Bernoulli **17**(1), 2011, 395–423
DOI: 10.3150/10-BEJ274

Statistical analysis of self-similar conservative fragmentation chains

MARC HOFFMANN[1] and NATHALIE KRELL[2]

[1]*Université Paris Est, Laboratoire d'Analyse et de Mathématiques Appliquées, CNRS-UMR 8050, 5, Boulevard Descartes, 77454 Marne-la-Vallée Cedex 2, France. E-mail: marc.hoffmann@univ-mlv.fr*
[2]*Université de Rennes 1, Institut de Recherche mathématique de Rennes, CNRS-UMR 6625, Campus de Beaulieu, Bâtiment 22, 35042 Rennes Cedex, France. E-mail: nathalie.krell@univ-rennes1.fr*

We explore statistical inference in self-similar conservative fragmentation chains when only approximate observations of the sizes of the fragments below a given threshold are available. This framework, introduced by Bertoin and Martinez [*Adv. Appl. Probab.* **37** (2005) 553–570], is motivated by mineral crushing in the mining industry. The underlying object that can be identified from the data is the step distribution of the random walk associated with a randomly tagged fragment that evolves along the genealogical tree representation of the fragmentation process. We compute upper and lower rates of estimation in a parametric framework and show that in the nonparametric case, the difficulty of the estimation is comparable to ill-posed linear inverse problems of order 1 in signal denoising.

Keywords: fragmentation chains; key renewal theorem; nonparametric estimation; parametric

1. Introduction

1.1. Motivation

Random fragmentation models, commonly used in a variety of physical models, have their theoretical roots in the works of Kolmogorov [11] and Filippov [8] (see also [1,5,12,13] and the references therein). Informally, we imagine an object that falls apart randomly as time passes. The resulting particles break independently of each other in a self-similar way. A thorough account on random fragmentation processes and chains is given in the book by Bertoin [5], a key reference for this paper.

In this work, we adopt the perspective of statistical inference. We focus on the quite specific class of *self-similar fragmentation chains*. The law of a self-similar fragmentation chain is determined by two components:

- the dislocation measure, which governs the way that the fragments split;
- the index of self-similarity, which determines the rate of splitting;

see the definition in Section 2.1. In this paper, we postulate a specific observation scheme, motivated by the mining industry, where the goal is to separate metal from non-valued components in large mineral blocks by a series of blasting, crushing and grinding operations. In this setting, one observes, approximately, the fragments arising from an initial block of size m only when they reach a size smaller than some screening threshold, say $\eta > 0$; see [6] and the references therein. Asymptotics are taken as the ratio $\varepsilon := \eta/m$ vanishes.

1.2. Organization and results of the paper

In Section 2, we recall the basic tools for the construction of conservative fragmentation chains, closely following the book by Bertoin [5]. For statistical purposes, our main tool is the empirical measure $\mathcal{E}_\varepsilon$ of the size of fragments when they reach a size smaller than a threshold ε in the limit $\varepsilon \to 0$. We highlight the fact that $\mathcal{E}_\varepsilon$ captures information about the dislocation measure through the Lévy measure π of a randomly tagged fragment associated with the fragmentation process.

In Section 3, we give a rate of convergence for the empirical measure $\mathcal{E}_\varepsilon$ toward its limit in Theorem 1, extending former results (under more stringent assumptions) of Bertoin and Martinez [6]. The rate is of the form $\varepsilon^{1/2-\ell(\pi)}$, where $\ell(\pi) > 0$ can be made arbitrarily small under suitable exponential moment conditions for π. We additionally consider the more realistic framework of observations with limited accuracy, where each fragment is actually known up to a systematic stochastic error of order $\sigma \ll \varepsilon$. We construct estimators related to functionals of π in the absolutely continuous case. In the parametric case (Theorem 3), we establish that the best achievable rate is $\varepsilon^{1/2}$, in the particular case of binary fragmentations, where a particle splits into two blocks at each step exactly. We construct a convergent estimator in a general setting (Theorem 2) with an error of order $\varepsilon^{1/2-\ell'(\pi)}$ for another $\ell'(\pi) > 0$ that can be made arbitrarily small under appropriate assumptions on the density of π near 0 and $+\infty$. In the nonparametric case, we construct an estimator that achieves (Theorem 4) a rate of the form $(\varepsilon^{1-\ell''(\pi)})^{s/(2s+3)}$, where $s > 0$ is the local smoothness of the density of π, up to appropriate rescaling. Except for the factor $\ell''(\pi) > 0$, we obtain the same rate as for ill-posed inverse problems of degree 1.

2. Statistical model

2.1. Fragmentation chains

A fragmentation chain can be constructed as follows. We start with a state space

$$\mathcal{S}^{\downarrow} := \left\{ \mathbf{s} = (s_1, s_2, \ldots), s_1 \geq s_2 \geq \cdots \geq 0, \sum_{i=1}^{\infty} s_i \leq 1 \right\}.$$

A point $\mathbf{s} \in \mathcal{S}^{\downarrow}$ is interpreted as the collection of (decreasing) sizes of fragments originating from a single (unit) mass. We also specify the following two quantities:

- a finite dislocation measure ν, that is, a finite measure $\nu(d\mathbf{s})$ on $\mathcal{S}^{\downarrow}$;
- a parameter of self-similarity, $\alpha \geq 0$.

A fragmentation chain with parameter of self-similarity α and dislocation measure ν is a Markov process $X = (X(t), t \geq 0)$ with value in $\mathcal{S}^{\downarrow}$. Its evolution can be described as follows: a fragment with size x lives for an exponential time with parameter $x^\alpha \nu(\mathcal{S}^{\downarrow})$ and then splits and gives rise to a family of smaller fragments distributed as $x\xi$, where ξ is distributed according to $\nu(\cdot)/\nu(\mathcal{S}^{\downarrow})$. We denote by $\mathbb{P}_m$ the law of X started from the initial configuration $(m, 0, \ldots)$ with $m \in (0, 1]$. Under $\mathbb{P}_m$, the law of X is entirely determined by α and $\nu(\cdot)$; see Theorem 3 of Bertoin [4]. To ensure that everything is well defined, the following assumptions on the dislocation measure $\nu(d\mathbf{s})$ of X are in force throughout the paper.

Assumption A. *We have* $\nu(\mathcal{S}^{\downarrow}) = 1$ *and* $\nu(s_1 \in (0,1)) = 1$.

In our setting, Assumption A is standard; see Bertoin [5]. We will repeatedly use the representation of fragmentation chains as random infinite marked trees. Let

$$\mathcal{U} := \bigcup_{n=0}^{\infty} \mathbb{N}^n$$

denote the infinite genealogical tree (with $\mathbb{N}^0 := \{\varnothing\}$) associated with X as follows: to each node $u \in \mathcal{U}$, we set a mark

$$(\xi_u, a_u, \zeta_u),$$

where ξ_u is the size of the fragment labeled by u, a_u is its birth-time and ζ_u is its life-time. We have the following identity between point measures on $(0, +\infty)$:

$$\sum_{i=1}^{\infty} 1_{\{X_i(t)>0\}} \delta_{X_i(t)} = \sum_{u \in \mathcal{U}} 1_{\{t \in [a_u, a_u+\zeta_u)\}} \delta_{\xi_u}, \qquad t \geq 0,$$

with $X(t) = (X_1(t), X_2(t), \ldots)$ and where δ_x denotes the Dirac mass at x. Finally, X has the following branching property: for every fragment $\mathbf{s} = (s_1, \ldots) \in \mathcal{S}^{\downarrow}$ and every $t \geq 0$, the distribution of $X(t)$ given $X(0) = \mathbf{s}$ is the same as the decreasing rearrangement of the terms of independent random sequences $X^{(1)}(t), X^{(2)}(t), \ldots$, where, for each i, $X^{(i)}(t)$ is distributed as $X(t)$ under $\mathbb{P}_{s_i}$.

2.2. Observation scheme

Keeping in mind the motivation of mineral crushing, we consider the fragmentation under $\mathbb{P} := \mathbb{P}_1$, initiated with a unique block of size $m = 1$, and we observe the process stopped at the time when all the fragments become smaller than some given threshold $\varepsilon > 0$, so we have data ξ_u, for every $u \in \mathcal{U}_\varepsilon$, with

$$\mathcal{U}_\varepsilon := \{u \in \mathcal{U}, \xi_{u-} \geq \varepsilon, \xi_u < \varepsilon\},$$

where we denote by $u-$ the parent of the fragment labeled by u. We will further assume that the total mass of the fragments remains constant through time, as follows.

Assumption B (Conservative property). *We have* $\nu(\sum_{i=1}^{\infty} s_i = 1) = 1$.

We next consider a test function $g(\cdot)$ integrated against the empirical measure

$$\mathcal{E}_\varepsilon(g) := \sum_{u \in \mathcal{U}_\varepsilon} \xi_u g(\xi_u/\varepsilon).$$

Indeed, under Assumption B, we have

$$\sum_{u \in \mathcal{U}_\varepsilon} \xi_u = 1 \qquad \mathbb{P}\text{-almost surely,} \tag{1}$$

so $\mathcal{E}_\varepsilon(g)$ appears as a weighted empirical version of $g(\cdot)$. Note that the empirical measure $\mathcal{E}_\varepsilon$ depends only on the size of the fragmentation and is thus independent of the self-similarity parameter α. Bertoin and Martinez show in [6], Corollary 1, that under mild assumptions on $\nu(\cdot)$, the random variable $\mathcal{E}_\varepsilon(g)$ converges to

$$\mathcal{E}(g) := \frac{1}{c(\nu)} \int_0^1 \frac{g(a)}{a} \int_{\mathcal{S}^\downarrow} \sum_{i=1}^{\infty} s_i 1_{\{s_i < a\}} \nu(\mathrm{d}\mathbf{s})\,\mathrm{d}a$$

in $L^1(\mathbb{P})$ as $\varepsilon \to 0$, with $c(\nu) = -\int_{\mathcal{S}^\downarrow} \sum_{i=1}^\infty s_i \log s_i \, \nu(\mathrm{d}\mathbf{s})$, tacitly assumed to be well defined. This suggests a strategy for recovering information about $\nu(\cdot)$ by choosing suitable test functions $g(\cdot)$. In Section 3.1, we will show that the convergence also holds in $L^2(\mathbb{P})$ and we will exhibit a rate of convergence, which is a crucial issue if statistical results are sought.

2.3. First estimates

From now on, we assume that we have data

$$X_\varepsilon := (\xi_u, u \in \mathcal{U}_\varepsilon) \tag{2}$$

and we specialize in the estimation of $\nu(\cdot)$. Clearly, the data give no information about the parameter of self-similarity α that we consider as a nuisance parameter. Assumptions A and B are in force. At this stage, we can relate $\mathcal{E}(g)$ to a more appropriate quantity by means of the so-called *tagged fragment* approach.

The randomly tagged fragment. Let us first consider the homogenous case $\alpha = 0$. Assume that we can "tag" a point at random according to a uniform distribution on the initial fragment and imagine that we can follow the evolution of the fragment that contains this point. Let us denote by $(\chi(t), t \geq 0)$ the process of the size of the fragment that contains the randomly chosen point. This fragment is a typical observation in our data set X_ε and it appears at time

$$T_\varepsilon := \inf\{t \geq 0, \chi(t) < \varepsilon\}.$$

Bertoin [5] shows that the process $\zeta(t) := -\log \chi(t)$ is a subordinator with Lévy measure

$$\pi(\mathrm{d}x) := \mathrm{e}^{-x} \sum_{i=1}^{\infty} \nu(-\log s_i \in \mathrm{d}x). \tag{3}$$

We can anticipate that the information we get from X_ε is actually information about the Lévy measure $\pi(\mathrm{d}x)$ of $\zeta(t)$ obtained via $\zeta(T_\varepsilon)$. The dislocation measure $\nu(\mathrm{d}\mathbf{s})$ and $\pi(\mathrm{d}x)$ are related by (3), which reads

$$\int_{\mathcal{S}^{\downarrow}} \sum_{i=1}^{\infty} s_i f(s_i) \nu(\mathrm{d}\mathbf{s}) = \int_{(0,+\infty)} f(\mathrm{e}^{-x}) \pi(\mathrm{d}x) \tag{4}$$

for any suitable $f(\cdot) : [0,1] \to [0,+\infty)$. In particular, by Assumption B and the fact that $\nu(\mathcal{S}^{\downarrow}) = 1$, $\pi(\mathrm{d}x)$ is a probability measure, hence $\zeta(t)$ is a compound Poisson process. Informally, a typical observation takes the form $\zeta(T_\varepsilon)$, which is the value of a subordinator with Lévy measure $\pi(\mathrm{d}x)$ at its first passage time strictly above $-\log \varepsilon$. The case $\alpha \neq 0$ is a bit more involved and reduces to the homogenous case by a time change; see Bertoin [4,5]. In terms of the limit of the empirical measure $\mathcal{E}_\varepsilon(g)$, we equivalently have

$$\mathcal{E}(g) = \frac{1}{c(\pi)} \int_0^1 \frac{g(a)}{a} \pi(-\log a, +\infty)\,\mathrm{d}a = \frac{1}{c(\pi)} \int_0^{+\infty} g(\mathrm{e}^{-x}) \pi(x, +\infty)\,\mathrm{d}x$$

with $c(\pi) = \int_{(0,+\infty)} x \pi(\mathrm{d}x)$. The representation of $\mathcal{E}(g)$ as an integral with respect to π will prove technically convenient. Except in the binary case (a particular case of interest, see Section 4.1), knowledge of $\pi(\cdot)$ does not, in general, allow us to recover $\nu(\cdot)$.

Measurements with limited accuracy. It is unrealistic to assume that we can observe exactly the sizes ξ_u of the fragments. This becomes even more striking if the dislocation splits at a given time into infinitely many fragments of non-zero size, a situation that we do not discard in principle. Therefore, we replace (2) by the more realistic observation scheme $X_{\varepsilon,\sigma} := (\xi_u^{(\sigma)}, u \in \mathcal{U}_{\varepsilon,\sigma})$ with

$$\mathcal{U}_{\varepsilon,\sigma} := \bigl\{ u \in \mathcal{U}, \xi_{u-}^{(\sigma)} \geq \varepsilon, \xi_u^{(\sigma)} < \varepsilon \bigr\}$$

and

$$\xi_u^{(\sigma)} := \xi_u + \sigma U_u. \tag{5}$$

The random variables $(U_u, u \in \mathcal{U})$ are identically distributed and account for a systematic experimental microstructure noise in the measurement of X_ε, independent of X_ε. We assume, furthermore, that for every $u \in \mathcal{U}$,

$$|U_u| \leq 1 \quad \text{and} \quad \mathbb{E}[U_u] = 0.$$

The noise level $0 \leq \sigma = \sigma(\varepsilon) \ll \varepsilon$ is assumed to be known and represents the accuracy level of the statistician. The observations $\xi_u + \sigma U_u$ are further discarded below a threshold $\sigma \leq t_\varepsilon \leq \varepsilon$, beyond which they become irrelevant, leading to the modified empirical measure

$$\mathcal{E}_{\varepsilon,\sigma}(g) := \sum_{u \in \mathcal{U}_{\varepsilon,\sigma}} 1_{\{\xi_u^{(\sigma)} \geq t_\varepsilon\}} \xi_u^{(\sigma)} g\bigl(\xi_u^{(\sigma)}/\varepsilon\bigr).$$

In the sequel, we take $t_\varepsilon = \gamma_0 \varepsilon$ for some (arbitrary) $0 < \gamma_0 < 1$ and assume further that $\sigma \leq \frac{1}{2} t_\varepsilon$.

3. Main results

3.1. A rate of convergence for the empirical measure

Definition 1. *For $\kappa > 0$, we say that a non-lattice probability measure $\pi(\mathrm{d}x)$ defined on $[0,+\infty)$ belong to $\Pi(\kappa)$ if $\int_{[0,+\infty)} \mathrm{e}^{\kappa x}\pi(\mathrm{d}x) < +\infty$. We set $\Pi(\infty) := \bigcap_{\kappa>0} \Pi(\kappa)$.*

For $m > 0$, let

$$\mathcal{C}(m) := \Big\{ g : [0,1] \to \mathbb{R}, \text{ continuous, } \|g\|_\infty := \sup_x |g(x)| \le m \Big\}$$

and

$$\mathcal{C}'(m) := \Big\{ g \in \mathcal{C}(m) : [0,1] \to \mathbb{R}, \text{ differentiable, } \|g'\|_\infty := \sup_x |g'(x)| \le m \Big\}.$$

Our first result exhibits explicit rates in the convergence $\mathcal{E}_\varepsilon(g) \to \mathcal{E}(g)$ as $\varepsilon \to 0$, extending Bertoin [5], Proposition 1.12.

Theorem 1. *We work under Assumptions A and B. Let $1 < \kappa \le \infty$ and assume that $\pi \in \Pi(\kappa)$.*

- *For every $m > 0$ and $1 \le \mu < \kappa$, we have*

$$\sup_{g \in \mathcal{C}(m)} \mathbb{E}\Big[\big(\mathcal{E}_\varepsilon(g) - \mathcal{E}(g)\big)^2\Big] = \mathrm{o}\big(\varepsilon^{\mu/(\mu+1)}\big). \tag{6}$$

- *The convergence (6) remains valid if we replace $\mathcal{E}_\varepsilon(\cdot)$ by $\mathcal{E}_{\varepsilon,\sigma}(\cdot)$ and $\mathcal{C}(m)$ by $\mathcal{C}'(m)$. The following additional error term must then be incorporated: for any $0 < \mu < \kappa$, we have*

$$\sup_{g \in \mathcal{C}'(m)} \mathbb{E}\Big[\big(\mathcal{E}_{\varepsilon,\sigma}(g) - \mathcal{E}_\varepsilon(g)\big)^2\Big] = \mathrm{o}(\varepsilon^{\mu/2}) + \mathcal{O}(\sigma\varepsilon^{-1}). \tag{7}$$

3.2. Statistical estimation

We study the estimation of $\pi(\cdot)$ by constructing estimators based on $\mathcal{E}_\varepsilon(\cdot)$ or, rather, $\mathcal{E}_{\varepsilon,\sigma}(\cdot)$. We need the following regularity assumption.

Assumption C. *The probability $\pi(\mathrm{d}x)$ is absolutely continuous with respect to the Lebesgue measure: $\pi(\mathrm{d}x) = \pi(x)\,\mathrm{d}x$. Moreover, its density function $x \rightsquigarrow \pi(x)$ is continuous on $(0,+\infty)$ and satisfies $\limsup_{x\to+\infty} \mathrm{e}^{\vartheta x}\pi(x) < +\infty$ for some $\vartheta \ge 1$.*

We distinguish two cases: the *parametric case*, where we estimate a linear functional of $\pi(\cdot)$ of the form

$$m_k(\pi) := \int_0^{+\infty} x^k \pi(x)\,\mathrm{d}x, \qquad k = 1, 2, \ldots,$$

and the *nonparametric case*, where we estimate the function $x \rightsquigarrow \pi(x)$ pointwise. In the latter case, it will prove convenient to assess the local smoothness properties of $\pi(\cdot)$ on a logarithmic scale. Henceforth, we consider the mapping

$$a \rightsquigarrow \beta(a) := a^{-1}\pi(-\log a), \qquad a \in (0,1). \tag{8}$$

In the nonparametric case, we estimate $\beta(a)$ for every $a \in (0,1)$.

3.3. The parametric case

Preliminaries. For $k \geq 1$, we estimate

$$m_k(\pi) := \int_0^{+\infty} x^k \pi(x)\,\mathrm{d}x = \int_0^1 \log(1/a)^k \beta(a)\,\mathrm{d}a$$

by the correspondence (8), implicitly assumed to be well defined. We first focus on the case $k=1$. Choose a sufficiently smooth test function $f(\cdot) : [0,1] \to \mathbb{R}$ such that $f(1)=0$ and let $g(a) := -af'(a)$. Clearly,

$$\begin{aligned} \mathcal{E}(g) &= \frac{1}{c(\pi)} \int_0^1 \frac{g(a)}{a} \pi(-\log a, +\infty)\,\mathrm{d}a \\ &= -\frac{1}{m_1(\pi)} \int_0^1 f'(a) \int_0^a \beta(u)\,\mathrm{d}u\,\mathrm{d}a = \frac{1}{m_1(\pi)} \int_0^1 f(a)\beta(a)\,\mathrm{d}a. \end{aligned} \tag{9}$$

Formally, taking $f(\cdot) \equiv 1$ would identify $1/m_1(\pi)$ since $\beta(\cdot)$ integrates to one, but this choice is forbidden by the boundary condition $f(1)=0$. We shall instead consider a family of regular functions that are close to the constant function 1 while satisfying $f(1)=0$.

Construction of the approximating functions. Let $f_\gamma : [0,1] \to \mathbb{R}$ with $0 < \gamma < 1$ be a family of smooth functions satisfying the following conditions:

- we have $f_\gamma(a) = 1$ for $a \leq 1-\gamma$ and $f_\gamma(1) = 0$;
- we have

$$\sup_{\gamma > 0} (\|f_\gamma\|_\infty + \gamma \|f'_\gamma\|_\infty + \gamma^2 \|f''_\gamma\|_\infty) < +\infty; \tag{10}$$

- for every $k \geq 1$ and some $\delta > 0$, we have

$$\begin{aligned} &\sup_{\gamma>0} \sup_{a\in(0,1)} \left\{ \gamma^2 |\log a|^k \big(a^{-1}|f_\gamma(1-a)| + |f'_\gamma(1-a)|\big) + \left(\frac{\gamma}{a}\right)^{1+\delta} f_\gamma(1-a) \right\} \\ &\quad < +\infty. \end{aligned} \tag{11}$$

The family $(f_\gamma, \gamma > 0)$ mimics the behaviour of the target function $f_0(a) = 1$ for $0 \leq a < 1$ and $f_0(1) = 0$ as $\gamma \to 0$. Condition (11) is technical (and probably not optimal). An explicit choice of a family $(f_\gamma, \gamma > 0)$ satisfying (10) and (11) is given by

$$f_\gamma(a) := \begin{cases} 1 & \text{if } a \leq 1-\gamma, \\ 10\left(\dfrac{1-a}{\gamma}\right)^3 - 15\left(\dfrac{1-a}{\gamma}\right)^4 + 6\left(\dfrac{1-a}{\gamma}\right)^5 & \text{if } 1-\gamma \leq a < 1, \\ 0 & \text{if } a = 1, \end{cases}$$

but other choices are obviously possible.

Construction of an estimator. We are now ready to give an estimator of the first moment $m_1(\pi)$ of π and, more generally, of any moment $m_k(\pi)$, $k \geq 1$. For a parametrization $\gamma := \gamma_\varepsilon \to 0$ to be specified later, we set

$$g_{\gamma_\varepsilon}(a) := -af'_{\gamma_\varepsilon}(a), \qquad a \in (0,1).$$

By Theorem 1, we expect $\mathcal{E}_{\sigma,\varepsilon}(g_{\gamma_\varepsilon})$ to be close to $\mathcal{E}(g_{\gamma_\varepsilon})$ which, in turn, is equal to $m_1(\pi)^{-1}\int_0^1 f_{\gamma_\varepsilon}(a)\beta(a)\,\mathrm{d}a$, by (9). Since $f_{\gamma_\varepsilon} \approx 1$ and $\beta(\cdot)$ is a density function, by appropriate regularity assumptions on π, we may further expect this last quantity to be close to $1/m_1(\pi)$. We therefore set

$$\widehat{m}_{1,\varepsilon} := \frac{1}{\mathcal{E}_{\varepsilon,\sigma}(g_{\gamma_\varepsilon})} \tag{12}$$

for an estimator of $m_1(\pi)$. More generally, for $k > 1$, we define successive moment estimators as follows. Set $h_{\gamma_\varepsilon}(a) := f_{\gamma_\varepsilon}(1-a)\log(1/a)^k$ and $\widetilde{g}_{\gamma_\varepsilon}(a) := -ah'_{\gamma_\varepsilon}(a)$. The same heuristics as before lead to the estimator

$$\widehat{m}_{k,\varepsilon} := \frac{\mathcal{E}_{\varepsilon,\sigma}(\widetilde{g}_{\gamma_\varepsilon})}{\mathcal{E}_{\varepsilon,\sigma}(g_{\gamma_\varepsilon})}.$$

Upper rates of convergence. We can describe the performance of $\widehat{m}_{k,\varepsilon}$ under an additional decay condition on $\pi(\cdot)$ near the origin.

Definition 2. *For $\kappa > 0$, we say that the probability $\pi(\cdot)$ belong to the class $\mathcal{R}(\kappa)$ if $\limsup_{x\to 0} x^{-\kappa+1}\pi(x) < +\infty$. We set $\mathcal{R}(\infty) := \bigcap_{\kappa>0}\mathcal{R}(\kappa)$.*

We obtain the following upper bound, under more stringent regularity assumptions on π than in Theorem 1.

Theorem 2. *We work under Assumptions A, B and C.*

- *For the estimation of $m_1(\pi)$, assume $\kappa_1 \geq 4$ and $\kappa_2 > 1$.*
- *For the estimation of $m_k(\pi)$ with $k \geq 2$, assume $\kappa_1 \geq 4$ and $\kappa_1 > \kappa_2 > 1$.*

For any $1 \leq \mu < \kappa_1$, let $\widehat{m}_{k,\varepsilon}$ be specified by $\gamma_\varepsilon := \varepsilon^{\mu/(\mu+1)(2\kappa_2+1)}$. The family

$$\left(\varepsilon^{-\mu/(\mu+1)}\right)^{\kappa_2/(2\kappa_2+1)}\left(\widehat{m}_{k,\varepsilon} - m_k(\pi)\right)$$

is tight provided that

$$\pi \in \Pi(\kappa_1) \cap \mathcal{R}(\kappa_2)$$

and $\sigma\varepsilon^{-3}$ remains bounded.

Some remarks: The convergence of $\widehat{m}_{k,\varepsilon}$ to $m_k(\pi)$ is of course no surprise, by (6). However, the dependence on ε in the test function $g_{\gamma_\varepsilon}(\cdot)$ (in particular, $g_{\gamma_\varepsilon}(\cdot)$ is unbounded as $\varepsilon \to 0$) requires a slight improvement of Theorem 1. This can be done thanks to Assumption C; see Proposition 2 in Section 5.3. The requirement $\sigma\varepsilon^{-3} = \mathcal{O}(1)$ ensures that the additional term coming from the approximation of $\mathcal{E}_\varepsilon(\cdot)$ by $\mathcal{E}_{\sigma,\varepsilon}(\cdot)$ is negligible.

Lower rates of convergence. Our next result shows that the exponent

$$\frac{\mu}{\mu+1}\frac{\kappa_2}{2\kappa_2+1} \le \frac{1}{2}$$

in the rate of convergence of Theorem 2 is nearly optimal, to within an arbitrarily small polynomial order.

Definition 3. *Let $\pi_0(\cdot)$ satisfy the assumptions of Theorem 2. The rate $0 < v_\varepsilon \to 0$ is a lower rate of convergence for estimating $m_k(\pi_0)$ if there exists a family $\pi_\varepsilon(\cdot)$ satisfying the assumptions of Theorem 2 and a constant $c > 0$ such that*

$$\liminf_{\varepsilon\to 0}\inf_{F_\varepsilon}\max_{\pi\in\{\pi_0,\pi_\varepsilon\}} \mathbb{P}[v_\varepsilon^{-1}|F_\varepsilon - m_k(\pi)| \ge c] > 0, \tag{13}$$

where the infimum is taken (for every ε) over all estimators constructed from $X_{\varepsilon,\sigma}$ at level ε.

Definition 3 expresses a kind of local min–max information bound: given $\pi_0(\cdot)$, one can find $\pi_\varepsilon(\cdot)$ such that no estimator can discriminate between $\pi_0(\cdot)$ and $\pi_\varepsilon(\cdot)$ at a rate faster than v_ε. We further restrict our attention to binary fragmentations; see Section 4.1. In that case, the dislocation measure satisfies $\nu(s_1 + s_2 \ne 1) = 0$ and, because of the conservation Assumption B, can be represented as

$$\nu(\mathrm{d}\mathbf{s}) = \rho(\mathrm{d}s_1)\delta_{1-s_1}(\mathrm{d}s_2), \tag{14}$$

where $\rho(\cdot)$ is a probability measure on $[1/2, 1]$.

Assumption D (Binary case). *The probability measure $\rho(\cdot)$ associated with $\pi(\cdot)$ is absolutely continuous and its density function is bounded away from zero.*

Theorem 3. *Assume that the fragmentation is binary and work under Assumption D. In the same setting as in Theorem 2, the rate $\varepsilon^{1/2}$ is a lower rate of convergence for estimating $m_k(\pi)$.*

3.4. The nonparametric case

Preliminaries. Under local smoothness assumptions on the function $\beta(\cdot)$, we estimate $\beta(a)$ for every $a \in (0, 1)$. Given $s > 0$, we say that $\beta(\cdot)$ belongs to the Hölder class $\Sigma(s)$ if there exists a constant $c > 0$ such that

$$\left|\beta^{(n)}(y) - \beta^{(n)}(x)\right| \le c|y - x|^{\{s\}}$$

with $s = n + \{s\}$, where n is a non-negative integer and $\{s\} \in (0, 1]$. We also need to relate $\beta(\cdot)$ to the decay of its corresponding Lévy measure $\pi(\cdot)$. Again abusing notation, we identify $\Pi(\kappa)$ with the set of $\beta(\cdot)$ such that $\mathrm{e}^x \beta(\mathrm{e}^{-x})\,\mathrm{d}x \in \Pi(\kappa)$, thanks to the inverse of (8), and likewise for $\mathcal{R}(\kappa)$.

Construction of an estimator. We construct an estimator of $\beta(\cdot)$ in the same way as for the parametric case: for $a \in (0, 1)$ and a normalizing factor $0 < \gamma_\varepsilon \to 0$, set

$$\varphi_{\gamma_\varepsilon,a}(x) := \gamma_\varepsilon^{-1}\varphi\bigl((x-a)/\gamma_\varepsilon\bigr),$$

where φ is a smooth function with support in $(0, 1)$ that satisfies the following oscillating property: for some integer $N \geq 1$,

$$\int_0^1 \varphi(a)\,\mathrm{d}a = 1, \qquad \int_0^1 a^k \varphi(a)\,\mathrm{d}a = 0, \qquad k = 1, \ldots, N. \tag{15}$$

The function $\varphi_{\gamma_\varepsilon,a}$ thus plays the role of a kernel centred around a. Set

$$h_{a,\varepsilon}(x) = -x\varphi'_{\gamma_\varepsilon,a}(x), \qquad x \in (0, 1).$$

We have

$$\mathcal{E}(h_{a,\varepsilon}) = \frac{1}{m_1(\pi)} \int_0^1 \varphi_{\gamma_\varepsilon,a}(x)\beta(x)\,\mathrm{d}x,$$

by (9). By letting $h_\varepsilon \to 0$ with an appropriate rate as $\varepsilon \to 0$, we expect this term to be close to $\beta(a)/m_1(\pi)$. Eventually, we can eliminate the denominator by means of our preliminary estimator $\widehat{m}_{1,\varepsilon}$. Our nonparametric estimator of $\beta(a)$ thus takes the form

$$\widehat{\beta}_\varepsilon(a) := \widehat{m}_{1,\varepsilon}\mathcal{E}_{\varepsilon,\sigma}(h_{a,\varepsilon}), \qquad a \in (0, 1),$$

where $\widehat{m}_{1,\varepsilon}$ is the estimator of $m_1(\pi)$ defined in (12).

Upper rates of convergence. We have the following result.

Theorem 4. *We work under Assumptions* A, B *and* C. *Let* $\kappa_1 \geq 4$ *and* $\kappa_2 > 1$. *For any* $1 \leq \mu < \kappa_1$, *let* $\widehat{\beta}_\varepsilon(\cdot)$ *be specified by* $\gamma_\varepsilon := \varepsilon^{\mu/(\mu+1)(2s+3)}$. *For every* $a \in (0, 1)$, *the family*

$$\bigl(\varepsilon^{-\mu/(\mu+1)}\bigr)^{s/(2s+3)}\bigl(\widehat{\beta}_\varepsilon(a) - \beta(a)\bigr)$$

is tight, *provided that*

$$\beta \in \Sigma(s) \cap \Pi(\kappa_1) \cap \mathcal{R}(\kappa_2)$$

for $0 < s < \min\{N, 3\kappa_2\}$ *and* $\sigma\varepsilon^{-3}$ *remains bounded.*

A proof of the (near) optimality, in the sense of the lower bound Definition 3 and in the spirit of Theorem 3, is presumably a delicate problem that lies beyond the scope of the paper; see Section 4.3.

4. Discussion

4.1. Binary fragmentations

The case of binary fragmentations is the simplest, yet is an important model of random fragmentation, where a particle splits into two blocks at each step (see, e.g., [7,8]). By using representation (14), if we further assume that $\rho(\mathrm{d}s_1) = \rho(s_1)\,\mathrm{d}s_1$ is absolutely continuous, then so is $\pi(\mathrm{d}x) = \pi(x)\,\mathrm{d}x$ and we have

$$\pi(x) = \mathrm{e}^{-2x}\bigl(\rho(\mathrm{e}^{-x})1_{[0,\log 2]}(x) + \rho(1-\mathrm{e}^{-x})1_{(\log 2,+\infty)}(x)\bigr) \tag{16}$$

for $x \in [0, +\infty)$ and

$$\beta(a) = a\bigl(\rho(a)1_{[1/2,1]}(a) + \rho(1-a)1_{[0,1/2)}(a)\bigr), \qquad a \in [0,1]. \tag{17}$$

In particular, the regularity properties of $\beta(\cdot)$ are obtained from the local smoothness of $\rho(\cdot)$ and its behaviour near 1. For instance, if $\rho(1-a) = \mathcal{O}(a^{\kappa-1})$ near the origin, for some $\kappa > 0$, then

$$\pi \in \Pi(\kappa) \cap \mathcal{R}(\kappa).$$

4.2. Concerning Theorem 1

Theorem 1 readily extends to error measurements of the form $\mathbb{E}[|\mathcal{E}_\varepsilon(g) - \mathcal{E}(g)|^p]$ with $1 \le p \le 2$. The rate becomes $\varepsilon^{-\mu p/2(\mu+1)}$ in (6) and $\sigma^p\varepsilon^{-p}$ in (7) under the less stringent condition $\mu < \kappa/2p$.

Generally speaking, in (6), we obtain the (normalized) rate $\varepsilon^{\mu/2(\mu+1)}$ for any $\mu < \kappa$. Intuitively, we have a number of observations that should be of order ε^{-1}, so the expected rate would rather be $\varepsilon^{1/2}$. Why can we not obtain the rate $\varepsilon^{1/2}$, or simply $\varepsilon^{\kappa/2(\kappa+1)}$? The proof in Section 5.2 shows that we lose quite a lot of information when applying Sgibnev's result (see Proposition 1 in Section 5.1) on the key renewal theorem for a random walk with step distribution $\pi(\cdot)$ in the limit $\log(1/\varepsilon) \to +\infty$.

Proposition 1 ensures that if $\pi(\cdot)$ has exponential moments up to order κ, then we can guarantee in the renewal theorem the rate $\mathrm{o}(\varepsilon^\mu)$ for any $\mu < \kappa$ with some uniformity in the test function, a crucial point for the subsequent statistical applications. It is presumably possible to improve this rate to $\mathcal{O}(\varepsilon^\kappa)$ by using Ney's result [14]. However, a careful glance at the proof of Theorem 1 shows that we would then lose an extra logarithmic term when replacing $\varepsilon^{\mu/2(\mu+1)}$ by $\varepsilon^{\kappa/(2\kappa+1)}$. More generally, exhibiting exact rates of convergence in Theorem 1 remains a delicate issue: the key renewal theorem is sensitive to a modification of the distribution outside a neighbourhood of $+\infty$; see, for example, Asmussen [2], page 196.

4.3. Concerning Theorems 2 and 4

In the parametric case, we obtain the rate

$$\bigl(\varepsilon^{\mu/(\mu+1)}\bigr)^{\kappa_2/(2\kappa_2+1)} \qquad \text{for all } \mu < \kappa_1,$$

which can be made arbitrary close to the lower bound $\varepsilon^{1/2}$ by assuming κ_1 and κ_2 to be large enough. The factor $\mu/(\mu+1)$ comes from Theorem 1, whereas the factor $\kappa_2/(2\kappa_2+1)$ arises when using the technical assumption $\pi \in \mathcal{R}(\kappa_2)$. We do not know how to improve this.

In the nonparametric case, the situation is a bit different than in the parametric case: we now obtain the rate

$$\left(\varepsilon^{\mu/(\mu+1)}\right)^{s/(2s+3)} \qquad \text{for all } \mu < \kappa_1 \tag{18}$$

for the estimation of $\beta(a)$ for any $a \in (0,1)$. In the limit $\kappa_1 \to +\infty$, it becomes $\varepsilon^{s/(2s+3)}$, which can be related to more classical models in the nonparametric literature. Informally, a function of d variables with degree of smoothness s observed in noise under the action of a smoothing operator of degree ν (e.g., ν-fold integration) can be recovered with optimal rate $\varepsilon^{s/(2s+2\nu+d)}$; see, for instance, [16]. Here, we have $d=1$ and $\nu=1$ by the representation (9), so formula (18) is consistent with the general nonparametric theory. This advocates in favour of the (near) optimality of the result in the sense of Definition 3, but a complete proof lies beyond the scope of the paper.

4.4. The Crump–Mode–Jagers alternative

As suggested by a referee, the statistical problem can be reformulated alternatively in terms of the Crump–Mode–Jagers (CMJ) branching process. Consider a transformed point process $(\tau_1, \tau_2, \dots)$ defined by $\tau_i = -\log s_i$ for $\mathbf{s} = (s_1, s_2, \dots) \in \mathcal{S}^{\downarrow}$. The sequence $(\tau_1, \tau_2, \dots)$ describes the consecutive ages at childbearing for the individual assumed to be born at time zero. In our setting, the resulting CMJ process is supercritical with Malthusian parameter 1 since $\mathrm{e}^{-\tau_1} + \mathrm{e}^{-\tau_2} + \cdots = 1$.

Let $\sigma_u = -\log \xi_u$. We may now interpret σ_u as the individual forming the coming generation at time $t = -\log \varepsilon$. The empirical measure $\mathcal{E}_\varepsilon$ now has the representation

$$\begin{aligned}\mathcal{E}_\varepsilon(g) &= \sum_{u \in \mathcal{U}, \sigma_u - \tau_u \le t < \sigma_u} \mathrm{e}^{-\sigma_u} g\left(\mathrm{e}^{-\sigma_u + t}\right) \\ &= \mathrm{e}^{-t} \sum_{u \in \mathcal{U}, \sigma_u - \tau_u \le t < \sigma_u} \mathrm{e}^{-\sigma_u + t} g\left(\mathrm{e}^{-\sigma_u + t}\right)\end{aligned}$$

and the last sum can be expressed in terms of a population size with random characteristics; see [10]. This yields another interpretation of our statistical approach in terms of branching processes, presumably more useful in other settings.

5. Proofs

We will repeatedly use the convenient notation $a_\varepsilon \lesssim b_\varepsilon$ if $0 < a_\varepsilon \le c b_\varepsilon$ for some constant $c > 0$ which may depend on $\pi(\cdot)$ and on the constant m appearing in the definition of the class $\mathcal{C}(m)$ or $\mathcal{C}'(m)$. Any other dependence on other ancillary quantities will be obvious from the context. A function $g \in \mathcal{C}(m)$ is tacitly defined on the whole real line by setting $g(a) = 0$ for $a \notin [0,1]$.

5.1. Preliminaries: Rates of convergence in the key renewal theorem

We state a special case of Sgibnev's result [15] on uniform rates of convergence in the key renewal theorem, an essential tool for this paper. Let $F(\mathrm{d}x)$ be a non-lattice probability distribution with positive mean m and renewal function $\mathbb{F} = \sum_{n=0}^{\infty} F^{n\star}$ with $F^{0\star} := \delta_0$, $F^{1\star} := F$ and $F^{(n+1)\star} := F \star F^{n\star}$, $n \geq 0$. We denote by $T(F)$ the σ-finite measure with density function

$$\int_{(x,+\infty)} F(\mathrm{d}u)1_{[0,+\infty)}(x) - \int_{(-\infty,x]} F(\mathrm{d}u)1_{(-\infty,0)}(x)$$

and define $T^2(F) := T(T(F))$. Let $\varphi(\cdot) : \mathbb{R} \to [0,+\infty)$ be a submultiplicative function, that is, such that $\varphi(0) = 1$, $\varphi(x+y) \leq \varphi(x)\varphi(y)$. We then have (see, e.g., [9], Section 6)

$$\begin{aligned} -\infty < r_1 := \lim_{x\to-\infty} \frac{\log\varphi(x)}{x} \\ \leq \lim_{x\to+\infty} \frac{\log\varphi(x)}{x} =: r_2 < +\infty. \end{aligned}$$

Assumption E. *We have $r_1 \leq 0 \leq r_2$ and there exists $r : \mathbb{R} \to \mathbb{R}$, an integrable function such that the following conditions are fulfilled*:

$$\sup_x |r(x)|\varphi(x) < +\infty, \qquad \lim_{|x|\to\infty} r(x)\varphi(x) = 0,$$

$$\lim_{x\to+\infty} \varphi(x) \int_{[x,+\infty)} r(u)\,\mathrm{d}u = \lim_{x\to-\infty} \varphi(x) \int_{(-\infty,x]} r(u)\,\mathrm{d}u = 0$$

and $\int_{\mathbb{R}} \varphi(x)T^2(F)(\mathrm{d}x) < \infty$. We call $\varphi(\cdot)$ a rate function and $r(\cdot)$ a dominating function.

Sgibnev's result takes the following form.

Proposition 1 ([15], Theorem 5.1). *We work under Assumption E. Then*

$$\lim_{|t|\to\infty} \varphi(t) \sup_{\psi, |\psi(x)|\leq|r(x)|} \left| \psi \star \mathbb{F}(t) - m^{-1} \int_{\mathbb{R}} \psi(x)\,\mathrm{d}x \right| = 0.$$

5.2. Proof of Theorem 1

Step 1: *A preliminary decomposition*. We first use the fact that for $\eta > \varepsilon$, during the fragmentation process, the unobserved state X_η necessarily anticipates the state X_ε. The choice $\eta = \eta(\varepsilon)$ will follow later. This yields the following representation:

$$\mathcal{E}_\varepsilon(g) = \sum_{v\in\mathcal{U}_\eta} \xi_v \sum_{w\in\mathcal{U}} 1_{\{\xi_v\tilde{\xi}_{w-}^{(v)} \geq \varepsilon, \xi_v\tilde{\xi}_w^{(v)} < \varepsilon\}} \tilde{\xi}_w^{(v)} g\big(\xi_v\tilde{\xi}_w^{(v)}/\varepsilon\big),$$

where, for each label $v \in \mathcal{U}_\eta$ and conditional on X_η, a new independent fragmentation chain $(\tilde{\xi}_w^{(v)}, w \in \mathcal{U})$ is started, thanks to the branching property; see Section 2.1. Now, define

$$\lambda_\eta(v) := 1_{\{\xi_{v-} \geq \eta, \xi_v < \eta\}} \xi_v$$

and

$$Y_\varepsilon(v, g) := \sum_{w \in \mathcal{U}} 1_{\{\xi_v \tilde{\xi}_{w-}^{(v)} \geq \varepsilon, \xi_v \tilde{\xi}_w^{(v)} < \varepsilon\}} \tilde{\xi}_w^{(v)} g\big(\xi_v \tilde{\xi}_w^{(v)}/\varepsilon\big).$$

We obtain the decomposition of $\mathcal{E}_\varepsilon(g) - \mathcal{E}(g)$ as the sum of a centred and a bias term:

$$\mathcal{E}_\varepsilon(g) - \mathcal{E}(g) = M_{\varepsilon,\eta}(g) + B_{\varepsilon,\eta}$$

with

$$M_{\varepsilon,\eta}(g) := \sum_{v \in \mathcal{U}} \lambda_\eta(v)\big(Y_\varepsilon(v, g) - \mathbb{E}[Y_\varepsilon(v, g) | \lambda_\eta(v)]\big)$$

and

$$B_{\varepsilon,\eta}(g) := \sum_{v \in \mathcal{U}} \lambda_\eta(v)\big(\mathbb{E}[Y_\varepsilon(v, g) | \lambda_\eta(v)] - \mathcal{E}(g)\big),$$

where we have used the conservative property (1) in order to incorporate the limit term $\mathcal{E}(g)$ into the sum in v.

Step 2: *The term* $M_{\varepsilon,\eta}(g)$. Conditional on the σ-field generated by the random variables $(1_{\{\xi_{v-} \geq \eta\}} \xi_v, v- \in \mathcal{U})$, the variables $(Y_\varepsilon(v, g), v \in \mathcal{U})$ are independent. Therefore,

$$\mathbb{E}[M_{\varepsilon,\eta}(g)^2] \leq \sum_{v \in \mathcal{U}} \mathbb{E}[\lambda_\eta(v)^2 \mathbb{E}[Y_\varepsilon(v, g)^2 | \lambda_\eta(v)]]. \tag{19}$$

Thus, we first need to control the conditional variance of $Y_\varepsilon(v, g)^2$ given $\lambda_\eta(v) = u$, for $0 \leq u \leq \eta$, since $\mathbb{P}$-almost surely, $\lambda_\eta(v) \leq \eta$. Moreover, we have $Y_\varepsilon(v, g) = 0$ on the event $\{\lambda_\eta(v) < \varepsilon\}$, hence we may assume that $\varepsilon \leq u \leq \eta$.

To this end, we will use the following representation property.

Lemma 1. *Let* $f(\cdot) : [0, +\infty) \to [0, +\infty)$. *Then*

$$\mathbb{E}\bigg[\sum_{v \in \mathcal{U}_\eta} \xi_v f(\xi_v)\bigg] = \mathbb{E}^\star[f(\chi(T_\eta))], \tag{20}$$

where $\chi(t) = \exp(-\zeta(t))$ *and* $(\zeta(t), t \geq 0)$ *is a subordinator with Lévy measure* $\pi(\cdot)$ *defined on an appropriate probability space* $(\Omega^\star, \mathbb{P}^\star)$ *and*

$$T_\eta := \inf\{t \geq 0, \zeta(t) > -\log \eta\}.$$

The proof readily follows the construction of the randomly tagged fragment as elaborated in the book by Bertoin [3] and is thus omitted. We plan to bound the right-hand side of (19) using Lemma 1. For $0 < \varepsilon \le u \le \eta$, we have

$$\mathbb{E}[Y_\varepsilon(v,g)^2|\lambda_\eta(v)=u] = \mathbb{E}\biggl[\biggl(\sum_{w\in\mathcal{U}_{\varepsilon/u}} \widetilde{\xi}_w^{(v)} g\bigl(\varepsilon u^{-1}\widetilde{\xi}_w^{(v)}\bigr)\biggr)^2 \Big| \lambda_\eta(v)=u\biggr]$$
$$\le \mathbb{E}\biggl[\sum_{w\in\mathcal{U}_{\varepsilon/u}} \widetilde{\xi}_w^{(v)} g\bigl(\varepsilon u^{-1}\widetilde{\xi}_w^{(v)}\bigr)^2 \Big| \lambda_\eta(v)=u\biggr],$$

where we have used Jensen's inequality combined with (1). Applying Lemma 1, we derive

$$\mathbb{E}[Y_\varepsilon(v,g)^2|\lambda_\eta(v)=u] \le \mathbb{E}^\star\bigl[g\bigl(u\varepsilon^{-1}\mathrm{e}^{-\zeta(T_{\varepsilon/u})}\bigr)^2\bigr]. \tag{21}$$

Let $U(\cdot)$ denote the renewal function associated with the subordinator $(\zeta(t), t\ge 0)$. By [3], Proposition 2, Chapter III, the right-hand side of (21) is equal to

$$\int_{[0,-\log(\varepsilon/u))} \mathrm{d}U(s) \int_{(-\log(\varepsilon/u)-s,+\infty)} g(u\varepsilon^{-1}\mathrm{e}^{-x-s})^2 \pi(\mathrm{d}x)$$
$$= \int_{[0,-\log(\varepsilon/u))} \mathrm{d}U(s) \int_{\mathcal{S}^\downarrow} \sum_{i=1}^\infty s_i 1_{\{s_i<\varepsilon u^{-1}\mathrm{e}^s\}} g(s_i u\varepsilon^{-1}\mathrm{e}^{-s})^2 \nu(\mathrm{d}\mathbf{s})$$
$$\lesssim \frac{1}{c(\pi)}\|g\|_\infty^2 \log(u/\varepsilon),$$

where we have successively used the representation (4) and the upper bound $U(s) \lesssim s/c(\pi)$; see, for instance, [3], Proposition 1, Chapter III. Therefore, for $\varepsilon \le u \le \eta$,

$$\mathbb{E}[Y_\varepsilon(v,g)^2|\lambda_\eta(v)=u] \lesssim \frac{1}{c(\pi)}\|g\|_\infty^2 \log(\eta/\varepsilon).$$

Going back to (19), since $\lambda_\eta(v)^2 \le \eta\lambda_\eta(v)$ and again using (1), we readily derive

$$\mathbb{E}[M_{\varepsilon,\eta}(g)^2] \lesssim \frac{1}{c(\pi)}\|g\|_\infty^2 \eta\log(\eta/\varepsilon) \lesssim \eta\log(\eta/\varepsilon). \tag{22}$$

Step 3: *The bias term* $B_{\varepsilon,\eta}(g)$. First, note that

$$\mathbb{E}[Y_\varepsilon(v,g)|\lambda_\eta(v)] = \xi_v^{-1}\mathbb{E}_{\xi_v}[\mathcal{E}_\varepsilon(g)],$$

$\mathbb{P}$-almost surely, so

$$B_{\varepsilon,\eta}(g) = \sum_{v\in\mathcal{U}} \lambda_\eta(v)\bigl(\xi_v^{-1}\mathbb{E}_{\xi_v}[\mathcal{E}_\varepsilon(g)] - \mathcal{E}(g)\bigr). \tag{23}$$

Conditioning on the mark of the parent $v- = \omega$ of v and applying the branching property, we get that $\mathbb{E}_{\xi_v}[\mathcal{E}_\varepsilon(g)]$ can be written as

$$\mathbb{E}_{\xi_v}\left[\sum_{\omega\in\mathcal{U}} 1_{\{\widehat{\xi}_\omega\geq\varepsilon\}}\widehat{\xi}_\omega \int_{\mathcal{S}^\downarrow}\sum_{i=1}^\infty 1_{\{\widehat{\xi}_\omega s_i<\varepsilon\}} s_i\, g(\widehat{\xi}_\omega s_i \varepsilon^{-1})\nu(\mathrm{d}\mathbf{s})\right],$$

where the $(\widehat{\xi}_w, w\in\mathcal{U})$ are the sizes of the marked fragments of a fragmentation chain with same dislocation measure $\nu(\cdot)$, independent of $(\xi_v, v\in\mathcal{U})$. Set

$$H_g(z) := \int_{\mathcal{S}^\downarrow}\sum_{i=1}^\infty 1_{\{s_i<\mathrm{e}^{-z}\}} s_i\, g(s_i\mathrm{e}^z)\nu(\mathrm{d}\mathbf{s}), \qquad z\geq 0.$$

It follows that $\mathbb{E}_{\xi_v}[\mathcal{E}_\varepsilon(g)]$ is equal to

$$\begin{aligned}&\mathbb{E}_{\xi_v}\left[\sum_{n=0}^\infty\sum_{|\omega|=n} 1_{\{\log\widehat{\xi}_\omega\geq\log\varepsilon\}}\widehat{\xi}_\omega H_g(\log\widehat{\xi}_\omega-\log\varepsilon)\right]\\ &\quad=\xi_v\mathbb{E}\left[\sum_{n=0}^\infty\sum_{|\omega|=n} 1_{\{\log\widehat{\xi}_\omega\geq\log(\varepsilon/\rho)\}}\widehat{\xi}_\omega H_g\big(\log\widehat{\xi}_\omega-\log(\varepsilon/\rho)\big)\right]_{\rho=\xi_v},\end{aligned}$$

by self-similarity, with the notation $|\omega|=n$ if $\omega=(\omega_1,\dots,\omega_n)\in\mathcal{U}$. Using [5], Proposition 1.6, we finally obtain

$$\mathbb{E}_{\xi_v}[\mathcal{E}_\varepsilon(g)] = \xi_v\sum_{n=0}^\infty \mathbb{E}\big[1_{\{S_n\leq\log(\rho/\varepsilon)\}}H_g\big(\log(\rho/\varepsilon)-S_n\big)\big]_{\rho=\xi_v},$$

where S_n is a random walk with step distribution $\pi(\mathrm{d}x)$. Note that this can also be written as

$$\xi_v^{-1}\mathbb{E}_{\xi_v}[\mathcal{E}_\varepsilon(g)] = \mathbb{F}\star\psi\big(\log(\xi_v/\varepsilon)\big), \tag{24}$$

where $\mathbb{F}=\sum_{n=0}^\infty \pi^{n\star}$ denotes the renewal measure associated with the probability measure π and $\psi(z)=1_{z\leq 0}H_g(-z)$. In order to bound

$$\xi_v^{-1}\mathbb{E}_{\xi_v}[\mathcal{E}_\varepsilon(g)]-\mathcal{E}(g),$$

we plan to apply a version of the renewal theorem with explicit rate of convergence as given in Sgibnev [15]; see Proposition 1 in Section 5.1. We take a rate function $\varphi(z):=\exp(\mu' z)$ for some arbitrary $\mu'<\kappa/2$, a dominating function $r(z):=\mathrm{e}^{-\kappa|z|}$ and set $F:=\pi$ in Proposition 1. We can write, for $z<0$,

$$H_g(-z) = 1_{\{z\leq 0\}}\int_{(-z,+\infty)} g(\mathrm{e}^{-x-z})\pi(\mathrm{d}x),$$

by (4). Since $g(\cdot)$ has support in [0, 1] and $\pi \in \Pi(\kappa)$,

$$|H_g(-z)| \le \int_{(-z,+\infty)} |g(\mathrm{e}^{-x-z})|\pi(\mathrm{d}x) \lesssim \mathrm{e}^{\kappa z}.$$

Therefore, $|1_{\{z\le 0\}} H_g(-z)| \lesssim r(z)$ for all $z \in \mathbb{R}$. Since $\kappa > 2\mu'$, Assumption E of Proposition 1 is readily checked. Now, let $A > 0$ (depending only on κ, m and $\pi(\cdot)$) such that, if $\log(\xi_v/\varepsilon) \ge A$, then, by Proposition 1,

$$\left|\xi_v^{-1}\mathbb{E}_{\xi_v}[\mathcal{E}_\varepsilon(g)] - \frac{1}{\mathbb{E}^\star[S_1]}\int_0^{+\infty} H_g(z)\,\mathrm{d}z\right| \le \left(\frac{\varepsilon}{\xi_v}\right)^{\mu'}. \tag{25}$$

We next note that

$$\frac{1}{\mathbb{E}^\star[S_1]}\int_0^{+\infty} H_g(z)\,\mathrm{d}z = \mathcal{E}(g).$$

Introducing the family of events $\{\log(\xi_v/\varepsilon) \ge A\}$ in the sum (23), we obtain the following decomposition:

$$B_{\varepsilon,\eta}(g)^2 \lesssim I + II$$

with

$$I := \sum_{v\in\mathcal{U}_\eta} \xi_v 1_{\{\log(\xi_v/\varepsilon)>A\}}\big(\xi_v^{-1}\mathbb{E}_{\xi_v}[\mathcal{E}_\varepsilon(g)] - \mathcal{E}(g)\big)^2$$

and

$$II := \sum_{v\in\mathcal{U}_\eta} \xi_v 1_{\{\log(\xi_v/\varepsilon)\le A\}}\big(\xi_v^{-1}\mathbb{E}_{\xi_v}[\mathcal{E}_\varepsilon(g)] - \mathcal{E}(g)\big)^2.$$

By (25), we have

$$I \le \varepsilon^{2\mu'} \sum_{v\in\mathcal{U}_\eta} 1_{\{-\log\xi_v < -A+\log(1/\varepsilon)\}}\xi_v \exp\big(2\mu'(-\log\xi_v)\big).$$

Integrating with respect to $\mathbb{P}$ and applying Lemma 1, in the same way as in step 2, we have

$$\begin{aligned}
\mathbb{E}[I] &\le \varepsilon^{2\mu'}\mathbb{E}^\star\big[\mathrm{e}^{2\mu'\zeta(T_\eta)}\big] \\
&= \varepsilon^{2\mu'}\int_{[0,-\log\eta)} \mathrm{d}U(s)\int_{(-\log\eta-s,+\infty)} \mathrm{e}^{2\mu'(s+x)}\pi(\mathrm{d}x) \\
&\le \varepsilon^{2\mu'}\int_{[0,-\log\eta)} \mathrm{e}^{2\mu' s}\,\mathrm{d}U(s) \lesssim (\varepsilon\eta^{-1})^{2\mu'}\log(1/\eta)
\end{aligned}$$

for small enough ε and where we have used $\pi \in \Pi(\kappa)$ with $2\mu' < \kappa$. For the term *II*, we first note that by (1) and self-similarity,

$$\mathbb{E}_{\xi_v}\Big[\sum_{u\in\mathcal{U}_\varepsilon} \widehat{\xi}_u\Big] = \xi_v, \qquad \mathbb{P}_{\xi_v}\text{-almost surely,}$$

hence

$$\big(\xi_v^{-1}\mathbb{E}_{\xi_v}[\mathcal{E}_\varepsilon(g)] - \mathcal{E}(g)\big)^2 \le 4\|g\|_\infty^2, \qquad \mathbb{P}_{\xi_v}\text{-almost surely.} \tag{26}$$

In the same way as for the term I, we derive

$$\begin{aligned}
\mathbb{E}[II] &\lesssim \mathbb{E}\Big[\sum_{v\in\mathcal{U}_\eta} \xi_v 1_{\{-\log\xi_v \ge -A+\log(1/\varepsilon)\}}\Big] \\
&= \mathbb{P}^\star[\zeta(T_\eta) \ge -A + \log(1/\varepsilon)] \\
&\le \int_{[0,-\log\eta)} \mathrm{d}U(s) \int_{(-A+\log(1/\varepsilon)-s,+\infty)} \pi(\mathrm{d}x) \\
&\lesssim \varepsilon^{\mu'} \log(1/\eta)
\end{aligned}$$

for small enough ε. Using all of the estimates together, we conclude that

$$\mathbb{E}[B_{\varepsilon,\eta}(g)^2] \lesssim \big(\varepsilon^{\mu'} + (\varepsilon\eta^{-1})^{2\mu'}\big)\log(1/\eta). \tag{27}$$

Step 4: *Proof of* (6). Using the estimates (22) and (27), we have

$$\begin{aligned}
\mathbb{E}\big[\big(\mathcal{E}_\varepsilon(g) - \mathcal{E}(g)\big)^2\big] &\lesssim \mathbb{E}[M_{\varepsilon,\eta}(g)^2] + \mathbb{E}[B_{\varepsilon,\eta}(g)^2] \\
&\lesssim \eta\log(\eta/\varepsilon) + (\varepsilon\eta^{-1})^{2\mu'}\log(1/\eta) + \varepsilon^{\mu'}\log(1/\eta).
\end{aligned}$$

The choice $\eta(\varepsilon) := \varepsilon^{2\mu'/(2\mu'+1)}$ yields the rate

$$\varepsilon^{\min\{2\mu'/(2\mu'+1),\mu'\}}\log(1/\varepsilon) \qquad \text{for any } 0 < \mu' < \kappa/2.$$

We thus obtain a rate of the form $\mathrm{o}(\varepsilon^{\mu/(\mu+1)})$ for any $1 \le \mu < \kappa$. The conclusion follows.

Step 5: *Proof of* (7). We plan to use the following decomposition:

$$\mathcal{E}_{\varepsilon,\sigma}(g) - \mathcal{E}_\varepsilon(g) = I + II$$

with

$$I := \sum_{u\in\mathcal{U}} \big(1_{\{\xi_{u-}^{(\sigma)}\ge\varepsilon,\xi_u^{(\sigma)}<\varepsilon\}} - 1_{\{\xi_{u-}\ge\varepsilon,\xi_u<\varepsilon\}}\big)\widetilde{\xi}_u^{(\sigma)} g\big(\xi_u^{(\sigma)}/\varepsilon\big)$$

and

$$II := \sum_{u\in\mathcal{U}_\varepsilon} \big(\widetilde{\xi}_u^{(\sigma)} g\big(\xi_u^{(\sigma)}/\varepsilon\big) - \xi_u g(\xi_u/\varepsilon)\big),$$

where we have set $\widetilde{\xi}_u^{(\sigma)} := \xi_u^{(\sigma)} 1_{\{\xi_u^{(\sigma)} \geq t_\varepsilon\}}$. Clearly,

$$\left|1_{\{\xi_{u-}^{(\sigma)} \geq \varepsilon, \xi_u^{(\sigma)} < \varepsilon\}} - 1_{\{\xi_{u-} \geq \varepsilon, \xi_u < \varepsilon\}}\right| \leq 1_{\{\xi_{u-}^{(\sigma)} \geq \varepsilon, \xi_{u-} < \varepsilon\}} + 1_{\{\xi_u^{(\sigma)} < \varepsilon, \xi_u \geq \varepsilon\}} + 1_{\{\xi_{u-} \geq \varepsilon, \xi_{u-}^{(\sigma)} < \varepsilon\}} + 1_{\{\xi_u < \varepsilon, \xi_u^{(\sigma)} \geq \varepsilon\}}.$$

Let $\delta > \sigma/\varepsilon$ and $\omega = u$ or $u-$. Since $|U_\omega| \leq 1$ for every ω, we can readily check that

$$\left\{\xi_\omega^{(\sigma)} \geq \varepsilon, \xi_\omega < \varepsilon\right\} \subset \{(1-\delta)\varepsilon \leq \xi_\omega < \varepsilon\}$$

and

$$\left\{\xi_\omega \geq \varepsilon, \xi_\omega^{(\sigma)} < \varepsilon\right\} \subset \{\varepsilon \leq \xi_\omega < (1+\delta)\varepsilon\}.$$

It follows that $|I| \leq III + IV$ with

$$III := \sum_{u \in \mathcal{U}} 1_{\{(1-\delta)\varepsilon \leq \xi_{u-} \leq \varepsilon(1+\delta)\}} \left|\widetilde{\xi}_u^{(\sigma)} g\big(\xi_u^{(\sigma)}/\varepsilon\big)\right|$$

and

$$IV := \sum_{u \in \mathcal{U}} 1_{\{(1-\delta)\varepsilon \leq \xi_u \leq (1+\delta)\varepsilon\}} \left|\widetilde{\xi}_u^{(\sigma)} g\big(\xi_u^{(\sigma)}/\varepsilon\big)\right|.$$

By choosing δ to be small enough, we may (and will) assume that $\widetilde{\xi}_u^{(\sigma)} \lesssim \xi_u$. Conditioning on the mark of the parent $u- = v$ of u, using the branching property, Jensen's inequality and the conservative Assumption 1, we conclude that $\mathbb{E}[III^2]$ is less than

$$\mathbb{E}\left[\sum_{v \in \mathcal{U}} 1_{\{(1-\delta)\varepsilon \leq \xi_v \leq \varepsilon(1+\delta)\}} \xi_v \int_{\mathcal{S}^\downarrow} \sum_{i=1}^{\infty} s_i\, g\big(\varepsilon^{-1}(\xi_v s_i + \sigma U_v)\big)^2 \nu(\mathrm{d}\mathbf{s})\right]$$
$$= \mathbb{E}\left[\sum_{\omega \in \mathcal{U}} 1_{\{(1-\delta)\varepsilon \leq \xi_\omega \leq \varepsilon(1+\delta)\}} \xi_\omega G_1(\xi_\omega)\right]$$

with

$$G_1(a) := \int_{\mathcal{S}^\downarrow} \sum_{i=1}^{\infty} s_i \mathbb{E}\big[g\big(\varepsilon^{-1}(a s_i + \sigma U)\big)^2\big] \nu(\mathrm{d}\mathbf{s})$$

and U distributed as the U_ω. Likewise,

$$\mathbb{E}[IV^2] \lesssim \mathbb{E}\left[\sum_{u \in \mathcal{U}} 1_{\{(1-\delta)\varepsilon \leq \xi_u \leq \varepsilon(1+\delta)\}} \xi_u G_2(\xi_u)\right]$$

with $G_2(a) := \mathbb{E}[g(\varepsilon^{-1}(a + \sigma U))^2]$. For $i = 1, 2$, the crude bound $|G_i(a)| \leq \|g\|_\infty^2$ and the genealogical representation argument used in step 3 enable us to bound either $\mathbb{E}[III^2]$ or $\mathbb{E}[IV^2]$

by

$$\|g\|_\infty^2 \sum_{n=0}^{\infty} \mathbb{P}^\star[-\log(1+\delta) \le S_n - \log(1/\varepsilon) \le -\log(1-\delta)],$$

where S_n is a random walk with step distribution $\pi(\cdot)$. We proceed as in step 3 and apply Proposition 1. The above term converges to

$$m_1(\pi)^{-1} \log\left(\frac{1+\delta}{1-\delta}\right) \lesssim \delta$$

uniformly in δ, provided that δ is bounded, at rate $\varepsilon^{\mu'}$ for any $0 < \mu' < \kappa/2$, and is thus of order $\delta + \varepsilon^{\mu'}$. We next turn to the term *II*. We have $II := V + VI + VII$ with

$$V := \sum_{u \in \mathcal{U}_\varepsilon} \xi_u \big(g\big(\xi_u^{(\sigma)}/\varepsilon\big) - g(\xi_u/\varepsilon)\big),$$

$$VI := \sigma \sum_{u \in \mathcal{U}_\varepsilon} U_u 1_{\{\xi_u^{(\sigma)} \ge t_\varepsilon\}} g\big(\xi_u^{(\sigma)}/\varepsilon\big),$$

$$VII := -\sum_{u \in \mathcal{U}_\varepsilon} \xi_u 1_{\{\xi_u^{(\sigma)} < t_\varepsilon\}} g\big(\xi_u^{(\sigma)}/\varepsilon\big).$$

From $g \in \mathcal{C}'(m)$, (1), Jensen's inequality and a Taylor expansion, we derive that

$$\mathbb{E}[V^2] \le \|g'\|_\infty^2 \sigma^2 \varepsilon^{-2}.$$

From $|U_u| \le 1$ and the inclusion $\{\xi_u^{(\sigma)} \ge t_\varepsilon\} \subset \{\xi_u \ge t_\varepsilon - \sigma\}$, we derive

$$\mathbb{E}[VI^2] \le \|g\|_\infty^2 \frac{\sigma^2}{(t_\varepsilon - \sigma)^2} \mathbb{E}\left[\left(\sum_{u \in \mathcal{U}_\varepsilon} \xi_u\right)^2\right] \lesssim \frac{\sigma^2}{\varepsilon^2},$$

where we have used the fact that $t_\varepsilon = \gamma_0 \varepsilon$ with $0 < \gamma_0 < 1$ and $\sigma \le t_\varepsilon/2$. Likewise, the inclusion $\{\xi_u^{(\sigma)} < t_\varepsilon\} \subset \{\xi_u \le t_\varepsilon + \sigma\}$ and Lemma 1 yield

$$\mathbb{E}[VII^2] \le \|g\|_\infty^2 \mathbb{P}^\star[-\log \chi(T_\varepsilon) > -\log(t_\varepsilon + \sigma)] \lesssim \varepsilon^{\mu'} \log(1/\varepsilon)$$

for any $0 < \mu' < \kappa/2$, along the same lines as for the bound of the right-hand side of (21) in step 2. Putting all of the estimates together with, for instance, $\delta := \sigma/2\varepsilon$, we finally obtain a rate of the form

$$\varepsilon^{\mu'} \log(1/\varepsilon) + \sigma \varepsilon^{-1} \qquad \text{for any } 0 < \mu' < \kappa/2,$$

which can be written as $\mathrm{o}(\varepsilon^{\mu/2}) + \mathcal{O}(\sigma \varepsilon^{-1})$ for any $0 < \mu < \kappa$. We thus obtain (7) and the proof of Theorem 1 is complete.

5.3. Proof of Theorem 2

Preliminaries. We begin with a technical lemma.

Lemma 2. *We work under Assumption* C. *Assume, moreover, that* $\pi \in \mathcal{R}(\kappa_2)$ *with* $\kappa_2 > 1$. *We have*

$$\sup_{a\in(0,1)} \beta(a) < +\infty.$$

Proof. By Assumption C, $x \rightsquigarrow \pi(x)$ is continuous on $(0,+\infty)$, hence $\beta(a) = a^{-1}\pi(-\log a)$ is continuous on $(0,1)$ and it suffices to show that $\beta(\cdot)$ is bounded in the vicinity of 0 and 1. By assumption, $\pi(x) \lesssim \mathrm{e}^{-\vartheta x}$ for some $\vartheta \geq 1$ near $+\infty$, so $\beta(a) \lesssim a^{\vartheta-1}$ near the origin and this term remains bounded as $a \to 0$. By assumption, we also have $\pi \in \mathcal{R}(\kappa_2)$, so $\pi(x) \lesssim x^{\kappa_2-1}$ near the origin, therefore $\beta(a) \lesssim (-\log a)^{\kappa_2-1}$ near 1 and this term remains bounded as $a \to 1$ since $\kappa_2 > 1$. □

Let $0 < b_\varepsilon \to 0$ as $\varepsilon \to 0$. For $m > 0$, define the class

$$\widetilde{\mathcal{C}}_{b_\varepsilon}(m) := \{g \in \mathcal{C}(m), |\mathrm{supp}(g)| \leq m b_\varepsilon\}.$$

We have the following extension of Theorem 1.

Proposition 2. *We work under Assumptions* A, B *and* C. *Assume that* $\pi \in \Pi(\kappa_1) \cap \mathcal{R}(\kappa_2)$ *with* $\kappa_1, \kappa_2 > 1$. *Then, for every* $1 \leq \mu < \kappa_1 + 1$,

$$\sup_{g\in\widetilde{\mathcal{C}}_{b_\varepsilon}(m)} \mathbb{E}\big[\big(\mathcal{E}_\varepsilon(g) - \mathcal{E}(g)\big)^2\big] = \mathrm{o}\big(\varepsilon^{\mu/(\mu+1)} b_\varepsilon\big).$$

Proof. We carefully revisit steps 2–4 of the proof of Theorem 1, under the additional Assumption C, and we write $g(\cdot) = g_\varepsilon(\cdot)$ to emphasize that $g(\cdot)$ may now depend on the asymptotics.

In step 2, the right-hand side of (21) is now bounded by the following chain of inequalities:

$$\begin{aligned}
&\int_0^{-\log(\varepsilon/u)} \mathrm{d}U(s) \int_{-\log(\varepsilon/u)-s}^{+\infty} g_\varepsilon(u\varepsilon^{-1}\mathrm{e}^{-x-s})^2 \pi(x)\,\mathrm{d}x \\
&\quad= \int_0^{-\log(\varepsilon/u)} \mathrm{d}U(s) \int_0^{\varepsilon u^{-1}\mathrm{e}^s} g_\varepsilon(xu\varepsilon^{-1}\mathrm{e}^{-s})^2 \beta(x)\,\mathrm{d}x \\
&\quad\leq \sup_{a\in(0,1)} \beta(a) u^{-1}\varepsilon \int_{[0,-\log(\varepsilon/u))} \mathrm{e}^s\,\mathrm{d}U(s) \int_0^1 g_\varepsilon(x)^2\,\mathrm{d}x \lesssim b_\varepsilon \log(u/\varepsilon),
\end{aligned}$$

where we have used Lemma 2, the fact that $|\,\mathrm{supp}(g_\varepsilon)| \lesssim b_\varepsilon$ and $U(s) \lesssim s/c(\pi)$ again. Therefore,

$$\mathbb{E}[Y_\varepsilon(v,g)^2 | \lambda_\eta(v) = u] \lesssim b_\varepsilon \log(\eta/\varepsilon),$$

hence

$$\mathbb{E}[M_{\varepsilon,\eta}(g)^2] \lesssim b_\varepsilon \eta \log(\eta/\varepsilon).$$

In step 3, we replace $g(\cdot)$ by $g_\varepsilon(\cdot)$ in $\mathcal{E}_\varepsilon(g)$ and $\mathcal{E}(g)$. We first consider the term I. We need to be careful when applying Proposition 1 because $H_{g_\varepsilon}(z)$ now depends on ε. By the Cauchy–Schwarz inequality, for $z < 0$,

$$\begin{aligned}|H_{g_\varepsilon}(-z)| &\le \left(\int_{-z}^{+\infty} g_\varepsilon(\mathrm{e}^{-x-z})^2 \pi(x)\,\mathrm{d}x\right)^{1/2} \left(\int_{-z}^{+\infty} \pi(x)\,\mathrm{d}x\right)^{1/2}\\ &\lesssim \mathrm{e}^{z/2}\left(\int_0^1 g_\varepsilon(y)^2 \beta(y\mathrm{e}^z)\,\mathrm{d}y\right)^{1/2} \mathrm{e}^{\kappa_1 z/2} \lesssim b_\varepsilon^{1/2}\mathrm{e}^{z(1+\kappa_1)/2},\end{aligned}$$

again using the fact that $\sup_a \beta(a) \lesssim 1$. We can therefore apply Proposition 1 when $0 < \mu' < (1+\kappa_1)/2$ with rate function $\varphi(z) := \exp(\mu' z)$, dominating function $r(z) := \mathrm{e}^{-(1+\kappa_1)|z|/2}$, test function $\psi(z) := b_\varepsilon^{-1/2} 1_{z\le 0} H_g(z)$ and $F := \pi$. We then obtain, along the same lines as in step 3, for $0 < \mu' < (1+\kappa_1)/2$, the estimate

$$\mathbb{E}[I] \lesssim b_\varepsilon^{1/2} (\varepsilon\eta^{-1})^{2\mu'} \log(1/\eta).$$

For the term *II*, it suffices to prove that both $\xi_v^{-1}\mathbb{E}_{\xi_v}[\mathcal{E}_\varepsilon(g_\varepsilon)]$ and $\mathcal{E}(g_\varepsilon)$ are smaller in order than $b_\varepsilon^{1/2}$; recall (26). For the first term, this follows from the previous bound on $H_{g_\varepsilon}(z)$ and the representation (24). For $\mathcal{E}(g_\varepsilon)$, since $\pi \in \Pi(\kappa_1)$ with $\kappa_1 > 1$, we have, successively,

$$\begin{aligned}|\mathcal{E}(g_\varepsilon)| &\le \frac{1}{c(\pi)} \int_0^1 \frac{|g_\varepsilon(a)|}{a} \int_{\log(1/a)}^{+\infty} \pi(x)\,\mathrm{d}x\,\mathrm{d}a\\ &\lesssim \int_0^1 |g_\varepsilon(a)| a^{\kappa_1 - 1}\,\mathrm{d}a \lesssim \int_0^1 |g_\varepsilon(a)|\,\mathrm{d}a \lesssim b_\varepsilon.\end{aligned}$$

We eventually obtain

$$\mathbb{E}[B_{\varepsilon,\eta}(g)^2] \lesssim b_\varepsilon\big(\varepsilon^{\mu'} + (\varepsilon\eta^{-1})^{2\mu'}\big)\log(1/\eta)$$

for any $0 < \mu' < (1+\kappa_1)/2$. The trade-off between $M_{\varepsilon,\eta}(g_\varepsilon)$ and $B_{\varepsilon,\eta}(g_\varepsilon)$ yields the rate

$$\varepsilon^{\max\{2\mu'/(2\mu'+1),\mu'\}} b_\varepsilon \qquad \text{for any } 0 < \mu' < (1+\kappa_1)/2,$$

which is of the form $\mathrm{o}(\varepsilon^{\mu/(\mu+1)} b_\varepsilon)$ for any $1 \le \mu < 1 + \kappa_1$, hence the result. □

Completion of proof of Theorem 2. By the representation formula (9), we can write

$$\mathcal{E}(g_{\gamma_\varepsilon}) - m_1(\pi)^{-1} = \frac{1}{m_1(\pi)} \int_{1-\gamma_\varepsilon}^1 \big(f_{\gamma_\varepsilon}(a) - 1\big)\beta(a)\,\mathrm{d}a,$$

where the integral is taken over $[1-\gamma_\varepsilon, 1]$ since $f_{\gamma_\varepsilon}(a) = 1$ on $[0, 1-\gamma_\varepsilon]$ and $\beta(\cdot)$ is a density function with respect to the Lebesgue measure on $(0, 1)$. We further have

$$\left|\int_{1-\gamma_\varepsilon}^{1} \big(f_{\gamma_\varepsilon}(a) - 1\big)\beta(a)\,\mathrm{d}a\right| \lesssim \int_0^{-\log(1-\gamma_\varepsilon)} \pi(x)\,\mathrm{d}x \lesssim \gamma_\varepsilon^{\kappa_2}$$

since $\|f_\gamma\|_\infty \lesssim 1$, by (10), $\pi \in \mathcal{R}(\kappa_2)$ and $-\log(1-x) \lesssim x$ for small enough $x \geq 0$. We deduce that

$$|\mathcal{E}(g_{\gamma_\varepsilon}) - m_1(\pi)^{-1}| \lesssim \gamma_\varepsilon^{\kappa_2}. \tag{28}$$

Next, for some $c > 0$, $\gamma_\varepsilon g_{\gamma_\varepsilon} \in \widetilde{\mathcal{C}}_{\gamma_\varepsilon}(c)$, hence, for any $0 < \mu < \kappa_1$, Proposition 2 entails that

$$\mathbb{E}[|\mathcal{E}_\varepsilon(g_{\gamma_\varepsilon}) - \mathcal{E}(g_{\gamma_\varepsilon})|] \lesssim \gamma_\varepsilon^{-1/2}\varepsilon^{\mu/(2\mu+2)}. \tag{29}$$

Moreover,

$$g'_{\gamma_\varepsilon}(a) = -f'_{\gamma_\varepsilon}(a) - a f''_{\gamma_\varepsilon}(a),$$

hence, by property (10), we have $\gamma_\varepsilon^2 g_{\gamma_\varepsilon} \in \mathcal{C}'(c)$ for some $c > 0$. Applying (7) of Theorem 1, we deduce that

$$\mathbb{E}[|\mathcal{E}_\varepsilon(g_{\gamma_\varepsilon}) - \mathcal{E}_{\varepsilon,\sigma}(g_{\gamma_\varepsilon})|] \lesssim \gamma_\varepsilon^{-2}[(\sigma\varepsilon^{-1})^{1/2} + \varepsilon^{\mu'/4}] \tag{30}$$

for any $0 < \mu' < \kappa_1$. The specification $\gamma_\varepsilon = \varepsilon^{\mu/(\mu+1)(2\kappa_2+1)}$ yields the correct rate for (28) and (29). The assumption that $\sigma\varepsilon^{-3}$ is bounded ensures that the term $\gamma_\varepsilon^{-2}(\sigma\varepsilon^{-1})^{1/2}$ in (30) is asymptotically negligible since $\kappa_2 \geq 1$. Using the fact that $\kappa_1 \geq 4$, the term $\gamma_\varepsilon^{-2}\varepsilon^{\mu/4}$ also proves negligible by taking μ' sufficiently close to 4. The conclusion readily follows for $\widehat{m}_{1,\varepsilon}$.

We now turn to higher moment estimators. Thanks to the proof for the case $k = 1$, it suffices to show that

$$m_1(\pi)\mathcal{E}_{\varepsilon,\sigma}(\widetilde{g}_{\gamma_\varepsilon}) \to \int_0^1 \left(\log\frac{1}{a}\right)^k \beta(a)\,\mathrm{d}a$$

in probability with the correct rate as $\varepsilon \to 0$. Note, first, that by representation (9),

$$\begin{aligned}\mathcal{E}(\widetilde{g}_{\gamma_\varepsilon}) &= \frac{1}{m_1(\pi)}\int_0^1 h_{\gamma_\varepsilon}(a)\beta(a)\,\mathrm{d}a\\ &= \frac{1}{m_1(\pi)}\int_0^1 f_{\gamma_\varepsilon}(1-a)\left(\log\frac{1}{a}\right)^k \beta(a)\,\mathrm{d}a,\end{aligned}$$

therefore

$$m_1(\pi)\mathcal{E}(\widetilde{g}_{\gamma_\varepsilon}) - \int_0^1 \left(\log\frac{1}{a}\right)^k \beta(a)\,\mathrm{d}a = \int_0^{\gamma_\varepsilon} \big(f_{\gamma_\varepsilon}(1-a) - 1\big)\left(\log\frac{1}{a}\right)^k \beta(a)\,\mathrm{d}a$$

since $f_{\gamma_\varepsilon}(1-a)=1$ if $a\geq\gamma_\varepsilon$. It follows that

$$\left|\int_0^{\gamma_\varepsilon}\big(f_{\gamma_\varepsilon}(1-a)-1\big)\left(\log\frac{1}{a}\right)^k\beta(a)\,\mathrm{d}a\right|\lesssim\int_0^{\gamma_\varepsilon}\left(\log\frac{1}{a}\right)^k\beta(a)\,\mathrm{d}a\lesssim\int_{\log 1/\gamma_\varepsilon}^{+\infty}x^k\pi(x)\,\mathrm{d}x$$
$$\lesssim\left(\int_{-\log\gamma_\varepsilon}^{+\infty}\pi(x)\,\mathrm{d}x\right)^{1-\delta'}\left(\int_0^{+\infty}x^{k/\delta'}\pi(x)\,\mathrm{d}x\right)^{\delta'}$$
$$\lesssim\gamma_\varepsilon^{\kappa_1(1-\delta')}$$

for any $0<\delta'<1$, by Hölder's inequality and where we have used the fact that $\pi\in\Pi(\kappa_1)$. The second integral in the last line is finite by Assumption C. Since the choice of δ' is free, the choice of γ_ε and the assumption that $\kappa_1>\kappa_2$ show that this term is asymptotically negligible with respect to $(\varepsilon^{\mu/(\mu+1)})^{\kappa_2/(2\kappa_2+1)}$. Therefore, it suffices to show that

$$\mathcal{T}_\varepsilon=\mathcal{E}_{\varepsilon,\sigma}(\widetilde{g}_{\gamma_\varepsilon})-\mathcal{E}(\widetilde{g}_{\gamma_\varepsilon})$$

has order $(\varepsilon^{\mu/(\mu+1)})^{\kappa_2/(2\kappa_2+1)}$. We split $\mathcal{T}_\varepsilon=\mathcal{T}_{\varepsilon,1}+\mathcal{T}_{\varepsilon,2}$ with

$$\mathcal{T}_{\varepsilon,1}=\mathcal{E}_{\varepsilon,\sigma}(\widetilde{g}_{\gamma_\varepsilon})-\mathcal{E}_\varepsilon(\widetilde{g}_{\gamma_\varepsilon})\quad\text{and}\quad\mathcal{T}_{\varepsilon,2}=\mathcal{E}_\varepsilon(\widetilde{g}_{\gamma_\varepsilon})-\mathcal{E}(\widetilde{g}_{\gamma_\varepsilon}).$$

Lemma 3. *There exists some constant $c>0$, independent of ε, such that*:

- *we have $\gamma_\varepsilon^2\widetilde{g}_{\gamma_\varepsilon}\in\mathcal{C}'(c)$;*
- *the decomposition*

$$\widetilde{g}_{\gamma_\varepsilon}(a)=q_{1,\gamma_\varepsilon}(a)+q_{2,\gamma_\varepsilon}(a)\tag{31}$$

holds, so that for any $0<\delta'<1$, we have $\gamma_\varepsilon^{\delta'}q_{1,\gamma_\varepsilon}\in\widetilde{\mathcal{C}}_{\gamma_\varepsilon}(c)$ and $\gamma_\varepsilon^{\delta'}q_{2,\gamma_\varepsilon}\in\mathcal{C}(c)$.

Proof. Tedious but straightforward computations show that

$$\begin{aligned}\widetilde{g}'_{\gamma_\varepsilon}(a)&=c_{k,1}a^{-1}(\log a)^{k-2}f_{\gamma_\varepsilon}(1-a)\\&\quad+[c_{k,2}(\log a)^k+c_{k,3}(\log a)^{k-1}]f'_{\gamma_\varepsilon}(1-a)+c_{k,4}a(\log a)^kf''_{\gamma_\varepsilon}(1-a)\end{aligned}$$

with explicit constants $c_{k,1}=(-1)^{k+1}k(k-1)$, $c_{k,2}=(-1)^k$, $c_{k,3}=(-1)^k(k+1)k$ and $c_{k,4}=(-1)^{k+1}$. Using property (11) of f_{γ_ε}, one readily checks that the four terms multiplied by γ_ε^2 are bounded. For the last term, corresponding to the constant $c_{k,4}$, the property (10) of f_{γ_ε} also shows that this term multiplied by γ_ε^2 has the correct order, so $\gamma_\varepsilon^2\widetilde{g}_\varepsilon\in\mathcal{C}'(c)$ for some $c>0$.

For the second part of the lemma, we have (31) with

$$q_{1,\gamma_\varepsilon}(a)=(-1)^kaf'_{\gamma_\varepsilon}(1-a)(\log a)^k$$

and

$$q_{2,\gamma_\varepsilon}(a)=(-1)^{k+1}f_{\gamma_\varepsilon}(1-a)k(\log a)^{k-1}.$$

By construction of f_{γ_ε}, we have $\mathrm{supp}(q_{1,\gamma_\varepsilon}) \subset [0, \gamma_\varepsilon]$. It follows that for any $0 < \delta' < 1$ and $a \in (0, 1)$, we have

$$|q_{1,\gamma_\varepsilon}(a)| \leq a^{\delta'}|\log a|^k a^{1-\delta'}|f'_{\gamma_\varepsilon}(1-a)| \lesssim \gamma_\varepsilon^{1-\delta'}\|f'_{\gamma_\varepsilon}\|_\infty \lesssim \gamma_\varepsilon^{-\delta'},$$

where we have used the fact that $\sup_{a\in(0,1)} a^{\delta'}|\log a|^k < +\infty$, the fact that $\mathrm{supp}(q_{1,\gamma_\varepsilon}) \subset [0, \gamma_\varepsilon]$ and property (10). We conclude that $\gamma_\varepsilon^{\delta'} q_{1,\gamma_\varepsilon} \in \widetilde{\mathcal{C}}_{\gamma_\varepsilon}(c)$ for some $c > 0$.

For the term q_{2,γ_ε}, we have, for any $a \in (0, \gamma_\varepsilon]$ and any $0 < \delta' < 1$,

$$|q_{2,\gamma_\varepsilon}(a)| \leq ka^{\delta'}|\log a|^{k-1}a^{1+\delta-\delta'}\left(\left(\frac{\gamma}{a}\right)^{1+\delta} f_\gamma(1-a)\right) \lesssim 1,$$

where we have again used the fact that $\sup_{a\in(0,1)} a^{\delta'}|\log a|^k < +\infty$ and property (11). For $a \geq \gamma_\varepsilon$, we directly have $|q_{2,\varepsilon}(a)| \lesssim |\log \gamma_\varepsilon|^{k-1}$, which is smaller in order than $\gamma_\varepsilon^{-\delta'}$ as $\varepsilon \to 0$. □

The first part of Lemma 3 enables us to apply (7) of Theorem 1: we obtain

$$\mathbb{E}[|\mathcal{T}_{\varepsilon,1}|] \lesssim \gamma_\varepsilon^{-2}[(\sigma\varepsilon^{-1})^{1/2} + \varepsilon^{\mu'/4}]$$

for any $0 < \mu' < \kappa_1$ and this term is asymptotically negligible in the same way as for (30). The second part of Lemma 3 enables us to apply Proposition 2 to the term q_{1,γ_ε} and Theorem 1 to the term q_{2,γ_ε}, respectively. It follows that

$$\begin{aligned}\mathbb{E}[|\mathcal{T}_{2,\varepsilon}|] &\leq \mathbb{E}[|q_{1,\gamma_\varepsilon}|] + \mathbb{E}[|q_{2,\gamma_\varepsilon}|] \\ &\lesssim \gamma_\varepsilon^{1/2}\gamma_\varepsilon^{-\delta'}\varepsilon^{\mu/2(\mu+1)} + \gamma_\varepsilon^{-\delta'}\varepsilon^{\mu/2(\mu+1)} \lesssim \gamma_\varepsilon^{-\delta'}\varepsilon^{\mu/2(\mu+1)}.\end{aligned}$$

One readily checks that the choice $\delta' < 1/2$ shows that this term is negligible. The proof of Theorem 2 is thus complete.

5.4. Proof of Theorem 3

Without loss of generality, we consider the homogeneous case with $\alpha = 0$. We may also assume that $\sigma = 0$ since adding experimental noise to the observation of the fragments only increases the error bounds.

Step 1: *An augmented experiment*. In the binary case, the dislocation measure $\nu(d\mathbf{s})$ is equivalently expressed via a probability measure on $[1/2, 1]$ with density function $a \rightsquigarrow \rho(a)$; see (14).

We prove a lower bound in the augmented experiment, where one can observe all of the sizes $\widetilde{X}_\varepsilon$ of the fragments until they become smaller than ε, namely,

$$\widetilde{X}_\varepsilon := \{\xi_u, \xi_{u-} \geq \varepsilon\} \cup \{\xi_u, u \in \mathcal{U}_\varepsilon\}.$$

Clearly, taking the infimum over all estimators based on $\widetilde{X}_\varepsilon$ instead of $X_\varepsilon = X_{\varepsilon,0}$ only reduces the lower bound.

For every $u \in \mathcal{U}_\varepsilon$, we have $\xi_{u-} \geq \varepsilon$. By the conservative Assumption B, there are at most ε^{-1} such ξ_{u-}, so $\operatorname{Card}\mathcal{U}_\varepsilon \leq 2\varepsilon^{-1}$. For every node $u \in \mathcal{U}$, the fragmentation process gives rise to two offspring with sizes $\xi_u U$ and $\xi_u(1-U)$, where U is a random variable independent of ξ_u with density function $\rho(\cdot)$. Therefore, the process of the sizes of the fragments in the enlarged experiment can be realized by fewer than

$$2\varepsilon^{-1}\left(1+\frac{1}{2}+\cdots+\frac{1}{2^{k(\varepsilon)}}\right) \leq \lfloor 4\varepsilon^{-1}\rfloor + 1 =: n(\varepsilon)$$

independent realizations of the law $\rho(\cdot)$, where $k(\varepsilon) := \log_2(2/\varepsilon)$, assumed to be an integer with no loss of generality.

In turn, Theorem 3 reduces to proving that $\varepsilon^{1/2}$ is a lower rate of convergence for estimating $m_k(\pi)$ based on the observation of an $n(\varepsilon)$-sample of the law $\rho(\cdot)$. The one-to-one correspondence between $\rho(\cdot)$ and $\pi(\cdot)$ is given in (16).

Step 2: *Construction of* π_ε. We write $\rho_\pi(\cdot)$ to emphasize the dependence on $\pi(\cdot)$. Let

$$\phi_k(a) := a\log(1/a)^k + (1-a)\log\bigl(1/(1-a)\bigr)^k, \qquad a \in [1/2, 1].$$

From (17), we have

$$m_k(\pi_0) = \int_{1/2}^1 \phi_k(a)\rho_{\pi_0}(a)\,\mathrm{d}a.$$

Let $0 < \tau < 1$. Choose a function $\psi_k(\cdot) : [1/2, 1] \to \mathbb{R}$ such that

$$\|\psi_k\|_\infty \leq \tau \inf_a \rho_{\pi_0}(a), \qquad \int_{1/2}^1 \psi_k(a)\,\mathrm{d}a = 0, \qquad r(k) := \int_{1/2}^1 \phi_k(a)\psi_k(a)\,\mathrm{d}a \neq 0,$$

a choice which is obviously possible thanks to Assumption D. For $\varepsilon > 0$, define

$$\rho_{\pi_\varepsilon}(a) := \rho_{\pi_0}(a) + \varepsilon^{1/2}\psi_k(a), \qquad a \in [1/2, 1].$$

(Therefore, (16) defines $\pi_\varepsilon(\cdot)$ unambiguously.) By construction, $\rho_{\pi_\varepsilon}(\cdot)$ is a density function on $[1/2, 1]$ and has a corresponding binary fragmentation with Lévy measure given by $\pi_\varepsilon(\cdot)$. Moreover,

$$m_k(\pi_\varepsilon) = m_k(\pi_0) + r(k)\varepsilon^{1/2}.$$

Step 3: *A two-point lower bound*. The following chain of arguments is fairly classical. We denote by $\widetilde{\mathbb{P}}_\pi$ the law of the independent random variables $(U_i, i = 1, \ldots, n(\varepsilon))$ with common density $\rho_\pi(\cdot)$ that we use to realize the augmented experiment.

Let F_ε be an arbitrary estimator of $m_k(\pi)$ based on $\widetilde{X}_\varepsilon$. Put $c := |r(k)|/2$. We have

$$\begin{aligned}
&\max_{\pi \in \{\pi_0, \pi_\varepsilon\}} \widetilde{\mathbb{P}}_\pi\bigl[\varepsilon^{-1/2}|F_\varepsilon - m_k(\pi)| \geq c\bigr] \\
&\quad \geq \tfrac{1}{2}\bigl(\widetilde{\mathbb{P}}_{\pi_0}\bigl[\varepsilon^{-1/2}|F_\varepsilon - m_k(\pi_0)| \geq c\bigr] + \widetilde{\mathbb{P}}_{\pi_\varepsilon}\bigl[\varepsilon^{-1/2}|F_\varepsilon - m_k(\pi_\varepsilon)| \geq c\bigr]\bigr) \\
&\quad \geq \tfrac{1}{2}\widetilde{\mathbb{E}}_{\pi_0}\bigl[1_{\{\varepsilon^{-1/2}|F_\varepsilon - m_k(\pi_0)| \geq c\}} + 1_{\{\varepsilon^{-1/2}|F_\varepsilon - m_k(\pi_\varepsilon)| \geq c\}}\bigr] - \tfrac{1}{2}\|\widetilde{\mathbb{P}}_{\pi_0} - \widetilde{\mathbb{P}}_{\pi_\varepsilon}\|_{\mathrm{TV}},
\end{aligned}$$

where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation distance between probability measures. By the triangle inequality, we have

$$\varepsilon^{-1/2}\big(|F_\varepsilon - m_k(\pi_0)| + |F_\varepsilon - m_k(\pi_\varepsilon)|\big) \geq |r(k)| = 2c,$$

so one of the two indicators within the expectation above must be equal to one with full $\widetilde{\mathbb{P}}_{\pi_0}$-probability. Therefore,

$$\max_{\pi\in\{\pi_0,\pi_\varepsilon\}} \widetilde{\mathbb{P}}_\pi[\varepsilon^{-1/2}|F_\varepsilon - m_k(\pi)| \geq c] \geq \tfrac{1}{2}(1 - \|\widetilde{\mathbb{P}}_{\pi_0} - \widetilde{\mathbb{P}}_{\pi_\varepsilon}\|_{\mathrm{TV}})$$

and Theorem 3 is proved if

$$\limsup_{\varepsilon\to 0} \|\widetilde{\mathbb{P}}_{\pi_0} - \widetilde{\mathbb{P}}_{\pi_\varepsilon}\|_{\mathrm{TV}} < 1. \tag{32}$$

By Pinsker's inequality, $\|\widetilde{\mathbb{P}}_{\pi_0} - \widetilde{\mathbb{P}}_{\pi_\varepsilon}\|_{\mathrm{TV}} \leq \frac{\sqrt{2}}{2}(\widetilde{\mathbb{E}}_{\pi_0}[\log \frac{\mathrm{d}\widetilde{\mathbb{P}}_{\pi_0}}{\mathrm{d}\widetilde{\mathbb{P}}_{\pi_\varepsilon}}])^{1/2}$ and

$$\begin{aligned}
\widetilde{\mathbb{E}}_{\pi_0}\left[\log \frac{\mathrm{d}\widetilde{\mathbb{P}}_{\pi_0}}{\mathrm{d}\widetilde{\mathbb{P}}_{\pi_\varepsilon}}\right] &= -\sum_{i=1}^{n(\varepsilon)} \widetilde{\mathbb{E}}_{\pi_0}\left[\log \frac{\rho_{\pi_\varepsilon}(U_i)}{\rho_{\pi_0}(U_i)}\right] \\
&= -\sum_{i=1}^{n(\varepsilon)} \widetilde{\mathbb{E}}_{\pi_0}\big[\log\big(1 + \varepsilon^{1/2}\psi_k(U_i)\rho_{\pi_0}(U_i)^{-1}\big) - \varepsilon^{1/2}\psi_k(U_i)\rho_{\pi_0}(U_i)^{-1}\big],
\end{aligned}$$

where we have used the fact that $\widetilde{\mathbb{E}}_{\pi_0}[\psi_k(U_i)\rho_{\pi_0}(U_i)^{-1}] = \int_{1/2}^1 \psi_k(a)\,\mathrm{d}a = 0$. We also have that the term $\varepsilon^{1/2}|\psi_k(U_i)\rho_{\pi_0}(U_i)^{-1}|$ is smaller than $\tau\varepsilon^{1/2}$. Hence, for small enough τ,

$$\big|-\log\big(1 + \varepsilon^{1/2}\psi_k(U_i)\rho_{\pi_0}(U_i)^{-1}\big) + \varepsilon^{1/2}\psi_k(U_i)\rho_{\pi_0}(U_i)^{-1}\big| \leq \tau^2\varepsilon.$$

Therefore $\|\widetilde{\mathbb{P}}_{\pi_0} - \widetilde{\mathbb{P}}_{\pi_\varepsilon}\|_{\mathrm{TV}} \leq \frac{\sqrt{2}}{2}\tau\varepsilon^{1/2}n(\varepsilon)^{1/2}$ and this quantity is bounded away from 1 by choosing τ small enough, uniformly in n, so (32) follows. The proof of Theorem 3 is thus complete.

5.5. Proof of Theorem 4

We plan to use the following decomposition:

$$\widehat{\beta}(a) - \beta(a) = \widehat{m}_{1,\varepsilon}\mathcal{E}_{\varepsilon,\sigma}(h_{a,\varepsilon}) - \beta(a) = I + II + III + IV$$

with

$$\begin{aligned}
I &:= \widehat{m}_{1,\varepsilon}\big(\mathcal{E}_{\varepsilon,\sigma}(h_{a,\varepsilon}) - \mathcal{E}_\varepsilon(h_{a,\varepsilon})\big), \\
II &:= \widehat{m}_{1,\varepsilon}\big(\mathcal{E}_\varepsilon(h_{a,\varepsilon}) - \mathcal{E}(h_{a,\varepsilon})\big), \\
III &:= \big(\widehat{m}_{1,\varepsilon} - m_1(\pi)\big)\mathcal{E}(h_{a,\varepsilon}), \\
IV &:= m_1(\pi)\mathcal{E}(h_{a,\varepsilon}) - \beta(a).
\end{aligned}$$

Considering *I* and *II*, the term $\widehat{m}_{1,\varepsilon}$ is bounded in probability by Theorem 2. By (7) in Theorem 1, together with the fact that $\gamma_\varepsilon^3 \varphi'_{\gamma_\varepsilon,a} \in \mathcal{C}'(\|\varphi''\|_\infty)$, we have

$$\mathbb{E}[|\mathcal{E}_\varepsilon(h_{a,\varepsilon}) - \mathcal{E}_{\varepsilon,\sigma}(h_{a,\varepsilon})|] \lesssim \gamma_\varepsilon^{-3}[(\sigma\varepsilon^{-1})^{1/2} + \varepsilon^{\mu'/4}] \tag{33}$$

for any $0 < \mu' < \kappa_1$. By construction, we have $\gamma_\varepsilon^2 \cdot \varphi'_{\gamma_\varepsilon,a}(\cdot) \in \widetilde{\mathcal{C}}_{\gamma_\varepsilon}(\|\varphi'\|_\infty)$. Therefore, by Proposition 2,

$$\mathbb{E}\big[\big(\mathcal{E}_\varepsilon(h_{a,\varepsilon}) - \mathcal{E}(h_{a,\varepsilon})\big)^2\big] \lesssim \gamma_\varepsilon^{-3}\varepsilon^{\mu/(\mu+1)}. \tag{34}$$

Considering *III*, note that for all $a \in (0,1)$, the function $\varphi_{\gamma_\varepsilon,a}(\cdot)$ has support in $(0,1)$ for sufficiently small ε since $\gamma_\varepsilon \to 0$. Using the representation (9), we then have

$$|\mathcal{E}(h_{a,\varepsilon})| = \left|\frac{1}{m_1(\pi)}\int_0^1 \varphi_{\gamma_\varepsilon,a}(u)\beta(u)\,\mathrm{d}u\right| \lesssim m_1(\pi)^{-1} \sup_{u\in(0,1)} \beta(u)$$

since $\int_0^1 \varphi_{\gamma_\varepsilon,a}(u)\,\mathrm{d}u = \int_0^1 \varphi(u)\,\mathrm{d}u = 1$. Recall that $\sup_{u\in(0,1)} \beta(u) \lesssim 1$, by Lemma 2. By Theorem 2, we conclude that III^2 has order

$$\varepsilon^{2\mu\kappa_2/(\mu+1)(2\kappa_2+1)} \tag{35}$$

in probability. For *IV*, we first note that $m_1(\pi)\mathcal{E}(h_{a,\varepsilon}) = \int_0^1 \varphi_{\gamma_\varepsilon,a}(u)\beta(u)\,\mathrm{d}u$, hence

$$IV^2 = \left(\int_0^1 \varphi_{\gamma_\varepsilon,a}(u)\beta(u)\,\mathrm{d}u - \beta(a)\right)^2.$$

The following argument is classical in nonparametric estimation: since $\beta \in \Sigma(s)$ with $s = n + \{s\}$, where n is a non-negative integer, by a Taylor expansion up to order n (recall that the number N of vanishing moments of $\varphi(\cdot)$, recall (15), satisfies $N > s$), we obtain

$$IV^2 \lesssim \gamma_\varepsilon^{2s}; \tag{36}$$

see, for instance, Tsybakov [16], Proposition 1.2. Combining (34) and (36), we see that the balance term $\gamma_\varepsilon = \varepsilon^{\mu/(\mu+1)(2s+3)}$ yields the correct rate for *II* and *IV*. Next, the condition $\kappa_2 \geq s/3$ ensures that the term (35) also has the correct order. Finally, the estimate (33) proves asymptotically negligible, thanks to the assumption that $\sigma\varepsilon^{-3}$ is bounded and using the fact that $\kappa_1 \geq 4$, in the same way as for (30) in the proof of Theorem 2. The proof of Theorem 4 is thus complete.

Acknowledgements

The careful reading, comments and suggestions of two referees helped to improve considerably an earlier version of this paper. This work was supported in part by the Agence Nationale de la Recherche, Grant No. ANR-08-BLAN-0220-01.

References

[1] Aldous, D.J. (1999). Deterministic and stochastic models for coalescence (aggregation, coagulation): A review of the mean-field theory for probabilists. *Bernoulli* **5** 3–48. MR1673235

[2] Asmussen, S. (2003). *Applied Probability and Queues*, 2nd edition. New York: Springer. MR1978607

[3] Bertoin, J. (1996). *Lévy Processes*. Cambridge: Cambridge Univ. Press. MR1406564

[4] Bertoin, J. (2002). Self similar fragmentations. *Ann. Inst. H. Poincaré. Probab. Statist.* **38** 319–340. MR1899456

[5] Bertoin, J. (2006). *Random Fragmentation and Coagulation Processes*. Cambridge: Cambridge Univ. Press. MR2253162

[6] Bertoin, J. and Martinez, S. (2005). Fragmentation energy. *Adv. in Appl. Probab.* **37** 553–570. MR2144567

[7] Brennan, M.D. and Durett, R. (1987). Splitting intervals. II. Limit laws for lengths. *Probab. Theory Related Fields* **75** 109–127. MR0879556

[8] Filippov, A.F. (1961). On the distribution of the sizes of particles which undergo splitting. *Theory Probab. Appl.* **6** 275–293.

[9] Hille, E. and Phillips, R.S. (1957). *Functional Analysis and Semi-Groups*. *Amer. Math. Soc. Colloquium Publications* **31**. Providence, RI: Amer. Math. Soc. MR0089373

[10] Jagers, P. and Nerman, O. (1984). The growth and composition of branching populations. *Adv. in Appl. Probab.* **16** 221–259. MR0742953

[11] Kolmogorov, A.N. (1941). Über das logaritmisch normale Verteinlungsgesetz der Dimensionen des Teilchen bei Zerstückelung. *C. R. (Doklady) Acad. Sci. USSR* **31** 99–101. MR0004415

[12] Krapivsky, P.L. and Ben-Naim, E. (1994). Scaling and multiscaling in models of fragmentation. *Phys. Rev. E* **50** 3502–3507.

[13] Krapivsky, P.L., Ben-Naim, E. and Grosse, I. (2004). Stable distributions in stochastic fragmentation. *J. Phys. A* **37** 2863–2880. MR2042981

[14] Ney, P. (1981). A refinement of the coupling method in renewal theory. *Stochastic Process. Appl.* **11** 11–26. MR0608004

[15] Sgibnev, M.S. (2002). Stone's decomposition of the renewal measure via Banach-algebraic techniques. *Proc. Amer. Math. Soc.* **130** 2425–2430. MR1897469

[16] Tsybakov, A.B. (2009). *Introduction to Non-Parametric Estimation*. New York: Springer.

Received March 2008 and revised January 2010