Vrije Universiteit Amsterdam
Faculty of Sciences

# MASTER THESIS

*Iva Kozáková*

# On Nearest Neighbor Model

*Department of Mathematics*

Supervisor: *Prof. Dr. R. W. J. Meester*
Degree Program: *Mathematics*

Amsterdam 2004

I would like to thank my supervisor for his guidance
and support in the preparation of this master thesis.

# Contents

# Chapter 1

# Introduction

At this place I would like to introduce the model and content of this paper. All the time we consider a Poisson process $\Pi$ in $\mathbb{R}^d$ with density one and with one extra point at the origin. We use the notation $\mathrm{NN}(d, 1)$ for the model on this Poisson process, where each point is connected to its nearest neighbor with a directed edge. The density of the process has no relevance in this model. Since the distances between any two pairs of points are a.s. different, the model is well defined. This also caused that there is almost sure no circuit longer than two. Two points connected to each other with a directed path will be called loop. Further the uniqueness of the distance implies that each component contains exactly one such loop (where "component" means a maximal set of connected points), for more details see Chapter 2.

The aim of this work is to develop some results which could help to estimate number of points in a component and probability that the component containing the origin has more than $n$ points.

Chapters 3 and 4 are based on the result in the paper [5] where the authors have estimated the probability that there exist a path starting at the origin and ending at some point farther than at distance $L$ from the origin. This article is still only a preprint so the reader can find it in the appendix. We enlarge upon a few steps in the proof, which could be not obviously seen by some readers of the article.

It will be proven in the Chapter 5 that the expected number of points in component in the model $\mathrm{NN}(d, 1)$ is bounded and grows to 4 as the dimension tends to infinity. And finally the Chapter 6 presents the main result of this thesis: the distribution of the component size in the limiting case that dimension tends to infinity. The component can be seen as two trees, which roots are connected in the loop. From each point in a component there is a path to the loop using just directed edges. In the limiting situation the number of sons of some given point (i.e. number of points for which the point is nearest neighbor) is independent of number of sons of arbitrary other point. And the number of sons in this limiting case has Poisson distribution with parametr depending only on the distance from the given point to the loop, as will be shown. Of course, the shape of the component has influence on the probability of its existence. Therefore we will discuss, which shape of the component (of the tree) is the most or least probable one. There exist an explicit formula expressing the probability that a component has exactly $n$ points (in the limiting case), but a better (more readable) estimate of this probability could be content of some following paper.

# Chapter 2

# Basic Definitions and Properties of Poisson Process and $\mathbf{NN(d,1)}$ Model

Denote $M$ the space of $\sigma$-finite counting measures on the $\sigma$-algebra $\mathcal{B}$ of Borel sets in $\mathbb{R}^d$ for which the measure of a point is 0 or 1. By $\mathcal{M}$ denote the smallest $\sigma$-algebra of subsets $M$ with respect to which all mappings

$$X_B : M \to 0, 1, ..., \infty \text{ with } m \to X_B(m) = m(B), B \in \mathcal{B}$$

are measurable.

A random element $\Pi$ defined on probability space $\{\Omega, \mathcal{F}, P\}$ and taking values in $\{M, \mathcal{M}\}$ is called a Poisson point process with intensity measure $\Lambda$, if

i. for every finite choice of disjoint set $B_1, \dots, B_n \in \mathcal{B}$ the random variables $\Pi(B_1), \dots, \Pi(B_n)$ are mutually independent

ii. for every $B \in \mathcal{B}$ with $\Lambda(B) < \infty$, $X(B)$ is a Poisson random variable with parameter $\Lambda(B)$, i.e.
$$P(\Pi(B) = k) = \frac{\Lambda(B)^k}{k!} e^{-\Lambda(B)}.$$

In this text we consider a homogeneous Poisson process, i.e. the intensity measure is in the form $\lambda V(B)$, where $\lambda$ is a constant, called simply density, and $V(B)$ is the Lebesgue measure of the set (volume). Then one particular point has zero measure and hence we can assume that the Poisson process has one extra point at the origin.

Sometimes it is useful to know the distribution of the nearest neighbor. Denote the nearest neighbor of $x_0$ as $x_1 = x_0 + W$. Then $W$ is uniformly distributed random variable on the $d$-dimensional sphere with radius $|W|$, and for its distribution and density function it holds

$$\begin{aligned} D_f(w) &= 1 - P(|W| > w) = 1 - P(\Pi(B(x_0, w)) = 0) = 1 - e^{-w^d \pi_d} \\ f(w) &= e^{-w^d \pi_d}. \end{aligned}$$

Since the distribution of the nearest neighbor is continuous, the distances between two distinct pairs of points are a.s. different. Suppose now for a while that we have a set of points $x_1, \dots, x_n$, $n > 2$ such that $x_i$ is the nearest neighbor of point $x_{i-1}$ and $x_1$ is the nearest neighbor of $x_n$. Then it has to be the situation that all the distances between two points are equal which happens with zero probability. Hence there is a.s. no such circuit longer then 2.

The nearest neighbor model $\mathrm{NN}(d,1)$ means a $\mathbb{R}^d$-valued Poisson process with directed edges which connect each point to its nearest neighbor. Maximal set of connected points is called a component.

From previous consideration it follows that in this model there can exist only circuits of length 2 (called loops) when two points are each other's nearest neighbors. Furthermore at most one loop can be in one component as it can be shown by simple contradiction: Suppose there are two loops $x_0 \leftrightarrow x_1$ and $x_{n-1} \leftrightarrow x_n$ connected by a path through points $x_2, \dots, x_{n-2}$, the path has length $n-2$ but only $n-3$ edges from the midpoints are available. On the other hand at least one loop has to be in each component. It is sufficient to prove that the origin is a.s. connected to a loop. Let $x_0$ be the point at the origin and $x_0, x_1, x_2, \dots$ be a sequence of points such that $x_{i+1}$ is the nearest neighbor of $x_i$ and $x_i \neq x_j$ for $i \neq j$. If the sequence is finite it has to be the case that the two last points are each others nearest neighbors. Probability that the sequence is infinite is zero. We will show it as in [1]. Let $r_i$ denote $|x_{i+1} - x_i|$. Suppose we know position of points $x_0, \dots, x_i$. Then the probability $P_{i+1}$ that $x_{i+1}$ is different from $x_0, \dots, x_i$ is equal to

$$\begin{aligned} P_{i+1} &= 1 - e^{V(B(x_i,r_i)\setminus \bigcup_{j=1}^{i-1} B(x_j,r_j))} \\ &\leq 1 - e^{V(B(x_i,r_i)} \\ &\leq 1 - e^{V(B(x_1,r_1)}. \end{aligned}$$

Hence given $x_1$ the probability of such sequence $x_2, \dots, x_i$ is equal to the product of the conditional probabilities $P_j$ and

$$\prod_{j=2}^{i} P_j \leq (1 - e^{V(B(x_1,r_1)})^{i-1},$$

which tends to 0 as $i \to \infty$. Since $r_1 > 0$ a.s., the probability that there exist an infinite path $x_0, x_1, x_2, \dots$, such that $x_{i+1}$ is the nearest neighbor of $x_i$, is zero and we have shown that each component contains exactly one loop. This fact we have used in the Section 5 for proving that the expected number of points in a component is finite, what implies that the existence of an infinite component has zero probability. Furthermore a component can be seen as a tree, where the loop is a root.

# Chapter 3

# Some Details of Proof by S. Nanda & C. M. Newman

In this chapter we strongly refer to particular parts of proof in [5]. We shall present some steps of their proof in more details, i.e. equation denoted in the paper as (5), the estimation used in equation (2) and the last inequality (13).

Consider the nearest neighbor model $\mathrm{NN}(d,1)$ and denote $\tau_d(L)$ the probability that there is a path starting at the origin and ending at a point $s$ with $|s| > L$. The following theorem and its proof has been treated:

**Theorem 1.** *There exist constants $c_1, c_2, L_0 \in (0,\infty)$ (depending only on $d$) such that*

$$e^{-c_1 L (\ln L)^{\frac{d-1}{d}}} \leq \tau_d(L) \leq e^{-c_2 L (\ln L)^{\frac{d-1}{d}}} \quad \textit{for } L \geq L_0,$$

*where $\tau_d(L)$ is the probability that there is a path starting at the origin and ending at a point $s$ with $|s| > L$.*

## Summation over Poisson Process (in Equality (5))

The task is to compute the probability (denoted by $p_d(n,L)$) that there exists a directed path from the origin longer than $L$ and touching exactly $n$ distinct points of the $d$-dimensional Poisson process $\Pi$ with density 1 in the model $\mathrm{NN}(d,1)$. Or equivalently the probability that there exist a sequence of points $s_1, s_2, ..., s_n$ such that $s_i$ is the nearest neighbor of $s_{i+1}$ (and $s_1$ is the nearest neighbor of the origin) and $|s_n| > L$. The probability of existence such $n$-tuple is in the article expressed by an integral and on following lines we will argue why we can write it in this way. In fact we shall prove that the expected value of a summation over Poisson process of a function (not only one-dimensional!) is equal to corresponding integral.

Let $S$ be defined following the article [5]. It means that $S$ is a subset of $(\mathbb{R}^d)^n$ containing points $(x_1, x_2, \ldots, x_n)$ which satisfy the following three conditions:

i. $|x_1| \geq |x_2| \geq |\ldots \geq |x_n|$

ii. $|s_n| \geq L$

iii. $s_i \notin \bigcup_{j=1}^{i} B_j$ for $i = 1, \ldots, n$

where $s_i = x_1 + x_2 + \ldots + x_i$ and $B_j = B(s_{j-1}, |x_j|)$ (open ball). Now we can write the indicator of the presence of any such directed path as the summation of indicators over $n$-tuples of points. In fact there can be a.s. at most one such $n$-tuple of n points which construct the directed path. Hence

$$p_d(n, L) = \mathrm{E} \sum_{(x_1, x_2, \ldots, x_n):\, x_i \in \Pi} I\{(x_1, x_2, \ldots, x_n) \text{ construct the directed path}\} \tag{3.1a}$$

$$= \mathrm{E} \sum_{(x_1, x_2, \ldots, x_n):\, x_i \in \Pi} I\{(x_1, x_2, \ldots, x_n) \in S \text{ \& no more points in } \bigcup_{j=1}^{n} B_j\} \tag{3.1b}$$

$$= \mathrm{E} \sum_{\substack{(x_1, x_2, \ldots, x_n):\, x_i \in \Pi, \\ (x_1, x_2, \ldots, x_n) \in S}} P(\text{no more points in } \bigcup_{j=1}^{n} B_j) \tag{3.1c}$$

$$= \mathrm{E} \sum_{\substack{(x_1, x_2, \ldots, x_n):\, x_i \in \Pi, \\ (x_1, x_2, \ldots, x_n) \in S}} e^{-V(\bigcup_{j=1}^{n} B_j)} \tag{3.1d}$$

$$= \int \cdots \int_{S} e^{-V(\bigcup_{j=1}^{n} B_j)} \, dx_1 \, dx_2 \, \ldots \, dx_n. \tag{3.1e}$$

The last equation follows from the following proposition, analogous to the one-dimensional case in [2] .

**Proposition 2.** *Let $f \in \mathrm{L}_1$ and $\Pi$ be a unit density Poisson process in $\mathbb{R}^d$. Then*

$$\mathrm{E} \sum_{\substack{(x_1, x_2, \ldots, x_n): \\ x_i \in \Pi, x_i \neq x_j}} f(x_1, x_2, \ldots, x_n) = \int \cdots \int_{(\mathbb{R}^d)^n} f(x_1, x_2, \ldots, x_n) \, dx_1 \, dx_2 \, \ldots \, dx_n. \tag{3.2}$$

*Proof.* Let us start with $f$ simple, in particular constant on n-dimensional intervals $A_{j_1} \times A_{j_2} \times \ldots \times A_{j_n}, j_i = 1 \ldots m$ for all $i, j$, and denote the value of the function there $f_{j_1 j_2 \ldots j_n} \geq 0$.

Then

$$\begin{aligned} \sum_{\substack{(x_1, x_2, \ldots, x_n): \\ x_i \in \Pi, \; x_i \neq x_j}} f(x_1, x_2, \ldots, x_n) &= \sum_{j_1=1}^{m} \sum_{\substack{j_2=1 \\ ji \neq jk \text{ for } i \neq k}}^{m} \ldots \sum_{j_n=1}^{m} f_{j_1 j_2 \ldots j_n} \mathrm{E}(N(A_{j_1}) N(A_{j_2}) \ldots N(A_{j_n})) + \\ &+ \sum_{\substack{j_1=1 \\ j_h = j_g,\text{ other not equal}}}^{m} \sum_{j_2=1}^{m} \ldots \sum_{j_n=1}^{m} f_{j_1 j_2 \ldots j_n} \mathrm{E}(N(A_{j_1}) \ldots N(A_{j_h}) \ldots (N(A_{j_g}) - 1) \ldots N(A_{j_n})) + \\ &+ \ldots + \sum_{j_1=1}^{m} \sum_{\substack{j_2=1 \\ j_1 = j_2 = \ldots = j_n}}^{m} \ldots \sum_{j_n=1}^{m} f_{j_1 j_2 \ldots j_n} \mathrm{E}(N(A_{j_1}) N(A_{j_2} - 1) \ldots (N(A_{j_n}) - n + 1)), \end{aligned} \tag{3.3}$$

where $N(A_{j_i})$ represents the number of points in the region $A_{j_i}$. Because of the Poisson distribution property the numbers of points in disjoint regions are independent and

$$\mathrm{E}N(A_{j_i}) = V(A_{j_i}) \tag{3.4a}$$

$$\mathrm{E}N(A_{j_i})(N(A_{j_i}) - 1) \ldots (N(A_{j_i}) - k + 1) = V(A_{j_i})^k. \tag{3.4b}$$

Hence

$$\mathrm{E} \sum_{\substack{(x_1,x_2,\ldots,x_n): \\ x_i\in\Pi, x_i\neq x_j}} f(x_1,x_2,\ldots,x_n) = \sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\cdots\sum_{j_n=1}^{m} f_{j_1 j_2 \ldots j_n} V(A_{j_1})V(A_{j_2})\ldots V(A_{j_n}) \tag{3.5}$$

which is equal to $\int\limits_{(\mathbb{R}^d)^n}\!\cdots\!\int f(x_1,x_2,\ldots,x_n)\,dx_1\,dx_2\,\ldots\,dx_n$.

Now arbitrary $f \geq 0$, $f \in \mathrm{L}_1$ can be constructed as a limit of nondecreasing sequence of simple functions such that

$$\int\limits_{(\mathbb{R}^d)^n}\!\cdots\!\int |f_m(x_1,x_2,\ldots,x_n) - f(x_1,x_2,\ldots,x_n)|\,dx_1\,dx_2\,\ldots\,dx_n \to 0.$$

Then by the monotone convergence theorem and by the dominated convergence theorem it follows

$$\sum_{\substack{(x_1,x_2,\ldots,x_n): \\ x_i\in\Pi, x_i\neq x_j}} f_m(x_1,x_2,\ldots,x_n) \to \sum_{\substack{(x_1,x_2,\ldots,x_n): \\ x_i\in\Pi, x_i\neq x_j}} f(x_1,x_2,\ldots,x_n) \tag{3.6a}$$

$$\mathrm{E} \sum_{\substack{(x_1,x_2,\ldots,x_n): \\ x_i\in\Pi, x_i\neq x_j}} f(x_1,x_2,\ldots,x_n) = \lim_{m\to\infty} \mathrm{E} \sum_{\substack{(x_1,x_2,\ldots,x_n): \\ x_i\in\Pi, x_i\neq x_j}} f_m(x_1,x_2,\ldots,x_n) \tag{3.6b}$$

$$= \int\limits_{(\mathbb{R}^d)^n}\!\cdots\!\int f(x_1,x_2,\ldots,x_n)\,dx_1\,dx_2\,\ldots\,dx_n. \tag{3.6c}$$

In the general case we can write $f = f^+ - f^-$, where $f^+, f^- \geq 0$ and $f^+, f^- \in \mathrm{L}_1$, and compute the expectation value as the sum of two of them. Hence the proposition is proven. □

## Limit Property of $\frac{(b)^n}{n!}e^{-cn(\frac{L}{n})^d}$

The noticed term estimates the connectivity function from bottom for each $n$. The authors of the article offer us some $n$ (as a function of $L$), which gives us the wanted lower bound. But why we choose exactly such $n$? Is it the best one? Therefore the task for this moment is to determine $n$ as a function of $L$ such that the function will tend to 0 as slow as possible, or equivalently, the logarithm of the function will tend to $-\infty$ as slow as possible. Using Stirling formula $\ln(n!) \approx n(\ln n - 1)$ we can rewrite the problem in the form

$$n(\ln n - 1 - \ln b) + cn\Big(\frac{L}{n}\Big)^d \to \infty \text{ as } L \to \infty,\ n = n(L). \tag{3.7}$$

It means that both summands should be of approximately same order. We can omit the constant $1 + \ln b$ and continue with this approximation:

$$L \approx n(\ln n)^{\frac{1}{d}} \tag{3.8a}$$

$$n \approx \frac{L}{(\ln n)^{\frac{1}{d}}} \approx \frac{L}{(\ln L - \frac{1}{d}\ln\ln n)^{\frac{1}{d}}} \approx \frac{L}{(\ln L - \sigma(\ln L))^{\frac{1}{d}}} \tag{3.8b}$$

$$n(\ln n) + cn\Big(\frac{L}{n}\Big)^d \approx \frac{L(\ln L - \frac{1}{d}\ln(\ln L - \sigma(\ln L)))}{(\ln L - \sigma(\ln L))^{\frac{1}{d}}} + \frac{cL(\ln L - \sigma(\ln L))}{(\ln L - \sigma(\ln L))^{\frac{1}{d}}} \tag{3.8c}$$

$$\approx L(\ln L - \sigma(\ln L))^{\frac{d-1}{d}} \tag{3.8d}$$

Hence the expression $\frac{(b)^n}{n!}e^{-cn(\frac{L}{n})^d}$ can be bounded by $e^{-c_1 L(\ln L)^{\frac{d-1}{d}}}$ (and nothing better for our object).

## Estimate in Inequality (13)

In the final part of the article there is used an upper bound of the generating function of a random variable with spatial exponential distribution. It's just a technical stuff, which takes us a few lines.

We would like to prove that for a random variable $W$ with the density function $e^{-\pi_d|w|^d}$, there exists $c$ such that it holds $\mathrm{E}(e^{r|W|}) \leq e^{cr^{\frac{d}{d-1}}}$ for large $r$. Let $\pi_d$ denote the volume of $d$-dimensional unit ball and thus the area of its surface is $d\pi_d$.

$$\mathrm{E}(e^{r|W|}) = \int_{\mathbb{R}^d} e^{r|w|-\pi_d|w|^d}\,dw = \int_{\mathbb{R}^+} e^{rx-\pi_d x^d} x^{d-1} d\pi_d\,dx \tag{3.9a}$$

$$= \int_0^{r^{\frac{1}{d-1}}} e^{rx-\pi_d x^d} x^{d-1} d\pi_d\,dx + \int_{r^{\frac{1}{d-1}}}^{\infty} e^{rx-\pi_d x^d} x^{d-1} d\pi_d\,dx \tag{3.9b}$$

$$\leq \int_0^{r^{\frac{1}{d-1}}} e^{r\,r^{\frac{1}{d-1}}-\pi_d r^{\frac{1}{d-1}}} r d\pi_d\,dx + \left[\frac{\pi_d}{\pi_d-1}e^{(1-\pi_d)x^d}\right]_{r^{\frac{1}{d-1}}}^{\infty} \tag{3.9c}$$

$$= e^{(1-\pi_d)r^{\frac{d}{d-1}}} r^{\frac{d}{d-1}} d\pi_d + e^{(1-\pi_d)r^{\frac{d}{d-1}}} \frac{\pi_d}{\pi_d-1} \tag{3.9d}$$

$$\leq e^{cr^{\frac{d}{d-1}}} \text{ for } r \geq r_0 \tag{3.9e}$$

In the second term in (3.9c) we used the fact that $rx - \pi_d x^d < (1-\pi_d)x^d$ for $x > r^{\frac{1}{d-1}}$.

What completes the proof of the last detail we wanted to mention and we can leave this nonhomogeneous chapter.

# Chapter 4

# Existence of Path Touching More than $n$ Points

From the article we have got some information about diameter of the component, about probability that the component reaches farther than something, which is clearly in proportion to the density of the process. Analogously we can ask for the estimate of probability that a path goes through a given number of points (which doesn't depend on the density in the nearest neighbor model). And this probability, that there exist a path which contains the origin and touches more than $n$ points, can be estimated following the similar way as in the article [5].

**Proposition 3.** *Consider the model* $\mathrm{NN}(d,1)$. *Let* $\tau_d(n)$ *be the probability that there is a path through origin touching more than* $n$ *points. Then there exist* $c_1$, $c_2$, $n_0$ *(depending only on* $d$*) such that for* $n \geq n_0$

$$e^{-c_1 n \ln n} \leq \tau_d(n) \leq e^{-c_2 n \ln n}. \tag{4.1}$$

*Proof.* This probability is bigger than the probability (denoted by $p_d(n)$) that there exist an oriented path starting at the origin, touching $n$ points and continuing from each point only in a cone of half angle $\leq \frac{\pi}{4}$.

Let $S$ be a subset of $(\mathbb{R}^d)^n$ containing points $(x_1, x_2, \ldots, x_n)$ satisfying the following two conditions:

i. $|x_1| \geq |x_2| \geq | \ldots \geq |x_n|$

ii. $s_i \notin \bigcup_{j=1}^{i} B_j$ for $i = 1, \ldots, n$

where $s_i = x_1 + x_2 + \ldots + x_i$ and $B_j = B(s_{j-1}, |x_j|)$ (open ball). Denote $\pi_d$ the volume of the $d$-dimensional unit ball then

$$p_d(n) = \int_S \cdots \int e^{-V(\bigcup_{i=j}^{n} B_j)}\, dx_1\, dx_2\, \ldots\, dx_n \tag{4.2}$$

$$\geq \int_S \cdots \int \prod_{j=1}^{n} e^{-\pi_d |x_j|^d}\, dx_1\, dx_2\, \ldots\, dx_n. \tag{4.3}$$

Let $W_1, .., W_n$ be i.i.d. $\mathrm{R}^d$-valued random variables with the density function $e^{-\pi_d |w|^d}$. Then

$$\begin{aligned} p_d(n) &\geq P(|W_1| \geq |W_2| \geq \ldots \geq |W_n| \text{ and all } |W_i| \text{ lie in the cone}) && (4.4a)\\ &= \frac{b_d(\theta)^n}{n!} && (4.4b)\\ &\approx e^{n(\ln(b_d(\theta)) - \ln n + 1)} && (4.4c)\\ &\geq e^{-c_1 n \ln n} \text{ for } n \geq n_0, && (4.4d) \end{aligned}$$

where $b_d(\theta)$ is the probability that $W_i$ falls in a cone of half angle $\theta \leq \frac{\pi}{4}$.

For the second inequality in (4.1) we use the fact that there has to be a path starting at the origin, touching more than $\lfloor \frac{n}{2} \rfloor$ points with at most one change in edge direction, say at the point $j$. We denote the probability of such path $p_d(m, j)$, where $m = \lfloor \frac{n}{2} \rfloor$. Define set $S'$ of $(x_1, \ldots, x_m)$ such that

i. $|x_1| \geq \ldots \geq |x_j|$ and $|x_{j+1}| \leq \ldots \leq |x_m|$

ii. $B_i \cap s_0, \ldots, s_m = s_{i-1}$ for $i = 1, .., j$ and
$B'_i \cap s_0, \ldots, s_m = s_i$ for $i = j+1, .., m$,

where again $s_i = x_1 + x_2 + \ldots + x_i$, $B_i = B(s_{i-1}, |x_i|)$ and $B'_i = B(s_i, |x_i|)$.

So we get

$$\begin{aligned} p_d(n) &\leq \sum_{j=0}^{m} p_d(m, j) && (4.5a)\\ &\leq \sum_{j=0}^{m} \int \cdots \int_{S'} e^{-V(\bigcup_{i=1}^{j} B_i \cup \bigcup_{i=j+1}^{m} B'_i)}\, dx_1\, dx_2 \ldots dx_m && (4.5b)\\ &\leq \sum_{j=0}^{m} \int \cdots \int_{S'} e^{-\frac{1}{|K_d|}(\sum_{i=1}^{j} V(B_i) + \sum_{i=j+1}^{m} V(B'_i))}\, dx_1\, dx_2 \ldots dx_m && (4.5c)\\ &= \sum_{j=0}^{m} \int \cdots \int_{S'} \prod_{j=1}^{m} e^{-\frac{1}{K_d}\pi_d |x_j|^d}\, dx_1\, dx_2 \ldots dx_m && (4.5d)\\ &= \sum_{j=0}^{m} \frac{(K_d)^m}{j!\,(m-1)!} && (4.5e)\\ &= \frac{(2K_d)^m}{m!} && (4.5f)\\ &\approx e^{\frac{n}{2}\ln(2K_d) - \frac{n}{2}(\ln n - \ln 2 - 1)} && (4.5g)\\ &\leq e^{-c_2 n \ln n} \text{ for } n \geq n_0, && (4.5h) \end{aligned}$$

where $K_d$ denotes so called kissing number, which gives the upper bound for the number of points with same nearest neighbor. The second inequality in (4.5b) can not be written as an equality because two paths can have common starting part, thus the events are not disjoint. This gives us the second bound and we are done. □

# Chapter 5

# Expected Number of Points in Component

In the model $\mathrm{NN}(d, 1)$ each point of the Poisson process is exactly in one component and each component contain a.s. exactly one loop (recall that the loop means the situation that two points are connected to each other with directed edges). Hence we would like to describe the relation between the density of loops in the process and expected number of points in a component (we can consider the component containing origin).

**Proposition 4.** *Consider the nearest neighbor model* $\mathrm{NN}(d, 1)$*. Then the expected number of points in a component is smaller or equal to* $4$*.*

*Proof.* Let $\Pi$ be a $d$-dimensional Poisson process with density 1 and $L$ be its subprocess with density $\lambda$ such that $L$ contains exactly one point from each loop and both of them have the same probability to be chosen. We will prove that the expected number of points in a component is smaller or equal to $\frac{1}{\lambda}$.

Let $B_n$ be the box $[-\frac{n}{2}, \frac{n}{2}] \times [-\frac{n}{2}, \frac{n}{2}] \times \ldots \times [-\frac{n}{2}, \frac{n}{2}] \subset \mathbb{R}^d$. Denote $M_n$ the number of points of $L$ in $B_n$ and divide the points of $\Pi$ in $B_n$ to components. Then there are some points connected to the loops outside the box, the number of them is, say, $\varepsilon$. Looking only on $B_n$, let $l_n$ denote the proportion of number of $\Pi$-points minus $\varepsilon$ to the number of $L$-points. Then the random variable $l_n$ tends to the expected number of points in a component as $n \to \infty$ (if the limit exists).

$$\lim_{n\to\infty} l_n = \lim_{n\to\infty} \frac{(\#\text{ points of } \Pi \in B_n) - \varepsilon}{(\#\text{ points of } L \in B_n)} \tag{5.1a}$$

$$\leq \lim_{n\to\infty} \frac{(\#\text{ points of } \Pi \in B_n)}{(\#\text{ points of } L \in B_n)} \tag{5.1b}$$

$$= \frac{\lim_{n\to\infty} \frac{1}{n^d}(\#\text{ points of } \Pi \in B_n)}{\lim_{n\to\infty} \frac{1}{n^d}(\#\text{ points of } L \in B_n)} \tag{5.1c}$$

Since the Poisson process and the subprocess $L$ are ergodic (see Section 5.1), both limits in the fraction tend to the according density due to ergodic theorem. Hence

$$\lim_{n\to\infty} l_n \leq \frac{1}{\lambda}. \tag{5.2}$$

Furthermore a point of $\Pi$, say origin, makes a loop, if it is the nearest neighbor of its nearest neighbor. Hence

$$P(\text{origin is in a loop}) = \mathrm{E}\sum_{x\in\Pi} I[\text{no more points in } B(0,|x|)\cup B(x,|x|)] \tag{5.3a}$$

$$= \int_{\mathbb{R}^d} e^{-V(B(0,|x|)\cup B(x,|x|))}\,dx \tag{5.3b}$$

$$\geq \int_{\mathbb{R}^d} e^{-2V(B(0,|x|)}\,dx. \tag{5.3c}$$

Since the area of the $d$-dimensional sphere is equal to $|x|^{d-1}d\pi_d$, where $\pi_d$ is the volume of $d$-dimensional unit ball, the integral can be reparametrized to the following form

$$\int_{\mathbb{R}+} e^{-2x^d\pi_d}x^{d-1}d\pi_d\,dx = \Big[-\frac{1}{2}e^{-2x^d\pi_d}\Big]_0^\infty = \frac{1}{2}. \tag{5.4}$$

From the definition of $L$ and previous result follows that

$$\lambda = \frac{1}{2}P(\text{origin is in a loop}) \geq \frac{1}{4} \tag{5.5}$$

$$\mathrm{E}(\#\text{ points in component}) \leq 4 \tag{5.6}$$

what completes the proof. □

Note that if the dimension goes to infinity, the probability in (5.3) tends to its upper bound and the expected component size is equal to 4 in the limit case.

## 5.1 Property of Subprocess $L$

Recall that the process $L$ has been defined as a subprocess of the Poisson process $\Pi$ such that it contains exactly one point from each loop and both loop points have the same probability to be chosen. It can be shown that this process is ergodic (even mixing), but not Poisson.

Let $T_x$ be a translation in $\mathbb{R}^d$ over the vector x: $T_x(y) = y + x$, for all $y \in \mathbb{R}^d$. Then $T_x$ induces a transformation $S_x : M \to M$ with $(S_x m)(B) = m(T - x^{-1}B), B \in \mathcal{B}$. We identify the space $(M, \mathcal{M})$ with $(\Omega, \mathcal{F})$ and then $P$ is a measure on $\mathcal{M}$. An event (resp. point process) is called stationary if its measure $P$ is invariant under $S_x$ for all $x \in \mathbb{R}^d$. Finally, denote $\mathcal{I}$ the $\sigma$-algebra of events which are invariant under the whole group $\{S_x : x \in \mathbb{R}^d\}$.

Then the stationary point process is said to be ergodic if $\mathcal{I}$ is trivial. And it is mixing if for arbitrary events $E, F \in \mathcal{F}$ it holds

$$P(S_x E \cup F) - P(E)P(F) \to 0, \text{ for } |x| \to \infty.$$

If the proces is mixing, then for any invariant event $E$ the term $P(S_x E \cup E) = P(E)$ tends to $P(E)^2$ and hence $E$ is trivial, which means that the process is also ergodic.

**Proposition 5.** *The subprocess $L$ defined above is mixing.*

*Proof.* Events depending on a realization of $L$ inside a bounded set can be approximated by events which depend only on a realization of $\Pi$ in some bigger bounded set. Since any event can be approximated by an event depending only on $L$ inside a bounded set, it can be also approximated by an event which depends only on the realization of $\Pi$ inside a bounded set. For $|x|$ big enough the shifted bounded set for $E$ and the other one for $F$ become disjoint and hence independent because of the Poisson process property. Thus the measure of the intersection tends to the product of measures (probabilities) of these events. □

Let us write at this place the formulation of the ergodic theorem, which we have used in Section 5.

**Theorem 6.** *Let $\{\Omega, \mathcal{F}, P\}$ be a probability space and let $\{S_x : x \in \mathbb{R}^d\}$ be a group of invertible commuting, measure-preserving transformations on $\Omega$. Let $f$ be a real measurable and $P$-integrable function on $\Omega$. Then*

$$\frac{1}{t^d} \int_{[0,t]^d} f(S_x(\omega)) \to \mathrm{E}(f|\mathcal{I})(\omega) \text{ a.s. as } t \to \infty,$$

*where $\mathcal{I}$ is the $\sigma$-algebra of events which are invariant under the whole group $\{S_x : x \in \mathbb{R}^d\}$.*

For the ergodic process the $\sigma$-algebra $\mathcal{I}$ is trivial and hence $\mathrm{E}(f|\mathcal{I}) = \mathrm{E}(f)$ a.s.

**Proposition 7.** *The subprocess $L$ is not a Poisson process.*

*Proof.* Let $x_0$ be in a loop. Then its nearest neighbor's position is a random variable $x_1 = x_0 + X$ with the following distribution and the density function:

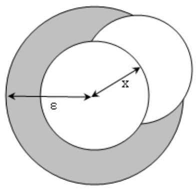
$$P(|X| \leq |x|) = \frac{\int_0^x e^{-v_d |y|^d}\, dy}{\int_0^\infty e^{-v_d |y|^d}\, dy}$$
$$f(X) = \frac{v_d}{\pi_d} e^{-v_d |X|^d},$$

where $\upsilon_d$ is the volume of two intersected unit balls $V(B(\mathbf{0},1) \cup B(\mathbf{1},1))$.

The subprocess $L$ has density $\lambda = \lambda_d = \frac{\pi_d}{2\upsilon_d}$. Assume (for contradiction) that $L$ is a Poisson process with this density. Then

$$P(0 \text{ points of } L \text{ in } B(x_0,\varepsilon)) = e^{-\lambda\varepsilon^d\pi_d}.$$

Let $x_0$ be a point of process $L$. Then the probability of the existence of another $L$-point in its surrounding is influenced by the position of its nearest neighbor in $\Pi$.




$$\begin{aligned} P(0 \text{ points of } L \text{ in } B(x_0,\varepsilon)) &= \int_0^\infty \frac{\upsilon_d}{\pi_d} e^{-\upsilon_d|x|^d}\, e^{-\lambda(V(B(x_0,\varepsilon)\setminus(B(x_0,|x|)\cup B(x_0+x,|x|))))}\, dx \\ &\geq \int_0^\varepsilon \frac{\upsilon_d}{\pi_d} e^{-\upsilon_d|x|^d}\, e^{-\lambda\pi_d(\varepsilon^d-|x|^d)}\, dx + \int_\varepsilon^\infty \frac{\upsilon_d}{\pi_d} e^{-\upsilon_d|x|^d}\, dx \\ &= e^{-\lambda\pi_d\varepsilon^d} \frac{\upsilon_d}{\upsilon_d-\lambda\pi_d} \left[-e^{-(\upsilon_d-\lambda\pi_d)|x|^d}\right]_0^\varepsilon + e^{-\upsilon_d\varepsilon^d} \\ &= e^{-\lambda\pi_d\varepsilon^d} \left( \frac{\upsilon_d(1-e^{-(\upsilon_d-\lambda\pi_d)\varepsilon^d})}{\upsilon_d-\lambda\pi_d} + e^{-(\upsilon_d-\lambda\pi_d)\varepsilon^d} \right) \end{aligned}$$

There exists $\varepsilon$ such that the inequality is not valid. It is sufficient to choose one which satisfies this condition:

$$\begin{aligned} \upsilon_d e^{-(\upsilon_d-\lambda\pi_d)\varepsilon^d} &< \lambda\pi_d \\ (\upsilon_d-\lambda\pi_d)\varepsilon^d &> -\ln(\frac{\lambda\pi_d}{\upsilon_d}) \\ \varepsilon^d &> \frac{\ln(\upsilon_d)-\ln(\lambda\pi_d)}{\upsilon_d-\lambda\pi_d} \end{aligned}$$

Hence we get for such $\varepsilon$ and some $\delta > 0$ the following contradiction:

$$e^{-\lambda\pi_d\varepsilon^d} \geq e^{-\lambda\pi_d\varepsilon^d}(1+\delta)$$

which implies that $L$ can not be a Poisson process. □

# Chapter 6

# Limiting process as $d \to \infty$

In this chapter we derive some estimates on the component size in the limiting case when the dimension grows over all bounds. It can be proven that the volume of the intersection of two balls (with radii bigger than the distance of their centers) in comparison to their volume become negligible as $d \to \infty$. This fact makes the limiting situation simpler than in finite dimensions.

## 6.1 Generations

For each point there exist unique path to the loop. In the Section 5 the probability that point is in a loop is treated. In the similar way we can provide also computations for the next generations, but now only for the limit case $d \to \infty$. Starting with the second generation - what is the probability that the point is not in a loop and its nearest neighbor is in a loop? Denote this probability by $G_2$ and define $G_3, G_4, \ldots$ analogously. As well we can say that the path from the point in the $i$-th generation to a loop contains $i$ distinct points.

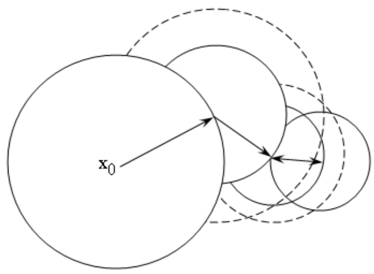



Figure 6.1: Possible constellation of points where the origin is in the third generation.

Denote the origin as $s_0 = x_0$ and $s_i := x_1 + \ldots + x_i$. Using the same idea as in the derivation of (3.1) we can express $G_k$ as the following integral. See Figure 6.1 to get better idea of the integration bounds used in the following expressions. Since omitting of intersections in exponents makes the integrated function smaller, we get the first inequality. Further the formula is divided into several terms as we gradually take out the intersections in integrating bounds.

$$
\begin{aligned}
G_k = &\lim_{d\to\infty} \int_{s_1\in\mathbb{R}^d} e^{-V(B(s_0,|x_1|))} \int_{s_2\in B(s_1,|x_1|)\setminus B(s_0,|x_1|)} e^{-V(B(s_1,|x_2|)\setminus B(s_0,|x_1|))} \\
&\int_{s_3\in B(s_2,|x_2|)\setminus(B(s_1,|x_2|)\cup B(s_0,|x_1|))} e^{-V(B(s_2,|x_3|)\setminus(B(s_0,|x_1|)\cup B(s_1,|x_2|)))} \dots \\
&\int_{s_k\in B(s_{k-1},|x_{k-1}|)\setminus\bigcup_{i=1}^{k-1} B(s_{i-1},|x_i|)} e^{-V(B(s_{k-1},|x_k|)\cup B(s_k,|x_k|)\setminus\bigcup_{i=0}^{k-1} B(s_{i-1},|x_i|))} \\
&ds_k \dots ds_3\, ds_2\, ds_1 \qquad (6.1)\\
\geq &\lim_{d\to\infty} \int_{s_1\in\mathbb{R}^d} e^{-\pi_d|x_1|^d} \int_{s_2\in B(s_1,|x_1|)\setminus B(s_0,|x_1|)} e^{-\pi_d|x_2|^d} \int_{s_3\in B(s_2,|x_2|)\setminus(B(s_1,|x_2|)\cup B(s_0,|x_1|))} \\
&e^{-\pi_d|x_3|^d} \dots \int_{s_k\in B(s_{k-1},|x_{k-1}|)\setminus\bigcup_{i=1}^{k-1} B(s_{i-1},|x_i|)} e^{-2\pi_d|x_k|} ds_k \dots ds_3\, ds_2\, ds_1 \qquad (6.2)\\
= &\lim_{d\to\infty} \int_{s_1\in\mathbb{R}^d} e^{-\pi_d|x_1|^d} \int_{s_2\in B(s_1,|x_1|)} \int_{s_3\in B(s_2,|x_2|)} e^{-\pi_d|x_3|^d} \dots \int_{s_k\in B(s_{k-1},|x_{k-1|})} \\
&e^{-2\pi_d|x_k|}\, ds_k \dots ds_3\, ds_2\, ds_1 - \sum_{j=2}^{k} \lim_{d\to\infty} A_j^k, \qquad (6.3)
\end{aligned}
$$

where

$$
\begin{aligned}
A_j^k = &\int_{s_1\in\mathbb{R}^d} e^{-\pi_d|x_1|^d} \int_{s_2\in B(s_1,|x_1|)} e^{-\pi_d|x_2|^d} \dots \int_{s_{j-1}\in B(s_{j-2},|x_{j-2}|)} e^{-\pi_d|x_{j-1}|^d} \\
&\int_{s_j\in B(s_{j-1},|x_{j-1}|)\cap\bigcup_{i=1}^{j-1} B(s_{i-1},|i_k|)} e^{-\pi_d|x_j|^d} \int_{s_{j+1}\in B(s_j,|x_j|)\setminus\bigcup_{i=1}^{j} B(s_{i-1},|x_i|)} e^{-\pi_d|x_{j+1}|^d} \\
&\dots \int_{s_k\in B(s_{k-1},|x_{k-1|})\setminus\bigcup_{i=1}^{k-1} B(s_{i-1},|x_i|)} e^{-2\pi_d|x_k|^d}\, ds_k \dots ds_2\, ds_1 \qquad (6.4)\\
\leq &\int_{s_1\in\mathbb{R}^d} e^{-\pi_d|x_1|^d} \int_{s_2\in B(s_1,|x_1|)} e^{-\pi_d|x_2|^d} \dots \int_{s_{j-1}\in B(s_{j-2},|x_{j-2}|)} e^{-\pi_d|x_{j-1}|^d} \\
&\int_{s_j\in B(s_{j-1},|x_{j-1}|)\cap\bigcup_{i=1}^{j-1} B(s_{i-1},|i_k|)} e^{-\pi_d|x_j|^d}\, 1\, ds_j \dots ds_2\, ds_1 \qquad (6.5)\\
\leq &\int_{s_1\in\mathbb{R}^d} e^{-\pi_d|x_1|^d} \int_{s_2\in B(s_1,|x_1|)} e^{-\pi_d|x_2|^d} \dots \int_{s_{j-1}\in B(s_{j-2},|x_{j-2}|)} e^{-\pi_d|x_{j-1}|^d} \\
&\left(\frac{V(B(s_{j-1},|x_{j-1}|)\cap\bigcup_{i=1}^{j-1} B(s_{i-1},|x_i|))}{V(B(s_{j-1},|x_{j-1}|)}\right) \int_{s_j\in B(s_{j-1},|x_{j-1|})} e^{-\pi_d|x_j|^d} \\
&ds_j \dots ds_2\, ds_1 \qquad (6.6)
\end{aligned}
$$

Since the integral over variables $j+1, ..., k$ is easily bounded by 1, we obtain the inequality (6.5). The inequality (6.6) is given by the fact that if $f(|x|)$ is decreasing in $|x|$ then for $y$ with $|y| \geq 1$ it holds

$$\int_{x \in B(0,1) \cap B(y,|y|)} f(|x|)\, dx \leq \left( \frac{V(B(0,1) \cap B(y,|y|))}{V(B(0,1))} \right) \int_{x \in B(0,1)} f(|x|)\, dx. \tag{6.7}$$

The volume of intersections (in the factor) can be estimated by the biggest possible intersection, which is the one with first ball (of radius $x_1$) in the situation that $x_{j-1}$ lies on the boundary.

$$\begin{aligned} \left( \frac{V(B(s_{j-1}, |x_{j-1}|) \cap \bigcup_{i=1}^{j-1} B(s_{i-1}, |x_i|))}{V(B(s_{j-1}, |x_{j-1}|))} \right) &\leq \left( j \frac{V(B(0, |x_{j-1}|) \cap B(s, |x_1|))}{V(B(0, |x_{j-1}|))} \right) \\ &= \left( j \frac{V(B(0,1) \cap B(s', \frac{|x_1|}{|x_{j-1}|}))}{V(B(0,1))} \right), \end{aligned} \tag{6.8}$$

where $s$, resp. $s'$ is an arbitrary point in $\mathbb{R}^d$ with $|s| = |x_1|$, resp. $|s'| = \frac{|x_1|}{|x_{j-1}|}$.

We reparametrize the integral in the same way as in Section 5. Again the area of the $d$-dimensional unit sphere is $d\pi_d$, where $\pi_d$ is the volume of the $d$-dimensional unit ball, in particular:

$$\pi_d = \int_{-1}^{1} \sqrt{(1-r)}^{(d-1)} \pi_{d-1} dr \tag{6.9a}$$

$$= \pi_{d-1} \int_0^{\pi} (\sin x)^d\, dx \tag{6.9b}$$

$$= \pi_{d-1} \sqrt{\pi}\, \frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d+2}{2})} \tag{6.9c}$$

$$= \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)}. \tag{6.9d}$$

Hence we can rewrite the integrals in spherical coordinates and substitute $y_i := \pi_d |x_i|^d$. Then

$$\begin{aligned} A_j^k \leq &\int_{|x_1| \in \mathbb{R}^+} e^{-\pi_d |x_1|^d} d\pi_d |x_1|^{d-1} \int_{|x_2| \leq |x_1|} e^{-\pi_d |x_2|^d} d\pi_d |x_2|^{d-1} \dots \int_{|x_{j-1}| \leq |x_{j-2}|} e^{-\pi_d |x_{j-1}|^d} d\pi_d |x_{j-1}|^{d-1} \\ &\left( \frac{j}{\pi_d} V\left( B(0,1) \cap B\left( s', \frac{|x_1|}{|x_{j-1}|} \right) \right) \right) \int_{|x_j| \leq |x_{j-1}|} e^{-\pi_d |x_j|^d} d\pi_d |x_j|^{d-1}\, d|x_j| \dots d|x_2|\, d|x_1| \end{aligned} \tag{6.10}$$

$$= \int_{y_1 \in \mathbb{R}^+} \int_{y_2 < y_1} \dots \int_{y_j < y_{j-1}} \left( \frac{j}{\pi_d} V\left( B(0,1) \cap B\left( s', \left( \frac{y_1}{y_{j-1}} \right)^{\frac{1}{d}} \right) \right) \right) e^{-\sum_{i=1}^{j} y_i}\, dy_j \dots dy_2\, dy_1$$

$$\leq \int_{y_1 \in \mathbb{R}^+} \int_{y_2 < y_1} \dots \int_{y_j < y_{j-1}} \left( \frac{j}{\pi_d} V\left( B(0,1) \cap B\left( s'', \frac{y_1}{y_{j-1}} \right) \right) \right) e^{-\sum_{i=1}^{j} y_i}\, dy_j \dots dy_2\, dy_1 \tag{6.11}$$

$$\leq \int_{y_1 \in \mathbb{R}^+} \int_{y_2 < y_1} \dots \int_{y_j < y_{j-1}} \frac{j(d-1)\big(1 - a(\frac{y_1}{y_{j-1}})\big)}{\sqrt{\pi}} \left( 1 - \left( a\left( \frac{y_1}{y_{j-1}} \right) \right)^2 \right)^{\frac{d-1}{2}} e^{-\sum_{i=1}^{j} y_i}\, dy_j \dots dy_2\, dy_1 \tag{6.12}$$

$\to 0.$

Since $A_j^k$'s are greater than 0, by the Lebesgue dominated convergence theorem the final limit equals 0. The only one thing which remains to be proved is the inequality (6.12). Let us formulate it in a proposition.

**Proposition 8.** *For each fixed real* $y \geq 1$

$$\frac{1}{\pi_d} V(B(0,1) \cap B(s,y)) \xrightarrow[d \to \infty]{} 0, \tag{6.13}$$

*where* $s$ *is an arbitrary point in* $\mathbb{R}^d$ *such that* $|s| = y$.

*Proof.* Let $a = a(y) > 0$ be the distance from 0 to the orthogonal projection of the spheres' intersection to the line segment joining the balls centers (which clearly depends only on the distance $y$). Then

$$\frac{1}{\pi_d} V(B(0,1) \cap B(s,y)) \leq \frac{2}{\pi_d} \int_a^1 \sqrt{1 - r^2}^{(d-1)} \pi_{d-1} \, dr \tag{6.14a}$$

$$= \frac{2\Gamma(\frac{d+2}{2})}{\sqrt{\pi}\,\Gamma(\frac{d+1}{2})} \int_a^1 \left(1 - r^2\right)^{\frac{d-1}{2}} dr \tag{6.14b}$$

$$\leq \frac{2\Gamma(\frac{d+3}{2})}{\sqrt{\pi}\,\Gamma(\frac{d+1}{2})} \int_a^1 \left(1 - r^2\right)^{\frac{d-1}{2}} dr \tag{6.14c}$$

$$= \frac{d+1}{\sqrt{\pi}} \int_a^1 \left(1 - r^2\right)^{\frac{d-1}{2}} dr \tag{6.14d}$$

$$\leq \frac{d+1}{\sqrt{\pi}} \int_a^1 \left(1 - a^2\right)^{\frac{d-1}{2}} dr \tag{6.14e}$$

$$= \frac{(d+1)(1-a)}{\sqrt{\pi}} \left(1 - a^2\right)^{\frac{d-1}{2}}. \tag{6.14f}$$

The term $(1 - a^2)$ is less than 1, hence the limit equals to 0 as $d \to \infty$. For the calculation of $\frac{\pi_{d-1}}{\pi_d}$ we used the identity (6.9). □

Since the summation over this upper bound for $G_k$ equals 1 (see (6.19)), we can write it as an equality:

$$G_k = \int_{y_1 \in \mathbb{R}^+} e^{-y_1} \int_{y_2 < y_1} e^{-y_2} \dots \int_{y_{k-1} < y_{k-2}} e^{-y_{k-1}} \int_{y_k < y_{k-1}} e^{-2y_k} \, dy_k \, dy_{k-1} \dots dy_2 \, dy_1. \tag{6.15}$$

We can see that in the computation for each following generation we did only one step more (one more integration and multiplication by $e^{-y}$), what implies following recurrent relationship:

$$G_i = \sum_{j=0}^{i} \frac{(-1)^j}{(j+1)!} G_{i-j-1}, \tag{6.16}$$

where we set $G_0 := 0$ and $G_{-1} := -1$.

Now we have enough information to express (and compute) the probability $G_i$, let us formulate it in a proposition.

**Proposition 9.** *The probability, that a point (the origin) is in the $i$-th generation in the limiting case* (NN$(d,1)$ *as* $d \to \infty$)*, is equal to*

$$G_i = \frac{i}{(i+1)!} \tag{6.17}$$

*Proof.* We will prove that the recurrent relationship holds for this formula of $G_i$. Suppose now for induction that $G_j = \frac{j}{(j+1)!}$ for $j < i$. Then

$$G_i = \sum_{j=0}^{i} \frac{(-1)^j}{(j+1)!} \frac{i-j-1}{(i-j)!} \tag{6.18a}$$

$$= i \sum_{j=0}^{i} \frac{(-1)^j}{(j+1)!\,(i-j)!} - \sum_{j=0}^{i} \frac{(-1)^j}{j!\,(i-j)!} = \tag{6.18b}$$

$$= i \left( \sum_{j=0}^{i+1} \frac{(-1)^{j-1}}{j!\,(i+1-j)!} + \frac{1}{(i+1)!} \right) - \frac{1}{i!} \sum_{j=0}^{i} (-1)^j \binom{i}{j} = \tag{6.18c}$$

$$= \frac{i}{(i+1)!}. \tag{6.18d}$$

Since $G_{-1}, G_0$ according to this formula are equal to the supposed ones, we have shown that the probability of a point being in the $i$-th generation is $\frac{i}{(i+1)!}$ (in the limiting process as $d \to \infty$). □

Before we continue, we compute some basic summations over $G_i$'s:

$$\sum_{i=1}^{\infty} G_i = \sum_{i=1}^{\infty} \frac{i}{(i+1)!} = \sum_{i=1}^{\infty} \left( \frac{1}{i!} - \frac{1}{(i+1)!} \right) = e - 1 - (e-2) = 1 \tag{6.19}$$

$$\sum_{i=1}^{\infty} iG_i = \sum_{i=0}^{\infty} \frac{i^2}{(i+1)!} = \sum_{i=0}^{\infty} \frac{1}{(i+1)!} + \sum_{i=1}^{\infty} \frac{1}{(i-1)!} - \sum_{i=0}^{\infty} \frac{1}{i!} = e - 1 \tag{6.20}$$

## 6.2 Tree Structure

In the limiting case of our model the possible descendant of a point is not influenced by other descendants or points in the previous generation and other branches of the component. It implies following conception: Starting with a box - we can estimate number of points of $i$-th (and $(i+1)$-st) generation in the box. The distribution of number of descendants (in the $(i+1)$-st generation) belonging to the one point can be approximated, due to the independency, by the binomial distribution and as the box size goes to infinity we get the Poisson distribution with parameter $\frac{G_{i+1}}{G_i}$.

$$\beta_{i-1} := \frac{G_i}{G_{i-1}} = \frac{i}{(i-1)(i+1)} \tag{6.21}$$

Define a random variable $U$ as a multiple composition of compound Poisson processes. In particular

$$U = 1 + \sum_{i_1=1}^{N_1} (1 + \sum_{i_2=1}^{N_2^{i_1}} (1 + \sum_{i_3=1}^{N_3^{i_1 i_2}} (1 + ...))), \tag{6.22}$$

where random variables $N_i^{\cdot\cdot\cdot}$ are independent and have Poisson distribution with parameter $\beta_i$. Then it holds for the generating function

$$\mathrm{E}e^{tU} = \exp(t + \beta_1(\mathrm{E}[e^{(t(1+\sum_{i_2=1}^{N_2^{\cdot\cdot\cdot}}(1+\ldots)))}] - 1)) \tag{6.23a}$$
$$= \exp(t + \beta_1(-1 + \exp(t + \beta_2(-1 + \exp(\ldots))))). \tag{6.23b}$$

Furthermore the number of points in the component is equal to $U_1 + U_2$, where both of them have the above defined distribution and are independent. In Chapter 5 we have shown that the expected size of the component in the limiting situation is 4 points. The following computation implies same result.

$$\mathrm{E}U = (\mathrm{E}e^{tU})'(0) = 1 + \beta_1 + \beta_1\beta_2 + \beta_1\beta_2\beta_3 + \ldots = 1 + \frac{1}{G_1}\sum_{i=2}^{\infty} G_i = 2 \tag{6.24}$$

Since $\beta_i$ tends to zero as $i \to \infty$, the generating function takes finite values and using Chernoff's bound for $t = 1$ we get

$$P(U \geq n) \leq e^{-n}(\mathrm{E}e^{tU})(1) \doteq 2.83e^{-n}. \tag{6.25}$$

Hence $e^{-cn}$, $c > 0$ is an upper bound for the probability that a component has more then $n$ points.

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $G_i$ | $\frac{1}{2}$ | $\frac{1}{3}$ | $\frac{1}{8}$ | $\frac{1}{30}$ | $\frac{1}{144}$ | $\frac{1}{840}$ | $\frac{1}{5760}$ |
| $\beta_i$ | $\frac{2}{3}$ | $\frac{3}{8}$ | $\frac{4}{15}$ | $\frac{5}{24}$ | $\frac{6}{35}$ | $\frac{7}{48}$ | $\frac{8}{63}$ |

Table 6.1: Numerical values of several first $G_i$'s and $\beta_i$'s

## 6.3 Component of Particular Size

There are several possible questions arising at this place. What is the probability that a component has $n$ points and how does the component look like? Which shape is the most (or the least) probable one? We do not present here complete answers of these questions, just some ideas.

First we define an spatial isomorphism on rooted trees, which becomes useful later. All the points in the tree are not labelled but the root is the only one identified point. There are several points connected directly to the root, call them first-generation sons and order them. If we remove the root, the tree falls into several parts, which can be seen as new trees, in which the first-generation sons have a role of root. Now come back to the original tree with ordered sons, two trees $t_1, t_2$ are said to be isomorphic if the trees of their sons are isomorphic, i.e. they have same number of sons and the tree of the first (resp. $i$-th) son of the $t_1$-root is isomorphic to the tree of the first (resp. $i$-th) son of the $t_2$-root. (Tree having one point is clearly isomorphic with another one point tree.) Denote $T_n$ the class of rooted trees on $n$ points not isomorphic in above defined way.

The size of the component was expressed by the random variable $U$. For simplicity, we will treat only the random variable $U$, which means exactly the number of points of the random rooted tree defined in (6.22). Furthermore the random variable $N_1$ means the number of points in the first generation. Let $J_1$ be the set of points in the first generation in the corresponding tree. Denote $N_2 = \sum_{i_1=1}^{N_1} N_2^{i_1}$ the number of points in second generation and $J_2$ the set of such points. Analogously define $N_3, J_3, N_4, J_4 \ldots$. Denote, for $j$ a point of a tree, $s(j)$ number of sons (descendants) of $j$. Naturally the sons can be ordered by the coordinates.

Now the probability, that the tree has exactly $n$ points, is equal to

$$P(U = n) = \sum_{t \in T_n} \prod_{i=1}^{n} \prod_{j \in J_i} \frac{1}{s(j)!} \beta_{i-1} e^{-\beta_i}, \tag{6.26}$$

where $beta_0 = 1$ and other variables are defined above. This formula can be obtained in the direct way using knowledge of Poisson distribution. Second possibility is to compute $n$-th derivation of the probability-generating function in 0, what I found not exactly easy for larger $n$.

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| $P(U = n)$ | 0 | 0.513 | 0.235 | 0.121 | 0.635 | 0.0329 | 0.0168 | 0.00846 |

Table 6.2: Numerical values of probability $P(U = n)$ for several first $n$

At this place it can be useful to know number of not isomorphic trees. (Note that the number of labelled rooted trees with $n$ points is $n^{n-1}$, see [3].) Following the same idea as in [3], denote $|T_n(m)|$ the number of trees with $m$ edges meeting at the origin. It holds $|T_n(1)| = |T_{n-1}|$. And $|T_n(2)| = |T_1||T_{n-2}| + |T_2||T_{n-3}| + ... + |T_{n-2}||T_1|$, because we distribute the $n-1$ points in two new trees and we do care about the order of the branches. Similarly for higher $m$. Therefore the final expression takes following form:

$$|T_n| = \sum_{u=1}^{n-1} \sum_{P} \prod_{j=1}^{u} |T_{k_j}|, \tag{6.27}$$

where $P$ is set of all possible composition at $n-1$ points of length $J$, i.e. set of $(k_1, k_2, \ldots, k_J)$ such that $k_i \geq 0$, $\sum_{j=1}^{u} k_j = n - 1$.

The expression (6.26) of the probability, that the component has exactly $n$ points, is a sum over trees of different shape (not isomorphic). Since the order of sons has no influence on the probability of particular tree, there exist more trees of same probability and the most and least probable tree can be not unique. Hence there should exist among the most (resp. least) probable trees such one that its sons are ordered by the number of points in their subtrees. This tree of particular shape can represent the class of trees with the same probability.

The least probable shape of the tree on $n$ points can be found easier. This minimal tree has to have only one point in $i$ first generations and the rest of the points with one common father - the point in the farthest one-point generation, see Figure 6.2. That is because any other branch starting in lower generation has greater probability than the situation, when we put this branch in the farthest generation. Applying this idea several times we get the described shape of the tree. Hence for an arbitrary shape of tree there exists one with lower probability, which responds our description. Note that this class includes also two spatial trees: the one with all points connected to the root and the other constructed by a single path (always only one son). Hence looking for the least probable tree is a task of finding $i^* = i^*(n) \in [1, n-1]$ as an argument of minimization of

$$\prod_{j=1}^{i} \beta_{j-1} e^{-\beta_j} \frac{1}{(n-i)!} (\beta_i e^{-\beta_{i+1}})^{n-i}. \tag{6.28}$$

Note that the $i^*$ is equal to 1 for $n \in [1, .., 4]$, to 2 for $n \in [5, .., 10]$ , to 3 for $n \in [11, .., 16]$ , to 4 for $n \in [17, .., 22]$ , to 5 for $n \in [23, .., 27]$ and to more then 5 for greater $n$.

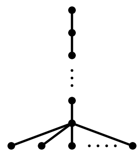


Figure 6.2: The shape of the tree on $n$ points with minimal probability

What can we say generally about the most probable tree? Since $\beta_i$'s are decreasing, the number of sons in lower generation (closer to the root) will be bigger or equal to the number of sons of point in farther generation. This and the zero-influence of the order allow us to shorten the number of trees we will maximize (resp. minimize) over. Such representative tree can be uniquely written as one of the (nonincreasing) partitions on $n-1$. So the $j$-tuple, $(k_1, \ldots, k_j)$ with $\sum_{i=1}^{j} k_i = n-1$ and $k_i$ nonincreasing, we read in following way: $k_1$ is number of sons of the root; $k_2, \ldots, k_{k_1}$ are the numbers of sons of the sons of the root and so we continue by generations as by lines. We provide the minimization in Maple (the source code is included at the end of this section). The most probable tree for several small $n$ can be seen in Figure 6.3. Unfortunately more explicit solution of this problem is a thing of future.

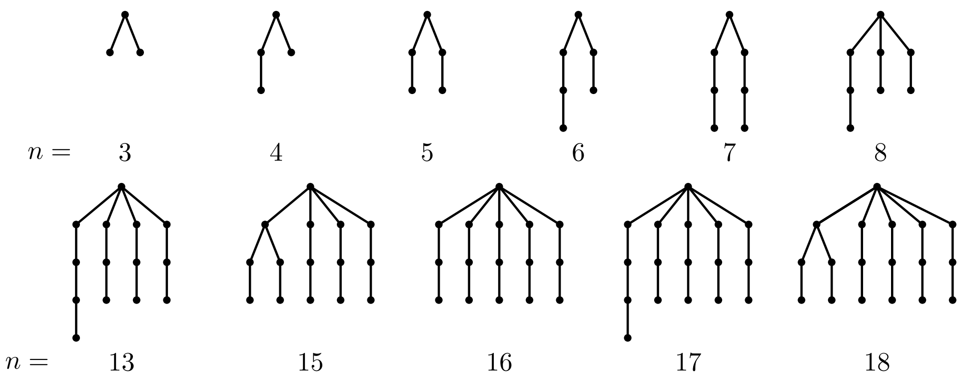



Figure 6.3: The shape of the tree with maximal probability for several $n$

Procedure for maximization (resp. minimization) in Maple:

```
> for i from 1 by 1 to 31 do beta[i]:=(1+i)/i/(i+2) end do;
> with(combinat):with(ListTools):

> for n from 1 by 1 to 30 do
> maxi:=0; maxid:=0;
>   a:=partition(n);
>   for j from 1 by 1 to numbpart(n) do
>     poc:=1; gen:=2; soucin:=exp(-2/3);poc2:=0; b:=Reverse(a[j]);
>     for k from 1 to nops(b) do
>        poc:=poc-1; poc2:=poc2+b[k];
>        soucin:=evalf(soucin/((b[k])!)*(beta[gen-1]*exp(-beta[gen]))^(b[k]));
>        if (poc=0) then poc:=poc2; poc2:=0; gen:=gen+1 end if;
>     end do;
>     if (evalf(soucin)>maxi) then maxi:=soucin; maxid:=b end if;
>     end do;
> maxii[n]:=maxi;
> maxidd[n]:=maxid;
> end do;

> for n from 1 by 1 to 30 do
> mini:=1; minid:=0;
>   a:=partition(n);
>   for j from 1 by 1 to numbpart(n) do
>     poc:=1; gen:=2; soucin:=exp(-2/3);poc2:=0; b:=(a[j]);
>     for k from 1 to nops(b) do
>        poc:=poc-1; poc2:=poc2+b[k];
>        soucin:=evalf(soucin/((b[k])!)*(beta[gen-1]*exp(-beta[gen]))^(b[k]) );
>        if (poc=0) then poc:=poc2; poc2:=0; gen:=gen+1 end if;
>     end do;
>     if (evalf(soucin)<mini) then mini:=soucin; minid:=b end if;
>     end do;
> minii[n]:=mini;
> minidd[n]:=minid;
> end do;
> for n from 1 to 30 do max[n]=maxii[n];maxd[n]=maxidd[n];
>                       min[n]=minii[n];mind[n]=minidd[n];
>                    end do;
```

# Appendix A

# Preprint of Article [5] (Required for Chapter 3)

**The Nearest Neighbor (NN$(d, 1)$) model - Estimates on the connectivity function:**
(S.Nanda and C.M. Newman, Dec. 1, 1997)

Basic Definitions and Results

Consider $X$, the density 1 Poisson process in $R^d$ with an "extra" Poisson particle located at the origin.

Let $\mathcal{G}_d$ denote the directed graph whose vertices are Poisson particles and in which there is a directed edge from $s \in X$ to $s' \in X$ if $s'$ is the NN of $s$. Let $\mathcal{G}$ denote the undirected NN$(d, 1)$ graph, as in Haagstrom and Meester.

$p_d(n, L)$ is the probability that there is a directed path in $\mathcal{G}_d$ starting at the origin, touching exactly $n$ distinct particles (besides the origin) and ending at a particle $s^{'}$ with $|s^{'}| > L$ (where $| - |$ denotes Euclidean length).

$\tau_d(L)$ is the probability that there is a path in $\mathcal{G}$ starting at the origin and ending at a particle $s^{'}$ with $|s^{'}| > L$.

Since in a Poisson process the distances between any two pairs of points are different a.s. each point has an a.s. unique NN. It is possible that two Poisson points are each others NN. In this case there is a closed mini-loop of directed edges between these two NN. In fact, since each Poisson point must have a NN, each cluster contains at least one such closed mini-loop. This would happen when a Poisson point $s$ is a NN of one or more other Poisson points, and when one of the latter is a NN of $s$ as well.

Further, each cluster contains exactly one closed mini-loop, as the existence of more than one such closed mini-loop would imply the existence of more than one NN for some Poisson point in the cluster, a contradiction. It is not possible to have any type of closed circuit other than the one described here as the existence of any closed circuit involving more than two Poisson points would contradict the fact that the lengths of successive directed edges are decreasing. The existence of an (a.s.) unique NN implies that along any path in $\mathcal{G}$ there is at most one change of direction (in the arrows of $\mathcal{G}_d$), and the change of direction takes place at one of the two Poisson points in a loop.

Theorem: There exist constants $C_1$, $C_2$, $L_0 \in (0, \infty)$ (depending only on d) such that

$$e^{-C_1 L(lnL)^{\frac{d-1}{d}}} \leq \tau_d(L) \leq e^{-C_2 L(lnL)^{\frac{d-1}{d}}} \quad for\, L \geq L_0.$$

The lower bound of the above theorem is based on the following. Note that $\tau_d(L) \geq p_d(n, L)$ for every $n$.

Proposition 1.1: There exist constants $b_1$ , $c_1 \in (0, \infty)$ (depending only on d) such that

$$p_d(n, L) \geq \frac{(b_1)^n}{n!} e^{-c_1 n (\frac{L}{n})^d}. \tag{A.1}$$

Remark: As will be seen from the proof of the proposition one may choose any $\theta \in (0, \frac{\pi}{4}]$ and take $b_1$ to be the probability that a uniformly distributed random point on the unit sphere $S^{d-1}$ falls into the "polar cap" of opening half-angle $\theta$. The corresponding $c_1$ may then be taken as $\frac{\pi_d}{(cos\theta)^d}$ where $\pi_d = V(B(0, 1))$, where $V$ denotes Euclidean volume and $B(x, r)$ denotes the (open) ball of radius $r$ centered at $x \in R^d$.

For the upper bound we need to bound not only $p_d(n, L)$ but also some closely related quantities that we now define.

For $j \in \{0, \dots, n\}$ we define $p_d(n, L, j)$ to be the probability that there are two directed paths in $\mathcal{G}_d$: one from 0 to some $s''$, touching exactly $j$ particles (besides 0) and one from some $s'$ to the same $s''$, touching exactly $n - j$ particles (besides $s''$) and such that $|s'| > L$. Thus $p_d(n, L) = p_d(n, L, n)$ and furthermore (by the properties of the directed graph $\mathcal{G}_d$ mentioned above and) by elementary arguments

$$\tau_d(L) \leq \sum_{n=1}^{\infty} \sum_{j=0}^{n} p_d(n, L; j). \tag{A.2}$$

The upper bound is based on the following:
Proposition 1.2: There exist constants $b_2$ , $c_2 \in (0, \infty)$ (depending only on d) such that

$$\begin{aligned} p_d(n, L, j) &\leq \frac{(b_2)^n}{j!(n-j)!} P(|W_1 + \cdots + W_n| \geq c_2 L) \\ &\leq \frac{(b_2)^n}{j!(n-j)!} P(|W_1| + \cdots + |W_n| \geq c_2 L) \end{aligned} \tag{A.3}$$

where $W_1, \cdots, W_n$ are i.i.d. $R^d$-valued random variables whose common probability density is $e^{-V(B(0,|w|))} = e^{-\pi_d |w|^d}$.

As a corollary to Prop 1.2 and the inequality (A.2) , we have the following.

Proposition 1.3: $\tau_d(L) \leq e^{2b_2} P(\mathcal{U} \geq c_2 L)$ where $\mathcal{U}$ is a random variable with $E(e^{r\mathcal{U}}) = e^{2b_2 E(e^{r|W_1|} - 1)}$.

Proof of Proposition 1.1: Given $(x_1, \dots, x_n) \in (R^d)^n$ we define $s_0 = 0$ and $s_i = x_1 + \cdots + x_i$ for $i = 1, \dots, n$. Let the set of points in $(R^d)^n$ satisfying the following three conditions be denoted by $\mathcal{S}$:

$$i)\ \ |x_1| \geq |x_2|.... \geq |x_n|$$
$$ii)\ \ |s_n| \geq L$$
$$iii)\ \ s_i \notin \bigcup_{j=1}^{i} B_j \text{ for } i = 1, \ldots, n$$

where $B_j = B(s_{j-1}, |x_j|)$ is the open ball centered at $s_{j-1}$ of radius $|x_j|$. Note that (because of $i)$ ) $iii)$ may be replaced by

$$iii^{'})\ \ B_l \bigcap \{s_0, \ldots, s_n\} = s_{l-1} \text{ for } l = 1, \ldots, n.$$

Then we have:

$$p_d(n, L) = \int_{\mathcal{S}} e^{-V(\bigcup_{j=1}^{n} B_j)} dx_1 \cdots dx_n. \tag{A.4}$$

Let $\mathcal{C}(\theta)$ be the polar cap of half angle $\theta \leq \frac{\pi}{4}$ in the unit sphere of $R^d$ with vertex at the origin; i.e.

$$\mathcal{C}(\theta) = \{x = (x^{(1)}, \ldots, x^{(d)}) \in R^d : |x| = 1 \text{ and } x^{(1)} \geq cos\theta\}.$$

Since

$$V(\bigcup_{j=1}^{n} B_j) \leq \sum_{j=1}^{n} V(B_j), \tag{A.5}$$

it follows (using the notation of Proposition 1.2, letting $S_i = W_1 + \cdots + W_i$ and $B_j = B(S_{j-1}, |W_j|)$) that:

$$\begin{aligned}
p_d(n, L) &\geq \int_{\mathcal{S}} \prod_{j=1}^{n} e^{-\pi_d |x_j|^d} dx_1 \cdots dx_n \\
&= P(|W_1| \geq \cdots \geq |W_n|, |\sum_{j=1}^{n} W_j| \geq L \text{ and } S_i \notin \bigcup_{j=1}^{i} B_j \text{ for each } i) \\
&\geq P(|W_1| \geq \cdots \geq |W_n|, |\sum_{j=1}^{n} W_j| \geq L \text{ and} \frac{W_i}{|W_i|} \in \mathcal{C}(\theta) \text{ for each } i) \\
&\geq (b(\theta))^n P(|W_1| \geq \cdots \geq |W_n|, \sum_{j=1}^{n} |W_j| cos(\theta) \geq L) \\
&= \frac{(b(\theta))^n}{n!} P(\sum_{j=1}^{n} |W_j| \geq (\frac{L}{cos\theta})) \\
&\geq \frac{(b(\theta))^n}{n!} [P(|W_1| \geq (\frac{L}{ncos(\theta)})]^n \\
&= \frac{(b(\theta))^n}{n!} e^{-(\frac{\pi_d}{(cos\theta)^d} n (\frac{L}{n})^d)}.
\end{aligned}$$

Here $b(\theta)$ represents the probability that $S_i$ lies in a cone of half angle $\theta \leq \frac{\pi}{4}$ with vertex at $S_{i-1}$. For $d = 2$ , $b(\theta) = \frac{\theta}{\pi} \leq \frac{1}{4}$.

To prove Proposition 1.2 we need Lemma 1.4 below which should be compared to (A.5). As a replacement for the equality (A.4), one can straightforwardly obtain the inequality,

$$p_d(n, L, j) \leq \int_{\mathcal{S}_j} e^{-V(\bigcup_{i=1}^{j} B_i \bigcup \bigcup_{i=j+1}^{n} B_i^{'})} dx_1 \cdots dx_n. \tag{A.6}$$

Here $B_i^{'} = B(s_i, |x_i|)$ and $\mathcal{S}_j$ is the set of points $(x_1, \ldots, x_n) \in (R^d)^n$ satisfying the three conditions (where again $s_0 = 0$ and $s_i = x_1 + \cdots + x_i$ ) :

$$i)\ |x_1| \geq \cdots \geq |x_j| \text{ and } |x_{j+1}| \leq \cdots \leq |x_n|$$

$$ii)\ |s_n| \geq L$$

$$iii^{'})\ B_l \bigcap \{s_o, \ldots, s_n\} = s_{l-1} \text{ for } l = 1, \ldots, j \text{ and}$$

$$B_l^{'} \bigcap \{s_o, \ldots, s_n\} = s_l \text{ for } l = j+1, \ldots, n.$$

<u>Lemma 1.4 :</u>There exists a finite constant $K_d$, depending only on the dimension $d$, such that

$$V(\bigcup_{i=1}^{j} B_i \bigcup \bigcup_{i=j+1}^{n} B_i^{'}) \geq \frac{1}{K_d}[\sum_{i=1}^{j} V(B_i) + \sum_{i=j+1}^{n} V(B_i^{'})]. \tag{A.7}$$

For $j = n$ we have $V(\bigcup_{i=1}^{n} B_i) \geq \frac{1}{K_d} \sum_{i=1}^{n} V(B_i)$.

*Proof* : Let $\tilde{B}_i = B_i$ for $i \leq j$ and $\tilde{B}_i = B_i^{'}$ for $i > j$ and let $q_i$ denote the center of $\tilde{B}_i$. Then from $iii^{'}$) we have that $q_i \notin \tilde{B}_l$ for $l \neq i$. Now

$$\begin{aligned} \sum_{l=1}^{n} V(\tilde{B}_l) &= \int_{R^d} \sum_{l=1}^{n} 1_{\tilde{B}_l}(x)\, dx \\ &= \int_{\bigcup_l \tilde{B}_l} \sum_i 1_{\tilde{B}_i}(x)\, dx \\ &\leq \int_{\bigcup_l \tilde{B}_l} K_d\, dx \\ &= K_d V(\bigcup_l \tilde{B}_l). \end{aligned} \tag{A.8}$$

The last inequality is a consequence of the fact that no point $x$ can lie in the intersection of more than $K_d$ of the balls $\tilde{B}_i$. To see this note that if this were not so then $x$ would be the closest point (among $\{x, q_1, \ldots q_n\}$) to more than $K_d$ of the $q_i$'s. But it is a standard (and fairly easy to show) fact that at most $K_d$ ( where $K_d$ is the kissing number) of the $q_i$'s can have a single point $x$ as their nearest neighbor (among the points $\{x, q_1, \ldots, q_n\}$).

¿From now on we call $K_d$ merely $K$.

_Remark_: For $d = 2\,,\ K = 6$.

Proof of Proposition 1.2: Starting with (A.6), we apply (A.7) to get

$$p_d(n, L, j) \le \int_{\mathcal{S}_j} e^{-\frac{1}{K}\sum_{i=1}^{j} V(B_i)} e^{-\frac{1}{K}\sum_{i=j+1}^{n} V(B_i')} dx_1 \cdots dx_n$$

$$= \int_{\mathcal{S}_j} e^{-\frac{1}{K}\pi_d \sum_{i=1}^{j} |x_i|^d} e^{-\frac{1}{K}\pi_d \sum_{i=j+1}^{n} |x_i|^d} dx_1 \cdots dx_n$$

$$= \int_{\mathcal{S}_j} \prod_{j=1}^{n} e^{-\frac{1}{K}\pi_d |x_j|^d} dx_1 \cdots dx_n.$$

By doing the change of variable $y_i = \frac{x_i}{(K)^{\frac{1}{d}}}$ for each $i$ the last integral becomes:

$$K^n \int_{\mathcal{S}'_j} \prod_{j=1}^{n} e^{-\pi_d |y_j|^d} dy_1 \cdots dy_n \tag{A.9}$$

where $\mathcal{S}'_j$ is the set of $(y_1, \ldots, y_n) \in (R^d)^n$ such that $(K^{\frac{1}{d}} y_1, \ldots, K^{\frac{1}{d}} y_n) \in \mathcal{S}_j$. To get an upper bound on (A.9) we simply drop the third condition in the definition of $\mathcal{S}_j$; i.e. we replace $\mathcal{S}'_j$ with $\mathcal{S}''_j$, the set of $(y_1, \ldots, y_n)$ satisfying only the two conditions:

$$i)\ \ |y_1| \ge \cdots \ge |y_j| \ \text{ and } \ |y_{j+1}| \le \cdots \le |y_n|$$

$$ii)\ \ |\sum_{i=1}^{n} K^{\frac{1}{d}} y_i| \ge L.$$

This yields the bound:

$$\begin{aligned} p_d(n, L, j) &\le \frac{(K)^n}{j!(n-j)!} P(|\sum_{i=1}^{n} W_i| \ge \frac{L}{K^{\frac{1}{d}}}) \\ &\le \frac{(K)^n}{j!(n-j)!} P(\sum_{i=1}^{n} |W_i| \ge \frac{L}{K^{\frac{1}{d}}}). \end{aligned}$$

Here the $W_i$'s are again i.i.d. random variables with common density $e^{-V(B(0,|w|))} = e^{-\pi_d |w|^d}$.

Proof of Proposition 1.3 : From (A.2) and (A.3) we have:

$$
\begin{aligned}
\tau_d(L) &\le \sum_{n=1}^{\infty}\sum_{j=0}^{n}\frac{(b_2)^n}{j!(n-j)!}P(|W_1+\cdots+W_n|\ge c_2L) \\
&= \sum_{n=1}^{\infty}\frac{(2b_2)^n}{n!}P(|W_1+\cdots+W_n|\ge c_2L) \\
&\le e^{2b_2}\sum_{n=0}^{\infty}e^{-2b_2}\frac{(2b_2)^n}{n!}P(|W_1|+\cdots+|W_n|\ge c_2L) \\
&= e^{2b_2}P(\mathcal{U}\ge c_2L). \qquad (A.10)
\end{aligned}
$$

The last equality follows by taking $\mathcal{U}$ to be the compound Poisson random variable $|W_1|+\cdots+|W_N|$ (where $N$ is independent of $|W_1|, |W_2|, \ldots$ and is Poisson with mean $2b_2$). By a standard calculation

$$E(e^{r\mathcal{U}}) = e^{2b_2E(e^{r|W_1|}-1)}. \qquad (A.11)$$

This completes the proof of Proposition 1.3.

The proof of the theorem now follows easily. Taking $n \approx \frac{L}{(lnL)^{\frac{1}{d}}}$ in (A.1) we get the lower bound in (A.1).

For the upper bound we use large deviation bounds on $P(\mathcal{U}\ge c_2L)$:

$$
\begin{aligned}
P(\mathcal{U}\ge c_2L) &\le \exp\{\inf_{r>0}\{ln(E(e^{r\mathcal{U}})-rL\}\} \\
&= \exp\{\inf_{r>0}\{2b_2E(e^{r|W_1|}-1)-rL\}\} \\
&\le \exp\{\inf_{r>0}\{2b_2{}'(e^{cr^{\frac{d}{d-1}}}-rL)\}. \qquad (A.12)
\end{aligned}
$$

The last inequality follows from the fact that $E(e^{r|W_1|}) \le e^{cr^{\frac{d}{d-1}}}$ for large $r$ (as can easily be shown). Taking $r \approx \alpha(lnL)^{\frac{d-1}{d}}$ for an appropriate constant $\alpha$ we get $\tau_d(L) \le e^{-C_2L(lnL)^{\frac{d-1}{d}}}$ for large L.

This completes the proof of the theorem.

# Bibliography

[1] O. Häggström, R. Meester: *Nearest neighbour and hard sphere models in continuum percolation*, Random Structures and Algorithm 9 (1996), p.295-315.

[2] J. F. C. Kingman: *Poisson Processes*, Oxford Clarendon Press 1995.

[3] C. W. Marshall: *Applied Graph Theory*, Wiley 1971.

[4] R. Meester, R. Roy: *Continuum Percolation*, Cambridge University Press 1996.

[5] S. Nanda, C. M. Newman: *The Nearest Neighbor Model* $\mathrm{NN}(d,1)$ - *Estimates on the connectivity function*, preprint