Title
Understanding Probabilistic Models Through Limit Theorems

Author
Lin, Jeffrey Thomas

Publication Date
2017

Peer reviewed|Thesis/dissertation

University of California

Los Angeles

Understanding Probabilistic Models Through Limit Theorems

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Mathematics

by

Jeffrey Thomas Lin

2017

Abstract of the Dissertation

Understanding Probabilistic Models

Through Limit Theorems

by

Jeffrey Thomas Lin

Doctor of Philosophy in Mathematics

University of California, Los Angeles, 2017

Professor Marek Biskup, Chair

Limit theorems are ubiquitous in probability theory. The present work samples contributions of the author at the interface of this theory with three distinct fields: interacting particle systems, exchangeable random variables, and long-range percolation.

In the theory of interacting particle systems, one often studies the stationary distributions, which are obtained as limiting distributions of the process. We will discuss a proof concerning characterization of these measures in the case of an attractive nearest neighbor translation invariant spin system on the integers.

Exchangeable sequences of random variables are mixtures of i.i.d. sequences, and the probability measure that determines the relative proportions of this mixture can be obtained as a limit from the exchangeable sequence itself. We will analyze the possibility of reconstructing this probability measure from only partial information about the exchangeable sequence.

A goal in long-range percolation is to understand how chemical distance scales with Euclidean separation. We will show that the limiting scaling behavior for a certain class of models is polylogarithmic. This will be an improvement on existing results.

The dissertation of Jeffrey Thomas Lin is approved.

Inwon Kim

Georg Menz

Thomas Liggett

Marek Biskup, Committee Chair

University of California, Los Angeles

2017

To my parents,

whose encouragement and guidance

have made this endeavor possible.

Table of Contents

Acknowledgments

Chapter 2 is a version of [14].

J. Lin, *A Correction Note: Attractive Nearest Neighbor Spin Systems on the Integers.* (ArXiv:1507.00678) (2014).

Chapter 3 is a version of [15].

J. Lin, *A Negative Answer to a Problem of Aldous on Determination of Exchangeable Sequences.* Electron. J. Probab. Vol. 21 no. 30, 1-26 (2016).

Chapter 4 is a version of [4].

M. Biskup, J. Lin, *Scaling of the Chemical Distance in Long Range Percolation.* Work in Progress.

I'd like to acknowledge Marek Biskup, my adviser and collaborator on the project discussed in Chapter 4. His direction and advice was invaluable in my progress in all three projects. His generosity with his time and energy knew no bounds. I have also reprinted a small portion of his work, [2], nearly verbatim.

I am thankful for funding for this project coming from the NSF award DMS-1407558.

I'd also like to acknowledge the encouragement from my parents to pursue an advanced degree, and the support I received from my parents while facing the many academic challenges of graduate school. Throughout my childhood and teenage years, I was encouraged to pursue understanding of the world around me, and to never give up. This attitude was necessary to my success in graduate school.

Finally, I'd like to thank the people who changed my life in connection to my coming out as gay in 2013. Had I not been able to come out, the stress and isolation of living in the closet would have likely prevented the completion of this work.

Vita

2007	Attended UCSD for undergraduate studies until transferring to California Institute of Technology in 2009.
2009	Intern at Lawrence Livermore National Laboratory studying x-ray crystallography
2010	E.T. Bell Prize for best written paper for the Summer Undergraduate Research Fellowship program at California Institute of Technology
2011	B.S. Mathematics and B.S. Physics, California Institute of Technology
2011-2016	Teaching Assistant, Mathematics Department UCLA. Lead discussion sections for calculus, linear algebra, and proofs based classes. Mentored students in smaller groups during 2 hours of office hours per week. Prepared graduate students for the real analysis qualifying exam.
2016	M.A. Mathematics UCLA
2016	National Security Agency Graduate Mathematics Program Internship. Implemented programs combining C++ and Python in a Linux environment to automate image processing and classification. Used knowledge of probability and statistics in order to improve image processing. Leveraged machine learning techniques, such as random forest, to solve classification problems. Collaboratively streamlined software development using Git version control. Prepared findings for presentation to the Institute for Defense Analysis at Bowie Center for Computing Sciences. Briefed Congressional staffers on technical results. Submitted detailed findings in an internally refereed technical paper.

CHAPTER 1

Introduction

Limit theorems are paramount in probability theory, appearing in areas such as percolation, Markov processes, large deviations, statistical mechanics, ergodic theory, universality of KPZ, and many more. The present work is an investigation of limit theorems in three different areas of probability theory: interacting particle systems, exchangeable random variables, and long-range percolation. In this introduction chapter, we will summarize these problems, and the next three chapters will cover them in detail. The final chapter is the conclusion.

1.1 Interacting Particle Systems Problem

Interacting particle systems are Feller Processes that take place on the compact state space W^S where W is a compact metric space, and S is an at most countable set, which we think of as particles. The model we discuss below can be applied to study frequently occurring systems such as TASEP, physical situations where particles arranged in a lattice have interacting spins, or other areas.

For our purposes, we will take $W = \{0, 1\}$, leading to what is called a spin system. The rate of transition is generally described using functions $c(\eta, x) \geq 0$ where $x \in S$ and $\eta \in W^S$. We think of the function c as describing the rate of change at site x from $\eta(x)$ to $1 - \eta(x)$ given that all of the coordinates are known to be $\{\eta(y)\}_{y \in S}$. This allows for the rate of change at a given site to depend on any of the coordinates. Feller processes exhibiting these transition rates exist, as shown in [13]. (Chapter 1, Chapter 3) Essentially, one defines the pregenerator Ω with codomain $C(W^S)$, the space of continuous functions defined on the state space of our process. Ω is an unbounded operator on a subspace of $C(W^S)$ given by the formula

$$\Omega f(\eta) = \sum_{x \in S} c(x,\eta)[f(\eta_x) - f(\eta)] \tag{1.1}$$

where $\eta_x = \eta$ on $S \setminus \{x\}$ and $\eta_x(x) = 1 - \eta(x)$. The pregenerator definition (1.1) applies to a dense set of f, under supremum norm, for which this series converges. Then one takes the closure of the graph of the pregenerator, and proves that this closure is a generator for a Feller process. All Feller processes can be represented by a generator. (see [13] Chapter 1)

There is a third way to represent Feller processes other than as Markov processes and via a generator. One can also use a semigroup.

Definition 1.1. *Let $T_t : C(W^S) \to C(W^S)$, $t \in [0,\infty)$ where each T_t is a bounded linear operator. If the family T_t obeys*

1. $T_0 = I$
2. $T_{t+s} = T_t \circ T_s$
3. *for each* $f \in C(W^S), ||T_t f - f||_\infty \to 0$ *as* $t \to 0$.
4. $T_t 1 = 1$
5. $T_t f \geq 0$ *whenever* $f \geq 0$.

then T_t is a Feller Semigroup.

Every Feller process has a unique Feller semigroup associated with it. ([13] Chapter 1 Section 1)

Let μ be a probability measure defined on W^S. Let us denote by μT_t the probability measure, also defined on W^S, given by

$$\int_{W^S} f d\mu T_t = \int_{W^S} T_t f d\mu \tag{1.2}$$

whenever $f \in C(W^S)$. Thus, μ is evolved under the semigroup for t time to become μT_t. If $t \geq 0 \Rightarrow \mu T_t = \mu$, then μ is called stationary.

Often in the theory of Markov Processes, one is concerned with the evolution of the process towards a stationary distribution. By Tychonoff's theorem, the state space of our spin system is compact, hence one can show that there always exists a stationary distribution.

Definition 1.2. *A spin system with Feller Semigroup T_t is called ergodic if there is exactly one stationary distribution ν for the process, and moreover for any μ a probability measure on W^S, we have $\mu T_t \to \nu$ weakly.*

In our case, we will assume a condition known as attractiveness ([13] p. 134).

Definition 1.3. *Given a spin system with rates $c(x, \eta)$ we call it attractive if whenever $\eta \leq \zeta$, we have $\forall x \in S$*

$$\eta(x) = \zeta(x) = 0 \quad \Rightarrow \quad c(x, \eta) \leq c(x, \zeta) \tag{1.3}$$

and

$$\eta(x) = \zeta(x) = 1 \quad \Rightarrow \quad c(x, \eta) \geq c(x, \zeta) \tag{1.4}$$

Intuitively, a spin system is attractive if flips at x that would cause the spin at x to agree with its environment are more likely to occur than flips at x that would cause disagreement with the environment.

Much is known about the theory of attractive spin systems. The attractiveness allows comparison of distributions to play a role. We say that two probability measures μ and ν have $\mu \leq \nu$ if all $f \in C(W^S)$ that are monotonic in each coordinate obey $\int f d\mu \leq \int f d\nu$. Attractiveness means that this (stochastic) ordering of measures is preserved by evolution of the system.

There is a maximum and minimum stationary distribution for any attractive spin system. (see [13] p.135) Sometimes they are called maximal and minimal respectively, because stochastic ordering is not a total linear order, but they are comparable to all other stationary distributions. We will adopt the notation of calling the maximal stationary distribution $\overline{\nu}$

and the minimal stationary distribution $\underline{\nu}$. These can be obtained by starting the system a.s. with either all 1 states or all 0 states respectively. Ergodicity for an attractive spin system is equivalent to $\overline{\nu} = \underline{\nu}$. ([13] p.136)

In our present problem, we will restrict not only to attractive spin systems, but even those with site set $S = \mathbb{Z}$ and assume that the rate of change at site $x \in \mathbb{Z}$ depends only on the current state of its two nearest neighbors. Moreover, we assume the rates are translation invariant. This means that $\forall y \in \mathbb{Z}$, $c(x-y, \{\eta(z-y)\}_{z\in\mathbb{Z}}) = c(x,\eta)$.

The main result of this work in the area of interacting particle systems is the correction of the proof for the following theorem, [12] and [13] (p. 152 Theorem 3.13 in [13]):

Theorem 1.4. *Suppose an attractive translation invariant nearest neighbor spin system on the integers satisfies $c(x,\eta) + c(x,\eta_x) > 0$ whenever $\eta(x-1) \neq \eta(x+1)$. Then for that spin system, the extremal invariant measures are precisely $\overline{\nu}, \underline{\nu}$*

1.2 Exchangeability Problem

The second subject discussed in this dissertation concerns exchangeable random variables.

Definition 1.5. *Let X_j be a sequence of real-valued random variables defined on the same probability space for $j \geq 1$. Call $X = (X_j)_{j\in\mathbb{N}}$ **exchangeable** if whenever $\pi : \mathbb{N} \to \mathbb{N}$ is a finite permutation, we have that*

$$(X_j)_{j\in\mathbb{N}} =_d (X_{\pi(j)})_{j\in\mathbb{N}}. \tag{1.5}$$

Exchangeable sequences are ubiquitous in mathematics. For instance, the sequence of draws from Pólya's urn is an exchangeable sequence, and exchangeability shows up in interacting particle systems. If X is a random variable, then the sequence $X, X, X, \ldots$ is exchangeable.

Exchangeable sequences are generalizations of sequences of iid random variables. In fact, de Finetti's theorem, which we review below, spells out exactly what kind of generalization.

De Finetti's theorem, and this discussion, applies only to infinite exchangeable sequences. (i.e. $X = (X_1, X_2, \dots)$)

In preparation for de Finetti's theorem, we need some definitions.

Definition 1.6. *Suppose Θ is a Borel probability measure on $P(\mathbb{R})$, the space of Borel probability measures on the reals, with vague topology. We call Θ a mixing measure. Say that X is **iid-**Θ if its distribution is given by*

$$\forall A \in \mathit{Borel}(\mathbb{R}^{\mathbb{N}}): \quad P(X \in A) = \int \theta^{\mathbb{N}}(A)\, d\Theta(\theta). \tag{1.6}$$

We also need

Definition 1.7. *Say that X is a **mixture of iids directed by** α, where α is a random Borel probability measure defined on the same probability space as X, if $\alpha^{\mathbb{N}}$ is a regular conditional distribution for X given $\sigma(\alpha)$.*

Now we state de Finetti's theorem.

Theorem 1.8 (de Finetti)**.** *The following are equivalent:*

1. *X is exchangeable*
2. *X is a mixture of iids directed by some α*
3. *X is iid-Θ for some Θ.*

Moreover, each X corresponds to a unique Θ and, up to distributional equality, a unique α in this way. Θ is the distribution of α. Exchangeable sequences that have the same distribution will have the same Θ, and every possible Θ describes an exchangeable sequence.

So iid sequences are exchangeable sequences with Θ a point mass. In general, the joint distribution of an exchangeable sequence is obtained by taking a weighted-integral-average of iid sequences. iid sequences themselves are obtained by only including one such a sequence in this weighted-integral-average.

In a sense, de Finetti's theorem can be regarded as saying that we can reconstruct the mixing measure, Θ from joint distributional information of the exchangeable sequence. We could ask if marginal information suffices.

The following question was posed by David Aldous in [1] (p. 20):

Question 1.9. *Let* $X = (X_j)_{j\in\mathbb{N}}$, $Y = (Y_j)_{j\in\mathbb{N}}$ *be exchangeable sequences of* $\mathbb{R}$*-valued random variables, with* $S_n = \sum_{j=1}^n X_j, T_n = \sum_{j=1}^n Y_j$*. Suppose that* $\forall n \geq 1$ *we have* $S_n =_d T_n$*. Does it follow that* $X =_d Y$*?*

The motivation for considering this question is reconstruction. It is well known that one cannot deduce the coupling of random variables just from knowing the marginals. We could ask if the situation changes if, in addition, we know a priori that the underlying joint distribution is exchangeable. However, to make such a question interesting, we should not consider the marginal distributions of the exchangeable sequence X itself, since they are all the same, but rather those of its partial sums. Since X is exchangeable, the distribution of a sum of distinct components of X depends only on the number of summands. Thus, the question can be seen as asking if the joint distribution of an exchangeable sequence can be deduced from the maximum amount of marginal information.

An answer to this question was known before the author's work. In [7], Evans and Zhou showed, via Fourier analytic techniques, that there exist two exchangeable sequences X and Y which have all the same marginals of their partial sums, but not the same joint distribution. However, by inspection of this method, one finds that X and Y are signed random variables. Since it is often true that an a priori nonnegative assumption helps reconstruction, it was natural to ask if sign matters in the present problem:

Question 1.10. *Let* $X = (X_j)_{j\in\mathbb{N}}$, $Y = (Y_j)_{j\in\mathbb{N}}$ *be exchangeable sequences of* $[0,\infty)$*-valued random variables, with* $S_n = \sum_{j=1}^n X_j, T_n = \sum_{j=1}^n Y_j$*. Suppose that* $\forall n \geq 1$ *we have* $S_n =_d T_n$*. Does it follow that* $X =_d Y$*?*

This question was also quite reasonable in light of the partial positive result in [7]. There, it was shown that the answer is affirmative with the additional assumption that the exchangeable sequences are mixtures of countably many nonnegative iid sequences. (i.e. their Θ's are concentrated on countable sets within $P([0,\infty))$.)

The present work proves that the answer to Question 1.10 is a resounding “no” by finding even a counterexample consisting of two exchangeable sequences X and Y which take values in the set $\{0, 1, 2, 3\}$.

Let us discuss the idea of our proof. We know that the joint distribution of a d-dimensional random vector can be recovered when the one-dimensional projections are known for all vectors in $\mathbb{R}^d$. One could consider a different, though related reconstruction problem of trying to recover the joint distribution of a d-dimensional random vector from knowing its one-dimensional projections along merely a subset of deterministic vectors in $\mathbb{R}^d$. We use the representation of an exchangeable sequence as an integral-combination of i.i.d distributions in order to reduce Question 1.10 to the above d-dimensional question for a specific subset. To solve the reduced question, we require an important result from the work of Cuesta-Albertos, Fraiman, and Ransford in [6]. In that work, the Paley-Wiener theorem [16] is used to produce two distinct probability measures with compact support that have the same one-dimensional projections along certain subsets. The resulting two measures are then appropriately transformed to suit our needs.

1.3 Long Range Percolation

The third and last subject discussed in this dissertation is long range percolation.

Let G be a graph. Percolation is the study of random subgraphs of G obtained by deleting some of the edges of G, but keeping all the vertices. The edges are deleted at random, and independently of one another. In our present model, the graph G is the complete graph on $\mathbb{Z}^d$ (hence ”long range”) and edges are kept with probability proportional to $\frac{1}{|x-y|^s}$ where x, y are the endpoints of the edge in question, and $s \in (d, 2d)$. The one exception to this rule is that nearest neighbor edges are present (kept) almost surely. In percolation theory in general, one is interested in the geometric properties of the random subgraph. We will focus on the chemical distance, which is defined as the graph distance between two points within the random subgraph. Because we always keep all nearest neighbor edges and our random graph is thus connected, the chemical distance is always finite.

A brief explanation of the restriction on s follows. Heuristically speaking, for large s, the model is much like nearest neighbor percolation. For small s values, the chemical distance function is bounded. Thus, the interesting behavior occurs in between the two extremes, which turns out to mean between d and $2d$. The critical cases $s = 2d$ and $s = d$ are of interest as well, but not addressed in this work. For readers interested in more details on other s values, see [3] for a summary.

In [2] and [3], Biskup establishes that

$$\lim_{L\to\infty} P(\log(L)^{\Delta-\epsilon} \leq D_{\text{dis}}(0, Le_1) \leq \log(L)^{\Delta+\epsilon}) = 1 \tag{1.7}$$

where D_{dis} is the chemical distance, e_1 is the first standard basis vector for $\mathbb{R}^d$, and the above holds for each $\epsilon > 0$. We will improve this to

Theorem 1.11. *There are $c, C > 0$ such that*

$$\lim_{L\to\infty} P(c * \log(L)^{\Delta} \leq D_{dis}(0, Le_1) \leq C * \log(L)^{\Delta}) = 1. \tag{1.8}$$

The strategy of proof in [2] and [3] relies on the observation that there is, with overwhelming likelihood, an edge $e = \langle x, y\rangle$ with x in a small neighborhood of 0 and y in a small neighborhood of Le_1. Then there is, with overwhelming likelihood, a pair of edges connecting a neighborhood of 0 to a neighborhood of x and a neighborhood of y to a neighborhood of Le_1. This iteration is continued until the probability of finding all the edges cannot be maintained close enough to 1. At each step, the size of the neighborhoods is microscopic compared to the previous scale, and the number of edges doubles. An appropriate choice of the size of the neighborhoods establishes the upper bound portion of (1.7), while the lower bound is established by an iterative method that effectively argues that optimizing paths have roughly an hierarchical form like the above.

This method does not appear to the author to be sufficient for the improvement we claim. Instead, the present work will derive an iteration inequality from similar considerations, and use it to obtain first and second moment estimates that force not only the bounds we claim, but also even a sort of scaling limit for a restricted notion of the distance.

This part of the dissertation, on long range percolation, is joint with Marek Biskup. A small portion of the corresponding text in the present dissertation is a nearly verbatim copy of our private communications.

CHAPTER 2

Interacting Particle Systems

2.1 Introduction

This section is copied nearly verbatim from [14], with only formatting and minor error corrections.

For the purposes of this note, an attractive translation invariant nearest neighbor spin system is a certain kind of Feller process defined on the compact state space $X = \{0,1\}^{\mathbb{Z}}$ with rates satisfying the attractiveness inequalities, and depending only on the nearest neighbors. (See [13], p.144-145 for more details.) Notice that such spin systems have rates entirely determined by $8 = 2^3$ parameters corresponding to the rates dependent on a given site and its two nearest neighbors. More concretely, the parameters are c_{ijk} where $i, j, k \in \{0,1\}$. For instance, c_{010} means the flip at any site x when the present spin of particle x is 1 and its two nearest neighbors both have spin 0. We will use η_x to denote the configuration that agrees with η at all sites other than x, and has the opposite spin at x. The notion of stochastic ordering, minimality, and maximality is discussed in the previous chapter.

T. Liggett proves in [12] and [13], (p. 152 Theorem 3.13) the following theorem:

Theorem 2.1. *Suppose an attractive translation invariant nearest neighbor spin system on the integers satisfies* $c(x,\eta) + c(x,\eta_x) > 0$ *whenever* $\eta(x-1) \neq \eta(x+1)$. *Then this spin system has only the minimal and maximal invariant measures (ordered stochastically) as the extremal invariant measures.*

We discuss the estimate in [13] of lemma 3.7 part (e). It is wrong, but the similar estimate $\epsilon \int g_{m,n}^{l+1} d\nu \leq (4Kl + 2\epsilon) \int g_{m,n}^{l} d\nu$ is valid. This change does not affect the proof moving forward, because the only time the estimate is used is in the proof of lemma 3.10 in [13], where it is used to derive that $\sup_{m\leq n}\int g^l_{m,n}d\nu < \infty$. This fact still holds with the new estimate. The same line of reasoning appears in [12] within the proof of lemma 2.2 there.

We will refer to [13] from now on.

2.2 A Correction

The problem with the estimate as written is at the top of p.150 of [13] in the discussion of bounding below the positive contribution to $\tilde{\Omega}g^l_{m,n}$. The argument there fails to consider the possibility that change in the γ coordinate at any of the x_i among the left and right endpoints in the $l+1$ length intervals may not only create a length l interval, but also destroy an adjacent length l interval. But this can only happen at at most $2g^l_{m,n}$ such sites, i.e. the left and right neighbors of intervals of length l. The bound below on the rate of the type of flip in question is still correctly stated as ϵ so as long as we replace $\epsilon(g^{l+1}_{m,n})$ by $\epsilon(g^{l+1}_{m,n} - 2g^l_{m,n})$ the estimate is correct. This results in the display reading

$$\tilde{\Omega}g^l_{m,n} \geq \epsilon(g^{l+1}_{m,n} - 2g^l_{m,n}) - 4Klg^l_{m,n}.$$

What has been shown is the statement that attractive translation invariant nearest neighbor spin systems with enough nonzero transition rates have the smallest possible set of extremal invariant measures.

CHAPTER 3

Exchangeable Random Variables

3.1 Introduction

This section is copied nearly verbatim from [15], with only formatting and minor error corrections.

In probability theory, there are many results concerning uniqueness of a distribution satisfying certain properties. For instance, there are the various moment problems, the inversion of the characteristic function, the inversion of the Laplace transform. Another kind of uniqueness result relates to exchangeable sequences of random variables. Given the joint distribution of an exchangeable vector $X = (X_n)_{n\in\mathbb{N}}$, it can be written as a "mixture" of iids in exactly one way.

We will prove altered versions of these types of results. Roughly speaking, we will assume only partial information, and make regularity assumptions to ensure that the resulting problem is well-defined. In the process, we show a trend of what types of obstructions there can be to such uniqueness results: arithmetic and algebraic structure. (These two notions are related because arithmetic relationships between exponents, say in a Laplace transform, will correspond to polynomial relationships between the exponentials, for instance the Laplace transforms.)

In [1], (p. 20) Aldous proposes the following question.

Question 3.1. *Let $X = (X_j)_{j\in\mathbb{N}}$, $Y = (Y_j)_{j\in\mathbb{N}}$ be exchangeable sequences of $\mathbb{R}$-valued random variables, with $S_n = \sum_{j=1}^{n} X_j, T_n = \sum_{j=1}^{n} Y_j$. Suppose that $\forall n \geq 1$ we have $S_n =_d T_n$. Does it follow that $X =_d Y$?*

Despite this question being over 30 years old, only partial progress has been made on it. In [7] Section 2, Evans and Zhou show the answer is negative in the class of signed random variables. Therefore, interest has somewhat shifted to the nonnegative case. In [7], it is shown that there is an affirmative answer if we know that X and Y are mixtures of countably many nonnegative iid sequences. We will show that the nonnegative restriction does not improve the situation in the absence of the additional assumption of [7]: X cannot be determined uniquely from this partial information even if $X_j \geq 0$. However, as we will see, uniqueness holds when the exchangeable sequences are mixtures of convex combinations of up to 3 iid sequences.

Uniqueness is also true for more complicated mixtures, so long as the iid sequences involved in the mixture are related "transcendentally". (It is the main purpose of this chapter to present results which make this theme precise, in the context of all uniqueness questions we will consider.) We may think of the heuristic, "mixtures of a small number of distributions exhibit uniqueness of X", as a special case of the heuristic about arithmetic and algebraic (polynomial) dependence. For instance, if the collection of allowed distributions in the mixture is small, then there cannot be too many arithmetic dependencies.

The first question naturally leads to the following continuous time analog, suggested by J. Černý.

Question 3.2. *Let S_t and T_t be mixtures of Lévy Processes. Suppose that $\forall t \geq 0$, $S_t =_d T_t$. Is $(S_t)_{t\geq 0}$ jointly equidistributed with $(T_t)_{t\geq 0}$?*

We will see that in the case of restricting to mixtures of Brownian Motions, the answer is yes. However, we will find a (possibly unexpected) parallel between a measure similar to the Levy measure implied in this problem and the mixing measure implied in Question 3.1. In particular, we will find that upon restricting to even just Poisson Processes, uniqueness fails.

For a Lévy Process, knowing the marginal at any time $t \neq 0$, such as $t = 1$, tells us the entire joint distribution of the process. However, this will not be true for a mixture of

Lévy Processes. We will see that in some cases, we could say S_t is unique just from making observations at $t \in \mathbb{N}$ and other times we require all $t \geq 0$. In the former case, the continuous time problem makes uniqueness of S_t more plausible than the discrete time case only because continuous time imparts infinite divisibility which limits the possible distributions that can occur in the mixture. In the latter case, the continuous time problems will tend to exhibit uniqueness over their discrete time counterparts not only because of infinite divisibility, but because observation at noninteger t's provide more information beyond that.

An example of where the fact that we make continuum observations makes a difference, at least in the argument, is the case of a mixture of normal distributions or Brownian motions. In this case, the continuum of observations destroys the arithmetic structure, making uniqueness a fact in the Brownian Motion case, as opposed to an open problem in the discrete time case.

In light of these questions, it is natural also to consider various random analogs of uniqueness problems from the classical theory of random variables. For instance, one such a question is posed below:

Question 3.3. *Let α, β be random Borel measures on $\mathbb{R}$. Suppose that their moments (which are random variables) $\int_{\mathbb{R}} x^n d\alpha$ and $\int_{\mathbb{R}} x^n d\beta$ are well-defined and equal in marginal distribution (i.e. for one n at a time). Do we have $\alpha =_d \beta$?*

Question 3.3 can be viewed as a randomization of classical uniqueness theorems, in the sense that the probability measure one tries to recover is deterministic in the classical theories, and we now take it to be random.

We will see that in all randomized versions of classical uniqueness problems we consider, knowing joint information is enough to assert uniqueness of the original random measure in distribution, but knowing marginal information is not enough. We will see that the existence of counterexamples again depends on the arithmetic and algebraic structure.

The layout of the chapter is as follows. We will treat Question 3.1 in Section 2, Question 3.2 in Section 3, and the questions similar to Question 3.3 in Section 4. Section 3 is fairly technical, but Section 4 does not depend on Section 3, so Section 3 could be omitted on a first pass. In each of the sections, we will first discuss the results required to make the questions precise and the definitions associated to the corresponding question. We also discuss the machinery to be used in the rest of the section. In subsections, we consider answers to the questions in various cases.

3.2 Discrete Time Exchangeability Problem

In this section, we consider Question 3.1 posed by Aldous. We keep the notation used in the statement of the question, so that $X = (X_j)$ and $Y = (Y_j)$ will denote sequences of exchangeable random variables.

First we state the following without proof:

Lemma 3.4 (Classical Bounded Moment Problem). *Let $V = (V_\gamma)$ and $Z = (Z_\gamma)$ be vectors indexed by the same set Γ of arbitrary cardinality. Thus V and Z take values in $\mathbb{R}^\Gamma$. Assume that for each $\gamma \in \Gamma$, V_γ, Z_γ are bounded real random variables. Then V and Z have the same distribution if and only if they have the same joint moments, i.e. $\forall\{\gamma_1, \ldots, \gamma_l\} \subset \Gamma$ $\forall r_1, \ldots, r_l \in \mathbb{N}$*

$$\mathbb{E}\left(\prod_{i=1}^{l} V_{\gamma_i}^{r_i}\right) = \mathbb{E}\left(\prod_{i=1}^{l} Z_{\gamma_i}^{r_i}\right) \tag{3.1}$$

.

We will have need throughout the chapter for the notion of a random measure. In general, we will use $P(\mathfrak{M})$ to denote the (Borel) probability measures defined on a standard Borel space $\mathfrak{M}$. Similarly, we define $M_+(\mathfrak{M})$ to be the collection of finite nonnegative Borel measures on $\mathfrak{M}$. In each case, give these spaces the measurable structure generated by the mapping $\mu \mapsto \mu(B), B \in \text{Borel}(\mathfrak{M})$. Since $\mathfrak{M}$ is standard Borel, so is $P(\mathfrak{M})$ with the topology given by vague convergence. The Borel σ algebra determined by vague convergence is an equivalent definition of the measurable structure given to $P(\mathfrak{M})$. (For these facts, see for instance [11].) A random measure is then a random variable taking values in either $M_+(\mathfrak{M})$

or $P(\mathfrak{M})$. We will always deal with Polish spaces when the topology is significant, and standard Borel spaces if not. We almost always deal with probability measures, i.e. except when dealing with Lévy measures.

From now on we will use

$$\forall s \geq 0,\, \mathcal{L}_\mu(s) = \int_{[0,\infty)} e^{-sx} d\mu(x) \tag{3.2}$$

to mean the Laplace transform (at $s \geq 0$) of a probability measure μ on $[0, \infty)$. Let $P^+ = \{\mu \in P(\mathbb{R}) | \mu \text{ is supported in } [0, \infty)\}$, which is closed in $P(\mathbb{R})$. If μ is random, then the Laplace transform will simply be a random variable.

In this section, we will use symbols $\alpha, \alpha_1, \alpha_2$ to denote random probability measures and $\Theta, \Theta_1, \Theta_2$ to denote elements of $P(P(\mathbb{R}))$. Symbols X_j, Y_j will denote real-valued random variables.

Definition 3.5. *For each $j \geq 1$, let X_j be a real-valued random variable. Call $X = (X_j)_{j \in \mathbb{N}}$* ***exchangeable*** *if whenever $\pi : \mathbb{N} \to \mathbb{N}$ is a finite permutation, we have that*

$$(X_j)_{j\in\mathbb{N}} =_d (X_{\pi(j)})_{j\in\mathbb{N}}. \tag{3.3}$$

Definition 3.6. *Say that X is* ***iid**-Θ if its distribution is given by*

$$\forall A \in Borel(\mathbb{R}^{\mathbb{N}}),\, P(X \in A) = \int \theta^{\mathbb{N}}(A)\, d\Theta(\theta) \tag{3.4}$$

.

Definition 3.7. *Say that X is a* ***mixture of iids directed by*** *α, where α is defined on the same probability space as X, if $\alpha^{\mathbb{N}}$ is a regular conditional distribution for X given $\sigma(\alpha)$.*

By definition of regular conditional distribution this is just the same as saying that whenever $A_1, \ldots, A_n \in Borel(\mathbb{R})$ we have for a probability 1 set of ω,

$$P(X_i \in A_i, 1 \leq i \leq n | \sigma(\alpha))(\omega) = \prod_i \alpha(\omega, A_i). \tag{3.5}$$

We now state the well-known theorem of de Finetti.

Theorem 3.8 (de Finetti). *The following are equivalent:*

1. *X is exchangeable*
2. *X is a mixture of iids directed by some α*
3. *X is iid-Θ for some Θ.*

Remark 3.9. *For each exchangeable law μ on $\mathbb{R}^{\mathbb{N}}$ there exists a unique law Θ on $P(\mathbb{R})$ for which any X with distribution μ is iid-Θ. This assignment $\mu \mapsto \Theta$ is a bijection from the exchangeable laws on $\mathbb{R}^{\mathbb{N}}$ to $P(P(\mathbb{R}))$. In fact, it is a homeomorphism. Also, if X is exchangeable then its directing measure is unique up to a.s. equality, and the distribution of the RV α is Θ. Therefore, if the (joint) distribution of X is determined, then the law of α is determined.*

In the literature, the phrase "mixing measure" refers to both Θ and α. We will use the terminology for Θ and call α the directing measure. In contexts when we have two exchangeable sequences X and Y, we denote by Θ_1, α_1 the mixing measure, directing measure (respectively) for X and Θ_2, α_2 the mixing measure, directing measure (respectively) for Y. Symbols $\alpha_1, \alpha_2, \alpha$ will always refer directing measures, not arbitrary random measures.

We set $S_n = \sum_{j=1}^{n} X_j$ and $T_n = \sum_{j=1}^{n} Y_j$. Our question is then whether $\forall n$, $S_n =_d T_n$ implies $\Theta_1 = \Theta_2$ (or $\alpha_1 =_d \alpha_2$).

Towards this, let us define:

Definition 3.10. *Let $L \subset P(\mathbb{R})$ be measurable. We will say that L is* ***mixture-restricting*** *provided that whenever Θ_1, Θ_2 are concentrated on L and $\forall n$, $S_n =_d T_n$, we have $\Theta_1 = \Theta_2$.*

To determine if a class L is mixture-restricting or not, we may use the following

Lemma 3.11. *A measurable class $L \subset P^+$ is mixture-restricting if and only if $\forall \alpha_1, \alpha_2$ a.s. L-valued, we have that*

$$\forall s \geq 0,\ \mathcal{L}_{\alpha_1}(s) =_d \mathcal{L}_{\alpha_2}(s) \tag{3.6}$$

implies $\Theta_1 = \Theta_2$.

Proof. We have that the statement $\forall n,\, S_n =_d T_n$ is equivalent to the statement

$$\forall n, s \geq 0,\, \mathbb{E}[e^{-sS_n}] = \mathbb{E}[e^{-sT_n}], \tag{3.7}$$

which, using the definition of a directing measure, is equivalent to

$$\forall n, s \geq 0,\, \mathbb{E}\left[\left(\int_{[0,\infty)} e^{-sx} d\alpha_1(x)\right)^n\right] = \mathbb{E}\left[\left(\int_{[0,\infty)} e^{-sx} d\alpha_2(x)\right)^n\right]. \tag{3.8}$$

By the classical bounded moment problem, this is equivalent to

$$\forall s \geq 0,\, \mathcal{L}_{\alpha_1}(s) =_d \mathcal{L}_{\alpha_2}(s). \tag{3.9}$$

□

We do not assert the analogous statement for MGFs (moment generating functions) because we have no need for it even though it is true under suitable boundedness hypotheses. We do not assert a statement for the characteristic function case because the characteristic function would be a complex-valued random variable, and one would need the joint distribution of the real and imaginary parts. The absence of an analogous lemma for characteristic functions is related to the absence of a bounded moment problem for complex-valued RVs that does not involve knowing any conjugate moments. This can be thought of as a heuristic reason for the shift in focus away from signed random variables in the recovery problems we consider in this chapter.

3.2.1 Known Results

In this section, we will discuss results known beforehand. Using the conditional SLLN, it follows that

Proposition 3.12. $P(\{0,1\})$ *is mixture-restricting.*

In Section 2 of [7], it is shown that

Proposition 3.13. *$P(\mathbb{R})$ is not mixture-restricting.*

In Section 3 of the same paper, it is shown that a positive result can still be salvaged.

Proposition 3.14. *If $\forall n$, $S_n =_d T_n$, and Θ_1, Θ_2 are purely atomic, concentrated in P^+ then $X =_d Y$. In other words, any countable subset of P^+ is mixture-restricting.*

Interestingly enough, Muntz's Theorem, which arises in Lemma 3.3 of [7], can be replaced by the theorem of complex analysis stating that holomorphic functions defined on a connected open set agree just as soon as they agree on a set of points that accumulates within the domain.

Their assumption is of a different nature than what we will consider; they restrict the size of the sets where Θ_1, Θ_2 are concentrated rather than requiring, as we will, that Θ_1, Θ_2 are concentrated on a nice set of distributions that may be a continuum. Considering that the general problem has been solved in the negative, we will primarily address restrictions of the latter type, hence the terminology of such a set L being "mixture-restricting". A particularly interesting case is when L is contained in P^+.

For example, let $\mathcal{P}$ denote the collection of distributions in $P(\mathbb{R})$ which are the distribution probability measures for Poisson RVs. (So $\mu \in \mathcal{P}$ should satisfy $\forall k \geq 0$, $\mu(\{k\}) = e^{-\lambda}\lambda^k/k!$ for some $\lambda \geq 0$.) Similarly, define $\mathcal{E}$ for exponentials, $\mathcal{G}$ for geometrics, and $\mathcal{B}_n$ for binomials with parameter n fixed and $p \in [0, 1]$ varying. It is an exercise in analysis to see that:

Remark 3.15. *$\mathcal{B}_n, \mathcal{G}, \mathcal{E}, \mathcal{P}$ are all mixture-restricting.*

3.2.2 The Relationship Between the Aldous Question and Arithmetic Properties of the Value Set

In this subsection, we will present one substantiation of the heuristic that the more arithmetic dependencies there are in the allowed value set for X_1, the harder it is to recover X uniquely. Fix $A = \{a_1, \ldots, a_N\}$ when $N \in \mathbb{N}$ or $A = \{a_1, \ldots, \}$ when $N = \infty$. Assume, for technical reasons, that A has no accumulation points in $\mathbb{R}$. We allow negative values in A. Define $F_A \subset P(\mathbb{R})$ to be those distributions that are supported in A. (We can say "supported" instead of "concentrated", because A is closed. F_A is closed because A is.) We set $L := F_A$. For X exchangeable, it is equivalent to say that $X_1 \in A$ a.s. or to say $\alpha \in L$ a.s. or to say Θ is supported on L.

Theorem 3.16. *F_A is mixture-restricting if A is linearly independent over $\mathbb{Q}$.*

Proof. It suffices to handle the case in which A is discrete, infinite, i.e. $N = \infty$, because subsets of mixture-restricting sets are always mixture-restricting. We have X, Y exchangeable sequences, which therefore comes with directing measures and mixing measures by Remark 3.9. Since we are trying to prove the mixture-restricting property, we assume $\forall n \geq 0$, $S_n =_d T_n$. (In future proofs of the mixture-restricting property, we not explicitly mention this.) We need to see that $\alpha_1 =_d \alpha_2$. It suffices to see that $(\alpha_1(\{a_j\}))_{j\in\mathbb{N}} =_d (\alpha_2(\{a_j\}))_{j\in\mathbb{N}}$

Suppose that $M \in \mathbb{N}$. Then by using the linear independence hypothesis for our value set and the properties of directing measures, we have that $\forall r_1, \ldots, r_M$, with $s = \sum_{j=1}^{M} r_i$,

$$\binom{s}{r_1, r_2, \ldots, r_M}\mathbb{E}(\prod_{i=1}^{M}(\alpha_1\{a_i\})^{r_i}) = \mathbb{P}(S_s = \sum_{i=1}^{s} r_i a_i)$$
$$= \mathbb{P}(T_s = \sum_{i=1}^{s} r_i a_i) = \binom{s}{r_1, r_2, \ldots, r_M}\mathbb{E}(\prod_{i=1}^{M}(\alpha_2\{a_i\})^{r_i}). \quad (3.10)$$

Here, the first and third equality hold because the linear independence hypothesis guarantees that the only way that a sum of elements from A can be $\sum_{i=1}^{s} r_i a_i$ is if r_i copies of a_i are used for each $i \leq M$.

Now Lemma 3.4 completes the proof. □

It is also acceptable to have a few arithmetic dependencies:

Proposition 3.17. *$F_{\{0,1,2\}}$ is mixture-restricting.*

More generally, let $\mu_1, \mu_2, \mu_3 \in P^+$ and let $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3$ denote their Laplace transforms respectively. Let

$$L := \{a_1\mu_1 + a_2\mu_2 + a_3\mu_3 | a_1 + a_2 + a_3 = 1, a_1, a_2, a_3 \geq 0\}. \tag{3.11}$$

Then L is mixture-restricting.

Proof. We prove the more general claim. We may assume that no strict subset of $\{\mu_1, \mu_2, \mu_3\}$ has convex hull containing all of μ_1, μ_2, μ_3 because subsets of mixture-restricting sets are mixture-restricting. From this, it follows that L is homeomorphic to $T_3 := \{(a, b, c)|a+b+c = 1, a, b, c \geq 0\}$ because the map taking $(a, b, c) \mapsto a\mu_1 + b\mu_2 + c\mu_3$ is invertible, and therefore because T_3 and L are both compact Hausdorff, they are homeomorphic via this assignment, hence measurably isomorphic.

Proceeding by way of Lemma 3.11, and using the homeomorphism above, we are reduced to showing that if U, V are T_3-valued random vectors for which $\forall s \geq 0$ we have

$$(\mathcal{L}_1(s), \mathcal{L}_2(s), \mathcal{L}_3(s)) \cdot U =_d (\mathcal{L}_1(s), \mathcal{L}_2(s), \mathcal{L}_3(s)) \cdot V \tag{3.12}$$

then $U =_d V$. That this is enough follows from the fact that when U, V are the pushforwards of α_1, α_2 via the homeomorphism above, the left side has the distribution of the (random) Laplace transform of α_1 evaluated at s, and the right side has the distribution of the (random) Laplace transform of α_2 evaluated at s.

Because U, V are probability vectors, it suffices to show that if U', V' are bounded random vectors in $\mathbb{R}^2$ with

$$(\mathcal{L}_1(s) - \mathcal{L}_3(s), \mathcal{L}_2(s) - \mathcal{L}_3(s)) \cdot U' =_d (\mathcal{L}_1(s) - \mathcal{L}_3(s), \mathcal{L}_2(s) - \mathcal{L}_3(s)) \cdot V' \tag{3.13}$$

then $U' =_d V'$.

We will check that the collection of v's of the form

$$v = c(\mathcal{L}_1(s) - \mathcal{L}_3(s), \mathcal{L}_2(s) - \mathcal{L}_3(s)) \tag{3.14}$$

cover a nonempty open set as $s \geq 0$ and $c \in \mathbb{R}$ vary. We know that $\mu_1 - \mu_3$ is a signed measure on $[0, \infty)$ and as is $\mu_2 - \mu_3$. We may regard $\mathcal{L}_1(s) - \mathcal{L}_3(s)$ and $\mathcal{L}_2(s) - \mathcal{L}_3(s)$ as the

Laplace transforms of these signed measures respectively. Therefore, a nonempty open set of v's are of the form (3.14). For instance, one may check that the derivatives of $\mathcal{L}_1(s) - \mathcal{L}_3(s)$ and $\mathcal{L}_2(s) - \mathcal{L}_3(s)$ must be different at some $s_0 > 0$, for if they were the same, then $\mu_1 - \mu_3$ and $\mu_2 - \mu_3$ would have the same Laplace transform.

Bivariate analytic functions agreeing on a nonempty open set in $\mathbb{R}^2$ have to agree, and we apply this to the bivariate MGF of U' and that of V'. □

3.2.3 Four Values are Too Many

We will see our first counterexample to the nonnegative Aldous problem, proving the claim

Proposition 3.18. *P^+ is not mixture-restricting.*

We will even show that

Theorem 3.19. *$F_{\{0,1,2,3\}}$ is not mixture-restricting.*

In order to prove this theorem, we must develop the following.

First we need some definitions

Definition 3.20. *Fix $d > 0$. A subset S of $\mathbb{R}^d$ is said to be **determining** if whenever U and V are bounded $\mathbb{R}^d$-valued random variables such that $\forall s \in S$, $s \cdot U =_d s \cdot V$ implies that $U =_d V$.*

As mentioned multiple times in the introduction, algebraic properties of certain restrictions will be important. Thus it is no surprise that we need to consider the 0 set of polynomials.

Definition 3.21. *A subset S of $\mathbb{R}^d$ is called a **projective variety** if there exists a degree-homogeneous polynomial p defined on $\mathbb{R}^d$ such that $p \neq 0$ and $p^{-1}(0) = S$.*

The next result, based on the work of Cuesta-Albertos, Fraiman, and Ransford in [6] (Theorems 3.1 and 3.5), and the subsequent proposition will be used throughout this chapter.

Lemma 3.22. *Fix $d \geq 1$. A subset S of $\mathbb{R}^d$ is determining if and only if it is not contained in any projective variety.*

We will borrow the ideas found in [6] in order to prove this lemma.

Proof. Fix $d \geq 1$ and $S \subset \mathbb{R}^d$. We assume that S is nonempty because the claim is true if S is empty.

Suppose S is not contained in any projective variety, and consider two bounded random variables U, V. For each $n \geq 0$, we define the polynomial

$$p(x) = \mathbb{E}[(x \cdot U)^n] - \mathbb{E}[(x \cdot V)^n] \tag{3.15}$$

which is homogeneous of degree n. Since p vanishes on S, it follows that p must be the 0 polynomial. It follows that $\forall x \in \mathbb{R}^d$ we have that all the moments of the real-valued, bounded random variable $x \cdot U$ are equal to those of $x \cdot V$. Hence $\forall x \in \mathbb{R}^d$ we have $x \cdot U =_d x \cdot V$, thus $U =_d V$.

For the converse, it suffices to assume that S is a projective variety and construct two different probability measures μ and ν with bounded support, defined on $\mathbb{R}^d$, such that $\forall s \in S$, $s \cdot \mu = s \cdot \nu$. Here, if U is μ distributed then $s \cdot \mu$ means the distribution probability measure of $s \cdot U$ and similarly for ν.

Define an auxiliary function $f : \mathbb{C}^d \to \mathbb{C}$ given by

$$f(z) := \prod_{j=1}^{d} \frac{\sin z_j - z_j}{z_j^3}. \tag{3.16}$$

Here, we have $z = (z_1, \ldots, z_d)$.

It is routine to verify that

(i)f is even, entire, and real valued when restricted to $\mathbb{R}^d$

(ii)f is of exponential type, i.e. there is a $C > 0$ such that $\forall z$ we have $|f(z)| \leq Ce^{\sum_{j=1}^{d} |z_j|}$

(iii)There is a $C > 0$ such that $\forall x \in \mathbb{R}^d$ we have $|f(x)| \leq C/(1 + ||x||^2)$.

(iv) we have $f(0) \neq 0$.

In what follows, all that matters is that f has the properties listed above.

Now, find p a homogeneous polynomial on $\mathbb{R}^d$ that is not a constant such that $S \subset p^{-1}(0)$. The reason p can be chosen to be not a constant is that S is nonempty. Define $g : \mathbb{R}^d \to \mathbb{R}$ by $g(x) = p(x)^2 f(x)^K$, where K is chosen large enough so that $g \in L^2(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$. This is possible by (iii). Let $h = \hat{g}$ be the Fourier transform of g. By Plancherel's theorem we have that $h \in L^2(\mathbb{R}^d)$ and h is real-valued since g is even and real-valued. Note also that h is bounded continuous. Moreover, the Paley-Wiener theorem [16] tells us that h is supported in a compact subset of $\mathbb{R}^d$. Here, we are applying Paley-Wiener to g which is exponential type, analytic because f is exponential type, analytic. (We extend the definition of p to $\mathbb{C}^d$.) But, depending on the convention of the Fourier transform, we already know the inverse Fourier transform of g in terms of h. This shows that h is 0 off of a compact set.

Define finite, positive Borel measures with compact support by

$$\mu = h^+ dx, \, \nu = h^- dx, \tag{3.17}$$

which are mutually singular and therefore not equal. Using the Fourier inversion theorem, there is a constant $c \neq 0$ depending on the conventions of the Fourier transform and its inverse, such that

$$\hat{\mu} - \hat{\nu} = cg = cp^2 f^K \text{ on } \mathbb{R}^d. \tag{3.18}$$

Evaluating this equality at $0 \in \mathbb{R}^d$ we find that

$$\mu(\mathbb{R}^d) - \nu(\mathbb{R}^d) = cp(0)^2 f(0)^K = 0. \tag{3.19}$$

Since μ and ν are both nonzero (because g is not zero, p is not zero, so h is not zero) and have the same total mass, we may renormalize if necessary to force them to be probability measures.

We also have that for any $x \in \mathbb{R}^d$ $x \cdot \mu = x \cdot \nu$ if and only if $\forall t \in \mathbb{R}$ we have $\hat{\mu}(tx) - \hat{\nu}(tx) = 0$ if and only if $\forall t \in \mathbb{R}$, $cp(tx)^2 f(tx)^K = 0$. Thus, $\forall x \in S$ we have $p(x) = 0$ and thus $x \cdot \mu = x \cdot \nu$.

This completes the proof that S is not determining, because μ and ν are distributions of bounded random vectors with values in $\mathbb{R}^d$ for which their projections along vectors in S agree in law, but they are not equidistributed with one another. $\square$

Part of the utility of presenting the above proof is to show how constructions in the sequel that use this lemma can be made explicit. We have now stated and proven what we need from the existing literature. We use the above to prove the following proposition, which will drive many constructions in this chapter.

Proposition 3.23. *There exist $U \neq_d V$ which take values in the unit tetrahedron*

$$T_4 := \{(x_0, x_1, x_2, x_3) | \forall j \in \{0, 1, 2, 3\}, \, x_j \geq 0, \sum_{j=0}^{3} x_j = 1\} \subset \mathbb{R}^4 \tag{3.20}$$

such that $\forall y \in \mathbb{R}$ we have $c_4(y) \cdot U =_d c_4(y) \cdot V$. Here, $c_4(y) = (1, y, y^2, y^3) \in \mathbb{R}^4$.

Proof. Let $c_3(y) = (1, y, y^2) \in \mathbb{R}^3$ for all $y \in \mathbb{R}$. By Lemma 3.22 applied to the collection of scalar multiples of points on the curve c_3, there exist $W = (W_1, W_2, W_3) \neq_d Z = (Z_1, Z_2, Z_3)$ bounded $\mathbb{R}^3$-valued RVs such that

$$\forall y \in \mathbb{R}, \, c_3(y) \cdot W =_d c_3(y) \cdot Z. \tag{3.21}$$

Let $H = \{(x, y, z) \in \mathbb{R}^3 | 1 \geq x \geq y \geq z \geq 0\}$. Observe that H has nonempty interior. Therefore, given any compact set $C \subset \mathbb{R}^3$, there exist $a \in \mathbb{R}$, $a \neq 0$, $b \in \mathbb{R}^3$ such that $aC + b \subset H$. Because (3.21) is unchanged by rescaling and translation, we may assume that W, Z are H-valued. Define

$$\begin{aligned} U_0 &= 1 - W_1 & V_0 &= 1 - Z_1 \\ U_1 &= W_1 - W_2 & V_1 &= Z_1 - Z_2 \\ U_2 &= W_2 - W_3 & V_2 &= Z_2 - Z_3 \\ U_3 &= W_3 - 0 & V_3 &= Z_3 - 0 \\ U &= (U_0, U_1, U_2, U_3), & V &= (V_0, V_1, V_2, V_3). \end{aligned} \tag{3.22}$$

Observe that U and V are T_4-valued. From linear algebra (invertibility of the linear transform linking (U_1, U_2, U_3) to W and (V_1, V_2, V_3) to Z) it follows that $(U_1, U_2, U_3) \neq_d (V_1, V_2, V_3)$ so that $U \neq_d V$. By (3.21) we have

$$\forall y \in \mathbb{R},\ c_3(y) \cdot (U_1 + U_2 + U_3, U_2 + U_3, U_3) =_d c_3(y) \cdot (V_1 + V_2 + V_3, V_2 + V_3, V_3). \tag{3.23}$$

Plugging in the explicit form of $c_3(y)$ yields

$$\forall y \in \mathbb{R},\ U_1 + (1+y)U_2 + (1+y+y^2)U_3 =_d V_1 + (1+y)V_2 + (1+y+y^2)V_3. \tag{3.24}$$

Multiplying by $(y-1)$, we obtain

$$\begin{aligned} \forall y \in \mathbb{R},\ &(y-1)U_1 + (y^2-1)U_2 + (y^3-1)U_3 \\ &=_d (y-1)V_1 + (y^2-1)V_2 + (y^3-1)V_3. \end{aligned} \tag{3.25}$$

Thus, by definition of U_0, V_0 we learn that

$$\forall y \in \mathbb{R},\ c_4(y) \cdot U =_d c_4(y) \cdot V. \tag{3.26}$$

We have shown that U, V have the required properties. □

Remark 3.24. *Notice that trying the above argument in one dimension lower would fail because the curve $c_2(y) = (1, y)$ is not contained in a projective variety. This prevents the above argument from contradicting our earlier discovery that $F_{\{0,1,2\}}$ is mixture-restricting.*

We are now ready to establish Theorem 3.19.

Proof of Theorem 3.19. Obtain U, V from Proposition 3.23. Define α_1 and α_2 so that

$$U =_d \Big(\alpha_1(\{0\}), \alpha_1(\{1\}), \alpha_1(\{2\}), \alpha_1(\{3\})\Big) \tag{3.27}$$

and

$$V =_d \Big(\alpha_2(\{0\}), \alpha_2(\{1\}), \alpha_2(\{2\}), \alpha_2(\{3\})\Big). \tag{3.28}$$

Strictly speaking, α_1, α_2 may not be directing measures for some exchangeable sequence, but the distribution of α_1, α_2 are still elements of $P(P(\mathbb{R}))$ and are thus mixing measures, from which we may extract directing measures that have the same distribution as α_1, α_2. Thus, we may assume without loss of generality that α_1, α_2 are already defined on an appropriate probability space so that they are directing measures. In the future, this argument will not be explicitly stated.

By a change of variables in $c_4(y)$ from Proposition 3.23 we have $\forall s \geq 0$,

$$\begin{aligned}
\mathcal{L}_{\alpha_1}(s) &= (e^{-0s}, e^{-1s}, e^{-2s}, e^{-3s}) \cdot \Big(\alpha_1(\{0\}), \alpha_1(\{1\}), \alpha_1(\{2\}), \alpha_1(\{3\})\Big) \\
&=_d (e^{-0s}, e^{-1s}, e^{-2s}, e^{-3s}) \cdot \Big(\alpha_2(\{0\}), \alpha_2(\{1\}), \alpha_2(\{2\}), \alpha_2(\{3\})\Big) \\
&= \mathcal{L}_{\alpha_2}(s)
\end{aligned} \tag{3.29}$$

However, we have that the corresponding Θ_1 and Θ_2 are distinct because U and V have different distributions, hence as do α_1 and α_2. Thus, Θ_1, Θ_2 have the required properties when we use Lemma 3.11 to prove that $F_{\{0,1,2,3\}}$ is not mixture-restricting. □

3.2.4 A Generalization of the Four-Value Case

We may regard the values $0, 1, 2, 3$ as independent sums of 0 copies of 1, 1 copy of 1, 2 copies of 1 and 3 copies of 1 respectively. We may replace the constant random variable 1 with any nonnegative distribution to obtain a generalization of Theorem 3.19.

Proposition 3.25. *Let $\mu \in P^+$ be nondegenerate and let $*$ denote convolution. Let L be the collection of convex combinations of $\delta_0, \mu, \mu * \mu, \mu * \mu * \mu$. Then L is not mixture-restricting. The same is true of the convex combinations of $\delta_0, \mu, \mu * \mu, \mu * \mu * \mu, \ldots$.*

The primary purpose of this subsection is to prepare for comparisons and analogies with material from subsection 3.3.2, for which the case in which μ is Poisson is important. The techniques themselves are not logical prerequisites for material in the sequel.

Proof. To imitate the last proof, we need a homeomorphism between T_4 and L. We propose the map assigning a vector $(a, b, c, d) \in T_4$ the element $a\delta_0 + b\mu + c\mu * \mu + d\mu * \mu * \mu$ of L. This map is surjective. To see it is injective, if there are $(a, b, c, d), (a', b', c', d')$ with $a\delta_0 + b\mu + c\mu * \mu + d\mu * \mu * \mu = a'\delta_0 + b'\mu + c'\mu * \mu + d'\mu * \mu * \mu$ then taking Laplace transforms we have that $\forall s \geq 0$,

$$a\mathcal{L}_\mu(s)^0 + b\mathcal{L}_\mu(s)^1 + c\mathcal{L}_\mu(s)^2 + d\mathcal{L}_\mu(s)^3 = a'\mathcal{L}_\mu(s)^0 + b'\mathcal{L}_\mu(s)^1 + c'\mathcal{L}_\mu(s)^2 + d'\mathcal{L}_\mu(s)^3. \quad (3.30)$$

Consider the operation of multiplying by $\mathcal{L}_\mu(s)$, then taking the derivative in s, and then dividing by $\frac{d}{ds}\mathcal{L}_\mu(s)$. This division is legitimate because $\frac{d}{ds}\mathcal{L}_\mu < 0$ for nondegenerate μ. Iterating this process on (3.30) as many times as we wish shows that $(a, b, c, d) = (a', b', c', d')$.

Bijective continuous maps between compact Hausdorff spaces are homeomorphisms.

Now, obtain U, V as in Proposition 3.23. Define

$$\alpha_1 = U_0\delta_0 + U_1\mu + U_2\mu * \mu + U_3\mu * \mu * \mu \quad (3.31)$$

and

$$\alpha_2 = V_0\delta_0 + V_1\mu + V_2\mu * \mu + V_3\mu * \mu * \mu. \tag{3.32}$$

We assume without loss of generality that α_1, α_2 are directing measures. This way, Θ_1, Θ_2 are the pushforwards of the distributions of U, V via the above homeomorphism. Since Θ_1, Θ_2 are the distributions of α_1, α_2 by Remark 3.9 we have that $\alpha_1 \neq_d \alpha_2$ and $\Theta_1 \neq \Theta_2$. However, the random Laplace transform of α_1 is

$$\mathcal{L}_{\alpha_1} = \left(\mathcal{L}_\mu^0, \mathcal{L}_\mu^1, \mathcal{L}_\mu^2, \mathcal{L}_\mu^3\right) \cdot U \tag{3.33}$$

and the Laplace transform of α_2 is

$$\mathcal{L}_{\alpha_2} = \left(\mathcal{L}_\mu^0, \mathcal{L}_\mu^1, \mathcal{L}_\mu^2, \mathcal{L}_\mu^3\right) \cdot V. \tag{3.34}$$

Vectors of the form $\left(\mathcal{L}_\mu^0, \mathcal{L}_\mu^1, \mathcal{L}_\mu^2, \mathcal{L}_\mu^3\right)(s)$ are a subset of the image of c_4. Thus, $\forall s$, α_1 and α_2 have random Laplace transforms evaluated at s that have the same distribution. Thus, by Lemma 3.11 it follows that L is not mixture-restricting. □

3.2.5 Another Generalization of the Four-Value Case

The primary purpose of this subsection is to prepare for comparisons and analogies with material from Subsection 3.3.2, and to expose some new techniques that are useful in proving results showing the relationship between arithmetic and algebraic dependencies versus uniqueness results. The techniques themselves are not logical prerequisites for material in the sequel.

We may instead replace the role of independent sums by regular sums of the random variable with itself. So we regard $0, 1, 2, 3$ as the sum of $0, 1, 2, 3$ copies of 1. If instead of 1, we use an arbitrary nonnegative random variable, we arrive at the following generalization.

Proposition 3.26. *Let $\mu \in P^+$ be such that $\mathcal{L}_\mu(s)$ is a rational function in s, and let $T \neq 0$ be a random variable with distribution μ. Fix $N \geq 3$. Let L be the convex combinations of* *the distributions of* $0T, 1T, 2T, \ldots, NT$. *i.e.* $L \subset P^+$ *is the set of probability measures that have Laplace transform of the form*

$$\sum_{j=0}^{N} b_j \mathcal{L}_\mu(js) \tag{3.35}$$

where

$$\sum_{j=0}^{N} b_j = 1 \tag{3.36}$$

and $\forall j \geq 0$, *we have* $b_j \geq 0$.

Then L *is not mixture-restricting.*

Independent sums of exponential random variables give rational Laplace transforms, for example.

Proof. It suffices to handle the case $N = 3$. Assume that $\mathcal{L}_\mu(s) = p(s)/q(s)$ with p, q polynomials sharing no common factor, and q having no zeros in $[0, \infty)$. We seek a homogeneous polynomial $r \neq 0$ of 3 variables for which

$$r\left(\frac{p(1s)}{q(1s)} - \frac{p(0s)}{q(0s)}, \frac{p(2s)}{q(2s)} - \frac{p(0s)}{q(0s)}, \frac{p(3s)}{q(3s)} - \frac{p(0s)}{q(30)}\right) = 0. \tag{3.37}$$

This is equivalent to

$$\begin{gathered} r(p(1s)q(0s)q(2s)q(3s) - p(0s)q(1s)q(2s)q(3s), p(2s)q(0s)q(1s)q(3s) \\ -p(0s)q(1s)q(2s)q(3s), p(3s)q(0s)q(1s)q(2s) - p(0s)q(1s)q(2s)q(3s)) = 0 \end{gathered} \tag{3.38}$$

Let us say that p has degree $n \geq 0$, q has degree $m \geq 0$, and r has degree $l > 0$. The space of polynomials in s of degree at most $l(3m + n)$ has dimension linear in l as a vector space over $\mathbb{R}$ whereas the space of polynomials in 3 variables that are homogeneous, of degree l is quadratic in l. Therefore, there exists l large enough such that the assignment of homogeneous degree l polynomials in 3 variables to polynomials of degree at most $l(3m+n)$ given by

$$\begin{aligned} r \mapsto r(&p(1s)q(0s)q(2s)q(3s) - p(0s)q(1s)q(2s)q(3s), p(2s)q(0s)q(1s)q(3s) \\ &-p(0s)q(1s)q(2s)q(3s), p(3s)q(0s)q(1s)q(2s) - p(0s)q(1s)q(2s)q(3s)) \end{aligned} \tag{3.39}$$

has nontrivial kernel.

Thus we have a nonzero homogeneous polynomial r for which $r(\mathcal{L}_\mu(1s)-\mathcal{L}_\mu(0s), \mathcal{L}_\mu(2s)-\mathcal{L}_\mu(0s), \mathcal{L}_\mu(3s)-\mathcal{L}_\mu(0s)) = 0$.

Now, we use Lemma 3.22 to find $W = (W_1, W_2, W_3) \neq_d Z = (Z_1, Z_2, Z_3)$ bounded random vectors for which $\forall s \geq 0$ we have

$$\begin{aligned} &(\mathcal{L}_\mu(1s) - \mathcal{L}_\mu(0s), \mathcal{L}_\mu(2s) - \mathcal{L}_\mu(0s), \mathcal{L}_\mu(3s) - \mathcal{L}_\mu(0s)) \cdot W \\ &= (\mathcal{L}_\mu(1s) - \mathcal{L}_\mu(0s), \mathcal{L}_\mu(2s) - \mathcal{L}_\mu(0s), \mathcal{L}_\mu(3s) - \mathcal{L}_\mu(0s)) \cdot Z. \end{aligned} \tag{3.40}$$

We then find C compact such that W, V are both C-valued, and $a \neq 0$, $b \in \mathbb{R}^3$ such that $aC + b \subset T_3' := \{(a,b,c) | a+b+c \leq 1, a,b,c \geq 0\}$. This is possible because C is compact and T_3' has nonempty interior. Thus, we may assume that W, Z were T_3'-valued to begin with. We now define U, V via

$$\begin{aligned} U_0 &= 1 - U_1 - U_2 - U_3 & V_0 &= 1 - Z_1 - Z_2 - Z_3 \\ U_1 &= W_1 & V_1 &= Z_1 \\ U_2 &= W_2 & V_2 &= Z_2 \\ U_3 &= W_3 & V_3 &= Z_3 \\ U &= (U_0, U_1, U_2, U_3), & V &= (V_0, V_1, V_2, V_3). \end{aligned} \tag{3.41}$$

Thus $U \neq_d V$ and $\forall s \geq 0$, we have

$$(\mathcal{L}_\mu(0s), \mathcal{L}_\mu(1s), \mathcal{L}_\mu(2s), \mathcal{L}_\mu(3s)) \cdot U$$
$$=_d (\mathcal{L}_\mu(0s), \mathcal{L}_\mu(1s), \mathcal{L}_\mu(2s), \mathcal{L}_\mu(3s)) \cdot V. \quad (3.42)$$

Definition 3.27. *Let μ_k denote the probability distribution of kT.*

We define $\alpha_1 = U_0\mu_0 + U_1\mu_1 + U_2\mu_2 + U_3\mu_3$ and $\alpha_2 = V_0\mu_0 + V_1\mu_1 + V_2\mu_2 + V_3\mu_3$ which are L-valued. Without loss of generality, we assume α_1 and α_2 are directing measures.

Since $T \neq 0$, we have that all of the μ_k are distinct, nondegenerate, and therefore have Laplace transforms with derivatives that are never 0. We aim to show that T_4 is homeomorphic to L through the map $(a, b, c, d) \mapsto a\mu_0 + b\mu_1 + c\mu_2 + d\mu_3$. This is surjective. Also, it is injective because if there are $(a, b, c, d), (a', b', c', d',)$ such that $a\mu_0 + b\mu_1 + c\mu_2 + d\mu_3 = a'\mu_0 + b'\mu_1 + c'\mu_2 + d'\mu_3$ then we may take Laplace transforms to obtain $\forall s \geq 0$

$$a\mathcal{L}_\mu(0s) + b\mathcal{L}_\mu(1s) + c\mathcal{L}_\mu(2s) + d\mathcal{L}_\mu(3s)$$
$$= a'\mathcal{L}_\mu(0s) + b'\mathcal{L}_\mu(1s) + c'\mathcal{L}_\mu(2s) + d'\mathcal{L}_\mu(3s). \quad (3.43)$$

We may take the derivative of this relation k times, then take $s \downarrow 0$, then divide by $[\frac{d^k}{ds^k}\mathcal{L}_\mu](0)$. Again, this operation shows us that $(a, b, c, d) = (a', b', c', d')$. Continuous bijections between compact Hausdorff spaces are always homeomorphisms.

Therefore, $U \neq_d V$ implies that $\alpha_1 \neq_d \alpha_2$ and $\Theta_1 \neq \Theta_2$. Also, $\forall s \geq 0$, $\mathcal{L}_{\alpha_1}(s) =_d \mathcal{L}_{\alpha_1}(s)$. This is because the left side is the left side of (3.42) in distribution and the right side is the right side of (3.42) in distribution. This suffices by Lemma 3.11.

□

In the following example, the answer to the Aldous problem actually is different than above despite the fact that the arithmetic dependencies among the μ_k still remain. We will use a very transcendental Laplace transform to make the arithmetic dependencies irrelevant.

Proposition 3.28. *Let T be Poisson distributed with parameter λ. Let $N \geq 1$. Let μ be the distribution of T, and define $\mu_0, \mu_1, \mu_2, \ldots, \mu_N$ as before. Define L as in Proposition 3.26. Then L is mixture-restricting.*

Proof. We will argue for $N = 3$, with the general case being similar. Let $\mathcal{L}(s) = e^{\lambda(e^{-s}-1)}$ denote the Laplace transform of μ. We already know that T_4 is homeomorphic to L in the natural way. (see the last proof) Therefore, it suffices to show that there cannot be any $U = (U_0, U_1, U_2, U_3) \neq_d V = (V_0, V_1, V_2, V_3)$ defined on T_4 for which $\forall s \geq 0$, $(\mathcal{L}(0s), \mathcal{L}(1s), \mathcal{L}(2s), \mathcal{L}(3s)) \cdot U = (\mathcal{L}(0s), \mathcal{L}(1s), \mathcal{L}(2s), \mathcal{L}(3s)) \cdot V$. We will in fact show that there is no homogeneous polynomial other than 0 that vanishes on the image of the curve $(\mathcal{L}(0s), \mathcal{L}(1s), \mathcal{L}(2s), \mathcal{L}(3s)) \in \mathbb{R}^4$ defined for $s \geq 0$. Suppose that $r \neq 0$ is such a homogeneous polynomial of degree l, say. Order the set M_l of monic monomials of total degree l in 4 variables by ordering lexicographically on the exponents, with the fourth variable taking highest priority, then the third, second, then first. This is a total ordering. Write a_w for the coefficient of any monic monomial w in r. Find the largest monic monomial with a nonzero coefficient in r. Call this monomial $m(x_0, x_1, x_2, x_3)$. Then we have

$$r(x_0, x_1, x_2, x_3) = a_m m(x_0, x_1, x_2, x_3) + \sum_{w<m\in M_l} a_w w(x_0, x_1, x_2, x_3) \tag{3.44}$$

with $a_m \neq 0$.

We have $\forall s \geq 0$

$$r(\mathcal{L}(0s), \mathcal{L}(1s), \mathcal{L}(2s), \mathcal{L}(3s)) = 0 \tag{3.45}$$

so by the theorem of complex analysis asserting the equality of holomorphic functions defined on the same connected open domain, agreeing on a set with an accumulation point within this domain, (3.45) holds also for $s < 0$. The term of $r(\mathcal{L}(0s), \mathcal{L}(1s), \mathcal{L}(2s), \mathcal{L}(3s))$ corresponding to m goes to ∞ as $s \to -\infty$ faster than any of the other terms, so the coefficient a_m is 0, contradiction. □

3.2.6 The Normal Case

The normal case is another case of significance to the next section on the continuous time analog of the present problem.

Lemma 3.29. *Let N be the collection of normal distributions, including the degenerate ones. For all $\mu \in N$ let $M(\mu)$ =mean of μ and let $V(\mu)$ = variance of μ. Then the map $(M, V) : N \to \mathbb{R} \times [0, \infty)$ is a homeomorphism, hence measurable isomorphism.*

Proof. This follows from convergence of types. □

Remark 3.30. *Let Θ be a mixing measure supported in N. Since Θ is then a probability measure on N, we can view M, V as random variables giving the (random) mean and variance of an element of N drawn with prior distribution Θ. Then the joint distribution of the corresponding exchangeable sequence X is given by $(X_i)_{\{i\in\mathbb{N}\}} =_d (A+B^{1/2}\mathcal{N}(0,1)_i)_{\{i\in\mathbb{N}\}}$ where the entire family $\{\mathcal{N}(0,1)_1, \mathcal{N}(0,1)_2, \ldots, (A, B)\}$ is independent (but the notation indicates A, B are not necessarily independent), $(A, B) =_d (M, V)$, and $\mathcal{N}(0,1)_i$ is normal with mean 0 and variance 1.*

See, for instance, p.29 of [1] regarding this remark.

We highlight in the remark that the distribution of (M, V) is calculated relative to Θ, and that it is necessary to use (A, B) instead of (M, V) when we deal with the independent normals because these normals and (A, B) are constructed on the same probability space.

We already specified that X, Y corresponds to Θ_1, Θ_2 and α_1, α_2 via Remark 3.9. For this subsection, when Θ_1, Θ_2 are supported on N, we will use (M_1, V_1) to indicate (M, V) defined on the probability space (N, Θ_1) and (M_2, V_2) to indicate (M, V) defined on the probability space (N, Θ_2).

We have the following transform inversion fact.

Lemma 3.31. *Suppose μ, ν be probability measures on $\mathbb{R} \times [0, \infty)$ such that $\forall t, s \in \mathbb{R} \times [0, \infty)$ we have*

$$\int_{\mathbb{R}\times[0,\infty)} e^{itx-sy} d\mu(x,y) = \int_{\mathbb{R}\times[0,\infty)} e^{itx-sy} d\nu(x,y). \tag{3.46}$$

Then $\mu = \nu$.

Notice that N is not a subset of P^+. The author is uncertain if N is mixture-restricting or not, which seems to rely on a generalization of Muntz's Theorem (see [8]) which would include the sequence of points $t_n = 1/n$ in the role of the values of the parameter at which the Laplace transform is known a priori. However, what is true is the following.

Proposition 3.32. *Let Θ_1, Θ_2 supported in N be given. Suppose that V_1, V_2 have finite MGF in some neighborhood around 0, and that $\forall n > 0$, $S_n =_d T_n$. Then $\Theta_1 = \Theta_2$.*

Proof. Let $(A, B), (A', B')$ have the same distributions as $(M_1, V_1), (M_2, V_2)$ respectively with the three random vectors/variables $(A, B), (A', B'), \mathcal{N}(0, 1)$ all independent. From $S_n =_d T_n$ we learn that $\forall n \geq 0$, $nA + (nB)^{1/2}\mathcal{N}(0,1) =_d nA' + (nB')^{1/2}\mathcal{N}(0,1)$ Computing the characteristic function of both sides reveals that $\forall n \geq 0, t \in \mathbb{R}$

$$\mathbb{E}[e^{itnA - t^2 nB/2}] = \mathbb{E}[e^{itnA' - t^2 nB'/2}] \tag{3.47}$$

or equivalently $\forall n > 0$, $t \in \mathbb{R}$

$$\mathbb{E}[e^{itA - \frac{t^2 B}{2n}}] = \mathbb{E}[e^{itA' - \frac{t^2 B'}{2n}}]. \tag{3.48}$$

Looking at (3.48) for fixed t and varying n, it follows that the convergence of the MGF in a neighborhood of 0 is precisely what is needed to be able to use complex analysis to conclude that for each fixed t, we have that $\forall s \geq 0$

$$\mathbb{E}[e^{itA - sB}] = \mathbb{E}[e^{itA' - sB'}] \tag{3.49}$$

Particularly, we are using the fact that holomorphic functions defined on a common connected open domain, agreeing on a set with a limit point in the domain must be equal. Namely, this limit point would be $s = 0$ regardless of which t was fixed. The MGF hypothesis is what allows $s = 0$ to be in the (interior of the) domain of these transforms.

Then, by Lemma 3.31 it follows that $(A, B) =_d (A', B')$ so that $(M_1, V_1) =_d (M_2, V_2)$ from which it follows by Lemma 3.29 that $\Theta_1 = \Theta_2$.

□

3.3 Continuous Time Exchangeability Problem

We could view the questions answered in the last section from the perspective of S_n, T_n. These are mixtures of partial sums of iid sequences. From this point of view, it is natural to consider mixtures of Lévy Processes, which are the continuous time analog. Recall that a Lévy process is an independent stationary increments process that is continuous in probability and starts at 0. We will use the notation S_t, T_t for mixtures of Lévy processes, after they are defined, in order to reflect this analogy. In order to aid our discussion, we recall:

Lemma 3.33 (Lévy Khintchine Formula). *Let $Z = (Z_t)_{t\geq 0}$ be a Lévy process. Then there exist unique β, σ^2, ν such that ν is a finite measure on $\mathbb{R}$ with $\nu(\{0\}) = 0$, $\sigma^2 \geq 0$, $\beta \in \mathbb{R}$ and $\forall u \in \mathbb{R}, t \geq 0$ we have*

$$\mathbb{E}[e^{iuZ_t}] = \exp\left\{iut\beta - \frac{u^2 t\sigma^2}{2} + t\int_{\mathbb{R}} \left(e^{iux} - 1 - \frac{iux}{1+x^2}\right)\frac{1+x^2}{x^2}d\nu(x)\right\}. \tag{3.50}$$

Furthermore, every $\beta \in \mathbb{R}, \sigma^2 \geq 0, \nu$ a finite measure on $\mathbb{R}$ with no atom at 0 corresponds to a unique (up to distributional equality) Lévy process with characteristic function given by (3.50).

We would now like to define the notion of a mixture of Lévy processes. For technical reasons, we downplay the role of exchangeability. For the moment, we also focus on characteristic functions in order to be able to state the definition before worrying about measurability concerns associated with generalizing (3.4). Then we will show how the definitions we make are directly analogous to those of the previous section. These claims are mostly for checking intuition about what a mixture of Lévy Processes should mean, but they will also be used in Subsection 3.3.2. (They will not be featured as prominently in the next subsection.)

Given parameters $\beta \in \mathbb{R}, \sigma^2 \geq 0, \nu$ finite measure on $\mathbb{R}$ with no atom at 0, we will use the notation $\phi_{\beta,\sigma^2,\nu,t_1,\ldots,t_n}$ for the n-variate characteristic function of $(Z_{t_1},\ldots,Z_{t_n})$ where the Lévy process Z_t is chosen for parameters β, σ^2, ν. We will use M_0^+ to denote the collection of nonnegative finite measures on $\mathbb{R}$ that vanish at $\{0\}$. From now on, we will always implicitly assume $(\beta, \sigma^2, \nu) \in \mathbb{R} \times [0,\infty) \times M_0^+$. We will use $L_{\beta,\sigma^2,\nu} := \left((Z_t)_{t\geq 0}\right)^*(\mathbb{P})$ to mean the pushforward of $\mathbb{P}$ (the probability measure on whichever space the process under study is defined on) via the Lévy process $Z = (Z_t)_{t\geq 0}$, i.e. the (joint) distribution of the Lévy Process.

Definition 3.34. *A **mixture of Lévy processes** is $S = (S_t)_{t\geq 0}$ such that there exist a probability measure Θ on $\mathbb{R} \times [0,\infty) \times M_0^+$ for which the joint characteristic function of S is specified by $\forall n \geq 1, \forall 0 \leq t_1 \leq \cdots \leq t_n, u_1,\ldots,u_n \in \mathbb{R}$ we have*

$$\mathbb{E}[e^{i\sum_{j=1}^n u_j S_{t_j}}] = \int_{\mathbb{R}\times[0,\infty)\times M^+} \phi_{\beta,\sigma^2,\nu,t_1,\ldots,t_n}(u_1,\ldots,u_n) d\Theta(\beta,\sigma^2,\nu) \tag{3.51}$$

We call Θ the mixing measure.

By using discrete time de Finetti, it follows that in this case Θ is uniquely determined by the distribution of S.

In the discrete time case, we were able to obtain a discrete time process (namely S_n) from the mixing measure Θ. It is reasonable to ask if the same can be done here.

Lemma 3.35. *Given a probability measure Θ on $\mathbb{R} \times [0,\infty) \times M_0^+$, there is a unique (up to joint distributional equality) stochastic process $S = (S_t)_{t\geq 0}$ for which Θ is the mixing measure.*

Proof. First we check existence. Restrict to finite dimensional distributions, using (3.51) to define these finite dimensional distributions. Then check Kolmogorov consistency and use Kolmogorov extension theorem.

The uniqueness up to distributional equality is built into the definition of mixture of Lévy processes. □

From Lévy continuity, stationarity of increments and the fact that distributional convergence to 0 is the same as in probability convergence to 0, it follows that all mixtures of Lévy processes are continuous in probability.

We will use the following notation: $S = (S_t)_{t\geq 0}, T = (T_t)_{t\geq 0}$ will be the mixture of Lévy processes, with mixing measures Θ_1, Θ_2. We will have no need for trying to define some analog of α_1, α_2 in this context.

The set $\mathcal{I}$ of infinitely divisible distributions is closed in $P(\mathbb{R})$, hence measurable.

We regard (3.50) as specifying a bijection between $\mathbb{R} \times [0, \infty) \times M_0^+$ and the collection $\mathfrak{L}$ of distributions of Lévy processes $\left((X_t)_{t\geq 0}\right)^*(\mathbb{P})$. So $\mathfrak{L} \subset P(\mathbb{R}^{[0,\infty)})$. $\mathfrak{L}$ is given the smallest σ algebra so that passage from an element of $\mathfrak{L}$ to its marginals is measurable from $\mathfrak{L}$ to $\mathcal{I}$. That is, $\forall t_0 \geq 0$, $(X_t)^*_{t\geq 0}(\mathbb{P}) \mapsto X^*_{t_0}(\mathbb{P})$ should be measurable. It follows from standard proofs of (3.50) that the bijection specified by (3.50) is a measurable isomorphism. There is also a natural measurable isomorphism between $\mathfrak{L}$ and $\mathcal{I}$ via $\left((X_t)_{t\geq 0}\right)^*(\mathbb{P}) \mapsto X_1^*(\mathbb{P})$.

Because $P(\mathbb{R})$ with vague convergence is a Polish space, and $\mathcal{I}$ is closed in $P(\mathbb{R})$, we have that $\mathfrak{L}$ is a standard Borel space. Therefore, our definition of a mixture of Lévy processes is entirely parallel to the notion of mixture from the last section.

Because of these observations, it is sensible to state and we have proven the following:

Lemma 3.36. *S is a mixture of Lévy processes if and only if there exists Θ a probability measure on $\mathfrak{L}$ for which $\forall A \subset \mathbb{R}^{[0,\infty)}$ product measurable,*

$$\mathbb{P}(S \in A) = \int_{\mathfrak{L}} \gamma(A) d\Theta(\gamma) \tag{3.52}$$

if and only if there exists Θ a probability measure on $\mathbb{R} \times [0, \infty) \times M_0^+$ for which $\forall A \subset \mathbb{R}^{[0,\infty)}$ product measurable,

$$\mathbb{P}(S \in A) = \int_{\mathbb{R}\times[0,\infty)\times M_0^+} L_{\beta,\sigma^2,\nu}(A) d\Theta(\beta, \sigma^2, \nu). \tag{3.53}$$

In any case, Θ is unique.

Thus, when speaking of Θ being a mixing measure or related topics, we will freely use these identifications. For example, we will allow ourselves to say "mixtures of Brownian motions". Also, we will no longer use the notation $\mathbb{R} \times [0, \infty) \times M_0^+$ and will use $\mathfrak{L}$ instead. These identifications needed to be measurable in order for it to be possible to discuss mixtures using any of the descriptions, reconciling with the intuition that Lévy processes are truly the same as their Lévy Khintchine parameters and as infinitely divisible distributions.

The interested reader can combine what we have done so far with [9] to see that being a mixture of Lévy processes is equivalent to being continuous in probability and satisfying a certain kind of exchangeable increments hypothesis.

Again, we will use a notion of a class being mixture-restricting to abbreviate our discussion.

Definition 3.37. *We will say that a measurable subset L of $\mathfrak{L}$ is* ***mixture-restricting*** *if whenever Θ_1, Θ_2 are concentrated on L and $\forall t \geq 0$, $S_t =_d T_t$ we have $S =_d T$.*

3.3.1 The Case of Brownian Motions

Recall that a Brownian motion is a Gaussian Lévy process, and can have drift and can proceed at any positive rate. (i.e. we only require that the variance at $t = 1$ is positive.) We denote the space of Brownian motions by $\mathrm{BM} \subset \mathfrak{L}$. BM corresponds to the requirement that the ν component of the Lévy Khintchine formula is 0. We claim that

Proposition 3.38. *BM is mixture-restricting.*

Proof. Using (3.51) for one value of t at a time, we have $\forall u \in \mathbb{R}, t \geq 0$

$$\begin{aligned}&\int_{\mathbb{R}\times[0,\infty)\times\{0\}} \exp\{iut\beta - tu^2\sigma^2/2\} d\Theta_1(\beta, \sigma, 0)\\ = &\int_{\mathbb{R}\times[0,\infty)\times\{0\}} \exp\{iut\beta - tu^2\sigma^2/2\} d\Theta_2(\beta, \sigma, 0).\end{aligned} \tag{3.54}$$

Lemma 3.31 now finishes the proof.

□

Notice that in the discrete time normal case, we had the last equation only for $t = 1/n$ but now we have it for all $t \geq 0$, which is important in eliminating the need for assumptions about convergence of MGFs. Because the discrete set of rationally related numbers $1/n$, arising via application of $r \mapsto 1/r$ to $\mathbb{N}$ in the proof of Proposition 3.32, is replaced with a continuum in the above proof, this can be thought of as a destruction of the arithmetic structure. As promised, this is a case in which passage to the continuous time problem implies not only the additional information of infinite divisibility, but crucially the observations at a continuum of times rather than only a discrete set.

3.3.2 A Poisson-Flavored Case

All functions of u of the form

$$\exp\left\{\int_{\mathbb{R}}(e^{iux} - 1)d\mu(x)\right\} \tag{3.55}$$

are characteristic functions of infinitely divisible distributions, as long as μ is a finite non-negative Borel measure. This can be seen by taking a vague limit of sums of independent Poisson Processes with various rates and jump sizes.

Call LISPP the subset of $\mathfrak{L}$ determined by (3.55). (Here LISPP stands for "limits of independent sums of Poisson Processes.") We may think of the elements of LISPP as "independent integrals" of Poisson Processes, which is a different notion than a mixture of Poisson Processes and also different from compound Poisson processes. By a calculation, we have

Lemma 3.39. *LISPP is measurable in $\mathfrak{L}$ because it is actually determined by the conditions $\int_{\mathbb{R}} \frac{1+x^2}{x^2} d\nu(x) < \infty, \beta = \int_{\mathbb{R}} \frac{1}{x} d\nu(x)$.*

It follows from the description of LISPP above that μ is uniquely determined by the infinitely divisible distribution. Moreover,

$$(X_t)^*_{t\geq 0}(\mathbb{P}) \in \text{LISPP} \mapsto \mu \in M_+(\mathbb{R}) \tag{3.56}$$

is a measurable isomorphism, which is defined on LISPP. Therefore, we may identify each element of LISPP with a nonnegative finite measure on $\mathbb{R}$ via this correspondence. Also, if LISPP^+ is the subset of LISPP corresponding to μ supported in $[0, \infty)$ (i.e. we only allow positive jump size Poisson Processes to enter the independent integral), then LISPP^+ is of course measurable in LISPP.

Since (3.55) specifies the distribution of a nonnegative infinitely divisible distribution for elements of LISPP^+, the Laplace transform can be calculated by analytic continuation: $\forall \mu$ finite Borel measure on $[0, \infty)$, we have the function

$$\exp\left\{\int_{[0,\infty)} (e^{-sx} - 1) d\mu(x)\right\} \tag{3.57}$$

of s is the Laplace transform of a member of LISPP^+, and moreover these are the only Laplace transforms of members of LISPP^+.

We will also refer to $\text{LISPP}_1, \text{LISPP}_1^+$ to denote the requirement that μ be a probability measure. These are also measurable subsets of LISPP.

If Θ is concentrated on $\text{LISPP}, \text{LISPP}^+, \text{LISPP}_1$ or LISPP_1^+, then μ can be regarded as a random measure defined on $\text{LISPP}, \text{LISPP}^+, \text{LISPP}_1$ or LISPP_1^+. We now show how the last section on the discrete problem can be embedded into the current problem.

Lemma 3.40. *Let L be a measurable subset of $LISPP_1^+$. Then L is mixture-restricting if and only if $\forall \Theta_1, \Theta_2$ concentrated on L, we have*

$$\forall s \geq 0,\ \mathcal{L}_{\mu_1}(s) =_d \mathcal{L}_{\mu_2}(s) \tag{3.58}$$

implies

$$\Theta_1 =_d \Theta_2. \tag{3.59}$$

Here, we regard $\mu \mapsto \mu$ as a map from $LISPP_1^+$ to $P([0,\infty))$, and we regard Θ_1, Θ_2 as giving the structure of a probability space to $LISPP_1^+$ in two different ways. We use μ_1, μ_2 to mean random probability measures with the same distribution as μ under Θ_1 and Θ_2 respectively.

Therefore, even though there is no ideological connection between the mixing measure of the last section and the measures μ from this section (which are similar to the Lévy Khintchine measure), at the level of the mathematical formalisms the problems are related.

Proof. Observe first that if S, T are mixtures from L, then $\forall t \geq 0$, $S_t =_d T_t$ if and only if $\forall t \geq 0, s \geq 0, n \geq 0$ we have

$$\mathbb{E}[\exp\{nt\int_{[0,\infty)}(e^{-sx}-1)d\mu_1(x)\}] = \mathbb{E}[\exp\{nt\int_{[0,\infty)}(e^{-sx}-1)d\mu_2(x)\}] \tag{3.60}$$

by using (3.51) for one value of t at a time. Then we know that (3.60) is equivalent to $\forall t \geq 0, s \geq 0, n \geq 0$

$$\mathbb{E}[\exp\{t\int_{[0,\infty)}(e^{-sx}-1)d\mu_1(x)\}^n] = \mathbb{E}[\exp\{t\int_{[0,\infty)}(e^{-sx}-1)d\mu_2(x)\}^n] \tag{3.61}$$

which, by the bounded moment problem is equivalent to $\forall t \geq 0, s \geq 0$

$$\exp\{t\int_{[0,\infty)}(e^{-sx}-1)d\mu_1(x)\} =_d \exp\{t\int_{[0,\infty)}(e^{-sx}-1)d\mu_2(x)\}. \tag{3.62}$$

But this last statement is equivalent to $\forall t \geq 0, s \geq 0$

$$t\int_{[0,\infty)}(e^{-sx}-1)d\mu_1(x) =_d t\int_{[0,\infty)}(e^{-sx}-1)d\mu_2(x) \tag{3.63}$$

which is the same as $\forall s \geq 0$

$$\int_{[0,\infty)}(e^{-sx}-1)d\mu_1(x) =_d \int_{[0,\infty)}(e^{-sx}-1)d\mu_2(x) \tag{3.64}$$

and therefore also the same as $\forall s \geq 0$

$$\int_{[0,\infty)} e^{-sx} d\mu_1(x) =_d \int_{[0,\infty)} e^{-sx} d\mu_2(x) \tag{3.65}$$

because μ_1, μ_2 are always probability measures.

We have that $S =_d T$ if and only if $\Theta_1 = \Theta_2$.

□

Remark 3.41. *Notice how it did not matter that we made a continuum of observations because the nonnegativity assumption LISPP$^+$ allowed us to use the Laplace transform. Since, when restricted to real arguments, exponentiation is invertible, we were able to cancel an exponentiation and then cancel the t. Therefore, a measurable subset of LISPP$_1^+$, when viewed as a subset of $\mathcal{I}$, is mixture-restricting if and only if it is mixture-restricting as in the last section. That is, to tell apart two mixtures of LISPP$_1^+$s, we only need to observe at natural number times. This manipulation was not available in the BM case because there we were dealing with complex exponentiation, which also forbids the use of the bounded moment problem above.*

For the next result, $\forall A \subset P(\mathbb{R})$ measurable, we use

$$\text{LISPP}(A) := \{(X_t)^*_{t\geq 0}(\mathbb{P}) \in \text{LISPP} | \mu \in A\} \tag{3.66}$$

where μ is the measure in (3.55) giving the characteristic function of X_1. For example, LISPP$F_{\{0,1,2,3\}}$ denotes the collection of Lévy processes that can be written as an independent sum of of a Poisson Process of rate 0 and jump size x_0, one of rate 1 and jump size x_1, one of rate 2 and jump size x_2, and one of rate 3 and jump size x_3 such that $(x_0, x_1, x_2, x_3) \in T_4$. Since we will always use A such that all measures in A are supported in $[0, \infty)$, our LISPP(A) will always be a subset of LISPP$_1^+$ so we will be able to use the above lemma.

The fact that the continuum of observations does not destroy any arithmetic structure suggests that the situation with LISPP$_1^+$ will be more nuanced than the situation with BM, where uniqueness of $S = (S_t)_{t\geq 0}$ held without further conditions. Indeed, we now know that the variety of possibilities of mixture-restricting classes is at least as much as that of the discrete time problem:

Theorem 3.42. *Let $A \subset P^+$ be measurable, and consist only of compactly supported measures. Then A is mixture-restricting (in the only sense that is available, i.e. from the discrete time problem) iff LISPP(A) is mixture-restricting (in either equivalently the sense of the present section or the last when LISPP(A) is viewed as a subset of $\mathcal{I} \subset P(\mathbb{R})$).*

Proof. This is a consequence of Lemma 3.40 □

Remark 3.43. *The upshot of this theorem is that any set of probability measures that is mixture-restricting when in the role of the allowed components of the mixture are also mixture-restricting when in the role of μ. One could iterate this. If A is mixture-restricting, then LISPP(A) is mixture-restricting, then LISPP(LISPP(A)) is mixture-restricting and so on. After all, thanks to the fact that knowing each S_n is enough, as long as we only concern ourselves with mixtures of Lévy processes, there is no difference between the continuous time and discrete time problems.*

Corollary 3.44. *Let ν_1, ν_2, ν_3 be probability measures on $[0, \infty)$. Let C be the convex hull of $\{\nu_1, \nu_2, \nu_3\}$. Then LISPP(C) is mixture-restricting. Also, if A is a discrete, countable set of real numbers that is linearly independent over $\mathbb{Q}$ then LISPP(F_A) is mixture-restricting. LISPP$(F_{\{0,1,2,3\}})$ is not mixture-restricting.*

Remark 3.45. *Of course, most of the other results of the last section could be generalized just as easily, but they are not as meaningful as the ones listed above in the context of mixtures of Lévy processes.*

Notice the comparison with Remark 3.15, which is a reasonable comparison because we already saw that in this case knowing all S_t is no different than knowing only the S_n. In Remark 3.15 we were concerned with mixtures of Poissons (with jump size 1 and rate λ) and we were mixing over different rates. There, uniqueness of S_n was true, which can also be deduced from our present machinery by applying Theorem 3.42 to the mixture-restricting set $A = \{\delta_x | x \in [0, \infty)\}$.

In general, uniqueness fails in the present setting for even jump sizes restricted to $\{0, 1, 2, 3\}$, because we allow the rates to vary as long as they add up to 1. Both this and Remark 3.15 are different than the situation in Proposition 3.28, where the rate was fixed, but the jump size was allowed to be $0, 1, 2$, or 3 and we were allowed to take convex combinations of these distributions before mixing them.

Another different situation occured in the context of Proposition 3.25 applied to Poisson distributions, where the jump size was fixed at 1, the rate was fixed at λ, but we allowed independent sums and convex combinations to come in before we take the mixture.

These show that the mixture-restricting property concerns not only the type of distributions used to build up a class, but also the specifics of how the class is assembled from those distributions.

3.4 A Class of Uniqueness Problems

From now on we will use

$$\forall s \in \mathbb{R},\, \mathcal{M}_\mu(s) = \int_{\mathbb{R}} e^{sx} d\mu(x) \tag{3.67}$$

to mean the MGF (at $s \in \mathbb{R}$) of a probability measure μ on $\mathbb{R}$. Here, μ may be random or deterministic. We will often make assumptions about finiteness of the MGF, which we will state as needed. We also use

$$\forall t \in \mathbb{R},\, \phi_\mu(t) = \int_{\mathbb{R}} e^{itx} d\mu(x) \tag{3.68}$$

to mean the characteristic function (at $t \in \mathbb{R}$) of a probability measure μ on $\mathbb{R}$. Here, μ may be random or deterministic.

In this section, we will discuss a variety of uniqueness problems regarding determining the distribution of a random probability measure from limited information. These problems will all run parallel to classical versions of various uniqueness results.

Definition 3.46. *Given a probability measure μ defined on $\mathbb{R}$, let $\mu_k := \int_{\mathbb{R}} x^k d\mu(x)$ denote the moment of order k of μ, when it exists. Define $\mathfrak{C}$ to be the collection of $\mu \in P(\mathbb{R})$ that obey the Carleman condition that all the moments exist, are finite, and $\sum_{j=0}^{\infty} 1/\mu_{2j}^{1/2j} = \infty$*

with the convention that $1/0 = \infty$*. Let* $M_{<\infty}$ *denote the collection of* $\mu \in P(\mathbb{R})$ *with finite MGF in some neighborhood of* 0*.*

Sometimes μ will denote a random measure.

We will sometimes make the assumption that the random measure μ is uniformly bounded a priori. This means there exists $M > 0$ such that μ is a.s. supported in $[-M, M]$.

We will use the notation $\mu_k := \int_{\mathbb{R}} x^k d\mu(x)$ to indicate the (random) moment of order k for μ, wherever this is defined.

For any of the classes in Definition 3.46 or for P^+, we will say μ is a member of that class if this holds a.s. Observe that all of these conditions are measurable. Notice this implies no uniformity, for instance each sample from $\mu \in M_{<\infty}$ may correspond to a different open interval about which the MGF is finite.

For now, let us assume that μ is deterministic and state the classical uniqueness results for comparison:

Proposition 3.47. *Let* μ, ν *be deterministic probability measures on* $\mathbb{R}$*. Then* $\mu = \nu$ *provided any of the following hold:*

1. $\phi_\mu = \phi_\nu$
2. $\mu, \nu \in P^+$ *and* $\mathcal{L}_\mu = \mathcal{L}_\nu$*.*
3. $\mu, \nu \in M_{<\infty}$ *and* $\mathcal{M}_\mu = \mathcal{M}_\nu$*.*
4. $\mu, \nu \in \mathfrak{C}$ *and* $\forall k \geq 0$, $\mu_k = \nu_k$

Let $S\phi, S\mathcal{L}, S\mathcal{M}$ denote respectively the collections of functions that arise from some μ as in case (1), (2), (3), with the value $+\infty$ possible in case (3). Let $SMOM$ denote the collection of sequences of real numbers that arise as the moments of some μ as in case (4), which we regard as a function of $k \geq 0$. On each of these spaces of functions, we use the Borel σ algebra generated by the evaluation maps. Then the content of the last theorem is that $\mu \mapsto \phi_\mu$ is a bijection from $P(\mathbb{R})$ to $S\phi$, $\mu \mapsto \mathcal{L}_\mu$ is a bijection from P^+ to $S\mathcal{L}$, $\mu \mapsto \mathcal{M}_\mu$

is a bijection from the $M_{<\infty}$ to $S\mathcal{M}$, and $\mu \mapsto (\mu_k)_{k\geq 0}$ is a bijection from $\mathfrak{C}$ to $SMOM$. Observe that all of these maps are measurable isomorphisms. It follows that

Proposition 3.48. *Let μ, ν be random probability measures. Then $\mu =_d \nu$ provided any of the following hold:*

1. $(\phi_\mu(t))_{t\in\mathbb{R}} =_d (\phi_\nu(t))_{t\in\mathbb{R}}$
2. $\mu, \nu \in P^+$ *and* $(\mathcal{L}_\mu(s))_{s\geq 0} =_d (\mathcal{L}_\nu(s))_{s\geq 0}$.
3. $\mu, \nu \in M_{<\infty}$ *and* $(\mathcal{M}_\mu(s))_(s))_{s\in\mathbb{R}} =_d (\mathcal{M}_\nu(s))_{s\in\mathbb{R}}$ *(which may be valued ∞ for some s and some sample points.)*
4. $\mu, \nu \in \mathfrak{C}$ *and* $(\mu_k)_{k\geq 0} =_d (\nu_k)_{k\geq 0}$

To summarize, under suitable conditions, knowing the characteristic function, laplace transform, MGF, or moments jointly tells us the joint distribution of a random measure. It is natural to ask what happens when this type of information is only known marginally. (We call these the marginal problems, as opposed to joint.) Actually, we can already provide an answer to 2 of these problems. One counterexample that is uniformly bounded will simultaneously witness the failure of both statements.

Theorem 3.49. *There exist μ, ν uniformly bounded random measures in P^+ such that $\mu \neq_d \nu$ while yet $\forall s \geq 0$, $\mathcal{L}_\mu(s) =_d \mathcal{L}_\nu(s)$ and $\forall s \in \mathbb{R}$, $\mathcal{M}_\mu(s) =_d \mathcal{M}_\nu(s)$.*

Proof. By Proposition 3.23, we obtain random variables $U, V \in T_4$ with $U \neq_d V$ such that $\forall y \in \mathbb{R}$, $c_4(y) \cdot U =_d c_4(y) \cdot V$. Define random probability measures μ, ν supported in $\{0, 1, 2, 3\}$ via $\mu = U_0\delta_0 + U_1\delta_1 + U_2\delta_2 + U_3\delta_3$ and $\nu = V_0\delta_0 + V_1\delta_1 + V_2\delta_2 + V_3\delta_3$. Upon calculating the transforms of these random measures, the fact that $\forall y \in \mathbb{R}$, $c_4(y) \cdot U =_d c_4(y) \cdot V$ translates into the fact that the (3') holds, and thus that (2') does as well. □

It is not difficult to believe that the moment problem will require a different argument to handle in the marginal case, but it may be surprising that the characteristic function marginal problem could not be handled in the above proof. That is because the characteristic function is a complex valued random variable. Indeed, we have been avoiding this situation partly because the methods of the last two sections cannot deal with complex valued uniqueness problems, and partly because we did not have to since the problem had already been solved in the negative, and special subclasses of interest already were nonnegative anyway. However, here we arrive at a case where the signed question has not been solved yet, and our methods with some adjustment will actually be applicable. We will present a proof of nonuniqueness for the marginal characteristic function case in Subsection 3.4.1. From this proof, it will be plausible that the arithmetic dependencies of the allowed support set is the culprit, because it reduces the analysis to one of polynomials, for which we will be able to use Lemma 3.22.

In Subsection 3.4.2 we will present a proof of nonuniqueness in the moment marginal case, even if the restriction is made to a finite support set, hence to the uniformly bounded hypothesis. In Subsection 3.4.3, we will show how the relevant arithmetic structure from Subsection 3.4.2 was multiplicative, and that therefore if the set of allowed values consists of coprime numbers, then uniqueness of the distribution of a random measure with a given set of moments does hold.

We display our results concerning random uniqueness problems in the following table. We remind the reader that the assumption for the MGF case is finite MGF in a neighborhood around 0 (a.s.), the assumption for Laplace transform is nonnegativity, there are no assumptions for characteristic function, and the Carleman condition is assumed to hold (a.s.) for the moment problem. The entries "yes" and "no" refer to whether or not uniqueness holds. All answers "no" come with a uniformly bounded (pair of) counterexamples.

	joint	marginal
MGF	yes	no
Nonnegative Laplace Transform	yes	no
Characteristic Function	yes	no
Moment Problem	yes	no

3.4.1 Uniqueness Fails for Characteristic Function Problem

In this subsection, we will use $z = x + iy$ to denote a complex number. We define $\forall n \geq 0$, $P_n(x, y) = \Re(z^n)$ and $Q_n(x, y) = \Im(z^n)$ which are homogeneous polynomials in the two variables x, y of degree n. By default, our polynomials will be defined on a Euclidean space $\mathbb{R}^d$ where $d \geq 1$ is the number of variables of the polynomial. We will make use of the fact that the dimension of the vector space of degree l, N-variable homogeneous polynomials with coefficients in $\mathbb{R}$ is $\binom{N+l-1}{l}$. This fact can be derived from recasting the counting problem as sorting l balls into N bins, which is equivalent to inserting $N-1$ partition barriers between l balls arranged in a line. There are thus $N+l-1$ objects total, all $N-1$ partition barriers are equivalent and all l balls are equivalent. The total number of ways of arranging these objects is $\binom{N+l-1}{l}$.

Moving on to the main program of this subsection, we first we need a lemma.

Lemma 3.50. *For $N, l \in \mathbb{N}$ large enough, there exists p, a nonzero degree l homogeneous polynomial in N variables such that $\forall s, t, x, y \in \mathbb{R}$ we have*

$$p(sP_0(x, y) + tQ_0(x, y), \ldots, sP_{N-1}(x, y) + tQ_{N-1}(x, y)) = 0. \tag{3.69}$$

Proof. We first note that if some N, l satisfies the conditions of the lemma, then all greater pairs would work as well, so our use of the phrase "large enough" is justified. We now only need to find one pair N, l for which (3.69) holds.

For any l, N, let $S_{l,N}$ denote the real vector space of polynomials in s, t, x, y of degree at most lN and let $T_{l,N}$ denote the real vector space of degree l homogeneous polynomials in N (commuting) variables.

Observe that

$$p(sP_0(x, y) + tQ_0(x, y), \ldots, sP_{N-1}(x, y) + tQ_{N-1}(x, y)) \tag{3.70}$$

defines a polynomial (not necessarily homogeneous) of degree at most lN in the 4 variables s, t, x, y. Define the corresponding evaluation map

$$\Phi : T_{l,N} \to S_{l,N} \tag{3.71}$$

via

$$\Phi(p) = p(sP_0(x,y) + tQ_0(x,y), \ldots, sP_{N-1}(x,y) + tQ_{N-1}(x,y)). \tag{3.72}$$

Observe that Φ is linear. The dimension of its codomain is $\sum_{j=0}^{lN} \binom{j+3}{3} \leq (lN+4)^4$. If we restrict to $l = N$ then the dimension of the domain is $\binom{2N-1}{N}$.

Therefore, for $N = l$ large enough, the kernel of Φ is nontrivial.

□

As before, we will construct a precursor to the counterexample by constructing merely bounded random vectors with the desired property, and then we will adjust them so as to turn them into random probability vectors. The following can be thought of as the complex analog of a part of the argument used in the proof of Proposition 3.23. The equality in distributions are meant for complex-valued random variables.

Lemma 3.51. *For N large enough, there exist*

$$U = (U_0, \ldots, U_{N-1}), V = (V_0, \ldots, V_{N-1}) \tag{3.73}$$

bounded random vectors taking values in $\mathbb{R}^N$ for which $\forall z \in \mathbb{C}$ we have

$$\sum_{j=0}^{N-1} z^j U_j =_d \sum_{j=0}^{N-1} z^j V_j \tag{3.74}$$

while yet $U \neq_d V$

Proof. Take N large enough so that the set

$$\{(sP_0(x,y)+tQ_0(x,y),\dots,sP_{N-1}(x,y)+tQ_{N-1}(x,y))\in\mathbb{R}^N|s,t,x,y\in\mathbb{R}\} \tag{3.75}$$

is contained in a projective variety. This is possible by the last lemma. By Lemma 3.22 we may find U,V bounded random vectors so that $\forall s,t,x,y$ we have

$$\begin{aligned}&(sP_0(x,y)+tQ_0(x,y),\dots,sP_{N-1}(x,y)+tQ_{N-1}(x,y))\cdot U\\ =_d\ &(sP_0(x,y)+tQ_0(x,y),\dots,sP_{N-1}(x,y)+tQ_{N-1}(x,y))\cdot V\end{aligned} \tag{3.76}$$

while yet $U\neq_d V$.

By the Cramér Wold device, we have that $\forall x,y\in\mathbb{R}$, the $\mathbb{R}^2$-valued random vectors

$$\big((P_0(x,y),\dots,P_{N-1}(x,y))\cdot U,(Q_0(x,y),\dots,Q_{N-1}(x,y))\cdot U\big) \tag{3.77}$$

and

$$\big((P_0(x,y),\dots,P_{N-1}(x,y))\cdot V,(Q_0(x,y),\dots,Q_{N-1}(x,y))\cdot V\big) \tag{3.78}$$

have the same distribution. But by real isomorphism of $\mathbb{R}^2$ with $\mathbb{C}$, this is the same as saying that

$$\begin{aligned}&(P_0(x,y),\dots,P_{N-1}(x,y))\cdot U+i((Q_0(x,y),\dots,Q_{N-1}(x-y))\cdot U)\\ =_d\ &(P_0(x,y),\dots,P_{N-1}(x,y))\cdot V+i((Q_0(x,y),\dots,Q_{N-1}(x,y))\cdot V)\end{aligned} \tag{3.79}$$

But by the definition of the P_n,Q_n this is the same as saying $\forall z\in\mathbb{C}$ we have

$$(z^0,\dots,z^{N-1})\cdot U=_d(z^0,\dots,z^{N-1})\cdot V \tag{3.80}$$

as complex-valued random variables. Thus, U and V have the required properties. □

Lemma 3.52. *There is N' large enough so that there exist*

$$U' = (U'_0, \ldots, U'_{N-1}) \neq_d V' = (V'_0, \ldots, V'_{N-1}), \tag{3.81}$$

both valued in

$$T_{N'} := \{(x_0, \ldots, x_{N'-1}) | \forall j : 0 \leq j \leq N'-1,\, x_j \geq 0, \sum_{j=0}^{N'-1} x_j = 1\}, \tag{3.82}$$

for which $\forall z \in \mathbb{C}$

$$\sum_{j=0}^{N'-1} z^j U'_j =_d \sum_{j=0}^{N'-1} z^j V'_j. \tag{3.83}$$

Proof. Take N, U, V from the last lemma. Set $N' = N + 1$. Notice that (3.74) still holds if we rescale or translate U, V in the same way. Consider $H := \{(x_0, \ldots, x_{N-1}) | 1 \geq x_0 \geq \cdots \geq x_{N-1} \geq 0\}$. Since H has nonempty interior, we may take $U, V \in H$ without loss of generality. Then when we define $U_{-1} = 1 = V_{-1}, U_N = 0 = V_N$, and $\forall j : 0 \leq j \leq N$, $U'_j = U_{j-1} - U_j, V'_j = V_{j-1} - V_j$, observe that U', V' are $T_{N'}$ valued. Furthermore, $U' \neq_d V'$. A calculation shows that we have $\forall z \in \mathbb{C}$

$$\begin{aligned} &(z^0, \ldots, z^{N-1}) \cdot (U'_1 + \cdots + U'_N, U'_2 + \cdots + U'_N, \ldots, U'_N) \\ =_d\, &(z^0, \ldots, z^{N-1}) \cdot (V'_1 + \cdots + V'_N, V'_2 + \cdots + V'_N, \ldots, V'_N). \end{aligned} \tag{3.84}$$

From this it follows that

$$\begin{aligned} &(z^0)U'_1 + (z^0 + z^1)U'_2 + \cdots + (z^0 + \cdots + z^{N-1})U'_N \\ =_d\, &(z^0)V'_1 + (z^0 + z^1)V'_2 + \cdots + (z^0 + \cdots + z^{N-1})V'_N. \end{aligned} \tag{3.85}$$

By multiplying by $(z-1)$, using the definition of $T_{N'}$, and adding 1 to both sides, we obtain $\forall z \in \mathbb{C}$

$$\sum_{j=0}^{N'-1} z^j U'_j =_d \sum_{j=0}^{N'-1} z^j V'_j \tag{3.86}$$

Thus, U' and V' have the required properties. □

We use the homeomorphism, hence measurable isomorphism of $T_{N'}$ with the collection of probability measures supported in $\{0, \ldots, N'-1\}$ given by $v \mapsto v_0\delta_0 + \cdots + v_{N'-1}\delta_{N'-1}$ in order to see the following:

Theorem 3.53. *For N' large enough, there are two random probability measures μ, ν with distinct distributions which are a.s. supported in $\{0, \ldots, N'-1\}$ for which $\forall t \in \mathbb{R}$, $\phi_\mu(t) =_d \phi_\nu(t)$.*

Proof. Obtain U', V', N' as in the last lemma, and set $\mu = U'_0\delta_0 + \cdots + U'_{N'-1}\delta_{N'-1}$ and $\nu = V'_0\delta_0 + \cdots + V'_{N'-1}\delta_{N'-1}$. Set $z = e^{it}$ in (3.83) so that its left side is the random characteristic function of μ and its right side is the random characteristic function of ν. This construction has the required properties.

□

3.4.2 Too Many Multiplicative Relationships Spoils Uniqueness of the Marginal Moment Problem

Let us see that 4 values is again enough to prove nonuniqueness in the marginal version of the random moment problem:

Theorem 3.54. *There exist μ, ν random measures a.s. supported in $\{1, 2, 4, 8\}$ for which $\forall k \geq 0$, $\mu_k =_d \nu_k$ while yet $\mu \neq_d \nu$.*

Proof. By the identification of T_4 with the collection of probability measures supported in $\{1, 2, 4, 8\}$ (given by $v \mapsto v_0\delta_1 + v_1\delta_2 + v_2\delta_4 + v_3\delta_8$) we have that it suffices to find two random probability vectors $U = (U_0, U_1, U_2, U_3), V = (V_0, V_1, V_2, V_3)$ for which $U \neq_d V$ and $\forall k \geq 0$

$$\forall k \geq 0,\ U_0 * 1^k + U_1 2^k + U_2 4^k + U_3 8^k =_d V_0 * 1^k + V_1 2^k + V_2 4^k + V_3 8^k. \tag{3.87}$$

That this suffices is because upon defining $\mu = U_0\delta_1 + U_1\delta_2 + U_2\delta_4 + U_3\delta_8$ and $\nu = V_0\delta_1 + V_1\delta_2 + V_2\delta_4 + V_3\delta_8$, which are random probability measures a.s. supported on $\{1, 2, 4, 8\}$, we would have that the left side of (3.87) is μ_k and the right side is ν_k.

We may write (3.87) as

$$\forall k \geq 0,\ U_0 * 2^{0k} + U_1 2^{1k} + U_2 2^{2k} + U_3 2^{3k} =_d V_0 * 2^{0k} + V_1 2^{1k} + V_2 2^{2k} + V_3 2^{3k}. \tag{3.88}$$

Pick U, V from Lemma 3.23.

□

3.4.3 How The Right Kind of Arithmetic Independence Can Help

Let $a_0, \dots, a_{N-1}$ be a list of nonzero natural numbers which are pairwise coprime.

Theorem 3.55. *If μ, ν are random probability measures which are a.s. supported in the set $\{a_0, \dots, a_{N-1}\}$ and $\forall k \geq 0$, $\mu_k =_d \nu_k$ then $\mu =_d \nu$.*

Proof. First we consider the homeomorphism between the space of probability measures supported on $\{a_0, \dots, a_{N-1}\}$ with $T_N := \{(x_0, \dots, x_{N-1}) | \forall j : 0 \leq j \leq N-1,\ x_j \geq 0, \sum_{j=0}^{N-1} x_j = 1\}$ given by $\theta \mapsto (\theta(\{a_0\}), \dots, \theta(\{a_{N-1}\}))$. It suffices to show that if U, V are T_N valued and $\forall k \geq 0$ we have

$$\sum_{j=0}^{N-1} (a_j)^k U_j =_d \sum_{j=0}^{N-1} (a_j)^k V_j \tag{3.89}$$

then $U =_d V$. Once this is shown, if μ, ν are a.s. supported on $\{a_0, \dots, a_{N-1}\}$, then upon defining

$$U = \Big(U_0, \dots, U_{N-1}\Big) = \Big(\mu(\{a_0\}), \dots, \mu(\{a_{N-1}\})\Big) \tag{3.90}$$

and

$$V = \Big(V_0, \ldots, V_{N-1}\Big) = \Big(\nu(\{a_0\}), \ldots, \nu(\{a_{N-1}\})\Big) \tag{3.91}$$

we would have that the left side of (3.89) is μ_k and the right side is ν_k.

By Lemma 3.22, it suffices to show that $\{(a_0^k, \ldots, a_{N-1}^k) | k \geq 0\}$ is not contained in any projective variety.

Suppose to the contrary that there exists some nonzero homogeneous polynomial p in N variables of degree $l > 0$ for which $\forall k \geq 0$, $p(a_0^k, \ldots, a_{N-1}^k) = 0$. Enumerate (without repetitions) the monomials $\{M_i\}_{j=1}^K$ of degree l, and write $p = \sum_{j=1}^K b_j M_j$ where $b_j \in \mathbb{R}$. For each j, $\exists C_j \in \mathbb{N}$ such that

$$\forall k \geq 0,\ M_j(a_0^k, \ldots, a_{N-1}^k) = (C_j)^k. \tag{3.92}$$

Because the M_j are distinct, and $a_0, \ldots, a_{N-1}$ are all coprime, we find that all the C_j are distinct. Find j_0 such that C_j is the largest, subject to the constraint that $b_j \neq 0$. This is a feasible optimization problem because $p \neq 0$. There will be only one optimal solution because the C_j are all distinct, and a solution exists because the list of C_j is finite. Consider $\lim_k p(a_0^k, \ldots, a_{N-1}^k)/C_{j_0}^k = b_{j_0} \neq 0$ for a contradiction with the fact that $\forall k \geq 0$, $p(a_0^k, \ldots, a_{N-1}^k)/C_{j_0}^k = 0$. This completes the proof.

□

CHAPTER 4

Long Range Percolation

Many real life and mathematical situations are modeled by random networks. For example, it is often stated that any two people are six handshakes apart. We would like to use a model to assess if this is true. We may model this by assuming that people have friends who are randomly selected, but more likely to be from their immediate locale than from halfway across the world. In other words, our model of such a problem could consist of a random graph, where the vertices represent people, and there is an edge for every friendship, or handshake. Edges are more likely to occur between nearby points than points separated by a great distance. The question of whether or not everybody is within six handshakes of one another, it turns out, depends on how quickly the friendship probability decays with geographical distance. The following work considers an intermediate rate of decay that does not quite lead to an upper bound on the number of handshakes between any two people, but rather a polylogarithmic (in their geographical distance) bound.

Another situation is network theory. Suppose that we have computers communicating to one another, and sometimes, a signal has to be sent between many intermediate computers before it reaches its final destination. But some communications are lost (at random) because of various kinds of failures of components or software. What is the most efficient path that a signal should travel? An analogous question could be asked for fluids flowing through pipes, with a probability of blockage.

Let us think of these two case studies as instances of the following abstract setting, called percolation. Fix a graph G. Percolation is the study of the connectivity of random subgraphs of G obtained by keeping all the vertices, and randomly keeping or discarding each edge independently from every other edge. Sometimes, a kept edge is referred to as "open" or "present." The word "open" refers to the fact that percolation models a fluid flowing through a random network where edges that are passable to the fluid are "open." Specific questions of interest in percolation theory concern the size of clusters, the (random) notions of distance, and other graph theoretic concepts. For a general survey of percolation, see [10].

Fix an integer $d \geq 1$. We will discuss a discrete problem first and then its continuous analog. In our case, we are dealing with G being the complete graph (no loops) on the set of vertices $\mathbb{Z}^d$. For edges $\langle x, y \rangle$ not between nearest neighbors, the probability that it is open is given by

$$p_{xy,\text{ dis}} = \min(\rho/|x-y|^s, 1) \tag{4.1}$$

for some $\rho > 0, s \in (d, 2d)$. Unqualified norm and modulus symbols refer to the Euclidean norm. If x, y are nearest neighbors, then the edge between them is open a.s. We will use

$$\gamma = \frac{s}{2d} \tag{4.2}$$

and

$$\Delta = (\log_2(1/\gamma))^{-1} \in (1, \infty). \tag{4.3}$$

We will use $D_{\text{dis}}(x, y)$ to mean the graph distance from $x \in \mathbb{Z}^d$ to $y \in \mathbb{Z}^d$ restricted to the subgraph imposed by a sampling from the percolation problem. This is often called the chemical distance.

The reason for the restriction $s \in (d, 2d)$ is that if $s < d$, then there exists a constant C such that $\forall x, y$ we have $D_{\text{dis}}(x, y) \leq C$. If $s > 2d$, then the problem resembles nearest neighbor percolation, where the asymptotic behavior of $D_{\text{dis}}(x, y)$ is linear in the underlying separation. As noted earlier, the selected condition on s, the rate of decay of the probability an edge is open, is intermediate in that $D_{\text{dis}}(x, y)$ scales polylogarithmically with $|x-y|$. For a summary of the cases outside of our selected range of s, see [3].

Sometimes, we will actually deal with a continuum percolation model instead, and then we eliminate the discrete subscripts. A broad overview of the setup is that edges appear randomly between any two $x, y \in \mathbb{R}^d$, according to a Poisson Point Process, and a walker trying to go from point a to point b seeks the quickest possible path. To travel from x to y requires $|x - y|$ time if it is done in the Euclidean space, but if it is done via a long range edge, it requires 1 unit of time. Let us make this more precise in what follows. Let Q be a Poisson Point Process on $\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \big| |x| > |y|\}$ with intensity measure $\frac{\rho}{|x-y|^s} dxdy$. We think of Q as a random subset of $\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \big| |x| > |y|\}$. Define P, a random subset of $\mathbb{R}^d \times \mathbb{R}^d$, by the rule $(x, y) \in P$ if and only if either $(x, y) \in Q$ or $(y, x) \in Q$. We think of P as a random set of unoriented edges. We chose the set $\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \big| |x| > |y|\}$ because this set, together with its image under the swapping of x and y coordinates, covers $\mathbb{R}^d \times \mathbb{R}^d$ except a null set under the intensity measure. We shall refer to P as a symmetrized Poisson Point Process with intensity measure $\frac{\rho}{|x-y|^s} dxdy$. For measurability reasons and for simplicity, we shall assume that any Poisson Point Processes mentioned in the sequel are exactly countably additive, not just almost surely. We define $\forall x, y \in \mathbb{R}^d$,

$$D(x, y) = \inf_{n \geq 0, x_1, \ldots, x_n, y_1, \ldots, y_n : \langle x_1, y_1 \rangle, \ldots, \langle x_n, y_n \rangle \in P} \{n + f(n, x, y, x_1, \ldots, x_n, y_1, \ldots, y_n)\} \tag{4.4}$$

where

$$f(0, x, y) = |y - x| \tag{4.5}$$

and $\forall n \geq 1$,

$$f(n, x, y, x_1, \ldots, x_n, y_1, \ldots, y_n) = |x - x_1| + |y - y_n| + \sum_{j=1}^{n-1} |x_{j+1} - y_j|. \tag{4.6}$$

Notice that although there could be infinitely many edges in a finite ball that connect points that are separated by less than one distance in Euclidean norm, such edges don't matter because an optimal path could never use them. We call the reader's attention to the fact that the underlying norm implied by the presence of all the nearest neighbor edges in the discrete case is the l^1 norm on $\mathbb{Z}^d$, whereas we are using the Euclidean norm in the continuum model for spherical symmetry. The distribution of $D(x, y)$ is translation and rotation invariant, and $D(x, y) \leq |x - y|$.

Considering the paths allowed by the infimum in the definition of D, it is possible for two balls that are separated by a considerable distance to have pairs of points where the shortest paths between those points each intersect the other ball. Thus, we use a restriction in order to guarantee independence in that setting. Define a restricted notion of the (continuum) distance D_{res} given by

$$D_{\text{res}}(x,y) = \inf_{n\geq 0, x_1,\ldots,x_n,y_1,\ldots,y_n \in B} \{n + f(n,x,y,x_1,\ldots,x_n,y_1,\ldots,y_n)\} \tag{4.7}$$

where the infimum is taken over all paths whose long-distance edges are entirely contained within the closed ball (all balls are closed here) B of radius $10|x-y|$ centered at x.

In [2] and [3], Biskup makes use of a path construction algorithm that proceeds down a binary hierarchy to furnish a path of the desired polylogarithmic asymptotics. Roughly speaking, given x and y that are to be connected by some path, the gap from x to y can be considered one gap. At each stage in the algorithm, for each gap from a to b, with overwhelming likelihood, an edge is found that is close to connecting a to b, but its two endpoints are slightly different from a and b, consuming the gap from a to b, but leaving behind two gaps. This algorithm is continued, with each gap leaving behind two smaller order gaps, until the probability of failure cannot be maintained close to 0. We would like to refine this strategy in the following lemma, stating it in a form that is more amenable to moment estimates, which we use in the sequel.

Lemma 4.1. *$\forall \alpha, \beta \in (0,1)$ such that $d\alpha + d\beta = s$, $\forall x \in \mathbb{R}^d$, there is an event $A_{x,\alpha,\beta}$ such that*

$$D_{res}(0,x) \leq_d D_{res}(0, Z_{x,\alpha}|x|^{\alpha}) + D'_{res}(0, Z'_{x,\beta}|x|^{\beta}) + 1 + |x|1_{A_{x,\alpha,\beta}}, \tag{4.8}$$

where on the right side, the entire families D_{res}, D'_{res} together with the two single random variables $Z_{x,\alpha}$ and $Z'_{x,\beta}$ are independent as a family of four objects. $Z_{x,\alpha}$ and $Z'_{x,\beta}$ have continuous densities that are nowhere 0. D', D are identically distributed and $P(A_{x,\alpha,\beta}) \leq k_1 e^{-k_2|x|^{d-s/2}}$ with k_1 and k_2 depending on α, β, s, d only. In the special case $\alpha = \gamma = \beta$ we have that $Z_{x,\gamma} =_d Z'_{x,\gamma}$ and the common distribution does not depend on x.

Remark 4.2. *Often times not all subscripts will be explicitly stated.*

The reader should consider that (4.8) is plausible because it effectively says that one special kind of path among the many possible is the kind of path that is formed by the following: take a long edge connecting a small neighborhood of 0 to a small neighborhood of x. This long edge exists with high probability, similarly to [2] and [3]. On the right side of (4.8) then, the two distances represent the distances from 0 and x to the endpoints of the long edge that reside in their neighborhoods. These are analogous to the gaps mentioned above. The reason there are Z's in the distances' arguments is that the endpoints of the long edge are random.

Proof. The outline of the proof is that we will define an optimization game, a decomposition of the intensity measure of P, the notion of an admissible path and the error event A_x. Lastly, we will use these to prove the lemma. The idea of an admissible path makes precise the kind of special path mentioned in the last remark.

First we define the optimization game and its associated random variables. We will have need for a special norm on $\mathbb{R}^d \times \mathbb{R}^d$, defined by

$$N(x,y) = (|x|^{2d} + |y|^{2d})^{\frac{1}{2d}}. \tag{4.9}$$

Let $\alpha, \beta \in (0,1)$ be such that $d\alpha + d\beta = s$. Assume $x \in \mathbb{R}^d$ is large enough so that both $|x|^{(\alpha+1)/2}$ and $|x|^{(\beta+1)/2}$ are bounded by $|x|/100$. Consider a symmetrized Poisson Point process P' like above P, except P' has intensity with a constant density $\rho(98|x|/100)^{-s}$. (x is fixed already.) Among all the present edges in this process P', we select the edge from x_1 to x_2 that minimizes $N(x_1 - 0, x_2 - x)$. Notice that reversing x_1 and x_2 may change $N(x_1 - 0, x_2 - x)$, so the optimization problem involves, among other things, choosing the correct orientation. This defines x_1 and x_2 almost surely. Thus, $X_{1,x}$ and $X_{2,x}$ are defined as the random variables that take the value x_1 and x_2 arising from this optimization respectively. The joint density of $X_{1,x}$ and $X_{2,x}$ as a function of two variables (which we also call x_1 and x_2) is

$$C_1\rho(98|x|/100)^{-s}e^{-C_2\rho(98|x|/100)^{-s}\left((|x_1-0|^{2d}+|x_2-x|^{2d})^{\frac{1}{2d}}\right)^{2d}} \tag{4.10}$$

thanks to our choice of N, for some constants C_1, C_2 that do not depend on x. This shows that $X_{1,x}$ and $X_{2,x}$ are independent. Notice that the distributions of $X_{1,x}$ and $X_{2,x} - x$ are spherically symmetric because the intensity measure is of constant density. Let us define $Z_{1,x,\alpha} = X_{1,x}/|x|^\alpha$ and $Z_{2,x,\beta} = (X_{2,x} - x)/|x|^\beta$. The two Zs are independent of one another. Sometimes we will replace the 1 or 2 subscripts with unprimed or primed superscripts respectively. Notice that in the special case of $\alpha = \gamma$ and therefore $\beta = \gamma$, we have that the distribution of the Zs are both the same, and also the same as x varies. In order to see this, it suffices to show that $Z_{x/|x|} =_d Z_x$. Observe that the marginal distribution of $X_{1,x}$, say, has density

$$\sqrt{C_1\rho(98|x|/100)^{-s}}e^{C_2\rho(98|x|/100)^{-s}|x_1|^{2d}}. \tag{4.11}$$

This means that for B Borel, we have

$$P(Z_x \in B) = \int_{|x|^\gamma B} \sqrt{C_1\rho(98|x|/100)^{-s}}e^{C_2\rho(98|x|/100)^{-s}|x_1|^{2d}}dx_1. \tag{4.12}$$

Applying the change of variables $y = x_1/|x|^\gamma$, we obtain

$$P(Z_x \in B) = \int_B \sqrt{C_1\rho(98|x|/100)^{-s}}e^{C_2\rho(98|x|/100)^{-s}|x|^{2d\gamma}|y|^{2d}}|x|^{\gamma d}dy. \tag{4.13}$$

After cancellation of some factors, we obtain

$$P(Z_x \in B) = \int_B \sqrt{C_1\rho(98/100)^{-s}}e^{C_2\rho(98/100)^{-s}|y|^{2d}}dy = P(Z_{x/|x|} \in B). \tag{4.14}$$

Observe the $|Z|$s have finite moment generating functions on the entire real line, and the Zs have densities that are everywhere nonzero and continuous.

Next we decompose the intensity measure for P into two pieces, which we think of as a Poisson thinning. Consider $\alpha, \beta \in (0,1)$ such that $d\alpha + d\beta = s$ and fixed x, y distant enough such that $|x-y|^{(1+\alpha)/2} \le |x-y|/100$, $|x-y| \ge 1$ and $|x-y|^{(1+\beta)/2} \le |x-y|/100$. Let us do a Poisson thinning in $B_x \times B_y$ where B_x is the ball of radius $|x-y|^{(1+\alpha)/2}$ around x and B_y is the ball of radius $|x-y|^{(1+\beta)/2}$ around y, considering the intensity to be the sum of $\rho(98|x-y|/100)^{-s}$ and the rest. (If necessary, change probability spaces to enable this thinning.) Let us say that an edge that is present in the $\rho(98|x-y|/100)^{-s}$ sample is red and the rest are blue. The notions of red edge and blue edge are relative to a choice of x and y, or often, relative to the choice of 0 and a choice of x.

Let us call a path $x, x_1, y_1, \ldots, x_n, y_n, y$ admissible for $\alpha, \beta \in (0,1)$ such that $d\alpha + d\beta = s$ if x and y are sufficiently distant as above, and the path has a red edge from B_x to B_y, say this edge goes between x_j and y_j. We then require that x_j is in B_x and y_j is in B_y. We also require that in the ordered list $x_1, y_1, x_2, y_2, \ldots x_n, y_n$, all elements to the left of x_j, inclusive, are in the ball centered at x of radius $10|x-y|^{(1+\alpha)/2}$, and the rest are in the ball centered at y of radius $10|x-y|^{(1+\beta)/2}$. Observe that admissible paths are allowed in the infimum in the definition of D_{res}.

$A_{\alpha,\beta,x}$ is the event that there are no admissible paths from 0 to x for the given α, β, or the condition that $|x|^{(1+\alpha)/2} > |x|/100$ or $|x| < 1$ or $|x|^{(1+\beta)/2} > |x|/100$. All conditions but the first are deterministic, and those deterministic conditions ensure x is large enough that all balls mentioned in the above paragraph centered at 0 are disjoint from those centered at x. The requirements also ensure that the Poisson intensity between the two balls of radii $|x-0|^{(1+\alpha)/2}$ around 0 and $|x-0|^{(1+\beta)/2}$ around x has intensity at least $\rho(98|x-0|/100)^{-s}dxdy$. Observe that $P(A_x) \leq k_1 e^{-k_2|x|^{d-s/2}}$. This is because the volume of $B_0 \times B_x$ is proportional to $|x|^{d+s/2}$ and the intensity measure for P' is proportional to $|x|^{-s}$.

With all the setup, we have thus ensured that $\forall \alpha, \beta \in (0,1)$ such that $d\alpha + d\beta = s$, $\forall x \in \mathbb{R}^d$,

$$D_{\text{res}}(0,x) \leq 1_{A^c} D_{\text{res}}(0,x) + |x| 1_A \tag{4.15}$$

and

$$1_{A^c} D_{\text{res}}(0,x) \leq 1_{A^c}(D_{\text{res}}(0, Y_{1,x}) + D'_{\text{res}}(x, Y_{2,x}) + 1). \tag{4.16}$$

where $Y_{1,x}, Y_{2,x}$ denotes the pair arising from the optimization game played with the intensity measure of P' on $B_0 \times B_x$ and the intensity measure of P elsewhere. Because the event A^c guarantees that this optimization occurs within $B_0 \times B_x$, $Y_{1,x}, Y_{2,x}$ can be replaced by $X_{1,x}, X_{2,x}$, with the qualification that these together with D_{res} and D'_{res} are independent. This last independence claim holds because D_{res} depends on only edges within B_0, D'_{res} only depends on edges within B_x, and the pair $(Y_{1,x}, Y_{2,x})$ depends only on edges of length of order $|x|$, which cannot possibly include any edges within B_0 nor within B_x due to the conditions in A^c on x. Therefore, when the two Ys are exchanged for the two Xs, which are independent, the coupling of the two Xs with the two D_{res}'s is also independent. Writing Z's instead of X's, we have

$$D_{\text{res}}(0,x) \leq_d D_{\text{res}}(0, Z_x|x|^{\alpha}) + D'_{\text{res}}(0, Z'_x|x|^{\beta}) + 1 + |x|1_A, \tag{4.17}$$

where on the right side, the entire families $D_{\text{res}}, D'_{\text{res}}$ together with the two single random variables Z_x and Z'_x are independent as a family of four objects. □

Other properties of the above objects that we shall use freely are:

1. The distribution of $D_{\text{res}}(0,x)$ depends only on $|x|$ and this distribution is continuous as a function of $|x|$.

2. $D_{\text{res}} \geq D$.

3. $D_{\text{res}}(x,y) \leq |x-y|$

   and

4. The distribution of D_{res} is translation invariant.

Often times, when one of the arguments of any of the objects $D, D_{\text{res}}, D_{\text{dis}}$ is 0, only the other argument is indicated. (e.g. $D(x) = D(0,x)$.) We will often make use of the following lemma as well.

Lemma 4.3. *For h being either D or D_{res}, the relations*

$$\forall c \leq 1, h(cx) \geq_d ch(x) \tag{4.18}$$

and

$$\forall c \geq 1, h(cx) \leq_d ch(x) \tag{4.19}$$

hold true.

Proof. We consider the case of $h = D$ and prove the first inequality. The rest is left to the reader. Assume $c \leq 1$. Consider a coupling of two continuum percolation processes together, where one process P_1 is according to a symmetrized Poisson Point Process with intensity measure $\frac{\rho}{|x-y|^s}dxdy$, and the second process P_2 simply takes each edge $\langle x, y\rangle$ occurring in P_1 and duplicates it as $\langle cx, cy\rangle$. Observe that every path from 0 to x in P_1 corresponds exactly with a path from 0 to cx in P_2. For the moment, let us use D_1 to mean the distance from 0 to x in P_1 and D_2 to mean the distance from 0 to cx in P_2. In both the expressions cD_1 and D_2, paths pay a cost of $c|a-b|$ to use the underlying Euclidean space to travel from point a to b in P_1, or ca to cb in the corresponding path in P_2. However, traversing an edge costs c for cD_1 and 1 for the corresponding edge in D_2. Therefore, $cD_1 \leq_d D_2$. The intensity measure for P_2 is dominated by the intensity measure for P_1 by a factor of $c^{s-2d} > 1$. From this, we deduce that $D_2 \leq_d D(cx)$. Putting this together yields the first result for $h = D$, and the rest are similar. □

Our program is to prove a lower bound for D, and then a scaling limit result for D_{res}, followed by an upper bound for D, with each of these steps using the last. However, it is most convenient to state the scaling limit result first:

Theorem 4.4. *There is a continuous function $\phi : (1, \infty) \to (0, \infty)$ such that $\phi(r) = \phi(r^\gamma)$ holds when $r > 1$, and for each $x \in \mathbb{R}^d$ nonzero, we have*

$$\frac{D_{res}(rx)}{\phi(r)log(r)^\Delta} \to 1 \tag{4.20}$$

in probability as $r \to \infty$.

Remark 4.5. *The reader is cautioned that ϕ does not include $r = 1$ in its domain.*

In [2] and [3], Biskup establishes that

$$\lim_{L\to\infty} P(\log(L)^{\Delta-\epsilon} \leq D_{\mathrm{dis}}(0, Le_1) \leq \log(L)^{\Delta+\epsilon}) = 1 \tag{4.21}$$

for each $\epsilon > 0$, where e_1 is the first standard basis vector for $\mathbb{R}^d$. We will improve this to

Theorem 4.6. *There are $c, C > 0$ such that*

$$\lim_{L\to\infty} P(c\log(L)^\Delta \leq D_{dis}(0, Le_1) \leq C\log(L)^\Delta) = 1. \tag{4.22}$$

The l^1 norm of the underlying nearest neighbor connections in the discrete case and the use of the Euclidean norm in the continuum case may strike the reader as not parallel. However, the choice of the norms used in the definition of f, and the choice between the continuum or discrete version of the problem are all immaterial, as Theorem 4.6 is equivalently valid for any choice. The validity of Theorem 4.6 is not affected by the choice of the Euclidean norm in the definition of either the intensity measure or $p_{xy,\,\mathrm{dis}}$. To see that the validity of the lower bound is unaffected by a change of the norm, compare to a model with a higher ρ if necessary. Similarly for upper bound. Also, in the continuum versions of this theorem, the validity is unaffected by whether the limit is taken over natural numbers L or real numbers L. We formalize this below in a lemma and a remark.

Lemma 4.7. *There are $\rho_u, c > 0$, and a coupling of the entire D with the entire D_{dis}, the latter with edge probabilities* $\min(\rho_u/|x-y|^s, 1)$*, such that*

$$cD_{dis}(x,y) \leq D(x,y) \tag{4.23}$$

pointwise, for every $x, y \in \mathbb{Z}^d$ with $|x-y| > 1$.

There are $\rho_l, C > 0$, and a coupling of the entire D with the entire D_{dis}, the latter with edge probabilities $\min(\rho_l/|x-y|^s, 1)$*, such that*

$$D(x,y) \leq CD_{dis}(x,y) \tag{4.24}$$

pointwise, for every $x, y \in \mathbb{Z}^d$ with $|x-y| > 1$.

Proof. We prove the first. If a path $x, x_1, y_1, \ldots, x_n, y_n, y$ is given with edges $\langle x_j, y_j, \rangle$ all present in the symmetrized Poisson Process with intensity $\rho/|x-y|^s$, then we refer to the $(x, x_1), (y_1, x_2), \ldots, (y_n, y)$ segments as gaps. An analogous definition is also used in discrete settings. We can assume for this proof, without loss of generality, that continuum gap traversal is done according to the l^1 norm. We will assume that in discrete settings, nearest neighbors are always connected. We describe below how to deal with other edges. From the continuum process, we split $\mathbb{R}^d$ into hypercubes of sidelength 1, and decree that two points are connected for the discrete process if and only if they are the lower left corners of two boxes in the grid that are connected by at least one edge in the continuum process of length at least 1. (edges of length less than 1 are never relevant anyway.) This means that the discrete process has independent edges with probability

$$p_{1,xy} = 1 - e^{-\int_{B\times B'} \frac{\rho}{|x-y|^s} 1_{|x-y|\geq 1} dxdy} \tag{4.25}$$

to be open where B and B' are two distinct hypercubes in the grid that are not nearest neighbors. We call D_1 the discrete chemical distance for this percolation with edge probabilities p_1. There exists ρ_u such that for all B, B' we have

$$1 - e^{-\int_{B\times B'} \frac{\rho}{|x-y|^s} 1_{|x-y|\geq 1} dxdy} \leq \min(\rho_u/|x-y|^s, 1) \tag{4.26}$$

where x and y are the lower left corners of the hypercubes B and B'. We call D_2 the discrete chemical distance for the percolation based on edge probabilities $p_2 = \min(\rho_u/|x-y|^s, 1)$. Since $p_1 \leq p_2$, we have $D_1 \geq_d D_2$. What remains is to compare D_1, which came with a coupling with D, to D itself.

For every path used in the infimum defining D, there is a corresponding discrete path that occurs in the infimum defining D_1. There is the issue that a gap may be much shorter than 1 unit of distance, but span multiple hypercubes. In any case, each gap can be overestimated by the discrete distance by no more than d units . If n edges are used, then there are $n+1$ gaps. Notice that if $n = 0$ then the number of gaps is 1 and that gap has length greater than 1, and this consideration leads to the need for an extra factor of $1/(d+1)$ in c. If $n > 0$, then the number of gaps is no more than double the number of long range edges. Therefore, this consideration inserts a factor of at most $1/(3(d+1))$ in c. Therefore, $D \geq cD_1 \geq_d cD_2$ with $c = 1/(3(d+1))$, which completes the proof of the first inequality

For the second inequality, the proof is similar, with the additional following detail when comparing D to D_1. In the continuum distance, there is the need to locally connect two endpoints of edges in the same hypercube. Thus, the continuum distance might involve paths that overestimate their discrete counterparts by no more than d per gap. As before, this is accounted for in the constant C, whether or not $n = 0$ where n is the number of gaps. $\square$

Remark 4.8. *Changes of the choice of norms in any of the definitions correspond to at worst a change in* ρ *(if the norm in the intensity density or edge probability is changed) or a constant multiplier on the chemical distance (if the norm for the underlying short ranged travel of gaps is changed). Neither type of change affects the existence of upper and lower bounds.*

4.1 The Lower Bound

We will borrow a method from [2] in order to prove the lower bound portion for the discrete problem. By the above reasoning, the lower bound will also hold for the continuum problem.

This section is copied nearly verbatim from section 3 in [2]. We will actually prove the lower bound assuming only that $p_{xy,\,\text{dis}} \leq \rho|x-y|^s$. The lower bound is a corollary of the next theorem

Theorem 4.9. *There are constants* $c_1, c_2 \in (0,\infty)$ *such that, for* $\Delta := 1/\log_2(2d/s)$, $x \neq 0$

$$P\big(D_{dis}(0,x) \leq n\big) \leq c_1 \left(\frac{e^{c_2\, n^{1/\Delta}}}{|x|}\right)^s, \qquad n \geq 1. \tag{4.27}$$

We present some lemmas:

Lemma 4.10. *Abbreviate* $B_k := B(0,k)$ *for* $k \geq 0$. *If* $k > 0$ *and* $x \in \mathbb{Z}^d$ *with* $x \neq 0$, *then*

$$P\big(D_{dis}(0,x) \leq k\big) \leq \rho\big(\frac{|x|}{k}\big)^{-s} \sum_{j=0}^{k} E|B_j|\, E|B_{k-j}| \tag{4.28}$$

Proof. If $D(0,x) \leq k$, then there exists a (vertex) self-avoiding path from 0 to x such that at least one edge has length at least $|x|/k$. If this edge occurs at the j-th step and it goes from vertex y to vertex z, then we must have $D(0,y) \leq j$ and $D(z,x) \leq k-j$. These events should actually occur spatially-disjointly. By conditioning on j and (y,z), the van den Berg-Kesten inequality (or, alternatively, a careful conditioning on the first part of the path) yields

$$P\big(D_{\text{dis}}(0,x) \leq k\big) \leq \sum_{j=1}^{k} \sum_{\substack{y,z\in\mathbb{Z}^d \\ |y-z|\geq|x|/k}} P\big(D_{\text{dis}}(0,y) \leq j\big)\, p_{yz,\,\text{dis}}\, P\big(D_{\text{dis}}(z,x) \leq k-j\big). \tag{4.29}$$

We can bound $p_{yz,\,\text{dis}} \le \rho(|x|/k)^{-s}$. Dropping the condition on $|y-z|$ we can now sum over y and z to get the right-hand side of (4.28). $\square$

Lemma 4.11. *There is an $a = a(d,s) \ge 1$ such that all $j \ge 1$, $K \ge j$ have the property that the bound $\forall x \in \mathbb{Z}^d$ such that $|x|/j \ge 1, P(D_{dis}(0,x) \le j) \le [K/|x|]^s$ implies that $E|B_j| \le aK^d$.*

Proof. Note that $|x| > K$ implies $|x|/j \ge 1$. Thus

$$E|B_j| = \sum_{x\in\mathbb{Z}^d} P\big(D_{\text{dis}}(0,x) \le j\big) \le \sum_{x\colon |x|\le K} 1 + \sum_{x\colon |x|>K} \left(\frac{K}{|x|}\right)^s. \tag{4.30}$$

It is easy to check that the first term is bounded by a constant $a_1 = a_1(d)$ times K^d, while the sum over $|x|^{-s}$ over $|x| > K$ is at most a constant $a_2 = a_2(d,s)$ times K^{d-s}. Putting these contributions together, the desired claim follows. $\square$

We need one more lemma.

Lemma 4.12. *For each $p > \frac{s+1}{2d-s}$ and each $c_0 > 0$ there is $C = C(p, c_0) \in (0,\infty)$ such that for each $c \ge c_0$ the sequence $\{K_n\}_{n\ge 0}$ given by*

$$K(n) := \frac{1}{C}(n+1)^{-p} e^{\,c\,n^{1/\Delta}}, \tag{4.31}$$

obeys

$$\rho\sum_{j=0}^{n} K(j)^d K(n-j)^d \le n^{-s} K(n)^s, \qquad n \ge 1. \tag{4.32}$$

Proof. Without loss of generality, to prove this lemma, we may assume $\rho = 1$.

Consider the function $\nu(x) := x^{1/\Delta} + (1-x)^{1/\Delta}$ and note that the exponentials in $K(j)^d K(n-j)^d$ combine into $\exp\{cn^{1/\Delta}\nu(j/n)d\}$. Note also that ν is maximized at $x := \frac{1}{2}$ where it equals $2^{1-1/\Delta} = s/d$. Let

$$\delta := s - d \max_{0\le x\le 1/4} \nu(x) \tag{4.33}$$

and observe that $\delta > 0$. Splitting the sum over j into the part when $|j - n/2| \le n/4$ or not, and using the symmetry $j \leftrightarrow n - j$ we thus get

$$\begin{aligned}\sum_{j=0}^{n} K(j)^d K(n-j)^d &\le 2 \sum_{j \le n/4} K(j)^d K(n-j)^d + \sum_{j\colon |n/2-j| \le n/4} K(j)^d K(n-j)^d \\ &\le 2 \sum_{j \le n/4} C^{-2d} \frac{e^{c\, n^{1/\Delta}(s-\delta)}}{(j+1)^{pd}(n-j+1)^{pd}} \\ &\quad + \sum_{j\colon |n/2-j| \le n/4} C^{-2d} \frac{e^{c\, n^{1/\Delta} s}}{(j+1)^{pd}(n-j+1)^{pd}}.\end{aligned} \tag{4.34}$$

Using that $j + 1 \ge (n + 1)/8$ and $n - j + 1 \ge (n + 1)/8$ for all integers j such that $|j - n/2| \le n/4$, we now get $\forall n > 0$,

$$\begin{aligned}\text{LHS of (4.32)} &\le 2(n+1)C^{-2d} e^{c\, n^{1/\Delta}(s-\delta)} + C^{-2d} 8^{2pd} (n+1)^{1-2pd} e^{c\, n^{1/\Delta} s} \\ &\le h(n)\, n^{-s} \Big[\frac{1}{C}(n+1)^{-p} e^{c\, n^{1/\Delta}}\Big]^s\end{aligned} \tag{4.35}$$

where

$$h(n) := C^{s-2d}\big(8^{2pd}(n+1)^{1-2pd} + 2(n+1)e^{-c\delta\, n^{1/\Delta}}\big)(n+1)^{s+ps}. \tag{4.36}$$

Observe that $1 - 2pd + s + ps < 0$ under the assumed condition on p and so the term multiplying C^{s-2d} is bounded uniformly in n for all $c > 0$. Given $c_0 > 0$, we can thus choose C so small that $h(n) \le 1$ holds for all $n \ge 1$ and all $c \ge c_0$. This defines $C(p, c_0)$ and proves the claim. □

Remark 4.13. *By choosing C even smaller if necessary, we may arrange for the previous lemma to be satisfied for the sequence K' given by $K'(n) = K(n)$ if $n > 0$, and $K'(0) = a^{-\frac{1-dq}{d}}$ where $q := \frac{2}{2d-s}$ and a is as in Lemma 4.11.*

Proof of Theorem 4.9. Let $p > \frac{s+1}{2d-s}$, set $q := \frac{2}{2d-s}$ and let $a = a(d, s)$ be as in Lemma 4.11. Pick $c_0 > 0$ and let $C(p, c_0)$ be as in the preceding remark. Finally, pick $c \ge c_0$ so large that

$$K(n) := \frac{1}{C(p, c_0)}(n+1)^{-p} e^{c\, n^{1/\Delta}} \ge a^q n \max\{1, \rho^{1/s}\}, \qquad n \ge 1. \tag{4.37}$$

We will show by induction that, for each $n \ge 1$, $x \ne 0$,

$$P\big(D_{\text{dis}}(0, x) \le n\big) \le \left(\frac{a^{-q} K(n)}{|x|}\right)^s. \tag{4.38}$$

Notice that this is trivially true for $|x| < a^{-q}K(n)$ and so we may thus always suppose that $|x| \geq a^{-q}K(n)$ which by (4.37) implies $|x| \geq n$.

For convenience, we define $K(0) = a^{-\frac{1-dq}{d}}$. This way $E[B_0] = 1 = a^{1-dq}K(0)^d$.

To start the induction we note that (4.38) holds for $n = 1$ as, for x away from the origin, $P(D_{\text{dis}}(0,x) \leq 1) = p_{0x,\text{ dis}}$ which is less than or equal to the right-hand side by the bound $a^{-q}K(1) \geq \rho^{1/s}$. So let us now suppose (4.38) holds for all $n \leq m \in \{1, 2, \dots\}$ and let us prove it for $n := m+1$. Notice that as we may assume $|x| \geq (m+1) \geq j$ for $j = 0, \dots, m+1$, Lemma 4.11 can be used for $E|B_j|$ with $K := a^{-q}K(j)$ for all $j = 1, \dots, m+1$. By Lemma 4.10 and Lemma 4.11 we thus get

$$P\big(D_{\text{dis}}(0,x) \leq m+1\big) \leq \rho\Big(\frac{|x|}{m+1}\Big)^{-s} a^{2-2dq} \sum_{j=0}^{m+1} K(j)^d K(m+1-j)^d. \tag{4.39}$$

The estimate even includes the $j = 0$ and $j = m+1$ terms in the sum because of the special definition of $K(0)$.

Invoking Lemma 4.12, the sum can be further bounded with the result

$$P\big(D_{\text{dis}}(0,x) \leq m+1\big) \leq a^{2-2dq}\left(\frac{K(m+1)}{|x|}\right)^s. \tag{4.40}$$

Since $2 - 2dq = -sq$, we get (4.38) for $n := m+1$. Thus (4.38) holds for all $n \geq 1$; choosing $c_1 := a^{-qs}C(p, c_0)^{-s}$ and $c_2 := c$ we then get also (4.27). □

We now prove the lower bound as a corollary.

Corollary 4.14. *There is $c > 0$ such that*

$$\lim_{L\to\infty} P(c * \log(L)^\Delta \leq D_{dis}(0, Le_1)) = 1. \tag{4.41}$$

Proof. For given x, in Theorem 4.9, plug in $n = \big\lfloor (\log(|x|)/c_2)^\Delta \big\rfloor$. □

4.2 The Scaling Limit

We will work towards a proof of Theorem 4.4.

We will prove first and second moment estimates that will be used in the sequel. At first we will only use (4.8) for $\alpha = \beta = \gamma$.

Since $Z_x = Z$ does not depend on x (distributionally) and has faster than exponentially decaying tails, we define the random variable

$$W := Z \prod_{j=1}^{\infty} |Z_j|^{\gamma^j}. \tag{4.42}$$

where the Z_j, Z are all iid.

Lemma 4.15. *The product defining W converges, W is finite, positive almost surely. W has probability density that is nonzero Lebesgue a.e. If Z and W are independent, then*

$$Z|W|^{\gamma} =_d W. \tag{4.43}$$

This last identity shows why we defined W this way. W is for absorbing copies of Z obtained from iterating (4.8).

Proof. Observe that $P(Z = 0) = 0$ since Z has a density. Taking the logarithm of the partial products defining W and applying the Kolmogorov Three Series Theorem shows that the product in W converges, and that $W \in (0, \infty)$. Since Z has continuous probability density that is nowhere zero, and W is defined as Z multiplied by a random variable that is a.s. never 0, we see that W has a density as well, and it is nonzero Lebesgue a.e. We can verify (4.43) with a calculation. $\square$

From (4.8), we derive $\forall r > 1$,

$$D_{\text{res}}(0, Wr) \leq_d D_{\text{res}}(0, |W|^{\gamma} Z r^{\gamma}) + D'_{\text{res}}(0, |W|^{\gamma} Z' r^{\gamma}) + 1 + r|W| 1_{A_{rW}} \tag{4.44}$$

where the joint distribution of the $W, Z, Z', D_{\text{res}}, D'_{\text{res}}$ is all independent, and also the entire family of A_x, though not necessarily independent of one another, are all independent from W.

Lemma 4.16. $\lim_n ED_{res}(0, Wr^{\gamma^{-n}})/2^n$ *exists. Moreover, there exist numbers $a_n \to 0$ such that $ED_{res}(0, Wr^{\gamma^{-n}})/2^n + a_n$ is a decreasing sequence. In particular, $ED_{res}(0, Wr^{\gamma^{-n}})/2^n$ is bounded uniformly in n and compact sets of r.*

Proof. Taking expectations of (4.44) with $r^{\gamma^{-n}}$ replacing r, and then dividing by 2^n, we obtain

$$ED_{\text{res}}(0, Wr^{\gamma^{-n}})/2^n \leq ED_{\text{res}}(0, Wr^{\gamma^{-n+1}})/2^{n-1} + 1/2^n + C/2^n \tag{4.45}$$

where C is a constant bound for the function $t \mapsto |t|k_1 e^{-k_2|t|^{d-s/2}}$. Setting $a_n = \sum_{j=n+1}^{\infty}(1+C)/2^n$ gives way to the decreasing property we require. Since $ED_{\text{res}}(0, Wr)$ is continuous and nonnegative, we know that $ED_{\text{res}}(0, Wr^{\gamma^{-n}})/2^n$ is bounded uniformly in n and compact sets of r. □

We define $L(r) = \lim_n ED_{\text{res}}(r^{\gamma^{-n}}W)/2^n$ which exists by Lemma 4.16. We now show a second moment estimate as part of the proof of

Theorem 4.17. *Let $r > 1$. Then for Lebesgue a.e. $x \in \mathbb{R}^d$, the limit*

$$D_{res}(r^{\gamma^{-n}}x)/2^n \to L(r) \tag{4.46}$$

in probability.

We can use (4.18) and (4.19) to derive

Corollary 4.18. *In the above, the same convergence can be claimed in probability for every x other than 0.*

Proof of Theorem 4.17. Square (4.44) to get $\forall r > 1$,

$$E(D_{\text{res}}(rW)^2) \leq 2E(D_{\text{res}}(r^\gamma W)^2) + 2E(E(D_{\text{res}}(r^\gamma Z|W|^\gamma)\big|W)^2) + F_0 \tag{4.47}$$

where for $r > 1$,

$$\begin{aligned} F_0 = F_0(r) = 1 + 4E[D_{\text{res}}(0, r|W|^\gamma Z)] \\ +E[(r|W|)^2 1_{A_{rW}}] + 2E[r|W|1_{A_{rW}}] + 4E[|rW|^{\gamma+1}|Z|1_{A_{rW}}]. \end{aligned} \tag{4.48}$$

We also define for $n > 0$,

$$F_n = F_n(r) = F_0(r^{\gamma^{-n}}). \tag{4.49}$$

Notice that each term of $F_n/2^n$ is bounded uniformly in n and compact sets of r, hence $F_n/2^n$ is bounded uniformly in n and compact sets of r, so that $F_n/4^n$ is summable uniformly on compact sets of $r > 1$. Here, we have used Lemma 4.16, the fact that W is independent of all random objects, the decay of $P(A_x)$, property 3, Cauchy Schwarz on the last term, and the faster than exponentially decaying tails of Z. We remind the reader that $A_x = A_{\gamma,\gamma,x}$ is the event that at least one of the following failure modes occur: there are no admissible paths from 0 to x for the given $\alpha = \gamma, \beta = \gamma$, or $|x|^{(1+\alpha)/2} > |x|/100$ or $|x| < 1$ or $|x|^{(1+\beta)/2} > |x|/100$.

Now we rewrite the second term using conditional variance, and then subtract suitable terms on both sides to get $\forall r > 1$

$$\begin{aligned}\mathrm{Var}(D_{\mathrm{res}}(rW)) \le 2\mathrm{Var}(D_{\mathrm{res}}(r^\gamma W)) + 2\mathrm{Var}(E(D_{\mathrm{res}}(r^\gamma Z|W|^\gamma)|W)^2) \\ + 4E(D_{\mathrm{res}}(r^\gamma W))^2 - E(D_{\mathrm{res}}(rW))^2 + F_0 \end{aligned} \quad (4.50)$$

Replacing W by $Z|W|^\gamma$ in the first two variances above and using that

$$\mathrm{Var}(X) = E(\mathrm{Var}(X|Y)) + \mathrm{Var}(E(X|Y)) \quad (4.51)$$

then gives

$$\begin{aligned}\mathrm{Var}(E(D_{\mathrm{res}}(rZ|W|^\gamma)|W)) + E(\mathrm{Var}(D_{\mathrm{res}}(rZ|W|^\gamma)|W)) \\ \le 4\mathrm{Var}(E(D_{\mathrm{res}}(r^\gamma Z|W|^\gamma)\big|W)) + 2E(\mathrm{Var}(D_{\mathrm{res}}(r^\gamma Z|W|^\gamma)|W)) \\ + 4E(D_{\mathrm{res}}(r^\gamma W))^2 - E(D_{\mathrm{res}}(rW))^2 + F_0 \end{aligned} \quad (4.52)$$

Denoting

$$\begin{aligned} A_n &:= \frac{1}{4^n}\mathrm{Var}(E(D_{\mathrm{res}}(r^{\gamma^{-n}}Z|W|^\gamma)|W)) \\ B_n &:= \frac{1}{4^n}E(\mathrm{Var}(D_{\mathrm{res}}(r^{\gamma^{-n}}Z|W|^\gamma)|W)) \\ C_n &:= \frac{1}{4^n}E(D_{\mathrm{res}}(r^{\gamma^{-n}}W))^2 \end{aligned} \quad (4.53)$$

we then get the inequality

$$A_n + B_n + C_n \le A_{n-1} + \frac{1}{2}B_{n-1} + C_{n-1} + \frac{F_n}{4^n} \quad (4.54)$$

because $F_n = F_n(r) = F_0(r^{\gamma^{-n}})$. Iterating shows

$$A_n + \frac{1}{2}B_n + C_n \leq A_0 + \frac{1}{2}B_0 + C_0 - \frac{1}{2}\sum_{k=1}^{n} B_k + \sum_{k=1}^{n} \frac{F_k}{4^k}. \tag{4.55}$$

Since all A_n, B_n, C_n are all positive and $F_n/4^n$ is summable in $n > 0$, the sum of B_k must remain bounded uniformly in n. We have thus proved:

Lemma 4.19. *We have*

$$\sum_{n=1}^{\infty} E(\mathit{Var}(2^{-n} D_{res}(r^{\gamma^{-n}} Z|W|^{\gamma})|W)) < \infty \tag{4.56}$$

We now consider again the independent copies $\tilde{D}$ and $\tilde{Z}$ of the quantities D and Z. Formula (4.56) then yields

$$\sum_{n=1}^{\infty} E\left[\left(\frac{\tilde{D}_{\text{res}}(r^{\gamma^{-n}}\tilde{Z}|W|^{\gamma})}{2^n} - \frac{D_{\text{res}}(r^{\gamma^{-n}}Z|W|^{\gamma})}{2^n}\right)^2\right] < \infty \tag{4.57}$$

Now pick a compact set $U \subset \mathbb{R}^d \setminus \{0\}$ with non-empty interior, let $\epsilon \in (0,1)$ and let us restrict the expectation to the event $\{Z|W|^{\gamma} \in U\} \cap \{\epsilon < |W| < 1/\epsilon\}$. Let $|U|$ denote the Lebesgue measure of U. From the fact that Z has a continuous nonzero density, it follows that

$$f(z)|w|^{-d\gamma} \geq c\frac{1}{|U|},\ z|w|^{\gamma} \in U,\ \epsilon < |w| < 1/\epsilon \tag{4.58}$$

for some constant $c = c(U, \epsilon) > 0$. Using this bound in the above expectation permits us to change variables to $x := z|w|^{\gamma}$ and conclude that for X uniform on U, and independent of all other random objects, we have

$$\sum_{n=1}^{\infty} E\left[\left(\frac{\tilde{D}_{\text{res}}(r^{\gamma^{-n}}\tilde{Z}|W|^{\gamma})}{2^n} - \frac{D_{\text{res}}(r^{\gamma^{-n}}X)}{2^n}\right)^2 \middle|\, \epsilon < |W| < 1/\epsilon\right] < \infty \tag{4.59}$$

where we also used that $P(\epsilon < |W| < 1/\epsilon) > 0$ for $\epsilon \in (0,1)$. Using Jensen's inequality, we can now pass the expectation over $\tilde{D}$, $\tilde{Z}$ and W inside the square to get

$$\sum_{n=1}^{\infty} E\left[\left(E\left[\frac{D_{\text{res}}(r^{\gamma^{-n}}Z|W|^{\gamma})}{2^n} \middle|\, \epsilon < |W| < 1/\epsilon\right] - \frac{D_{\text{res}}(r^{\gamma^{-n}}X)}{2^n}\right)^2\right] < \infty \tag{4.60}$$

By Monotone Convergence Theorem, this implies

$$\frac{D_{\rm res}(r^{\gamma^{-n}}X)}{2^n} - E\left[\frac{D_{\rm res}(r^{\gamma^{-n}}Z|W|^\gamma)}{2^n}\,\middle|\,\epsilon < |W| < 1/\epsilon\right] \underset{n\to\infty}{\longrightarrow} 0, \text{a.s.} \tag{4.61}$$

with the exceptional set not depending on ϵ.

The above reasoning (namely, the fact that $A_n + B_n + C_n$ is bounded) shows

$$\tilde{c} := \sup_{n\geq 1} E\left(\left(\frac{D_{\rm res}(r^{\gamma^{-n}}Z|W|^\gamma)}{2^n}\right)^2 \right) < \infty \tag{4.62}$$

and so, denoting $q_\epsilon := 1 - P(\epsilon < |W| < 1/\epsilon)$, from Cauchy-Schwarz we have

$$\left|(1-q_\epsilon)E\left[\frac{D_{\rm res}(r^{\gamma^{-n}}Z|W|^\gamma)}{2^n}\,\middle|\,\epsilon < |W| < 1/\epsilon\right] - E\left[\frac{D_{\rm res}(r^{\gamma^{-n}}W)}{2^n}\right]\right| \leq \sqrt{\tilde{c}q_\epsilon} \tag{4.63}$$

Since $q_\epsilon \to 0$ as $\epsilon \downarrow 0$, we thus conclude that, a.s. for Lebesgue a.e. $x \in U$, we have

$$\lim_{n\to\infty} \frac{D_{\rm res}(r^{\gamma^{-n}}x)}{2^n} = \lim_{n\to\infty} E\left[\frac{D_{\rm res}(r^{\gamma^{-n}}W)}{2^n}\right], \tag{4.64}$$

As this holds for any choice of U as above, the claim follows. □

In the proof of Theorem 4.17, we already provided most of the details for another useful fact, which we now formalize.

Lemma 4.20. $E[(D_{res}(r^{\gamma^{-n}}W))^2]/4^n$ *is uniformly bounded in* n *and compact sets of* $r > 1$.

Proof. From the above, we have

$$A_n + B_n + C_n \leq A_0 + B_0 + C_0 + \sum_{k=1}^\infty \frac{F_k}{4^k}. \tag{4.65}$$

A_0, B_0, C_0 are all continuous in r, and the series is uniformly convergent on compact sets of r, hence $E[(D_{\rm res}(r^{\gamma^{-n}}W))^2]/4^n = A_n + B_n + C_n$ is bounded uniformly in n and compact sets of r. □

We will need a third moment bound.

Lemma 4.21. $E[(D_{res}(r^{\gamma^{-n}}W))^3]/8^n$ *is uniformly bounded in* n *and compact sets of* $r > 1$.

Proof. By cubing (4.44) and taking conditional expectation on W and then expectation, using Holder's inequality on the cross terms on the right side, and using X_n to denote $D_{\text{res}}(r^{\gamma^{-n}}W)$, we find that

$$E[X_n^3] \leq 8E[X_{n-1}^3] + E_n. \tag{4.66}$$

Here E_n, similar to F_n, is the error term, where any term in it that involves the family A_x is uniformly bounded in r and n, and terms that do not involve the A_x's can only involve X_n up to the second moment, which are bounded uniformly on compact sets of $r > 1$ by the previous lemma. Therefore, $E_n/4^n$ is uniformly bounded on compact sets of r, much as $F_n/2^n$ is uniformly bounded on compact sets of r. Therefore, $E_n/8^n$ is uniformly summable on compact sets of r, much as $F_n/4^n$ is uniformly summable on compact sets of r. Thus, $D_{\text{res}}(r^{\gamma^{-n}}W)/2^n$ is an L^3 bounded family, namely the bound on the third moment can be taken to be $E[X_0^3] + \sum_{k=1}^{\infty} E_k/8^k$. □

For convenience,

Definition 4.22. *Fix $r > 1$. Any of the x above that admit* (4.46) *is called a* ***valid multiplier*** *for r.*

Valid multipliers are never 0.

Theorem 4.17 gives way to limits in probability along the sparse sequences $r^{\gamma^{-n}}$. We now derive some properties of $L(r)$ in order to generalize to other sequences.

Lemma 4.23. *$L(r^\gamma) = L(r)/2$, $L(r) > 0$, and $L(r)$ is continuous.*

Proof. We leave the first claim to the reader. $L(r) > 0$ for every $r > 1$ because of the lower bound established in the previous section, and because of property 2. This is the only place where the lower bound of the last section is used. We move towards a proof of the continuity of $L(r)$.

Observe that $L(r)$ is a limit of continuous functions, hence Borel measurable, as a function of $r > 1$.

Let $\alpha, \beta \in (0,1)$ be such that $\alpha+\beta = s/d$. Define θ so that $(1-\theta)\gamma = \alpha$ and $(1+\theta)\gamma = \beta$. By (4.8), we have

$$D_{\text{res}}(0, Wr) \leq_d D_{\text{res}}(0, |W|^\alpha Z_{Wr,\alpha} r^\alpha) + D'_{\text{res}}(0, |W|^\beta Z'_{Wr,\beta} r^\beta) + 1 + rW 1_{A_{rW}} \tag{4.67}$$

where the α, β decorations on Z, Z' indicate that they are for α and β, thus representing distributions different from Z and Z' that we have seen so far in this section, and thus not interacting with W in the same way. Additionally, the first subscripts arise from the fact that, without knowing $\alpha = \gamma$, the Z and Z' may depend on Wr, as per equation (4.8). Here, W is independent from all other random objects appearing. Thus, we still know that $Z_{Wr,\alpha}, Z'_{Wr,\beta}$ have distributions that have the same null sets as the Lebesgue measure on $\mathbb{R}^d$. Thus, we can take (4.67) with $r^{\gamma^{-n}}$ substituted in for r and divide by 2^n and take the limit in probability as $n \to \infty$ to conclude that $L(r) \leq L(r^\alpha) + L(r^\beta)$. The coupling across different n can be done in any way the reader pleases. This is the same as

$$L(r) \leq \frac{L(r^{1-\theta}) + L(r^{1+\theta})}{2}. \tag{4.68}$$

By varying α, β among possible choices, or equivalently varying θ over a sufficiently small open interval centered around 0, this implies that the function $g(r) := L(e^r)$ defined for $r > 0$ is midpoint convex on sufficiently small open intervals about each point in its domain of definition. Since L is Borel, so is g. Thus, by [5], which contains the proof that midpoint convexity implies continuity for real-valued Lebesgue measurable functions on the real line, we see that $g(r)$ is continuous. Hence $L(r)$ is continuous in r. □

Lemma 4.24. *We have*

$$\frac{ED_{res}(rW)}{L(r)} \to 1 \tag{4.69}$$

as $r \to \infty$ and the limit in Theorem 4.16 is uniform on compact sets of $r > 1$.

Proof. Let $h_n(r) := ED_{\text{res}}(0, Wr^{\gamma^{-n}})/2^n$, so we know $h_n \to L(r)$ uniformly for compact sets of r, by the previous lemma, Lemma 4.16 and Dini's Theorem. By applying this to the compact set $[e^\gamma, e]$, we find that $\forall \epsilon > 0, \exists n_0 \geq 1$ such that $\forall n \geq n_0, r \in [e^\gamma, e]$ we have

$|h_n(r) - L(r)| \leq \epsilon$. This is the same as saying

$$\sup_{n \geq n_0} \sup_{e^{\gamma^{-n+1}} \leq r \leq e^{\gamma^{-n}}} \left| \frac{ED_{\text{res}}(rW)}{2^n} - L(r^{\gamma^n}) \right| \leq \epsilon \tag{4.70}$$

This is the same as

$$\sup_{n \geq n_0} \sup_{e^{\gamma^{-n+1}} \leq r \leq e^{\gamma^{-n}}} \left| \frac{ED_{\text{res}}(rW)}{2^n} - \frac{L(r)}{2^n} \right| \leq \epsilon. \tag{4.71}$$

Notice that $\phi(r) := L(r)/\log(r)^{\Delta}$ is periodic in the sense that $L(r) = L(r^{\gamma})$ because $\gamma^{\Delta} = 1/2$. Therefore $\phi(r)$ is bounded away from both 0 and ∞. From this, we can conclude that the $L(r)/2^n$ term above is bounded away from both 0 and ∞ as n and r vary over their allowed domains. Specifically, $1/(2\max\phi) \leq 2^n/L(r) \leq 2/\min(\phi)$, so that the bounds have no dependence on ϵ, r, n. We thus conclude that

$$\sup_{n \geq n_0} \sup_{e^{\gamma^{-n+1}} \leq r \leq e^{\gamma^{-n}}} \left| \frac{ED_{\text{res}}(rW)}{L(r)} - 1 \right| \leq 2\epsilon/\min(\phi). \tag{4.72}$$

This is the same as

$$\sup_{r \geq e^{\gamma^{-n_0+1}}} \left| \frac{ED_{\text{res}}(rW)}{L(r)} - 1 \right| \leq 2\epsilon/\min(\phi). \tag{4.73}$$

Thus, we have learned that

$$\frac{ED_{\text{res}}(rW)}{L(r)} \to 1 \tag{4.74}$$

as $r \to \infty$. □

Remark 4.25. *Observe that the ϕ defined here has the required continuity and periodicity properties in the statement of Theorem 4.4.*

Dini's Theorem was paramount in establishing convergence of first moments in the continuum parameter r rather than the discrete parameter n. We proceed to show how the same tool can be used to deal with the second moments.

Lemma 4.26. *We have*

$$Var\Big(\frac{D_{res}(rW)}{L(r)}\Big) \to 0 \tag{4.75}$$

as $r \to \infty$.

Proof. Fix $r > 1$. Almost every x (Lebesgue) is a valid multiplier, so W is a valid multiplier with probability 1. Therefore, Theorem 4.17 with a copy of W that is independent of all random objects plugged in implies that

$$D_{\mathrm{res}}(r^{\gamma^{-n}} W)/2^n \to L(r) \tag{4.76}$$

almost surely.

Let us define $Y_n = X_n/2^n$, where X_n was defined in the proof of Lemma 4.21. The L^3 bound there implies that Y_n^2 is uniformly integrable and Y_n is uniformly integrable, so the fact that $Y_n \to L(r)$ in probability implies that $\mathrm{Var}(Y_n) \to 0$ so that $A_n + B_n \to 0$ pointwise in r. We would like to extract uniformity in r on compacts, so r is no longer fixed at this point.

From the proof of Theorem 4.17, we learn that the sequence $A_n + B_n + C_n - \sum_{k=1}^n F_k/4^k$ is decreasing, and we already know its limit is $L(r)^2 - \sum_{k=1}^\infty F_k/4^k$, which is continuous in r. Dini's Theorem tells us that the convergence is thus uniform on compact sets of r. Because C_n converges uniformly to $L(r)^2$ on compact sets of r and the partial sums of $F_k/4^k$ converge uniformly to their limit on compact sets of r, we thus conclude that $A_n + B_n$ converges to 0 uniformly on compact sets of r.

By applying this uniform convergence to the compact set $[e^\gamma, e]$ we see that given $\epsilon > 0$, there exists $n_0 > 0$ such that

$$\sup_{n \geq n_0} \sup_{e^\gamma \leq r \leq e} \mathrm{Var}[D_{\mathrm{res}}(r^{\gamma^{-n}} W)/2^n] < \epsilon \tag{4.77}$$

so that

$$\sup_{n \geq n_0} \sup_{e^{\gamma^{-n+1}} \leq r \leq e^{\gamma^{-n}}} \mathrm{Var}[D_{\mathrm{res}}(rW)/L(r)] \leq 2(\min \phi)^{-1} \epsilon. \tag{4.78}$$

Thus, $D_{\mathrm{res}}(rW)/L(r)$ has variance converging to 0. □

We now prove Theorem 4.4.

Proof of Theorem 4.4. So far, we know $D_{\mathrm{res}}(rW)/L(r)$ has variances converging to 0 and expectations converging to 1, so it converges in probability to 1 as $r \to \infty$.

Let $x \in \mathbb{R}^d$ be nonzero, and let $t > 0$. Find $\epsilon \in (0, |x|)$ such that $c_1 = (1+\epsilon/|x|)(1-t) < 1$ and $c_2 = (1-\epsilon/|x|)(1+t) > 1$. We know that as $r \to \infty$,

$$P(D_{\text{res}}(0, Wr)/L(r) < c_1 \big| |W| \in [|x|, |x| + \epsilon) \to 0 \tag{4.79}$$

and

$$P(D_{\text{res}}(0, Wr)/L(r) > c_2 \big| |W| \in (|x| - \epsilon, |x|] \to 0. \tag{4.80}$$

Observe that $|x|/|W| \leq 1$ in the first of these, and ≥ 1 in the second. Therefore, using (4.18) and (4.19),

$$\begin{aligned} &P(D_{\text{res}}(0, xr)/L(r) < 1 - t) \\ &\leq P(D_{\text{res}}(0, Wr)|x|/(|W|L(r)) < 1 - t \big| |W| \in [|x|, |x| + \epsilon)) \\ &\leq P(D_{\text{res}}(0, Wr)/L(r) < c_1 \big| |W| \in [|x|, |x| + \epsilon)) \to 0 \end{aligned} \tag{4.81}$$

and

$$\begin{aligned} &P(D_{\text{res}}(0, xr)/L(r) > 1 + t) \\ &\leq P(D_{\text{res}}(0, Wr)|x|/(|W|L(r)) > 1 + t \big| |W| \in (|x| - \epsilon, |x|]) \\ &\leq P(D_{\text{res}}(0, Wr)/L(r) > c_2 \big| |W| \in (|x| - \epsilon, |x|]) \to 0 \end{aligned} \tag{4.82}$$

Together, these imply the required convergence in probability. □

We provide a proof of the upper bound within Theorem 4.6. It suffices to do it for D instead of D_{dis} by Lemma 4.7. In fact, it suffices to do it for D_{res} instead of D.

Proof of Upper Bound: Pick $|x| = 1$ in 4.4. Using the definition of convergence in probability, we conclude that $P(D_{\text{res}}(0, rx)/L(r) > 2) \to 0$. We have thus proved the upper bound with $C = 2 \max \phi$. □

CHAPTER 5

Conclusion

This work has explored a sampling of limit theorems in a diverse array of subfields of probability theory. Often times, limiting objects are easier to understand than the sequences that approximate them. It is certainly true that stationary distributions, mixing measures, the continuum version of the long range percolation problem, and the scaling limit of the long range percolation chemical distances obey this general idea. This dissertation provides a survey of how limiting objects could be used to better understand probabilistic processes.

References

[1] D. Aldous, *Exchangeability and Related Topics.* Springer Berlin Heidelberg, pages 1-198 (1985).

[2] M. Biskup, *Graph Diameter in Long-Range Percolation.* Random Structures and Algorithms Vol. 39 No. 2, p. 210-227 (2011).

[3] M. Biskup, *On the scaling of the chemical distance in long range percolation models.* Ann. Probab. Vol 32 No. 4, p.2938-2977 (2004).

[4] M. Biskup, J. Lin, *Scaling of the Chemical Distance in Long Range Percolation.* Work in Progress.

[5] H. Blumberg, *On Convex Functions.* Trans. Amer. Math. Soc. Vol. 20, p. 40-44 (1919).

[6] J. Cuesta-Albertos, R. Fraiman, T. Ransford, *A Sharp Form of the Cramér-Wold Theorem.* J. Theor. Probab. **20**, 20117209 (2007).

[7] S. Evans, X. Zhou, *Identifiability of Exchangeable Sequences with Identically Distributed Partial Sums.* Elect. Comm. in Probab. **4**, 91713 (1999).

[8] W. Feller, *On Muntz' theorem and completely monotone functions.* Amer. Math. Monthly **75**, 34217350 (1968).

[9] D. Freedman, *de Finetti's theorem in continuous time.* IMS Lecture Notes Monogr. Ser. **30**, 831798 (1996).

[10] G. Grimmett, *Percolation.* 2nd ed. Springer-Verlag Berlin Heidelberg, New York, 1989.

[11] A. Kechris, *Classical Descriptive Set Theory*, Springer-Verlag, New York, 1995.

[12] T. Liggett, *Attractive Nearest Neighbor Spin Systems on the Integers.* Ann. Probability Vol. 6 No. 4, p. 629-636 (1978).

[13] T. Liggett, *Interacting Particle Systems.* Springer Berlin Heidelberg, p.143-152 and Ch. 1 (1985).

[14] J. Lin, *A Correction Note: Attractive Nearest Neighbor Spin Systems on the Integers.* (ArXiv:1507.00678). (2014)

[15] J. Lin, *A Negative Answer to a Problem of Aldous on Determination of Exchangeable Sequences.* Electron. J. Probab. Vol. 21 no. 30, 1-26 (2016).

[16] E. Stein, G. Weiss *Introduction to Fourier Analysis on Euclidean Spaces.* Princeton University Press, New Jersey (1971).