \leq a < b$ and $|Y_1| \in \mathfrak{M}_{(0,\infty)}$. Consider*

$$\hat{f}_n(x) = \frac{1}{2\pi} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\mathcal{M}_n[|Y_T|](\frac{\gamma+H-1+iv}{H})}{\mathcal{M}[|Y_1|](\frac{\gamma+H-1+iv}{H})} x^{-\gamma-iv}\, dv$$

from (3.2.6) *for $x > 0$ and $\max\{1-H, a\} < \gamma < b$ as an estimator for f_T. We have*

$$\operatorname{Var}[x^\gamma \hat{f}_n(x)] \leq \frac{\mathcal{M}[|Y_1|]((2\gamma-2)/H+1)\mathcal{M}[T](2\gamma-1)}{4\pi^2 n} \times \left(\int_{-1/h_n}^{1/h_n} \frac{1}{|\mathcal{M}[|Y_1|](\frac{\gamma+H-1+iv}{H})|} dv \right)^2$$

for all $n \in \mathbb{N}$ and all $x > 0$.

PROOF. Let $x > 0$, $n \in \mathbb{N}$. Then we have

$$\begin{aligned}
\mathrm{Var}[x^\gamma \hat{f}_n(x)] &= \mathrm{Var}\left[\frac{1}{2\pi}\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\mathcal{M}_n[|Y_T|](\frac{\gamma+H-1+iv}{H})}{\mathcal{M}[|Y_1|](\frac{\gamma+H-1+iv}{H})} x^{-iv}\, dv\right] \\
&\le \frac{1}{4\pi^2}\left(\int_{-1/h_n}^{1/h_n} \sqrt{\mathrm{Var}\left[\frac{\mathcal{M}_n[|Y_T|](\frac{\gamma+H-1+iv}{H})}{\mathcal{M}[|Y_1|](\frac{\gamma+H-1+iv}{H})} x^{-iv}\right]} dv\right)^2 \\
&= \frac{1}{4\pi^2}\left(\int_{-1/h_n}^{1/h_n} \frac{\sqrt{\mathrm{Var}\left[\mathcal{M}_n[|Y_T|](\frac{\gamma+H-1+iv}{H})\right]}}{|\mathcal{M}[|Y_1|](\frac{\gamma+H-1+iv}{H})|} dv\right)^2
\end{aligned}$$

$$= \frac{1}{4\pi^2 n}\left(\int_{-1/h_n}^{1/h_n} \frac{\sqrt{\mathrm{Var}\left[|Y_T|^{\frac{\gamma-1+iv}{H}}\right]}}{|\mathcal{M}[|Y_1|](\frac{\gamma+H-1+iv}{H})|} dv\right)^2. \tag{3.6.1}$$

Here we used Lemma 2.1.8 for the "$\le$"-sign, Lemma 2.1.7 for the second "="-sign and Lemma 2.2.12 for the last "="-sign. In order to get a bound on $\mathrm{Var}\left[|Y_T|^{\frac{\gamma-1+iv}{H}}\right]$ we apply Fubini's theorem and the self-similarity of Y to get

$$\begin{aligned}
\mathrm{Var}\left[|Y_T|^{\frac{\gamma-1+iv}{H}}\right] &\le \mathrm{E}[(|Y_T|)^{(2\gamma-2)/H}] \\
&= \int_0^\infty \mathrm{E}[(|Y_t|)^{(2\gamma-2)/H}] f_T(t) dt \\
&= \mathrm{E}[(|Y_1|)^{(2\gamma-2)/H}] \int_0^\infty t^{2\gamma-2} f_T(t) dt \\
&= \mathcal{M}[|Y_1|]((2\gamma-2)/H+1)\mathcal{M}[T](2\gamma-1),
\end{aligned}$$

which (together with (3.6.1)) gives the desired bound on $\mathrm{Var}[x^\gamma \hat{f}_n(x)]$. □

3.7. Application to Bessel Processes

To make further statements we need to specify Y in the general setting of section 3.1, where Y was only supposed to be a known self-similar process with càdlàg paths. In this section we choose Y to be a Bessel process $BES = (BES_t)_{t\ge 0}$ starting in 0 with dimension $d \ge 1$. Throughout this thesis non-integer d are allowed. Note that the case $d = 1$ leads to the absolute value of the one-dimensional Brownian motion and was already considered in [**6**]. We refer the reader to [**31**] for detailed information about Bessel processes. Here we shall only state some basic properties.

DEFINITION 3.7.1. *A Bessel process with dimension $d \geq 1$ started in $x \geq 0$ is a Markov process with the semi-group $(P_t)_{t\geq 0}$ of Markov kernels on $\mathbb{R}_+$ defined by*

$$P_t(x,B) := \int_B p_t(x,y)dy \quad \text{for all } B \in \mathcal{B}(\mathbb{R}_+), x \in \mathbb{R}_+, t \geq 0,$$

where

$$p_t(x,y) = \frac{1}{t}\left(\frac{y}{x}\right)^{d/2-1} y \exp\left(-\frac{x^2+y^2}{2t}\right) I_{d/2-1}\left(\frac{xy}{t}\right) \quad \text{for } x,t>0 \tag{3.7.1}$$

and

$$I_\nu(x) = \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(\nu+k+1)}\left(\frac{x}{2}\right)^{2k+\nu} \quad \text{for } \nu \geq -1, x>0$$

is a modified Bessel function.

See for instance [**31**] for the proof of the fact that $(P_t)_{t\geq 0}$ indeed constitutes a semi-group. The following well-known properties of Bessel processes allow us to construct the estimator (3.2.6).

THEOREM 3.7.2. *Let $BES = (BES_t)_{t\geq 0}$ be a Bessel process with dimension $d \geq 1$ started in 0.*

(i) The density of BES_t is given by:

$$f_t(y) = \frac{2^{1-\frac{d}{2}} t^{-\frac{d}{2}}}{\Gamma(d/2)} y^{d-1} e^{-\frac{y^2}{2t}}, \quad y \geq 0.$$

(This follows from (3.7.1) *with $x \to 0$, see* [**31**, page 446]*)*

(ii) BES is self-similar with scaling parameter $H = \frac{1}{2}$ (see [**31**, page 446]*).*

(iii) BES is a Feller process with continuous paths (see [**31**, page 446]*).*

(iv) In Example 2.2.5 we calculated for $\mathrm{Re}(s) \geq 1-d$*:*

$$\mathcal{M}[BES_1](s) = \frac{1}{\Gamma(d/2)}\Gamma\left(\frac{s+d-1}{2}\right) 2^{\frac{s-1}{2}}.$$

(v) BES can be realized as the Euclidean norm of a d-dimensional Brownian motion, if $d \in \mathbb{N}$ (see [**31**, page 446]*).*

Looking at (3.2.6) we use Theorem 3.7.2 and obtain

$$\hat{f}_n(x) = \frac{1}{2\pi}\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\mathcal{M}_n[|BES_T|](\frac{\gamma+1/2-1+iv}{1/2})}{\mathcal{M}[|BES_1|](\frac{\gamma+1/2-1+iv}{1/2})} x^{-\gamma-iv}\, dv$$

$$= \frac{1}{2\pi}\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\Gamma\left(\frac{d}{2}\right)\frac{1}{n}\sum_{k=1}^{n} X_k^{2(\gamma-1+iv)}}{\Gamma\left(\gamma+\frac{d}{2}-1+iv\right)2^{\gamma-1+iv}} x^{-\gamma-iv}\, dv \tag{3.7.2}$$

as an estimator for the density $f_T(x)$ of a stopping time $T \geq 0$ for $x > 0$ and $\max\{1/2, a\} < \gamma < b$, where a, b are such that $T \in \mathfrak{M}_{(a,b)}$ and $X_1, ..., X_n$ are independent samples of BES_T. With our major result Theorem 3.7.3 we shall derive the convergence rates for (3.7.2). Recall the weighted mean squared risk

$$\mathrm{MSE}_\gamma(x) = x^{2\gamma}\,\mathrm{E}[|f_T(x) - \hat{f}_n(x)|^2]$$

from the introduction to Section 3.3.

THEOREM 3.7.3. *Let $BES = (BES_t)_{t\geq 0}$ be a Bessel process with dimension $d \in \mathbb{R}$, $d \geq 1$. Let $T \geq 0$ be a stopping time with density f_T independent of BES. Suppose $T \in \mathfrak{M}_{(a,b)}$ with $0 \leq a < b$ and consider*

$$\hat{f}_n(x) = \frac{1}{2\pi} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\Gamma\left(\frac{d}{2}\right) \frac{1}{n} \sum_{k=1}^n X_k^{2(\gamma-1+iv)}}{\Gamma\left(\gamma + \frac{d}{2} - 1 + iv\right) 2^{\gamma-1+iv}} x^{-\gamma-iv}\, dv$$

for $x > 0$ and $\max\{1/2, a\} < \gamma < b$ from (3.7.2) as an estimator for f_T.

(i) If $T \in \mathcal{C}(\beta, L, \gamma)$ for some $\beta, L > 0$ and $\gamma \in (a, b)$ with $2\gamma - 1 \in (a, b)$ and $\gamma > (4-d)/4$, then

$$\mathrm{MSE}_\gamma(x) \leq C_{L,d,\gamma} \left(\frac{1}{n} e^{\pi/h_n} + e^{-2\beta/h_n} \right) \tag{3.7.3}$$

holds for all $x > 0$ and some positive constant $C_{L,d,\gamma}$ (given by (3.7.11) in the proof) depending only on L, γ, d as well as T. Moreover, taking

$$h_n = \frac{\pi + 2\beta}{\log n}$$

in (3.7.3), one has the polynomial convergence rate

$$\sqrt{\mathrm{MSE}_\gamma(x)} \lesssim n^{\frac{-\beta}{\pi+2\beta}},\ n \to \infty$$

for all $x > 0$.

(ii) If $T \in \mathcal{D}(\beta, L, \gamma)$ for some $\beta, L > 0$ and $\gamma \in (a, b)$ with $2\gamma - 1 \in (a, b)$ and $\gamma > (4-d)/4$, then for all $x > 0$:

$$\mathrm{MSE}_\gamma(x) \leq C_{L,d,\gamma} \left(\frac{1}{n} e^{\pi/h_n} + h_n^{2\beta} \right) \tag{3.7.4}$$

with the same constant $C_{L,d,\gamma}$ as in (i). Choosing

$$h_n = \frac{\pi}{\log n - 2\beta \log\log n}$$

in (3.7.3) yields the logarithmic convergence rate

$$\sqrt{\mathrm{MSE}_\gamma(x)} \lesssim (\log n)^{-\beta},\ n \to \infty$$

for all $x > 0$.

PROOF. Let $x > 0$. By (3.3.4),

$$\mathrm{MSE}_\gamma(x) = \mathrm{Var}[x^\gamma \hat{f}_n(x)] + \left(x^\gamma \,\mathrm{E}[(f_T(x) - \hat{f}_n(x))]\right)^2. \tag{3.7.5}$$

For the variance we can achieve the same bound in parts (i) and (ii) of the claim. The bias needs to be treated separately.
We use the upper bound on variance obtained in Lemma 3.6.1 with $H = 1/2$ to get

$$\mathrm{Var}[x^\gamma \hat{f}_n(x)] \leq \frac{C_0(\gamma, d)}{4\pi^2 n} \left(\int_{-1/h_n}^{1/h_n} \frac{1}{|\mathcal{M}[BES_1](2\gamma - 1 + 2iv)|} dv \right)^2, \tag{3.7.6}$$

where

$$C_0(\gamma, d) := \mathcal{M}[BES_1](4\gamma - 3)\mathcal{M}[T](2\gamma - 1).$$

Since we already calculated

$$\mathcal{M}[BES_1](s) = \frac{1}{\Gamma(d/2)} \Gamma\left(\frac{s + d - 1}{2}\right) 2^{\frac{s-1}{2}} \tag{3.7.7}$$

for $\mathrm{Re}(s) \geq 0$ in Example 2.2.5, we can write

$$C_0(\gamma, d) = \frac{1}{\Gamma(d/2)} \Gamma\left(\frac{d}{2} + 2\gamma - 2\right) 2^{2\gamma - 2} \mathcal{M}[T](2\gamma - 1)$$

and see that is $C_0(\gamma, d)$ finite by the assumption $\gamma \geq (4 - d)/4$ and $2\gamma - 1 \in (a, b)$. We continue inequality (3.7.6) by firstly plugging in (3.7.7) and secondly applying Lemma 2.4.4(ii) (with $\delta = 1$, $U = 1/h_n$ and $\alpha = \gamma - 1 + \frac{d}{2}$ there):

$$\begin{aligned}
\mathrm{Var}[x^\gamma \hat{f}_n(x)] &\leq \frac{C_0(\gamma, d)}{4\pi^2 n} \left(\int_{-1/h_n}^{1/h_n} \frac{\Gamma\left(\frac{d}{2}\right)}{|\Gamma\left(\gamma - 1 + \frac{d}{2} + iv\right) 2^{\gamma - 1 + iv}|} dv \right)^2 \\
&\leq \frac{C_0(\gamma, d) \Gamma\left(\frac{d}{2}\right)^2}{\pi^2 2^{2\gamma} n} \left(\int_{-1/h_n}^{1/h_n} \frac{1}{|\Gamma\left(\gamma - 1 + \frac{d}{2} + iv\right)|} dv \right)^2 \\
&\leq \frac{C_0(\gamma, d) \Gamma\left(\frac{d}{2}\right)^2}{\pi^2 2^{2\gamma} n} (C_1(d, \gamma) + C_2 e^{\frac{\pi}{2h_n}})^2
\end{aligned} \tag{3.7.8}$$

for some positive constants $C_1(d, \gamma)$ and C_2 obtainable from the proof of Lemma 2.4.4(ii).
Now we prove part (i) of the claim. Let $T \in \mathcal{C}(\beta, L, \gamma)$. We plug in (3.7.8) and (3.5.4) from Lemma 3.5.2(i) into (3.7.5) to obtain

$$\mathrm{MSE}_\gamma(x) \leq \frac{C_0(\gamma, d) \Gamma\left(\frac{d}{2}\right)^2}{\pi^2 2^{2\gamma} n} (C_1(d, \gamma) + C_2 e^{\frac{\pi}{2h_n}})^2 + L^2 \frac{e^{-2\beta/h_n}}{4\pi^2}. \tag{3.7.9}$$

Inequality (3.7.9) can be simplified via following technical Lemma.

LEMMA 3.7.4. *Let $a,b,c,d > 0$ and $(x_n)_{n\in\mathbb{N}}$, $(y_n)_{n\in\mathbb{N}}$ sequences of positive numbers. If $x_n \to \infty$ for $n \to \infty$, then we have*

$$\frac{a}{n}(b+cx_n)^2 + dy_n \le \max\{a(b+c)^2, d\}(x_n^2/n + y_n)$$

for all $n \ge n_0$ for some $n_0 \in \mathbb{N}$.

PROOF. Choose n_0 such that $x_n \ge 1$ for all $n \ge n_0$. Then we have

$$\begin{aligned}\frac{a}{n}(b+cx_n)^2 + dy_n &= \frac{a}{n}(b/x_n + c)^2 x_n^2 + dy_n \\ &\le a(b+c)^2 x_n^2/n + dy_n \\ &\le \max\{a(b+c)^2, d\}(x_n^2/n + y_n)\end{aligned}$$

for all $n \ge n_0$. □

Continuing with the proof of Theorem 3.7.3 we choose $x_n = e^{\frac{\pi}{2h_n}}$, $y_n = e^{-2\beta/h_n}$, $a = \frac{C_0(\gamma,d)\Gamma\left(\frac{d}{2}\right)^2}{\pi^2 2^{2\gamma}}$, $b = C_1(d,\gamma)$, $c = C_2$ and $d = \frac{L^2}{4\pi^2}$ in Lemma 3.7.4. Recall that $h_n \to 0$ so that $x_n \to \infty$ for $n \to \infty$. Then (3.7.9) implies

$$\text{MSE}_\gamma(x) \le C_{L,d,\gamma}\left(\frac{1}{n}e^{\pi/h_n} + e^{-2\beta/h_n}\right) \tag{3.7.10}$$

for all n greater than some $n_0 \in \mathbb{N}$ and

$$\begin{aligned}C_{L,d,\gamma} &:= \max\left\{\frac{C_0(\gamma,d)\Gamma\left(\frac{d}{2}\right)^2(C_1(d,\gamma)+C_2)^2}{\pi^2 2^{2\gamma}}, \frac{L^2}{4\pi^2}\right\} \\ &= \max\left\{\frac{2^{\gamma-1}}{\pi^2}\Gamma\left(\frac{d}{2}\right)\Gamma\left(2\gamma+\frac{d}{2}-2\right)\mathcal{M}[T](2\gamma-1)(C_1(d,\gamma)+C_2)^2, \frac{L^2}{4\pi^2}\right\}.\end{aligned} \tag{3.7.11}$$

Looking at (3.7.10) choose h_n in such a way that $\frac{1}{n}e^{\pi/h_n} = e^{-2\beta/h_n}$. This is accomplished by the choice $h_n = \frac{\pi+2\beta}{\log n}$, which yields

$$\begin{aligned}C_{L,d,\gamma}\left(\frac{1}{n}e^{\pi/h_n} + e^{-2\beta/h_n}\right) &= C_{L,d,\gamma}\left(\frac{1}{n}\exp\left(\frac{\pi\log n}{\pi+2\beta}\right) + \exp\left(\frac{-2\beta\log n}{\pi+2\beta}\right)\right) \\ &= C_{L,d,\gamma}\left(n^{\frac{\pi}{\pi+2\beta}-1} + n^{\frac{-2\beta}{\pi+2\beta}}\right) \\ &= 2C_{L,d,\gamma}n^{-\frac{2\beta}{\pi+2\beta}},\end{aligned}$$

and allows us to conclude that for $n \to \infty$:

$$\sqrt{\text{MSE}_\gamma(x)} \le \sqrt{2C_{L,d,\gamma}n^{-\frac{2\beta}{\pi+2\beta}}} \lesssim n^{\frac{-\beta}{\pi+2\beta}}.$$

This concludes the proof of part (i) of the claim. We next turn to (ii).
For $T \in \mathcal{D}(\beta, L, \gamma)$ we obtain the same bound on the variance as previously. For

the bias we take the bound obtained in Lemma 3.5.2(ii). Plugging these bounds into (3.7.5) gives

$$\begin{aligned}\mathrm{MSE}_\gamma(x) &\le \frac{C_0(\gamma,d)2^{-d-\gamma}}{\pi^2\Gamma\left(\frac{d}{2}\right)^2 n}(C_1(d,\gamma)+C_2 e^{\frac{\pi}{2h_n}})^2+\left(\frac{L}{2\pi}h_n^\beta\right)^2\\ &\le C_{L,d,\gamma}\left(\frac{1}{n}e^{\pi/h_n}+h_n^{2\beta}\right)\end{aligned}$$

with the same constant $C_{L,d,\gamma}$ as in (i) and Lemma 3.7.4 as justification for the second inequality sign. Next choose h_n such that $\frac{1}{n}e^{\pi/h_n} = h_n^{2\beta}$, that is to say $h_n = \frac{\pi}{\log n - 2\beta \log\log n}$, which gives

$$\begin{aligned}\frac{1}{n}e^{\pi/h_n}+h_n^{2\beta} &= \frac{\exp(\log n - 2\beta\log\log n)}{n}+\frac{\pi^{2\beta}}{(\log n - 2\beta\log\log n)^{2\beta}}\\ &= \log^{-2\beta} n + \pi^{2\beta}\left(\log n - 2\beta\log\log n\right)^{-2\beta}.\end{aligned}$$

Hence, we conclude that

$$\begin{aligned}\sqrt{\mathrm{MSE}_\gamma(x)} &\le \sqrt{C_{L,d,\gamma}\left(\log^{-2\beta} n + \pi^{2\beta}\left(\log n - 2\beta\log\log n\right)^{-2\beta}\right)}\\ &\lesssim (\log n)^{-\beta}\end{aligned}$$

for $n \to \infty$. □

In Section 3.4 we showed that the class $\mathcal{C}(\beta, L, \gamma)$ (and therefore $\mathcal{D}(\beta, L, \gamma)$) includes such well-known families of distributions as Gamma, Weibull, Beta, log-normal, inverse Gaussian, F for all $\beta \in (0, \pi/2)$. So, if T belongs to one of those families, we have $(a, b) = (0, \infty)$ in Theorem 3.7.3 and the statement is true for any $\gamma > \max\{1/2, (4-d)/4\}$. If $d \ge 2$, then we only require $\gamma > 1/2$.
The influence of the choice parameter γ in inequalities (3.7.3) and (3.7.4) is ambiguous. Note that $\mathrm{MSE}_\gamma(x)$ is decreasing in γ for $x \in (0, 1)$ and increasing in γ for $x > 1$. Based on this fact alone, we would choose a small γ whenever we want to estimate f_T near the origin, and we would choose a large γ whenever we want to estimate f_T for arguments greater 1. This observation about γ does not depend on the setting of this section and can be made whenever we consider error bounds with respect to MSE_γ. The influence of γ on the constant $C_{L,d,\gamma}$ is not determinable because the latter depends on the unknown stopping time T (see (3.7.11)).

3.8. Asymptotic Normality, Unbiasedness and Variance in Context of Bessel Processes

Note that estimator (3.7.2) can be written as

$$\hat{f}_n(x) = \frac{1}{n}\sum_{k=1}^{n} Z_{n,k}, \tag{3.8.1}$$

with

$$Z_{n,k} := \frac{\Gamma\left(\frac{d}{2}\right)}{\pi} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{X_k^{2(\gamma-1+iv)}}{\Gamma\left(\gamma+d/2-1+iv\right)2^{\gamma+iv}} x^{-\gamma-iv}\,dv \tag{3.8.2}$$

for $x > 0,\ d \geq 1,\ h_n > 0$ with $h_n \to 0$ and $\gamma \in (a,b)$, where $a, b \in \mathbb{R}$ are such that $T \in \mathfrak{M}_{(a,b)}$ and $X_1, ..., X_n$ are independent samples of BES_T. Since $\hat{f}_n$ is a sum of independent identically distributed random variables, we can show that (under mild assumptions on f_T) $\hat{f}_n$ is asymptotically normal, that is for some positive sequence $(\nu_n)_{n\in\mathbb{N}}$

$$\sqrt{n}\nu_n^{-1/2}(\hat{f}_n(x) - \mathrm{E}[\hat{f}_n(x)])$$

converges in distribution to the standard normal distribution for all $x > 0$.
We begin by investigating the absolute moments of $Z_{n,1}$.

Lemma 3.8.1. *Let $f_T \in \mathfrak{M}_{(a,b)}$ for some $0 \leq a < b$ and $(h_n)_{n\in\mathbb{N}} \subset \mathbb{R}_+$ with $h_n \to 0$ as $n \to \infty$. If there is a $\gamma \in (a,b)$ such that $2\gamma - 1 \in (a,b)$, $\gamma > (4-d)/4$ and $(\gamma-1)j+1 \in (a,b)$, then the j-th absolute moment of $Z_{n,1}$ (see (3.8.2)) is asymptotically bounded by*

$$\mathrm{E}[|Z_{n,1}|^j] \lesssim \begin{cases} h_n^{\gamma+d/2-3/2} e^{\pi j/(2h_n)}, & \text{if } \gamma+d/2-3/2 < 0 \\ e^{\pi j/(2h_n)}, & \text{if } \gamma+d/2-3/2 \geq 0 \end{cases} \tag{3.8.3}$$

as $n \to \infty$ for all $j \in \mathbb{R}_+$. The constant implied by $\lesssim$ may depend on j, d and γ. Note that $Z_{n,1}$ depends on h_n, d and γ. In particular, we claim that all absolute moments of $Z_{n,1}$ exist for all $n \in \mathbb{N}$ greater than some $n_0 \in \mathbb{N}$.

For the special case $(a,b) = (0,1)$ and $d = j = 1$ this result is mentioned in [**6**, Proof of Proposition 4.1] but without an extensive proof, which we provide here.

Proof. Case 1: $\gamma + d/2 - 3/2 \geq 0$. Using Jensen inequality for the first inequality sign, Lemma 2.4.4(ii) for the second and the fact

$$\mathcal{M}[X_1](s) = \mathcal{M}[T]((s+1)/2)\mathcal{M}[BES_1](s)$$

(cf. (3.2.2) with $H = 1/2,\ Y = BES$) for the last equality sign we estimate

$$\begin{aligned}
\mathrm{E}[|Z_{n,1}|^j] =&\ \mathrm{E}\left[\left|\frac{\Gamma\left(\frac{d}{2}\right)}{\pi}\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{X_1^{2(\gamma-1+iv)}}{\Gamma\left(\gamma+d/2-1+iv\right)2^{\gamma+iv}} x^{-\gamma-iv}\,dv\right|^j\right] \\
\leq&\ \frac{\Gamma\left(\frac{d}{2}\right)^j}{(2^\gamma \pi x^\gamma)^j}\mathrm{E}[X_1^{2(\gamma-1)j}] \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{1}{|\Gamma\left(\gamma+d/2-1+iv\right)|^j}\,dv \\
\leq&\ \frac{\Gamma\left(\frac{d}{2}\right)^j}{(2^\gamma \pi x^\gamma)^j}\mathcal{M}[X_1](2(\gamma-1)j+1)(C_{\gamma,d,j}+C_j e^{\pi j/(2h_n)}) \\
=&\ \frac{\Gamma\left(\frac{d}{2}\right)^j}{(2^\gamma \pi x^\gamma)^j}\mathcal{M}[T]((\gamma-1)j+1)\mathcal{M}[BES_1](2(\gamma-1)j+1) \\
&\times (C_{\gamma,d,j}+C_j e^{\pi j/(2h_n)}),
\end{aligned}$$

where $C_{\gamma,d,j}$ and C_j are some positive constants from Lemma 2.4.4. This implies the claim since the appearing Mellin transforms are finite by assumptions on γ.
Case 2: $\gamma + d/2 - 3/2 < 0$. We obtain

$$\begin{aligned}
\mathrm{E}[|Z_{n,1}|^j] \leq&\ \frac{\Gamma\left(\frac{d}{2}\right)^j}{(2^\gamma \pi x^\gamma)^j}\mathcal{M}[T]((\gamma-1)j+1)\mathcal{M}[BES_1](2(\gamma-1)j+1) \\
&\times C_j h_n^{\gamma+d/2-3/2} e^{\pi j/(2h_n)}
\end{aligned}$$

with the same arguments as in case 1. The only difference is that we apply Lemma 2.4.4(i) instead of Lemma 2.4.4(ii). Hence, (3.8.11) holds in this case as well. □

Suppose $d \geq 2$ in Lemma 3.8.1. Then the assumption $\gamma > (4-d)/4$ is redundant. Moreover, $\gamma + d/2 - 3/2 \geq 0$ holds and we always have the smaller bound

$$\mathrm{E}[|Z_{n,1}|^j] \lesssim e^{\pi j/(2h_n)}. \tag{3.8.4}$$

for $n \to \infty$.
The following lemma is only an auxiliary result used in the proof of Theorem 3.8.3, which will the main result of this section.

LEMMA 3.8.2. *Let $h_n \to 0$ for $n \to \infty$ and $\rho_n := h_n^{-\alpha}$, where $0 < \alpha < 1/2$. Then*

$$\begin{aligned}
(1/h_n - s)\log(1/h_n - s) - (1/h_n - r)\log(1/h_n - r) - (r - s) =&\ (r-s)\log(1/h_n) \\
&+ \mathcal{O}(\rho_n^2 h_n)
\end{aligned}$$

for all $0 < r, s < \rho_n$ and $n \to \infty$.

PROOF. In the Taylor expansion

$$x\log(x) = a\log(a) + (\log(a)+1)(x-a) + \frac{1}{2a}(x-a)^2 - \int_a^x \frac{(x-t)^2}{6t^2}dt$$

for $x, a > 0$ choose $a = 1/h_n$, and $x = 1/h_n - s$ or $x = 1/h_n - r$ respectively. For both choices the remainder is of order $\mathcal{O}(h_n^2\rho_n^3)$ and we get

$$\begin{aligned}(1/h_n - s)\log(1/h_n - s) &- (1/h_n - r)\log(1/h_n - r) - (r-s)\\ &= (r-s)\log(1/h_n) + h_n s^2/2 - h_n r^2/2 + \mathcal{O}(h_n^2\rho_n^3)\\ &= (r-s)\log(1/h_n) + \mathcal{O}(\rho_n^2 h_n).\end{aligned}$$

This concludes the proof. □

We are now ready to prove the main result of this section.

THEOREM 3.8.3. *Let $f_T \in \mathfrak{M}_{(a,b)}$ for some $0 \le a < b$. Suppose there is a $\gamma \in (a,b)$ such that $2\gamma - 1 \in (a,b)$, $\gamma > (4-d)/4$ and $(\delta+2)\gamma - \delta - 1 \in (a,b)$ for some $\delta > 0$ and*

$$\int_{-\infty}^{\infty} |\mathcal{M}[T](2\gamma - 1 + iv)|dv < \infty. \tag{3.8.5}$$

If we choose $h_n \sim \log^{-1}(n)$ in (3.7.2) then we have

$$\sqrt{n}\nu_n^{-1/2}(\hat{f}_n(x) - \mathrm{E}[\hat{f}_n(x)]) \xrightarrow{d} \mathcal{N}(0,1) \tag{3.8.6}$$

for all $x > 0$, where

$$\nu_n = \frac{c\Gamma\left(\frac{d}{2}\right)}{2\pi^3 x^{2\gamma}} h_n^{2\gamma-d+3} e^{\frac{\pi}{h_n}} \log^{-2}(1/h_n)(1+o(1)) \tag{3.8.7}$$

with

$$c = (\pi/2)\Gamma(2\gamma + (d-4)/2)\mathcal{M}[T](2\gamma-1) > 0. \tag{3.8.8}$$

Let us make a few remarks before we present the proof. As we mentioned in the end of Section 3.7, we can often assume $(a,b) = (0,\infty)$ so that the choice of γ is only restricted by $\gamma > \max\{1/2, (4-d)/4\}$. If $(a,b) = (0,\infty)$, then a suitable δ can always be found, for instance, any $\delta > 2\gamma/(1-\gamma)$ is valid. If additionally $d \ge 2$, then the statement is true for all $\gamma > 1/2$.
The choice $h_n \sim \log^{-1}(n)$ may not be the only possible one. Other choices may also yield (3.8.6) (perhaps even with a better convergence rate (see Theorem 3.8.5), but we are not aware of them.

PROOF OF THEOREM 3.8.3. We roughly imitate the proof of an analogous result for the special case $d = 1$, $(a,b) = (0,1)$ found in [**6**]. In distinction from [**6**] we do not restrict ourselves to the case $x = 1$ in the proof and provide the specific form of ν_n for all $x > 0$.

Throughout this proof we use brief notations like $\{u \geq r\}$ and $\{|u-v| \leq r\}$ for the respective sets $\{u \in \mathbb{R} | u \geq r\}$ and $\{(u,v) \in \mathbb{R}^2 || u-v| \leq r\}$ ($r \in \mathbb{R}$), and similar ones. Symbols $\mathbb{1}_{u \geq r}$ and $\mathbb{1}_{|u-v| \leq r}$ denote the indicator functions of such sets.

The sequence $(Z_{n,k})_{k \in \mathbb{N}}$ is square-integrable (see Lemma 3.8.1) and i.i.d. for all $n \in \mathbb{N}$. So it suffices to show the Lyapunov condition, i.e. for a $\delta > 0$:

$$\lim_{n\to\infty} \frac{\mathrm{E}[|Z_{n,1} - \mathrm{E}[Z_{n,1}]|^{2+\delta}]}{n^{\delta/2}(\mathrm{Var}[Z_{n,1}])^{1+\delta/2}} = 0. \tag{3.8.9}$$

By [**5**, page 239] the claim (3.8.6) follows from (3.8.9) with $\nu_n = \mathrm{Var}[Z_{n,1}]$.

Let $x > 0$. Note that

$$\mathrm{E}[Z_{n,1}] \to f_T(x), \quad n \to \infty$$

by monotone convergence and (3.2.4) (if we choose $H = 1/2$ and $Y = BES$ there). So, (3.8.9) holds if we can prove, that $\mathrm{Var}[Z_{n,1}] \to \infty$ and

$$\lim_{n\to\infty} \frac{\mathrm{E}[|Z_{n,1}|^{2+\delta}]}{n^{\delta/2}(\mathrm{Var}[Z_{n,1}])^{1+\delta/2}} = 0. \tag{3.8.10}$$

In any case of Lemma 3.8.1 (for $j = \delta + 2$) we have

$$\mathrm{E}[|Z_{n,1}|^{2+\delta}] \lesssim h_n^{-c} e^{\pi(2+\delta)/(2h_n)} \tag{3.8.11}$$

as $n \to \infty$ for all $\delta \in \mathbb{R}_+$ and some $c > 0$.
Now we investigate the asymptotic behavior of $\mathrm{Var}[Z_{n,1}]$. Looking at (3.8.2) we use Definition 2.1.4, pull out constants and integral signs using Lemmas 2.1.6 and 2.1.10 and then calculate the covariance by the rule $\mathrm{Cov}[X,Y] = \mathrm{E}[X\overline{Y}] - \mathrm{E}[X]\overline{\mathrm{E}[Y]}$ (Lemma 2.1.6) to obtain

$$\begin{aligned}
\mathrm{Var}[Z_{n,1}] &= \mathrm{Cov}[Z_{n,1}, Z_{n,1}] \\
&= \frac{\Gamma\left(\frac{d}{2}\right)^2}{\pi^2} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\mathrm{Cov}[X_1^{2\gamma-1+iv}, X_1^{2\gamma-1+iu}]dvdu}{(2x)^{2\gamma+i(v-u)}\Gamma\left(\gamma+d/2-1+iv\right)\Gamma\left(\gamma+d/2-1-iu\right)} \\
&= \frac{\Gamma\left(\frac{d}{2}\right)^2}{\pi^2} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\mathrm{E}[X_1^{4\gamma-4+2i(v-u)}]dvdu}{(2x)^{2\gamma+i(v-u)}\Gamma\left(\gamma+d/2-1+iv\right)\Gamma\left(\gamma+d/2-1-iu\right)} \\
&\quad - \frac{\Gamma\left(\frac{d}{2}\right)^2}{\pi^2} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{(2x)^{-2\gamma-i(v-u)}\,\mathrm{E}[X_1^{2\gamma-2+2iv}]\,\mathrm{E}[X_1^{2\gamma-2-2iu}]dvdu}{\Gamma\left(\gamma+d/2-1+iv\right)\Gamma\left(\gamma+d/2-1-iu\right)} \\
&= \frac{\Gamma\left(\frac{d}{2}\right)^2}{\pi^2} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\mathcal{M}[X_1](4\gamma+2i(v-u)-3)dvdu}{(2x)^{2\gamma+i(v-u)}\Gamma\left(\gamma+d/2-1+iv\right)\Gamma\left(\gamma+d/2-1-iu\right)} \\
&\quad - \frac{\Gamma\left(\frac{d}{2}\right)^2}{\pi^2} \left| \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\mathcal{M}[X_1](2\gamma-1+2iv)}{(2x)^{\gamma+iv}\Gamma\left(\gamma+d/2-1+iv\right)} dv \right|^2 =: R_1 - R_2.
\end{aligned}$$

By Example 2.2.5(iii) we can estimate

$$R_2 \leq \frac{1}{\pi^2 x^{2\gamma}} \left(\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \left|\mathcal{M}[T](\gamma+iv)\right| dv \right)^2 < C < \infty$$

for some $C > 0$ and further

$$R_1 = \frac{\Gamma\left(\frac{d}{2}\right)}{2\pi^2 x^{2\gamma}} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\mathcal{M}[T](2\gamma-1+i(v-u))\Gamma\left(2\gamma-2+d/2+i(v-u)\right)}{x^{i(v-u)}\Gamma\left(\gamma+d/2-1+iv\right)\Gamma\left(\gamma+d/2-1-iu\right)} dvdu.$$

Our strategy now is to decompose the double integral defining R_1 into pieces that are easy to estimate. To that end let $\rho_n := h_n^{-\alpha}$, where $0 < \alpha < 1/2$ and define

$$I^1_{n,\rho_n} := \int\limits_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \int\limits_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \mathbb{1}_{|v-u|\geq\rho_n} \frac{\mathcal{M}[T](2\gamma-1+i(v-u))\Gamma\left(2\gamma-2+\frac{d}{2}+i(v-u)\right)}{x^{i(v-u)}\Gamma\left(\gamma+\frac{d-2}{2}+iv\right)\Gamma\left(\gamma+\frac{d-2}{2}-iu\right)} dvdu.$$

By Lemma 2.4.3 there are $C_1, C_2 > 0$ such that

$$|\Gamma(\gamma+d/2-1+iv)| \geq C_1\mathbb{1}_{|v|\leq 2} + C_2\mathbb{1}_{|v|>2}|v|^{\gamma-1+(d-1)/2}e^{-\pi|v|/2}$$
$$|\Gamma(\gamma+d/2-1-iu)| \geq C_1\mathbb{1}_{|u|\leq 2} + C_2\mathbb{1}_{|u|>2}|u|^{\gamma-1+(d-1)/2}e^{-\pi|u|/2}$$

and $K_1, K_2 > 0$ such that

$$\Gamma(2\gamma-2+d/2+i(v-u)) \leq K_1\mathbb{1}_{|v-u|\leq 2} + K_2\mathbb{1}_{|v-u|>2}|v-u|^{2(\gamma-1)+\frac{d-1}{2}}e^{-\pi|v-u|/2}.$$

With the help of these inequalities we deduce

$$\begin{aligned}
|I^1_{n,\rho_n}| &\lesssim \int\limits_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \int\limits_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\mathbb{1}_{|v-u|\geq\rho_n}|v-u|^{2(\gamma-1)+\frac{d-1}{2}}e^{-\frac{\pi}{2}|v-u|}|\mathcal{M}[T](2\gamma-1+i(v-u))|dvdu}{(\mathbb{1}_{|v|\leq 2}+\mathbb{1}_{|v|>2}|v|^{\gamma-1+\frac{d-1}{2}}e^{-\frac{\pi}{2}|v|})(\mathbb{1}_{|u|\leq 2}+\mathbb{1}_{|u|>2}|u|^{\gamma-1+\frac{d-1}{2}}e^{-\frac{\pi}{2}|u|})} \\
&\leq h_n^{-2|\gamma-1|-\frac{d-1}{2}}e^{-\pi\rho_n/2} \int\limits_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \int\limits_{-\frac{1}{h_n}}^{\frac{1}{h_n}} |\mathcal{M}[T](2\gamma-1+i(v-u))| \\
&\quad \times \left(1+|u|^{|\gamma-1|+\frac{d-1}{2}}e^{\frac{\pi}{2}|u|}+|v|^{|\gamma-1|+\frac{d-1}{2}}e^{\frac{\pi}{2}|v|}+|vu|^{|\gamma-1|+\frac{d-1}{2}}e^{\frac{\pi}{2}(|v|+|u|)}\right)dvdu \\
&\leq h_n^{-2|\gamma-1|-\frac{d-1}{2}}e^{-\pi\rho_n/2}(1+2h_n^{-|\gamma-1|-\frac{d-1}{2}}e^{\pi/(2h_n)}+h_n^{-2|\gamma-1|-(d-1)}e^{\pi/h_n}) \\
&\quad \times \int\limits_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \int |\mathcal{M}[T](2\gamma-1+iv)|dvdu \\
&\lesssim h_n^{-3|\gamma-1|-d+2}e^{\pi(\frac{1}{2h_n}-\frac{\rho_n}{2})} + h_n^{-4|\gamma-1|-(3d-5)/2}e^{\pi(\frac{1}{h_n}-\frac{\rho_n}{2})}
\end{aligned}$$

for $n\to\infty$. (Note (3.8.5) for the last $\lesssim$-sign. Also note that appearing additive constants are negligible, because the last expression diverges to infinity.) Similarly, we estimate

$$\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}}\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \mathbb{1}_{|u|\le\frac{1}{h_n}-\rho_n}\mathbb{1}_{|v-u|\ge\rho_n}\frac{\mathcal{M}[T](2\gamma-1+i(v-u))\Gamma\left(2\gamma+\frac{d-4}{2}+i(v-u)\right)}{x^{i(v-u)}\Gamma\left(\gamma+\frac{d-2}{2}+iv\right)\Gamma\left(\gamma+\frac{d-2}{2}-iu\right)}dvdu$$

$$\begin{aligned}
&\lesssim h_n^{-2|\gamma-1|-\frac{d-1}{2}}e^{-\pi\rho_n/2}\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}}\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}}\mathbb{1}_{|u|\le\frac{1}{h_n}-\rho_n}|\mathcal{M}[T](2\gamma-1+i(v-u))|\\
&\quad\times\left(1+|u|^{|\gamma-1|+\frac{d-1}{2}}e^{\pi|u|/2}+|v|^{|\gamma-1|+\frac{d-1}{2}}e^{\pi|v|/2}+|vu|^{|\gamma-1|+\frac{d-1}{2}}e^{\pi(|v|+|u|)/2}\right)dvdu\\
&\le h_n^{-2|\gamma-1|-\frac{d-1}{2}}e^{-\pi\rho_n/2}(1+h_n^{-|\gamma-1|-\frac{d-1}{2}}e^{\pi/(2h_n)}+h_n^{-|\gamma-1|-\frac{d-1}{2}}e^{\pi/(2h_n-\rho_n)}\\
&\quad+h_n^{-2|\gamma-1|-(d-1)}e^{\pi/h_n})\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}}\int_{\mathbb{R}}|\mathcal{M}[T](2\gamma-1+iv)|dvdu\\
&\lesssim h_n^{-3|\gamma-1|-d+2}e^{\pi(\frac{1}{2h_n}-\frac{\rho_n}{2})}+h_n^{-3|\gamma-1|-d+2}e^{\pi(\frac{1}{2h_n}-\rho_n)}+h_n^{-4|\gamma-1|-(3d-5)/2}e^{\pi(\frac{1}{h_n}-\frac{3\rho_n}{2})}\\
&\lesssim h_n^{-l}e^{\pi(\frac{1}{h_n}-\rho_n)}
\end{aligned}\tag{3.8.12}$$

for some $l\ge 0$. By interchanging the roles of v and u in the last inequality chain,

$$\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}}\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \mathbb{1}_{|v|\le\frac{1}{h_n}-\rho_n}\mathbb{1}_{|v-u|\ge\rho_n}\frac{\mathcal{M}[T](2\gamma-1+i(v-u))\Gamma\left(2\gamma+\frac{d-4}{2}+i(v-u)\right)}{x^{i(v-u)}\Gamma\left(\gamma+\frac{d-2}{2}+iv\right)\Gamma\left(\gamma+\frac{d-2}{2}-iu\right)}dvdu \lesssim h_n^{-l}e^{\pi(\frac{1}{h_n}-\rho_n)}. \tag{3.8.13}$$

Combine (3.8.12) and (3.8.13) to obtain

$$\begin{aligned}
I^2_{n,\rho_n} &:= \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}}\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}}\mathbb{1}_{|v-u|\le\rho_n}\frac{\mathcal{M}[T](2\gamma-1+i(v-u))\Gamma\left(2\gamma+\frac{d-4}{2}+i(v-u)\right)}{x^{i(v-u)}\Gamma\left(\gamma+\frac{d-2}{2}+iv\right)\Gamma\left(\gamma+\frac{d-2}{2}-iu\right)}dvdu\\
&\lesssim \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}}\int_{-\frac{1}{h_n}}^{\frac{1}{h_n}}\mathbb{1}_{|v|\ge\frac{1}{h_n}-\rho_n}\mathbb{1}_{|u|\ge\frac{1}{h_n}-\rho_n}\mathbb{1}_{|v-u|\le\rho_n}\\
&\quad\times\frac{\mathcal{M}[T](2\gamma-1+i(v-u))\Gamma\left(2\gamma+\frac{d-4}{2}+i(v-u)\right)}{x^{i(v-u)}\Gamma\left(\gamma+\frac{d-2}{2}+iv\right)\Gamma\left(\gamma+\frac{d-2}{2}-iu\right)}dvdu+\mathcal{O}(h_n^{-l}e^{\pi(\frac{1}{h_n}-\rho_n)})\\
&=: I^3_{n,\rho_n}+\mathcal{O}(h_n^{-l}e^{\pi(\frac{1}{h_n}-\rho_n)}).
\end{aligned}$$

Next, we examine the asymptotic behavior of the integral I^3_{n,ρ_n}. To this end, we take advantage of Stirling's formula (Lemma 2.4.1)

$$\Gamma\left(\gamma+\frac{d-2}{2}+iv\right)=\left(\gamma+\frac{d-2}{2}+iv\right)^{\gamma+\frac{d-3}{2}+iv}e^{-\gamma-\frac{d-2}{2}-iv}\sqrt{2\pi}(1+\mathcal{O}(v^{-1}))$$

for $v\to\infty$. Consider the integrand of I^3_{n,ρ_n}. In the denominator it holds by means of the identity $\log(iv)=\log(v)+\frac{i\pi}{2}$ that

$$\begin{aligned}&\Gamma(\gamma+d/2-1+iv)\Gamma(\gamma+d/2-1-iu)\\&\quad=2\pi\exp(iv\log v-iu\log u-i(v-u))\\&\qquad\times\exp\left(-\frac{\pi}{2}(u+v)+\left(\gamma+\frac{d-3}{2}\right)(\log(v)+\log(u))\right)(1+\mathcal{O}(v^{-1})+\mathcal{O}(u^{-1}))\end{aligned}$$

for $u,v\to\infty$. On the set

$$\left\{|u|\geq\frac{1}{h_n}-\rho_n\right\}\cap\left\{|v|\geq\frac{1}{h_n}-\rho_n\right\}\cap\{|v-u|\leq\rho_n\}\cap\{v\geq 0,u\geq 0\}$$

we define $u=1/h_n-r$, $v=1/h_n-s$ with $0<r,s<\rho_n$, $|r-s|<\rho_n$ to obtain

$$\begin{aligned}&\Gamma(\gamma+d/2-1+iv)\Gamma(\gamma+d/2-1-iu)\\&\qquad=2\pi\exp(i(1/h_n-s)\log(1/h_n-s)-i(1/h_n-r)\log(1/h_n-r)-i(r-s))\\&\qquad\quad\times h_n^{-2\gamma-d+3}e^{-\pi/h_n}e^{\pi(r+s)/2}(1+\mathcal{O}(h_n))(1+\mathcal{O}(\rho_nh_n)+\mathcal{O}(\rho_n^2h_n))^{\gamma+(d-3)/2}.\end{aligned}$$

Note that due to the choice of ρ_n, we have $\rho_nh_n\to 0$ and $\rho_n^2h_n\to 0$. Using Lemma 3.8.2 we derive

$$\begin{aligned}\Gamma\left(\gamma+d/2-1+iv\right)\Gamma\left(\gamma+d/2-1-iu\right)&=2\pi h_n^{-2\gamma-d+3}\exp(i(r-s)\log(1/h_n))\\&\quad\times e^{-\pi/h_n}e^{\pi(r+s)/2}(1+O(\rho_n^2h_n)).\end{aligned}$$

Analogously, on the set

$$\left\{|u|\geq\frac{1}{h_n}-\rho_n\right\}\cap\left\{|v|\geq\frac{1}{h_n}-\rho_n\right\}\cap\{|v-u|\leq\rho_n\}\cap\{v\leq 0,u\leq 0\}$$

we define $u=-1/h_n+r$, $v=-1/h_n+s$ with $0<r,s<\rho_n$, $|r-s|<\rho_n$ to obtain

$$\begin{aligned}\Gamma\left(\gamma+d/2-1+iv\right)\Gamma\left(\gamma+d/2-1-iu\right)&=2\pi h_n^{-2\gamma-d+3}\exp(-i(r-s)\log(1/h_n))\\&\quad\times e^{-\pi/h_n}e^{\pi(r+s)/2}(1+O(\rho_n^2h_n)).\end{aligned}$$

Hence, I^3_{n,ρ_n} can be decomposed as follows:

$$\begin{aligned}
I^3_{n,\rho_n} &= \frac{h_n^{2\gamma+d-3}}{2\pi}\exp(\pi/h_n)\int_0^{\rho_n}\int_0^{\rho_n}\mathbb{1}_{\{|r-s|\le\rho_n\}}x^{i(s-r)}\\
&\quad\times\exp(-\pi(r+s)/2)\Gamma(2\gamma+(d-4)/2+i(r-s))\\
&\quad\times\mathcal{M}[T](2\gamma-1+i(r-s))\exp(i(s-r)\log(1/h_n))(1+O(\rho_n^2h_n))^{-1}drds\\
&= \frac{h_n^{2\gamma+d-3}}{2\pi}\exp(\pi/h_n)\{\mathrm{Re}[I^4_{n,\rho_n}]+O(\rho_n^2h_n)]\},
\end{aligned}$$

where

$$\begin{aligned}
I^4_{n,\rho_n} &:= \int_0^{\rho_n}\int_0^{\rho_n}\mathbb{1}_{\{|r-s|\le\rho_n\}}\exp(-\pi(r+s)/2)\Gamma(2\gamma+(d-4)/2+i(r-s))\\
&\quad\times\mathcal{M}[T](2\gamma-1+i(r-s))\exp(i(s-r)\log(1/h_n))x^{i(r-s)}drds\\
&= \int_0^{\rho_n}e^{-\pi v}R_n(v)dv
\end{aligned}$$

with
(3.8.14)

$$R_n(v) := \int_0^{\rho_n-v} e^{-\pi u/2}x^{iu}\Gamma(2\gamma+(d-4)/2+iu)\mathcal{M}[T](2\gamma-1+iu)\exp(iu\log(1/h_n))du.$$

The Fourier type integral in (3.8.14) allows a series representation via the following lemma:

LEMMA 3.8.4. *Let $\alpha<\beta$. If $f:(\alpha,\beta)\to\mathbb{C}$ is N times continuously differentiable ($N\in\mathbb{N}$), then we have the expansion*

$$\int_\alpha^\beta f(u)e^{ixu}du = B_N(x)-A_N(x)+o(x^{-N})$$

for $x\to\infty$, where

$$A_N(x) = \sum_{n=0}^{N-1} i^{n-1}f^{(n)}(\alpha)x^{-n-1}e^{ix\alpha}$$

$$B_N(x) = \sum_{n=0}^{N-1} i^{n-1}f^{(n)}(\beta)x^{-n-1}e^{ix\beta}.$$

PROOF. Repeated integration by parts, see [**17**, page 47]. □

Coming back to (3.8.14), we choose

$$f(u) := e^{-\pi u/2}x^{iu}\Gamma(2\gamma+(d-4)/2+iu)\mathcal{M}[T](2\gamma-1+iu)$$

and $N = 2$ in Lemma 3.8.4 to get

$$\begin{aligned} R_n(v) &= if(0)\log^{-1}(1/h_n) + f'(0)\log^{-2}(1/h_n) + \mathcal{O}(\log^{-3}(1/h_n)) \\ &= i\Gamma\left(2\gamma + \frac{d-4}{2}\right)\mathcal{M}[T](2\gamma-1)\log^{-1}(1/h_n) \\ &\quad - \frac{d}{du}[e^{-\pi u/2}x^{iu}\Gamma\left(2\gamma + \frac{d-4}{2} + iu\right)\mathcal{M}[T](2\gamma-1+iu)]\Big|_{u=0}\log^{-2}(1/h_n) \\ &\quad + \mathcal{O}(\log^{-3}(1/h_n)) \end{aligned}$$

uniformly in v. Note that

$$\mathcal{M}[T](2\gamma - 1 + iu)$$

and

$$\frac{d}{du}\mathcal{M}[T](2\gamma-1+iu) = i\mathcal{M}[log(\cdot)f_T(\cdot)](2\gamma-1+iu)$$

vanish for $u \to \infty$ by Lemma 3.4.3. Likewise do

$$\Gamma(2\gamma + (d-4)/2 + iu) \quad \text{and} \quad \tfrac{d}{du}\Gamma(2\gamma + (d-4)/2 + iu)$$

for $u \to \infty$ by Lemma 2.4.3. So, the term B_2 in Lemma 3.8.4 is at most of order $\mathcal{O}(e^{-\pi(\rho_n - v)/2})$ and is negligible compared to $\mathcal{O}(\log^{-3}(1/h_n))$. Thus,

$$\begin{aligned} \mathrm{Re}[I^4_{n,\rho_n}] &= (c\log^{-2}(1/h_n) + \mathcal{O}(\log^{-3}(1/h_n)))(1/\pi - e^{-\pi\rho_n}/\pi) \\ &= \frac{c}{\pi}\log^{-2}(1/h_n) + \mathcal{O}(\log^{-3}(1/h_n)) \end{aligned}$$

holds with

$$\begin{aligned} c :=& -\mathrm{Re}\left[\frac{d}{du}e^{-\pi u/2}x^{iu}\Gamma(2\gamma + (d-4)/2 + iu)\Big|_{u=0}\right] \\ =& -\mathrm{Re}[-\pi/2\Gamma(2\gamma + (d-4)/2 + iu)\mathcal{M}[T](2\gamma-1) \\ &\quad + ix^{-1}\Gamma(2\gamma + (d-4)/2)\mathcal{M}[T](2\gamma-1) \\ &\quad + i\Gamma(2\gamma + (d-4)/2)\psi(2\gamma + (d-4)/2)\mathcal{M}[T](2\gamma-1) \\ &\quad + i\Gamma(2\gamma + (d-4)/2)\mathcal{M}[\log(\cdot)f_T(\cdot)](2\gamma-1)] \\ =& \frac{\pi}{2}\Gamma(2\gamma + (d-4)/2)\mathcal{M}[T](2\gamma-1), \end{aligned}$$

where $\psi(s) = \Gamma'(s)/\Gamma(s)$ for $\mathrm{Re}(s) > 0$ is the digamma-function. So, we have $c > 0$, if $\mathcal{M}[T](2\gamma - 1)$ is positive, which is true for all nonnegative stopping times except for $T = 0$ or $T = \infty$ almost surely, which is not allowed, because T has a Lebesgue density.

Summing up the auxiliary quantities introduced above we get

$$
\begin{aligned}
\mathrm{Var}[Z_{n,1}] &= R_1 - R_2 \\
&= \frac{\Gamma\left(\frac{d}{2}\right)}{2\pi^2 x^{2\gamma}}(I^1_{n,\rho_n} + I^2_{n,\rho_n}) + \mathcal{O}(1) \\
&= \frac{\Gamma\left(\frac{d}{2}\right)}{2\pi^2 x^{2\gamma}}(I^3_{n,\rho_n} + \mathcal{O}(h_n^{-l} e^{\pi(\frac{1}{h_n}-\rho_n)})) + \mathcal{O}(1) \\
&= \frac{\Gamma\left(\frac{d}{2}\right)}{2\pi^2 x^{2\gamma}}(h_n^{2\gamma-d+3} e^{\frac{\pi}{h_n}}\{\mathrm{Re}[I^4_{n,\rho_n}] + \mathcal{O}(\rho_n^2 h_n)\} + \mathcal{O}(h_n^{-l} e^{\pi(\frac{1}{h_n}-\rho_n)})) + \mathcal{O}(1) \\
&= \frac{\Gamma\left(\frac{d}{2}\right)}{2\pi^2 x^{2\gamma}}\Big[h_n^{2\gamma-d+3} e^{\frac{\pi}{h_n}}\{\frac{c}{\pi}\log^{-2}(1/h_n) + \mathcal{O}(\log^{-3}(1/h_n)) + \mathcal{O}(\rho_n^2 h_n)\} \\
&\quad + \mathcal{O}(h_n^{-l} e^{\pi(\frac{1}{h_n}-\rho_n)})\Big] + \mathcal{O}(1)
\end{aligned}
$$

and thus,

$$
\mathrm{Var}[Z_{n,1}] = \frac{c\Gamma\left(\frac{d}{2}\right)}{2\pi^3 x^{2\gamma}} h_n^{2\gamma-d+3} e^{\frac{\pi}{h_n}} \log^{-2}(1/h_n)(1+o(1)) \tag{3.8.15}
$$

for $n \to \infty$. Combining (3.8.15) and (3.8.11) we finally get

$$
\begin{aligned}
\frac{\mathrm{E}[|Z_{n,1}|^{2+\delta}]}{n^{\delta/2}(\mathrm{Var}[Z_{n,1}])^{1+\delta/2}} &\lesssim \frac{h_n^{-c} e^{\pi(2+\delta)/(2h_n)}}{n^{\delta/2}(h_n^{2\gamma+d-3}\log^{-2}(1/h_n)\exp(\pi/h_n))^{1+\delta/2}} \\
&= \frac{h_n^{-c}}{n^{\delta/2}(h_n^{2\gamma+d-3}\log^{-2}(1/h_n))^{1+\delta/2}} \\
&\to 0
\end{aligned}
$$

for $n \to \infty$, if we choose $h_n \sim \log^{-1}(n)$. So the Lyapunov condition is satisfied and the claim follows using (3.8.15). □

Incidentally, we proved that under assumptions of Theorem 3.8.3,

$$
\mathrm{Var}[\hat{f}_n(x)] = \mathrm{Var}[Z_{n,1}]\ \frac{1}{n} = \frac{c\Gamma\left(\frac{d}{2}\right)}{2\pi^3 x^{2\gamma}} h_n^{2\gamma-d+3} e^{\frac{\pi}{h_n}} \log^{-2}(1/h_n)(1+o(1))\ \frac{1}{n}. \tag{3.8.16}
$$

Recall that in the proof of Theorem 3.7.3 we only showed (see (3.7.8))

$$
\mathrm{Var}[\hat{f}_n(x)] \le \frac{C_0(\gamma,d)\Gamma\left(\frac{d}{2}\right)^2}{\pi^2 2^{2\gamma} x^{2\gamma}}(C_1(d,\gamma) + C_2 e^{\frac{\pi}{2h_n}})^2\ \frac{1}{n}
$$

for some constants C_0, C_1, C_2. With (3.8.16) we now have an even better bound on the variance of $\hat{f}_n(x)$.
Since all moments of $Z_{n,k}$ exist by Lemma 3.8.1, it is possible to give a Berry-Esseen type error estimate for the convergence in (3.8.6). This is a new result even for dimension $d = 1$.

THEOREM 3.8.5. *Let $f_T \in \mathfrak{M}_{(a,b)}$ for some $0 \leq a < b$. Suppose there is a $\gamma \in (a,b)$ such that $2\gamma - 1 \in (a,b)$, $\gamma > (4-d)/4$, $3\gamma - 2 \in (a,b)$ and*

$$\int_{-\infty}^{\infty} |\mathcal{M}[T](2\gamma - 1 + iv)| dv < \infty. \tag{3.8.17}$$

Fix some $x > 0$. Denote by F_n the distribution function of

$$\sqrt{n}\nu_n^{-1/2}(\hat{f}_n(x) - \mathrm{E}[\hat{f}_n(x)])$$

(where $\hat{f}_n(x)$ is defined by (3.7.2) and $\nu_n = n \operatorname{Var}[\hat{f}_n(x)]$ is given by (3.8.7)) and by Φ the distribution function of the standard normal distribution. Define the distance

$$\rho_n := \sup_{y \in \mathbb{R}} |F_n(y) - \Phi(y)|.$$

If we choose $h_n \sim \log^{-1}(n)$ in (3.7.2) then we have

$$\rho_n \lesssim \begin{cases} n^{-1/2}(\log n)^{4\gamma - d + 3} \log^3(\log(n)), & \text{if } \gamma + d/2 - 3/2 < 0 \\ n^{-1/2}(\log n)^{3(2\gamma - d + 3)/2} \log^3(\log(n)), & \text{if } \gamma + d/2 - 3/2 \geq 0 \end{cases} \tag{3.8.18}$$

for $n \to \infty$.

PROOF. Let $x > 0$ and $n \in \mathbb{N}$. Consider the representation (3.8.1) of $\hat{f}_n(x)$ as mean of independently identically distributed variables $Z_{n,1}, ..., Z_{n,n}$. Berry-Esseen Theorem (see [**20**, Korollar 4.2.15]) states

$$\rho_n \leq \frac{6\,\mathrm{E}[|Z_{n,1} - \mathrm{E}[Z_{n,1}]|^3]}{(\operatorname{Var}[Z_{n,1}])^{3/2} n^{1/2}}. \tag{3.8.19}$$

First, apply Minkowski inequality, then triangle inequality and finally Lyapunov inequality to obtain

$$\begin{aligned} \mathrm{E}[|Z_{n,1} - \mathrm{E}[Z_{n,1}]|^3] &\leq ((\mathrm{E}[|Z_{n,1}|^3])^{1/3} + |\mathrm{E}[Z_{n,1}]|)^3 \\ &\leq ((\mathrm{E}[|Z_{n,1}|^3])^{1/3} + \mathrm{E}[|Z_{n,1}|])^3 \\ &\leq (2(\mathrm{E}[|Z_{n,1}|^3])^{1/3})^3 \\ &\lesssim \mathrm{E}[|Z_{n,1}|^3] \end{aligned}$$

for $n \to \infty$. Then we choose $j = 3$ in Lemma 3.8.1 to get

$$\mathrm{E}[|Z_{n,1} - \mathrm{E}[Z_{n,1}]|^3] \lesssim \begin{cases} h_n^{\gamma + d/2 - 3/2} e^{3\pi/(2h_n)}, & \text{if } \gamma + d/2 - 3/2 < 0 \\ e^{3\pi/(2h_n)}, & \text{if } \gamma + d/2 - 3/2 \geq 0 \end{cases} \tag{3.8.20}$$

for $n \to \infty$. During the proof of Theorem 3.8.3 we already showed that under assumptions on γ above,

$$n \operatorname{Var}[Z_{n,1}] = \nu_n = \frac{c\Gamma\left(\frac{d}{2}\right)}{2\pi^3 x^{2\gamma}} h_n^{2\gamma - d + 3} e^{\frac{\pi}{h_n}} \log^{-2}(1/h_n)(1 + o(1)) \tag{3.8.21}$$

for some $c > 0$. Now plug (3.8.20) and (3.8.21) in (3.8.19). Then choose

$$h_n \sim \log^{-1}(n)$$

. This leads to two cases:
Case 1: $\gamma + d/2 - 3/2 < 0$. We have

$$\begin{aligned} \rho_n &\lesssim \frac{h_n^{\gamma+d/2-3/2} e^{3\pi/(2h_n)}}{(h_n^{2\gamma-d+3} e^{\frac{\pi}{h_n}} \log^{-2}(1/h_n)(1+o(1)))^{3/2} n^{1/2}} \\ &= \frac{(\log n)^{\gamma+d/2-3/2} n^{3\pi/2}}{(\log n)^{3(-2\gamma+d-3)/2} n^{3\pi/2} \log^{-3}(\log(n))(1+o(1))^{3/2} n^{1/2}} \\ &\lesssim n^{-1/2} (\log n)^{4\gamma-d+3} \log^3(\log(n)) \end{aligned}$$

for $n \to \infty$.
Case 2: $\gamma + d/2 - 3/2 \geq 0$. We have

$$\begin{aligned} \rho_n &\lesssim \frac{e^{3\pi/(2h_n)}}{(h_n^{2\gamma-d+3} e^{\frac{\pi}{h_n}} \log^{-2}(1/h_n)(1+o(1)))^{3/2} n^{1/2}} \\ &= \frac{n^{3\pi/2}}{(\log n)^{3(-2\gamma+d-3)/2} n^{3\pi/2} \log^{-3}(\log(n))(1+o(1))^{3/2} n^{1/2}} \\ &\lesssim n^{-1/2} (\log n)^{3(2\gamma-d+3)/2} \log^3(\log(n)) \end{aligned}$$

for $n \to \infty$. This concludes the proof. □

We do not know if the rates in (3.8.18) are optimal. Different choices of h_n and different estimates in the proof may yield smaller bounds. Note that the signs of the powers $4\gamma - d + 3$ and $3(2\gamma - d + 3)/2$ in (3.8.18) are ambiguous and depend on the relative positions of γ and d. However, if $d \geq 2$ then we only have the case $\gamma + d/2 - 3/2 \geq 0$ and the power of the logarithm is positive.
In order to achieve a high convergence rate we would choose γ as small as possible (under the given restrictions in Theorem 3.8.3). However, a small γ may result in a large variance of $\hat{f}_n$ because the influence of γ is ambiguous there (see (3.8.7)).

3.9. Simulation Study

In this Section we test our estimator (3.7.2) with some simulated data. Consider a Bessel process with dimension $d = 5$ and a Gamma$(2,1)$ distributed stopping time T, i.e. T has the density

$$f(x) = xe^{-x}, \quad x \geq 0.$$

We write the estimator (3.7.2) as a function of $1/h_n$:

$$\hat{f}[1/h_n](x) := \hat{f}_n(x) = \frac{1}{2\pi} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\Gamma\left(\frac{d}{2}\right) \frac{1}{n} \sum_{k=1}^n X_k^{2(\gamma-1+iv)}}{\Gamma\left(\gamma + \frac{d}{2} - 1 + iv\right) 2^{\gamma-1+iv}} x^{-\gamma-iv} \, dv$$

Next we choose $\gamma = 0.7$. The question of the appropriate choice for the cut-off parameter $1/h_n$ arises here. From Theorem 3.7.3 we know

$$1/h_n = \log(n)/(\pi + 2\beta) \tag{3.9.1}$$

to be the optimal choice for densities from class $\mathcal{C}(\beta, L, \gamma)$. But β is unknown in practice. Corollary 3.4.6(ii) suggests that often $\beta \in (0, \pi)$. Thus, we can guess from (3.9.1) that some $1/h_n \in (0.7, 2.2)$ would be a good choice for $n = 1000$.

Next, we compute $\hat{f}[1/h_n](x)$ for $n = 1000$ and different values of the cut-off parameter $1/h_n$. On the left-hand side of Figure 2 we can see the loss

$$\sup_{x \in \mathbb{R}_+} \{|\hat{f}[1/h_n](x) - f_T(x)|\} \tag{3.9.2}$$

approximated by

$$\sup_{x \in \{0.1, 0.11, \dots, 9.99, 10\}} \{|\hat{f}[1/h_n](x) - f_T(x)|\}$$

as a function of $1/h_n$ with minimum at $1/h_n^\star \approx 2.7$. Denote

$$f_n^o(x) := \hat{f}[1/h_n^\star](x).$$

Since f_n^o contains information about the unknown f_T, it is not an estimator. We call it an *oracle* for f_T. "This is the "best forecast" of $[f_T]$, which is, however, inaccessible: in order to construct it, we would need an "oracle" that knows $[f_T]$"([**37**, page 60]). We call $1/h_n^\star$ the oracle choice of $1/h_n$.

Figure 2 demonstrates that an appropriate choice of $1/h_n$ is crucial for the performance of $\hat{f}_n$, but for practical reasons we wish to assume as little as possible about the unknown density f_T. Hence, we propose a data driven choice of $1/h_n$ based on the quasi-optimality approach proposed in [**4**]. The same was done in [**6**] for the case $d = 1$. We will show that it leads to reasonable results in our model as well. The implementation goes as follows.

Firstly, we choose a sequence of bandwidths $1/h_n^1, \dots, 1/h_n^N$ and calculate estimators $\hat{f}[1/h_n^1](x), \dots, \hat{f}[1/h_n^N](x)$. Secondly, we determine $l^\star = \arg\min_{l=1,\dots,N} d(l)$ with

$$\begin{aligned} d(l) :&= \int_0^\infty |f[1/h_n^{l+1}](x) - f[1/h_n^l](x)| dx \\ &\approx \sum_{i=0}^{90} |f[1/h_n^{l+1}](0.1 + 0.01i) - f[1/h_n^l](0.1 + 0.01i)|. \end{aligned}$$

Denote the adaptive estimator constructed in this way by

$$\tilde{f}_n := f[1/h_n^{l^\star}].$$

Note that we avoid the evaluation of $\hat{f}_n(x)$ for $x \in (0, 0.1)$ since the calculation is very unstable there. We also forego evaluation for $x > 10$ which is reasonable in the case of a Gamma-distributed stopping time. (We have $\hat{f}_n(x) \approx 0$ for $x > 10$.)

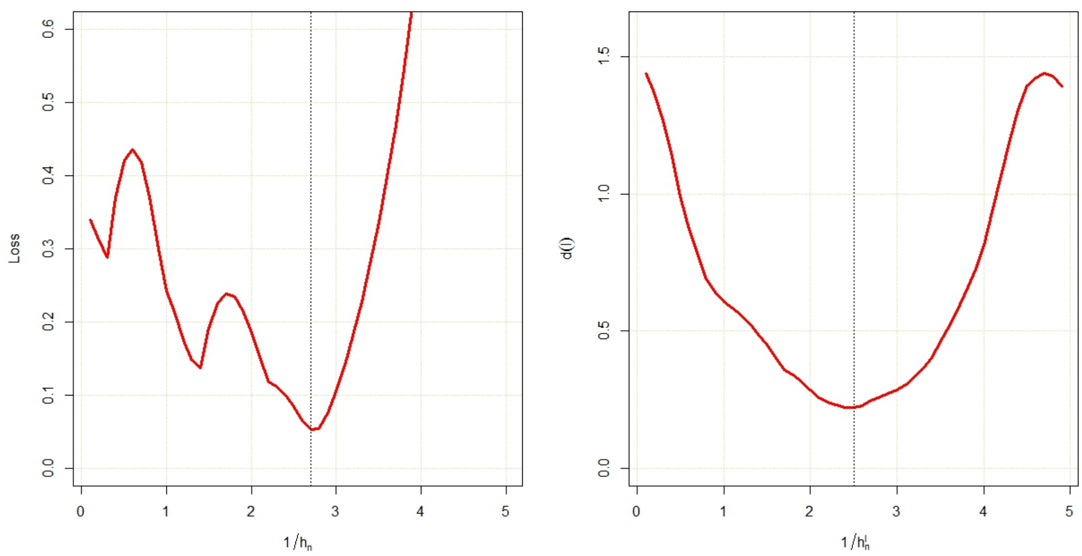

FIGURE 2. Left: the loss $\sup_{x\in\mathbb{R}_+}\{|\hat{f}[1/h_n](x)-f_T(x)|\}$ as a function of the cut-off parameter $1/h_n$. Right: objective function $d(l)$ as a function of $1/h_n^l$.

This needs to be adjusted if T is known to likely assume values greater 10. Other loss functions may also be considered instead of (3.9.2) and $d(l)$.
Based on our guess $1/h_n \in (0.7, 2.2)$ from above, we take $1/h_n^l = 0.1 \times l$ for $l = 1, ..., 50$ for our simulation study. Right-hand side of Figure 2 shows the objective function $d(l)$ to assume its minimum at $1/h_n^l \approx 2.5$.

In order to test the performance of $\tilde{f}_n$ we let it compete against the oracle f_n^o associated with the oracle choice $1/h_n^\star$ of $1/h_n$. We compute each estimate based on 100 independent samples of BES_T of size $n \in \{1000, 5000, 10000, 50000\}$. On the left-hand side of Figure 3 we see the box-plot of the loss

$$\sup_{x\in\mathbb{R}_+}\{|f_n^o(x)-f_T(x)|\} \approx \sup_{x\in\{0.1,0.11,\ldots,9.99,10\}}\{|f_n^o(x)-f_T(x)|\}$$

produced by f_n^o and on the right-hand side the corresponding loss

$$\sup_{x\in\mathbb{R}_+}\{|\tilde{f}_n(x)-f_T(x)|\} \approx \sup_{x\in\{0.1,0.11,\ldots,9.99,10\}}\{|\tilde{f}_n(x)-f_T(x)|\}$$

associated with the adaptive estimator. A comparison of the two suggests that the performance of $\tilde{f}_n$ is acceptable in this setting.

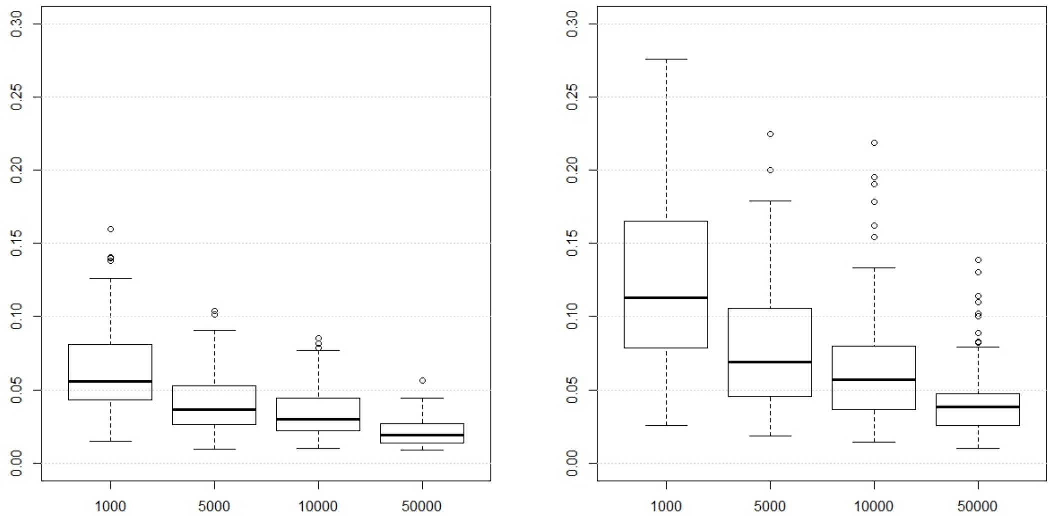

FIGURE 3. Left: box plots of the loss $\sup_{x\in\mathbb{R}_+}\{|f_n^o(x)-f_T(x)|\}$ for different sample sizes. Right: box plots of the loss $\sup_{x\in\mathbb{R}_+}\{|\tilde{f}_n(x) - f_T(x)|\}$ for different sample sizes.

Let us now demonstrate the performance of our adaptive estimator for different distributions of T. As examples we consider Exponential, Gamma, Inverse-Gaussian, Weibull, Log-Normal, and F distributions. To construct the estimate (3.7.2) we choose $d = 5$, $\gamma = 0.7$, and $1/h_n = 1/h_n^{l^\star}$. Figure 4 shows the densities of the six distributions and their 50 respective estimates based on 50 independent samples of BES_T of size $n = 500$.
We can see that the error is particularly large in the neighborhood of 0. That is because our estimator is not defined in 0 and $\lim\limits_{x\to 0}\hat{f}_n(x)$ does not exist for fixed n. Note also that the variance of our estimator is large for small x (see (3.8.21)).

3.10. Some Other Self-Similar Processes

After we have extensively discussed the properties of the estimator constructed in Section 3.2 in the context of Bessel processes, we now turn to other examples of processes that allow the estimation procedure of Section 3.2.

3.10.1. Normally Distributed Processes. Let $Y = (Y_t)_{t\geq 0}$ be a self-similar process with scaling parameter $H > 0$, càdlàg paths and Y_1 standard normally distributed. As an example consider a fractional Brownian motion mentioned in the beginning of Chapter 3. By [**8**], this is a self-similar Gaussian process that admits a version whose sample paths are càdlàg. Let $T \geq 0$ be a stopping time

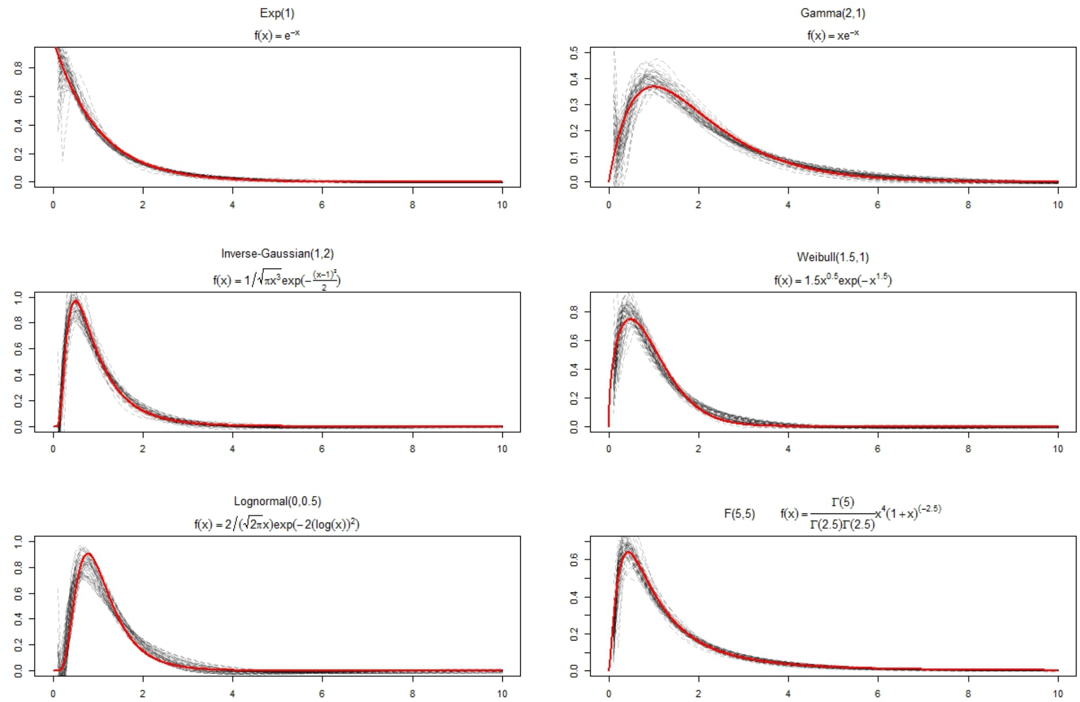

FIGURE 4. Estimated densities (red) and their 50 respective estimates (grey) for the sample size $n = 1000$

with density f_T independent of Y. We wish to estimate f_T non-parametrically based on i.i.d. samples $X_1, \dots, X_n$ of Y_T.

REMARK 3.10.1. *We can easily generalize the setting above to the case where $Y_1 \sim \mathcal{N}(0, \sigma^2)$ with $\sigma^2 > 0$ by considering the process $(\tilde{Y}_t)_{t\geq 0} := (Y_t/\sigma)_{t\geq 0}$ (that is again self-similar with scaling parameter H) and modifying our observations to $\tilde{X}_i := X_i/\sigma$ with the result $\tilde{X}_i \overset{d}{=} \tilde{Y}_T$ for $i = 1, ..., n$.*

Taking $d = 1$ in Example 2.2.5 gives

$$\mathcal{M}[|Y_1|](s) = \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{s}{2}\right) 2^{\frac{s+1}{2}}, \quad \mathrm{Re}(s) > 0.$$

Thus, estimator (3.2.6) assumes the form

$$\hat{f}_n(x) = \frac{1}{2\sqrt{\pi}} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\frac{1}{n}\sum_{k=1}^n X_k^{(\gamma-1+iv)/H}}{\Gamma\left(\frac{\gamma+H-1+iv}{2H}\right) 2^{\frac{\gamma+2H-1+iv}{2H}}} x^{-\gamma-iv}\, dv \tag{3.10.1}$$

for $x > 0$ and $\max\{1-H, a\} < \gamma < b$. We can prove a convergence result for this estimator, similar to Theorem 3.7.3. Recall that in this thesis we consider convergence with respect to the weighted mean squared risk

$$\mathrm{MSE}_\gamma(x) = x^{2\gamma}\,\mathrm{E}[|f_T(x) - \hat{f}_n(x)|^2]$$

from the introduction to Section 3.3.

THEOREM 3.10.2. *Let $f_T \in \mathfrak{M}_{(a,b)}$ for some $0 \le a < b$.*

(i) If $T \in \mathcal{C}(\beta, L, \gamma)$ for some $\beta, L > 0$ and $\gamma \in (\max\{a, 1-H, 3/4\}, b)$, then

$$\mathrm{MSE}_\gamma(x) \lesssim \begin{cases} \frac{1}{n} e^{\pi/(2Hh_n)} + e^{-2\beta/h_n}, & \text{if } \gamma \ge 1 \\ \frac{1}{n} h_n^{(\gamma-1)/H} e^{\pi/(2Hh_n)} + e^{-2\beta/h_n}, & \text{if } \gamma < 1 \end{cases} \tag{3.10.2}$$

for $n \to \infty$ and all $x > 0$. If

$$h_n = \begin{cases} \frac{\pi + 4H\beta}{2H \log n}, & \text{if } \gamma \ge 1 \\ \frac{\pi + 4H\beta}{2H \log n - 2(\gamma-1)\log\log n}, & \text{if } \gamma < 1 \end{cases}, \tag{3.10.3}$$

then we obtain the polynomial convergence rate

$$\sqrt{\mathrm{MSE}_\gamma(x)} \lesssim \begin{cases} n^{-\frac{2H\beta}{\pi + 4H\beta}}, & \text{if } \gamma \ge 1 \\ n^{-\frac{2H\beta}{\pi + 4H\beta}} (\log n)^{\frac{(1-\gamma)2\beta}{\pi + 4H\beta}}, & \text{if } \gamma < 1 \end{cases} \tag{3.10.4}$$

for $n \to \infty$ and all $x > 0$.

(ii) If $T \in \mathcal{D}(\beta, L, \gamma)$ for some $\beta, L > 0$ and $\gamma \in (\max\{a, 1-H, 3/4\}, b)$, then

$$\mathrm{MSE}_\gamma(x) \lesssim \begin{cases} \frac{1}{n} e^{\pi/(2Hh_n)} + h_n^{2\beta}, & \text{if } \gamma \ge 1 \\ \frac{1}{n} h_n^{(\gamma-1)/H} e^{\pi/(2Hh_n)} + h_n^{2\beta}, & \text{if } \gamma < 1 \end{cases} \tag{3.10.5}$$

for $n \to \infty$ and all $x > 0$. If

$$h_n = \begin{cases} \frac{\pi/(2H)}{\log n - 2\beta \log\log n}, & \text{if } \gamma \ge 1 \\ \frac{\pi/(2H)}{\log n - (2\beta + (\gamma-1)/H)\log\log n}, & \text{if } \gamma < 1 \end{cases}, \tag{3.10.6}$$

then we obtain the logarithmic convergence rate

$$\sqrt{\mathrm{MSE}_\gamma(x)} \lesssim (\log n)^{-\beta} \tag{3.10.7}$$

for $n \to \infty$ and all $x > 0$.

PROOF. (i) Let $T \in \mathcal{C}(\beta, L, \gamma)$. The upper bound on variance from Lemma 3.6.1 leads to

$$\mathrm{Var}[x^\gamma \hat{f}_n(x)] \le \frac{C_0(\gamma, H)}{4\pi^2 n} \left(\int_{-1/h_n}^{1/h_n} \frac{1}{|\mathcal{M}[|Y_1|]((\gamma + H - 1 + iv)/H)|} dv \right)^2, \tag{3.10.8}$$

where

$$\begin{aligned} C_0(\gamma, H) &:= \mathcal{M}[|Y_1|](4\gamma - 3)\mathcal{M}[T](2\gamma - 1) \\ &= \frac{1}{\sqrt{\pi}}\Gamma\left(\frac{4\gamma - 3}{2}\right) 2^{2\gamma-1}\mathcal{M}[T](2\gamma - 1) \end{aligned}$$

is finite by the assumption $\gamma \geq 3/4$.
From (3.10.8) and Lemma 2.4.4 it follows that

$$\begin{aligned} \operatorname{Var}[x^\gamma \hat{f}_n(x)] &\leq \frac{C_0(\gamma, H)}{4\pi n} \left(\int_{-1/h_n}^{1/h_n} \left| \frac{1}{\Gamma\left(\frac{\gamma+H-1}{2H} + i\frac{v}{2H}\right)} 2^{-\frac{\gamma+2H-1}{2H} - i\frac{v}{2H}} \right| dv \right)^2 \\ &\leq \frac{C_0(\gamma, H)H^2}{\pi 2^{\frac{\gamma+2H-1}{2H}} n} \left(\int_{-1/(2Hh_n)}^{1/(2Hh_n)} \left| \frac{1}{\Gamma\left(\frac{\gamma+H-1}{2H} + iv\right)} \right| dv \right)^2 \\ &\leq \frac{C_0(\gamma, H)H^2}{\pi 2^{\frac{\gamma+2H-1}{2H}} n} \times \begin{cases} (C_1(\gamma, H) + C_2 e^{\frac{\pi}{4Hh_n}})^2, & \text{if } \gamma \geq 1 \\ C_1^2 \left(2Hh_n\right)^{\frac{\gamma-1}{H}} e^{\frac{\pi}{2Hh_n}}, & \text{if } \gamma < 1 \end{cases}. \end{aligned}$$

Combining this bound on the variance with the bound on the bias from Lemma 3.5.2(i) we obtain (3.10.2). Plugging (3.10.3) into (3.10.2) gives the rate (3.10.4).
(ii) For $T \in \mathcal{D}(\beta, L, \gamma)$ we obtain the same bound on the variance as in (i). For the bias we take the bound obtained in Lemma 3.5.2(ii). Adding bias and variance gives (3.10.5). Plugging (3.10.6) into (3.10.5) gives the rate (3.10.7). □

Taking $H = 1/2$ in Theorem 3.10.2 we obtain the same rates as for Bessel processes (see Theorem 3.7.3). For smaller H the rate is worse and for greater H it is better. Note that we work with observations of $|Y_T|$ rather than Y_T. In Chapter 6 we will show that the rates (3.10.4) and (3.10.7) are optimal for $H \in (0, 2)$ in a sense to be specified there, but only if $|Y_T|$ is observed. We do not know if the rates can be improved by observing Y_T directly or for $H \geq 2$.

3.10.2. Gamma Distributed Processes. Let $Y = (Y_t)_{t\geq 0}$ be a self-similar process with scaling parameter $H > 0$, càdlàg paths and Y_1 Gamma-distributed with shape parameter $\sigma > 0$ and rate parameter $r = 1$, i.e Y_1 has Lebesgue density

$$f_1(x) = \frac{x^{\sigma-1}e^{-x}}{\Gamma(\sigma)}, \quad x \geq 0. \tag{3.10.9}$$

Let $T \geq 0$ be a stopping time with density f_T independent of Y. Again, the aim is to estimate f_T non-parametrically based on i.i.d. samples $X_1, \dots, X_n$ of Y_T.

Remark 3.10.3. *We can easily generalize the setting above to the case, where Y_1 is Gamma-distributed with shape parameter $\sigma > 0$ and rate parameter $r > 0$,* *i.e.* Y_1 *has density*

$$f_1(x) = \frac{r^\sigma}{\Gamma(\sigma)} x^{\sigma-1} e^{-rx}, \quad x \geq 0. \tag{3.10.10}$$

We can reduce this model to the previous case $r = 1$ *by considering the process* $(\tilde{Y}_t)_{t\geq 0} := (rY_t)_{t\geq 0}$ *and modifying our observations to* $\tilde{X}_i := rX_i$ *with the result that* $\tilde{X}_i \stackrel{d}{=} \tilde{Y}_T$ *for* $i = 1, ..., n$.

As an example consider the so-called *square of a Bessel process* with dimension d starting at 0. This is a self-similar $\mathbb{R}_+$-valued process with scaling parameter $H = 1$ and continuous paths. It is equal in distribution to the square of a Bessel process with dimension d starting at 0. See [**31**, Chapter XI, §1] for further details about these processes. If we choose $r = 1/2$ and $\sigma = d/2$ for $d \geq 1$, then the density of a squared Bessel process with dimension d at time $t = 1$ is given by (3.10.10).
Let us present the estimation procedure. From Example 2.2.5 we have

$$\mathcal{M}[|Y_1|](s) = \frac{\Gamma(s+\sigma-1)}{\Gamma(\sigma)} \quad \mathrm{Re}(s) > \sigma - 1.$$

Thus, estimator (3.2.6) takes the form

$$\hat{f}_n(x) = \frac{\Gamma(\sigma)}{2\pi} \int_{-\frac{1}{h_n}}^{\frac{1}{h_n}} \frac{\frac{1}{n}\sum_{k=1}^n X_k^{(\gamma-1+iv)/H}}{\Gamma\left(\frac{\sigma H+\gamma-1+iv}{H}\right)} x^{-\gamma-iv}\, dv \tag{3.10.11}$$

for $x > 0$ and $\max\{1-\sigma H, a\} < \gamma < b$. We can prove a convergence result for this estimator, that is similar to Theorems 3.7.3 and 3.10.2.

THEOREM 3.10.4. *Let* $f_T \in \mathfrak{M}_{(a,b)}$ *for some* $0 \leq a < b$.

(i) If $T \in \mathcal{C}(\beta, L, \gamma)$ *for some* $\beta, L > 0$ *and* $\gamma \in (\max\{a, 1-\sigma H, 1-\sigma/4\}, b)$ *with* $2\gamma - 1 \in (a, b)$, *then*

$$\mathrm{MSE}_\gamma(x) \lesssim \begin{cases} \frac{1}{n} e^{\frac{\pi}{Hh_n}} + e^{-2\beta/h_n}, & \text{if } \gamma \geq 1-\sigma H + H/2 \\ \frac{1}{n} h_n^{\frac{2(\gamma+\sigma H-1)}{H}-1} e^{\frac{\pi}{Hh_n}} + e^{-2\beta/h_n}, & \text{if } \gamma < 1-\sigma H + H/2 \end{cases} \tag{3.10.12}$$

for $n \to \infty$ *and all* $x > 0$. *We choose*

$$h_n = \begin{cases} \frac{\pi/H+2\beta}{\log n}, & \text{if } \gamma \geq 1-\sigma H + H/2 \\ \frac{\pi/H+2\beta}{\log n - \left(1-\frac{2(\gamma+\sigma H-1)}{H}\right)\log\log n}, & \text{if } \gamma < 1-\sigma H + H/2 \end{cases} \tag{3.10.13}$$

to obtain the rate

$$\sqrt{\mathrm{MSE}_\gamma(x)} \lesssim \begin{cases} n^{-\frac{\beta}{\pi/H+2\beta}}, & \text{if } \gamma \geq 1-\sigma H + H/2 \\ n^{-\frac{\beta}{\pi/H+2\beta}} (\log n)^{-\frac{\beta}{\pi/H+2\beta}\left(\frac{2(\gamma+\sigma H-1)}{H}-1\right)}, & \text{if } \gamma < 1-\sigma H + H/2 \end{cases} \tag{3.10.14}$$

for $n \to \infty$ and all $x > 0$.

(ii) If $T \in \mathcal{D}(\beta, L, \gamma)$ for some $\beta, L > 0$ and $\gamma \in (\max\{a, 1-\sigma H, 1-\sigma/4\}, b)$ with $2\gamma - 1 \in (a, b)$, then

$$\mathrm{MSE}_\gamma(x) \lesssim \begin{cases} \frac{1}{n} e^{\frac{\pi}{Hh_n}} + h_n^{2\beta}, & \text{if } \gamma \geq 1-\sigma H + H/2 \\ \frac{1}{n} h_n^{\frac{2(\gamma+\sigma H-1)}{H}-1} e^{\frac{\pi}{Hh_n}} + h_n^{2\beta}, & \text{if } \gamma < 1-\sigma H + H/2 \end{cases} \tag{3.10.15}$$

for $n \to \infty$ and all $x > 0$. Choosing

$$h_n = \begin{cases} \frac{\pi/H}{\log n - 2\beta \log\log n}, & \text{if } \gamma \geq 1-\sigma H + H/2 \\ \frac{\pi/H}{\log n - (2\beta+1-2(\gamma+\sigma H-1)/H)\log\log n}, & \text{if } \gamma < 1-\sigma H + H/2 \end{cases} \tag{3.10.16}$$

yields the rate

$$\sqrt{\mathrm{MSE}_\gamma(x)} \lesssim (\log n)^{-\beta} \tag{3.10.17}$$

for $n \to \infty$ and all $x > 0$.

PROOF. (i) Let $T \in \mathcal{C}(\beta, L, \gamma)$. The upper bound on variance from Lemmas 3.6.1 and 2.4.4 give

$$\begin{aligned} \mathrm{Var}[x^\gamma \hat{f}_n(x)] &\leq \frac{C_0(\gamma, H, \sigma)\Gamma^2(\sigma)}{4\pi^2 n} \left(\int_{-1/h_n}^{1/h_n} \left| \frac{1}{\Gamma\left(\frac{\gamma+H-1+iv}{H} + \sigma - 1\right)} \right| dv \right)^2 \\ &\leq \frac{C_0(\gamma, H, \sigma) H^2 \Gamma^2(\sigma)}{4\pi^2 n} \left(\int_{-1/(Hh_n)}^{1/(Hh_n)} \left| \frac{1}{\Gamma\left(\frac{\gamma+\sigma H-1}{H} + iv\right)} \right| dv \right)^2 \\ &\lesssim \begin{cases} \frac{1}{n} e^{\frac{\pi}{Hh_n}}, & \text{if } \gamma \geq 1-\sigma H + H/2 \\ \frac{1}{n} h_n^{\frac{2(\gamma+\sigma H-1)}{H}-1} e^{\frac{\pi}{Hh_n}}, & \text{if } \gamma < 1-\sigma H + H/2 \end{cases}, \end{aligned}$$

where

$$C_0(\gamma, H, \sigma) := \frac{\Gamma(4\gamma + \sigma - 4)}{\Gamma(\sigma)} \mathcal{M}[T](2\gamma - 1)$$

is finite by the assumption $\gamma \geq 1 - \sigma/4$ and $2\gamma - 1 \in (a, b)$.
Combining this bound on the variance with the bound on the bias from Lemma 3.5.2(i) we obtain (3.10.12). And plugging (3.10.13) into (3.10.12) gives the rate (3.10.14).
(ii) For $T \in \mathcal{D}(\beta, L, \gamma)$ we obtain the same bound on the variance as in (i). For the bias we take the bound obtained in Lemma 3.5.2(ii). Adding bias and variance gives (3.10.15). And plugging (3.10.16) into (3.10.15) gives the rate (3.10.17). □

In Chapter 6 we will discuss optimality of the rates (3.10.14) and (3.10.17). As in Theorem 3.10.2 we see again that greater scaling parameters H yield better convergence rates (at least for stopping times in $\mathcal{C}(\beta, L, \gamma)$). Comparing the rate

$n^{-\frac{\beta}{\pi/H+2\beta}}$ in (3.10.14) with $n^{-\frac{\beta}{\pi/(2H)+2\beta}}$ which is our result for norms of normally distributed processes in an analogous setting (see Theorem 3.10.2) we observe that the rate achieved for norms of Gaussian processes is the better one.

CHAPTER 4

Processes Associated with a Convolution Semi-group

In this chapter we generalize the problem of estimating a stopping time based on observations of a stopped random process to the case of *Lévy processes on noncompact Sturm-Liouville hypergroups*. These processes include the already discussed Bessel processes, but do not coincide with classical Lévy processes. Instead, all of the above are special cases of so-called *Lévy processes on commutative hypergroups*. A classical Lévy process with state space $\mathbb{R}^d$ ($d \in \mathbb{N}$) is characterized by its stationary independent increments. If we try to generalize this notion to processes on a hypergroup K as their state space, we encounter a problem. As the difference of two elements of K is generally not defined, it is unclear how to interpret an increment $X_t - X_s$ for $s < t$ of a random process $X = (X_t)_{t \geq 0}$ with state space K. To overcome this problem we use the characterization of a Lévy process via its associated convolution semi-group (with respect to the convolution "$*$" associated with K), i.e. we describe X as a time-homogeneous Markov process with transition probabilities

$$P(X_t \in A | X_s = x) = (\mu_{t-s} * \delta_x)(A) \tag{4.0.1}$$

for $t < s$, $x \in K$, Borel sets $A \subseteq K$ and some family of probability measures $(\mu_t)_{t \geq 0}$ on K.
Throughout this chapter we roughly follow the outlines in [**10**] and [**30**] to provide a basic understanding of hypergroups and the processes associated with them.

4.1. Introduction to Hypergroups

We refer the reader to [**10**] or [**22**] for a thorough introduction into the topic of this section. Here we only summarize definitions and facts necessary for our purposes.
A hypergroup $(K, *)$ consists of a locally compact Hausdorff space K and a map $* : M_b(K) \times M_b(K) \to M_b(K)$ that is bilinear and associative on the Banach space $M_b(K)$ of all bounded regular Borel measures on K satisfying the following conditions

(i) For all $x, y \in K$ we have $\delta_x * \delta_y \in M^1(K)$ and $\text{supp}(\delta_x * \delta_y)$ is compact;
(ii) The mapping $K \times K \to M^1(K)$, $(x, y) \mapsto \delta_x * \delta_y$ is continuous with respect to the weak topology on $M^1(K)$;

(iii) The mapping $(x,y) \mapsto \mathrm{supp}(\delta_x * \delta_y)$ is continuous from $K \times K$ into the space of compact subsets of K with respect to the Michael topology (cf. [**10**, Section 1.1] and [**26**]);
(iv) There exists a (necessarily unique) neutral element $e \in K$ such that $\delta_e * \delta_x = \delta_x * \delta_e = \delta_x$ for all $x \in K$;
(v) There exists a (necessarily unique) involution (a homeomorphism $x \mapsto x^-$ of K onto itself with the property $(x^-)^- = x$ for all $x \in K$) such that $(\delta_x * \delta_y)^- = \delta_{y^-} * \delta_{x^-}$ for all $x, y \in K$ where μ^- denotes the image of μ under involution;
(vi) For $x, y \in K$, $e \in \mathrm{supp}(\delta_x * \delta_y)$ if and only if $x = y^-$.

Once the convolution of two point measures is established the convolution of arbitrary measures $\mu, \nu \in M_b(K)$ is necessarily given by

$$(\mu * \nu)(A) = \iiint \mathbb{1}_A(z)\ d(\delta_x * \delta_y)(z) d\mu(x) d\nu(y) \tag{4.1.1}$$

for $A \in \mathcal{B}(K)$ and integration with respect to $\mu * \nu$ is determined as

$$\int_K f(z)\ d(\mu * \nu)(z) = \iiint f(z)\ d(\delta_x * \delta_y)(z) d\mu(x) d\nu(y)$$

for $f \in C_c(K)$. We call $(K, *)$ *commutative*, if $*$ is commutative. $(K, *)$ is called *hermitian*, if $x^- = x$ for all $x \in K$. We often denote a hypergroup $(K, *)$ briefly by K.

DEFINITION 4.1.1. *Let $(K, *)$ be a commutative hypergroup.*
(i) A nontrivial measure ω on K is called Haar measure if $\delta_x * \omega = \omega$ for all $x \in K$.
(ii) The dual space $\widehat{K}$ of K is defined by

$$\widehat{K} := \{\alpha \in C_b(K) | \alpha \not\equiv 0,\ \alpha(x * y^-) = \alpha(x)\overline{\alpha(y)} \text{ for all } x, y \in K\},$$

where $C_b(K)$ is the class of continuous bounded functions $K \to \mathbb{C}$, $\overline{\alpha(y)}$ denotes the complex conjugate of $\alpha(y)$ and

$$\alpha(x * y^-) := \int_K \alpha(z) d(\delta_x * \delta_{y^-})(z).$$

(iii) The Fourier transforms of $f \in \mathcal{L}^1(K, \omega)$ and $\mu \in M_b(K)$ are given by

$$\hat{f}(\alpha) := \int_K \overline{\alpha(x)} f(x) d\omega(x) \qquad \text{and} \qquad \hat{\mu}(\alpha) := \int_K \bar{\alpha} d\mu$$

for $\alpha \in \widehat{K}$. If a random variable X has distribution μ, we will denote its Fourier transform by $\mathcal{F}_r[X] := \hat{\mu}$.

$\widehat{K}$ is endowed with the topology of uniform convergence on compact sets and it is a locally compact Hausdorff space with respect to this topology. There is a unique measure π on $\widehat{K}$ such that the Fourier transform on $\mathcal{L}^1(K, \omega) \cap \mathcal{L}^2(K, \omega)$ extends uniquely to an isometric isomorphism between $\mathcal{L}^2(K, \omega)$ and $\mathcal{L}^2(\widehat{K}, \pi)$.

In particular, the Fourier transform is unique for bounded measures. We call π *Plancherel measure*. Notice that $\mathrm{supp}(\pi)$ is sometimes a proper subset of $\widehat{K}$. See [**10**] for proofs of the facts in this paragraph.

THEOREM 4.1.2. *If $(K,*)$ is a commutative hypergroup, then*

$$\widehat{(\mu*\nu)}(\alpha)=\hat{\mu}(\alpha)\hat{\nu}(\alpha)$$

for all $\mu,\nu\in M_b(K)$ and $\alpha\in\widehat{K}$.

PROOF. Let $\alpha\in\widehat{K}$. For $\mu,\nu\in M_b(K)$

$$\begin{aligned}\widehat{(\mu*\nu)}(\alpha)&=\int_K\bar{\alpha}(z)\ d(\mu*\nu)(z)\\&=\int_K\int_K\int_K\bar{\alpha}(z)\ d(\delta_x*\delta_y)(z)d\mu(x)d\nu(y)\\&=\int_K\bar{\alpha}(x)\ d\mu(x)\int_K\bar{\alpha}(y)\ d\nu(y)\\&=\hat{\mu}(\alpha)\hat{\nu}(\alpha),\end{aligned}$$

which shows that the Fourier transform is multiplicative. □

Let us consider some examples of hypergroups.

EXAMPLE 4.1.3. *Every locally compact group $(G,\cdot)$ is a hypergroup where the convolution is defined by*

$$(\mu*\nu)(A):=\iint\mathbb{1}_A(x\cdot y)d\mu(x)d\nu(y)$$

for $\mu,\nu\in M^b(G)$ and any Borel set A.

EXAMPLE 4.1.4. *Consider the group $(\mathbb{R}_+,\cdot)$ with the canonical topology of open sets on $\mathbb{R}_+$ and multiplication of real numbers as group operation. Define the convolution $*:=\odot$ by*

$$(\mu\odot\nu)(A)=\iint\mathbb{1}_A(xy)d\mu(x)d\nu(y)$$

*for $\mu,\nu\in M_b(\mathbb{R}_+)$, $A\in\mathcal{B}(\mathbb{R}_+)$. The operation $\odot$ is called Mellin convolution (cf. [**12**]). The pair $(\mathbb{R}_+,\odot)$ constitutes a commutative hypergroup with the following (easy to check) properties:*

(i) *The Haar measure is given by $\omega:=\frac{1}{t}dt$, where dt represents the Lebesgue measure on $\mathbb{R}_+$;*
(ii) *The dual space is identifiable with*

$$\widehat{\mathbb{R}}_+:=\{\varphi_\alpha|\alpha\in i\mathbb{R}\},$$

where $\varphi_\alpha:\mathbb{R}_+\to\mathbb{C}$, $\varphi_\alpha(x)=x^{-\alpha}$;

(iii) *The Fourier transforms of* $f \in \mathcal{L}^1(\mathbb{R}_+, \frac{1}{t}dt)$ *and* $\mu \in M_b(\mathbb{R}_+)$ *are given by their Mellin transforms*

$$\hat{f}(\alpha) = \textstyle\int_{\mathbb{R}_+} x^{\alpha-1} f(x)dx \qquad \text{and} \qquad \hat{\mu}(\alpha) := \textstyle\int_{\mathbb{R}_+} x^{\alpha} d\mu(x).$$

See Section 2.2 for further details about Mellin transforms.

We are now ready to introduce the notion of a Lévy process that is associated with a convolution semi-group on a hypergroup.

4.2. Lévy Processes on Commutative Hypergroups

Lévy processes on commutative hypergroups constitute a large class of processes including classical Lévy processes as described for example in [**32**] as well as Bessel processes.

Definition 4.2.1. *Let* $(\mu_t)_{t\geq 0}$ *be a family of probability measures on a hypergroup* K.

(i) $(\mu_t)_{t\geq 0}$ *is called a convolution semi-group on* K, *if* $\mu_s * \mu_t = \mu_{s+t}$ *for all* $s, t \geq 0$ *with* $\mu_0 = \delta_\varepsilon$, *and if* $\mu_t \to \delta_\varepsilon$ *weakly as* $t \to 0$.

(ii) K*-valued Markov process* $Y = (Y_t)_{t\geq 0}$ *is called a Lévy process on* K *associated with* $(\mu_t)_{t\geq 0}$, *if its transition probabilities satisfy*

$$P(X_t \in A | X_s = x) = (\mu_{t-s} * \delta_x)(A)$$

for all $0 \leq s \leq t,\ x \in K$ *and* $A \in \mathcal{B}(K)$.

By Lemma 4.1.2 and the uniqueness of the Fourier transform, the condition $\mu_s * \mu_t = \mu_{s+t}$ is equivalent to $\hat{\mu}_s \hat{\mu}_t = \hat{\mu}_{s+t}$ for all $s, t \geq 0$.
In the following we use the term *classical Lévy process* for processes associated with a classical convolution semi-group on $\mathbb{R}^d$ (in the sense of [**32**]). By *Lévy processes* we always mean the general concept of Definition 4.2.1. By [**30**] such processes are Feller processes and admit an equivalent càdlàg version. Thus, in the following we can assume without loss of generality that a Lévy process $Y = (Y_t)_{t\geq 0}$ on a hypergroup $(K, *)$ has càdlàg paths. And so it is feasible to consider Y_T for a random time T as a random variable and do statistics based on observations of it.
A convolution semi-group (and the associated Lévy process) on a given hypergroup can be characterized by its *negative definite function*. There are various definitions of this concept (cf. [**11**]) of which we will provide two.

Definition 4.2.2. *A locally bounded measurable function* $\psi : \widehat{K} \to \mathbb{C}$ *is called strongly negative definite, if* $\psi(\mathbb{1}_K) \geq 0$ *and if there is a* $\mu_t \in M_b(K)$ *with* $\exp(-t\psi) = \hat{\mu}_t$ *for all* $t > 0$. *We denote the class of these functions by* $N_B^{(s)}(\widehat{K})$.

For the second definition denote the class of compactly supported measures on $\widehat{K}$ by $\mathcal{M}_c(\widehat{K})$ and the class of compactly supported continuous functions on $\widehat{K}$ by $C_c(\widehat{K})$.

DEFINITION 4.2.3. *Let* $T_2 := \{c\delta_{\mathbb{1}_K} + g\pi \in \mathcal{M}_c(\widehat{K}) | c \in \mathbb{C},\ g \in C_c(\widehat{K})\}$. *A continuous function* $\psi : \widehat{K} \to \mathbb{C}$ *is called weakly negative definite, if*

(i) *For all* $\mu \in T_2$ *with* $\check{\mu} \geq 0$ *and* $\check{\mu}(e) = 0$ *we have* $\int_{\widehat{K}} \psi d\mu \leq 0$;

(ii) $\psi(\mathbb{1}_K) \geq 0$ *and* $\psi(\bar{\alpha}) = \overline{\psi(\alpha)}$ *hold for all* $\alpha \in \widehat{K}$.

We denote the class of these functions by $N_{T_2}^{(w)}(\widehat{K})$.

By [**11**] we have $N_B^{(s)}(\widehat{K}) \subseteq N_{T_2}^{(w)}(\widehat{K})$ but the two classes coincide if their elements are restricted to $\mathrm{supp}(\pi)$. That is to say the following:

THEOREM 4.2.4. *If* $(\mu_t)_{t\geq 0}$ *is a convolution semi-group on* $(K, *)$, *then there is a unique* $\psi \in N_B^{(s)}(\widehat{K})$ *with* $\hat{\mu}_t = e^{-t\psi}$ *for all* $t > 0$. *Conversely, if we have a* $\psi \in N_{T_2}^{(w)}(\widehat{K})$, *then there is a unique convolution semi-group* $(\mu_t)_{t\geq 0}$ *on* $(K, *)$ *with* $\hat{\mu}_t\big|_{\mathrm{supp}(\pi)} = e^{-t\psi}\big|_{\mathrm{supp}(\pi)}$ *for all* $t > 0$.

Theorem 4.2.4 is called *Schoenberg correspondence* (cf. [**10**] and [**11**]). It shows a one-to-one correspondence between negative definite functions and a convolution semi-groups. It was first established by [**38**, Theorem 3.7].

4.3. Noncompact Sturm-Liouville Hypergroups

In this section we consider Lévy processes that are associated with a hypergroup structure on $K = \mathbb{R}_+$. All known examples of commutative hypergroups on $\mathbb{R}_+$ allow a representation as a so-called *Sturm-Liouville hypergroup* (see [**10**, page 201]). We will refer to Lévy processes associated to these hypergroups as *Sturm-Liouville processes*. Every Sturm-Liouville hypergroup is characterized by a so-called *Sturm-Liouville function*, that is a continuous mapping $A : \mathbb{R}_+ \to \mathbb{R}$ which is differentiable and strictly positive on $(0, \infty)$.
For a Sturm-Liouville function A we define the *Sturm-Liouville operator* L on $\mathcal{C}^2((0, \infty))$ by

$$Lf := L_A f := -f'' - \frac{A'}{A} f'.$$

Using these notations we adopt the following definition from [**41**]:

DEFINITION 4.3.1. *Let* $(\mathbb{R}_+, *)$ *be a hypergroup with*

$$\mathrm{supp}(\delta_x * \delta_y) = [|x - y|, x + y] \quad \textit{for } x, y \in \mathbb{R}_+.$$

It will be called Sturm-Liouville hypergroup if for every $f \in \mathcal{C}^\infty(\mathbb{R}_+)$ *the function* $u_f \in \mathcal{C}^\infty(\mathbb{R}_+ \times \mathbb{R}_+)$ *defined by*

$$u_f(x, y) := \int_{\mathbb{R}_+} f d(\delta_x * \delta_y) \quad x, y \in \mathbb{R}_+$$

is two times differentiable and satisfies the partial differential equation

$$L_x u_f(x, y) = L_y u_f(x, y), \quad u_x(0, y) = 0 \quad \textit{for } x, y > 0,$$

where the indexes refer to parital derivatives with respect to the respective indexes.

For the following suppose that A allows the representation

$$\frac{A'}{A}(x) = \frac{\alpha_0}{x} + \alpha_1(x) \tag{4.3.1}$$

for all x in a neighborhood of 0, where $\alpha_0 \in \mathbb{R}_+$ and $\alpha_1 : \mathbb{R}_+ \to \mathbb{R}$ are such that

- **(SL1)** one of the following statements is true:
  - **(SL1.1)** $\alpha_0 > 0$ and $\alpha_1 \in C^\infty(\mathbb{R})$, where α_1 is an odd function (this implies $A(0) = 0$);
  - **(SL1.2)** $\alpha_0 = 0$ and $\alpha_1 \in C^1(\mathbb{R}_+)$ (this implies $A(0) > 0$).

We assume further that A is such that

- **(SL2)** there is a $\beta \in \mathcal{C}^1(\mathbb{R}_+)$ such that $\beta(0) \geq 0$, $\frac{A'}{A} - \beta$ is nonnegative and decreasing on $(0, \infty)$, and $q := \frac{1}{2}\beta' - \frac{1}{4}\beta^2 + \frac{A'}{2A}\beta$ is decreasing on $(0, \infty)$.

If $\mu(\Omega) \leq 1$ for a measure μ on a measurable space $(\Omega, \mathcal{A})$, then we call μ subprobability measure. The following properties of Sturm-Liouville hypergroups are well-known, for example from [**10**].

THEOREM 4.3.2. *Let* $(\mathbb{R}_+, *(A))$ *be a noncompact Sturm-Liouville hypergroup. Assume additionally* (4.3.1) *and (SL2). Then*

(i) The limit

$$\rho := \frac{1}{2} \lim_{x\to\infty} \frac{A'}{A}(x)$$

exists and is a nonnegative real number. ρ *is called index of* $(\mathbb{R}_+, *(A))$*;*

(ii) $\frac{A'}{A} \geq 0$ *holds on* $(0, \infty)$*;*

(iii) The dual space of $(\mathbb{R}_+, *(A))$ *consists of real-valued functions* φ_λ $(\lambda \in \mathbb{R}_+ \cup i[0, \rho])$*, where* φ_λ *is the unique solution of the initial value problem*

$$L_A\varphi_\lambda = (\lambda^2 + \rho^2)\varphi_\lambda, \;\; \varphi_\lambda(0) = 1, \;\; \varphi_\lambda'(0) = 0;$$

This implies that the Fourier transform associated with $(\mathbb{R}_+, *(A))$ *is given by*

$$\hat{\mu}(\lambda) := \int_{\mathbb{R}_+} \varphi_\lambda(x) \; d\mu(x)$$

for $\mu \in M_b(\mathbb{R}_+)$*;*

(iv) For all $\lambda \in \mathbb{R}_+ \cup i(0, \rho]$ *we have* $||\varphi_\lambda||_\infty \leq 1$*;*

(v) $(\mathbb{R}_+, *(A))$ *allows a Laplace representation, i.e. for all* $x \in \mathbb{R}_+$ *there is a subprobability measure* τ_x *on* $[-x, x]$ *such that for all* $\lambda \in \mathbb{C}$

$$\varphi_\lambda(x) = \int_{-x}^{x} \cos(\lambda t) \; d\tau_x(t).$$

If $\rho = 0$*, then* τ_x *is a probability measure.*

If a random variable X has distribution μ, then we denote the Fourier transform of X by $\mathcal{F}_r[X] := \hat{\mu}$.
Before we give some examples of noncompact Sturm-Liouville hypergroups, we need to introduce some special functions.

DEFINITION 4.3.3. *Let $\alpha > 1/2$. Denote by J_α the Bessel function of the first kind, defined by*

$$J_\alpha(x) := \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\Gamma(k+\alpha+1)} \left(\frac{x}{2}\right)^{2k+\alpha}, \quad x > 0.$$

The modified Bessel function $\Lambda_\alpha : \mathbb{R}_+ \to \mathbb{R}$ is defined by

$$\Lambda_\alpha(x) := \begin{cases} 2^\alpha \Gamma(\alpha+1) x^{-\alpha} J_\alpha(x), & \text{for } x > 0 \\ 1 \ (= \lim\limits_{x\to 0} \Lambda_\alpha(x)), & \text{for } x = 0 \end{cases}. \tag{4.3.2}$$

Both functions J_α and Λ_α are well-known in the literature. See, for example, [**1**, Chapter 9] or [**18**] for their analytical properties.

Let us now list some examples of noncompact Sturm-Liouville hypergroups.

EXAMPLE 4.3.4. *(i) Bessel-Kingman hypergroups (see* [**10**] *and* [**23**]*) with*

$$A(x) := x^{2\alpha-1}, \quad x \in \mathbb{R}_+$$

for some $\alpha > -1/2$.
For point measures δ_x and δ_y with $x, y \in \mathbb{R}_+$ convolution is defined by

$$(\delta_x *_\alpha \delta_y)(A) := \frac{\Gamma(\alpha+1)}{\sqrt{\pi}\Gamma(\alpha+1/2)} \int_0^\pi \mathbb{1}_A(\sqrt{x^2+y^2-2xy\cos(t)}) \sin^{2\alpha}(t)dt$$

for $A \in \mathcal{B}(\mathbb{R}_+)$. For $\mu, \nu \in M_b(\mathbb{R}_+)$ convolution is defined by (4.1.1).
The dual space is given by $\{\varphi_\lambda | \lambda \geq 0\}$, where $\varphi_\lambda(x) := \Lambda_\alpha(\lambda x)$ for $x \geq 0$.
The Fourier transform on a Bessel-Kingman hypergroup of $\mu \in M_b(\mathbb{R}_+)$ is given by

$$\hat{\mu}(\lambda) = \int_{\mathbb{R}_+} \Lambda_\alpha(\lambda x) d\mu(x).$$

(ii) Jacobi hypergroups, where

$$A(x) := \sinh^{2\alpha+1}(x) \cosh^{2\beta+1}(x)$$

for all $x \in \mathbb{R}_+$ and some $\alpha \geq \beta > -\frac{1}{2}$.
(iii) Square hypergroup, where

$$A(x) := (1+x)^2$$

for all $x \in \mathbb{R}_+$.

(iv) Two point support hypergroups, where

$$A(x) := a\cosh^2(bx)$$

for all $x \in \mathbb{R}_+$ and some $a > 0,\ b \geq 0$.

The next Theorem is an analogue to the Lévy-Khinchin formula for classical convolution groups.

THEOREM 4.3.5. *Let $(\mu_t)_{t\geq 0}$ a convolution semi-group on a noncompact Sturm-Liouville hypergroup. Then the Fourier transform of μ_t is given for all $t \geq 0,\ u > 0$ by*

$$\hat{\mu}_t(u) = \exp(-t\psi(u)), \tag{4.3.3}$$

where

$$\psi(\lambda) = \psi(0) + c(\lambda^2 + \rho^2) + \int\limits_{\mathbb{R}_+\setminus\{0\}} (1 - \varphi_\lambda(x))\ d\nu(x) \tag{4.3.4}$$

with a unique constant $c \geq 0$ and a Lévy measure ν (i.e. ν is a measure on $\mathbb{R}_+$ which satisfies $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}_+} \max\{1, x^2\} d\nu(x) < \infty$). We call the couple (c, ν) the characteristic of $(\mu_t)_{t\geq 0}$ (or of a Lévy process associated with $(\mu_t)_{t\geq 0}$).

PROOF. By Theorem 4.2.4 ψ is the unique strongly negative definite function associated with $(\mu_t)_{t\geq 0}$. And by [**11**] ψ allows the representation above. □

REMARK 4.3.6. *If $\rho = 0$, then we have $\psi(0) = 0$. That is because Theorem 4.3.2(iii) implies $\phi_{i\rho} = 1$ (cf.* [**10**, page 223]*) and therefore,*

$$\exp(-\psi(0)) = \hat{\mu}_1(0) = \int_{\mathbb{R}_+} \phi_0(x) d\mu_1(x) = \mu_1(\mathbb{R}_+) = 1,$$

which implies the claim.

LEMMA 4.3.7. *Let $\rho = 0$. If $\int_1^\infty x d\nu(x) < \infty$, then ψ is continuously differentiable and the derivative is given by*

$$\psi'(\lambda) = 2c\lambda - \int\limits_{\mathbb{R}_+\setminus\{0\}} \frac{d}{d\lambda}\varphi_\lambda(x)\ d\nu(x).$$

PROOF. The claim follows from (4.3.4) by interchanging differentiation with integration. This interchange is valid by dominated convergence (see [**14**, page 148]). In order to apply that, note that the derivative $\frac{d}{d\lambda}\varphi_\lambda$ exists by Theorem 4.3.2 on $\mathbb{R}_+$. Moreover, it follows (considering Theorem 4.3.2) that

$$\left|\frac{d}{d\lambda}(1 - \varphi_\lambda(x))\right| = \left|\frac{d}{d\lambda}\int\limits_{-x}^{x} \cos(\lambda t) d\tau_x(t)\right| = \left|\int\limits_{-x}^{x} t\sin(\lambda t) d\tau_x(t)\right| \leq \begin{cases} x^2, & x \leq 1 \\ x, & x > 1 \end{cases}$$

for all $\lambda, x \in \mathbb{R}_+$, where τ_x is some probability measure. This gives us an integrable bound with respect to ν. □

4.4. Lévy Processes with Spherical Symmetry

In this section we discuss the class of multi-dimensional spherically symmetric Lévy processes. These processes are particularly suitable to describe the movement of particles in the experiment described in the introduction. Particles in that experiment move around "freely", that is without any preference for direction, since there is no external force (like gravitation) driving them to one side. This property can be made mathematically precise by the concept of *spherical symmetry.* Let us first discuss the spherical symmetry of a random vector. For this we follow the outlines in [**23**] and [**25**].

For the following definition denote by $\mathrm{SO}(d)$ the set of orthogonal $\mathbb{R}^{d\times d}$-matrices with determinant 1 (that is the set of matrices that act as a rotation).

DEFINITION 4.4.1. *Let $\mu \in M^1(\mathbb{R}^d)$, $d \in \mathbb{N}$. We call μ spherically symmetric (or say μ has spherical symmetry or μ is rotationally invariant), if we have*

$$\mu(A) = \mu(O(A))$$

for all $A \in \mathcal{B}(\mathbb{R}^d)$ and $O \in SO(d)$. A random vector $X = (X_1, ..., X_d)$ is called spherically symmetric, if its distribution is spherically symmetric.

The distribution of a spherically symmetric vector $X = (X_1, ..., X_d)$ is determined by that of its length

$$|X| = \left(\sum_{i=1}^{d} X_i^2\right)^{1/2}.$$

In particular the characteristic function of X is given by

$$\mathcal{F}[X](t) = \mathrm{E}[e^{i<t,X>}] = \mathrm{E}[\Lambda_\alpha(|t||X|)], \tag{4.4.1}$$

where $t = (t_1, ..., t_d) \in \mathbb{R}^d$, $\alpha = d/2 - 1$ and Λ_α is defined by (4.3.2). The proof of this fact can be found in [**25**]. In view of (4.4.1) Lord points out in [**25**] that while investigating spherically symmetric random vectors it is often sufficient to work with their lengths.

Let us now transfer the concept of spherical symmetry to classical Lévy processes. We call a classical d-dimensional Lévy process $Y = (Y_t)_{t\geq 0}$ with the associated convolution semi-group $(\mu_t)_{t\geq 0}$ spherically symmetric, if μ_t has spherical symmetry for each $t \geq 0$. By (4.4.1), such a process satisfies

(4.4.2)

$$\mathrm{E}[\Lambda_\alpha(|r||Y_{s+t}|)] = \mathcal{F}[Y_{s+t}](r) = \mathcal{F}[Y_s](r)\mathcal{F}[Y_t](r) = \mathrm{E}[\Lambda_\alpha(|r||Y_s|)\Lambda_\alpha(|r||Y_t|)]$$

for all $s, t \geq 0$, $r \in \mathbb{R}^d$ and $\alpha = d/2 - 1$. Denote by $|\mu_t|$ the image measure of μ_t under the Euclidean norm ($t \geq 0$). Identity (4.4.2) implies

$$\widehat{|\mu_{s+t}|}(r) = \widehat{|\mu_s|}(r)\widehat{|\mu_t|}(r), \quad s, t, r \in \mathbb{R}_+,$$

where $\widehat{|\mu_t|}(r) = \int_{\mathbb{R}_+} \Lambda_\alpha(rx) d|\mu_t|(x)$ is the Fourier transform of $|\mu_t|$ associated with the Bessel-Kingman hypergroup (see Example 4.3.4(i)), i.e. $(|\mu_t|)_{t\geq 0}$ is a convolution semi-group on the Bessel-Kingman hypergroup. Hence, the process $||Y||_2 := (||Y_t||_2)_{t\geq 0}$ is a Lévy process on the Bessel-Kingmann hypergroup associated with $(|\mu_t|)_{t\geq 0}$. Thus, in order to perform statistical inference for classical d-dimensional spherically symmetric Lévy processes it is sufficient to work with their lengths. This provides us the key to solving multi-dimensional problems like the one described in the introduction.

CHAPTER 5

Estimation for Lévy Processes on Sturm-Liouville Hypergroups

Let $(\mu_t)_{t\geq 0}$ be a convolution semi-group associated with a strongly negative definite function $\psi \in N_B^{(s)}(\widehat{K})$ on a noncompact Sturm-Liouville hypergroup with index $\rho = 0$. Consider the Lévy process $Y = (Y_t)_{t>0}$ associated with $(\mu_t)_{t\geq 0}$ in the sense of Definition 4.2.1. We assume that ψ (and consequently the distribution of Y) is known explicitly by its characterizing parameters c and ν (cf. Theorem 4.3.5). Our goal is to estimate the density f_T of a nonnegative stopping time T independent of Y based on i.i.d. observations $X_1, ..., X_n$ of Y_T. This can be achieved in a similar way to the case of classical Lévy processes described in [**7**].

5.1. Construction of the Estimator

Our approach is based on the following basic fact which connects the Fourier transform $\mathcal{F}_r$ of a random variable X with its classical Laplace transform

$$\mathcal{L}[X](t) := \mathrm{E}[e^{-tX}], \quad t \in \mathbb{R}.$$

LEMMA 5.1.1. *Given the setting above we have*

$$\mathcal{F}_r[Y_T](\lambda) = \mathcal{L}[T](\psi(\lambda))$$

for all $\lambda \in \mathbb{R}_+$.

PROOF. Let T be a random variable on a probability space $(\Omega_1, \mathcal{A}_1, P_1)$ and Y a process on $(\Omega_2, \mathcal{A}_2, P_2)$. (cf. Remark 3.1.1). Denote by P the product measure of P_1 and P_2. Then Theorem 4.2.4 and Fubini's theorem imply

$$\begin{aligned}
\mathcal{F}_r[Y_T](\lambda) &= \int_K \varphi_\lambda(x) dP^{Y_T}(x) \\
&= \int_{\Omega_2}\int_{\Omega_1} \varphi_\lambda(Y_{T(\omega_2)}(\omega_1)) dP_1(\omega_1) dP_2(\omega_2) \\
&= \int_{\Omega_2} \mathcal{F}_r[Y_{T(\omega_2)}](\lambda) dP_2(\omega_2) \\
&= \int_{\Omega_2} e^{-T(\omega_2)\psi(\lambda)} dP_2(\omega_2) \\
&= \mathcal{L}[T](\psi(\lambda))
\end{aligned}$$

for all $\lambda \in \mathbb{R}_+$, where φ_λ are the functions which constitute the dual space of $\mathbb{R}_+$ (see Theorem 4.3.2(iii)). □

In order to construct an estimator for f_T we consider the Mellin transform of the Laplace transform $\mathcal{M}[\mathcal{L}[T]]$ of T. The following lemma describes how this object is related to the Mellin transform of T. This relation is well-known (see, for instance, [**27**, page 3]).

LEMMA 5.1.2. *Let $f_T \in \mathfrak{M}_{(0,1)}$. Then we have*

$$\mathcal{M}[\mathcal{L}[T]](z) = \mathcal{M}[T](1-z) \cdot \Gamma(z)$$

for all $0 < \operatorname{Re}(z) < 1$.

PROOF. Let $0 < \operatorname{Re}(z) < 1$. By $f_T \in \mathfrak{M}_{(0,1)}$ we have

$$\mathrm{E}\left[\int_0^\infty |u^{z-1}e^{-uT}|\ du\right] = \mathrm{E}[|T^{(1-z)-1}|] \int_0^\infty |x^{z-1}e^{-x}|\ dx < \infty,$$

which allows us to interchange the order of integration in (5.1.1) by Fubini's theorem. With the change of variables $x := uT$ in (5.1.2) it follows that

$$\begin{aligned} \mathcal{M}[\mathcal{L}[T]](z) &= \int_0^\infty u^{z-1}\, \mathrm{E}[e^{-uT}]\ du \\ &= \mathrm{E}\left[\int_0^\infty u^{z-1}e^{-uT}\ du\right] \qquad (5.1.1) \\ &= \mathrm{E}\left[\frac{1}{T}\int_0^\infty \left(\frac{x}{T}\right)^{z-1} e^{-x}\ dx\right] \qquad (5.1.2) \\ &= \mathrm{E}[T^{(1-z)-1}] \int_0^\infty x^{z-1}e^{-x}\ dx \\ &= \mathcal{M}[T](1-z)\cdot\Gamma(z), \end{aligned}$$

which is our claim.

□

Another representation of $\mathcal{M}[\mathcal{L}[T]]$ is given by

THEOREM 5.1.3. *Let $(Y_t)_{t\geq 0}$ be a Sturm-Liouville process associated with a strongly negative definite function ψ and characteristic (c,ν). If $\int_1^\infty x d\nu(x) < \infty$*

and $\lim_{u\to\infty} \psi(u) = \infty$*, then*

$$\mathcal{M}[\mathcal{L}[T]](z) = \int_0^\infty \psi(\lambda)^{z-1} \mathcal{F}_r[Y_T](\lambda)\psi'(\lambda)\, d\lambda$$

for $0 < \mathrm{Re}(z) < 1$.

PROOF. The function $u^{z-1}\mathcal{L}[T](u)$ is continuous in u on $(0,\infty)$ and ψ is continuously differentiable on $(0,\infty)$ by Lemma 4.3.7. Hence, by the change of variables $u := \psi(\lambda)$, we have

$$\begin{aligned} \mathcal{M}[\mathcal{L}[T]](z) &= \int_0^\infty u^{z-1}\mathcal{L}[T](u)du \\ &= \int_0^\infty \psi(\lambda)^{z-1}\mathcal{L}[T](\psi(\lambda))\psi'(\lambda)\, d\lambda. \end{aligned} \tag{5.1.3}$$

Note that $\psi(\infty) = \infty$ by assumption and $\psi(0) = 0$ by Remark 4.3.6. (So the integration interval does not change after the change of variables.) The identity (5.1.3) implies the assertion considering $\mathcal{L}[T](\psi(\lambda)) = \mathcal{F}_r[Y_T](\lambda)$ for all $\lambda \geq 0$ by Lemma 5.1.1. □

The assumption $\lim_{u\to\infty} \psi(u) = \infty$ in Theorem 5.1.3 is satisfied if we have $c > 0$.

Provided the requirements of Lemmas 5.1.2 and 5.1.3 are met, we have

$$\mathcal{M}[T](z) = \frac{\mathcal{M}[\mathcal{L}[T]](1-z)}{\Gamma(1-z)} = \frac{\int_0^\infty \psi(\lambda)^{-z}\mathcal{F}_r[Y_T](\lambda)\psi'(\lambda)\, d\lambda}{\Gamma(1-z)} \tag{5.1.4}$$

for $0 < \mathrm{Re}(z) < 1$. By Mellin inversion this implies

$$f_T(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\int_0^\infty \psi(\lambda)^{-\gamma-iv}\mathcal{F}_r[Y_T](\lambda)\psi'(\lambda)\, d\lambda}{\Gamma(1-\gamma-iv)} x^{-\gamma-iv}\, dv. \tag{5.1.5}$$

Using the approximation

$$\mathcal{F}_r[Y_T](\lambda) = \mathrm{E}[\varphi_\lambda(Y_T)] \approx \frac{1}{n}\sum_{k=1}^n \varphi_\lambda(X_k)$$

we define for $\frac{1}{2} < \gamma < 1$:

$$\hat{f}_n(x) := \frac{1}{2\pi}\int_{-U_n}^{U_n} \frac{\int_0^{A_n} \psi(\lambda)^{-\gamma-iv}\psi'(\lambda)\frac{1}{n}\sum_{k=1}^n \varphi_\lambda(X_k)\, d\lambda}{\Gamma(1-\gamma-iv)} x^{-\gamma-iv}\, dv \tag{5.1.6}$$

as an estimator for the density of a stopping time $T \geq 0$ based on samples $X_1, ..., X_n$ of Y_T. $(U_n)_{n\in\mathbb{N}}$ and $(A_n)_{n\in\mathbb{N}}$ are regularizing sequences to be chosen later such that $A_n, U_n \to \infty$ for $n \to \infty$. They (the sequences) make sure that the integrals defining $\hat{f}_n(x)$ converge. The real-valued functions φ_λ which define the dual space of the underlying Sturm-Liouville hypergroup are determined by the initial value problem in Theorem 4.3.2(iii). We assume that they are either known explicitly (like in our example in Section 5.3) or can be calculated numerically.

5.2. Convergence

The goal of this section is to show consistency of estimator (5.1.6), and to derive its convergence rate. The main result will be presented in Theorem 5.2.6. Before we can prove that, we need a number of auxiliary results.

We begin by showing an asymptotic bound on the Fourier transform of a Sturm-Liouville process stopped at a random time.

LEMMA 5.2.1. *Let $(Y_t)_{t\geq 0}$ be a Sturm-Liouville process with associated strongly negative definite function ψ and characteristic (c,ν), $c>0$. If f_T is essentially bounded, then*

$$|\mathcal{F}_r[Y_T](\lambda)| \lesssim \lambda^{-2}$$

holds for $\lambda \to \infty$.

PROOF. Lemma 5.1.1 implies

$$|\mathcal{F}_r[Y_T](\lambda)| = |\mathcal{L}[T](\psi(\lambda))| = \left|\int_0^\infty e^{-\psi(\lambda)t} f_T(t)dt\right|.$$

Let $B>0$ such that $f_T \leq B$ almost everywhere. By Lemma 5.2.2 we have $\psi(\lambda) \gtrsim \lambda^2$ for $\lambda\to\infty$. Hence, we conclude that

$$|\mathcal{F}_r[Y_T](\lambda)| \leq B\left|\int_0^\infty e^{-\psi(\lambda)t}dt\right| = B\,|\psi(\lambda)|^{-1} \lesssim \lambda^{-2}$$

for $\lambda\to\infty$. □

Next, we investigate the asymptotic behavior of the strongly negative definite function of a given Sturm-Liouville process.

LEMMA 5.2.2. *Let $(Y_t)_{t\geq 0}$ be a Sturm-Liouville process with associated strongly negative definite function ψ and characteristic (c,ν). Suppose $\int_0^\infty x^2\,d\nu(x)<\infty$ and $c>0$. Then it follows that*

(i) $\psi(\lambda)\gtrsim\lambda^2$ and $|\psi'(\lambda)|\lesssim\lambda$ for $\lambda\to\infty$;
(ii) $\psi(\lambda)\gtrsim\lambda^2$ and $|\psi'(\lambda)|\lesssim\lambda$ for $\lambda\to 0$.

PROOF. Theorem (4.3.5) yields

$$\psi(\lambda) = c\lambda^2 + \int_{(0,\infty)} (1-\varphi_\lambda(x))\,d\nu(x) \gtrsim \lambda^2 \tag{5.2.1}$$

for $\lambda \to \infty$ and $\lambda \to 0$, because the value of the integral in (5.2.1) is positive. For the derivative we use the Laplace representation (see Theorem 4.3.2(v))

$$\varphi_\lambda(x) = \int_{-x}^{x} \cos(\lambda t)\ d\tau_x(t). \tag{5.2.2}$$

with some probability measure τ_x. Lemma 4.3.7 and (5.2.2) imply

$$\begin{aligned}
|\psi'(\lambda)| &= \left| 2c\lambda - \int_{(0,\infty)} \frac{d}{d\lambda}\varphi_\lambda(t)\ d\nu(x) \right| \\
&\le 2c\lambda + \int_{(0,\infty)} \int_{-x}^{x} |t\sin(\lambda t)|\ d\tau_x(t) d\nu(x) \\
&\le 2c\lambda + \lambda \int_{(0,\infty)} \int_{-x}^{x} t^2\ d\tau_x(t) d\nu(x) \\
&\le 2c\lambda + \lambda \int_{(0,\infty)} x^2\ d\nu(x).
\end{aligned}$$

Thus $|\psi'(\lambda)| \lesssim \lambda$ for $\lambda \to \infty$ and $\lambda \to 0$. $\square$

In the next step we consider an integral similar to the inner integral in (5.1.6). It will be used later to derive a bound on the variance of (5.1.6).

LEMMA 5.2.3. *Let $(Y_t)_{t\ge 0}$ be a Sturm-Liouville process with associated strongly negative definite function ψ and characteristic (c,ν). Suppose $\int_0^\infty x^2 d\nu(x) < \infty$ and $c > 0$. If $(A_n)_{n\in\mathbb{N}}$ is a sequence with $A_n \to \infty$ for $n \to \infty$, then we have*

$$\int_1^{A_n} |\psi(\lambda)|^{-\gamma} |\psi'(\lambda)|\ d\lambda \lesssim A_n^{2(1-\gamma)}$$

for $n \to \infty$ for all $\gamma \in (1/2, 1)$.

PROOF. By Lemma 5.2.2(ii) there are $C \ge 0, \lambda_0 \le 1$ such that

$$|\psi(\lambda)|^{-\gamma} |\psi'(\lambda)| \le C\lambda^{1-2\gamma}$$

for all $\lambda \le \lambda_0$. Hence,

$$\begin{aligned}
\int_0^1 |\psi(\lambda)|^{-\gamma}|\psi'(\lambda)|\ d\lambda &= \int_0^{\lambda_0} |\psi(\lambda)|^{-\gamma}|\psi'(\lambda)|\ d\lambda + \int_{\lambda_0}^1 |\psi(\lambda)|^{-\gamma}|\psi'(\lambda)|\ d\lambda \\
&\le C\int_{\lambda_0}^1 \lambda^{1-2\gamma}\ d\lambda + (1-\lambda_0) \sup_{\lambda\in[0,\lambda_0]} \{|\psi(\lambda)|^{-\gamma}|\psi'(\lambda)|\} \\
&= \frac{C}{2(1-\gamma)}(1-\lambda_0^{2(1-\gamma)}) + (1-\lambda_0) \sup_{\lambda\in[0,\lambda_0]} \{|\psi(\lambda)|^{-\gamma}|\psi'(\lambda)|\} \\
&< C_1 \qquad (5.2.3)
\end{aligned}$$

holds for some $C_1 > 0$.
By Lemma 5.2.2(i) there are $K \geq 0, \lambda_1 \geq 1$ such that for all $\lambda \geq \lambda_1$ we have:

$$|\psi(\lambda)|^{-\gamma}|\psi'(\lambda)| \leq K\lambda^{1-2\gamma}.$$

Choose $n_0 \in \mathbb{N}$ such that $A_n \geq \lambda_1$ for all $n \geq n_0$. Then we have

$$\begin{aligned}\int_1^{A_n} |\psi(\lambda)|^{-\gamma}|\psi'(\lambda)|\, d\lambda &= \int_1^{\lambda_1} |\psi(\lambda)|^{-\gamma}|\psi'(\lambda)|\, d\lambda + \int_{\lambda_1}^{A_n} |\psi(\lambda)|^{-\gamma}|\psi'(\lambda)|\, d\lambda \\ &\leq (\lambda_1 - 1) \sup_{\lambda\in[1,\lambda_1]} \{|\psi(\lambda)|^{-\gamma}|\psi'(\lambda)|\} + K \int_{\lambda_1}^{A_n} \lambda^{1-2\gamma}\, d\lambda \\ &\lesssim \frac{1}{2(1-\gamma)}(A_n^{2(1-\gamma)} - \lambda_1^{2(1-\gamma)}) \\ &\lesssim A_n^{2(1-\gamma)} \qquad (5.2.4)\end{aligned}$$

for $n \to \infty$ and $\gamma \in (1/2, 1)$. A combination of (5.2.3) and (5.2.4) yields the claim. □

Next, denote the inner integral of (5.1.6) without the Σ-sign and $1/n$ by

$$\Phi_n(z, X_k) := \int_0^{A_n} [\psi(\lambda)]^{z-1}\varphi_\lambda(X_k)\psi'(\lambda)\, d\lambda$$

with $z = 1-\gamma-iv$ and $k = 1, ..., n$. In the following lemma we consider the mean and the variance of these random variables.

Lemma 5.2.4. *If $(A_n)_{n\in\mathbb{N}}$ is a sequence satisfying $A_n \to \infty$ for $n \to \infty$, then*

(i) $\mathrm{E}\left[\Phi_n(z, X_1)\right] = \int_0^{A_n} [\psi(\lambda)]^{z-1}\mathcal{F}_r[X_1](\lambda)\psi'(\lambda)\, d\lambda$ *and*
(ii) $\sqrt{\mathrm{Var}\left[\Phi_n(z, X_1)\right]} \leq \int_0^{A_n} |\psi(\lambda)|^{\mathrm{Re}(z)-1}|\psi'(\lambda)|\sqrt{\mathrm{Var}[e^{i\lambda X_1}]}\, d\lambda$

for all $n \in \mathbb{N}$ and $\mathrm{Re}(z) \in (0, 1)$.

Proof. Let $n \in \mathbb{N}$ and $\mathrm{Re}(z) \in (0, 1)$.
(i) Fubini's theorem and Lemma 5.1.1 yield

$$\begin{aligned}\mathrm{E}\left[\Phi_n(z, X_1)\right] &= \mathrm{E}\left[\int_0^{A_n} [\psi(\lambda)]^{z-1}\varphi_\lambda(X_1)\psi'(\lambda)\, d\lambda\right] \\ &= \int_0^{A_n} [\psi(\lambda)]^{z-1}\, \mathrm{E}[\varphi_\lambda(X_1)]\psi'(\lambda)\, d\lambda \\ &= \int_0^{A_n} [\psi(\lambda)]^{z-1}\mathcal{F}_r[X_1](\lambda)\psi'(\lambda)\, d\lambda.\end{aligned}$$

(ii) By means of Lemmas 2.1.8 and 2.1.7 we have

$$\begin{aligned}\sqrt{\mathrm{Var}\left[\Phi_n(z,X_1)\right]} &= \sqrt{\mathrm{Var}\left[\int_0^{A_n}[\psi(\lambda)]^{z-1}\varphi_\lambda(X_1)\psi'(\lambda)\ d\lambda\right]}\\ &\le \int_0^{A_n}\sqrt{\mathrm{Var}\left[[\psi(\lambda)]^{z-1}\varphi_\lambda(X_k)\psi'(\lambda)\right]}\ d\lambda\\ &= \int_0^{A_n}|\psi(\lambda)|^{\mathrm{Re}(z)-1}|\psi'(\lambda)|\sqrt{\mathrm{Var}[\varphi_\lambda(X_1)]}\ d\lambda.\end{aligned}$$

Thus both claims are shown. □

For the final lemma before our main result we consider an object which is associated with the bias of estimator (5.1.6). Denote it by

$$I_n := \frac{1}{2\pi}\int_{-U_n}^{U_n}\frac{\int_{A_n}^{\infty}[\psi(\lambda)]^{-\gamma-iv}\mathcal{F}_r[Y_T](\lambda)\psi'(\lambda)\ d\lambda}{\Gamma(1-\gamma-iv)}x^{-\gamma-iv}\ dv.$$

The following lemma provides an upper bound on this expression.

LEMMA 5.2.5. *Let $(Y_t)_{t\ge 0}$ be a Sturm-Liouville process with associated strongly negative definite function ψ and characteristic (c,ν). Suppose that $\int_0 \infty x^2 d\nu(x) < \infty$ and $c>0$. Let $(U_n)_{n\in\mathbb{N}}$ and $(A_n)_{n\in\mathbb{N}}$ be sequences with $A_n, U_n\to\infty$ for $n\to\infty$. If f_T is essentially bounded, then for all $x>0$ we have*

$$I_n \lesssim x^{-\gamma}U_n^{\gamma-1/2}e^{U_n\pi/2}A_n^{-2\gamma}$$

as $n\to\infty$.

PROOF. Triangle inequality implies:

$$|I_n| \le \frac{x^{-\gamma}}{2\pi}\int_{-U_n}^{U_n}\frac{\int_{A_n}^{\infty}|\psi(\lambda)|^{-\gamma}|\mathcal{F}_r[Y_T](\lambda)||\psi'(\lambda)|\ d\lambda}{|\Gamma(1-\gamma-iv)|}\ dv.$$

According to Lemma 5.2.2 we have $|\psi(\lambda)| \gtrsim \lambda^2$ and $|\psi'(\lambda)| \lesssim \lambda$ for $\lambda\to\infty$. Thus

$$\begin{aligned} I_n &\lesssim x^{-\gamma}\int_{-U_n}^{U_n}\frac{\int_{A_n}^{\infty}\lambda^{-2\gamma+1}|\mathcal{F}_r[Y_T](\lambda)|\ d\lambda}{|\Gamma(1-\gamma-iv)|}\ dv, \quad n\to\infty\\ &= x^{-\gamma}\int_{-U_n}^{U_n}\frac{1}{|\Gamma(1-\gamma-iv)|}\ dv\int_{A_n}^{\infty}\lambda^{-2\gamma+1}|\mathcal{F}_r[Y_T](\lambda)|\ d\lambda. \end{aligned} \tag{5.2.5}$$

Lemma 2.4.4 provides an estimate for the first integral in (5.2.5) and Lemma 5.2.1 for the second one. Overall,

$$I_n \lesssim x^{-\gamma}U_n^{\gamma-1/2}e^{U_n\pi/2}\int_{A_n}^{\infty}\lambda^{-2\gamma-1}\ d\lambda = x^{-\gamma}U_n^{\gamma-1/2}e^{U_n\pi/2}A_n^{-2\gamma}.$$

is true for $\gamma\in(1/2,1)$ and $n\to\infty$. This concludes the proof. □

Now we have all the auxiliary results that we need in order to give some convergence rates of estimator (5.1.6). Recall the classes of functions

$$\mathcal{C}(\beta, L, \gamma) = \left\{ f \in \mathfrak{M}_\gamma \middle| \int e^{\beta|v|} \left|\mathcal{M}[f](\gamma + iv)\right| dv \leq L \right\}$$

and

$$\mathcal{D}(\beta, L, \gamma) = \left\{ f \in \mathfrak{M}_\gamma \middle| \int (1 + |v|^\beta) \left|\mathcal{M}[f](\gamma + iv)\right| dv \leq L \right\}$$

for $\beta, L > 0$ and $\gamma \in \mathbb{R}$ from Section 3.4 and the weighted mean squared risk

$$\mathrm{MSE}_\gamma(x) = x^{2\gamma}\, \mathrm{E}[|f_T(x) - \hat{f}_n(x)|^2].$$

If the density of the estimated stopping time is in the class $\mathcal{C}(\beta, L, \gamma)$ for suitable β, L, γ, then we have the following convergence result.

THEOREM 5.2.6. *Let $(Y_t)_{t\geq 0}$ be a Sturm-Liouville process with characteristic (c, ν). Suppose that $\int_0 \infty x^2 d\nu(x) < \infty$ and $c > 0$. Let $T \geq 0$ be a stopping time independent of $(Y_t)_{t\geq 0}$ with an essentially bounded density f_T. Moreover let $f_T \in \mathcal{C}(\beta, L, \gamma)$ for some $\gamma \in (1/2, 1)$, $\beta > 0$ and $L > 0$. If we choose*

$$A_n = n^{1/4} \quad \text{and} \quad U_n = \frac{\gamma}{2\beta + \pi} \log n - \frac{2\gamma - 1}{2\beta + \pi} \log\log n$$

in (5.1.6), *then we get the convergence rate*

$$\sqrt{\mathrm{MSE}_\gamma(x)} \lesssim n^{-\frac{\beta\gamma}{2\beta+\pi}} (\log n)^{\beta \frac{2\gamma-1}{2\beta+\pi}} \tag{5.2.6}$$

for all $x > 0$ and $n \to \infty$.

PROOF. Let $x > 0$. For the bias of $\hat{f}_n$ we have

$$\begin{aligned}
|\,\mathrm{E}[\hat{f}_n(x)] - f_T(x)| &= \left| \frac{1}{2\pi} \int_{-U_n}^{U_n} \frac{\int_0^{A_n} [\psi(\lambda)]^{-\gamma - iv} \mathcal{F}_r[Y_T](\lambda) \psi'(\lambda)\, d\lambda}{\Gamma(1 - \gamma - iv)} x^{-\gamma - iv}\, dv \right.\\
&\qquad \left. - \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{M}[T](\gamma + iv) x^{-\gamma - iv} dv \right| \\
&\leq \frac{1}{2\pi} \left| \int_{-U_n}^{U_n} \frac{\int_{A_n}^{\infty} [\psi(\lambda)]^{-\gamma - iv} \mathcal{F}_r[Y_T](\lambda) \psi'(\lambda)\, d\lambda}{\Gamma(1 - \gamma - iv)} x^{-\gamma - iv}\, dv \right| \\
&\qquad + \frac{x^{-\gamma}}{2\pi} \int_{\{|v| > U_n\}} |\mathcal{M}[T](\gamma + iv)| dv \\
&=: (*)_1 + (*)_2
\end{aligned}$$

by Lemma 5.2.4, the Mellin inversion formula and (5.1.4). In Lemma 5.2.5 we already showed

$$(*)_1 \lesssim x^{-\gamma} U_n^{\gamma - 1/2} e^{U_n \pi/2} A_n^{-2\gamma}$$

for $n \to \infty$. And from $f_T \in \mathcal{C}(\beta, L, \gamma)$ we get

$$(*)_2 \le \frac{x^{-\gamma}}{2\pi} e^{-\beta U_n} \int_{\{|v|>U_n\}} |\mathcal{M}[T](\gamma + iv)| e^{\beta|v|} dv \le \frac{x^{-\gamma} L}{2\pi} e^{-\beta U_n}.$$

For the variance of estimator (5.1.6) we conclude by means of Lemma 2.1.8, that

$$\begin{aligned}
\operatorname{Var}[\hat{f}_n(x)] &= \operatorname{Var}\left[\frac{1}{2\pi n} \sum_{k=1}^{n} \int_{-U_n}^{U_n} \frac{\Phi_n(1-\gamma-iv, X_k)}{\Gamma(1-\gamma-iv)} x^{-\gamma-iv}\, dv\right] \\
&= \frac{1}{(2\pi)^2 n} \operatorname{Var}\left[\int_{-U_n}^{U_n} \frac{\Phi_n(1-\gamma-iv, X_1)}{\Gamma(1-\gamma-iv)} x^{-\gamma-iv} dv\right] \\
&\le \frac{1}{(2\pi)^2 n} \left[\int_{-U_n}^{U_n} \sqrt{\operatorname{Var}\left[\frac{\Phi_n(1-\gamma-iv, X_1)}{\Gamma(1-\gamma-iv)} x^{-\gamma-iv}\right]} dv\right]^2 \\
&= \frac{1}{(2\pi)^2 n} x^{-2\gamma} \left[\int_{-U_n}^{U_n} \frac{\sqrt{\operatorname{Var}\left[\Phi_n(1-\gamma-iv, X_1)\right]}}{|\Gamma(1-\gamma-iv)|} dv\right]^2.
\end{aligned}$$

Lemmas 5.2.4 and 2.4.4 imply (Note that $\sqrt{\operatorname{Var}[\varphi_\lambda(X_1)]} \le 1$ because $\|\varphi_\lambda\|_\infty = 1$)

$$\begin{aligned}
\operatorname{Var}[\hat{f}_n(x)] &\le \frac{x^{-2\gamma}}{(2\pi)^2 n} \left(\int_{-U_n}^{U_n} \frac{\int_0^{A_n} |\psi(\lambda)|^{-\gamma} |\psi'(\lambda)| \sqrt{\operatorname{Var}[\varphi_\lambda(X_1)]}\, d\lambda}{|\Gamma(1-\gamma-iv)|} dv\right)^2 \\
&\le \frac{x^{-2\gamma}}{\pi^2 n} \left(C U_n^{\gamma-1/2} e^{U_n \pi/2} \int_0^{A_n} |\psi(\lambda)|^{-\gamma} |\psi'(\lambda)|\, d\lambda\right)^2
\end{aligned}$$

As a consequence of Lemma 5.2.3 it follows that

$$x^{2\gamma} \operatorname{Var}[\hat{f}_n(x)] \lesssim \frac{1}{n} \left(U_n^{\gamma-1/2} e^{U_n \pi/2} A_n^{2(1-\gamma)}\right)^2, \quad n \to \infty. \tag{5.2.7}$$

Combining estimates for the terms $(*)_1$, $(*)_2$ and (5.2.7) yields

$$\begin{aligned}
\sqrt{\operatorname{MSE}_\gamma(x)} &= \sqrt{x^{2\gamma} \operatorname{Var}[\hat{f}_n(x)] + (x^\gamma(\operatorname{E}[\hat{f}_n(x)] - f_T(x)))^2} \\
&\le \sqrt{x^{2\gamma} \operatorname{Var}[\hat{f}_n(x)] + (x^\gamma((*)_1 + (*)_2))^2} \\
&\le x^\gamma \sqrt{\operatorname{Var}[\hat{f}_n(x)]} + x^\gamma((*)_1 + (*)_2) \\
&\lesssim \frac{1}{\sqrt{n}} U_n^{\gamma-1/2} e^{U_n\pi/2} A_n^{2(1-\gamma)} + e^{-\beta U_n} + U_n^{\gamma-1/2} e^{U_n \pi/2} A_n^{-2\gamma}.
\end{aligned}$$

By choosing

$$A_n = n^{1/4} \quad \textit{and} \quad U_n = \frac{\gamma}{2\beta+\pi} \log n - \frac{2\gamma-1}{2\beta+\pi} \log\log n$$

we get

$$\begin{aligned}\sqrt{\mathrm{MSE}_\gamma(x)} &\lesssim U_n^{\gamma-1/2}e^{U_n\pi/2}\left(\frac{1}{\sqrt{n}}A_n^{2(1-\gamma)}+A_n^{-2\gamma}\right)+e^{-\beta U_n}\\ &\lesssim (\log n)^{\gamma-1/2}\left(n^{\frac{\gamma}{2\beta+\pi}}(\log n)^{\frac{1-2\gamma}{2\beta+\pi}}\right)^{\pi/2}(n^{\frac{1}{2}(1-\gamma)-\frac{1}{2}}+n^{-\frac{1}{2}\gamma})\\ &\quad+\left(n^{\frac{\gamma}{2\beta+\pi}}(\log n)^{\frac{1-2\gamma}{2\beta+\pi}}\right)^{-\beta}\\ &= n^{\frac{\gamma}{2\beta+\pi}\frac{\pi}{2}-\frac{\gamma}{2}}(\log n)^{\frac{1-2\gamma}{2\beta+\pi}\frac{\pi}{2}+\gamma-\frac{1}{2}}+n^{-\frac{\beta\gamma}{2\beta+\pi}}(\log n)^{\beta\frac{2\gamma-1}{2\beta+\pi}}\\ &\lesssim n^{-\frac{\beta\gamma}{2\beta+\pi}}(\log n)^{\beta\frac{2\gamma-1}{2\beta+\pi}},\end{aligned}$$

which is the claimed order of convergence. □

REMARK 5.2.7. *By letting* $\gamma \to 1$ *in* (5.2.6) *we obtain the convergence rate* $n^{-\frac{\beta}{2\beta+\pi}}\log^{\frac{\beta}{2\beta+\pi}} n$. *This is the same rate (up to a logarithmic factor) we derived for Bessel processes in Theorem 3.7.3(i). In Chapter 6 we show that this rate is optimal for Bessel processes. Since the class of Sturm-Liouville processes contains Bessel processes (see Section 5.3), it follows that* $n^{-\frac{\beta}{2\beta+\pi}}$ *is the optimal rate here as well.*

Next, let the density of the estimated stopping time be in the class $\mathcal{D}(\beta, L, \gamma)$ for suitable β, L, γ. Then the following convergence result holds.

THEOREM 5.2.8. *Let* $(Y_t)_{t\geq 0}$ *be a Sturm-Liouville process with characteristic* (c,ν). *Suppose that* $\int_0 \infty x^2 d\nu(x) < \infty$ *and* $c > 0$. *Let* $T \geq 0$ *be a stopping time independent of* $(Y_t)_{t\geq 0}$ *with an essentially bounded density* f_T. *Moreover let* $f_T \in \mathcal{D}(\beta, L, \gamma)$ *for some* $\gamma \in (1/2, 1)$, $\beta > 0$ *and* $L > 0$. *If we choose*

$$A_n = n^{1/4} \quad \text{and} \quad U_n = \frac{\gamma}{\pi}\log n - \frac{2\beta+2\gamma-1}{\pi}\log\log n$$

in (5.1.6), *then we get the convergence rate*

$$\sqrt{\mathrm{MSE}_\gamma(x)} \lesssim (\log n)^{-\beta} \tag{5.2.8}$$

for all $x > 0$ *and* $n \to \infty$.

PROOF. Let $x > 0$. This proof is analogue to the one of Theorem 5.2.6 save for the estimation of the term $(*)_2$. Here, from $f_T \in \mathcal{D}(\beta, L, \gamma)$ it follows that

$$(*)_2 \leq \frac{x^{-\gamma}}{2\pi(1+U_n^\beta)}\int_{\{|v|>U_n\}}|\mathcal{M}[T](\gamma+iv)|(1+|v|^\beta)dv \leq \frac{x^{-\gamma}L}{2\pi(1+U_n^\beta)}.$$

For the error term in question this leads to

$$\begin{aligned}\sqrt{\mathrm{MSE}_\gamma(x)} &\leq x^\gamma \sqrt{\mathrm{Var}[\hat{f}_n(x)]} + x^\gamma((*)_1 + (*)_2)\\ &\lesssim U_n^{\gamma-1/2} e^{U_n\pi/2}\left(\frac{1}{\sqrt{n}} A_n^{2(1-\gamma)} + A_n^{-2\gamma}\right) + U_n^{-\beta}.\end{aligned}$$

The choices

$$A_n = n^{1/4} \quad \textit{and} \quad U_n = \frac{\gamma}{\pi}\log n - \frac{2\beta+2\gamma-1}{\pi}\log\log n$$

yield

$$\begin{aligned}\sqrt{\mathrm{MSE}_\gamma(x)} &\lesssim (\log n)^{\gamma-1/2}\left(n^{\frac{\gamma}{\pi}}(\log n)^{\frac{1-2\beta-2\gamma}{\pi}}\right)^{\pi/2} n^{-\frac{\gamma}{2}} + (\log n)^{-\beta}\\ &\lesssim (\log n)^{-\beta},\end{aligned}$$

which is the claimed order of convergence in this case. □

Again, we were able to recover the rate $(\log n)^{-\beta}$ that we also had for Bessel processes in an analogous setting (see Theorem 3.7.3(ii)). Whether this rate is optimal for $\mathcal{D}(\beta, L, \gamma)$ remains an open question due to reasons described in the end of Chapter 6.

5.3. Application to a Bessel Process

As a comparison to Section 3.9 consider again a Gamma(2, 1) distributed stopping time T, i.e. T has the density

$$f(x) = xe^{-x}, \quad x \geq 0.$$

As mentioned in [**10**], [**23**] and [**18**], every Bessel process with dimension d has a representation as a Lévy process on a Bessel-Kingman hypergroup $(\mathbb{R}_+, *_\alpha)$ with index $\alpha = d/2 - 1 \geq -1/2$ (cf. Example 4.3.4(i)). To be more precise consider the *Rayleigh convolution semi-group* $(\mu_t)_{t\geq 0}$ on a Bessel-Kingman hypergroup defined by $\mu_t := \rho_{\frac{1}{\sqrt{2t}}}$ for all $t \geq 0$, where ρ_b are given for all $b \geq 0$ by their Lebesgue densities

$$f_b(x) := \frac{2b^{2\alpha+2}x^{2\alpha+2}}{\Gamma(\alpha+1)} e^{-b^2x^2}$$

for all $x \in \mathbb{R}_+$.
A Lévy process on $(\mathbb{R}_+, *_\alpha)$ associated with $(\mu_t)_{t\geq 0}$ is equivalent to a Bessel process of dimension $d := 2\alpha+2$ as introduced in Section 3.7 (see [**18**, page 52] for a proof). The Fourier transform on a Bessel-Kingman hypergroup of a probability measure μ is given by

$$\hat{\mu}(\lambda) = \int_{\mathbb{R}_+} \Lambda_\alpha(\lambda x) d\mu(x).$$

It is easy to see (using the definition of J_α) that the modified Bessel function Λ_α introduced in Definition 4.3.3 allows the series representation

$$\Lambda_\alpha(x) = \sum_{k=0}^{\infty} \frac{(-1)^k \Gamma(\alpha+1)}{2^{2k} k! \Gamma(\alpha+k+1)} x^{2k}, \quad x \geq 0. \tag{5.3.1}$$

Moreover, by [**39**, page 48] we also have the representation

$$\Lambda_\alpha(x) = \frac{2\Gamma(\alpha+1)}{\pi\Gamma(\alpha+1/2)} \int_0^1 \cos(xy)(1-y^2)^{\alpha-1/2}\, dy, \quad x \geq 0. \tag{5.3.2}$$

The associated strongly negative definite function ψ allows the representation

$$\psi(\lambda) = c\lambda^2 + \int_{\mathbb{R}_+} (1 - \Lambda_\alpha(\lambda x)) d\nu(x), \quad \lambda \geq 0$$

with $c \geq 0$ and a Lévy measure ν.
Taking $c = 1/2,\ \nu = 0$ yields the strongly negative definite function corresponding to a Bessel process of dimension $d = 2\alpha + 2$ (cf. [**18**, page 58]). Applying these facts to our estimator (5.1.6) gives

$$\hat{f}_n(x) = \frac{1}{2\pi} \int_{-U_n}^{U_n} \frac{\int_0^{A_n} \left(\frac{\lambda^2}{2}\right)^{-\gamma-iv} \lambda\phi_n(\lambda)\, d\lambda}{\Gamma(1-\gamma-iv)} x^{-\gamma-iv}\, dv, \tag{5.3.3}$$

with $\phi_n(\lambda) := \frac{1}{n}\sum_{k=1}^n \Lambda_\alpha(\lambda X_k)$ and $\gamma \in (1/2, 1)$ as an estimator for the density of a stopping time $T \geq 0$ based on samples $X_1, ..., X_n$ of BES_T. In order to compute the inner integral numerically we use

Lemma 5.3.1. *In the setting above we have for $\gamma \in (0,1)$ the decomposition*

$$\begin{aligned} \int_0^{A_n} \left(\frac{\lambda^2}{2}\right)^{-\gamma-iv} \lambda\phi_n(\lambda)\, d\lambda = &\int_0^{A_n} \left(\frac{\lambda^2}{2}\right)^{-\gamma-iv} \lambda[\phi_n(\lambda) - e^{-m_n\lambda^2/2}]\, d\lambda \\ &+ m_n^{\gamma-1+iv}\Gamma(1-\gamma-iv) + \mathcal{O}(m_n^{\gamma-1}e^{-m_n A_n^2/2}), \end{aligned} \tag{5.3.4}$$

where $m_n := \frac{1}{2n(1+\alpha)}\sum_{k=1}^n X_k^2$. Moreover,

$$m_n \to \mathrm{E}[T]\frac{\mathrm{E}[Y_1^2]}{2(1+\alpha)} = 2 \tag{5.3.5}$$

holds almost surely for $n \to \infty$ and

$$\phi_n(\lambda) - e^{-m_n\lambda^2/2} = \mathcal{O}(\lambda^4) \tag{5.3.6}$$

for $\lambda \to 0$.

PROOF. We have

$$\int_0^{A_n} \left(\frac{\lambda^2}{2}\right)^{-\gamma-iv} \lambda e^{-m_n\lambda^2/2}\, d\lambda = m_n^{\gamma-1+iv}\Gamma(1-\gamma-iv) - \int_{A_n^2/2}^{\infty} \lambda^{-\gamma-iv}e^{-m_n\lambda}d\lambda$$
$$= m_n^{\gamma-1+iv}\Gamma(1-\gamma-iv) + \mathcal{O}(m_n^{\gamma-1}\exp -A_n^2/2),$$

which implies (5.3.4). Convergence in (5.3.5) follows from the law of large numbers and Lemma 3.1.2. The asymptotic relation (5.3.6) is a consequence of (5.3.1) and the exponential series. □

The asymptotic relation (5.3.6) guarantees that the integrand on the right-hand side of (5.3.4) has no singularity in 0 (unlike the left-hand side). This lack of singularities makes the whole computation much more stable. A similar decomposition is used in [**7**] to compute estimates in the case of a classical Lévy process.

In order to compare the performance of this estimator to the one discussed in section 3.9 we choose $\gamma = 0.7$ and $\alpha = 3/2$ (which is equivalent to $d = 5$). Note that

$$\Lambda_{3/2}(x) = \frac{3\sin(x) - 3x\cos(x)}{x^3}$$

for $x \in \mathbb{R}_+$.
Choose $A_n = n^{1/4}$ in accordance with Theorem 5.2.6. We employ the adaptive procedure described in Section 3.9 to find a suitable choice for the cut off parameter U_n, that is to say we calculate estimators $f_n^{(1)}, ..., f_n^{(40)}$ from 5.3.3 using the decomposition (5.3.4) and replacing U_n with $U_n^l = 0.1 \times l$ for $l = 1, ..., 40$ and determine $l^\star = \text{argmin}_l d(l)$ with

$$d(l) = \int_0^\infty |f_n^{(l+1)}(x) - f_n^{(l)}(x)|dx \approx \sum_{i=0}^{90} |f_n^{(l+1)}(0.1 + 0.01i) - f_n^{(l)}(0.1 + 0.01i)|.$$

$f_n^{(l^\star)}$ is our estimator of f_T. We then calculate $f_n^{(l^\star)}$ one hundred times based on different samples of size $n = 1000$ each. Left-hand side of Figure 5 shows the Gamma(2,1) density and its 100 estimators. Right-hand side of Figure 5 depicts the box-plot of the loss

$$\sup_{x\in\mathbb{R}_+} \{|f_n^{(l^\star)}(x) - f_T(x)|\} \approx \sup_{x\in\{0.1,0.11,...,9.99,10\}} \{|f_n^{(l^\star)}(x) - f_T(x)|\}.$$

Note that estimation near 0 is difficult for the same reasons that we mentioned at the end of Section 3.9.
We can now compare the performance of this estimation procedure with the one based on self-similarity (see Section 3.9). When we consider Figure 4 (top right) and the associated Figure 5, it is difficult to see a difference. A comparison of the box plots of the loss (Figure 3 (right) and Figure 5) reveals that estimation based on self-similarity performs better: We have a greater median loss of circa 0.14 from the hypergroup approach against 0.11 from the self-similarity approach

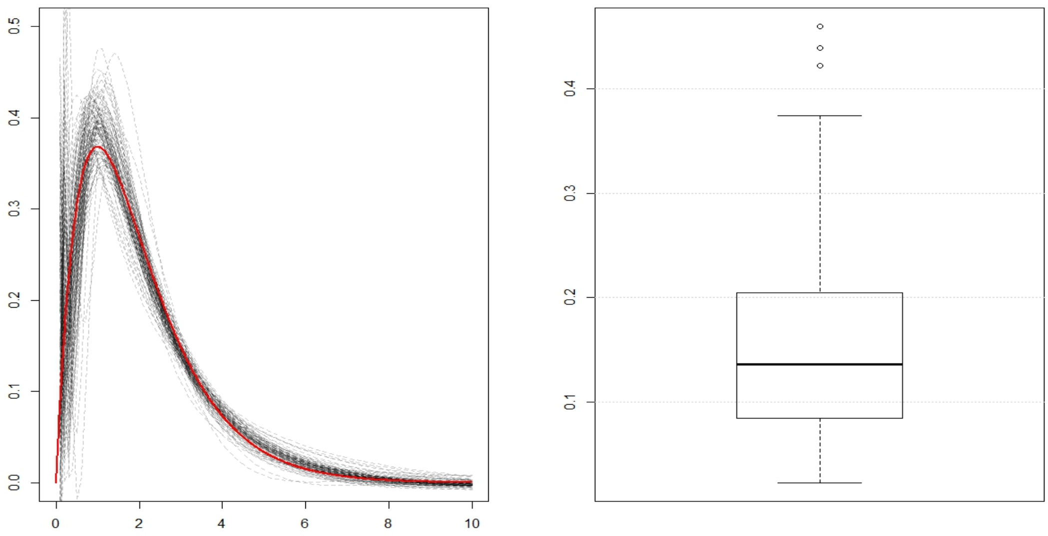

FIGURE 5. Left: Estimated Gamma density (red) and 100 independent copies of $f_n^{(l^\star)}$ (grey) for the sample size $n = 1000$. Right: box plot of the loss $\sup_{x\in\mathbb{R}_+}\{|f_n^{(l^\star)}(x) - f_T(x)\}$.

with a wider range (loss of up to circa 0.49 against 0.28). This difference can (at least in part) be explained by the numerical difficulties discussed in this section and does not devalue the general ideas of Section 5.1.

CHAPTER 6

Optimality

So far we only established upper bounds on the risk $\sqrt{\text{MSE}_\gamma(x)}$ defined by (3.3.3). The question arises whether theses bounds can be improved. To answer this we introduce the notion of optimal rate of convergence roughly following the outline of [**37**, Chapter 2]. Application of this theory to our setting reveals that the upper bounds provided above are optimal in the sense described below. A similar optimality result was obtained in [**6**] for the case where a one-dimensional Brownian motion is observed. We will build on their work to obtain analogous results for observation types considered in this thesis.

6.1. Minimax Risk

Let Θ be a non-parametric class of functions containing the function f we wish to estimate. In this thesis $\Theta = \mathcal{C}(\beta, L, \gamma)$ or $\Theta = \mathcal{D}(\beta, L, \gamma)$ (see Section 3.4). Consider the distance

$$d(f,g) := d_x(f,g) := x^\gamma |f(x) - g(x)| \tag{6.1.1}$$

for $f, g \in \Theta$ and some fixed $x, \gamma > 0$. The performance of an estimator $\hat{f}_n$ of f is measured by the maximum risk of this estimator on Θ:

$$r(\hat{f}_n) := \sup_{f\in\Theta} E[d^2(\hat{f}_n, f)] = \sup_{f\in\Theta} \text{MSE}_\gamma\,.$$

In Sections 3.7 and 3.10 we established upper bounds on $r(\hat{f}_n)$, that is, inequalities of the form

$$\sup_{f\in\Theta} E[d^2(\hat{f}_n, f)] \le C\psi_n^2$$

for different estimators $\hat{f}_n$, some positive sequences ψ_n with $\psi_n \to 0$ for $n \to \infty$ and some constants $C < \infty$. The aim of this section is to complement these upper bounds by the corresponding lower bounds

$$\sup_{f\in\Theta} E[d^2(\hat{f}_n, f)] \ge c\psi_n^2$$

(for sufficiently large n) for all estimators $\hat{f}_n$ (that is, all measurable functions of our observations $X_1, ..., X_n$) and some positive constant c. To this end, it is useful to define the minimax risk associated with the distance d:

$$\mathcal{R}_n^\star := \inf_{\hat{f}_n} \sup_{f\in\Theta} \text{E}[d^2(\hat{f}_n, f)],$$

where the infimum is over all estimators. The upper bounds established in Sections 3.7 and 3.10 imply that there exists a constant $C < \infty$ such that

$$\limsup_{n\to\infty} \psi_n^{-2}\mathcal{R}_n^\star \le C \tag{6.1.2}$$

for a sequence ψ_n with $\psi_n \to 0$ as $n \to \infty$. The corresponding lower bounds claim that there exists a constant $c > 0$ such that

$$\liminf_{n\to\infty} \psi_n^{-2}\mathcal{R}_n^\star \ge c \tag{6.1.3}$$

holds for the same sequence $(\psi_n)_{n\in\mathbb{N}}$.

Definition 6.1.1. *A positive sequence $(\psi_n)_{n\in\mathbb{N}}$ is called an optimal rate of convergence of estimators on (Θ, d) if (6.1.2) and (6.1.3) hold. An estimator $f_n^\star$ satisfying*

$$\sup_{f\in\Theta} E[d^2(f_n^\star, f)] \le C'\psi_n^2,$$

where $(\psi_n)_{n\in\mathbb{N}}$ is the optimal rate of convergence and $C' < \infty$ is a constant, is called a rate optimal estimator on (Θ, d).

Optimal rates of convergence are defined within a multiplicative constant. In fact, if ψ_n is an optimal rate of convergence, then any sequence ψ_n' satisfying

$$0 < \liminf_{n\to\infty}(\psi_n/\psi_n') \le \limsup_{n\to\infty}(\psi_n/\psi_n') < \infty$$

is also called an optimal rate of convergence.

6.2. Reduction to two Hypotheses

The following argument is useful in obtaining lower bounds on $\mathcal{R}_n^\star$. By Markov inequality,

$$\begin{aligned} E[\psi_n^{-2}d^2(\hat f_n, f)] &= x^{2\gamma}E[\psi_n^{-2}|\hat f_n(x) - f(x)|^2] \\ &\ge x^{2\gamma}P(|\hat f_n(x) - f(x)| \ge \psi_n) \end{aligned}$$

for all $x > 0$. Since $x^{2\gamma} > 0$ for all $x, \gamma > 0$ and since the optimal rate ψ_n is unique up to a multiplicative constant, instead of searching for a lower bound on the minimax risk $\mathcal{R}_n^\star$, it is sufficient to find a lower bound on the minimax probabilities of the form

$$\inf_{\hat f_n}\sup_{f\in\Theta} P(|\hat f_n(x) - f(x)| \ge \varepsilon\psi_n)$$

for some $\varepsilon > 0$.
To find a lower bound on the minimax probabilities note that

$$\inf_{\hat f_n}\sup_{f\in\Theta} P(|\hat f_n(x) - f(x)| \ge \varepsilon\psi_n) \ge \inf_{\hat f_n} \max_{f\in\{f_{n,0}, f_{n,1}\}} P(|\hat f_n(x) - f(x)| \ge \varepsilon\psi_n) \tag{6.2.1}$$

holds for any $f_{n,0}, f_{n,1} \in \Theta$, $n \in \mathbb{N}$, $\varepsilon > 0$. Later we will choose $f_{n,0}$ and $f_{n,1}$ in an appropriate way and call them *hypotheses*. We call a *test* any measurable function

$\Psi : (X_1, ..., X_n)(\Omega) \to \{0,1\}$, where $X_1, ..., X_n$ are the observations in the model. Suppose that $f_{n,0}$ and $f_{n,1}$ can be chosen such that for some $x, \varepsilon > 0$

$$|f_{n,0}(x) - f_{n,1}(x)| \geq 2\varepsilon\psi_n \tag{6.2.2}$$

for all n large enough. By the triangle inequality,

$$|f_{n,k}(x) - f_{n,j}(x)| \leq |\hat{f}_n(x) - f_{n,k}(x)| + |\hat{f}_n(x) - f_{n,j}(x)| \tag{6.2.3}$$

for $j, k = 0, 1$ and any estimator $\hat{f}_n$. Using (6.2.2) and then (6.2.3) we obtain

$$\begin{aligned} P(|\hat{f}_n(x) - f_{n,j}(x)| \geq \varepsilon\psi_n) &\geq P(2|\hat{f}_n(x) - f_{n,j}(x)| \geq \min_{k \neq j} |f_{n,k}(x) - f_{n,j}(x)|) \\ &\geq P(|\hat{f}_n(x) - f_{n,j}(x)| \geq \min_{k \neq j} |\hat{f}_n(x) - f_{n,k}(x)|) \\ &= P(\arg\min_{k=0,1} |\hat{f}_n(x) - f_{n,k}(x)| \neq j) \end{aligned}$$

for $j = 0, 1$. In other words

$$P(|\hat{f}_n(x) - f_{n,j}(x)| \geq \varepsilon\psi_n) \geq P(\Psi^\star \neq j), \quad j = 0, 1, \tag{6.2.4}$$

where $\Psi^\star = \arg\min_{k=0,1} |\hat{f}_n(x) - f_{n,k}(x)|$ is the *minimum distance test*. It follows from (6.2.1) and (6.2.4) that if we can construct hypotheses $f_{n,0}$ and $f_{n,1}$ satisfying (6.2.2), then

$$\inf_{\hat{f}_n} \sup_{f \in \Theta} P(|\hat{f}_n(x) - f(x)| \geq \varepsilon\psi_n) \geq \inf_{\hat{f}_n} \max_{f \in \{f_{n,0}, f_{n,1}\}} P(|\hat{f}_n(x) - f(x)| \geq \varepsilon\psi_n) \geq p_e,$$

where

$$p_e := \inf_{\Psi} \max_{j=0,1} P(\Psi \neq j)$$

and $\inf_\Psi$ denotes the infimum over all tests in a model with n observation. For the sake of brevity we omit that p_e depends on n. The quantity p_e is called the *minimax probability of error* for the problem of testing two hypotheses $f_{n,0}$ and $f_{n,1}$. Summing up this section we just proved the following theorem.

THEOREM 6.2.1. *Let Θ be a parameter class of real valued functions, $x > 0$ and d_x as in* (6.1.1). *Let $\hat{f}_n$ be an estimator for $f \in \Theta$ satisfying* (6.1.2) *for some sequence $(\psi_n)_{n \in \mathbb{N}}$. If there are two hypotheses $f_{n,0}$, $f_{n,1} \in \Theta$ such that for some $\varepsilon > 0$ we have*

$$d_x(f_{n,0}, f_{n,1}) \geq 2\varepsilon\psi_n \tag{6.2.5}$$

for all n large enough and if

$$p_e \geq c' \tag{6.2.6}$$

holds for some $c' > 0$ independent of n, then $(\psi_n)_{n \in \mathbb{N}}$ is the optimal rate of convergence for (Θ, d_x).

Tsybakov provides various lower bounds for p_e in [**37**] that can be used to show (6.2.6). The following is of interest to us.

THEOREM 6.2.2. *Let Θ be a parameter class and $n \in \mathbb{N}$. Let $q_{n,j}$ be the common density of n observations associated with parameter $f_{n,j} \in \Theta$, $j = 0, 1$. If*

$$\chi^2(q_{n,1}, q_{n,0}) := \int \frac{(q_{n,1}(x) - q_{n,0}(x))^2}{q_{n,0}(x)} dx \le \alpha \tag{6.2.7}$$

for some $\alpha < \infty$ (independent of n), then

$$p_e \ge \max\left\{\frac{\exp(-\alpha)}{4}, \frac{1 - \sqrt{\alpha/2}}{2}\right\}$$

for all n large enough. (Recall that p_e depends on n.)

PROOF. See [**37**, Theorem 2.2(iii)]. □

In our model of a stopped process Y (see Section 3.1) we consider a class of densities of stopping times as Θ. For two stopping times T_0 and T_1 the densities $q_{n,0}$ and $q_{n,1}$ from Theorem 6.2.2 correspond to the distributions of $(X_1^1, ..., X_n^1)$ and $(X_1^2, ..., X_n^2)$ respectively, where $X_i^j \sim |Y_{T_j}|$ $(j = 0, 1,\ i = 1, ..., n)$. We already mentioned at the end of Subsection 3.10.1 that we can only show optimality if the sign of our observations is unknown. This is, of course, not an issue for Bessel and squared Bessel processes which are nonnegative.
$\chi^2(\cdot, \cdot)$ is called χ^2-*divergence*. If p is the density of a product measure $P = \bigotimes_{i=1}^n P_i$ with marginal densities p_i and if q is the density of a product measure $Q = \bigotimes_{i=1}^n Q_i$ with marginal densities q_i, then

$$\chi^2(p, q) = \prod_{i=1}^n (1 + \chi^2(p_i, q_i)) - 1$$

(cf. [**37**, page 86]). Since we have i.i.d. observations in our models, in order to check (6.2.7) it is sufficient to show

$$(1 + \chi^2(p_{1,n}, p_{0,n}))^n - 1 \le \alpha \tag{6.2.8}$$

for some $\alpha < \infty$ (independent of n), where $p_{j,n}$ is the density of one of n observations of $|Y_{T_j}|$ $(j = 0, 1)$.

6.3. Optimality for the Class $\mathcal{C}$

In order to show that the rates obtained in Theorems 3.7.3(i), 3.10.2(i) and 3.10.4(i) are optimal in the minimax sense we need to construct hypotheses $f_{n,0}$ and $f_{n,1}$ as in Theorem 6.2.1. For this purpose we can follow the construction in [**7**], where optimality is shown for the model, where the absolute value of a one-dimensional Brownian motion is observed.
Define for $\nu > 1$ and $M > 0$ two auxiliary functions

$$q(x) = \frac{\nu \sin(\pi/\nu)}{\pi} \frac{1}{1 + x^\nu}, \quad x \ge 0$$

and

$$\rho_M(x) = \frac{1}{\sqrt{2\pi}} e^{-\log^2(x)/2} \frac{\sin(M\log(x))}{x}, \quad x \geq 0.$$

The following lemma provides some properties of the functions q and ρ_M.

Lemma 6.3.1. *The function q is a probability density on $\mathbb{R}_+$ with the Mellin transform*

$$\mathcal{M}[q](z) = \frac{\sin(\pi/\nu)}{\sin(\pi z/\nu)}, \quad 0 < \mathrm{Re}(z) < \nu. \tag{6.3.1}$$

The Mellin transform of the function ρ_M is given by

$$\mathcal{M}[\rho_M](z) = \frac{e^{(z-1+iM)^2/2} - e^{(z-1-iM)^2/2}}{2i}, \quad z \in \mathbb{C}. \tag{6.3.2}$$

Proof. The formula for $\mathcal{M}[q](z)$ can be found in [**27**]. Representation (6.3.2) follows by the change of variables $y = \log(x)$ and completing the square. See also [**7**, Lemma 6.2] for more details. □

Set now for any $M > 0$ and some $\delta > 0$,

$$f_{0,M}(x) := q(x), \quad f_{1,M}(x) := q(x) + \delta(q \odot \rho_M)(x), \tag{6.3.3}$$

for $x \geq 0$, where $q \odot \rho_M$ stands for the Mellin convolution of q and ρ_M (see (2.2.1)). If we choose M appropriately depending on n (see below), then $f_{0,M}$ and $f_{1,M}$ are the two hypotheses which we need in order do apply Theorem 6.2.1. The following lemma will help us verify condition (6.2.5).

Lemma 6.3.2. *For any $M > 0$ and some $\delta > 0$ not depending on M the function $f_{1,M}$ is a probability density satisfying*

$$\sup_{x \geq 0} |f_{0,M}(x) - f_{1,M}(x)| \gtrsim \exp(-M\pi/\nu), \quad M \to \infty.$$

Moreover, $f_{0,M}$ and $f_{1,M}$ are in $\mathcal{C}(\beta, L, \gamma)$ for all $0 < \beta < \pi/\nu$ and $\gamma > 0$ with L depending on γ.

Proof. See [**7**, Lemma 6.3]. □

Looking further towards applying Theorem 6.2.1 let us consider the densities $p_{M,0}$ and $p_{M,1}$ of an observation associated with the hypotheses $f_{0,M}$ and $f_{1,M}$, respectively. This step is somewhat different from [**7**], because we consider a different model. At this point we have to differentiate between the models we discussed so far.

6.3.1. Observation of a Bessel Process. In this subsection we prove that the rate $n^{\frac{-\beta}{\pi+2\beta}}$ from Theorem 3.7.3(i) is optimal for $(\mathcal{C}(\beta, L, \gamma), d_x)$ in the sense of Definition 6.1.1 for all $\beta \in (0, \pi)$, $\gamma > 0$ with L depending on γ. By Remark 5.2.7 this implies that the rate in Theorem 5.2.6(i) is also optimal.
Let $T_{0,M}$ and $T_{1,M}$ be two random variables with respective densities $f_{0,M}$ and $f_{1,M}$ given by (6.3.3). The density of the random variable $BES_{T_{i,M}}$, $i = 0, 1$ is obtained via Lemma 3.1.3 from the density of BES_t (given in Lemma 3.7.2(i)). We have

$$p_{i,M}(x) = \frac{2^{1-\frac{d}{2}}}{\Gamma(d/2)} x^{d-1} \int_0^\infty \lambda^{-d/2} e^{-\frac{x^2}{2\lambda}} f_{i,M}(\lambda) d\lambda, \quad x > 0, d \geq 1, i = 0, 1.$$

For the Mellin transform of $p_{i,M}$ we use self-similarity of BES and 3.7.2(iv) to get

$$\begin{aligned} \mathcal{M}[p_{i,M}](s) &= \mathcal{M}[BES_1](s)\mathcal{M}[T_{i,M}]((s+1)/2) \\ &= \frac{1}{\Gamma(d/2)} \Gamma\left(\frac{s+d-1}{2}\right) 2^{\frac{s+1}{2}} \mathcal{M}[f_{i,M}]((s+1)/2), \quad i = 0, 1 \end{aligned} \tag{6.3.4}$$

for $\mathrm{Re}(s) > 1 - d$ and $\mathrm{Re}(s) \in (-1, 2\nu + 1)$.

LEMMA 6.3.3. *The χ^2-distance between the densities $p_{0,M}$ and $p_{1,M}$ fulfills*

$$\chi^2(p_{1,M}, p_{0,M}) \lesssim M^{\nu+d-2} e^{-M\pi(1+2/\nu)}$$

for $M \to \infty$ and all $d \geq 1$, $\nu > 1$.

PROOF. This proof is similar to the one of [**7**, Lemma 6.4], where the special case $d = 1$ is treated.

Step 1. We show

$$p_{0,M}(x) \gtrsim x^{-2\nu+1}, \quad x \to \infty \tag{6.3.5}$$

for $d \geq 1$, $\nu > 1$.
To this end define $c_{\nu,d} := \frac{2^{1-\frac{d}{2}}}{\Gamma(d/2)} \frac{\nu \sin(\pi/\nu)}{\pi}$ and perform the change of variables $y = 1/\lambda$ in the second equality sign of the following calculation and $z = y\frac{x^2}{2}$ in the sixth to

obtain

$$
\begin{aligned}
p_{0,M}(x) &= c_{\nu,d} x^{d-1} \int_0^\infty \lambda^{-d/2} e^{-\frac{x^2}{2\lambda}} \frac{1}{1+\lambda^\nu} d\lambda \\
&= c_{\nu,d} x^{d-1} \int_0^\infty y^{d/2} e^{-y\frac{x^2}{2}} \frac{1}{y^2(1+y^{-\nu})} dy \\
&= c_{\nu,d} x^{d-1} \int_0^\infty e^{-y\frac{x^2}{2}} \frac{y^{\nu+d/2-2}}{y^\nu+1} dy \\
&= c_{\nu,d} x^{d-1} \int_0^\infty e^{-y\frac{x^2}{2}} y^{\nu+d/2-2} \left(1 - \frac{y^\nu}{y^\nu+1}\right) dy \\
&= c_{\nu,d} x^{d-1} \left(\int_0^\infty e^{-y\frac{x^2}{2}} y^{\nu+d/2-2} dy - R \right) \\
&= c_{\nu,d} x^{d-1} \left(2^{\nu+d/2-1} x^{-2\nu-d} \int_0^\infty e^{-z} z^{\nu+d/2-2} dz - R \right) \\
&= c_{\nu,d} \left(\Gamma(\nu+d/2-1) 2^{\nu+d/2-1} x^{-2\nu+1} - x^{d-1} R \right) \qquad (6.3.6)
\end{aligned}
$$

for $d \geq 1,\ \nu > 1$ with

$$
R := \int_0^\infty e^{-y\frac{x^2}{2}} \frac{y^{2\nu+d/2-2}}{y^\nu+1} dy.
$$

Now we investigate the asymptotic behavior of R. It is easy to see with elementary calculus that $\frac{y^{\nu-1/2}}{1+y^\nu} \leq 1$ for all $y > 0$, $\nu > 1$. Thus,

$$
\begin{aligned}
R &= \int_0^\infty e^{-y\frac{x^2}{2}} y^{\nu+d/2-3/2} \frac{y^{\nu-1/2}}{y^\nu+1} dy \\
&\leq \int_0^\infty e^{-y\frac{x^2}{2}} y^{\nu+d/2-3/2} dy \\
&= 2^{\nu+d/2} \Gamma(\nu+d/2-1/2) x^{-2\nu-d+1}.
\end{aligned}
$$

This means $R = \mathcal{O}(x^{-2\nu-d+1})$ for $x \to \infty$. Together with (6.3.6) this implies (6.3.5).

Step 2. We show

$$\int_0^\infty x^a(p_{1,M}(x)-p_{0,M}(x))^2dx \lesssim M^{\frac{2d+a-3}{2}}e^{-M\pi(1+2/\nu)} \tag{6.3.7}$$

for $M \to \infty$ all $d, \nu > 1,\ a \in \{0, 2\nu-1\}$.
Due to Theorems 2.2.4(i), 2.2.4(ii), (6.3.4) and (6.3.3) we have

$$\begin{aligned}\mathcal{M}[p_{1,M}-p_{0,M}](s) &= \frac{1}{\Gamma(d/2)}\Gamma\left(\frac{s+d-1}{2}\right)2^{\frac{s+1}{2}}\mathcal{M}[f_{1,M}-f_{0,M}]((s+1)/2)\\ &= \frac{\delta}{\Gamma(d/2)}\Gamma\left(\frac{s+d-1}{2}\right)2^{\frac{s+1}{2}}\mathcal{M}[q\odot\rho_M]((s+1)/2)\end{aligned} \tag{6.3.8}$$

for $\mathrm{Re}(s) \in (0, 2\nu+1)$. Choose

$$f(x) := x^a(p_{1,M}(x)-p_{0,M}(x)) \quad \text{and} \quad g(x) := p_{1,M}(x)-p_{0,M}(x)$$

in Theorem 2.2.9, then apply Theorem 2.2.4(iv) and (6.3.8) to get

$$\begin{aligned}&\int_0^\infty x^a(p_{0,M}(x)-p_{1,M}(x))^2dx\\ &\qquad= \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\mathcal{M}[p_{1,M}-p_{0,M}](z)\mathcal{M}[p_{1,M}-p_{0,M}](1+a-z)dz\\ &\qquad= \frac{\delta^2 2^{\frac{3+a}{2}}}{\pi i\Gamma(d/2)^2}\int_{\gamma-i\infty}^{\gamma+i\infty}\Gamma\left(\frac{z+d-1}{2}\right)\mathcal{M}[q\odot\rho_M]\left(\frac{z+1}{2}\right)\\ &\qquad\quad\times\Gamma\left(\frac{a+d-z}{2}\right)\mathcal{M}[q\odot\rho_M]\left(\frac{2+a-z}{2}\right)dz\end{aligned} \tag{6.3.9}$$

for $\gamma \in (0,1)$ in the case $a=0$ and for $\gamma \in (1, 2\nu)$ in the case $a = 2\nu-1$, where $\mathcal{M}[q\odot\rho_M] = \mathcal{M}[q]\mathcal{M}[\rho_M]$. Due to (6.3.2), we can estimate

$$|\mathcal{M}[\rho_M](u+iv)| \le e^{\frac{(u-1)^2}{2}}\frac{\varphi(v+M)+\varphi(v-M)}{2} \tag{6.3.10}$$

with $\varphi(v) = e^{-\frac{v^2}{2}}$. Due to (6.3.1) and the simple inequality

$$|\sin(x+iy)|^{-1} \le 2e^{-|y|}$$

for $x \ge \pi/6,\ y \in \mathbb{R}$ we have

$$|\mathcal{M}[q](u+iv)| = \frac{|\sin(\pi/\nu)|}{|\sin(\pi u/\nu + i\pi v/\nu)|} \le ce^{-\pi|v|/\nu} \tag{6.3.11}$$

for all $v \in \mathbb{R}$ and some constant c depending on u and ν. Next, use Lemma 2.4.3 in (6.3.9) to estimate the gamma terms, then plug in (6.3.10) and (6.3.11) to obtain

$$\begin{aligned}
&\int_0^\infty x^a(p_{1,M}(x) - p_{0,M}(x))^2 dx \\
&\lesssim \int\limits_{\{|v|\geq 2\}} \left|\Gamma\left(\frac{d+\gamma-1+iv}{2}\right)\mathcal{M}[q\odot\rho_M]\left(\frac{\gamma+1+iv}{2}\right)\right. \\
&\qquad \left.\times\,\Gamma\left(\frac{a+d-\gamma-iv}{2}\right)\mathcal{M}[q\odot\rho_M]\left(\frac{a-\gamma+2-iv}{2}\right)\right| dv + S_0 \\
&\lesssim \int\limits_{\{|v|\geq 2\}} |v|^{\frac{d+\gamma-2}{2}} e^{-|v|\pi/4} \left|\mathcal{M}[q\odot\rho_M]\left(\frac{\gamma+1+iv}{2}\right)\right| \\
&\qquad \times |v|^{\frac{a+d-\gamma-1}{2}} e^{-|v|\pi/4} \left|\mathcal{M}[q\odot\rho_M]\left(\frac{a-\gamma+2-iv}{2}\right)\right| dv + S_0 \\
&\lesssim \int\limits_{\{|v|\geq 2\}} |v|^{\frac{2d+a-3}{2}} e^{-|v|\pi/2-|v|\pi/\nu}(e^{-(v/2+M)^2/2} + e^{-(v/2-M)^2/2})^2 dv + S_0 \\
&\lesssim |2M|^{\frac{2d+a-3}{2}} e^{-|2M|\pi/2-|2M|\pi/\nu} + S_0 \\
&\lesssim M^{\frac{2d+a-3}{2}} e^{-M\pi(1+2/\nu)} + S_0 \qquad (6.3.12)
\end{aligned}$$

for $M \to \infty$ with

$$\begin{aligned}
S_0 := &\int\limits_{\{|v|\leq 2\}} \left|\Gamma\left(\frac{d+\gamma-1+iv}{2}\right)\mathcal{M}[q\odot\rho_M]\left(\frac{\gamma+1+iv}{2}\right)\right. \\
&\qquad \left.\times\,\Gamma\left(\frac{a+d-\gamma-iv}{2}\right)\mathcal{M}[q\odot\rho_M]\left(\frac{a-\gamma+2-iv}{2}\right)\right| \, dv.
\end{aligned}$$

S_0 turns out to be asymptotically negligible. To see this note that the gamma terms are maximal in $v = 0$. Hence, (6.3.10) and (6.3.11) imply

$$\begin{aligned}
S_0 &\lesssim \int\limits_{-2}^{2} \left|\mathcal{M}[q\odot\rho_M]\left(\frac{\gamma+1+iv}{2}\right)\mathcal{M}[q\odot\rho_M]\left(\frac{a-\gamma+2-iv}{2}\right)\right| \, dv \\
&\lesssim \int\limits_{-2}^{2} (e^{-(v/2+M)^2/2} + e^{-(v/2-M)^2/2})^2 \, dv \\
&\leq 4(e^{-(1+M)^2/2} + e^{-(-1-M)^2/2})^2 \\
&\lesssim e^{-M^2}
\end{aligned}$$

for $M \to \infty$. Thus, (6.3.12) implies (6.3.7).

Step 3. Now we show the claim. By (6.3.5) there is $x_0 > 0$ such that

$$p_{0,M}(x) \geq Cx^{-2\nu+1}$$

holds for all $x \geq x_0$ and some $C > 0$. Since $p_{0,M}(x) \geq c$ on $[0, x_0]$ for some $c > 0$, we can estimate

$$\begin{aligned}
\chi^2(p_{1,M}, p_{0,M}) &= \int_0^\infty \frac{(p_{M,1}(x) - p_{M,0}(x))^2}{p_{M,0}(x)} dx \\
&\leq c \int_0^{x_0} (p_{M,1}(x) - p_{M,0}(x))^2 dx + C \int_{x_0}^{\infty} x^{2\nu-1}(p_{M,1}(x) - p_{M,0}(x))^2 dx \\
&\leq c \int_0^{\infty} (p_{M,1}(x) - p_{M,0}(x))^2 dx + C \int_0^{\infty} x^{2\nu-1}(p_{M,1}(x) - p_{M,0}(x))^2 dx.
\end{aligned}$$

Finally, (6.3.7) implies

$$\chi^2(p_{1,M}, p_{0,M}) \lesssim M^{\frac{2d-3}{2}} e^{-M\pi(1+2/\nu)} + M^{\nu+d-2} e^{-M\pi(1+2/\nu)},$$

where $M^{\nu+d-2}e^{-M\pi(1+2/\nu)}$ is the dominating term for $M \to \infty$. This completes the proof. □

Let us now discuss why the rate $n^{\frac{-\beta}{\pi+2\beta}}$ from Theorem 3.7.3(i) is optimal for the class $\mathcal{C}(\beta, L, \gamma)$. Lemma 6.3.3 shows that

$$(1 + \chi^2(p_{1,M}, p_{0,M}))^M - 1 \lesssim (1 + M^{\nu+d-2}e^{-M\pi(1+2/\nu)})^M - 1 \lesssim (1 + M^{-1})^M - 1 \leq \alpha$$

for $M \to \infty$, all $d \geq 1$, $\nu > 1$ and some $\alpha < \infty$. So, we have (6.2.8) and Theorem 6.2.2 implies $p_e \geq e^{-\alpha}/4 > 0$. Hence, condition (6.2.6) from Theorem 6.2.1 is satisfied. Choose

$$M = \frac{1}{\pi(1 + 2/\nu)} \log(n).$$

In order to check (6.2.5), apply Lemma 6.3.2 to get

$$\begin{aligned}
d_x(f_{M,0}, f_{M,1}) &= |f_{M,0}(x) - f_{M,1}(x)| \\
&\gtrsim \exp(-M\pi/\nu) \\
&= \exp\left(-\frac{1}{2+\nu} \log(n)\right) \\
&= n^{-\frac{\pi/\nu}{\pi+2\pi/\nu}}
\end{aligned}$$

for some $x > 0$ and all $\nu > 1$. Now set $\beta = \pi/\nu$ for $\nu > 1$. Theorem 6.2.1 implies that $n^{-\frac{\beta}{\pi+2\beta}}$ is the optimal rate for $\mathcal{C}(\beta, L, \gamma)$ for all $\beta \in (0, \pi)$, $\gamma > 0$ with L depending on γ.

6.3.2. Observation of a Self-Similar Gaussian Process. In this subsection we prove that the rate $n^{\frac{-2H\beta}{\pi+4H\beta}}$ from Theorem 3.10.2(i) is optimal for $(\mathcal{C}(\beta,\gamma,L),d_x)$ in the sense of Definition 6.1.1 for all $\beta\in(0,\pi)$, $\gamma>0$ with L depending on γ. Our strategy here is the same as in Subsection 6.3.1. In fact, there we considered an arbitrary dimension $d\geq 1$ and $H=1/2$. Here we consider a scaling parameter $H\in(0,2)$ and $d=1$.
Let $T_{0,M}$ and $T_{1,M}$ be two random variables with densities $f_{0,M}$ and $f_{1,M}$ given by (6.3.3) respectively. Let $(Y_t)_{t\geq 0}$ be as in Subsection 3.10.1, that is a $\mathbb{R}$-valued self-similar process with scaling factor $H\in(0,2)$, càdlàg paths and standard normally distributed at time 1. Hence, the marginal densities of $(|Y_t|)_{t\geq 0}$ are given by

$$f_t(x)=\frac{2}{t^H\sqrt{2\pi}}e^{-\frac{x^2}{2t^{2H}}},\quad x>0.$$

By Lemma 3.1.3, the density of $|Y_{T_{i,M}}|$, $i=0,1$ is then given by

$$p_{i,M}(x)=\frac{2}{\sqrt{2\pi}}\int_0^\infty \lambda^{-H}e^{-\frac{x^2}{2\lambda^{2H}}}f_{i,M}(\lambda)d\lambda,\quad x>0, H\in(0,2), i=0,1.$$

For the Mellin transform of $p_{i,M}$ we get

$$\begin{aligned}\mathcal{M}[p_{i,M}](s)&=\mathcal{M}[|Y_1|](s)\mathcal{M}[T_{i,M}](Hs-H+1)\\ &=\frac{2^{s/2}}{\sqrt{2\pi}}\Gamma\left(\frac{s}{2}\right)\mathcal{M}[f_{i,M}](Hs-H+1),\quad i=0,1\end{aligned}\tag{6.3.13}$$

for $\mathrm{Re}(s)>0$ and $\mathrm{Re}(s)\in(\frac{H-1}{H},\frac{H+\nu}{H})$.

LEMMA 6.3.4. *The χ^2-distance between the densities $p_{0,M}$ and $p_{1,M}$ fulfills*

$$\chi^2(p_{1,M},p_{0,M})\lesssim M^{\frac{\nu-1}{2H}}e^{-M\pi(1/(2H)+2/\nu)}$$

for $M\to\infty$ and all $H\in(0,2)$, $\nu>1$.

PROOF. This proof is again similar to the one of [**7**, Lemma 6.4], where the special case $H=1/2$ is treated.

Step 1. We show

$$p_{0,M}(x)\gtrsim x^{\frac{1-\nu-H}{H}},\quad x\to\infty.\tag{6.3.14}$$

To this end we perform the change of variables $y=\lambda^{-2H}$ in the second equality sign of the following calculation and $z=y\frac{x^2}{2}$ in the sixth to obtain

$$
\begin{aligned}
p_{0,M}(x) &= \frac{2}{\sqrt{2\pi}} \frac{\nu \sin(\pi/\nu)}{\pi} \int\limits_0^\infty \lambda^{-H} e^{-\frac{x^2}{2\lambda^{2H}}} \frac{1}{1+\lambda^\nu} d\lambda \\
&= \frac{1}{H\sqrt{2\pi}} \frac{\nu \sin(\pi/\nu)}{\pi} \int\limits_0^\infty y^{-\frac{H+1}{2H}} e^{-y\frac{x^2}{2}} \frac{1}{1+y^{-\frac{\nu}{2H}}} dy \\
&= \frac{1}{H\sqrt{2\pi}} \frac{\nu \sin(\pi/\nu)}{\pi} \int\limits_0^\infty y^{\frac{\nu-H-1}{2H}} e^{-y\frac{x^2}{2}} \frac{1}{y^{\frac{\nu}{2H}}+1} dy \\
&= \frac{1}{H\sqrt{2\pi}} \frac{\nu \sin(\pi/\nu)}{\pi} \int\limits_0^\infty y^{\frac{\nu-H-1}{2H}} e^{-y\frac{x^2}{2}} \left(1 - \frac{y^{\frac{\nu}{2H}}}{y^{\frac{\nu}{2H}}+1}\right) dy \\
&= \frac{1}{H\sqrt{2\pi}} \frac{\nu \sin(\pi/\nu)}{\pi} \left(\int\limits_0^\infty y^{\frac{\nu-H-1}{2H}} e^{-y\frac{x^2}{2}} dy - R\right) \\
&= \frac{1}{H\sqrt{2\pi}} \frac{\nu \sin(\pi/\nu)}{\pi} \left(2^{\frac{\nu+H-1}{2H}} x^{\frac{-2\nu-2H+2}{2H}} \int\limits_0^\infty z^{\frac{\nu+H-1}{2H}} e^{-z} dz - R\right) \\
&= \frac{1}{H\sqrt{2\pi}} \frac{\nu \sin(\pi/\nu)}{\pi} \left(\Gamma\left(\frac{\nu+H-1}{2H}\right) 2^{\frac{\nu+H-1}{2H}} x^{\frac{1-\nu-H}{H}} - R\right), \qquad (6.3.15)
\end{aligned}
$$

where

$$
R := \int_0^\infty y^{\frac{\nu-H-1}{2H}} e^{-y\frac{x^2}{2}} \frac{y^{\frac{\nu}{2H}}}{y^{\frac{\nu}{2H}}+1} dy.
$$

We need to show that R tends to zero faster than $x^{\frac{1-\nu-H}{H}}$ for $x \to \infty$. Elementary calculus shows that $\frac{y^{\frac{\nu}{2H}-\frac{1}{H}}}{y^{\frac{\nu}{2H}}+1} \le 1$ for all $y > 0$, $H \in (0,2)$, $\nu > 1$. Hence,

$$
\begin{aligned}
R &= \int_0^\infty y^{\frac{\nu+H-1}{2H}} e^{-y\frac{x^2}{2}} \frac{y^{\frac{\nu}{2H}-\frac{1}{H}}}{y^{\frac{\nu}{2H}}+1} dy \\
&\le \int_0^\infty y^{\frac{\nu+H-1}{2H}} e^{-y\frac{x^2}{2}} dy \\
&= 2^{\frac{\nu+3H-1}{2H}} \Gamma\left(\frac{\nu+3H-1}{2H}\right) x^{\frac{-\nu-3H+1}{H}}.
\end{aligned}
$$

This means $R = \mathcal{O}(x^{\frac{1-\nu-H}{H}-2})$, $x \to \infty$. Together with (6.3.15) this implies (6.3.14).

Step 2. We show

$$\int_0^\infty x^a(p_{1,M}(x)-p_{0,M}(x))^2dx \lesssim M^{\frac{a-1}{2}}e^{-M\pi(1/(2H)+2/\nu)} \tag{6.3.16}$$

for $M\to\infty$ and all $\nu>1$, $H\in(0,2)$, $a\in\{0,\frac{\nu+H-1}{H}\}$.
Due to Theorem 2.2.4(i), (2.2.4)(ii), (6.3.13) and (6.3.3) we have

$$\begin{aligned}\mathcal{M}[p_{1,M}-p_{0,M}](s) &= \frac{2^{s/2}}{\sqrt{2\pi}}\Gamma\left(\frac{s}{2}\right)\mathcal{M}[f_{1,M}-f_{0,M}](Hs-H+1)\\ &= \frac{\delta 2^{s/2}}{\sqrt{2\pi}}\Gamma\left(\frac{s}{2}\right)\mathcal{M}[q\odot\rho_M](Hs-H+1)\end{aligned} \tag{6.3.17}$$

for $\mathrm{Re}(s)\in(0,\infty)\cap(\frac{H-1}{H},\frac{H+\nu}{H})$.
Now choose $f(x):=x^a(p_{1,M}(x)-p_{0,M}(x))$ and $g(x):=p_{1,M}(x)-p_{0,M}(x)$ in Theorem 2.2.9, then apply Theorem 2.2.4iv) and (6.3.17) to get

$$\begin{aligned}&\int_0^\infty x^a(p_{0,M}(x)-p_{1,M}(x))^2dx\\ &\quad=\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\mathcal{M}[p_{1,M}-p_{0,M}](z)\mathcal{M}[p_{1,M}-p_{0,M}](1-z+a)dz\\ &\quad=\frac{\delta^2 2^{\frac{a+1}{2}}}{(2\pi)^2 i}\int_{\gamma-i\infty}^{\gamma+i\infty}\Gamma\left(\frac{z}{2}\right)\mathcal{M}[q\odot\rho_M]\,(Hz-H+1)\\ &\qquad\times\Gamma\left(\frac{1-z+a}{2}\right)\mathcal{M}[q\odot\rho_M]\,(1+Ha-Hz)\,dz\end{aligned} \tag{6.3.18}$$

for $\gamma\in(0,a+1)\cap(\frac{H-1}{H},\frac{H+\nu}{H})\cap(\frac{1+Ha-\nu}{H},\frac{1+Ha}{H})$, where

$$\mathcal{M}[q\odot\rho_M]=\mathcal{M}[q]\mathcal{M}[\rho_M].$$

It is straightforward to check that a suitable γ can be found for all parameters as in (6.3.16). This is, where we require the constraint $H<2$. Otherwise we would get a problem in the case $a=0$, where we require $\frac{H-1}{H}<\frac{1}{H}$, which is only true for $H<2$.

Next, use Lemma 2.4.3 in (6.3.18) to estimate the gamma terms, then plug in (6.3.10) and (6.3.11) to obtain

$$\begin{aligned}
&\int_0^\infty x^a (p_{1,M}(x) - p_{0,M}(x))^2 dx \\
&\quad \lesssim \int\limits_{\{|v|\geq 2\}} \left| \Gamma\left(\frac{\gamma}{2} + \frac{v}{2}i\right) \mathcal{M}[q \odot \rho_M]\left(H\gamma - H + 1 - Hvi\right) \right. \\
&\qquad \left. \times\, \Gamma\left(\frac{1-\gamma+a}{2} - \frac{v}{2}i\right) \mathcal{M}[q \odot \rho_M]\left(1 + Ha - H\gamma - Hvi\right) \right| dv + S_0 \\
&\quad \lesssim \int\limits_{\{|v|\geq 2\}} |v|^{\frac{\gamma-1}{H}} e^{-|v|\pi/4} \left|\mathcal{M}[q \odot \rho_M]\left(1 + iHv\right)\right| \\
&\qquad \times |v|^{\frac{a-\gamma}{H}} e^{-|v|\pi/4} \left|\mathcal{M}[q \odot \rho_M]\left(1 + Ha - H - iHv\right)\right| dv + S_0 \\
&\quad \lesssim \int\limits_{\{|v|\geq 2\}} |v|^{\frac{a-1}{H}} e^{-|v|\pi/2 - 2H|v|\pi/\nu} (e^{-(Hv+M)^2/2} + e^{-(Hv-M)^2/2})^2 dv + S_0 \\
&\quad \lesssim |M/H|^{\frac{a-1}{2}} e^{-M\pi/(2H) - 2M\pi/\nu} + S_0
\end{aligned}$$

$$\lesssim M^{\frac{a-1}{2}} e^{-M\pi(1/(2H)+2/\nu)} + S_0 \tag{6.3.19}$$

for $M \to \infty$, where

$$\begin{aligned}
S_0 := \int\limits_{\{|v|\leq 2\}} &\left| \Gamma\left(\frac{\gamma}{2} + \frac{v}{2}i\right) \mathcal{M}[q \odot \rho_M]\left(H\gamma - H + 1 - Hvi\right) \right. \\
&\left. \times\, \Gamma\left(\frac{1-\gamma+a}{2} - \frac{v}{2}i\right) \mathcal{M}[q \odot \rho_M]\left(1 + Ha - H\gamma - Hvi\right) \right| \, dv.
\end{aligned}$$

S_0 is asymptotically negligible. To see this note that the gamma terms are maximal in $v = 0$. Hence, due to (6.3.10) and (6.3.11),

$$\begin{aligned}
S_0 &\lesssim \int\limits_{-2}^{2} \left|\mathcal{M}[q \odot \rho_M]\left(H\gamma - H + 1 - Hvi\right) \mathcal{M}[q \odot \rho_M]\left(1 + Ha - H\gamma - Hvi\right)\right| \, dv \\
&\lesssim \int\limits_{-2}^{2} (e^{-(Hv+M)^2/2} + e^{-(Hv-M)^2/2})^2 \, dv \\
&\leq 4(e^{-(2H+M)^2/2} + e^{-(-2H-M)^2/2})^2 \\
&\lesssim e^{-M^2}
\end{aligned}$$

for $M \to \infty$. Thus, (6.3.19) implies (6.3.16).

Step 3. Now we show the claim. By (6.3.14) there is $x_0 > 0$ such that

$$p_{0,M}(x) \geq Cx^{\frac{1-\nu-H}{H}}$$

for all $x \geq x_0$ and some $C > 0$. Since $p_{0,M}(x) \geq c$ on $[0, x_0]$ for some $c > 0$,

$$\begin{aligned}\chi^2(p_{1,M}, p_{0,M}) &= \int_0^\infty \frac{(p_{M,1}(x) - p_{M,0}(x))^2}{p_{M,0}(x)} dx \\ &\leq c \int_0^{x_0} (p_{M,1}(x) - p_{M,0}(x))^2 dx + C \int_{x_0}^{\infty} x^{\frac{\nu+H-1}{H}} (p_{M,1}(x) - p_{M,0}(x))^2 dx \\ &\leq c \int_0^{\infty} (p_{M,1}(x) - p_{M,0}(x))^2 dx + C \int_0^{\infty} x^{\frac{\nu+H-1}{H}} (p_{M,1}(x) - p_{M,0}(x))^2 dx.\end{aligned}$$

Finally, (6.3.16) implies

$$\chi^2(p_{1,M}, p_{0,M}) \lesssim M^{-1/2} e^{-M\pi(1/(2H)+2/\nu)} + M^{\frac{\nu-1}{2H}} e^{-M\pi(1/(2H)+2/\nu)},$$

where $M^{\frac{\nu-1}{2H}} e^{-M\pi(1/(2H)+2/\nu)}$ is the dominating term for $M \to \infty$. This completes the proof. □

Let us now discuss why the rate $n^{-\frac{2H\beta}{\pi+4H\beta}}$ from Theorem 3.10.2(i) is optimal for the class $\mathcal{C}(\beta, \gamma, L)$. Lemma 6.3.4 shows that

$$(1+\chi^2(p_{1,M}, p_{0,M}))^M - 1 \lesssim (1+M^{\frac{\nu-1}{2H}} e^{-M\pi(1/(2H)+2/\nu)})^M - 1 \lesssim (1+M^{-1})^M - 1 \leq \alpha$$

for $M \to \infty$, all $d \geq 1$, $H \in (0, 2)$ and some $\alpha < \infty$. So, we have (6.2.8) and Theorem 6.2.2 implies $p_e \geq e^{-\alpha}/4 > 0$. Hence, condition (6.2.6) from Theorem 6.2.1 is satisfied. Choose

$$M = \frac{1}{\pi(1/2H + 2/\nu)} \log(n).$$

In order to check (6.2.5) apply Lemma 6.3.2 to get

$$\begin{aligned}d_x(f_{M,0}, f_{M,1}) &= |f_{M,0}(x) - f_{M,1}(x)| \\ &\gtrsim \exp(-M\pi/\nu) \\ &= \exp\left(-\frac{1}{\nu(1/(2H) + 2/\nu)} \log(n)\right) \\ &= n^{-\frac{2H\pi/\nu}{\pi+4H\pi/\nu}}\end{aligned}$$

for some $x > 0$ and all $\nu > 1$. Now set $\beta = \pi/\nu$ ($\nu > 1$). Theorem 6.2.1 implies that $n^{-\frac{\beta}{\pi+2\beta}}$ is the optimal rate for $\mathcal{C}(\beta, \gamma, L)$ up to a logarithmic factor for all $\beta \in (0, \pi)$, $\gamma > 0$ with L depending on γ.

6.3.3. Observation of a Squared Bessel Process. In this subsection we prove that the rate $n^{-\frac{\beta}{\pi/H+2\beta}}$ from Theorem 3.10.4(i) is optimal for $(\mathcal{C}(\beta, \gamma, L), d_x)$ in the sense of Definition 6.1.1 for all $\beta \in (0, \pi)$, $\gamma > 0$ with L depending on γ. Our strategy here is the same as in Subsections 6.3.1 and 6.3.2.
Let $T_{0,M}$ and $T_{1,M}$ be two random variables with densities $f_{0,M}$ and $f_{1,M}$ given by (6.3.3) respectively. Let $(Y_t)_{t\geq 0}$ be as in Subsection 3.10.2, that is a $\mathbb{R}_+$-valued self-similar process with scaling factor $H \in (0, 2)$, càdlàg paths and Gamma-distributed with shape parameter $\sigma > \max\{0, 1 - 1/H\}$ and rate parameter $r = 1$ at time 1. Hence, the marginal densities of $(Y_t)_{t\geq 0}$ are given by

$$f_t(x) = \frac{1}{t^{H\sigma}\Gamma(\sigma)} x^{\sigma-1} e^{-\frac{x}{t^H}}, \quad x > 0.$$

By Lemma 3.1.3, the density of $Y_{T_{i,M}}$, $i = 0, 1$ is then given by

$$p_{i,M}(x) = \frac{1}{\Gamma(\sigma)} x^{\sigma-1} \int_0^\infty \lambda^{-H\sigma} e^{-\frac{x}{\lambda^H}} f_{i,M}(\lambda) d\lambda, \quad x, H, \sigma > 0, i = 0, 1.$$

For the Mellin transform of $p_{i,M}$ we get

$$\begin{aligned} \mathcal{M}[p_{i,M}](s) &= \mathcal{M}[Y_1](s)\mathcal{M}[T_{i,M}](Hs - H + 1) \\ &= \frac{\Gamma(s + \sigma - 1)}{\Gamma(\sigma)} \mathcal{M}[f_{i,M}](Hs - H + 1), \quad i = 0, 1 \end{aligned} \tag{6.3.20}$$

for $\mathrm{Re}(s) > 1 - \sigma$ and $\mathrm{Re}(s) \in (\frac{H-1}{H}, \frac{\nu+H-1}{H})$.

LEMMA 6.3.5. *The χ^2-distance between the densities $p_{0,M}$ and $p_{1,M}$ fulfills*

$$\chi^2(p_{1,M}, p_{0,M}) \lesssim M^{2\sigma-2+\frac{\nu+H-1}{H}} e^{-M\pi(1/H+2/\nu)}$$

for $M \to \infty$ and all $H \in (0, 2)$ with $H < \frac{1}{1-\sigma}$ and $\nu > 1$.

PROOF. **Step 1**. We show

$$p_{0,M}(x) \gtrsim x^{1/H-\nu/H-1}, \quad x \to \infty. \tag{6.3.21}$$

To this end define $c_{\nu,H,\sigma} := \frac{1}{\Gamma(\sigma)} \frac{\nu \sin(\pi/\nu)}{\pi}$ and perform the change of variables $y = \lambda^{-H}$ in the second equality sign of the following calculation and $z = yx$ in the sixth to obtain

$$\begin{aligned}
p_{0,M}(x) &= c_{\nu,H,\sigma} x^{\sigma-1} \int_0^\infty \lambda^{-H\sigma} e^{-\frac{x}{\lambda^H}} \frac{1}{1+\lambda^\nu} d\lambda \\
&= \frac{c_{\nu,H,\sigma}}{H} x^{\sigma-1} \int_0^\infty y^{\sigma-1-1/H} e^{-yx} \frac{1}{1+y^{-\nu/H}} dy \\
&= \frac{c_{\nu,H,\sigma}}{H} x^{\sigma-1} \int_0^\infty y^{\sigma-1-1/H+\nu/H} e^{-yx} \frac{1}{1+y^{\nu/H}} dy \\
&= \frac{c_{\nu,H,\sigma}}{H} x^{\sigma-1} \int_0^\infty y^{\sigma-1-1/H+\nu/H} e^{-yx} \left(1 - \frac{y^{\nu/H}}{y^{\nu/H}+1}\right) dy \\
&= \frac{c_{\nu,H,\sigma}}{H} x^{\sigma-1} \left(\int_0^\infty y^{\sigma-1-1/H+\nu/H} e^{-yx} dy - R\right) \\
&= \frac{c_{\nu,H,\sigma}}{H} x^{\sigma-1} \left(x^{-\sigma+1/H-\nu/H} \int_0^\infty z^{\sigma-1-1/H+\nu/H} e^{-z} dz - R\right) \\
&= \frac{c_{\nu,H,\sigma}}{H} \left(\Gamma\left(\sigma - 1/H + \nu/H\right) x^{1/H-\nu/H-1} - x^{\sigma-1} R\right), \qquad (6.3.22)
\end{aligned}$$

where

$$R := \int_0^\infty y^{\sigma-1-1/H+\nu/H} e^{-yx} \frac{y^{\nu/H}}{y^{\nu/H}+1} dy.$$

We need to show that $x^{\sigma-1}R$ tends faster to zero than $x^{1/H-\nu/H-1}$ for $x \to \infty$. Elementary calculus shows that $\frac{y^{\nu/H-2/H}}{y^{\nu/H}+1} \leq 1$ for all $y > 0$, $H \in (0,2)$, $\nu > 1$. Thus,

$$\begin{aligned}
R &= \int_0^\infty y^{\sigma-1+1/H+\nu/H} e^{-yx} \frac{y^{\nu/H-2/H}}{y^{\nu/H}+1} dy \\
&\leq \int_0^\infty y^{\sigma-1+1/H+\nu/H} e^{-yx} dy \\
&= \Gamma\left(\sigma + 1/H + \nu/H\right) x^{-\sigma-1/H-\nu/H}.
\end{aligned}$$

This means $x^{\sigma-1}R = \mathcal{O}(x^{-1/H-\nu/H-1})$, $x \to \infty$. Together with (6.3.22) this implies (6.3.21).

Step 2. We show

$$\int_0^\infty x^a (p_{1,M}(x) - p_{0,M}(x))^2 dx \lesssim M^{2\sigma+a-2} e^{-M\pi(1/H+2/\nu)} \qquad (6.3.23)$$

for $M \to \infty$ and all $\nu > 1$, $H \in (0,2)$, $\sigma > \max\{0, 1-1/H\}$ and $a \in \{0, -1/H + \nu/H + 1\}$.

Due to Theorem 2.2.4(i), (2.2.4)(ii), (6.3.20) and (6.3.3) we have

$$\begin{aligned}\mathcal{M}[p_{1,M}-p_{0,M}](s) &= \frac{\Gamma(s+\sigma-1)}{\Gamma(\sigma)}\mathcal{M}[f_{1,M}-f_{0,M}](Hs-H+1)\\ &= \frac{\delta\Gamma(s+\sigma-1)}{\Gamma(\sigma)}\mathcal{M}[q\odot\rho_M](Hs-H+1) \qquad (6.3.24)\end{aligned}$$

for $\mathrm{Re}(s)\in(1-\sigma,\infty)\cap(\frac{H-1}{H},\frac{\nu+H-1}{H})$.
Choose $f(x):=x^a(p_{1,M}(x)-p_{0,M}(x))$ and $g(x):=p_{1,M}(x)-p_{0,M}(x)$ in Theorem 2.2.9, then apply Theorem 2.2.4iv) and (6.3.24) to get

$$\begin{aligned}&\int_0^\infty x^a(p_{0,M}(x)-p_{1,M}(x))^2dx\\ &\qquad = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\mathcal{M}[p_{1,M}-p_{0,M}](z)\mathcal{M}[p_{1,M}-p_{0,M}](1-z+a)dz\\ &\qquad = \frac{\delta^2}{2\pi\Gamma(\sigma)^2 i}\int_{\gamma-i\infty}^{\gamma+i\infty}\Gamma\left(z+\sigma-1\right)\mathcal{M}[q\odot\rho_M]\left(Hz-H+1\right)\\ &\qquad\quad \times\Gamma\left(a+\sigma-z\right)\mathcal{M}[q\odot\rho_M]\left(1+Ha-Hz\right)dz \qquad (6.3.25)\end{aligned}$$

for $\gamma\in(1-\sigma,a+\sigma)\cap(\frac{H-1}{H},\frac{\nu+H-1}{H})\cap(\frac{1+Ha-\nu}{H},\frac{Ha+1}{H})$. It is straightforward to show that a suitable γ exists, provided $H<2$ and $H<\frac{1}{1-\sigma}$. If we have $\sigma>\frac{1}{2}$, then $H<\frac{1}{1-\sigma}$ is obsolete.
Next, use Lemma 2.4.3 in (6.3.25) to estimate the gamma terms, then plug in (6.3.10) and (6.3.11) to obtain

$$\begin{aligned}&\int_0^\infty x^a(p_{1,M}(x)-p_{0,M}(x))^2dx\\ &\lesssim \int_{\{|v|\geq 2\}}|\Gamma\left(\gamma+\sigma-1+iv\right)\mathcal{M}[q\odot\rho_M]\left(H\gamma-H+1+Hvi\right)\\ &\quad\times\Gamma\left(a+\sigma-\gamma-iv\right)\mathcal{M}[q\odot\rho_M]\left(Ha-H\gamma+1-Hvi\right)|\,dv+S_0\\ &\lesssim \int_{\{|v|\geq 2\}}|v|^{\gamma+\sigma-3/2}e^{-|v|\pi/2}\left|\mathcal{M}[q\odot\rho_M]\left(H\gamma-H+1+Hvi\right)\right|\\ &\quad\times|v|^{a+\sigma-\gamma-1/2}e^{-|v|\pi/2}\left|\mathcal{M}[q\odot\rho_M]\left(Ha-H\gamma+1-Hvi\right)\right|dv+S_0\\ &\lesssim \int_{\{|v|\geq 2\}}|v|^{2\sigma+a-2}e^{-|v|\pi-2H|v|\pi/\nu}(e^{-(Hv+M)^2/2}+e^{-(Hv-M)^2/2})^2dv+S_0\\ &\lesssim |M/H|^{2\sigma+a-2}e^{-M\pi/H-2M\pi/\nu}+S_0\\ &\lesssim M^{2\sigma+a-2}e^{-M\pi(1/H+2/\nu)}+S_0 \qquad (6.3.26)\end{aligned}$$

for $M \to \infty$, where

$$S_0 := \int_{\{|v|\leq 2\}} |\Gamma(\gamma+\sigma-1+iv)\,\mathcal{M}[q\odot\rho_M]\,(H\gamma-H+1+Hvi) \times \Gamma(a+\sigma-\gamma-iv)\,\mathcal{M}[q\odot\rho_M]\,(Ha-H\gamma+1-Hvi)|\ dv.$$

S_0 turns out to be asymptotically negligible. To see this note that the gamma terms are maximal in $v=0$. Hence, due to (6.3.10) and (6.3.11),

$$\begin{aligned} S_0 &\lesssim \int_{-2}^{2} |\mathcal{M}[q\odot\rho_M]\,(H\gamma-H+1+Hvi)\,\mathcal{M}[q\odot\rho_M]\,(Ha-H\gamma+1-Hvi)|\ dv \\ &\lesssim \int_{-2}^{2} (e^{-(Hv+M)^2/2}+e^{-(Hv-M)^2/2})^2\ dv \\ &\leq 4(e^{-(2H+M)^2/2}+e^{-(-2H-M)^2/2})^2 \\ &\lesssim e^{-M^2} \end{aligned}$$

for $M\to\infty$. Thus, (6.3.12) implies (6.3.7).

Step 3. Now we show the claim. By (6.3.21) there is $x_0>0$ such that

$$p_{0,M}(x) \geq Cx^{\frac{1-\nu-H}{H}}$$

for all $x\geq x_0$ and some $C>0$. Since $p_{0,M}(x)\geq c$ on $[0,x_0]$ for some $c>0$,

$$\begin{aligned} \chi^2(p_{1,M},p_{0,M}) &= \int_0^\infty \frac{(p_{M,1}(x)-p_{M,0}(x))^2}{p_{M,0}(x)}dx \\ &\leq c\int_0^{x_0}(p_{M,1}(x)-p_{M,0}(x))^2dx + C\int_{x_0}^{\infty} x^{\frac{\nu+H-1}{H}}(p_{M,1}(x)-p_{M,0}(x))^2dx \\ &\leq c\int_0^{\infty}(p_{M,1}(x)-p_{M,0}(x))^2dx + C\int_{0}^{\infty} x^{\frac{\nu+H-1}{H}}(p_{M,1}(x)-p_{M,0}(x))^2dx. \end{aligned}$$

Finally, (6.3.23) implies

$$\chi^2(p_{1,M},p_{0,M}) \lesssim M^{2\sigma-2}e^{-M\pi(1/H+2/\nu)} + M^{2\sigma-2+\frac{\nu+H-1}{H}}e^{-M\pi(1/H+2/\nu)},$$

where $M^{2\sigma-2+\frac{\nu+H-1}{H}}e^{-M\pi(1/H+2/\nu)}$ is the dominating term for $M\to\infty$. This completes the proof. □

Let us now discuss why the rate $n^{-\frac{\beta}{\pi/H+2\beta}}$ from Theorem 3.10.4(i) is optimal for the class $\mathcal{C}(\beta,\gamma,L)$. Lemma 6.3.4 shows that

$$\begin{aligned}(1+\chi^2(p_{1,M},p_{0,M}))^M-1 &\lesssim (1+M^{2\sigma-2+\frac{\nu+H-1}{H}}e^{-M\pi(1/H+2/\nu)})^M-1\\ &\lesssim (1+M^{-1})^M-1\\ &\leq \alpha\end{aligned}$$

for $M\to\infty$, all $d\geq 1$, $H\in(0,2)$ and some $\alpha<\infty$. So, we have (6.2.8) and Theorem 6.2.2 implies $p_e\geq e^{-\alpha}/4>0$. Hence, condition (6.2.6) from Theorem 6.2.1 is satisfied. Choose

$$M=\frac{1}{\pi(1/H+2/\nu)}\log(n).$$

In order to check (6.2.5) apply Lemma 6.3.2 to get

$$\begin{aligned}d_x(f_{M,0},f_{M,1}) &= |f_{M,0}(x)-f_{M,1}(x)|\\ &\gtrsim \exp(-M\pi/\nu)\\ &= \exp\left(-\frac{1}{\nu(1/H+2/\nu)}\log(n)\right)\\ &= n^{-\frac{\pi/nu}{\pi/H+2\pi/\nu}}\end{aligned}$$

for some $x>0$ and all $\nu>1$. Now set $\beta=\pi/\nu$ ($\nu>1$). Theorem 6.2.1 implies that $n^{-\frac{\beta}{\pi/H+2\beta}}$ is the optimal rate for $\mathcal{C}(\beta,\gamma,L)$ up to a logarithmic factor for all $\beta\in(0,\pi)$, $\gamma>0$ with L depending on γ.

6.4. Optimality for the Class $\mathcal{D}$

It remains an open question whether the rate $(\log n)^{-\beta}$ appearing in Theorems 3.7.3(ii), 3.10.2(ii), 3.10.4(ii) and 5.2.8 is optimal for the class $\mathcal{D}(\beta,L,\gamma)$. In order to prove optimality in the same fashion as we did for $\mathcal{C}(\beta,L,\gamma)$ we would need to find hypotheses $q_{n,0}$ and $q_{n,1}$ in $\mathcal{D}(\beta,L,\gamma)$ satisfying the conditions of Theorem 6.2.1. For the case of an observed Bessel processes with dimension $d=1$ the authors of [**7**] suggest the following construct which, even after a thorough contemplation, we find incomprehensible:
Define for any $\nu>1$ and $M>0$,

$$q(x):=[2\Gamma(\nu)]^{-1}\times\begin{cases}\log^{\nu-1}(1/x), & 0\leq x\leq 1\\ x^{-2}\log^{\nu-1}(x), & x>1\end{cases}$$

and

$$\rho_M(x):=\frac{1}{2\pi}e^{-\frac{\log^2(x)}{2}}\frac{\sin(M\log(x))}{x\log(x)},\quad x\geq 0.$$

Set now for any $M>0$ and some $\delta>0$,

$$q_{0,M}(x):=q(x),\quad q_{1,M}(x):=(1-\delta\zeta_M)q(x)+\delta(q\odot\rho_M)(x),$$

where $q \odot \rho_M$ is defined by (2.2.1) and

$$\zeta_M = \frac{1}{\sqrt{2\pi}} \int_{-M}^{M} e^{-\frac{x^2}{2}} dx.$$

M and ν are later chosen depending on n and β respectively. For instance, the following (see [**7**, Lemma 6.6]) is claimed using these definitions:

LEMMA 6.4.1. *For any $M > 0$ and some $\delta > 0$ not depending on M, the function $q_{1,M}$ is a probability density satisfying*

$$\sup_{x\in(1-\varrho,1+\varrho)} |q_{0,M}(x)-q_{1,M}(x)| = |\cos(\pi\nu/2)|M^{-\nu+1}+\mathcal{O}(M^{-\nu}), \quad M \to \infty, \tag{6.4.1}$$

where $\varrho > 0$ is a fixed number.

In fact,

$$|q_{0,M}(x) - q_{1,M}(x)| = |-\delta\zeta_M q(x) + \delta(q \odot \rho_M)(x)|$$

and $(q \odot \rho_M)(x)$ is shown in [**7**, Proof of Lemma 6.6] to behave like the right hand side of (6.4.1). But clearly, we have

$$\delta\zeta_M q(x) \geq C > 0$$

for some $C > 0$, all $M > 0$ and all $x \neq 1$. Thus, (6.4.1) is unclear.

The search for other hypotheses and a proof of optimality for $\mathcal{D}(\beta, L, \gamma)$ should be the focus of a further investigation.

APPENDIX

R Source Code

Below we implement the adaptive estimators presented in Sections 3.9 and 5.3 using the freeware R version 3.4.2 (2017-09-28). We need the packages “pracma”, “sfsmisc” and “SuppDists”.

Estimator for Bessel Processes Based on Self-Similarity

(With Oracle Cut-Off Parameter, See Section 3.9)

```
sup.dist <- function(x1, x2) {max(abs(x1-x2))}
l1.dist<- function(f1,f2){sum(abs(f1-f2))/0.1}
d <- 5
gamma <- 0.7
h_n <- seq(0.1,4,0.1)
n <- 1000
minloss2<- matrix(10,nrow=100,ncol=4)
arg<- seq(-4,4,0.001)
schaetervonxh<-matrix(0,nrow = 100,ncol = length(h_n))

#Simulate Observations
BeobT <- rgamma(n,shape=2, scale = 1) #Here T Gamma distributed
BeobBES1 <- sqrt(rchisq(n,d))
BeobX<- sqrt(BeobT)*BeobBES1

integrand <- function(v){sum(BeobX^(2*(gamma+1i*v-1)))
   /gammaz(gamma+d/2-1+1i*v)/2^(gamma+1i*v) }
integrand<-Vectorize(integrand)
integrand.werte<- integrand(arg)

schaetzerspec <- function(x,h){
integrand1 <-{Re(integrand.werte*x^(-gamma-1i*arg) )}
ans<-gamma(d/2)/(pi*n)* integrate.xy(arg,integrand1,-h,h)
return(ans)
}
#schaetzerspec(x,h) returns the estimated density with cut-off
   parameter h in x.
```

Estimator for Bessel Processes Based on Self-Similarity

(With Adaptive Cut-Off Parameter, See Section 3.9)

```
sup.dist <- function(x1, x2) {max(abs(x1-x2))}
l1.dist<- function(f1,f2){sum(abs(f1-f2))/0.1}
n<-1000
d <- 5
gamma <- 0.7
h_n <- seq(0.1,4,0.1)
minloss2<- matrix(10,nrow=100,ncol=4)
loss2<-rep(10,(length(h_n)-1))
arg<- seq(-4,4,0.001) #arguments for the integrand
schaetervonxh<-matrix(0,nrow = 100,ncol = length(h_n))

#Simulate observations of T
BeobT <- rexp(n,1) #Here T exponentially distributed

#Simulate observations of BES_1 and BES_T
BeobBES1 <- sqrt(rchisq(n,d))
BeobX<- sqrt(BeobT)*BeobBES1

integrand <- function(v) {sum(BeobX^(2*(gamma+1i*v-1)))
    /gammaz(gamma+d/2-1+1i*v)/2^(gamma+1i*v) }
integrand<-Vectorize(integrand)
integrand.werte<- integrand(arg)
schaetzerspec <- function(x,h){
integrand1 <-{Re(integrand.werte*x^(-gamma-1i*arg) )}
ans<-gamma(d/2)/(pi*n)* integrate.xy(arg,integrand1,-h,h)
return(ans)
}
chippy<-Vectorize(schaetzerspec)
for (j in 1:length(h_n)) {
schaetervonxh[,j]<- chippy(0.1*(1:100),h_n[j])
}
for (j in 1:(length(h_n)-1)) {
loss2[j]<- l1.dist(schaetervonxh[,j+1],schaetervonxh[,j] )
}
#schaetervonxh[,which.min(loss2)] returns the estimated density
    evaluated in 0.1*(1:100)
```

Estimator for Bessel Processes Based on Hypergroup Theory

(with Adaptive Cut-Off Parameter, See Section 5.3)

```
Lambda <- function(x){
return((3*sin(x) - 3*x*cos(x))/x^3)
}
Lambda<-Vectorize(Lambda,vectorize.args = "x")
l1.dist<- function(f1,f2){integrate.xy(x,abs(f1-f2),0.1,10)}

alpha <- 3/2
gamma <- 0.7
n <- 1000
x<- 0.1*(1:100)
A_n <- n^(1/4)
U_n <- seq(0.1,4,0.1)
loss2vonU<-rep(10,length(U_n)-1)
schaetzervonxU<-matrix(0,nrow=100,ncol=length(U_n))
arg<- seq(-4,4,0.01)

for (k in 1:100) {
#Simulate observations of T
BeobT <- rgamma(n,shape=2, scale = 1) #Here T Gamma distributed
BeobBES1 <- sqrt(rchisq(n,d))
BeobX<- sqrt(BeobT)*BeobBES1

m_n<- sum(BeobX^2)/n/(1+alpha)/2 #approximately-equal 2d

inneresintegral <- function(v){
integrand1re <-
   function(lambda){Re((lambda^2/2)^(-gamma-1i*v)*lambda*(1/n* sum(
   Lambda(lambda*BeobX) ) -exp(-m_n*(lambda^2/2)) ))}
integrand1re <- Vectorize( integrand1re,vectorize.args = "lambda")
integrand1im <-
   function(lambda){Im((lambda^2/2)^(-gamma-1i*v)*lambda*(1/n* sum(
   Lambda(lambda*BeobX) ) -exp(-m_n*(lambda^2/2)) ))}
integrand1im <- Vectorize( integrand1im,vectorize.args = "lambda")
ans<-(integrate(integrand1re,0,A_n)$value
   +1i*integrate(integrand1im,0,A_n)$value +
   gammaz(1-gamma-1i*v)*m_n^(gamma-1+1i*v))/gammaz(1-gamma-1i*v)
return(ans)
}

inneresintegral<-Vectorize(inneresintegral) #ingegrand from -4 to 4
outerint.werte<- inneresintegral(arg)
```

```
schaetzerallg <- function(s,U){
integrandre<-Re(outerint.werte*s^(-gamma-1i*arg))
ans<-1/(2*pi)* (integrate.xy(arg,integrandre,-U,U))
return(ans)
}

#schaetzerallg(s,U) returns the estimated density in s for a given
   cut-off parameter U
chippy<-Vectorize(schaetzerallg,vectorize.args = "s")

for (j in 1:length(U_n)) {
schaetzervonxU[,j]<- chippy(x,U_n[j])
}
for (l in 1:(length(U_n)-1)) {
loss2vonU[l]<- l1.dist(schaetzervonxU[,l+1],schaetzervonxU[,l])
}

#schaetzervonxU[,which.min(loss2vonU)] returns the estimated density
   evaluated in 0.1*(1:100)
```

List of Symbols

$\cdot^-$	involution in a hypergroup
$\bar{\cdot}$	complex conjugate
$\stackrel{d}{=}$	equality in law for random variables; equality of all finite-dimensional distributions for processes
$\lesssim,\ \gtrsim,\ \sim$	asymptotically less/greater/equivalent (see Section 2.3)
$\mathbb{1}_A$	characteristic function of the set A, i.e. $\mathbb{1}_A(x) := \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$
$\mathcal{B}$	Borel σ-algebra on $\mathbb{R}$
$\mathcal{B}(K)$	Borel σ-algebra on the locally compact space K
$C_b(K)$	space of continuous bounded functions $f : K \to \mathbb{C}$
$C_c(K)$	space of continuous functions $f : K \to \mathbb{C}$ with compact support
$C^k(K)$	space of k-times continuously differentiable functions $f : K \to \mathbb{C}$ ($k \in \mathbb{N}$)
δ_x	point measure in x
e	neutral element of a hypergroup
$\mathcal{F}[f](y)$	$:= \int_{-\infty}^{\infty} f(u)e^{iuy}du$ classical Fourier transform of a function f
$\mathcal{F}[X](y)$	$:= \mathrm{E}[e^{iXy}]$ classical Fourier transform of a random variable X
$\hat{\mu}$	hypergroup Fourier transform of a measure μ
$\mathcal{F}_r[X]$	hypergroup Fourier transform of a random variable X
K	locally compact Hausdorff space
$\widehat{K}$	dual space of a commutative hypergroup K, see Definition 4.1.1
i.i.d.	independently identically distributed
$\mathcal{L}[X](t)$	$:= \mathrm{E}[e^{-tX}]$ the Laplace transform of a random variable X.
$\mathcal{L}^p(K,\omega)$	space of functions $f : K \to \mathbb{C}$ with finite p-norm with respect to ω
$L^2_{\mathbb{C}}$	$= \{X \mid X$ is a complex random variable with $\mathrm{E}[\lvert X\rvert^2] < \infty\}$ (see Section 2.1)
$M^1(K)$	space of probability measures on K

$M_b(K)$	space of bounded regular Borel measures on K
$\mathfrak{M}_{(a,b)}$	space of functions with well defined Mellin transform (cf. Definition 2.2.1)
$\mathcal{M}[f](s)$	$:= \int_0^\infty f(x)x^{s-1}dx$ Mellin transform of a function f
$\mathcal{M}[X](s)$	$:= \mathrm{E}[X^{s-1}]$ Mellin transform of a random variable X
$\mathbb{N}$	set of natural numbers
$\mathbb{N}_0$	$:= \mathbb{N} \cup \{0\}$
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$
ω	Haar measure
π	Plancherel measure
$\mathbb{R}_+$	$= [0, \infty)$, set of nonnegative real numbers
$N_B^{(s)}(\widehat{K}),\ N_{T_2}^{(w)}(\widehat{K})$	classes of negative definite functions on K, see Definitions 4.2.2 and 4.2.3
$\mathrm{supp}(\cdot)$	support of a measure

Bibliography

[1] M. Abramowitz and I. Stegun. *Handbook of mathematical functions with formulas, graphs and mathematical tables*. Dover Publications, 1972.

[2] H. Andersen, M. Hojbjerre, D. Sorensen, and P. Eriksen. *Linear and Graphical Models for the Multivariate Complex Normal Distribution*. de Gruyter, 2002.

[3] G. Andrews, R. Askey, and R. Roy. *Special Functions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1999.

[4] F. Bauer and M. Reiß. Regularization independent of the noise level: an analysis of quasi-optimality. *Inverse Problems*, 24(5):055009, 2008.

[5] H. Bauer. *Wahrscheinlichkeitstheorie*. de Gruyter, 2002.

[6] D. Belomestny and J. Schoenmakers. Statistical Skorohod embedding problem: Optimality and asymptotic normality. *Statistics and Probability Letters*, 104:169 – 180, 2015.

[7] D. Belomestny and J. Schoenmakers. Statistical inference for time-changed Lévy processes via Mellin transform approach. *Stochastic Processes and their Applications*, 126(7):2092 – 2122, 2016.

[8] F. Biagini, Y. Hu, B. Oksendal, and T. Zhang. *Stochastic Calculus for Fractional Brownian Motion and Applications*. Probability and Its Applications. Springer, 2008.

[9] N. Bleistein and R. Handelsman. *Asymptotic Expansions of Integrals*. Dover Publications, 1986.

[10] W. Bloom and H. Heyer. *Harmonic Analysis of Probability Measures on Hypergroups*. de Gruyter, 1995.

[11] W. Bloom and H. Heyer. Negative definite functions and convolution semigroups of probability measures on an commutative hypergroup. *Probability and Mathematical Statistics*, 16:157 – 176, 1996.

[12] P. Butzer and S. Jansche. A direct approach to the Mellin transform. *Journal of Fourier Analysis and Applications*, 3(4):325–376, 1997.

[13] O. Chybiryakov, N. Demni, L. Gallardo, M. Rösler, M. Voit, and M. Yor. *Harmonic and stochastic analysis of Dunkl processes*. Hermann Mathématiques, 2008.

[14] J. Elstrodt. *Maß- und Integrationstheorie*. Springer, 2007.

[15] P. Embrechts and M. Maejima. An introduction to the theory of self-similar stochastic processes. *International Journal of Modern Physics B*, 14(12n13):1399–1420, 2000.

[16] H. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications. Springer, 1996.

[17] A. Erdélyi. *Asymptotic Expansions*. Dover Books on Mathematics. Dover Publications, 1956.

[18] U. Finckh. *Beiträge zur Wahrscheinlichkeitstheorie auf einer Kingman-Struktur*. Eberhard-Karls-Universität Tübingen, 1986. unv. Diss.

[19] P. Flajolet, X. Gourdon, P. Dumas, D. T. D. Knuth, N. G. D. Bruijn, and H. Mellin. Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, 144:3–58, 1995.

[20] P. Gänssler and W. Stute. *Wahrscheinlichkeitstheorie*. Hochschultext. Springer, 1977.

[21] J. Hadamard. *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. Dover Books on Science. Yale University Press, 1923.

[22] R. Jewett. Spaces with an abstract convolution of measures. *Advances in Mathematics*, 18(1):1 – 101, 1975.

[23] J. F. C. Kingman. Random walks with spherical symmetry. *Acta Mathematica*, 109(1):11–53, 1963.

[24] J. Lamperti. Semi-stable stochastic processes. *Transactions of the American Mathematical Society*, 104:62–78, 1962.

[25] R. D. Lord. The use of the hankel transform in statistics. *Biometrika*, 41(1/2):44–55, 1954.

[26] E. Michael. Topologies on spaces of subsets. *Transactions of the American Mathematical Society*, 71:152–182, 1951.

[27] F. Oberhettinger. *Tables of Mellin Transforms*. Springer, 2012.

[28] J. Obloj. The Skorokhod embedding problem and its offspring. *Probability Surveys*, 1, 2004.

[29] V. Pipiras and M. S. Taqqu. *Long-Range Dependence and Self-Similarity*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2017.

[30] R. Renzsch and M. Voit. Lévy processes on commutative hypergroups. *Contemp. Math.*, 261:83 – 105, 2000.

[31] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer, 1999.

[32] K. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1999.

[33] Y. Sinai. Self-similar probability distributions. *Theory of Probability and Its Applications*, 21(1):64–80, 1976.

[34] A. Skorokhod. *Issledovaniya po teorii sluchainykh protsessov (Stokhasticheskie differentsialnye uravneniya i predelnye teoremy dlya protsessov Markova)*. 1961.

[35] A. Skorokhod. *Studies in the Theory of Random Processes*. Addison-Wesley Publishing Company, 1965.

[36] M. Taqqu. Weak convergence to fractional Brownian motion and the Rosenblatt process. *Probability Theory and Related Fields*, 31:287–302, 1975.

[37] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2008.

[38] M. Voit. Positive and negative definite functions on the dual space of a commutative hypergroup. *Analysis*, 9(4):371–388, 1989.

[39] G. Watson. *A Treatise on the Theory of Bessel Functions*. Cambridge Mathematical Library. Cambridge University Press, 1995.

[40] A. Zayed. *Handbook of Function and Generalized Function Transformations*. Applied mathematics. Taylor and Francis, 1996.

[41] H. Zeuner. One-dimensional hypergroups. *Advances in Mathematics*, 76(1):1 – 18, 1989.