Poisson convergence in stochastic geometry via generalized size-bias coupling

Inaugural Dissertation
of the Faculty of Science,
University of Bern

Presented by

Federico Pianoforte
from Italy

Supervisor of the doctoral thesis:

Prof. Dr. Matthias Schulte

Institute of Mathematics,
Hamburg University of Technology

Poisson convergence in stochastic geometry via generalized size-bias coupling

Inaugural Dissertation
of the Faculty of Science,
University of Bern

Presented by

Federico Pianoforte
from Italy

Supervisor of the doctoral thesis:

Prof. Dr. Matthias Schulte

Institute of Mathematics,
Hamburg University of Technology

Accepted by the Faculty of Science

Bern, 17.09.2021

The dean:
Prof. Dr. Zoltan Balogh

Acknowledgement

First of all, I would like to thank my PhD advisor Matthias Schulte for his patience and support. Without Matthias' guidance, I would not have finished this manuscript.

I am grateful to all my colleagues from the Institute of Mathematical Statistics and Actuarial Science of the University of Bern. I deeply appreciate their kindness and I will never forget all the encouragement they gave me during this experience.

I am very thankful to Dominic Schuhmacher for refereeing this thesis. My sincere thanks go to Chinmoy Bhattacharjee and Sam Allen for reading some parts of this thesis and making valuable comments, and to Riccardo Turin for the collaboration during the last year.

Finally, I would like to thank all the teachers and professors who contributed to my educational training.

Abstract

This dissertation aims to investigate several aspects of the Poisson convergence: Poisson approximation, multivariate Poisson approximation, Poisson process approximation and weak convergence to a Poisson process.

The size-bias coupling is a powerful tool that, when combined with the Chen-Stein method, leads to many general results on Poisson approximation. We define an approximate size-bias coupling for integer-valued random variables by introducing error terms, and we combine it with the Chen-Stein method to compare the distributions of integer-valued random variables and Poisson random variables. In particular, we provide explicit bounds on the pointwise difference between the cumulative distribution functions. By these findings, we show approximation results in the Kolmogorov distance for minimal circumscribed radii and maximal inradii of stationary Poisson-Voronoi tessellations. Moreover, we compare the distributions of Poisson random variables and U-statistics with underlying Poisson processes or binomial point processes, which, in particular, allows us to approximate the rescaled minimum Euclidean distance between pairs of points of a Poisson process with midpoint in an observation window by an exponentially distributed random variable using the Kolmogorov distance.

A multivariate version of the size-bias coupling is employed to investigate the Gaussian approximation for random vectors by L. Goldstein and Y. Rinott. We extend the notion of approximate size-bias coupling for random variables to random vectors, and we combine it with the Chen-Stein method to investigate the multivariate Poisson approximation in the Wasserstein distance and the Poisson process approximation in a new metric defined herein. As an application, we obtain a bound on the Wasserstein distance between the sum of m-dependent Bernoulli random vectors and a Poisson random vector. Moreover, we consider point processes of U-statistic structure, that is, point processes that, once evaluated on a measurable set, become U-statistics. For point processes of U-statistic structure with an underlying Poisson process, we establish a Poisson process approximation result that is the analogue of the one shown by L. Decreusefond, M. Schulte, and C. Thäle with the Kantorovich–Rubinstein distance replaced by our new metric.

General criteria for the weak convergence of locally finite point processes to a Poisson process are derived from the relation between probabilities of two consecutive values of a Poisson random variable. P. Calka and N. Chenavier studied the limiting behavior of characteristic radii of homogeneous Poisson-Voronoi tessellations. By our general results, we extend and improve their findings by showing Poisson process convergence for point processes constructed using inradii and circumscribed radii of inhomogeneous Poisson-Voronoi tessellations.

Contents

Chapter 1

Introduction

The study of the asymptotic behavior of random elements and the derivation of approximation results are of great interest among probabilists and have considerable importance in all scientific fields, from astrophysics to genetics. Indeed, they permit the study of complicated distributions using simpler ones and hence they can be used to express non-trivial problems in a simple way with a certain level of accuracy. The Poisson distribution is often used to model the occurrences of events, that is, situations involving the sum of indicator functions. It is well known, for example, that the sum of n independent identically distributed Bernoulli trials with success probability $p = p(n) > 0$ such that $pn \sim \lambda$ can be approximated using a Poisson distribution with mean λ when n is large. Nowadays, Poisson random variables, and more generally multivariate Poisson random vectors and Poisson processes, are used to model many situations involving a large number of rare events, not necessarily independent. For instance, they are employed in extreme values theory, time series analysis, stochastic geometry and theory of summation (see e.g. [26, 49]).

This thesis investigates the limit behavior of several random elements, mostly taken from stochastic geometry problems. It establishes both limit theorems and (non-asymptotic) approximation results with Poisson random variables, Poisson random vectors and Poisson processes as limits.

In the first part of this dissertation, we compare the distributions of integer-valued random variables and Poisson random variables. For $\lambda > 0$ and a random variable X with values in $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, a possible way to study the distance between the distributions of X and a Poisson random variable P_λ with mean λ is via a sequence of error terms $q_k, k \in \mathbb{N}_0$, given by the equation

$$q_{k-1} = k\mathbb{P}(X = k) - \lambda\mathbb{P}(X + Z = k - 1), \quad k \in \mathbb{N}, \tag{1.1}$$

for some random variable Z with values in $\mathbb{Z}$ and defined on the same probability space as X. Intuitively, if $|q_k|, k \in \mathbb{N}_0$, are small and $|Z|$ is zero with high probability, then X behaves approximately like P_λ, while if both are zero, it follows the same distribution as P_λ. If $\mathbb{P}(Z + X \geq 0) = 1$, one can show that

$$\mathbb{E}[Xf(X)] = \lambda\mathbb{E}[f(X + Z + 1)] + \sum_{k=1}^{\infty} f(k)q_{k-1} \tag{1.2}$$

for all measurable f such that $\mathbb{E}[|Xf(X)|] < \infty$. Observe that if $q_k = 0$ for all $k \in \mathbb{N}_0$ and $\mathbb{E}[X] = \lambda$, the random variable $X+Z+1$ in the previous equation has the size-biased distribution of X and the condition $\mathbb{P}(Z + X \geq 0) = 1$ is always satisfied. In this case the total variation distance between X and P_λ is bounded by

$$d_{TV}(X, P_\lambda) = \sup_{A \subset \mathbb{N}_0} \left|\mathbb{P}(X \in A) - \mathbb{P}(P_\lambda \in A)\right| \leq (1 \wedge \lambda)\mathbb{E}[|Z|],$$

where $a \wedge b = \min\{a, b\}$ for $a, b \in \mathbb{R}$. This inequality was first proven in [9] for the sum of Bernoulli random variables and later generalized to all non-negative integer-valued random variables (see e.g. [60, Theorem 4.13]). On the other hand, if the error terms q_k are not zero, by (1.2) we can interpret the random variable $X + Z + 1$ as an approximate size-bias coupling of X. In this case, we prove that

$$d_{TV}(X, P_\lambda) \leq (1 \wedge \lambda)\mathbb{E}[|Z|] + \left(1 \wedge \frac{1}{\sqrt{\lambda}}\right) \sum_{k=0}^{\infty} |q_k|.$$

Observe that, in contrast to the previous classical result, for this bound it is not required that $\mathbb{E}[X] = \lambda$. The employment of the error terms is sometimes necessary, as it is not always possible to find an exact size-bias coupling, or equivalently, a Z for which the q_k defined by (1.1) with $\lambda = \mathbb{E}[X]$ are zero for all k.

By means of the previous results, and more generally using Poisson approximation results, one can study the weak convergence for minima or maxima of collections of random variables. This is possible because the tail distribution of an exponential random variable E_1 with mean 1 can be expressed using $\mathbb{P}(P_\lambda = 0), \lambda \geq 0$. Hence the difference between the tail distribution functions of the minimum of a collection of non-negative random variables $Y_1, \dots Y_n, n \in \mathbb{N}$, and E_1 can be bounded using $X_\lambda = \sum_{j=1}^n \mathbf{1}\{Y_j \leq \lambda\}$ by

$$\left|\mathbb{P}\Big(\min_{j=1,\dots,n} Y_j > \lambda\Big) - \mathbb{P}(E_1 > \lambda)\right| = |\mathbb{P}(X_\lambda = 0) - \mathbb{P}(P_\lambda = 0)| \leq d_{TV}(X_\lambda, P_\lambda). \qquad (1.3)$$

Then, by showing that for any fixed $\lambda \geq 0$, the total variation distance between X_λ and P_λ goes to 0 as n increases to infinity, it is possible to prove that the minimum of $Y_1, \dots, Y_n$ converges weakly to E_1. Similar arguments also apply if we replace E_1 by e.g. a Weibull random variable and P_λ by $P_{f(\lambda)}$ for some positive real-valued function f. However, since the total variation distance between X_λ and P_λ usually depends on λ, the previous inequality does not permit the differences on the left-hand side to be bounded uniformly in $\lambda \geq 0$. As a result, the estimates for the total variation distance do not lead to a bound for the Kolmogorov distance

$$d_K\Big(\min_{j=1,\dots,n} Y_j, E_1\Big) = \sup_{\lambda \geq 0} \left|\mathbb{P}\Big(\min_{j=1,\dots,n} Y_j > \lambda\Big) - \mathbb{P}(E_1 > \lambda)\right|$$

between the minimum of $Y_1, \dots, Y_n$ and E_1. To overcome this problem, we prove for $\lambda > 0$ that the difference of the probabilities at 0 can be bounded by

$$\begin{aligned}|\mathbb{P}(X_\lambda = 0) - \mathbb{P}(P_\lambda = 0)| \leq \frac{1}{\lambda}\mathbb{E}[|Z|] + (1 \wedge \lambda)\mathbb{E}\big[|Z|\mathbf{1}\{X_\lambda - Z_- = 0\}\big] \\ + \left(1 \wedge \frac{1}{\lambda}\right)|q_0| + \left(1 \wedge \frac{1}{\lambda^2}\right) \sum_{k=1}^{\infty} |q_k|,\end{aligned}$$

where Z_- is the negative part of a random variable Z in $\mathbb{Z}$ and the sequence $q_k, k \in \mathbb{N}_0$, is defined by (1.1) with $X = X_\lambda$, and Z and λ as above. As we shall see in many examples, for large λ and a proper choice of Z, this inequality establishes a better bound than the total variation distance for the left-hand side of (1.3), and it permits the approximation of minima or maxima of collections of random variables by suitable distributions in the Kolmogorov distance.

In this work, we also estimate the pointwise difference between the cumulative distribution functions of X_λ and P_λ. Since the probability that the m-smallest element of $Y_1, \ldots, Y_n$ is greater than a given threshold λ corresponds to $\mathbb{P}(X_\lambda \leq m-1)$ for all $m = 1, \ldots, n$, one can then derive Poisson approximation results for the m-smallest (largest) element of $Y_1, \ldots, Y_n$ by our general results and similar arguments to the one employed to study the distribution of $\min_{j=1,\ldots,n} Y_j$.

For a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with values in $\mathbb{N}_0^d$, $d \in \mathbb{N}$, and $(\lambda_1, \ldots, \lambda_d) \in [0,\infty)^d$, a multivariate version of (1.1) is given by the equation

$$\begin{aligned} q^{(i)}_{k_1,\ldots,k_i} &= k_i \mathbb{P}\big((X_1, \ldots, X_i) = (k_1, \ldots, k_i)\big) \\ &\quad - \lambda_i \mathbb{P}\big((X_1, \ldots, X_i) + \mathbf{Z}^{(i)} = (k_1, \ldots, k_{i-1}, k_i - 1)\big), \end{aligned} \tag{1.4}$$

for $k_1, \ldots, k_i \in \mathbb{N}_0$ with $k_i \neq 0$, and $i = 1, \ldots, d$, where $\mathbf{Z}^{(i)}$ is a random vector with values in $\mathbb{Z}^i$ defined on the same probability space as $\mathbf{X}$. Intuitively, when the error terms $q^{(i)}_{k_1,\ldots,k_i}$ are small in absolute value and $\mathbf{Z}^{(i)}$ is the null vector with high probability for all i, $(X_1, \ldots, X_d)$ behaves like a Poisson random vector $\mathbf{P}$ with mean $\mathbb{E}[\mathbf{P}] = (\lambda_1, \ldots, \lambda_d)$, that is, a random vector in $\mathbb{N}_0^d$ whose components are independent and Poisson distributed random variables with means $\lambda_i, i = 1, \ldots, d$. Let $\mathrm{Lip}^d(1)$ denote the set of Lipschitz functions $g : \mathbb{N}_0^d \to \mathbb{R}$ with Lipschitz constant bounded by 1 with respect to the metric induced by the 1-norm. In the second part of this dissertation, we derive an explicit bound on the Wasserstein distance

$$d_W(\mathbf{X}, \mathbf{P}) = \sup_{g \in \mathrm{Lip}^d(1)} \big|\mathbb{E}[g(\mathbf{X})] - \mathbb{E}[g(\mathbf{P})]\big|$$

between $\mathbf{X}$ and $\mathbf{P}$ that depends on the sum of the absolute values of the error terms and on the L^1-norm of the components of the random vectors $\mathbf{Z}^{(i)}, i = 1, \ldots, d$. In [28], the size-bias coupling for random vectors is defined as a particular case of the more general definition of size-bias coupling for a collection of non-negative random variables and is employed in the Gaussian approximation of random vectors. Similarly to what we have seen for the one-dimensional case, the family of random vectors

$$\mathbf{Y}^{(i)} = \mathbf{Z}^{(i)} + (X_1, \ldots, X_{i-1}, X_i + 1), \quad i = 1, \ldots, d,$$

can be interpreted as an approximate size-bias coupling of $\mathbf{X}$, and in the case when $\mathbf{Z}^{(i)}, i = 1, \ldots, d$, are such that the error terms $q^{(i)}_{k_1,\ldots,k_i}$ defined by (1.4) with $(\lambda_1, \ldots, \lambda_d) = \mathbb{E}[\mathbf{X}]$ are all zeros, they fulfill a slightly different requirement to that in the definition of size-bias coupling from [28].

In this work, many problems from stochastic geometry are considered. In stochastic geometry, one is often interested in random geometric structures, such as random tessellations or geometric random graphs. These structures depend on random points in a measurable space, which is often convenient to describe using a point process. Intuitively, a point process ξ on a measurable space $(\mathbb{X}, \mathcal{X})$ is a random collection of points in $\mathbb{X}$, and for any measurable set $B \in \mathcal{X}$, the random variable $\xi(B)$ gives the number of elements from ξ that are in B. When the point process ξ is constructed from n independent identically distributed (i.i.d.) random points, then ξ is called a binomial point process.

In many situations, the random configurations consists of a random number of points, which could also be infinite. For these, a better model is given by a Poisson process. A point process η on a measure space $(\mathbb{X}, \mathcal{X}, \lambda)$ is a Poisson process with intensity measure λ if, for all measurable and disjoint sets $B_1, \dots, B_k$, $\eta(B_1), \dots, \eta(B_k)$ are independent and Poisson distributed random variables with means $\lambda(B_1), \dots, \lambda(B_k)$.

The distance between the distributions of a point process ξ and a Poisson process η on $\mathbb{X}$ is usually measured using the total variation distance or the Kantorovich-Rubistein distance, which are the analogues of the total variation and the Wasserstein distance for point processes. In this work, we define a new distance between point processes ξ and ζ on $\mathbb{X}$ with finite intensity measure as

$$d_\pi(\xi, \zeta) = \sup_{(A_1,\dots,A_d)\in\mathcal{X}^d_{\text{disj}},\, d\in\mathbb{N}} d_W\big((\xi(A_1), \dots, \xi(A_d)), (\zeta(A_1), \dots, \zeta(A_d))\big),$$

where for $d \in \mathbb{N}$, $\mathcal{X}^d_{\text{disj}}$ denotes the set of all d-tuples of disjoint measurable sets in $\mathbb{X}$. Then, by applying the bound mentioned above for the Wasserstein distance to $d_W(\mathbf{X}, \mathbf{P})$ with $\mathbf{X} = (\xi(A_1), \dots, \xi(A_d))$ and $\mathbf{P} = (\eta(A_1), \dots, \eta(A_d))$, we derive a Poisson process approximation result for point processes with finite intensity measure.

In a large variety of applications in the literature, the point processes are defined on a locally compact second countable Hausdorff space S and are considered to be locally finite, which means that they take finite values on compact sets (almost surely). In other words, they are random elements in the space of locally finite counting measures on S. Since this space is Polish, for these point processes it is also possible to derive asymptotic results using the weak convergence. In the literature, several classical results are available to establish weak convergence (see e.g. [33]). Following the intuition given at the beginning of the introduction about (1.1), in the last part of this dissertation we show that a tight sequence of locally finite point processes ξ_n, $n \in \mathbb{N}$, on S satisfies

$$\lim_{n\to\infty} k\mathbb{P}(\xi_n(B) = k) - \lambda(B)\mathbb{P}(\xi_n(B) = k - 1) = 0, \quad k \in \mathbb{N},$$

for any $B \subset S$ in a certain family of sets and some locally finite measure λ, if and only if ξ_n converges in distribution to a Poisson process with intensity measure λ. As a consequence of this result, we obtain a general criterion for the weak convergence of locally finite point processes constructed from Poisson processes and binomial point processes.

The non-asymptotic general results mentioned above are obtained by the Chen-Stein method. This method is a powerful tool for computing an error bound when approximating probability distributions by the Poisson distribution. It is worth mentioning that our multivariate Poisson approximation results are established by applying the Chen-Stein method to each component of the random vectors.

To demonstrate the versatility of our findings we apply them to several examples. The applications can be summarized as follows:

- ***U*-statistics**: A U-statistic is defined as the sum of a real-valued k-variate symmetric function h evaluated over all possible combinations of k distinct points from a random sample. Bounds for the Poisson approximation of U-statistics constructed from binomial point processes or Poisson processes with h being either 0 or $1/k!$ are obtained in [51] and [66]. Moreover, limit theorems for the extreme values of U-statistics were considered in [36], though without providing approximation results with respect to any distance. We derive similar Poisson approximation results for U-statistics and establish explicit bounds on the pointwise difference between the cumulative distribution functions of a U-statistic and a Poisson random variable. Our theoretical results permit the approximation of minima or maxima of U-statistics by suitable distributions in the Kolmogorov distance. As an application of our findings, we approximate the minimum inter-point distance between the points of a Poisson process with midpoint in an observation window by an exponential distribution.

The accuracy of the Poisson process approximation of point processes of U-statistic structure, i.e. point processes that, once evaluated on a measurable set, become U-statistics, is investigated in [22]. For an underlying Poisson process, we establish the analogue of [22, Theorem 3.1] with the Kantorovich–Rubinstein distance replaced by the distance d_π.

- **Poisson-Voronoi tessellations**: Given a cell of a Voronoi tessellation, the circumscribed radius is the smallest radius for which the ball centered at the nucleus contains the cell, while the inradius is the largest radius for which the ball centered at the nucleus is contained in the cell. For a homogeneous Poisson-Voronoi tessellation generated by a stationary Poisson process with intensity $t > 0$, the limiting distributions as $t \to \infty$ of the maximal inradius and the minimal circumscribed radius of cells with nucleus within an observation window were derived in [15]. In our work, we extend these findings in two directions. Firstly, we prove Poisson process convergence of point processes constructed from inradii (circumscribed radii) of inhomogeneous Poisson-Voronoi tessellations. This generalizes the mentioned results from [15] to inhomogeneous Poisson-Voronoi tessellations and allows us to deal with the m-th largest (or smallest) value or combinations of several order statistics. Secondly, we derive approximation results for certain transforms of the maximal inradius and the minimal circumscribed radius of cells with nucleus in an observation window for stationary Poisson-Voronoi tessellations.

- **k-runs:** A k-run means at least k successes in a row in a sequence of trials. When the successes are generated by i.i.d. Bernoulli random variables, it is shown in [45] that the difference between the probability that there are no more than v non-overlapping k-runs among n trials, and $\mathbb{P}(P_\alpha \leq v)$ for a certain Poisson random variable P_α, after taking the supremum over all $k = 1, \dots, n$, behaves asymptotically like $O(\log n/n)$. We improve this result by finding an explicit (non-asymptotic) bound for the supremum of the difference.

  The Poisson approximation of the number of non-overlapping k-runs in a sequence of n i.i.d. Bernoulli random variables has been investigated by several authors; see e.g. the survey [46]. It is known that the first arrival time of a k-run in a sequence of i.i.d. Bernoulli random variables multiplied by the probability of having a k-run converges weakly to an exponentially distributed random variable as the success probability converges to zero. We extend this result to the situation when the Bernoulli random variables are weakly dependent. Moreover, we show that the rescaled starting points of the k-runs behave like a Poisson process if the success probabilities converge to zero and if some independence assumptions are satisfied.

- **Multinomial distribution**: The multivariate Poisson approximation of multinomial random vectors, and more generally of sums of independent Bernoulli random vectors, has already been investigated by many authors using the total variation distance; see e.g. [57] and references therein. In contrast to what is usually done in the literature, we assume that the Bernoulli random vectors are m-dependent, and we study the multivariate Poisson approximation of their sum in the Wasserstein distance.

This thesis is organised in the following way. In Chapter 2, we introduce some important properties of point processes and present the Chen-Stein method and size-bias coupling. The remaining chapters are based on the following papers:

- *Pianoforte and Schulte 2021*: Poisson approximation with applications to stochastic geometry. Preprint.
- *Pianoforte and Turin 2021*: Multivariate Poisson and Poisson process approximations with applications to Bernoulli sums and U-statistics. Preprint.
- *Pianoforte and Schulte 2021*: Criteria for Poisson process convergence with applications to inhomogeneous Poisson-Voronoi tessellations. Preprint.

In Chapter 3, we compare the distributions of integer-valued random variables and Poisson random variables. We consider the total variation and the Wasserstein distance and provide, in particular, explicit bounds on the pointwise difference between the cumulative distribution functions. In Chapter 4, we investigate the multivariate Poisson approximation of random vectors in the Wasserstein distance and the Poisson process approximation of point processes with finite intensity measure in the new metric d_π. Finally, in Chapter 5, we study the weak convergence of point processes to a Poisson process.

Chapter 2

Preliminaries

This chapter is organised as follows. In the first section, we provide some basic notation that will be used hereafter without necessarily being defined. In the second section, we introduce point processes and locally finite point processes, following the approaches in the textbooks [38] and [32], respectively. In the third section, we illustrate the Chen-Stein method, mirroring what is done in [52]. Finally, the fourth section is devoted to an introduction to size-bias coupling, which summarizes some results given in the survey [60].

2.1 Notation

Let $(\mathbb{X}, \mathcal{X})$ be a measurable space. The integral of a measurable function $f : \mathbb{X} \to \mathbb{R}$ with respect to a measure μ on $\mathbb{X}$ is written as

$$\int_{\mathbb{X}} f(x) d\mu(x).$$

We say that f belongs to $L^1(\mu)$ if

$$\int_{\mathbb{X}} |f(x)| d\mu(x) < \infty.$$

Throughout this work, $\mathbb{N}$ is the set of positive integers and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. For $n \in \mathbb{N}$, we denote by μ^n the n-fold product measure of μ on the space $\mathbb{X}^n$ endowed with the σ-field generated by $\mathcal{X}^n$. Analogously, given two measures μ_1 and μ_2 defined on measurable spaces $(\mathbb{X}_1, \mathcal{X}_1)$ and $(\mathbb{X}_2, \mathcal{X}_2)$, respectively, we write $\mu_1 \times \mu_2$ for the product measure of μ_1 and μ_2 on $\mathbb{X}_1 \times \mathbb{X}_2$ endowed with the σ-field generated by $\mathcal{X}_1 \times \mathcal{X}_2$. A measure μ on $(\mathbb{X}, \mathcal{X})$ is said to be σ-finite if $\mathbb{X}$ can be written as countable union of measurable sets $A_i, i \in \mathbb{N}$, such that $\mu(A_i) < \infty$ for all $i \in \mathbb{N}$. The Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, where $\mathcal{B}(\mathbb{R}^d)$ stands for the Borel σ-field, is denoted by λ_d, and we use the shorthand notation dx for the integration with respect to the Lebesgue measure.

For a finite set A, we write $|A|$ for its cardinality. We use the shorthand notations $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$ for $a, b \in \mathbb{R}$, and we indicate by X_+ and X_- the positive and negative part of a random variable X, respectively. Whenever we write $\alpha > 0$, it is understood that $\alpha \in (0, \infty)$. By $\xrightarrow{d}$, we denote the convergence in distribution, and by $\overset{d}{=}$, the equality in distribution. $\mathbb{P}_Z$ stands for the probability distribution of a random element Z in $(\mathbb{X}, \mathcal{X})$.

In Chapter 4, since we will focus on the components of vectors, for convenience an element in $\mathbb{R}^d, d \in \mathbb{N}$, is denoted using the bold notation $\mathbf{x} = (x_1, \ldots, x_d)$.

2.2 Point processes

2.2.1 Point processes on a measurable space

Throughout this subsection, we always assume that $\mathbb{X}$ is a measurable space endowed with a σ-field $\mathcal{X}$. We define $\mathbf{N}_{\mathbb{X}}$ as the space of all σ-finite counting measures on $\mathbb{X}$. The set $\mathbf{N}_{\mathbb{X}}$ is equipped with the σ-field $\mathcal{N}_{\mathbb{X}}$ generated by the collection of all subsets of $\mathbf{N}_{\mathbb{X}}$ of the form

$$\{\mu \in \mathbf{N}_{\mathbb{X}} \ : \ \mu(B) = k\}, \quad B \in \mathcal{X}, k \in \mathbb{N}_0.$$

This means that $\mathcal{N}_{\mathbb{X}}$ is the smallest σ-field on $\mathbf{N}_{\mathbb{X}}$ such that $\mu \mapsto \mu(B)$ is measurable for all $B \in \mathcal{X}$.

Definition 2.2.1. *A point process on $\mathbb{X}$ is a random element in $(\mathbf{N}_{\mathbb{X}}, \mathcal{N}_{\mathbb{X}})$.*

From now on, we write $(\Omega, \mathcal{F}, \mathbb{P})$ for the underlying probability space. If ξ is a point process on $\mathbb{X}$ and $B \in \mathcal{X}$, we denote by $\xi(B)$ the mapping $\omega \mapsto \xi(\omega)(B)$. The intensity measure of a point process ξ on $\mathbb{X}$ is the measure μ defined by $\mu(B) = \mathbb{E}[\xi(B)], B \in \mathcal{X}$. A point process ξ is said to be finite if $\xi(\mathbb{X}) < \infty$ almost surely.

In order to study the distribution of a point process it is convenient to consider its finite dimensional distributions (see [38, Proposition 2.10]).

Proposition 2.2.2. *Let ξ and ζ be point processes on $\mathbb{X}$. Then, $\xi \overset{d}{=} \zeta$ if and only if*

$$(\xi(B_1), \ldots, \xi(B_m)) \overset{d}{=} (\zeta(B_1), \ldots, \zeta(B_m))$$

for all $m \in \mathbb{N}$ and all pairwise disjoint sets $B_1, \ldots, B_m \in \mathcal{X}$.

In [38, Proposition 2.10] is also established that two point processes have the same distribution if their Laplace functionals coincide or if, the integrals of any positive measurable function on $\mathbb{X}$ with respect to the point processes have the same distribution.

Next, let us introduce binomial point processes and Poisson processes. We denote by δ_x the Dirac measure concentrated at $x \in \mathbb{X}$.

Definition 2.2.3. *For $n \in \mathbb{N}$, let $X_1, \ldots, X_n$ be i.i.d. random elements in $\mathbb{X}$. Then, we call the random element*

$$\beta_n = \sum_{i=1}^{n} \delta_{X_i}$$

in $\mathbf{N}_{\mathbb{X}}$ a binomial point process.

Note that, if the X_i are distributed according to a probability measure $\mathbb{Q}$, then $\beta_n(A)$ follows a binomial distribution with parameters n and $\mathbb{Q}(A)$ for any $A \in \mathcal{X}$.

Definition 2.2.4. *Let λ be a σ-finite measure on $\mathbb{X}$. A Poisson process with intensity measure λ is a point process on $\mathbb{X}$ with the following properties:*

(i) $\eta(B)$ follows a Poisson distribution with mean $\lambda(B)$ for all $B \in \mathcal{X}$.

(ii) $\eta(B_1), \ldots, \eta(B_n)$ are independent for disjoint sets $B_1, \ldots, B_n \in \mathcal{X}$, $n \in \mathbb{N}$.

The properties (i) and (ii) are not independent of each other. In fact, [38, Theorem 6.10] and [38, Theorem 6.12] show that under certain extra conditions, either of the defining properties of the Poisson process implies the other.

[38, Theorem 3.6] establishes that for any σ-finite measure λ on $\mathbb{X}$ there exists a Poisson process η on $\mathbb{X}$ with intensity measure λ. Furthermore, [38, Corollary 3.7] proves that there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting random elements $X_1, X_2, \ldots$ in $\mathbb{X}$ and a random variable τ in $\mathbb{N}_0 \cup \{\infty\}$ such that

$$\eta \overset{d}{=} \sum_{n=1}^{\tau} \delta_{X_n}, \tag{2.1}$$

with the standard convention $\eta = 0$ if $\tau = 0$. Thus, we can think of η as the random set $(X_n)_{n=1}^{\tau}$ in $\mathbb{X}$, where it is allowed that two different elements in $(X_n)_{n=1}^{\tau}$ are the same element in $\mathbb{X}$. In the following we always consider Poisson processes represented as in (2.1). More generally, a point process is said to be proper if it can be written as in (2.1). By [38, Corollary 6.5], we obtain that any point process on a Borel space satisfying some σ-finite conditions (e.g. locally finiteness) is proper. Recall that a Borel space is a measurable space that can be identified to a Borel subset of $\mathbb{R}$ by a measurable bijection. An example of a Borel space is any locally compact second countable Hausdorff space.

For any non-negative measurable function $f : \mathbb{X} \to [0, \infty)$ and a proper point process ξ on $\mathbb{X}$, we indicate the random integral of f with respect to ξ by

$$\xi(f) = \sum_{x \in \xi} f(x) = \int_{\mathbb{X}} f(x) d\xi(x),$$

From [38, Proposition 2.7] it follows that $\xi(f)$ is a random variable. Since ξ is proper, for almost surely every $\omega \in \Omega$, we may write $\xi(\omega) = (x_i)_{i \in I}$ for some $x_i = x_i(\omega) \in \mathbb{X}$ and with $I = I(\omega)$ at most countable. Given a non-negative measurable function $g : \mathbb{X}^k \to [0, \infty)$ with $\mathbb{X}^k$ equipped with the σ-field generated by $\mathcal{X}^k, k \in \mathbb{N}$, we define

$$\sum_{(x_1,\ldots,x_k) \in \xi^k_{\neq}} g(x_1, \ldots, x_k) = \sum_{\substack{i_1,\ldots,i_k \in I: \\ i_{l_1} \neq i_{l_2} \forall l_1, l_2 \in \{1,\ldots,k\}, l_1 \neq l_2}} g(x_{i_1}, \ldots, x_{i_k}),$$

Thus, $\xi^k_{\neq}$ represents the set of all k-tuples of distinct points from ξ (where it is possible that two distinct elements from ξ are the same element in $\mathbb{X}$). For $k = 1$, we use the convention $\xi = \xi^1_{\neq}$. The random sum $\sum_{(x_1,\ldots,x_k) \in \xi^k_{\neq}} g(x_1, \ldots, x_k)$ corresponds to $\xi^{(k)}(g)$, where $\xi^{(k)}$ is the so called k-th factorial measure of ξ (see [38, Section 4.2]), which is a proper point process on $\mathbb{X}^k$ because ξ is proper. Whence, again by [38, Proposition 2.7], $\sum_{(x_1,\ldots,x_k) \in \xi^k_{\neq}} g(x_1, \ldots, x_k)$ is a random variable.

A characterization for the distribution of a Poisson process is given by the Mecke equation (see [38, Theorem 4.1]).

Proposition 2.2.5. *Let λ be an σ-finite measure on $\mathbb{X}$ and η be a point process on $\mathbb{X}$. Then η is a Poisson process with intensity measure λ if and only if*

$$\mathbb{E}\Big[\sum_{x \in \eta} f(x, \eta)\Big] = \int_{\mathbb{X}} \mathbb{E}[f(x, \eta + \delta_x)] d\lambda(x)$$

for all non-negative measurable functions $f : \mathbb{X} \times \mathbf{N}_{\mathbb{X}} \to [0, \infty)$.

Finally, we state the multivariate version of the Mecke formula (see [38, Theorem 4.4]).

Proposition 2.2.6. *Let η be a Poisson process on $\mathbb{X}$ with σ-finite intensity measure λ. Then,*

$$\mathbb{E}\Big[\sum_{(x_1,\ldots,x_k)\in\eta^k_{\neq}} f(x_1,\ldots,x_k,\eta)\Big] = \int_{\mathbb{X}} \mathbb{E}[f(x_1,\ldots,x_k,\eta+\sum_{i=1}^k \delta_{x_i})]\, d\lambda^k(x_1,\ldots,x_k)$$

for all non-negative measurable functions $f : \mathbb{X}^k \times \mathbf{N}_{\mathbb{X}} \to [0,\infty)$ and $k \in \mathbb{N}$.

2.2.2 Locally finite point processes

Let S be a locally compact second countable Hausdorff space, abbreviated as lcscH space, equipped with the Borel σ-field $\mathcal{S}$. A topological space is second countable if its topology has a countable basis, and it is locally compact if every point has an open neighborhood whose topological closure is compact.

We denote by $\mathcal{N}(S)$ the space of all locally finite counting measures on S. Recall that a measure μ on S is locally finite if, for any $x \in S$ there exists an open set $A \subset S$ with $x \in A$ such that $\mu(A) < \infty$. $\mathcal{N}(S)$ is equipped with the corresponding trace σ-field of $\mathbf{N}_S$. The σ-field of $\mathcal{N}(S)$ coincides with the Borel σ-field for the vague topology, which is generated by the mappings

$$\pi_f : \mu \mapsto \mu(f) = \int_S f(x) d\mu(x), \quad f \in C_K^+(S),$$

where $C_K^+(S)$ denotes the set of non-negative and continuous functions with compact support (see [32, Theorem A2.3-(iv)]). $\mathcal{N}(S)$ endowed with the vague topology is a Polish space, that is, a separable completely metrizable topological space (see [32, Theorem A2.3-(i)]).

Definition 2.2.7. *A locally finite point process on S is a random element in $\mathcal{N}(S)$ equipped with the trace σ-field of $\mathbf{N}_S$.*

Note that any point process on S with locally finite intensity measure is locally finite. We denote by $\widehat{\mathcal{S}}$ the family of relatively compact Borel sets from S. For a point process ξ on S, we define

$$\widehat{\mathcal{S}}_\xi = \{B \in \widehat{\mathcal{S}} \,:\, \xi(\partial B) = 0 \text{ a.s.}\},$$

where ∂B indicates the boundary of B. Observe that, if λ denotes the intensity measure of ξ, then

$$\widehat{\mathcal{S}}_\xi = \widehat{\mathcal{S}}_\lambda = \{B \in \widehat{\mathcal{S}} \,:\, \lambda(\partial B) = 0\}.$$

A locally finite point process ξ is said to be simple if

$$\mathbb{P}(\xi(\{x\}) \leq 1 \text{ for all } x \in S) = 1.$$

We say that a measure λ on S is non-atomic if $\lambda(\{x\}) = 0$ for all $x \in S$. It is possible to verify whether a Poisson process is simple by checking if its intensity measure is non-atomic (see [38, Proposition 6.9]).

Lemma 2.2.8. *Let η be a Poisson process on S with locally finite intensity measure λ. Then η is simple if and only if λ is non-atomic.*

Since $\mathcal{N}(S)$ is a Polish space, it follows from [32, Theorem 16.3] that a sequence of locally finite point processes is tight if and only if it is relatively compact in distribution, i.e., every subsequence has a further subsequence that converges in distribution. [32, Lemma 16.4] guarantees that continuous mappings preserve tightness, and [32, Lemma 16.15] gives the following tightness criterion.

Lemma 2.2.9. *Let $\xi_1, \xi_2, \ldots$ be locally finite point processes on S. Then the sequence $\xi_n, n \in \mathbb{N}$, is tight if and only if $\xi_n(B), n \in \mathbb{N}$, is tight in $\mathbb{R}$ for every $B \in \widehat{\mathcal{S}}$.*

In the last part of this subsection, we give some general weak convergence criteria for point processes. The proof of the following result is given in [32, Theorem 16.16].

Proposition 2.2.10. *Let $\xi, \xi_1, \xi_2, \ldots$ be locally finite point processes on S. Then these conditions are equivalent:*

(i) $\xi_n \xrightarrow{d} \xi$.

(ii) $\xi_n(f) \xrightarrow{d} \xi(f)$ *for all* $f \in C_K^+(S)$.

(iii) $(\xi_n(B_1), \ldots, \xi_n(B_k)) \xrightarrow{d} (\xi(B_1), \ldots, \xi(B_k))$ *for all* $B_1, \ldots, B_k \in \widehat{\mathcal{S}}_\xi, k \in \mathbb{N}$.

If ξ is a simple point process, it is also equivalent that

(iv) $\xi_n(B) \xrightarrow{d} \xi(B)$ for all $B \in \widehat{\mathcal{S}}_\xi$.

A non-empty class $\mathcal{U}$ of subsets of S is called a ring if it is closed under finite unions and intersections, as well as under proper differences. A ring $\mathcal{U}$ is said to be dissecting if any open set $G \subset S$ can be written as a countable union of sets in $\mathcal{U}$, and every relatively compact set $B \in \widehat{\mathcal{S}}$ is covered by finitely many sets in $\mathcal{U}$.

When $\xi, \xi_1, \xi_2, \ldots$ are locally finite point processes and ξ is simple, $\xi_n \xrightarrow{d} \xi$ follows already from the one-dimensional convergence $\xi_n(U) \xrightarrow{d} \xi(U)$ with U restricted to a dissecting ring $\mathcal{U} \subset \widehat{\mathcal{S}}_\xi$. In fact, this is a consequence of the following result (see [33, Theorem 4.15]).

Proposition 2.2.11. *Let $\xi, \xi_1, \xi_2, \ldots$ be locally finite point processes on S, where ξ is simple, and fix a dissecting ring $\mathcal{U} \subset \widehat{S}_\xi$. Then $\xi_n \xrightarrow{d} \xi$ if and only if*

(i) $\mathbb{P}(\xi_n(U) = 0) \to \mathbb{P}(\xi(U) = 0)$, $U \in \mathcal{U}$.

(ii) $\limsup_{n\to\infty} \mathbb{P}(\xi_n(U) > 1) \leq \mathbb{P}(\xi(U) > 1)$, $U \in \mathcal{U}$.

2.3 The Chen-Stein method

The Stein method is a technique employed to investigate the accuracy of the approximation to one distribution by another in various metrics. C. Stein in [69] initially conceived it to study the approximation to the normal distribution for the sum of dependent random variables. L. H. Y. Chen modified the Stein method to obtain approximation results for the Poisson distribution; for this reason the Stein method applied to the problem of Poisson approximation is referred to as the Chen-Stein method. For a detailed and more general introduction into Stein's method, we refer to [40, 60], and for the Chen-Stein method for Poisson approximation to [10, 16, 60].

Let Lip(1) denote the set of all Lipschitz functions $g : \mathbb{N}_0 \to \mathbb{R}$ with Lipschitz constant bounded by 1, and let P_λ be a Poisson random variable with mean $\lambda \geq 0$. For any fixed $g \in \mathrm{Lip}(1)$, the solution of Stein's equation for the Poisson approximation is a function $f_g : \mathbb{N}_0 \to \mathbb{R}$ with $f_g(0) = 0$ that satisfies

$$\lambda f_g(i+1) - i f_g(i) = g(i) - \mathbb{E}[g(P_\lambda)], \quad i \in \mathbb{N}_0. \tag{2.2}$$

The function f_g can be obtained by solving recursively (2.2) for $i = 0, 1, \ldots$. An explicit expression for this solution is given in [5, Lemma 1]. In particular for $g = \mathbf{1}_A$ with $A \subset \mathbb{N}_0$, one has the following representation for f_g (see [60, Lemma 4.2]).

Lemma 2.3.1. *For any $\lambda > 0$ and $A \subset \mathbb{N}_0$ the unique solution f_A of*

$$\lambda f_A(i+1) - i f_A(i) = \mathbf{1}\{i \in A\} - \mathbb{P}(P_\lambda \in A), \quad i \in \mathbb{N}_0, \tag{2.3}$$

with $f_A(0) = 0$ is given by

$$f_A(i) = \frac{e^\lambda (i-1)!}{\lambda^i} \big[\mathbb{P}(P_\lambda \in A \cap \{0, 1, \ldots, i-1\}) - \mathbb{P}(P_\lambda \in A)\mathbb{P}(P_\lambda \leq i-1)\big], \quad i \in \mathbb{N}.$$

From now on, we denote by f_A the solution of the Stein equation (2.2) for $g = \mathbf{1}_A$ with $A \subset \mathbb{N}_0$. Let X be a random variable with values in $\mathbb{N}_0$. The idea of the Chen-Stein method for the Poisson approximation of X is to plug X in (2.2) and to take the expectation, which yields

$$\mathbb{E}[\lambda f_g(X+1) - X f_g(X)] = \mathbb{E}[g(X)] - \mathbb{E}[g(P_\lambda)].$$

So we can control the difference between the expectations of $g(X)$ and $g(P_\lambda)$ on the right-hand side by estimating the term on the left-hand side. This requires some bounds on the solution of (2.2). These bounds are also called Stein's factors or magic factors, where the latter name derives from the fact that they tend to decrease as the mean λ of P_λ increases. For a function $h : \mathbb{N}_0 \to \mathbb{R}$ we define $\Delta h : \mathbb{N}_0 \to \mathbb{R}$ by $\Delta h(i) = h(i+1) - h(i)$. The solution of the Stein equation (2.2) has the following bounds (see [11, Theorem 1.1]).

We use the shorthand notation $a \wedge b = \min\{a, b\}$ for $a, b \in \mathbb{R}$.

Lemma 2.3.2. *For any $\lambda > 0$ and $g \in \mathrm{Lip}(1)$, let f_g be the solution of (2.2). Then,*

$$\max_{i \in \mathbb{N}_0} |f_g(i)| \leq 1 \quad \text{and} \quad \max_{i \in \mathbb{N}} |\Delta f_g(i)| \leq 1 \wedge \frac{8}{3\sqrt{2e\lambda}} \leq 1 \wedge \frac{1.1437}{\sqrt{\lambda}}.$$

Since $f_g(0) = 0$, Lemma 2.3.2 implies for $\lambda > 0$ and $g \in \mathrm{Lip}(1)$ that

$$\max_{i \in \mathbb{N}_0} |f_g(i)| \leq 1 \quad \text{and} \quad \max_{i \in \mathbb{N}_0} |\Delta f_g(i)| \leq 1. \tag{2.4}$$

Moreover, the solution of (2.3) for $A \subset \mathbb{N}_0$ has the following magic factors (see [60, Lemma 4.4]).

Lemma 2.3.3. *For f_A as in Lemma 2.3.1,*

$$\max_{i \in \mathbb{N}_0} |f_A(i)| \leq 1 \wedge \frac{1}{\sqrt{\lambda}} \quad \text{and} \quad \max_{i \in \mathbb{N}_0} |\Delta f_A(i)| \leq 1 \wedge \frac{1}{\lambda}.$$

In Chapter 3, we are interested, in particular, in the solutions of the Stein equation f_A with $A = \{0, \ldots, v\}, v \in \mathbb{N}_0$. For these, in Section 3.2, we derive similar - potentially sharper magic factors than those mentioned above.

2.4 Size-bias coupling

The size-bias coupling first appeared in the context of Stein's method for Gaussian approximation in [28]. In this section, we only present the main properties of size-bias coupling focusing on its application to Poisson approximation; for a more detailed introduction, we refer to [3, 60].

Definition 2.4.1. *For a random variable $Y \geq 0$ with $\mu = \mathbb{E}[Y] > 0$, we say that the random variable Y^s has the size-bias distribution of Y if for all measurable functions $f : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}[|Yf(Y)|] < \infty$, we have*

$$\mathbb{E}[Yf(Y)] = \mu\mathbb{E}[f(Y^s)].$$

Moreover, we say that Y^s is a size-bias coupling of Y if it is defined on the same probability space as Y.

The size-bias distribution of a random variable $Y \geq 0$ with finite mean always exists (see [60, Proposition 3.18]). In particular, for a non-negative integer-valued random variable, one obtains the following result (see [60, Corollary 3.19]).

Lemma 2.4.2. *If $X \geq 0$ is an integer-valued random variable with $\lambda = \mathbb{E}[X] > 0$, then a random variable X^s with the size-bias distribution of X is such that*

$$\mathbb{P}(X^s = k) = \frac{k\mathbb{P}(X = k)}{\lambda}.$$

The first inequality in the next proposition is a classical result (see [60, Theorem 4.13]), whose proof is based on the Chen-Stein method and size-bias coupling, while the second inequality, which gives a Poisson approximation result in the Wasserstein distance, can be derived by combining the proof of [60, Theorem 4.13] with [11, Theorem 1.1]. Given two non-negative integer-valued random variables Y_1 and Y_2, the total variation distance between Y_1 and Y_2 is defined as

$$d_{TV}(Y_1, Y_2) = \sup_{A \subset \mathbb{N}_0} |\mathbb{P}(Y_1 \in A) - \mathbb{P}(Y_2 \in A)|,$$

and the Wasserstein distance between Y_1 and Y_2 is given by

$$d_W(Y_1, Y_2) = \sup_{g \in \text{Lip}(1)} |\mathbb{E}[g(Y_1)] - \mathbb{E}[g(Y_2)]|.$$

Proposition 2.4.3. *Let $X \geq 0$ be an integer-valued random variable with $\lambda = \mathbb{E}[X] > 0$, and let P_λ be a Poisson random variable with mean λ. If X^s is a size-bias coupling of X, then*

$$d_{TV}(X, P_\lambda) \leq (1 \wedge \lambda)\mathbb{E}[|X + 1 - X^s|]$$

and

$$d_W(X, P_\lambda) \leq (1.1437\sqrt{\lambda} \wedge \lambda)\mathbb{E}[|X + 1 - X^s|].$$

From the previous proposition it follows that the total variation distance between X and P_λ is small if the absolute value of the difference between $X + 1$ and X^s has small expectation. For sums of random variables, standard techniques to construct size-bias couplings are available. A general result in this direction is [60, Proposition 3.21]. The following lemma is a corollary of such proposition for the situation when the random variables take values in $\{0, 1\}$, (see [60, Corollary 3.24]).

Lemma 2.4.4. *Let $X_1, \ldots, X_n$ be zero-one random variables and also let $p_i = \mathbb{P}(X_i = 1)$. For each $i = 1, \ldots, n$, let $(X_j^{(i)})_{j \neq i}$ have the distribution of $(X_j)_{j \neq i}$ conditional on $X_i = 1$. If $X = \sum_{i=1}^n X_i$, $\lambda = \mathbb{E}[X] > 0$, and I is chosen independent of all else with $\mathbb{P}(I = i) = p_i/\lambda$, then $X^s = \sum_{j \neq I} X_j^{(I)} + 1$ has the size-bias distribution of X.*

Finally, we introduce the notion of size-bias coupling for random vectors. This definition was employed in [28] to study the multivariate Gaussian approximation. For a random vector $\mathbf{Y} = (Y_1, \ldots, Y_d)$ in $\mathbb{R}^d, d \in \mathbb{N}$, we write $\mathbb{E}[\mathbf{Y}] = (\mathbb{E}[Y_1], \ldots, \mathbb{E}[Y_d])$ for the mean of $\mathbf{Y}$.

Definition 2.4.5. *Let $\mathbf{Y} = (Y_1, \ldots, Y_d)$ be a random vector with values in $[0, \infty)^d, d \in \mathbb{N}$, with mean $\mathbb{E}[\mathbf{Y}] = (\mu_1, \ldots, \mu_d) \in (0, \infty)^d$. A family of random vectors $\mathbf{Y}^{(i)}, i = 1, \ldots, d$, with values in $\mathbb{R}^d$ is a size-bias coupling of $\mathbf{Y}$ if the random vectors $\mathbf{Y}^{(i)}$ are defined on the same probability space as $\mathbf{Y}$, and, for each i and all measurable functions $f : \mathbb{R}^d \to \mathbb{R}$ such that $\mathbb{E}[|Y_i f(\mathbf{Y})|] < \infty$, they satisfy*

$$\mathbb{E}[Y_i f(\mathbf{Y})] = \mu_i \mathbb{E}[f(\mathbf{Y}^{(i)})].$$

Chapter 3

Poisson approximation

This chapter is a slightly modified and adjusted version of the following preprint article jointly written with Matthias Schulte:

F. Pianoforte and M. Schulte. Poisson approximation with applications to stochastic geometry. arXiv:2104.02528, 2021.

Abstract. In this chapter, we compare the distributions of integer-valued random variables and Poisson random variables. We consider the total variation and the Wasserstein distance and provide, in particular, explicit bounds on the pointwise difference between the cumulative distribution functions. Special attention is dedicated to estimating the difference when the cumulative distribution functions are evaluated at 0. This permits to approximate the minimum (or maximum) of a collection of random variables by a suitable random variable in the Kolmogorov distance. The main theoretical results are obtained by combining the Chen-Stein method with size-bias coupling and a generalization of size-bias coupling for integer-valued random variables developed herein. A wide variety of applications are then discussed with a focus on stochastic geometry.

Acknowledgments. This research was supported by the Swiss National Science Foundation (grant number 200021_175584). We would like to thank Fraser Daly for some valuable comments.

3.1 Introduction and main results

Let X be a random variable taking values in $\mathbb{N}_0$ and let P_λ be a Poisson random variable with mean $\lambda > 0$. In this chapter, we employ Stein's method, size-bias coupling and a generalization of size-bias coupling for integer-valued random variables developed herein to compare the distributions of X and P_λ. We derive upper bounds on the total variation distance

$$d_{TV}(X, P_\lambda) = \sup_{A \subset \mathbb{N}_0} |\mathbb{P}(X \in A) - \mathbb{P}(P_\lambda \in A)|$$

and the Wasserstein distance

$$d_W(X, P_\lambda) = \sup_{g \in \mathrm{Lip}(1)} |\mathbb{E}[g(X)] - \mathbb{E}[g(P_\lambda)]|$$

between X and P_λ, where Lip(1) denotes the set of all Lipschitz functions $g : \mathbb{N}_0 \to \mathbb{R}$ with Lipschitz constant bounded by 1. In addition, we establish bounds on the pointwise differences

$$\big|\mathbb{P}(X \leq v) - \mathbb{P}(P_\lambda \leq v)\big|, \quad v \in \mathbb{N}_0,$$

between the cumulative distribution functions of X and P_λ, which are smaller than those for the total variation distance. Particular attention is paid to the case $v = 0$. This permits to approximate the minimum (or maximum) of a collection of random variables by a suitable random variable in the Kolmogorov distance. For example, let λ_d denote the Lebesgue measure on $\mathbb{R}^d$, let k_d stand for the volume of the d-dimensional unit ball, and let η_t be a Poisson process on $\mathbb{R}^d$ with intensity measure $t\lambda_d$, $t > 0$. From the aforementioned bounds for $v = 0$ we deduce that the random variable Y_t given by

$$Y_t = \min_{(x,y)\in\eta^2_{t,\neq} \,:\, \frac{x+y}{2}\in[0,1]^d} 2^{-1}t^2 k_d \|x-y\|^d,$$

which is the rescaled minimum (Euclidean) distance between pairs of points of η_t with midpoint in $[0,1]^d$, satisfies

$$0 \leq \mathbb{P}(Y_t > u) - \mathbb{P}(E_1 > u) \leq \frac{81}{t} \tag{3.1}$$

for $u \geq 0$ (see Theorem 3.3.4), where E_1 denotes an exponentially distributed random variable with mean 1. This is possible because $\mathbb{P}(Y_t > u)$ can be written as $\mathbb{P}(X_u = 0)$ with

$$X_u = \frac{1}{2} \sum_{(x,y)\in\eta^2_{t,\neq}} \mathbf{1}\Big\{\frac{x+y}{2} \in [0,1]^d, 2^{-1}t^2 k_d \|x-y\|^d \in [0,u]\Big\}$$

and $\mathbb{P}(E_1 > u) = \mathbb{P}(P_u = 0)$. By estimating $|\mathbb{P}(X_u = 0) - \mathbb{P}(P_u = 0)|$ uniformly for all $u \geq 0$, one obtains (3.1), which provides a bound on the Kolmogorov distance

$$d_K(Y_t, E_1) = \sup_{u\in\mathbb{R}} |\mathbb{P}(Y_t > u) - \mathbb{P}(E_1 > u)|$$

between Y_t and E_1.

Let us now give precise formulations of our main results. We use the shorthand notation $a \wedge b = \min\{a, b\}$ for $a, b \in \mathbb{R}$, and we indicate by W_+ and W_- the positive and negative part of a random variable W, respectively.

Theorem 3.1.1. *Let X be a random variable taking values in $\mathbb{N}_0$ and let P_λ be a Poisson random variable with mean $\lambda = \mathbb{E}[X] > 0$. Assume there exists a random variable Z defined on the same probability space as X with values in $\mathbb{Z}$ such that*

$$i\mathbb{P}(X = i) = \lambda\mathbb{P}(X + Z = i - 1), \quad i \in \mathbb{N}, \tag{3.2}$$

is satisfied. Then,

$$d_{TV}(X, P_\lambda) \leq (1 \wedge \lambda)\mathbb{E}[|Z|] \quad \text{and} \quad d_W(X, P_\lambda) \leq (1.1437\sqrt{\lambda} \wedge \lambda)\mathbb{E}[|Z|]. \tag{3.3}$$

Furthermore for all $m \in \mathbb{N}_0$,

$$|\mathbb{P}(X = 0) - \mathbb{P}(P_\lambda = 0)| \leq \frac{m!}{\lambda^m}\mathbb{E}[|Z|] + \sum_{k=0}^{m-1} \left(\frac{\lambda}{k+1} \wedge \frac{k!}{\lambda^k}\right) \mathbb{E}\big[|Z|\mathbf{1}\{X - Z_- = k\}\big] \tag{3.4}$$

and for all $v \in \mathbb{N}$,

$$|\mathbb{P}(X \leq v) - \mathbb{P}(P_\lambda \leq v)| \leq \frac{(v+1)^2}{\lambda}\mathbb{E}[|Z|] + \mathbb{E}\big[|Z|\mathbf{1}\{X - Z_- \leq v\}\big]. \tag{3.5}$$

Recall that for a random variable $Y \geq 0$ with $\mu = \mathbb{E}[Y] > 0$, a random variable Y^s on the same probability space as Y is a size-bias coupling of Y if it satisfies

$$\mathbb{E}[Yf(Y)] = \mu\mathbb{E}[f(Y^s)] \tag{3.6}$$

for all measurable f such that $\mathbb{E}[|Yf(Y)|] < \infty$. Thus, (3.2) implies that $X + Z + 1$ is a size-bias coupling of X so that we can replace Z by $X^s - X - 1$ with a size-bias coupling X^s of X in Theorem 3.1.1. In this form the bounds in (3.3) were already presented in Proposition 2.4.3.

Remark 3.1.2. *Let X be as in Theorem 3.1.1 and assume that* (3.2) *is satisfied.*

(i) The last expressions on the right-hand sides of (3.4) *and* (3.5) *can be further bounded using the inequalities*

$$\mathbb{E}[|Z|\mathbf{1}\{X - Z_- = k\}] \leq \mathbb{E}[Z_-] + \mathbb{E}[Z_+\mathbf{1}\{X = k\}], \quad k \in \mathbb{N}_0,$$
$$\mathbb{E}\big[|Z|\mathbf{1}\{X - Z_- \leq v\}\big] \leq \mathbb{E}[Z_-] + \mathbb{E}[Z_+\mathbf{1}\{X \leq v\}], \quad v \in \mathbb{N}.$$

(ii) From (3.6) *with $f(x) = x$ we obtain $\lambda\mathbb{E}[X^s] = \mathbb{E}[X^2]$ so that $Z = X^s - X - 1$ yields*

$$\mathbb{E}[Z] = \frac{1}{\lambda}\big\{\mathrm{Var}(X) - \lambda\big\}. \tag{3.7}$$

The next result constitutes our main achievement and generalizes Theorem 3.1.1. Instead of assuming that Z satisfies (3.2) exactly, we allow error terms on the right-hand side of (3.2).

Theorem 3.1.3. *Let X be an integrable random variable with values in $\mathbb{N}_0$ and let P_λ be a Poisson random variable with mean $\lambda > 0$. Let Z be a random variable defined on the same probability space as X with values in $\mathbb{Z}$, and let $q_i, i \in \mathbb{N}_0$, be the sequence given by*

$$q_{i-1} = i\mathbb{P}(X = i) - \lambda\mathbb{P}(X + Z = i - 1), \quad i \in \mathbb{N}. \tag{3.8}$$

Then,

$$d_{TV}(X, P_\lambda) \leq (1 \wedge \lambda)\mathbb{E}[|Z|] + \left(1 \wedge \frac{1}{\sqrt{\lambda}}\right)\sum_{i=0}^{\infty}|q_i| \tag{3.9}$$

and

$$d_W(X, P_\lambda) \leq \lambda\mathbb{E}[|Z|] + \sum_{i=0}^{\infty}|q_i|. \tag{3.10}$$

Moreover, if $\mathbb{P}(X + Z \geq 0) = 1$, then

$$d_W(X, P_\lambda) \leq (1.1437\sqrt{\lambda} \wedge \lambda)\mathbb{E}[|Z|] + \sum_{i=0}^{\infty}|q_i|, \tag{3.11}$$

for all $m \in \mathbb{N}_0$,

$$|\mathbb{P}(X = 0) - \mathbb{P}(P_\lambda = 0)| \leq \frac{m!}{\lambda^m}\mathbb{E}\big[|Z|\big] + \sum_{k=0}^{m-1}\left(\frac{\lambda}{k+1} \wedge \frac{k!}{\lambda^k}\right)\mathbb{E}\big[|Z|\mathbf{1}\{X - Z_- = k\}\big] + \left(1 \wedge \frac{1}{\lambda}\right)|q_0| + \left(1 \wedge \frac{1}{\lambda^2}\right)\sum_{i=1}^{\infty}|q_i| \tag{3.12}$$

and for all $v \in \mathbb{N}$,

$$|\mathbb{P}(X \leq v) - \mathbb{P}(P_\lambda \leq v)| \leq \frac{(v+1)^2}{\lambda}\mathbb{E}[|Z|] + \mathbb{E}\big[|Z|\mathbf{1}\{X - Z_- \leq v\}\big] + \left(1 \wedge \frac{1}{\sqrt{\lambda}}\right)\sum_{i=0}^{\infty}|q_i|. \tag{3.13}$$

Note that Theorem 3.1.1 is a special case of Theorem 3.1.3. Indeed, if $q_i = 0$ for all $i \in \mathbb{N}_0$, (3.8) becomes (3.2) and the bounds in Theorem 3.1.3 simplify to those in Theorem 3.1.1. In this situation $X + Z + 1$ is a size-bias coupling of X. Thus, we can think of $X + Z + 1$ with Z satisfying (3.8) as a generalization of size-bias coupling. In order to have good bounds in Theorem 3.1.3, the error terms q_i, $i \in \mathbb{N}_0$, should be small. The important advantage of Theorem 3.1.3 compared to Theorem 3.1.1 is that one only needs to construct an approximate size-bias coupling instead of an exact size-bias coupling.

For our work the so-called magic factors or Stein's factors play a crucial role. These are bounds on the solutions of the Stein equation, which lead to the factors involving λ in our results. Since different classes of test functions have different magic factors, the upper bounds for the differences between $\mathbb{P}(X \leq v)$ and $\mathbb{P}(P_\lambda \leq v)$ for $v \in \mathbb{N}_0$ in Theorems 3.1.1 and 3.1.3 are of a better order in λ than those for the total variation distance or the Wasserstein distance. This observation is essential for obtaining approximation results in the Kolmogorov distance as (3.1) since it allows to bound the right-hand sides of (3.4), (3.5), (3.12) and (3.13) uniformly in λ. For a different Poisson approximation result where one has a better order in λ for the difference of the probabilities at zero than for the total variation distance we refer the reader to [1, Theorem 1].

To demonstrate the versatility of our general main results we apply them to several examples. In particular, we deduce bounds as (3.1), where we compare minima or maxima of collections of dependent random variables with random variables having an exponential, Weibull or Gumbel distribution.

We study the Poisson approximation of U-statistics constructed from an underlying Poisson or binomial point process (see Subsections 3.3.1 and 3.3.2). As application of our main finding on U-statistics with Poisson input, Theorem 3.3.3, we consider the minimum inter-point distance problem discussed at the beginning of the introduction and establish the bound (3.1) for the exponential approximation in Kolmogorov distance.

Our next example is the Poisson approximation of the number of non-overlapping k-runs in a sequence of n i.i.d. Bernoulli random variables. By a k-run one means at least k successes in a row. Here, we use Theorem 3.1.1 to bound the difference between the probability that among n trials there are no more than v non-overlapping k-runs and $\mathbb{P}(P_\alpha \leq v)$ for a certain Poisson random variable P_α; this bound is remarkable because it does not depend on k, i.e., the number of required successes in a row.

For stationary Poisson-Voronoi tessellations we consider statistics related to circumscribed radii and inradii. We use the inequality (3.12) in Theorem 3.1.3 to compare a transform of the minimal circumscribed radius of the cells with the nucleus in an observation window with a Weibull random variable in Kolmogorov distance. For this example we use the full generality of Theorem 3.1.3 since we construct a coupling that satisfies (3.8), but is not a size-bias coupling. By applying the inequality (3.4) in Theorem 3.1.1 we approximate a transform of the maximal inradius of the cells with the nucleus in an observation window by a Gumbel random variable in the Kolmogorov distance.

A crucial contribution of this work to stochastic geometry is that we provide bounds with respect to the Kolmogorov distance for the distributional approximation of some minima and maxima. The limiting distributions of the minimal distance between the points of a Poisson process and of large inradii and small circumscribed radii of Poisson-Voronoi tessellations have been studied before in e.g. [15, 19, 65, 66]. Some of these works provide quantitative bounds for the difference of the distribution functions at a fixed $u \in \mathbb{R}$, which depend on u. Thanks to our general Poisson approximation results Theorem 3.1.1 and Theorem 3.1.3, we are able to derive uniform bounds for all $u \in \mathbb{R}$. An alternative approach to deducing such results via Poisson approximation is to apply directly Stein's method for the exponential, Weibull or Gumbel distribution; see e.g. [60] for more details on Stein's method for exponential approximation.

In [51], a general result for the Poisson approximation of statistics of Poisson processes is derived by combining the Chen-Stein method and a kind of size-bias coupling and applied to study some statistics of inhomogeneous random graphs such as isolated vertices. Requiring some (stochastic) ordering assumptions between a random variable and its size-bias coupling leads to Poisson approximation results. In a similar spirit to our work, these ordering conditions were relaxed in [21]. For some recent Poisson process convergence results related to stochastic geometry we refer the reader to [47] and Chapters 4 and 5.

Other noteworthy general results derived in this chapter are lower and upper bounds on the probability that X equals 0, which are given in Proposition 3.2.3 and Corollary 3.2.4. Informally, they say that $\mathbb{P}(X = 0)$ can be bounded from above or below by $e^{-\lambda}$ for some $\lambda > 0$ if the random variable Z and the sequence $q_i, i \in \mathbb{N}_0$, in Theorem 3.1.1 and Theorem 3.1.3 satisfy certain conditions on their signs; for Z as in Theorem 3.1.1, it is understood that $q_i = 0$ for all $i \in \mathbb{N}_0$. These results sometimes allow us to remove the absolute values from the left-hand sides of (3.4) and (3.12).

The proof of Theorem 3.1.3 is based on the Chen-Stein method and the coupling in (3.8). Using the solution of the Stein equation for the Poisson distribution, we derive in Proposition 3.2.2 a new expression for the difference $|\mathbb{E}[g(P_\lambda)] - \mathbb{E}[g(X)]|$ for any $g \in \mathrm{Lip}(1)$. Taking in Proposition 3.2.2, the supremum over all functions in Lip(1) (or all indicator functions) establishes a different way to represent the Wasserstein distance (or the total variation distance). Moreover, choosing $g = \mathbf{1}\{\cdot \le v\}$ with $v \in \mathbb{N}$ gives a new expression for $|\mathbb{P}(X \le v) - \mathbb{P}(P_\lambda \le v)|$. These identities are then manipulated and combined with the magic factors and the coupling in (3.8) to prove Theorem 3.1.3.

Before we present our applications in Section 3.3, we prove our main results in the next section.

3.2 Proofs of the results of Section 3.1

This section contains the proofs of Theorem 3.1.1 and Theorem 3.1.3. To this end, we employ the Stein equation for Poisson random variables. Recall that for any fixed $g \in \mathrm{Lip}(1)$ and $\lambda > 0$, the solution of the Stein equation is a function $f_g : \mathbb{N}_0 \to \mathbb{R}$ with $f_g(0) = 0$ that satisfies

$$\lambda f_g(i+1) - i f_g(i) = g(i) - \mathbb{E}[g(P_\lambda)], \quad i \in \mathbb{N}_0, \tag{3.14}$$

where P_λ denotes a Poisson random variable with mean $\lambda > 0$. We denote by f_A the solution of (3.14) for $g = \mathbf{1}_A$ with $A \subset \mathbb{N}_0$. That is, for any $\lambda > 0$ and $A \subset \mathbb{N}_0$, the function f_A is the unique solution of

$$\lambda f_A(i+1) - i f_A(i) = \mathbf{1}\{i \in A\} - \mathbb{P}(P_\lambda \in A), \quad i \in \mathbb{N}_0, \tag{3.15}$$

such that $f_A(0) = 0$. For any indicator or Lipschitz function g, bounds on the solution of the Stein equation f_g are given in Section 2.3. We now derive similar - potentially sharper

- magic factors for the special cases f_A with $A = \{0, \ldots, v\}, v \in \mathbb{N}_0$. Similar bounds for sets A that are singletons were deduced for the translated Poisson approximation in [56, Lemma 3.7].

Lemma 3.2.1. *Let $f_{\{0\}}$ be the unique solution of* (3.15) *for $A = \{0\}$. Then,*

$$|f_{\{0\}}(i)| \leq \begin{cases} 1 \wedge \frac{1}{\lambda}, & \text{if } i = 1, \\ 1 \wedge \frac{1}{\lambda^2}, & \text{if } i \geq 2, \end{cases} \tag{3.16}$$

and for all $i \in \mathbb{N}$,

$$\Delta f_{\{0\}}(i) \leq 0. \tag{3.17}$$

Furthermore for all $i, n \in \mathbb{N}$ with $i \geq n$,

$$|\Delta f_{\{0\}}(i)| \leq \frac{1}{n} \wedge \frac{(n-1)!}{\lambda^n}. \tag{3.18}$$

Let $f_{\{0,\ldots,v\}}$ be the unique solution of (3.15) *for $A = \{0, \ldots, v\}$ with $v \in \mathbb{N}$ and $v \leq \lambda$. Then for all $i \geq v + 2$,*

$$\Delta f_{\{0,\ldots,v\}}(i) \leq 1 \wedge \frac{(v+1)^2}{\lambda^2}. \tag{3.19}$$

Proof. Obviously, the upper bound 1 in (3.16) follows from Lemma 2.3.3. Lemma 2.3.1 yields for $i \in \mathbb{N}$ that

$$f_{\{0\}}(i) = \frac{(i-1)!}{\lambda^i}(1 - \mathbb{P}(P_\lambda \leq i-1)) = \frac{(i-1)!}{\lambda^i} \sum_{m=i}^{\infty} \frac{\lambda^m}{m!} e^{-\lambda} = \sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{(i+\ell)!}(i-1)! e^{-\lambda}. \tag{3.20}$$

This implies (3.16) for $i = 1, 2$, and yields for $i \geq 3$ that

$$f_{\{0\}}(i) = \sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{(i+\ell)!}(i-1)! e^{-\lambda} = \frac{1}{\lambda^2} \sum_{\ell=0}^{\infty} \frac{\lambda^{\ell+2}}{(\ell+2)!} \frac{(i-1)!(\ell+2)!}{(i+\ell)!} e^{-\lambda}.$$

Thus, the elementary inequalities

$$\frac{(i-1)!(\ell+2)!}{(i+\ell)!} = \frac{(i-1)!}{(\ell+3) \cdot \ldots \cdot (\ell+i)} \leq \frac{2(i-1)!}{i!} \leq 1$$

establish (3.16) for $i \geq 3$. From (3.20) we also obtain for $n \in \mathbb{N}$,

$$\begin{aligned} \Delta f_{\{0\}}(i) &= \sum_{\ell=0}^{\infty} \left(\frac{\lambda^\ell}{(i+1+\ell)!} i! - \frac{\lambda^\ell}{(i+\ell)!}(i-1)! \right) e^{-\lambda} \\ &= \sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{(i+1+\ell)!} \big(i! - (i+1+\ell)(i-1)! \big) e^{-\lambda} \\ &= -\sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{(i+1+\ell)!} (\ell+1)(i-1)! e^{-\lambda} \\ &= -\sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{(n+\ell)!} \frac{(\ell+1)(n+\ell)!(i-1)!}{(i+1+\ell)!} e^{-\lambda}, \end{aligned}$$

which proves (3.17). For $i, n \in \mathbb{N}$ with $i \geq n$ the elementary inequalities

$$\frac{(\ell+1)(n+\ell)!(i-1)!}{(i+1+\ell)!} \leq \frac{(n+\ell)!(i-1)!}{(i+\ell)!} \leq (n-1)!$$

lead to

$$\left|\Delta f_{\{0\}}(i)\right| \leq (n-1)! e^{-\lambda} \sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{(n+\ell)!}.$$

Now the observations that

$$\sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{(n+\ell)!} \leq \frac{e^\lambda}{\lambda^n} \quad \text{and} \quad \sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{(n+\ell)!} \leq \frac{1}{n!} \sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{\ell!} \frac{\ell! n!}{(n+\ell)!} \leq \frac{1}{n!} \sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{\ell!} \leq \frac{e^\lambda}{n!}$$

show (3.18). Finally assume $\lambda \geq v$. By Lemma 2.3.1, we obtain for $i \geq v+2$,

$$\Delta f_{\{0,\ldots,v\}}(i) = e^\lambda \mathbb{P}(P_\lambda \in \{0, \ldots, v\}) \Delta f_{\{0\}}(i).$$

Then (3.18) with $n = v + 2$ implies that

$$|\Delta f_{\{0,\ldots,v\}}(i)| \leq \frac{(v+1)!}{\lambda^{v+2}} \sum_{\ell=0}^{v} \frac{\lambda^\ell}{\ell!} = \frac{(v+1)!}{\lambda^2} \sum_{\ell=0}^{v} \frac{\lambda^{\ell-v}}{\ell!} \leq \frac{(v+1)^2}{\lambda^2},$$

where we used the inequality $\lambda^{\ell-v}/\ell! \leq 1/v!$ for $\ell = 0, \ldots, v$ and $\lambda \geq v$ in the last step. This and Lemma 2.3.3 establish (3.19). □

The next proposition compares the distributions of an integer-valued random variable and a Poisson distributed random variable.

Proposition 3.2.2. *Let X be an integrable random variable taking values in $\mathbb{N}_0$, let $\lambda > 0$, and define*

$$\mathcal{D}(i) = i\mathbb{P}(X = i) - \lambda \mathbb{P}(X = i-1), \quad i \in \mathbb{N}.$$

Then, for all $g \in \mathrm{Lip}(1)$,

$$\mathbb{E}[g(P_\lambda)] - \mathbb{E}[g(X)] = \sum_{i=1}^{\infty} f_g(i) \mathcal{D}(i),$$

where f_g is the solution of (3.14).

Proof. It follows from (3.14) and the definition of $\mathcal{D}(i)$, $i \in \mathbb{N}$, that

$$\begin{aligned}
\mathbb{E}[g(P_\lambda)] - \mathbb{E}[g(X)] &= \mathbb{E}[X f_g(X) - \lambda f_g(X+1)] = \sum_{i=0}^{\infty} \mathbb{P}(X = i)(i f_g(i) - \lambda f_g(i+1)) \\
&= \sum_{i=1}^{\infty} \mathbb{P}(X = i) i f_g(i) - \sum_{i=1}^{\infty} \mathbb{P}(X = i-1) \lambda f_g(i) = \sum_{i=1}^{\infty} f_g(i) \mathcal{D}(i),
\end{aligned}$$

which gives the desired result. □

From Proposition 3.2.2 we derive the identity

$$\mathbb{P}(P_\lambda \in A) - \mathbb{P}(X \in A) = \sum_{i=1}^{\infty} f_A(i) \mathcal{D}(i), \quad A \subset \mathbb{N}_0,$$

where f_A is the solution of (3.15). We are now in position to show Theorem 3.1.3.

Proof of Theorem 3.1.3. It follows from (3.8) that

$$\mathcal{D}(i) = i\mathbb{P}(X = i) - \lambda\mathbb{P}(X = i-1) = \lambda\mathbb{P}(X + Z = i-1) - \lambda\mathbb{P}(X = i-1) + q_{i-1}, \quad i \in \mathbb{N}.$$

Thus, Proposition 3.2.2 yields for $g \in \mathrm{Lip}(1)$ that

$$\begin{aligned} &\mathbb{E}[g(P_\lambda)] - \mathbb{E}[g(X)] \\ &= \lambda \sum_{i=1}^{\infty} f_g(i)\big(\mathbb{P}(X+Z = i-1) - \mathbb{P}(X = i-1)\big) + \sum_{i=1}^{\infty} f_g(i) q_{i-1} =: H_g + Q_g. \end{aligned} \tag{3.21}$$

With $f_g(0) = 0$ and the convention $f_g(i) = 0$ for $i < 0$, we obtain

$$\begin{aligned} H_g &= \lambda \sum_{j\in\mathbb{Z}\setminus\{0\}} \sum_{i=1}^{\infty} f_g(i)\big(\mathbb{P}(X = i-1-j, Z = j) - \mathbb{P}(X = i-1, Z = j)\big) \\ &= \lambda \sum_{j\in\mathbb{Z}\setminus\{0\}} \sum_{i\in\mathbb{Z}} f_g(i)\big(\mathbb{P}(X = i-1-j, Z = j) - \mathbb{P}(X = i-1, Z = j)\big) \\ &= \lambda \sum_{j\in\mathbb{Z}\setminus\{0\}} \sum_{i\in\mathbb{Z}} f_g(i+j)\mathbb{P}(X = i-1, Z = j) - f_g(i)\mathbb{P}(X = i-1, Z = j) \\ &= \lambda \sum_{j\in\mathbb{Z}\setminus\{0\}} \sum_{i\in\mathbb{Z}} \big(f_g(i+j) - f_g(i)\big)\mathbb{P}(X = i-1, Z = j) \\ &= \lambda \sum_{j\in\mathbb{Z}\setminus\{0\}} \sum_{i=1}^{\infty} \big(f_g(i+j) - f_g(i)\big)\mathbb{P}(X = i-1, Z = j), \end{aligned} \tag{3.22}$$

where we used that X takes only values in $\mathbb{N}_0$ in the last step. The triangle inequality implies that

$$\begin{aligned} |H_g| &\le \lambda \max_{i\in\mathbb{N}_0} |\Delta f_g(i)| \sum_{j\in\mathbb{Z}\setminus\{0\}} \sum_{i=1}^{\infty} |j|\mathbb{P}(X = i-1, Z = j) \\ &= \lambda \max_{i\in\mathbb{N}_0} |\Delta f_g(i)| \sum_{j\in\mathbb{Z}\setminus\{0\}} |j|\mathbb{P}(Z = j) = \lambda \max_{i\in\mathbb{N}_0} |\Delta f_g(i)|\mathbb{E}[|Z|]. \end{aligned}$$

Furthermore, we have

$$|Q_g| \le \max_{i\in\mathbb{N}} |f_g(i)| \sum_{i=0}^{\infty} |q_i|. \tag{3.23}$$

Then combining (2.4) in Section 2.3 and the bounds on $|H_g|$ and $|Q_g|$ establishes (3.10). Moreover, from Lemma 2.3.3 and the bounds on $|H_g|$ and $|Q_g|$ with $g = \mathbf{1}_A$ for $A \subset \mathbb{N}_0$, we obtain (3.9).

Next, we notice that

$$\begin{aligned} |H_g| \le\ & \lambda \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} |f_g(i+j) - f_g(i)|\mathbb{P}(X = i-1, Z = j) \\ &+ \lambda \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} |f_g(i-j) - f_g(i)|\mathbb{P}(X = i-1, Z = -j). \end{aligned} \tag{3.24}$$

The assumption $\mathbb{P}(X+Z\geq 0)=1$ implies that $\mathbb{P}(X=i-1, Z=-j)=0$ for all $i,j\in\mathbb{N}$ with $i\leq j$. Hence, we obtain

$$\begin{aligned}
&\lambda\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}|f_g(i-j)-f_g(i)|\mathbb{P}(X=i-1,Z=-j)\\
&=\lambda\sum_{j=1}^{\infty}\sum_{i=j+1}^{\infty}\big|f_g(i-j)-f_g(i)\big|\mathbb{P}(X=i-1,Z=-j)\\
&=\lambda\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}\big|f_g(i)-f_g(i+j)\big|\mathbb{P}(X=i+j-1,Z=-j).
\end{aligned}\tag{3.25}$$

From (3.24), (3.25) and the triangle inequality it follows that

$$\begin{aligned}
|H_g|&\leq\lambda\max_{i\in\mathbb{N}}|\Delta f_g(i)|\sum_{j=1}^{\infty}j\Big(\sum_{i=1}^{\infty}\mathbb{P}(X=i-1,Z=j)+\sum_{i=1}^{\infty}\mathbb{P}(X=i+j-1,Z=-j)\Big)\\
&\leq\lambda\max_{i\in\mathbb{N}}|\Delta f_g(i)|\sum_{j\in\mathbb{Z}\setminus\{0\}}|j|\mathbb{P}(Z=j)=\lambda\max_{i\in\mathbb{N}}|\Delta f_g(i)|\mathbb{E}[|Z|].
\end{aligned}$$

Together with (3.21) and (3.23), this implies that

$$|\mathbb{E}[g(P_\lambda)]-\mathbb{E}[g(X)]|\leq\lambda\max_{i\in\mathbb{N}}|\Delta f_g(i)|\mathbb{E}[|Z|]+\max_{i\in\mathbb{N}_0}|f_g(i)|\sum_{i=0}^{\infty}|q_i|.$$

Hence, Lemma 2.3.2 establishes (3.11).

Combining (3.21), (3.24) and (3.25) with $g=\mathbf{1}_A$ for $A\subset\mathbb{N}_0$ yields

$$\mathbb{P}(P_\lambda\in A)-\mathbb{P}(X\in A)=:H_A+Q_A\tag{3.26}$$

where $H_A=H_g$ and $Q_A=Q_g$ with $g=\mathbf{1}_A$, and

$$\begin{aligned}
|H_A|\leq\ &\lambda\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}|f_A(i+j)-f_A(i)|\mathbb{P}(X=i-1,Z=j)\\
&+\lambda\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}|f_A(i)-f_A(i+j)|\mathbb{P}(X=i+j-1,Z=-j)=:H_A^{(1)}+H_A^{(2)}.
\end{aligned}$$

For $A=\{0\}$, by (3.18) in Lemma 3.2.1 with $n=i$ for $i\leq m$ and $n=m+1$ for $i\geq m+1$, we have

$$\begin{aligned}
H_{\{0\}}^{(1)}&\leq\sum_{j=1}^{\infty}\sum_{i=1}^{m}\Big(\frac{\lambda}{i}\wedge\frac{(i-1)!}{\lambda^{i-1}}\Big)j\mathbb{P}(X=i-1,Z=j)+\sum_{j=1}^{\infty}\sum_{i=m+1}^{\infty}\frac{m!}{\lambda^m}j\mathbb{P}(X=i-1,Z=j)\\
&=\sum_{k=0}^{m-1}\Big(\frac{\lambda}{k+1}\wedge\frac{k!}{\lambda^k}\Big)\mathbb{E}[Z_+\mathbf{1}\{X=k\}]+\frac{m!}{\lambda^m}\mathbb{E}[Z_+\mathbf{1}\{X\geq m\}].
\end{aligned}$$

Again (3.18) in Lemma 3.2.1 with $n = i$ for $i \le m$ and $n = m+1$ for $i \ge m+1$ leads to

$$
\begin{aligned}
H_{\{0\}}^{(2)} &\le \sum_{j=1}^{\infty}\sum_{i=1}^{m}\left(\frac{\lambda}{i} \wedge \frac{(i-1)!}{\lambda^{i-1}}\right) j\mathbb{P}(X = i+j-1, Z = -j) \\
&\quad + \sum_{j=1}^{\infty}\sum_{i=m+1}^{\infty}\frac{m!}{\lambda^m} j\mathbb{P}(X = i+j-1, Z = -j) \\
&= \sum_{i=1}^{m}\left(\frac{\lambda}{i} \wedge \frac{(i-1)!}{\lambda^{i-1}}\right)\mathbb{E}[Z_-\mathbf{1}\{X+Z = i-1\}] + \frac{m!}{\lambda^m}\mathbb{E}[Z_-\mathbf{1}\{X+Z \ge m\}] \\
&= \sum_{k=0}^{m-1}\left(\frac{\lambda}{k+1} \wedge \frac{k!}{\lambda^k}\right)\mathbb{E}[Z_-\mathbf{1}\{X+Z = k\}] + \frac{m!}{\lambda^m}\mathbb{E}[Z_-\mathbf{1}\{X+Z \ge m\}].
\end{aligned}
$$

From (3.16) in Lemma 3.2.1 it follows that

$$
|Q_{\{0\}}| \le \left(1 \wedge \frac{1}{\lambda}\right)|q_0| + \left(1 \wedge \frac{1}{\lambda^2}\right)\sum_{i=1}^{\infty}|q_i|.
$$

Combining (3.26) and the bounds on $|Q_{\{0\}}|, H_{\{0\}}^{(1)}$ and $H_{\{0\}}^{(2)}$ completes the proof of (3.12).

For $\lambda < v$, (3.13) follows directly from (3.9). By Lemma 2.3.3 for $i \le v+1$ and (3.19) in Lemma 3.2.1 for $i \ge v+2$, we obtain

$$
\begin{aligned}
H_{\{0,\dots,v\}}^{(1)} &\le (1 \wedge \lambda)\sum_{j=1}^{\infty}\sum_{i=1}^{v+1} j\mathbb{P}(X = i-1, Z = j) + \sum_{j=1}^{\infty}\sum_{i=v+2}^{\infty}\frac{(v+1)^2}{\lambda} j\mathbb{P}(X = i-1, Z = j) \\
&= (1 \wedge \lambda)\mathbb{E}[Z_+\mathbf{1}\{X \le v\}] + \frac{(v+1)^2}{\lambda}\mathbb{E}[Z_+\mathbf{1}\{X \ge v+1\}]
\end{aligned}
$$

and

$$
\begin{aligned}
H_{\{0,\dots,v\}}^{(2)} &\le (1 \wedge \lambda)\sum_{j=1}^{\infty}\sum_{i=1}^{v+1} j\mathbb{P}(X = i+j-1, Z = -j) \\
&\quad + \sum_{j=1}^{\infty}\sum_{i=v+2}^{\infty}\frac{(v+1)^2}{\lambda} j\mathbb{P}(X = i+j-1, Z = -j) \\
&= (1 \wedge \lambda)\mathbb{E}[Z_-\mathbf{1}\{X+Z \le v\}] + \frac{(v+1)^2}{\lambda}\mathbb{E}[Z_-\mathbf{1}\{X+Z \ge v+1\}].
\end{aligned}
$$

Moreover, Lemma 2.3.3 yields

$$
|Q_{\{0,\dots,v\}}| \le \max_{i \in \mathbb{N}_0}|f_{\{0,\dots,v\}}(i)|\sum_{i=0}^{\infty}|q_i| \le \left(1 \wedge \frac{1}{\sqrt{\lambda}}\right)\sum_{i=0}^{\infty}|q_i|.
$$

Combining (3.26) with $A = \{0,\dots,v\}$ and the bounds on $|Q_{\{0,\dots,v\}}|, H_{\{0,\dots,v\}}^{(1)}$ and $H_{\{0,\dots,v\}}^{(2)}$ establishes (3.13). □

Next we derive Theorem 3.1.1 from Theorem 3.1.3.

Proof of Theorem 3.1.1. It follows from (3.2) that X and Z satisfy (3.8) with $\lambda = \mathbb{E}[X]$ and $q_i = 0$ for $i \in \mathbb{N}_0$ and that

$$
\lambda = \mathbb{E}[X] = \sum_{k=1}^{\infty} k\mathbb{P}(X = k) = \sum_{k=1}^{\infty}\lambda\mathbb{P}(X+Z = k-1) = \lambda\mathbb{P}(X+Z \ge 0),
$$

whence $\mathbb{P}(X+Z\geq 0)=1$. This allows us to apply Theorem 3.1.3 which proves (3.3), (3.4) and (3.5). $\square$

The next result provides some inequalities for the probability that a non-negative integer-valued random variable equals zero.

Proposition 3.2.3. *Let X be an integrable random variable with values in $\mathbb{N}_0$ and $\lambda>0$. Consider a random variable Z defined on the same probability space as X with values in $\mathbb{Z}$, and let $(q_i)_{i\in\mathbb{N}_0}$ be the sequence given by*

$$q_{i-1}=i\mathbb{P}(X=i)-\lambda\mathbb{P}(X+Z=i-1),\quad i\in\mathbb{N}.$$

a) If Z is non-negative and $q_i\leq 0$ for $i\in\mathbb{N}_0$,

$$\mathbb{P}(X=0)\geq e^{-\lambda}.$$

b) If Z is non-positive, $\mathbb{P}(X+Z\geq 0)=1$ and $q_i\geq 0$ for $i\in\mathbb{N}_0$,

$$\mathbb{P}(X=0)\leq e^{-\lambda}.$$

Proof. It follows from (3.21) and (3.22) for $f=\mathbf{1}_{\{0\}}$ as well as $\mathbb{P}(P_\lambda=0)=e^{-\lambda}$ that

$$\begin{aligned}&e^{-\lambda}-\mathbb{P}(X=0)\\&=\lambda\sum_{j\in\mathbb{Z}\setminus\{0\}}\sum_{i=1}^{\infty}\big(f_{\{0\}}(i+j)-f_{\{0\}}(i)\big)\mathbb{P}(X=i-1,Z=j)+\sum_{i=1}^{\infty}f_{\{0\}}(i)q_{i-1}.\end{aligned}$$

By the assumption that $Z\geq 0$ (resp. $Z\leq 0$ and $\mathbb{P}(X+Z\geq 0)=1$) the first sum on the right-hand side runs only over $j\geq 1$ (resp. $j\leq -1$ and the inner sum runs over all $i\in\mathbb{N}$ with $i+j\geq 1$). Together with

$$f_{\{0\}}(i+j)-f_{\{0\}}(i)\leq 0 \text{ for } i,j\geq 1 \quad\text{and}\quad f_{\{0\}}(i+j)-f_{\{0\}}(i)\geq 0 \text{ for } j\leq -1,\ i+j\geq 1,$$

which follows from (3.17) in Lemma 3.2.1, and the assumptions on $(q_i)_{i\in\mathbb{N}_0}$, this leads to the desired results. $\square$

Since (3.2) is a special case of (3.8) with $\mathbb{P}(X+Z\geq 0)=1$ (see the proof of Theorem 3.1.1), the following corollary is a direct consequence of Proposition 3.2.3.

Corollary 3.2.4. *Let X be a random variable taking values in $\mathbb{N}_0$ and let $\lambda=\mathbb{E}[X]>0$. Assume there exists a random variable Z such that* (3.2) *is satisfied.*

a) If Z is non-negative,

$$\mathbb{P}(X=0)\geq e^{-\lambda}.$$

b) If Z is non-positive,

$$\mathbb{P}(X=0)\leq e^{-\lambda}.$$

3.3 Applications

In this section, we discuss several applications of our general main results, Theorem 3.1.1 and Theorem 3.1.3. We consider problems from the following topics: U-statistics, k-runs, Voronoi tessellations. Throughout this section, by P_λ we always denote a Poisson random variable with mean $\lambda\geq 0$.

3.3.1 U-statistics of binomial point processes

Let $(\mathbb{X}, \mathcal{X})$ be a measurable space. We consider a binomial point process β_n on $\mathbb{X}$ of $n \in \mathbb{N}$ independent points in $\mathbb{X}$ that are distributed according to a probability measure K. Let $\ell \in \mathbb{N}$ and let $h : \mathbb{X}^\ell \to \{0, 1\}$ be a measurable symmetric function. In the following we study the U-statistic

$$S = \frac{1}{\ell!} \sum_{(x_1,...,x_\ell)\in\beta^\ell_{n,\neq}} h(x_1, ..., x_\ell),$$

where $\beta^\ell_{n,\neq}$ denotes the set of all ℓ-tuples of distinct points of β_n. We refer to the monographs [34, 39] for more details on U-statistics and their applications in statistics. A straightforward computation shows that

$$\lambda := \mathbb{E}[S] = \frac{(n)_\ell}{\ell!} \int_{\mathbb{X}^\ell} h(x_1, \ldots, x_\ell) dK^\ell(x_1, \ldots, x_\ell),$$

where $(n)_\ell$ stands for the ℓ-th descending factorial.

In this subsection, we establish bounds on the Poisson approximation of S in the total variation and Wasserstein distances. We also provide bounds on the pointwise difference between the cumulative distribution functions of S and P_λ. To this end, we define

$$r = \max_{1\leq i\leq \ell-1} (n)_{2\ell-i} \int_{\mathbb{X}^i} \left(\int_{\mathbb{X}^{\ell-i}} h(x_1, \ldots, x_\ell) dK^{\ell-i}(x_{i+1}, \ldots, x_\ell) \right)^2 dK^i(x_1, \ldots, x_i)$$

for $\ell \geq 2$, and put $r = 0$ for $\ell = 1$. Moreover for $n \geq 2\ell$, we define

$$\tilde{S} = \frac{1}{\ell!} \sum_{(x_1,...,x_\ell)\in\beta^\ell_{n-2\ell,\neq}} h(x_1, ..., x_\ell).$$

Theorem 3.3.1. *Let $n \geq 2\ell$ and let S, $\lambda > 0$, r and $\tilde{S}$ be as above. Then,*

$$d_{TV}(S, P_\lambda) \leq (1 \wedge \lambda)\left(\frac{2^\ell r}{\ell!\lambda} + \frac{2\ell^2\lambda}{n}\right) \quad \text{and} \quad d_W(S, P_\lambda) \leq (1.1437\sqrt{\lambda} \wedge \lambda)\left(\frac{2^\ell r}{\ell!\lambda} + \frac{2\ell^2\lambda}{n}\right). \tag{3.27}$$

Moreover, for all $m \in \mathbb{N}$,

$$\left|\mathbb{P}(S = 0) - e^{-\lambda}\right| \leq \left[\sum_{k=0}^{m-1} \left(\frac{\lambda}{k+1} \wedge \frac{k!}{\lambda^k}\right)\mathbb{P}(\tilde{S} \leq k) + \frac{m!}{\lambda^m}\right]\left(\frac{2^\ell r}{\ell!\lambda} + \frac{2\ell^2\lambda}{n}\right) \tag{3.28}$$

and for all $v \in \mathbb{N}$,

$$|\mathbb{P}(S \leq v) - \mathbb{P}(P_\lambda \leq v)| \leq \left[\frac{(v+1)^2}{\lambda} + \mathbb{P}(\tilde{S} \leq v)\right]\left(\frac{2^\ell r}{\ell!\lambda} + \frac{2\ell^2\lambda}{n}\right). \tag{3.29}$$

The bound on the Wasserstein distance in (3.27) slightly improves that in [22, Theorem 7.1] since it has a better order in λ. The bound for the total variation distance was also derived in [66, Proposition 2] by rewriting [8, Theorem 2]. By means of (3.28), one can study for some measurable symmetric function $g : \mathbb{X}^\ell \to \mathbb{R}$ the maximum (minumum) of $g(p)$ over all $p \in \beta^\ell_{n,\neq}$, which is called U-max-statistic (U-min-statistic). This is possible because for any $u \in \mathbb{R}$, the probability that $\max_{p\in\beta^\ell_{n,\neq}} g(p)$ is less than u can be written as the probability that $\sum_{p\in\beta^\ell_{n,\neq}} \mathbf{1}\{g(p) \geq u\}$ equals 0. Limit theorems for

U-max-statistics were considered in [36], yet without providing approximation results with respect to any distance; see also [42]. In contrast to these works, (3.28) may lead to approximation results in the Kolmogorov distance; see Theorem 3.3.4 in Subsection 3.3.3 and the discussion below it. To the best of our knowledge, the last two inequalities presented in Theorem 3.3.1 have no analogues in the literature.

From now on assume that $n \geq \ell$. Let χ be a point process of ℓ random points $X'_1, \ldots, X'_\ell$ in $\mathbb{X}$ that are independent of β_n and distributed such that

$$\mathbb{P}((X'_1, \ldots, X'_\ell) \in A) = \frac{(n)_\ell}{\ell! \lambda} \int_{\mathbb{X}^\ell} \mathbf{1}\{(x_1, \ldots, x_\ell) \in A\} h(x_1, \ldots, x_\ell) dK^\ell(x_1, \ldots, x_\ell)$$

for all A from the product σ-field generated by $\mathcal{X}^\ell$. Now we define

$$S' = -h(X'_1, \ldots, X'_\ell) + \frac{1}{\ell!} \sum_{(x_1, ..., x_\ell) \in (\beta_{n-\ell} \cup \chi)^\ell_{\neq}} h(x_1, ..., x_\ell).$$

Proposition 3.3.2. *For all $n \geq \ell$ and $k \in \mathbb{N}$,*

$$k\mathbb{P}(S = k) = \lambda \mathbb{P}(S' = k - 1).$$

Proof. We have that

$$k\mathbb{P}(S = k) = \mathbb{E}[k\mathbf{1}\{S = k\}] = \frac{1}{\ell!} \mathbb{E} \sum_{(x_1, \ldots, x_\ell) \in \beta^\ell_{n,\neq}} h(x_1, \ldots, x_\ell) \mathbf{1}\{S = k\}.$$

Using the fact that for any measurable map $g : \mathbb{X}^u \times \mathbf{N}_{\mathbb{X}} \to [0, \infty)$ with $u \in \mathbb{N}$,

$$\mathbb{E} \sum_{(x_1, \ldots, x_u) \in \beta^u_{n,\neq}} g(x_1, \ldots, x_u, \beta_n) = (n)_u \int_{\mathbb{X}^u} \mathbb{E}[g(x_1, \ldots, x_u, \beta_{n-u} + \sum_{i=1}^{u} \delta_{x_i})] dK^u(x_1, \ldots, x_u),$$

we obtain

$$\begin{aligned} k\mathbb{P}(S = k) &= \frac{(n)_\ell}{\ell!} \int_{\mathbb{X}^\ell} h(x_1, \ldots, x_\ell) \mathbb{P}\Big(\frac{1}{\ell!} \sum_{(y_1, \ldots, y_\ell) \in (\beta_{n-\ell} \cup \{x_1, \ldots, x_\ell\})^\ell_{\neq}} h(y_1, \ldots, y_\ell) = k \Big) \\ &\qquad \times dK^\ell(x_1, \ldots, x_\ell) \\ &= \lambda \mathbb{P}(S' + h(X'_1, \ldots, X'_\ell) = k) = \lambda \mathbb{P}(S' = k - 1), \end{aligned}$$

where we used $h(X'_1, \ldots, X'_\ell) = 1$ in the last step. This concludes the proof. □

Proof of Theorem 3.3.1. Suppose $n \geq 2\ell$. Our goal is to apply Theorem 3.1.1 with $Z = S' - S$, which satisfies the assumption (3.2) by Proposition 3.3.2. We define $s : \mathbf{N}_{\mathbb{X}} \to \mathbb{R}$ by

$$s(\nu) = \frac{1}{\ell!} \sum_{(x_1, \ldots, x_\ell) \in \nu^\ell_{\neq}} h(x_1, \ldots, x_\ell)$$

so that $S = s(\beta_n)$ and $S' = s(\beta_{n-\ell} + \chi) - h(X'_1, \ldots, X'_\ell)$. By the monotonicity of s, we have

$$\begin{aligned} |Z| = |S' - S| &= |s(\beta_{n-\ell} + \chi) - h(X'_1, \ldots, X'_\ell) - s(\beta_{n-\ell}) - (s(\beta_n) - s(\beta_{n-\ell}))| \\ &\leq (s(\beta_{n-\ell} + \chi) - h(X'_1, \ldots, X'_\ell) - s(\beta_{n-\ell})) + s(\beta_n) - s(\beta_{n-\ell}). \end{aligned} \tag{3.30}$$

Together with

$$\begin{aligned} &s(\beta_{n-\ell}+\chi) - h(X_1',\dots,X_\ell') - s(\beta_{n-\ell}) + s(\beta_n) - s(\beta_{n-\ell}) \\ &= s(\beta_{n-\ell}+\chi) - h(X_1',\dots,X_\ell') - s(\beta_n) + 2(s(\beta_n) - s(\beta_{n-\ell})) = Z + 2(s(\beta_n) - s(\beta_{n-\ell})) \end{aligned} \tag{3.31}$$

this implies

$$\mathbb{E}[|Z|] \leq \mathbb{E}[Z] + 2\mathbb{E}[s(\beta_n) - s(\beta_{n-\ell})].$$

From (3.7) in Remark 3.1.2 we know that

$$\mathbb{E}[Z] = \frac{1}{\lambda}(\mathrm{Var}(S) - \lambda) = \frac{1}{\lambda}(\mathbb{E}[S^2] - \lambda^2 - \lambda).$$

Thus, it follows from [22, Lemma 6.1] and the definition of r that

$$\mathbb{E}[Z] \leq \frac{2^\ell r}{\ell!\lambda}.$$

A straightforward computation shows that

$$\mathbb{E}[s(\beta_n) - s(\beta_{n-\ell})] = \left(1 - \frac{(n-\ell)_\ell}{(n)_\ell}\right)\lambda = \frac{(n)_\ell - (n-\ell)_\ell}{(n)_\ell}\lambda \leq \frac{\ell^2(n-1)_{\ell-1}}{(n)_\ell}\lambda = \frac{\ell^2\lambda}{n}.$$

Combining the previous estimates yields

$$\mathbb{E}[|Z|] \leq \frac{2^\ell r}{\ell!\lambda} + \frac{2\ell^2\lambda}{n}$$

so that (3.27) follows from (3.3).

Let $k \in \mathbb{N}$ be fixed. Note that $S \geq s(\beta_{n-\ell})$ and $S' \geq s(\beta_{n-\ell})$. If $Z \geq 0$, this implies

$$\mathbf{1}\{S - Z_- \leq k\} = \mathbf{1}\{S \leq k\} \leq \mathbf{1}\{s(\beta_{n-\ell}) \leq k\}.$$

For $Z \leq 0$ we obtain

$$\mathbf{1}\{S - Z_- \leq k\} = \mathbf{1}\{S + Z \leq k\} = \mathbf{1}\{S' \leq k\} \leq \mathbf{1}\{s(\beta_{n-\ell}) \leq k\}.$$

Combing the two cases leads to

$$\mathbf{1}\{S - Z_- = k\} \leq \mathbf{1}\{S - Z_- \leq k\} \leq \mathbf{1}\{s(\beta_{n-\ell}) \leq k\}.$$

Together with (3.30) we obtain

$$\begin{aligned} &\mathbb{E}[|Z|\mathbf{1}\{S - Z_- = k\}] \\ &\leq \mathbb{E}[\mathbf{1}\{s(\beta_{n-\ell}) \leq k\}(s(\beta_{n-\ell}+\chi) - h(X_1',\dots,X_\ell') - s(\beta_{n-\ell}) + s(\beta_n) - s(\beta_{n-\ell}))]. \end{aligned} \tag{3.32}$$

For $u \in \{1,\dots,\ell-1\}$ and $g : \mathbb{X}^u \to [0,\infty)$, we have

$$\begin{aligned} &\mathbb{E}[\mathbf{1}\{s(\beta_{n-\ell}) \leq k\} \sum_{(x_1,\dots,x_u)\in\beta^u_{n-\ell,\neq}} g(x_1,\dots,x_u)] \\ &= (n-\ell)_u \int_{\mathbb{X}^u} \mathbb{P}(s(\beta_{n-\ell-u} + \sum_{i=1}^u \delta_{x_i}) \leq k) g(x_1,\dots,x_u) dK^u(x_1,\dots,x_u) \\ &\leq \mathbb{P}(s(\beta_{n-2\ell}) \leq k)(n-\ell)_u \int_{\mathbb{X}^u} g(x_1,\dots,x_u) dK^u(x_1,\dots,x_u) \\ &= \mathbb{P}(s(\beta_{n-2\ell}) \leq k) \mathbb{E} \sum_{(x_1,\dots,x_u)\in\beta^u_{n-\ell,\neq}} g(x_1,\dots,x_u), \end{aligned} \tag{3.33}$$

where the inequality follows from the monotonicity of s. Because of

$$s(\beta_{n-\ell}+\chi)-h(X_1',\dots,X_\ell')-s(\beta_{n-\ell})=\sum_{u=1}^{\ell-1}\sum_{(x_1,\dots,x_u)\in\beta^u_{n-\ell,\neq}}\tilde{h}_u(x_1,\dots,x_u;\chi)$$

and

$$s(\beta_n)-s(\beta_{n-\ell})=\sum_{u=1}^{\ell-1}\sum_{(x_1,\dots,x_u)\in\beta^u_{n-\ell,\neq}}\overline{h}_u(x_1,\dots,x_u;\beta_n\setminus\beta_{n-\ell})$$

with suitable functions $\tilde{h}_u$ and $\overline{h}_u$, $u\in\{1,\dots,\ell-1\}$, we can rewrite the second factor on the right-hand side of (3.32) as sum of U-statistics with respect to $\beta_{n-\ell}$. Now an application of (3.33) and (3.31) yield

$$\begin{aligned}&\mathbb{E}[|Z|\mathbf{1}\{S-Z_-=k\}]\\&\le\mathbb{P}(s(\beta_{n-2\ell})\le k)\mathbb{E}[s(\beta_{n-\ell}+\chi)-h(X_1',\dots,X_\ell')-s(\beta_{n-\ell})+s(\beta_n)-s(\beta_{n-\ell})]\\&=\mathbb{P}(s(\beta_{n-2\ell})\le k)\big(\mathbb{E}[Z]+2\mathbb{E}[s(\beta_n)-s(\beta_{n-\ell})]\big).\end{aligned}$$

Bounding the second factor on the right-hand side as above leads to

$$\mathbb{E}[|Z|\mathbf{1}\{S-Z_-=k\}]\le\mathbb{P}(s(\beta_{n-2\ell})\le k)\left(\frac{2^\ell r}{\ell!\lambda}+\frac{2\ell^2\lambda}{n}\right).$$

Thus, (3.28) and (3.29) are immediate consequences of (3.4) and (3.5). □

3.3.2 *U*-statistics of Poisson processes

In this subsection, we study the Poisson approximation of U-statistics, where one sums over all ℓ-tuples of distinct points of a Poisson process instead of those of a binomial point process as in the previous subsection. In this case, the summation can run over infinitely many ℓ-tuples. As the results for U-statistics with binomial input in Subsection 3.3.1, the theory developed herein permits to study extreme value problems arising in stochastic geometry. For example, in the next subsection, we employ our main result for U-statistics with Poisson input to investigate the limiting behavior of the minimum inter-point distance between the points of a Poisson process in $\mathbb{R}^d$.

Let $(\mathbb{X},\mathcal{X})$ be a measurable space and let η be a Poisson process with a σ-finite intensity measure L on $\mathbb{X}$. For a fixed $\ell\in\mathbb{N}$ and a symmetric measurable function $h:\mathbb{X}^\ell\to\{0,1\}$ that is integrable with respect to L^ℓ we consider the U-statistic

$$S=\frac{1}{\ell!}\sum_{(x_1,\dots,x_\ell)\in\eta^\ell_{\neq}}h(x_1,\dots,x_\ell),$$

where $\eta^\ell_{\neq}$ denotes the set of all ℓ-tuples of distinct points of η. It follows from the multivariate Mecke formula that

$$\lambda:=\mathbb{E}[S]=\frac{1}{\ell!}\int_{\mathbb{X}^\ell}h(x_1,\dots,x_\ell)dL^\ell(x_1,\dots,x_\ell).$$

We define

$$r=\max_{1\le i\le\ell-1}\int_{\mathbb{X}^i}\left(\int_{\mathbb{X}^{\ell-i}}h(x_1,\dots,x_\ell)dL^{\ell-i}(x_{i+1},\dots,x_\ell)\right)^2dL^i(x_1,\dots,x_i)$$

for $\ell\ge2$, and put $r=0$ for $\ell=1$. The expression r is used to quantify the accuracy of the Poisson approximation for S and it is the analogue of r given in Subsection 3.3.1 for binomial U-statistics.

Theorem 3.3.3. *Let S, $\lambda > 0$ and r be as above. Then,*

$$d_{TV}(S, P_\lambda) \leq \left(1 \wedge \frac{1}{\lambda}\right)\frac{2^\ell r}{\ell!} \quad \text{and} \quad d_W(S, P_\lambda) \leq \left(1 \wedge \frac{1.1437}{\sqrt{\lambda}}\right)\frac{2^\ell r}{\ell!}. \tag{3.34}$$

Moreover, for all $m \in \mathbb{N}$,

$$0 \leq \mathbb{P}(S = 0) - e^{-\lambda} \leq \left[\sum_{k=0}^{m-1}\left(\frac{1}{k+1} \wedge \frac{k!}{\lambda^{k+1}}\right)\mathbb{P}\big(S \leq k\big) + \frac{m!}{\lambda^{m+1}}\right]\frac{2^\ell r}{\ell!} \tag{3.35}$$

and for all $v \in \mathbb{N}$,

$$|\mathbb{P}(S \leq v) - \mathbb{P}(P_\lambda \leq v)| \leq \left[\frac{(v+1)^2}{\lambda} + \mathbb{P}(S \leq v)\right]\frac{2^\ell r}{\ell!\lambda}. \tag{3.36}$$

The result for the total variation distance in (3.34) was shown in [66, Proposition 1], which improved [65, Proposition 4.1], and in [51, Section 8]. The bound for the Wasserstein distance in (3.34) was also derived in [51, Section 8] and has a slightly better order in λ than that in [22, Theorem 7.1]. To the best of our knowledge, the other inequalities presented in Theorem 3.3.3 have no analogues in the literature.

Proof of Theorem 3.3.3. We follow a similar approach as in the proof of Theorem 3.3.1. For $\ell = 1$, Theorem 3.3.3 is a direct consequence of [38, Theorem 5.1], whence we assume $\ell \geq 2$ from now on.

Let χ be a point process of ℓ random points $X_1', \dots, X_\ell'$ that are independent of η and distributed according to

$$\mathbb{P}((X_1', \dots, X_\ell') \in A) = \frac{1}{\ell!\lambda}\int_{\mathbb{X}^\ell} \mathbf{1}\{(x_1, \dots, x_\ell) \in A\}h(x_1, \dots, x_\ell)dL^\ell(x_1, \dots, x_\ell)$$

for A from the product σ-field generated by $\mathcal{X}^\ell$. We define

$$S' = -h(X_1', \dots, X_\ell') + \frac{1}{\ell!}\sum_{(x_1,\dots,x_\ell)\in(\eta\cup\chi)^k_{\neq}} h(x_1, \dots, x_\ell).$$

For $k \in \mathbb{N}$ the multivariate Mecke formula implies that

$$\begin{aligned}
&k\mathbb{P}(S = k) = \mathbb{E}[S\mathbf{1}\{S = k\}] \\
&= \frac{1}{\ell!}\mathbb{E}\sum_{(x_1,\dots,x_\ell)\in\eta^\ell_{\neq}} h(x_1, \dots, x_\ell)\mathbf{1}\left\{\frac{1}{\ell!}\sum_{(y_1,\dots,y_\ell)\in\eta^\ell_{\neq}} h(y_1, \dots, y_\ell) = k\right\} \\
&= \frac{1}{\ell!}\int_{\mathbb{X}^\ell} h(x_1, \dots, x_\ell)\mathbb{P}\left(\frac{1}{\ell!}\sum_{(y_1,\dots,y_\ell)\in(\eta\cup\{x_1,\dots,x_\ell\})^\ell_{\neq}} h(y_1, \dots, y_\ell) = k\right)dL^\ell(x_1, \dots, x_\ell) \\
&= \lambda\mathbb{P}(S' + h(X_1', \dots, X_\ell') = k) = \lambda\mathbb{P}(S' = k-1),
\end{aligned}$$

where we used $h(X_1', \dots, X_\ell') = 1$ in the last step. Thus, we see that S satisfies the hypothesis of Theorem 3.1.1 with $Z = S' - S \geq 0$.

Next we compute the expressions on the right-hand sides of the bounds in Theorem 3.1.1. Let $k \in \mathbb{N}$ be fixed. Define $s(\nu) = \frac{1}{\ell!}\sum_{(x_1,\dots,x_\ell)\in\nu^\ell_{\neq}} h(x_1, \dots, x_\ell)$ for $\nu \in \mathbf{N}_\mathbb{X}$ and note that $S = s(\eta)$. Since

$$s(\nu + \chi + \delta_x) - s(\nu + \delta_x) \geq s(\nu + \chi) - s(\nu) \quad \text{and} \quad \mathbf{1}\{s(\nu + \delta_x) \leq k\} \leq \mathbf{1}\{s(\nu) \leq k\}$$

for all $\nu \in \mathbf{N}_{\mathbb{X}}$ and $x \in \mathbb{X}$, by [38, Theorem 20.4] we obtain

$$\mathbb{E}\big[Z\mathbf{1}\{S \leq k\}\big] \leq \mathbb{E}[Z]\mathbb{P}(S \leq k).$$

Together with $Z \geq 0$, we have

$$\mathbb{E}[|Z|\mathbf{1}\{S - Z_- = k\}] = \mathbb{E}[Z\mathbf{1}\{S = k\}] \leq \mathbb{E}[Z\mathbf{1}\{S \leq k\}] \leq \mathbb{E}[Z]\mathbb{P}(S \leq k). \tag{3.37}$$

Furthermore, from Remark 3.1.2-(ii) it follows that

$$\mathbb{E}[Z] = \frac{1}{\lambda}\big\{\mathrm{Var}(S) - \lambda\big\} = \frac{1}{\lambda}\big\{\mathbb{E}[S^2] - \lambda^2 - \lambda\big\}. \tag{3.38}$$

Then, from $Z \geq 0$ and [22, Lemma 6.1] we deduce

$$\mathbb{E}[|Z|] = \mathbb{E}[Z] \leq \sum_{i=1}^{\ell-1} \frac{1}{\ell!\lambda}\binom{\ell}{i} r \leq \frac{2^\ell}{\ell!\lambda} r. \tag{3.39}$$

Finally, combining this bound with (3.3) shows (3.34), while (3.4) and (3.5) together with (3.37) and (3.39) lead to (3.35), where the first inequality is a consequence of Corollary 3.2.4 a), and (3.36). □

3.3.3 The distances between the points of a Poisson process

We consider random points in $\mathbb{R}^d$ distributed according to a Poisson process. For any pair of these points with the midpoint in a bounded measurable set $W \subset \mathbb{R}^d$, we take a transform of the Euclidean distance, and we study the Poisson approximation for the number of times that these quantities belong to a certain range of values. More importantly, we consider the exponential approximation for a transform of the minimal distance between pairs of points with midpoint in W.

Let η_t be a Poisson process on $\mathbb{R}^d$ with intensity measure $t\lambda_d$, $t > 0$, where we denote by λ_d the d-dimensional Lebesgue measure. For convenience, we assume $\lambda_d(W) = 1$; nonetheless, the following arguments are valid for every W with a positive and finite volume. Define

$$\xi_t = \frac{1}{2} \sum_{(x,y)\in\eta^2_{t,\neq}} \mathbf{1}\Big\{\frac{x+y}{2} \in W\Big\}\delta_{2^{-1}t^2 k_d\|x-y\|^d}, \qquad t > 0,$$

$$Y_t = \min_{(x,y)\in\eta^2_{t,\neq}:\frac{x+y}{2}\in W} 2^{-1}t^2 k_d\|x-y\|^d, \quad t > 0,$$

where $\|\cdot\|$ denotes the Euclidean norm.

Theorem 3.3.4. *Let ξ_t and Y_t be as above for $t > 0$. Let γ be a Poisson process on $[0,\infty)$ with the restriction of the Lebesgue measure to $[0,\infty)$ as intensity measure. Then for all $u \geq 0$ and all measurable $B \subset [0,u]$,*

$$d_{TV}(\xi_t(B), \gamma(B)) \leq (1 \wedge u)\frac{8u}{t} \tag{3.40}$$

and

$$0 \leq \mathbb{P}(Y_t > u) - e^{-u} \leq \frac{81}{t}. \tag{3.41}$$

The minimal distance between the points of a Poisson process was also considered in [15, 22, 65, 66], sometimes formulated as minimal edge length of the random geometric graph or the minimal inradius of a Poisson-Voronoi tessellation. The important achievement of Theorem 3.3.4 is that a rate of convergence for the Kolmogorov distance is provided in (3.41). So far it was only possible to prove bounds on the difference between $\mathbb{P}(Y_t > u)$ and e^{-u} that depend on $u > 0$ (see e.g. [65, Theorem 2.4] or [66, Corollary 3]).

In the works mentioned above all pairs of points are considered such that one or both points belong to W. Our approach, where we only require that the midpoint of the points is in W, can be extended to this different way of counting, but one might get additional terms in the bounds since $\mathbb{E}[\xi_t([0,u])]$ is not necessarily u due to boundary effects.

In [65, 66], beside Poisson approximation results for the number of inter-point distances below a given threshold it was shown that the point process of rescaled inter-point distances converges weakly to a Poisson process. By (3.40) and Proposition 2.2.10, we can also deduce that ξ_t converges weakly to γ as $t \to \infty$.

The related problem of small distances between the points of a binomial point process was first studied in [67]. Because of the similarity to Theorem 3.3.3, we believe that by applying Theorem 3.3.1 it is possible to prove a similar result to Theorem 3.3.4 for an underlying binomial point process.

By using in the proof of Theorem 3.3.4 the corresponding bound of Theorem 3.3.3 for the Wasserstein distance, one can obtain the counterpart of (3.40) for the Wasserstein distance with a different power in u and the same rate of convergence in t.

Proof of Theorem 3.3.4. First, we show that the intensity measure of the point process ξ_t is the restriction of the Lebesgue measure to $[0,\infty)$. Let $v_t = \big(\frac{2u}{k_d t^2}\big)^{1/d}$. The change of variable $z = \frac{x+y}{2}$ yields

$$\begin{aligned}\mathbb{E}[\xi_t([0,u])] &= \frac{t^2}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}\Big\{\frac{x+y}{2} \in W\Big\}\mathbf{1}\{\|x-y\| \le v_t\}\,dydx \\ &= 2^{d-1}t^2 \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}\{z \in W\}\mathbf{1}\{2\|x-z\| \le v_t\}\,dzdx \\ &= 2^{d-1}t^2 \int_{W}\int_{\mathbb{R}^d} \mathbf{1}\{2\|x-z\| \le v_t\}\,dxdz = u.\end{aligned}$$

For $B \subset [0,u]$ with $u > 0$ define

$$r_t(B) = t\int_{\mathbb{R}^d}\bigg(t\int_{\mathbb{R}^d}\mathbf{1}\Big\{\frac{x+y}{2} \in W\Big\}\mathbf{1}\{2^{-1}t^2k_d\|x-y\|^d \in B\}\,dy\bigg)^2 dx.$$

Again from the change of variable $z = \frac{x+y}{2}$, it follows that

$$\begin{aligned}r_t(B) \le r_t([0,u]) &= 2^{2d}t^3\int_{\mathbb{R}^d}\bigg(\int_W \mathbf{1}\{2\|x-z\| \le v_t\}\,dz\bigg)^2 dx \\ &\le 2^{2d}t^3\int_{\mathbb{R}^d}\bigg(\int_{\mathbb{R}^d}\mathbf{1}\{2\|x-z\| \le v_t\}\,dz\int_W \mathbf{1}\{2\|x-\tilde{z}\| \le v_t\}\,d\tilde{z}\bigg)dx \\ &= 2^{d+1}ut\int_{\mathbb{R}^d}\int_W \mathbf{1}\{2\|x-\tilde{z}\| \le v_t\}\,d\tilde{z}dx \\ &= 2^{d+1}ut\int_W\int_{\mathbb{R}^d} \mathbf{1}\{2\|x-\tilde{z}\| \le v_t\}\,dxd\tilde{z} = \frac{4u^2}{t}.\end{aligned}$$

Therefore (3.34) in Theorem 3.3.3 with $h(x,y) = \mathbf{1}\{\frac{x+y}{2} \in W\}\mathbf{1}\{2^{-1}t^2 k_d \|x-y\|^d \in B\}$ yields for $B \subset [0,u]$ that

$$d_{TV}(\xi_t(B), \gamma(B)) \leq \left(1 \wedge \frac{1}{u}\right) 2r_t(B) \leq (1 \wedge u)\frac{8u}{t}.$$

From (3.38) and (3.39) in the proof of Theorem 3.3.3 with $S = \xi_t([0,u])$, $r = r_t([0,u])$ and h as above, we know that

$$\mathrm{Var}(\xi_t([0,u])) \leq \mathbb{E}[\xi_t([0,u])] + 2r_t([0,u]) = u + \frac{8u^2}{t}.$$

Thus it follows from the Chebyshev inequality that

$$\mathbb{P}(\xi_t([0,u]) = 0) \leq \mathbb{P}(|\xi_t([0,u]) - u| \geq u) \leq \frac{\mathrm{Var}(\xi_t([0,u]))}{u^2} = \frac{1}{u} + \frac{8}{t}.$$

Together with (3.35) in Theorem 3.3.3 with $m = 1$ and straightforward arguments, this leads to

$$\begin{aligned} 0 \leq \mathbb{P}(\xi_t([0,u]) = 0) - e^{-u} = \mathbb{P}(Y_t > u) - e^{-u} &\leq \left[\frac{1}{u}\mathbb{P}(\xi_t([0,u]) = 0) + \frac{1}{u^2}\right]\frac{8u^2}{t} \qquad (3.42) \\ &\leq \left(\frac{1}{u^2} + \frac{8}{ut} + \frac{1}{u^2}\right)\frac{8u^2}{t} = \frac{16}{t} + \frac{64u}{t^2} \end{aligned}$$

so that

$$\sup_{u \in [0,t]} |\mathbb{P}(Y_t > u) - e^{-u}| \leq \frac{80}{t}.$$

Thus, we have

$$\mathbb{P}(Y_t > t) \leq \frac{80}{t} + e^{-t} \leq \frac{81}{t}$$

and

$$\sup_{u \in [0,\infty)} |\mathbb{P}(Y_t > u) - e^{-u}| \leq \max\Big\{\sup_{u \in [0,t]} |\mathbb{P}(Y_t > u) - e^{-u}|, \mathbb{P}(Y_t > t), e^{-t}\Big\} \leq \frac{81}{t},$$

which combined with the left-hand side of (3.42) completes the proof. □

3.3.4 k-runs in a sequence of i.i.d. Bernoulli random variables

Consider n independent and identically distributed Bernoulli random variables. A k-head run is defined as an uninterrupted sequence of k successes, where k is a positive integer. For example, for $k = 1$, one simply studies the successes, while for $k = 2$, one considers the occurrence of two consecutive successes in a row. Several authors have investigated the number of k-head runs in a sequence of Bernoulli random variables; see e.g. the book [4]. In this subsection, we discuss the Poisson approximation of the number of non-overlapping k-runs among n i.i.d. Bernoulli random variables, denoted by $S_{n,k}$. In particular, we obtain an explicit bound on the pointwise difference between the cumulative distribution functions of $S_{n,k}$ and $P_{\mathbb{E}[S_{n,k}]}$ that is independent from the number k of required successes in a row.

Let $k \in \mathbb{N}$ and X_j, $j \in \mathbb{N}_0$, be a sequence of independent and Bernoulli distributed random variables with parameter $0 < p \leq 1/2$. We denote by $I^{(i)}$ with $i \in \mathbb{N}_0$ the random variable

$$I^{(i)} = \mathbf{1}\{X_{i-1} = 0, X_i = 1, \ldots, X_{i+k-1} = 1\},$$

where $X_{-1} = 0$. For $k \leq n$ the number $S_{n,k}$ of non-overlapping k-runs in $X_0, \ldots, X_{n-1}$ is given by

$$S_{n,k} = \sum_{i=0}^{n-k} I^{(i)}. \tag{3.43}$$

Theorem 3.3.5. *Let $S_{n,k}$ be the random variable given by* (3.43) *with $k, n \in \mathbb{N}, k \leq n$. Then,*

$$d_{TV}\big(S_{n,k}, P_{\mathbb{E}[S_{n,k}]}\big) \leq (2k+1)\big(1 \wedge \mathbb{E}[S_{n,k}]\big)p^k. \tag{3.44}$$

Moreover, for $v \in \mathbb{N}_0$ and $n \geq 2$,

$$\max_{1 \leq k \leq n} |\mathbb{P}(S_{n,k} \leq v) - \mathbb{P}(P_{\mathbb{E}[S_{n,k}]} \leq v)| \leq 40(v+2)^2 \frac{\log n}{n}. \tag{3.45}$$

The bound (3.44) was shown in [46, Corollary 15] as a consequence of [1, Theorem 1]. The Poisson approximation for $S_{n,k}$ is also investigated in e.g. [2, 10, 27, 35]. The explicit bound in (3.45) on the pointwise difference between the cumulative distribution functions of $S_{n,k}$ and $P_{\mathbb{E}[S_{n,k}]}$ does not depend on the number k of required successes in a row. Hence, (3.45) improves [45, Corollary 3.23] and [46, Corollary 16] because we found an explicit bound. Furthermore, since the proof of Theorem 3.3.5 is based on Theorem 3.1.1, by applying the second inequality of (3.3) in Theorem 3.1.1, it is possible to attain a bound on the Wasserstein distance between $S_{n,k}$ and $P_{\mathbb{E}[S_{n,k}]}$.

For the proof of Theorem 3.3.5 we define

$$U_\ell = \sum_{i=0 \vee (\ell-k)}^{(n-k)\wedge(\ell+k)} I^{(i)}, \quad \ell = 0, \ldots, n-k,$$

where $a \vee b = \max\{a, b\}$ for any $a, b \in \mathbb{R}$, and let Y be a random variable independent from $X_j, j \in \mathbb{N}_0$, and with distribution given by

$$\mathbb{P}(Y = \ell) = \frac{\mathbb{E}[I^{(\ell)}]}{\mathbb{E}[S_{n,k}]}, \quad \ell = 0, \ldots, n-k.$$

The next proposition is derived by a standard construction of size-bias couplings (see Lemma 2.4.4).

Proposition 3.3.6. *Let $k, n \in \mathbb{N}$ with $k \leq n$. For any $m \in \mathbb{N}$,*

$$m\mathbb{P}(S_{n,k} = m) = \mathbb{E}[S_{n,k}]\mathbb{P}(S_{n,k} - U_Y = m-1).$$

Proof. Let $\ell \in \{0, \ldots, n-k\}$ and $m \in \mathbb{N}$ be fixed. Then, we have

$$\mathbb{E}[I^{(\ell)}\mathbf{1}\{S_{n,k} - I^{(\ell)} = m-1\}] = \mathbb{E}[I^{(\ell)}\mathbf{1}\{S_{n,k} - U_\ell = m-1\}].$$

Since $I^{(\ell)}$ and $S_{n,k} - U_\ell$ are independent, it follows that

$$\begin{aligned} m\mathbb{P}(S_{n,k} = m) &= \sum_{\ell=0}^{n-k} \mathbb{E}[I^{(\ell)}\mathbf{1}\{S_{n,k} = m\}] = \sum_{\ell=0}^{n-k} \mathbb{E}[I^{(\ell)}\mathbf{1}\{S_{n,k} - I^{(\ell)} = m-1\}] \\ &= \sum_{\ell=0}^{n-k} \mathbb{E}[I^{(\ell)}]\mathbb{P}(S_{n,k} - U_\ell = m-1) = \mathbb{E}[S_{n,k}]\mathbb{P}(S_{n,k} - U_Y = m-1), \end{aligned}$$

which concludes the proof. □

Remark 3.3.7. *Since $U_Y \geq 0$, from Corollary 3.2.4 b) it follows that $\mathbb{P}(S_{n,k} = 0) \leq e^{-\mathbb{E}[S_{n,k}]}$. Thus, straightforward calculations imply that*

$$\mathbb{P}(S_{n,k} = 0) \leq \exp\big(- (n-k+1)p^k(1-p)\big).$$

Proof of Theorem 3.3.5. From (3.3) in Theorem 3.1.1 and Proposition 3.3.6, it follows that

$$d_{TV}(S_{n,k}, P_{\mathbb{E}[S_{n,k}]}) \leq (1 \wedge \mathbb{E}[S_{n,k}])\mathbb{E}[U_Y] \leq (2k+1)(1 \wedge \mathbb{E}[S_{n,k}])p^k,$$

where we used $\mathbb{E}[U_\ell] \leq (2k+1)p^k$ for $\ell = 0, \dots, n-k$ in the last step. This proves (3.44).

Let $n \geq 2$ be fixed. Since $(2k+1)p^k, k \geq 1$, is decreasing in k for any $p \leq 1/2$, by (3.44) we deduce for $k \geq 2 \log n$ that

$$|\mathbb{P}(S_{n,k} \leq v) - \mathbb{P}(P_{\mathbb{E}[S_{n,k}]} \leq v)| \leq (2k+1)p^k \leq (4\log n + 1)2^{-2\log n} \leq \frac{4\log n + 1}{n}. \tag{3.46}$$

Let $k < 2\log n$. From (3.4) in Theorem 3.1.1 with $m = 1$ for $v = 0$ and (3.5) in Theorem 3.1.1 for $v \in \mathbb{N}$, it follows that

$$|\mathbb{P}(S_{n,k} \leq v) - \mathbb{P}(P_{\mathbb{E}[S_{n,k}]} \leq v)| \leq \frac{(v+1)^2\mathbb{E}[U_Y]}{\mathbb{E}[S_{n,k}]} + \mathbb{E}[U_Y \mathbf{1}\{S_{n,k} - U_Y \leq v\}]. \tag{3.47}$$

From $0 \leq U_\ell \leq 2$ for $\ell \in \{0, \dots, n-k\}$ and the definition of Y it follows that

$$\begin{aligned}\mathbb{E}[U_Y \mathbf{1}\{S_{n,k} - U_Y \leq v\}] \leq \mathbb{E}[U_Y \mathbf{1}\{S_{n,k} \leq v+2\}] &= \mathbb{E}\sum_{\ell=0}^{n-k} \frac{\mathbb{E}[I^{(\ell)}]}{\mathbb{E}[S_{n,k}]}\mathbb{E}[U_\ell \mathbf{1}\{S_{n,k} \leq v+2\}] \\ &\leq \frac{p^k}{\mathbb{E}[S_{n,k}]}\mathbb{E}\sum_{\ell=0}^{n-k}\sum_{i=0\vee(\ell-k)}^{(n-k)\wedge(\ell+k)} I^{(i)}\mathbf{1}\{S_{n,k} \leq v+2\}.\end{aligned}$$

Thus, by the inequality

$$\sum_{\ell=0}^{n-k}\sum_{i=0\vee(\ell-k)}^{(n-k)\wedge(\ell+k)} a_i \leq (2k+1)\sum_{m=0}^{n-k} a_m, \quad a_0, \dots, a_{n-k} \geq 0,$$

we obtain

$$\mathbb{E}[U_Y \mathbf{1}\{S_{n,k} - U_Y \leq v\}] \leq \frac{(2k+1)p^k}{\mathbb{E}[S_{n,k}]}\mathbb{E}[S_{n,k}\mathbf{1}\{S_{n,k} \leq v+2\}] \leq \frac{(2k+1)p^k(v+2)}{\mathbb{E}[S_{n,k}]}.$$

Together with (3.47) and the inequalities

$$\mathbb{E}[S_{n,k}] \geq (n-k+1)p^k/2 \quad \text{and} \quad \mathbb{E}\big[U_Y\big] \leq (2k+1)p^k,$$

this shows for $k < 2\log n$ and $n > 4\log n$ that

$$\begin{aligned}|\mathbb{P}(S_{n,k} \leq v) - \mathbb{P}(P_{\mathbb{E}[S_{n,k}]} \leq v)| &\leq \frac{2(v+1)^2(2k+1)}{n-k+1} + \frac{2(v+2)(2k+1)}{n-k+1} \\ &\leq \frac{4(v+2)^2(4\log n + 1)}{n - 2\log n} \leq \frac{40(v+2)^2\log n}{n},\end{aligned}$$

where we used the inequalities $4\log n + 1 \leq 5\log n$ and $n - 2\log n \geq n/2$ for $n > 4\log n$ in the last step. Combining this and (3.46) establishes (3.45) for $n > 4\log n$. In conclusion, note that $n > 4\log n$ for $n > 10$, and for $2 \leq n \leq 10$, the right-hand side of (3.45) is greater than 1. Thus, (3.45) holds for all $n \geq 2$. □

3.3.5 Minimal circumscribed radii of stationary Poisson-Voronoi tessellations

In this subsection, we consider circumscribed radii of stationary Poisson-Voronoi tessellations. The aim is to continue the work started in [15] by proving that the Kolmogorov distance between a transform of the minimal circumscribed radius and a Weibull random variable converges to 0 at a rate of $1/t^{1/(d+1)}$ when the intensity t of the underlying Poisson process goes to infinity.

For any locally finite counting measure ν on $\mathbb{R}^d$, we denote by $N(x,\nu)$ the Voronoi cell with nucleus $x \in \mathbb{R}^d$ generated by $\nu + \delta_x$, that is

$$N(x,\nu) = \left\{ y \in \mathbb{R}^d \, : \, \|y - x\| \leq \|y - x'\|, x \neq x' \in \nu \right\},$$

where $\|\cdot\|$ denotes the Euclidean norm. Voronoi tessellations, i.e., tessellations consisting of Voronoi cells $N(x,\nu)$, $x \in \nu$, arise in different fields such as biology [53], astrophysics [54] and communication networks [12]. For more details on Poisson-Voronoi tessellations, i.e., Voronoi tessellations generated by an underlying Poisson process, we refer the reader to e.g. [14, 44, 61]. We denote by $\mathbf{B}(x,r)$ the open ball centered at $x \in \mathbb{R}^d$ with radius $r > 0$. The circumscribed radius of the Voronoi cell $N(x,\nu)$ is defined as

$$C(x,\nu) = \inf \left\{ R \geq 0 \, : \, \mathbf{B}(x,R) \supset N(x,\nu) \right\},$$

i.e., the circumscribed radius is the smallest radius for which the ball centered at the nucleus contains the cell.

Throughout this subsection we consider the stationary Poisson-Voronoi tessellation generated by a Poisson process η_t on $\mathbb{R}^d$ with intensity measure $t\lambda_d$, $t > 0$, where λ_d is the d-dimensional Lebesgue measure. Let $W \subset \mathbb{R}^d$ be a measurable set with $\lambda_d(W) = 1$. For any Voronoi cell $N(x,\eta_t)$ with $x \in \eta_t \cap W$, we take the circumscribed radius of the cell, and we define the point process ξ_t on the positive half line as

$$\xi_t = \sum_{x \in \eta_t \cap W} \delta_{\alpha_2 k_d t^{(d+2)/(d+1)} C(x,\eta_t)^d}. \tag{3.48}$$

Here k_d denotes the volume of the d-dimensional unit ball, and the constant $\alpha_2 > 0$ is given by

$$\alpha_2 = \left(\frac{2^{d(d+1)}}{(d+1)!} p_{d+1} \right)^{1/(d+1)} \tag{3.49}$$

with

$$p_{d+1} := \mathbb{P}\Big(N\Big(0, \sum_{j=1}^{d+1} \delta_{Y_j}\Big) \subseteq \mathbf{B}(0,1)\Big), \tag{3.50}$$

where $Y_1, \ldots, Y_{d+1}$ are independent and uniformly distributed random points in $\mathbf{B}(0,2)$. We denote by T_t the first arrival time of ξ_t, i.e.,

$$T_t = \min_{x \in \eta_t \cap W} \alpha_2 k_d t^{(d+2)/(d+1)} C(x,\eta_t)^d, \tag{3.51}$$

which is - up to a rescaling - the d-th power of the minimal circumscribed radius of the cells with nucleus in W. Recall that a random variable Y has a Weibull distribution if its cumulative distribution function is given by $\mathbb{P}(Y \leq u) = 1 - e^{-(u/s)^k}$ for $u \geq 0$, and 0 otherwise; $k > 0$ is the shape parameter and $s > 0$ is the scale parameter.

Theorem 3.3.8. *Suppose $t \geq 1$. Let ξ_t and T_t be the point process and the random variable given by* (3.48) *and* (3.51)*, respectively. Let Y be a Weibull distributed random variable with shape parameter $d+1$ and scale parameter 1. Then, there exist constants $C_{\mathrm{TV}}, C_K > 0$ only depending on d such that*

$$d_{TV}\big(\xi_t([0,u]), P_{u^{d+1}}\big) \leq C_{\mathrm{TV}} \frac{u^{d+2}}{t^{1/(d+1)}} \tag{3.52}$$

for $u > 0$, and

$$d_K(T_t, Y) \leq \frac{C_K}{t^{1/(d+1)}}. \tag{3.53}$$

Note that explicit formulas for the constants C_{TV} and C_K are given in the proof of Theorem 3.3.8. In [15, Theorem 1, Equation (2d)], the weak convergence of T_t to Y as $t \to \infty$ is shown. For an underlying inhomogeneous Poisson process, the weak convergence of ξ_t to a Poisson process and the weak convergence of T_t to Y are proven in Subsection 5.3.3. Although we only consider stationary Poisson processes, we believe that the arguments employed in this subsection may also establish similar results on the minimal circumscribed radius for more general Poisson processes with a different rate of convergence in t under some constraints on the density (e.g. Hölder continuity). To the best of our knowledge, the present work is the first time the rates of convergence for the Poisson approximation of $\xi_t([0,u])$ and the Weibull approximation of T_t have been addressed.

The proof of Theorem 3.3.8 requires several preparations. We set

$$s_t = \alpha_2 k_d t^{(d+2)/(d+1)}.$$

Let M_t denote the intensity measure of ξ_t, and let the quantities $\widehat{M}_t$ and θ_t on $[0,\infty)$ be defined by

$$\widehat{M}_t([0,u]) = t \int_W \mathbb{E}\big[\mathbf{1}\{s_t C(x, \eta_t + \delta_x)^d \leq u\} \mathbf{1}\{\eta_t\big(\mathbf{B}\big(x, 4(u/s_t)^{1/d}\big)\big) = d+1\}\big]\, dx,$$

$$\theta_t([0,u]) = t \int_W \mathbb{E}\big[\mathbf{1}\{s_t C(x, \eta_t + \delta_x)^d \leq u\} \mathbf{1}\{\eta_t\big(\mathbf{B}\big(x, 4(u/s_t)^{1/d}\big)\big) > d+1\}\big]\, dx$$

for $u \geq 0$. For $x \in W$ and $u \geq 0$ we have

$$\eta_t(\mathbf{B}(x, 2(u/s_t)^{1/d})) \geq d+1 \quad \text{whenever} \quad s_t C(x, \eta_t + \delta_x)^d \leq u. \tag{3.54}$$

This is the case since $s_t C(x, \eta_t + \delta_x)^d \leq u$ implies that the nuclei of the neighboring cells of x are in $\mathbf{B}(x, 2(u/s_t)^{1/d})$ and each Voronoi cell has at least $d+1$ neighboring cells. From the Mecke formula and (3.54) it follows that

$$M_t([0,u]) = \widehat{M}_t([0,u]) + \theta_t([0,u]), \quad u \geq 0.$$

Lemma 3.3.9. *For all $u > 0$ and $t > 0$,*

$$\widehat{M}_t([0,u]) = u^{d+1} \exp\Big(-\frac{4^d u}{\alpha_2 t^{1/(d+1)}}\Big),\ \theta_t([0,u]) \leq \frac{2^{d(d+3)}}{\alpha_2 p_{d+1}} \frac{u^{d+2}}{t^{1/(d+1)}} \text{ and } M_t([0,u]) \leq \frac{u^{d+1}}{p_{d+1}}.$$

Proof. First we compute $\widehat{M}_t([0,u])$. From (3.54) and the definition of p_{d+1} in (3.50) we derive

$$\begin{aligned} \widehat{M}_t([0,u]) = t \int_W & e^{-2^d k_d t u/s_t} \frac{\big(2^d k_d t u/s_t\big)^{d+1}}{(d+1)!} p_{d+1} \\ & \times \mathbb{P}\big(\eta_t\big(\mathbf{B}\big(x, 4(u/s_t)^{1/d}\big) \setminus \mathbf{B}\big(x, 2(u/s_t)^{1/d}\big)\big) = 0\big)\, dx. \end{aligned}$$

Substituting $s_t = \alpha_2 k_d t^{(d+2)/(d+1)}$ and $\alpha_2 = \big(\frac{2^{d(d+1)}}{(d+1)!} p_{d+1}\big)^{1/(d+1)}$ into the previous equation implies that the right-hand side equals

$$\begin{aligned} u^{d+1} \int_W \exp\Big(- \frac{2^d u}{\alpha_2 t^{1/(d+1)}} - t\lambda_d\big(\mathbf{B}\big(x, 4(u/s_t)^{1/d}\big) \setminus \mathbf{B}\big(x, 2(u/s_t)^{1/d}\big)\big)\Big) dx \\ = u^{d+1} \exp\Big(- \frac{2^d u}{\alpha_2 t^{1/(d+1)}} - \frac{2^d u}{\alpha_2 t^{1/(d+1)}}(2^d - 1)\Big) = u^{d+1} \exp\Big(- \frac{4^d u}{\alpha_2 t^{1/(d+1)}}\Big), \end{aligned}$$

which completes the first part of the proof.

For $u > 0$, we have

$$\theta_t([0, u]) \leq t \int_W \mathbb{E}\big[\mathbf{1}\big\{\eta_t\big(\mathbf{B}\big(x, 4(u/s_t)^{1/d}\big)\big) > d + 1\big\}\big] dx = t \sum_{k=d+2}^{\infty} e^{-\beta_t} \frac{\beta_t^k}{k!}$$

with $\beta_t = 4^d k_d t u/s_t$. Elementary calculations imply that

$$\begin{aligned} t \sum_{k=d+2}^{\infty} e^{-\beta_t} \frac{\beta_t^k}{k!} = t\beta_t^{d+2} \sum_{k=d+2}^{\infty} e^{-\beta_t} \frac{\beta_t^{k-d-2}}{k!} = t\beta_t^{d+2} \sum_{\ell=0}^{\infty} e^{-\beta_t} \frac{\beta_t^{\ell}}{(\ell + d + 2)!} \\ \leq \frac{t\beta_t^{d+2}}{(d+2)!} = \frac{t\big(4^d k_d t u/s_t\big)^{d+2}}{(d+2)!}. \end{aligned}$$

Substituting $s_t = \alpha_2 k_d t^{(d+2)/(d+1)}$ and $\alpha_2 = \big(\frac{2^{d(d+1)}}{(d+1)!} p_{d+1}\big)^{1/(d+1)}$ into the latter term yields

$$\theta_t([0, u]) \leq \frac{2^{d(d+3)}}{\alpha_2 p_{d+1}} \frac{u^{d+2}}{t^{1/(d+1)}},$$

which is the desired result.

From the Mecke formula, (3.54) and the same arguments as above, we obtain

$$\begin{aligned} M_t([0, u]) \leq t \int_W \mathbb{P}(\eta_t(\mathbf{B}(x, 2(u/s_t)^{1/d})) \geq d + 1) dx = t \sum_{k=d+1}^{\infty} \frac{(2^d k_d t u/s_t)^k}{k!} e^{-2^d k_d t u/s_t} \\ \leq \frac{t(2^d k_d t u/s_t)^{d+1}}{(d+1)!} = \frac{2^{d(d+1)} k_d^{d+1} t^{d+2} u^{d+1}}{k_d^{d+1} \frac{2^{d(d+1)} p_{d+1}}{(d+1)!} (d+1)! t^{d+2}} = \frac{u^{d+1}}{p_{d+1}}, \end{aligned}$$

which concludes the proof. □

We now provide a statement that will be employed in the proof of the subsequent proposition. This result is a direct consequence of Lemma 5.3.14 in Subsection 5.3.3.

Lemma 3.3.10. *Let $x_0, \dots, x_{d+1} \in \mathbb{R}^d$ be in general position (i.e., no k-dimensional affine subspace of $\mathbb{R}^d$ with $k \in \{0, \dots, d-1\}$ contains more than $k+1$ of the points) and assume that $N(x_0, \sum_{j=0}^{d+1} \delta_{x_i})$ is bounded. Then $N(x_i, \sum_{j=0}^{d+1} \delta_{x_i})$ is unbounded for any $i \in \{1, \dots, d+1\}$.*

Next we construct a random variable that satisfies (3.8) for $\xi_t([0, u])$ with remainder terms $q_i, i \in \mathbb{N}_0$, which vanish as $t \to \infty$. By B^c we denote the complement of $B \subset \mathbb{R}^d$ and by $\eta_t|_B$ the restriction of η_t to B.

Proposition 3.3.11. *Let X be uniformly distributed in W and independent of η_t. Then for $u > 0$,*

$$k\mathbb{P}(\xi_t([0,u]) = k) = \widehat{M}_t([0,u])\mathbb{P}(\xi_t([0,u]) + Z_{t,u} = k-1) + q_{k-1}(t,u), \quad k \in \mathbb{N},$$

with

$$Z_{t,u} = \xi_t\big(\eta_t|_{\mathbf{B}(X,4(u/s_t)^{1/d})^c}\big)([0,u]) - \xi_t([0,u])$$

and

$$\begin{aligned} q_i(t,u) = t\int_W \mathbb{E}\Big[&\mathbf{1}\{s_tC(x,\eta_t+\delta_x)^d \le u\}\mathbf{1}\{\eta_t\big(\mathbf{B}\big(x,4(u/s_t)^{1/d}\big)\big) > d+1\} \\ &\times \mathbf{1}\Big\{\sum_{y\in\eta_t\cap W}\mathbf{1}\{s_tC(y,\eta_t+\delta_x)^d \le u\} = i\Big\}\Big]dx \end{aligned}$$

for $i \in \mathbb{N}_0$.

Proof. The Mecke equation implies for $k \in \mathbb{N}$ that

$$\begin{aligned} k\mathbb{P}(\xi_t([0,u]) = k) &= t\int_W \mathbb{E}\big[\mathbf{1}\{s_tC(x,\eta_t+\delta_x)^d \le u\}\mathbf{1}\{\xi_t(\eta_t+\delta_x)([0,u]) = k\}\big]dx \\ &= t\int_W \mathbb{E}\Big[\mathbf{1}\{s_tC(x,\eta_t+\delta_x)^d \le u\}\mathbf{1}\Big\{\sum_{y\in\eta_t\cap W}\mathbf{1}\{s_tC(y,\eta_t+\delta_x)^d \le u\} = k-1\Big\}\Big]dx. \end{aligned}$$

Now we divide the integral in

$$\begin{aligned} A_k + q_{k-1}(t,u) := t\int_W \mathbb{E}\Big[&\mathbf{1}\{s_tC(x,\eta_t+\delta_x)^d \le u\}\mathbf{1}\{\eta_t\big(\mathbf{B}\big(x,4(u/s_t)^{1/d}\big)\big) = d+1\} \\ &\times \mathbf{1}\Big\{\sum_{y\in\eta_t\cap W}\mathbf{1}\{s_tC(y,\eta_t+\delta_x)^d \le u\} = k-1\Big\}\Big]dx \\ + t\int_W \mathbb{E}\Big[&\mathbf{1}\{s_tC(x,\eta_t+\delta_x)^d \le u\}\mathbf{1}\{\eta_t\big(\mathbf{B}\big(x,4(u/s_t)^{1/d}\big)\big) > d+1\} \\ &\times \mathbf{1}\Big\{\sum_{y\in\eta_t\cap W}\mathbf{1}\{s_tC(y,\eta_t+\delta_x)^d \le u\} = k-1\Big\}\Big]dx. \end{aligned}$$

Then, it is enough to show that $A_k = \widehat{M}_t([0,u])\mathbb{P}(\xi_t([0,u]) + Z_{t,u} = k-1)$. In order to simplify the notation throughout this proof, we write

$$\mathbf{B}_2(x) := \mathbf{B}\big(x, 2(u/s_t)^{1/d}\big) \quad \text{and} \quad \mathbf{B}_4(x) := \mathbf{B}\big(x,4(u/s_t)^{1/d}\big), \quad x \in \mathbb{R}^d.$$

In case there are only $d+1$ points of η_t in $\mathbf{B}_4(x)$, we have by (3.54) that $s_tC(x,\eta_t+\delta_x)^d \le u$ only if the $d+1$ elements of η_t belong to $\mathbf{B}_2(x)$. Therefore we obtain

$$\begin{aligned} A_k = t\int_W \mathbb{E}\Big[&\mathbf{1}\{s_tC(x,\eta_t+\delta_x)^d \le u\}\mathbf{1}\{\eta_t(\mathbf{B}_4(x)\setminus\mathbf{B}_2(x)) = 0, \eta_t(\mathbf{B}_2(x)) = d+1\} \\ &\times \mathbf{1}\Big\{\sum_{y\in\eta_t\cap W}\mathbf{1}\{s_tC(y,\eta_t+\delta_x)^d \le u\} = k-1\Big\}\Big]dx. \end{aligned} \tag{3.55}$$

The observation that

$$s_tC(y,\eta_t+\delta_x)^d \le u \quad \text{if and only if} \quad s_tC\big(y,(\eta_t+\delta_x)|_{\mathbf{B}_2(y)}\big)^d \le u \tag{3.56}$$

for $y \in \eta_t$ establishes that

$$A_k = t \int_W \mathbb{E}\Big[\mathbf{1}\big\{s_t C(x, \eta_t + \delta_x)^d \le u\big\}\mathbf{1}\big\{\eta_t(\mathbf{B}_4(x) \setminus \mathbf{B}_2(x)) = 0, \eta_t(\mathbf{B}_2(x)) = d+1\big\} \\ \times \mathbf{1}\Big\{\xi_t(\eta_t|_{\mathbf{B}_4(x)^c})([0,u]) + \sum_{y \in \eta_t \cap \mathbf{B}_2(x) \cap W} \mathbf{1}\big\{s_t C(y, \eta_t + \delta_x)^d \le u\big\} = k-1\Big\}\Big] dx.$$

Suppose that $s_t C(x, \eta_t + \delta_x)^d \le u$ and that there are exactly $d+1$ points $y_1, \dots, y_{d+1}$ of η_t in $\mathbf{B}_2(x)$ and $\eta_t \cap \mathbf{B}_4(x) \cap \mathbf{B}_2(x)^c = \emptyset$. From Lemma 3.3.10 it follows that the Voronoi cells $N(y_i, \eta_t|_{\mathbf{B}_4(x)} + \delta_x), i = 1, \dots, d+1$, are unbounded. In particular, we have

$$C(y_i, \eta_t + \delta_x) > (u/s_t)^{1/d}, \quad i = 1, \dots, d+1.$$

Together with the same arguments used to show (3.55) and independence, this implies that

$$\begin{aligned} A_k &= t \int_W \mathbb{E}\big[\mathbf{1}\big\{s_t C(x, \eta_t + \delta_x)^d \le u\big\}\mathbf{1}\{\eta_t(\mathbf{B}_4(x) \setminus \mathbf{B}_2(x)) = 0, \eta_t(\mathbf{B}_2(x)) = d+1\} \\ &\qquad \times \mathbf{1}\{\xi_t(\eta_t|_{\mathbf{B}_4(x)^c})([0,u]) = k-1\}\big] dx \\ &= t \int_W \mathbb{E}\big[\mathbf{1}\big\{s_t C(x, \eta_t + \delta_x)^d \le u\big\}\mathbf{1}\{\eta_t(\mathbf{B}_4(x)) = d+1\}\big] \\ &\qquad \times \mathbb{P}\big(\xi_t(\eta_t|_{\mathbf{B}_4(x)^c})([0,u]) = k-1\big) dx. \end{aligned}$$

Then, because the expectation in the latter equation does not depend on the choice of $x \in W$, we have that

$$\begin{aligned} A_k &= \widehat{M_t}([0,u]) \int_W \mathbb{P}\big(\xi_t\big(\eta_t|_{\mathbf{B}_4(x)^c}\big)([0,u]) = k-1\big) dx \\ &= \widehat{M_t}([0,u]) \mathbb{P}(\xi_t([0,u]) + Z_{t,u} = k-1) \end{aligned}$$

with

$$Z_{t,u} = \xi_t\big(\eta_t|_{\mathbf{B}_4(X)^c}\big)([0,u]) - \xi_t([0,u]).$$

This and $\mathbf{B}_4(X) = \mathbf{B}\big(X, 4(u/s_t)^{1/d}\big)$ give the desired conclusion. □

Lemma 3.3.12. *For $u > 0$, $t > 0$ and $Z_{t,u}$ as in Proposition 3.3.11,*

$$\mathbb{E}[|Z_{t,u}|] \le \frac{6^d}{\alpha_2 p_{d+1}} \frac{u^{d+2}}{t^{(d+2)/(d+1)}}.$$

Proof. For $x \in W$ it follows from the observation in (3.56) that

$$0 \le \xi_t([0,u]) - \xi_t\big(\eta_t|_{\mathbf{B}(x, 4(u/s_t)^{1/d})^c}\big)([0,u]) \le \sum_{y \in \eta_t \cap W \cap \mathbf{B}(x, 6(u/s_t)^{1/d})} \mathbf{1}\{s_t C(y, \eta_t)^d \le u\}.$$

By the Mecke formula and the stationarity of η_t, we obtain

$$\begin{aligned} \mathbb{E} \sum_{y \in \eta_t \cap W \cap \mathbf{B}(x, 6(u/s_t)^{1/d})} \mathbf{1}\{s_t C(y, \eta_t)^d \le u\} &\le t\lambda_d(W \cap \mathbf{B}(x, 6(u/s_t)^{1/d}))\mathbb{P}(s_t C(0, \eta_t + \delta_0)^d \le u) \\ &\le \frac{6^d u}{\alpha_2 t^{(d+2)/(d+1)}} t\mathbb{P}(s_t C(0, \eta_t + \delta_0)^d \le u). \end{aligned}$$

From Lemma 3.3.9 we deduce

$$t\mathbb{P}(s_t C(0, \eta_t + \delta_0)^d \le u) = M_t([0,u]) \le \frac{u^{d+1}}{p_{d+1}},$$

which proves the assertion. □

Lemma 3.3.13. *For $u > 0$ and $t > 0$,*

$$\mathbb{P}(T_t > u) = \mathbb{P}(\xi_t([0,u]) = 0) \le e^{-\widehat{M}_t([0,u])}.$$

Proof. The first identity is obvious. Let $Z_{t,u}$ be the random variable defined in Proposition 3.3.11. Since $Z_{t,u} \le 0$, $\mathbb{P}(\xi_t([0,u]) + Z_{t,u} \ge 0) = 1$ and $q_i(t,u) \ge 0$ for all $i \in \mathbb{N}_0$, the inequality follows from Proposition 3.2.3 b). □

In the next lemma, we combine the results obtained above and Theorem 3.1.3 to derive intermediate bounds on the quantities considered in Theorem 3.3.8.

Lemma 3.3.14. *For $u > 0$ and $t > 0$,*

$$d_{TV}\big(\xi_t([0,u]), P_{\widehat{M}_t([0,u])}\big) \le \frac{6^d}{\alpha_2 p_{d+1}} \frac{u^{d+2}}{t^{(d+2)/(d+1)}} + \theta_t([0,u]) \tag{3.57}$$

and

$$0 \le e^{-\widehat{M}_t([0,u])} - \mathbb{P}(T_t > u) \le \left(1 + \frac{1}{\widehat{M}_t([0,u])}\right) \frac{6^d}{\alpha_2 p_{d+1}} \frac{u^{d+2}}{t^{(d+2)/(d+1)}} + \frac{2\theta_t([0,u])}{\widehat{M}_t([0,u])^2}. \tag{3.58}$$

Proof. From Proposition 3.3.11 it follows that the assumptions of Theorem 3.1.3 are satisfied. Then, (3.9) in Theorem 3.1.3 yields

$$d_{TV}(\xi_t([0,u]), P_{\widehat{M}_t([0,u])}) \le (1 \wedge \widehat{M}_t([0,u]))\mathbb{E}[|Z_{t,u}|] + \big(1 \wedge \widehat{M}_t([0,u])^{-1/2}\big)\theta_t([0,u])$$

so that (3.57) follows from Lemma 3.3.12.

Let us now prove (3.58). From Lemma 3.3.13, (3.12) in Theorem 3.1.3 with $m = 1$ and $\sum_{i=1}^{\infty} q_i(t,u) \le \theta_t([0,u])$ we obtain

$$0 \le e^{-\widehat{M}_t([0,u])} - \mathbb{P}(T_t > u) \le \frac{\mathbb{E}\big[|Z_{t,u}|\big]}{\widehat{M}_t([0,u])} + \mathbb{E}\big[|Z_{t,u}|\big] + \frac{q_0(t,u)}{\widehat{M}_t([0,u])} + \frac{\theta_t([0,u])}{\widehat{M}_t([0,u])^2}.$$

The first two terms on the right-hand side can be bounded by Lemma 3.3.12. Recall that

$$q_0(t,u) = t \int_W \mathbb{E}\Big[\mathbf{1}\big\{s_t C(x, \eta_t + \delta_x)^d \le u\big\}\mathbf{1}\big\{\eta_t\big(\mathbf{B}\big(x, 4(u/s_t)^{1/d}\big)\big) > d+1\big\} \\ \times \mathbf{1}\Big\{\sum_{y \in \eta_t \cap W} \mathbf{1}\{s_t C(y, \eta_t + \delta_x)^d \le u\} = 0\Big\}\Big] dx.$$

Since the product of the first two indicator functions is increasing with respect to additional points, while the third indicator function is decreasing, it follows from [38, Theorem 20.4] that

$$q_0(t,u) \le t \int_W \mathbb{E}\Big[\mathbf{1}\big\{s_t C(x, \eta_t + \delta_x)^d \le u\big\}\mathbf{1}\big\{\eta_t\big(\mathbf{B}\big(x, 4(u/s_t)^{1/d}\big)\big) > d+1\big\}\Big] \\ \times \mathbb{P}\Big(\sum_{y \in \eta_t \cap W} \mathbf{1}\{s_t C(y, \eta_t + \delta_x)^d \le u\} = 0\Big) dx.$$

Now Lemma 3.3.13 and the elementary inequality $ve^{-v} \le 1$ for $v \ge 0$ lead to

$$q_0(t,u) \le \theta_t([0,u])\mathbb{P}(\xi_t([0,u]) = 0) \le \theta_t([0,u])e^{-\widehat{M}_t([0,u])} \le \frac{\theta_t([0,u])}{\widehat{M}_t([0,u])},$$

which concludes the proof. □

Proof of Theorem 3.3.8. Let $u > 0$ be fixed. From (3.57) in Lemma 3.3.14, Lemma 3.3.9 and $t \geq 1$ it follows that

$$d_{TV}\big(\xi_t([0,u]), P_{\widehat{M}_t([0,u])}\big) \leq \frac{6^d}{\alpha_2 p_{d+1}} \frac{u^{d+2}}{t^{(d+2)/(d+1)}} + \theta_t([0,u]) \leq \frac{6^d + 2^{d(d+3)}}{\alpha_2 p_{d+1}} \frac{u^{d+2}}{t^{1/(d+1)}}. \tag{3.59}$$

Using a well-known bound for the total variation distance between two Poisson distributed random variables, Lemma 3.3.9 and the inequality $1 - e^{-v} \leq v$ for $v \geq 0$, we obtain

$$d_{TV}\big(P_{u^{d+1}}, P_{\widehat{M}_t([0,u])}\big) \leq u^{d+1} - \widehat{M}_t([0,u]) = u^{d+1}\left(1 - \exp\left(-\frac{4^d u}{\alpha_2 t^{1/(d+1)}}\right)\right) \leq \frac{4^d u^{d+2}}{\alpha_2 t^{1/(d+1)}}.$$

Now the triangle inequality yields

$$d_{TV}\big(\xi_t([0,u]), P_{u^{d+1}}\big) \leq \frac{3 \cdot 2^{d(d+3)}}{\alpha_2 p_{d+1}} \frac{u^{d+2}}{t^{1/(d+1)}},$$

which proves (3.52).

Let us now show (3.53). From (3.59) and Lemma 3.3.13 we have that, for $u \in [0,1]$,

$$0 \leq e^{-\widehat{M}_t([0,u])} - \mathbb{P}(T_t > u) \leq \frac{2^{d(d+3)+1}}{\alpha_2 p_{d+1}} \frac{1}{t^{1/(d+1)}}.$$

In the following we consider the case $1 \leq u \leq t^{1/(2d+2)}\tau$ with $\tau = \alpha_2/4^d$. From Lemma 3.3.9 and $t \geq 1$ we obtain

$$u^{d+1} \geq \widehat{M}_t([0,u]) \geq \frac{u^{d+1}}{e}. \tag{3.60}$$

Together with Lemma 3.3.13, (3.58) in Lemma 3.3.14, Lemma 3.3.9 and $u \geq 1$ we obtain

$$\begin{aligned} 0 \leq e^{-\widehat{M}_t([0,u])} - \mathbb{P}(T_t > u) &\leq \left(1 + \frac{1}{\widehat{M}_t([0,u])}\right) \frac{6^d}{\alpha_2 p_{d+1}} \frac{u^{d+2}}{t^{(d+2)/(d+1)}} + \frac{2\theta_t([0,u])}{\widehat{M}_t([0,u])^2} \\ &\leq (1+e) \frac{6^d}{\alpha_2 p_{d+1}} \frac{u^{d+2}}{t^{(d+2)/(d+1)}} + 2e^2 \frac{1}{u^{2d+2}} \frac{2^{d(d+3)}}{\alpha_2 p_{d+1}} \frac{u^{d+2}}{t^{1/(d+1)}}. \end{aligned}$$

Using $1 \leq u^{d+2} \leq t^{(d+2)/(2d+2)} \alpha_2^{d+2} / 4^{d(d+2)}$, $t \geq 1$ and the definition of α_2 in (3.49), we deduce

$$\begin{aligned} 0 \leq e^{-\widehat{M}_t([0,u])} - \mathbb{P}(T_t > u) &\leq \frac{(1+e)6^d}{4^{d(d+2)}} \frac{\alpha_2^{d+1}}{p_{d+1}} \frac{1}{t^{1/(d+1)}} + 2e^2 \frac{1}{u^{2d+2}} \frac{2^{d(d+3)}}{\alpha_2 p_{d+1}} \frac{u^{d+2}}{t^{1/(d+1)}} \\ &\leq \frac{1}{t^{1/(d+1)}} + \frac{2^{d(d+3)+4}}{\alpha_2 p_{d+1}} \frac{1}{t^{1/(d+1)}} \end{aligned}$$

so that

$$\sup_{u \in [0, t^{1/(2d+2)}\tau]} |e^{-\widehat{M}_t([0,u])} - \mathbb{P}(T_t > u)| \leq \left[1 + \frac{2^{d(d+3)+4}}{\alpha_2 p_{d+1}}\right] \frac{1}{t^{1/(d+1)}}.$$

Moreover, by Lemma 3.3.9, (3.60) and elementary arguments we obtain for $0 \leq u \leq t^{1/(2d+2)}\tau$ that

$$\begin{aligned} 0 \leq e^{-\widehat{M}_t([0,u])} - e^{-u^{d+1}} &\leq \left[u^{d+1} - \widehat{M}_t([0,u])\right] e^{-\widehat{M}_t([0,u])} \\ &\leq \frac{4^d u^{d+2}}{\alpha_2 t^{1/(d+1)}} e^{-u^{d+1}e^{-1}} \leq \frac{4^d e^{\frac{d+2}{d+1}}}{\alpha_2 t^{1/(d+1)}} \leq \frac{2^{2d+3}}{\alpha_2 t^{1/(d+1)}}, \end{aligned}$$

where we used the inequalities $1-e^{-x}\le x$ and $e^{-x^{d+1}}x^{d+2}\le 1$ for $x\ge 0$. This implies that

$$\sup_{u\in[0,t^{1/(2d+2)}\tau]}|e^{-u^{d+1}}-\mathbb{P}(T_t>u)|\le\left[1+\frac{2^{d(d+3)+4}}{\alpha_2 p_{d+1}}+\frac{2^{2d+3}}{\alpha_2}\right]\frac{1}{t^{1/(d+1)}}.$$

On the other hand, $x^2e^{-x^{d+1}}\le 1$ for $x\ge 0$ leads to

$$\exp\left(-(t^{1/(2d+2)}\tau)^{d+1}\right)\le\frac{1}{(t^{1/(2d+2)}\tau)^2}\le\frac{16^d}{\alpha_2^2}\frac{1}{t^{1/(d+1)}}.$$

Combining the two previous inequalities gives a bound for $\mathbb{P}\left(T_t>t^{1/(2d+2)}\tau\right)$ and it implies

$$\begin{aligned}&\sup_{u\in[0,\infty)}|e^{-u^{d+1}}-\mathbb{P}(T_t>u)|\\&\le\max\Big\{\sup_{u\in[0,t^{1/(2d+2)}\tau]}|e^{-u^{d+1}}-\mathbb{P}(T_t>u)|,\mathbb{P}\left(T_t>t^{1/(2d+2)}\tau\right),\exp\left(-(t^{1/(2d+2)}\tau)^{d+1}\right)\Big\}\\&\le\left[1+\frac{2^{d(d+3)+4}}{\alpha_2 p_{d+1}}+\frac{2^{2d+3}}{\alpha_2}+\frac{16^d}{\alpha_2^2}\right]\frac{1}{t^{1/(d+1)}}.\end{aligned}$$

Now the identity $\mathbb{P}(T_t>0)=1$ concludes the proof. □

3.3.6 Maximal inradii of stationary Poisson-Voronoi tessellations

In this subsection, we consider the inradii of stationary Poisson-Voronoi tessellations. Recall that the inradius of a cell is the largest radius for which the ball centered at the nucleus is contained in the cell. The aim is to continue the work started in [15] by proving that the Kolmogorov distance between a transform of the largest inradius and a Gumbel random variable converges to 0 at a rate of $\log(t)/\sqrt{t}$ as the intensity t of the underlying Poisson process goes to infinity. More details on Poisson-Voronoi tessellations are given in Subsection 3.3.5.

Let $W\subset\mathbb{R}^d$ be a measurable set with Lebesgue measure $\lambda_d(W)=1$. Let η_t be a Poisson process on $\mathbb{R}^d$ with intensity measure $t\lambda_d, t>0$. Consider the measurable function $h_t: W\times\mathcal{N}(\mathbb{R}^d)\to\mathbb{R}$ defined as

$$h_t(x,\mu)=\min_{y\in\mu\setminus\{x\}}tk_d\|x-y\|^d-\log(t),$$

where $\mu\setminus\{x\}$ denotes $\mu-\delta_x$ if $x\in\mu$, and μ otherwise, k_d is the volume of the d-dimensional unit ball, and $\|\cdot\|$ denotes the Euclidean norm. Note that for any $x\in\eta_t$, $\min\{\|x-y\|\ :\ y\in\eta_t\setminus\{x\}\}$ is twice the inradius of the Voronoi cell with nucleus x generated by η_t. Then, the random variable

$$T_t=\max_{x\in\eta_t\cap W}h_t(x,\eta_t) \tag{3.61}$$

is a transform of the maximal inradius over the cells with nucleus in W. We define the point process ξ_t as

$$\xi_t=\xi_t(\eta_t)=\sum_{x\in\eta_t\cap W}\delta_{h_t(x,\eta_t)}. \tag{3.62}$$

Recall that a random variable Y has a standard Gumbel distribution if its cumulative distribution function is given by $\mathbb{P}(Y\le u)=e^{-e^{-u}}$ for $u\in\mathbb{R}$.

Theorem 3.3.15. *Suppose $t > e^2$. Let T_t and ξ_t be the random variable and the point process given by* (3.61) *and* (3.62)*, respectively. Let Y be a random variable with a standard Gumbel distribution. Then,*

$$d_{TV}\big(\xi_t((u,\infty)), P_{e^{-u}}\big) \leq 2^d \frac{u + \log(t)}{e^{u/2}\sqrt{t}} + \frac{u + \log(t)}{e^u t} \tag{3.63}$$

for $u > -\log(t)$, and

$$d_K(T_t, Y) \leq [2^{d+2}(4^d + 2^d + 2) + 1] \frac{\log(t)}{\sqrt{t}}. \tag{3.64}$$

The main achievement of Theorem 3.3.15 is the rate of convergence for the Kolmogorov distance in (3.64). In [15, Theorem 1, Equation (2a)], the weak convergence of T_t to a Gumbel random variable is proven. For $d = 2$ one obtains from the proof of [19, Proposition 8] that for any fixed $u \in \mathbb{R}$ the difference between $\mathbb{P}(T_t \leq u)$ and $\mathbb{P}(Y \leq u)$ behaves like $O(\log(t)/\sqrt{t})$, where the constant hidden in the big-O-notation depends on u. However this result does not permit the difference between $\mathbb{P}(T_t \leq u)$ and $\mathbb{P}(Y \leq u)$ to be bounded uniformly in $u \in \mathbb{R}$, whence it does not lead to a bound for the Kolmogorov distance. Note that [19, Proposition 8] concerns the maximal inradii of planar Gauss-Voronoi tessellations, which are generated by a Poisson cluster process and include planar Poisson-Voronoi tessellations as a special case. For this model it is shown that for any fixed $u \in \mathbb{R}$, $|\mathbb{P}(T_t \leq u) - \mathbb{P}(Y \leq u)|$ behaves like $O(\log(t)^{-1/2})$, where the big-O-term depends on u.

For an underlying inhomogeneous Poisson process, the weak convergence of ξ_t to a Poisson process and the weak convergence of T_t to Y are established in Subsection 5.3.2, and for an underlying inhomogeneous binomial point process, the weak convergence of T_t to Y is studied in [31, Theorem 1]. As for the results stated in Subsection 3.3.5 about the minimal circumscribed radius, we believe that similar arguments as in this subsection could lead to comparable results with a different rate of convergence in t for the maximal inradius of a Voronoi tessellation generated by an inhomogeneous Poisson processes under some constraints on the density.

Counting cells whose inradius is larger than a given value is equivalent to counting isolated vertices in random geometric graphs. The related problem of finding the longest edge of a k-nearest neighbor graph or a minimal spanning tree is studied, for example, in [49, Chapter 8] or [50] for underlying finite Poisson processes or binomial point processes, where one needs to take care of boundary effects.

Since the proof of Theorem 3.3.15 is based on Theorem 3.1.1, together with the second inequality of (3.3) in Theorem 3.1.1, the same arguments used to show (3.63) may also lead to a bound on the Wasserstein distance between $\xi_t((u,\infty))$ and $P_{e^{-u}}$.

For the proof of Theorem 3.3.15 we introduce some notation. By M_t we denote the intensity measure of ξ_t. For $u > -\log(t)$, set

$$v_t = v_t(u) = \left(\frac{u + \log(t)}{t k_d}\right)^{1/d}. \tag{3.65}$$

Then, for $u > -\log(t)$ we have

$$\begin{aligned} M_t((u,\infty)) &= t \int_W \mathbb{E}\big[\mathbf{1}\{h_t(x, \eta_t + \delta_x) > u\}\big] dx = t \int_W \mathbb{P}\big(\eta_t(\mathbf{B}(x, v_t)) = 0\big) dx \\ &= t \int_W e^{-t v_t^d k_d} dx = t e^{-u - \log(t)} = e^{-u}. \end{aligned}$$

Let X be a uniformly distributed random vector in W independent of η_t. In the next proposition we show that for each $u > -\log(t)$, and for an opportune choice of a random ball $\mathbf{B}$ centered at X, the random variable $\xi_t(\eta_t|_{\mathbf{B}^c})((u,\infty)) - \xi_t((u,\infty))$ satisfies (3.2) for $\xi_t((u,\infty))$, where $\eta_t|_{\mathbf{B}^c}$ denotes the restriction of η_t to the complement of $\mathbf{B}$.

Proposition 3.3.16. *For any $t > e$ and $u > -\log(t)$,*

$$k\mathbb{P}(\xi_t((u,\infty)) = k) = M_t((u,\infty))\mathbb{P}(\xi_t((u,\infty)) + Z_t(u) = k-1), \quad k \in \mathbb{N},$$

where the random variable $Z_t(u)$ is defined as

$$Z_t(u) = \xi_t(\eta_t|_{\boldsymbol{B}(X,v_t)^c})((u,\infty)) - \xi_t((u,\infty))$$

with $v_t = v_t(u)$ given by (3.65).

Proof. Let $B = (u,\infty)$ with $u > -\log(t)$. The Mecke equation yields for $k \in \mathbb{N}$ that

$$k\mathbb{P}(\xi_t(B) = k) = t\int_W \mathbb{E}\big[\mathbf{1}\{h_t(x,\eta_t+\delta_x) > u\}\mathbf{1}\{\xi_t(\eta_t+\delta_x)(B) = k\}\big]dx.$$

Since $h_t(x,\eta_t+\delta_x) > u$ if and only if $\eta_t(\mathbf{B}(x,v_t)) = 0$, the right-hand side equals

$$\begin{aligned}
&t\int_W \mathbb{E}\big[\mathbf{1}\{\eta_t(\mathbf{B}(x,v_t)) = 0\}\mathbf{1}\{\xi_t(\eta_t|_{\mathbf{B}(x,v_t)^c})(B) = k-1\}\big]dx \\
&= t\int_W \mathbb{P}\big(\eta_t(\mathbf{B}(x,v_t)) = 0\big)\mathbb{E}\big[\mathbf{1}\{\xi_t(\eta_t|_{\mathbf{B}(x,v_t)^c})(B) = k-1\}\big]dx \\
&= e^{-u}\int_W \mathbb{P}\big(\xi_t(\eta_t|_{\mathbf{B}(x,v_t)^c})(B) = k-1\big)dx.
\end{aligned}$$

Hence, elementary arguments lead to

$$\begin{aligned}
k\mathbb{P}(\xi_t(B) = k) &= M_t(B)\mathbb{P}\big(\xi_t(\eta_t|_{\mathbf{B}(X,v_t)^c})(B) = k-1\big) \\
&= M_t(B)\mathbb{P}\big(\xi_t(B) + Z_t(u) = k-1\big),
\end{aligned}$$

which is the desired conclusion. ☐

Proof of Theorem 3.3.15. Suppose $u > -\log(t)$ and let $Z_t(u)$ be as in Proposition 3.3.16. We can rewrite $Z_t(u)$ as

$$\begin{aligned}
Z_t(u) &= \xi_t(\eta_t|_{\mathbf{B}(X,v_t)^c})((u,\infty)) - \xi_t((u,\infty)) \\
&= \sum_{z\in\eta_t\cap W\cap\mathbf{B}(X,2v_t)\cap\mathbf{B}(X,v_t)^c} \mathbf{1}\{h_t(z,\eta_t|_{\mathbf{B}(X,v_t)^c}) > u\} - \mathbf{1}\{h_t(z,\eta_t) > u\} \\
&\quad - \sum_{z\in\eta_t\cap\mathbf{B}(X,v_t)\cap W} \mathbf{1}\{h_t(z,\eta_t) > u\} \\
&=: Z'_{t,X}(u) - Z''_{t,X}(u),
\end{aligned}$$

where $Z'_{t,X}(u)$ and $Z''_{t,X}(u)$ are non-negative. For a fixed $x \in W$, the Mecke formula and short computations yield

$$\begin{aligned}
\mathbb{E}[Z'_{t,X}(u)] &\leq \mathbb{E}\Big[\sum_{z\in\eta_t\cap\mathbf{B}(x,2v_t)\cap\mathbf{B}(x,v_t)^c} \mathbf{1}\big\{h_t(z,\eta_t|_{\mathbf{B}(x,v_t)^c}) > u\big\}\Big] \\
&= t\int_{\mathbf{B}(x,2v_t)\cap\mathbf{B}(x,v_t)^c} \mathbb{P}\big(\eta_t(\mathbf{B}(z,v_t)\cap\mathbf{B}(x,v_t)^c) = 0\big)dz \\
&\leq t\int_{\mathbf{B}(x,2v_t)\cap\mathbf{B}(x,v_t)^c} e^{-tv_t^d k_d/2}dz \leq 2^d(u+\log(t))e^{-(u+\log(t))/2} = 2^d\frac{u+\log(t)}{e^{u/2}\sqrt{t}}
\end{aligned} \tag{3.66}$$

and, similarly,

$$\begin{aligned}\mathbb{E}[Z''_{t,X}(u)] &\le \mathbb{E}\Big[\sum_{z\in\eta_t\cap\mathbf{B}(x,v_t)} \mathbf{1}\big\{h_t(z,\eta_t)>u\big\}\Big] = t\int_{\mathbf{B}(x,v_t)} \mathbb{P}\big(\eta_t(\mathbf{B}(z,v_t))=0\big)dz \\ &\le t\int_{\mathbf{B}(x,v_t)} e^{-tv_t^d k_d}dz \le (u+\log(t))e^{-u-\log(t)} = \frac{u+\log(t)}{e^u t}.\end{aligned} \tag{3.67}$$

It follows from the triangle inequality that

$$\mathbb{E}[|Z_t(u)|] \le 2^d\frac{u+\log(t)}{e^{u/2}\sqrt{t}} + \frac{u+\log(t)}{e^u t}.$$

Then, by the first inequality of (3.3) in Theorem 3.1.1, we obtain (3.63).

Let us now show (3.64). We consider the cases $u\ge 0$, $u\in[-\log(\log(t)),0)$ and $u<-\log(\log(n))$ separately. Because of $ue^{-u}\le 1$ and $ue^{-u/2}\le 1$ for $u\ge 0$ and $\log(t)\ge 1$, by (3.63) we have

$$d_{TV}\big(\xi_t((u,\infty)),P_{e^{-u}}\big) \le (2^{d+1}+2)\frac{\log(t)}{\sqrt{t}}$$

for $u\ge 0$, which proves (3.64) for $u\ge 0$.

In the following let $u\in[-\log(\log(t)),0)$ be fixed. Since $Z_t(u)=Z'_{t,X}(u)-Z''_{t,X}(u)$ and the terms on the right-hand side are both non-negative, we obtain that

$$Z_t(u)_+ \le Z'_{t,X}(u) \quad\text{and}\quad Z_t(u)_- \le Z''_{t,X}(u).$$

Combining these inequalities and (3.4) in Theorem 3.1.1 with $m=1$ establishes

$$\begin{aligned}&|\mathbb{P}(T_t\le u)-\mathbb{P}(P_{e^{-u}}=0)| = |\mathbb{P}(\xi_t((u,\infty))=0)-\mathbb{P}(P_{e^{-u}}=0)| \\ &\le e^u\mathbb{E}[|Z_t(u)|] + \mathbb{E}[|Z_t(u)|\mathbf{1}\{\xi_t((u,\infty))-Z_t(u)_-=0\}] \\ &\le e^u\mathbb{E}[|Z_t(u)|] + \mathbb{E}[Z'_{t,X}(u)\mathbf{1}\{\xi_t((u,\infty))=0\}] + \mathbb{E}[Z''_{t,X}(u)].\end{aligned}$$

Moreover, by (3.66) and (3.67) we have

$$\mathbb{E}[Z'_{t,X}(u)] \le 2^d\frac{u+\log(t)}{e^{u/2}\sqrt{t}} \le 2^d\frac{\log(t)}{e^{u/2}\sqrt{t}}$$

and

$$\mathbb{E}[Z''_{t,X}(u)] \le \frac{u+\log(t)}{e^u t} \le \frac{(\log(t))^2}{t} \le \frac{\log(t)}{\sqrt{t}}.$$

Thus the identity $Z_t(u)=Z'_{t,X}(u)-Z''_{t,X}(u)$ with $Z'_{t,X}(u),Z''_{t,X}(u)\ge 0$ implies that

$$|\mathbb{P}(T_t\le u)-\mathbb{P}(P_{e^{-u}}=0)| \le (2^d+2)\frac{\log(t)}{\sqrt{t}} + \mathbb{E}[Z'_{t,X}(u)\mathbf{1}\{\xi_t((u,\infty))=0\}]. \tag{3.68}$$

For $x\in W$ we define

$$\xi_{t,x}((u,\infty)) = \sum_{z\in\eta_t\cap W\cap\mathbf{B}(x,4v_t)^c} \mathbf{1}\{h_t(z,\eta_t)>u\}.$$

Since, for any $x\in W$, $\mathbf{1}\{\xi_t((u,\infty))=0\}\le\mathbf{1}\{\xi_{t,x}((u,\infty))=0\}$ and $Z'_{t,x}(u)$ and $\mathbf{1}\{\xi_{t,x}((u,\infty))=0\}$ are independent, we have

$$\begin{aligned}\mathbb{E}[Z'_{t,X}(u)\mathbf{1}\{\xi_t((u,\infty))=0\}] &\le \int_W \mathbb{E}[Z'_{t,x}(u)\mathbf{1}\{\xi_{t,x}((u,\infty))=0\}]dx \\ &= \int_W \mathbb{E}[Z'_{t,x}(u)]\mathbb{P}(\xi_{t,x}((u,\infty))=0)dx.\end{aligned} \tag{3.69}$$

For $x \in W$, the Markov and the triangle inequalities, (3.63) and $e^{u/2}\sqrt{t} \geq 1$ imply that

$$\begin{aligned}
\mathbb{P}(\xi_{t,x}((u,\infty)) = 0) &\leq \mathbb{P}(\xi_t((u,\infty)) = 0) + \mathbb{P}\Big(\sum_{z \in \eta_t \cap \mathbf{B}(x,4v_t)} \mathbf{1}\{h_t(z,\eta_t) > u\} > 0 \Big) \\
&\leq 2^d \frac{\log(t)}{e^{u/2}\sqrt{t}} + \frac{\log(t)}{e^u t} + e^{-e^{-u}} + \mathbb{E}\Big[\sum_{z \in \eta_t \cap \mathbf{B}(x,4v_t)} \mathbf{1}\big\{h_t(z,\eta_t) > u\big\} \Big] \\
&\leq (2^d + 1)\frac{\log(t)}{e^{u/2}\sqrt{t}} + e^{-e^{-u}} + \mathbb{E}\Big[\sum_{z \in \eta_t \cap \mathbf{B}(x,4v_t)} \mathbf{1}\big\{h_t(z,\eta_t) > u\big\} \Big].
\end{aligned}$$

Similar arguments as used in (3.67) and $e^{u/2}\sqrt{t} \geq 1$ lead to

$$\mathbb{E}\Big[\sum_{z \in \eta_t \cap \mathbf{B}(x,4v_t)} \mathbf{1}\big\{h_t(z,\eta_t) > u\big\} \Big] \leq \frac{4^d(u + \log(t))}{e^u t} \leq \frac{4^d \log(t)}{e^{u/2}\sqrt{t}}.$$

Since $\log(t)e^u \geq 1$ and $\frac{\log(t)^2}{\sqrt{t}} \leq 4$ for $t > e^2$, we obtain

$$\frac{\log(t)}{e^{u/2}\sqrt{t}} \leq \frac{\log(t)^2 e^u}{e^{u/2}\sqrt{t}} \leq \frac{\log(t)^2}{\sqrt{t}} e^{u/2} \leq 4e^{u/2}.$$

Together with $\exp(-e^{-u} - u/2) \leq 1$, which follows from $u < 0$, we have shown

$$\mathbb{P}(\xi_{t,x}((u,\infty)) = 0) \leq 4(4^d + 2^d + 1)e^{u/2} + e^{u/2} \leq (4(4^d + 2^d + 1) + 1)e^{u/2}$$

so that, by (3.66) and (3.69),

$$\begin{aligned}
\mathbb{E}[Z'_{t,X}(u)\mathbf{1}\{\xi_t((u,\infty)) = 0\}] &\leq (4(4^d + 2^d + 1) + 1)e^{u/2} \frac{2^d \log(t)}{e^{u/2}\sqrt{t}} \\
&= (2^{d+2}(4^d + 2^d + 1) + 2^d)\frac{\log(t)}{\sqrt{t}}.
\end{aligned}$$

Combining this with (3.68) leads to

$$\begin{aligned}
\Big|\mathbb{P}(T_t \leq u) - e^{-e^{-u}}\Big| &\leq (2^{d+2}(4^d + 2^d + 1) + 2^d + 2^d + 2)\frac{\log(t)}{\sqrt{t}} \\
&\leq 2^{d+2}(4^d + 2^d + 2)\frac{\log(t)}{\sqrt{t}},
\end{aligned} \tag{3.70}$$

which establishes (3.64) for $u \in [-\log(\log(t)), 0)$.

Finally for $u < -\log(\log(t))$ we have

$$\mathbb{P}(T_t \leq u) \leq \mathbb{P}(T_t \leq -\log(\log(t))),$$

which by (3.70) and the triangle inequality is bounded by

$$2^{d+2}(4^d + 2^d + 2)\frac{\log(t)}{\sqrt{t}} + \frac{1}{t}.$$

Therefore elementary arguments lead to

$$\sup_{u < -\log(\log(t))} \big|\mathbb{P}(T_t \leq u) - e^{-e^{-u}}\big| \leq [2^{d+2}(4^d + 2^d + 2) + 1]\frac{\log(t)}{\sqrt{t}},$$

which concludes the proof of (3.64). □

Remark 3.3.17. *Note that the integral in the middle of* (3.66) *cannot be bounded with a better exponent for* t. *Indeed, using substitution, we can rewrite the integral as*

$$\frac{u+\log(t)}{k_d}\int_{\boldsymbol{B}(0,2)\cap\boldsymbol{B}(0,1)^c} e^{-(u+\log(t))\frac{\lambda_d(\mathbf{B}(y,1)\cap\mathbf{B}(0,1)^c)}{k_d}}\,dy.$$

For any sufficiently small $\varepsilon > 0$ *there exists a set* $A \subset \boldsymbol{B}(0,2)\cap\boldsymbol{B}(0,1)^c$ *with* $\lambda_d(A) > 0$ *such that the ratio in the exponent is at least* $(1+\varepsilon)/2$ *for all* $y \in A$. *This provides a lower bound of the order* $\log(t)t^{-(1+\varepsilon)/2}$.

Chapter 4

Multivariate Poisson and Poisson process approximations

This chapter is a slightly modified and adjusted version of the following preprint article jointly written with Riccardo Turin:

F. Pianoforte and R. Turin. Multivariate Poisson and Poisson process approximations with applications to Bernoulli sums and U-statistics. arXiv:2105.01599, 2021.

Abstract. In this chapter, we derive quantitative limit theorems for multivariate Poisson and Poisson process approximations. Employing the solution of Stein's equation for Poisson random variables, we obtain an explicit bound for the multivariate Poisson approximation of random vectors in the Wasserstein distance. The bound is then utilized in the context of point processes to provide a Poisson process approximation result in terms of a new metric called d_π defined herein, which is the supremum over all Wasserstein distances between random vectors obtained by evaluating the point processes on arbitrary collections of disjoint sets. As applications, the multivariate Poisson approximation of the sum of m-dependent Bernoulli random vectors, the Poisson process approximation of point processes of U-statistic structure and the Poisson process approximation of point processes with Papangelou intensity are considered. Our bounds in d_π are as good as those already available in the literature.

Acknowledgments. This research was supported by Swiss National Science Foundation (grant number 200021_175584). The authors would like to thank Chinmoy Bhattacharjee, Ilya Molchanov and Matthias Schulte for valuable comments.

4.1 Introduction and main results

Let $\mathbf{X} = (X_1, \dots, X_d)$ be an integrable random vector taking values in $\mathbb{N}_0^d$, $d \in \mathbb{N}$, and let $\mathbf{P} = (P_1, \dots, P_d)$ be a Poisson random vector, that is, a random vector with independent and Poisson distributed components. The first result of this chapter is an upper bound on the Wasserstein distance

$$d_W(\mathbf{X}, \mathbf{P}) = \sup_{g \in \mathrm{Lip}^d(1)} \left| \mathbb{E}[g(\mathbf{X})] - \mathbb{E}[g(\mathbf{P})] \right|$$

between $\mathbf{X}$ and $\mathbf{P}$, where $\mathrm{Lip}^d(1)$ denotes the set of Lipschitz functions $g : \mathbb{N}_0^d \to \mathbb{R}$ with Lipschitz constant bounded by 1 with respect to the metric induced by the 1-norm,

$|\mathbf{x}|_1 = \sum_{i=1}^d |x_i|$ for $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$.

The accuracy of the multivariate Poisson approximation has mostly been studied in terms of the total variation distance; among others we mention [1, 5, 6, 10, 25, 58, 59]. In contrast, we consider the Wasserstein distance. Note that, since the indicator functions defined on $\mathbb{N}_0^d$ are Lipschitz continuous, for random vectors in $\mathbb{N}_0^d$ the Wasserstein distance dominates the total variation distance, and it is not hard to find sequences that converge in total variation distance but not in Wasserstein distance. Our goal is to extend the approach developed in Chapter 3 for the Poisson approximation of random variables to the multivariate case.

Throughout this chapter, for any $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ and index $1 \leq j \leq d$, we denote by $x_{1:j}$ and $x_{j:d}$ the subvectors $(x_1, \dots, x_j)$ and $(x_j, \dots, x_d)$, respectively. Moreover, to simplify the notation, we use the convention $\mathbb{E}|Y| = \mathbb{E}[|Y|]$ for any random variable Y.

Theorem 4.1.1. *Let $\mathbf{X} = (X_1, \dots, X_d)$ be an integrable random vector with values in $\mathbb{N}_0^d$, $d \in \mathbb{N}$, and let $\mathbf{P} = (P_1, \dots, P_d)$ be a Poisson random vector with $\mathbb{E}[\mathbf{P}] = (\lambda_1, \dots, \lambda_d) \in [0, \infty)^d$. For $1 \leq i \leq d$, consider any random vector $\mathbf{Z}^{(i)} = (Z_1^{(i)}, \dots, Z_i^{(i)})$ in $\mathbb{Z}^i$ defined on the same probability space as $\mathbf{X}$, and define*

$$q_{m_{1:i}}^{(i)} = m_i \mathbb{P}\big(X_{1:i} = m_{1:i}\big) - \lambda_i \mathbb{P}\big(X_{1:i} + \mathbf{Z}^{(i)} = (m_{1:i-1}, m_i - 1)\big) \tag{4.1}$$

for $m_{1:i} \in \mathbb{N}_0^i$ with $m_i \neq 0$. Then,

$$d_W(\mathbf{X}, \mathbf{P}) \leq \sum_{i=1}^d \left(\lambda_i \mathbb{E}\big|Z_i^{(i)}\big| + 2\lambda_i \sum_{j=1}^{i-1} \mathbb{E}\big|Z_j^{(i)}\big| + \sum_{\substack{m_{1:i} \in \mathbb{N}_0^i \\ m_i \neq 0}} \left| q_{m_{1:i}}^{(i)} \right| \right). \tag{4.2}$$

For a random variable X, Equation (4.1) corresponds to the condition required in Theorem 3.1.3. There, sharper bounds on the Wasserstein distance for the case of random variables are shown. However, Theorem 4.1.1 tackles the case of random vectors instead of just considering random variables.

In order to give an interpretation of the hypothesis in Theorem 4.1.1, for a random vector $\mathbf{X} = (X_1, \dots, X_d)$ in $\mathbb{N}_0^d, d \in \mathbb{N}$, let us consider the family of random vectors

$$\mathbf{Y}^{(i)} = (X_{1:i-1}, X_i + 1) + \mathbf{Z}^{(i)}, \quad i = 1, \dots, d, \tag{4.3}$$

where $\mathbf{Z}^{(i)}, i = 1, \dots, d$, are defined as in Theorem 4.1.1. Under the additional condition $\mathbb{P}(X_{1:i} + \mathbf{Z}^{(i)} \in \mathbb{N}_0^i) = 1$, a sequence of real numbers $q_{m_{1:i}}^{(i)}, m_{1:i} \in \mathbb{N}_0^i$ with $m_i \neq 0$, satisfies Equation (4.1) if and only if

$$\mathbb{E}[X_i f(X_{1:i})] = \lambda_i \mathbb{E}[f(\mathbf{Y}^{(i)})] + \sum_{m_{1:i} \in \mathbb{N}_0^i,\, m_i \neq 0} q_{m_{1:i}}^{(i)} f(m_{1:i}) \tag{4.4}$$

for all functions $f : \mathbb{N}_0^i \to \mathbb{R}$ such that $\mathbb{E}\,|X_i f(X_{1:i})| < \infty$. When the elements $q_{m_{1:i}}^{(i)}$ are all zeros and $\mathbb{E}[X_i] = \lambda_i$, (4.4) becomes

$$\mathbb{E}[X_i f(X_{1:i})] = \mathbb{E}[X_i]\mathbb{E}[f(\mathbf{Y}^{(i)})]. \tag{4.5}$$

In this case, by taking the sum over all $m_{1:i} \in \mathbb{N}_0^d$ with $m_i \neq 0$ in (4.1), we obtain that the condition $\mathbb{P}(X_{1:i} + \mathbf{Z}^{(i)} \in \mathbb{N}_0^i) = 1$ is always satisfied. Recall that, for a random variable

$X \geq 0$ with $\mathbb{E}[X] > 0$, a random variable X^s defined on the same probability space as X is a size bias coupling of X if it satisfies

$$\mathbb{E}[Xf(X)] = \mathbb{E}[X]\mathbb{E}[f(X^s)] \tag{4.6}$$

for all measurable $f : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}\,|Xf(X)| < \infty$. Therefore, if the $q^{(i)}_{m_{1:i}}$ are all zeros for any $i = 1, \dots, d$ and $\mathbb{E}[\mathbf{X}] = (\lambda_1, \dots, \lambda_d)$, the random vectors $\mathbf{Y}^{(i)}, i = 1, \dots, d$, can be seen as a size bias coupling of $\mathbf{X}$, as they are defined on the same probability space as $\mathbf{X}$ and satisfy (4.5), which corresponds to (4.6) in the one-dimensional case. Note that this suggests a definition of size bias coupling of random vectors that is slightly different from the one introduced by Definition 2.4.5. Following this interpretation, when $\mathbb{E}[\mathbf{X}] = (\lambda_1, \dots, \lambda_d)$ and the random vectors $\mathbf{Z}^{(i)}$ are chosen such that the $q^{(i)}_{m_{1:i}}$ are not zero, we can think of the random vectors $\mathbf{Y}^{(i)}$ defined by (4.3) as an approximate size bias coupling of $\mathbf{X}$, where instead of assuming that $\mathbf{Y}^{(i)}$ satisfies (4.5) exactly, we allow error terms $q^{(i)}_{m_{1:i}}$ and obtain (4.4). This is an important advantage of Theorem 4.1.1, since one does not need to find an exact size bias coupling (in the sense of (4.5)), it only matters that the error terms $q^{(i)}_{m_{1:i}}$ are sufficiently small and that the random vectors $\mathbf{Z}^{(i)}$ are the null vector with high probability.

The second main contribution of this chapter concerns Poisson process approximation of point processes with finite intensity measure. Let $(\mathbb{X}, \mathcal{X})$ be a measurable space and consider a point process ξ and a Poisson process η on $\mathbb{X}$ with finite intensity measure. For any choice of subsets $A_1, \dots, A_d \in \mathcal{X}$, the random vectors $(\xi(A_1), \dots, \xi(A_d))$ and $(\eta(A_1), \dots, \eta(A_d))$ take values in $\mathbb{N}_0^d$ (almost surely). Thus, Theorem 4.1.1 provides bounds on the Wasserstein distance

$$d_W((\xi(A_1), \dots, \xi(A_d)), (\eta(A_1), \dots, \eta(A_d)))$$

for all $A_1, \dots, A_d \in \mathcal{X}$ and $d \in \mathbb{N}$. Then, by taking the supremum of over all arbitrary collections $(A_1, \dots, A_d)$ of disjoint sets, these bounds permit the comparison of the distributions of ξ and η.

Definition 4.1.2. *Let ξ and ζ be point processes on $\mathbb{X}$ with finite intensity measure. The distance d_π between ξ and ζ is defined as*

$$d_\pi(\xi, \zeta) = \sup_{(A_1,\dots,A_d)\in\mathcal{X}^d_{\text{disj}},\, d\in\mathbb{N}} d_W\big((\xi(A_1), \dots, \xi(A_d)), (\zeta(A_1), \dots, \zeta(A_d))\big),$$

where

$$\mathcal{X}^d_{\text{disj}} = \{(A_1, \dots, A_d) \in \mathcal{X}^d \,:\, A_i \cap A_j = \emptyset, i \neq j\}.$$

That d_π is a probability distance between the distributions of point processes with finite intensity measure follows immediately from its definition and Proposition 2.2.2.

To the best of our knowledge, this is the first time the distance d_π is defined and employed in Poisson process approximation. We believe that it is possible to extend d_π to larger classes of point processes by restricting $\mathcal{X}^d_{\text{disj}}$ to suitable families of sets. For example, for locally finite point processes on a locally compact second countable Hausdorff space, we may define the distance d_π by replacing $\mathcal{X}^d_{\text{disj}}$ with the family of d-tuples of disjoint and relatively compact Borel sets. However, this falls out of the scope of this chapter. Let us now state our main theoretical result on Poisson process approximation.

Theorem 4.1.3. *Let ξ be a point process on $\mathbb{X}$ with finite intensity measure, and let η be a Poisson process on $\mathbb{X}$ with finite intensity measure λ. For any i-tuple $(A_1,\dots,A_i)\in\mathcal{X}^i_{\text{disj}}$ with $i\in\mathbb{N}$, consider a random vector $\mathbf{Z}^{A_{1:i}}=(Z_1^{A_{1:i}},\dots,Z_i^{A_{1:i}})$ defined on the same probability space as ξ with values in $\mathbb{Z}^i$, and define*

$$\begin{aligned} q^{A_{1:i}}_{m_{1:i}} = m_i\mathbb{P}\big((\xi(A_1),\dots,\xi(A_i)) = m_{1:i}\big) \\ - \lambda(A_i)\mathbb{P}\big((\xi(A_1),\dots,\xi(A_i)) + \mathbf{Z}^{A_{1:i}} = (m_{1:i-1}, m_i - 1)\big) \end{aligned} \tag{4.7}$$

for $m_{1:i}\in\mathbb{N}_0^i$ with $m_i\neq 0$. Then,

$$d_\pi(\xi,\eta)\le \sup_{(A_1,\dots,A_d)\in\mathcal{X}^d_{\text{disj}},d\in\mathbb{N}} \sum_{i=1}^d \left(\sum_{\substack{m_{1:i}\in\mathbb{N}_0^i\\ m_i\neq 0}} \big|q^{A_{1:i}}_{m_{1:i}}\big| + 2\lambda(A_i)\sum_{j=1}^i \mathbb{E}\big|Z_j^{A_{1:i}}\big| \right).$$

The Poisson process approximation has mostly been studied in terms of the total variation distance in the literature; see e.g. [2, 5, 7, 13, 17, 63, 64] and references therein. In contrast, [22, 23] deal with Poisson process approximation using the Kantorovich–Rubinstein distance. In Proposition 4.2.5, we establish that the total variation distance

$$d_{TV}(\xi,\zeta)=\sup_{B\in\mathcal{N}_{\mathbb{X}}}|\mathbb{P}(\xi\in B)-\mathbb{P}(\zeta\in B)|$$

between two point processes ξ and ζ on $\mathbb{X}$ with finite intensity measure is bounded from above by $d_\pi(\xi,\zeta)$. Moreover, since $d_\pi(\xi,\zeta)\ge|\mathbb{E}[\xi(\mathbb{X})]-\mathbb{E}[\zeta(\mathbb{X})]|$, Example 2.2 in [22] provides a sequence of point processes $(\zeta_n)_{n\in\mathbb{N}}$ that converges in total variation distance to a point process ζ even though $d_\pi(\zeta_n,\zeta)\to\infty$ as n goes to infinity. This shows that d_π is stronger than d_{TV} in the sense that convergence in d_π implies convergence in total variation distance, but not vice versa. The Kantorovich-Rubinstein distance between two point processes ξ and ζ is defined as the optimal transportation cost between their distributions, when the cost function is the total variation distance between measures; see [22, Equation 2.5]. When the configuration space $\mathbb{X}$ is a locally compact second countable Hausdorff space (lcscH), which is indeed the case considered in [22] and [23], the Kantorovich duality theorem ([70, Theorem 5.10]) yields an equivalent definition for this metric:

$$d_{KR}(\xi,\zeta)=\sup|\mathbb{E}[h(\xi)]-\mathbb{E}[h(\zeta)]|$$

where the supremum runs over all measurable functions $h:\mathbf{N}_{\mathbb{X}}\to\mathbb{R}$ that are 1-Lipschitz with respect to the total variation distance between measures and make $h(\xi)$ and $h(\zeta)$ integrable. For a lcscH space $\mathbb{X}$, we prove in Lemma 4.2.6 that $d_\pi\le 2d_{KR}$. The constant 2 in this inequality cannot be improved, as shown by the following simple example: let $\mathbb{X}=\{a,b\}$ with $\mathcal{X}=\{\emptyset,\{a\},\{b\},\mathbb{X}\}$, and let δ_a and δ_b be deterministic point processes corresponding to the Dirac measures centered at a and b, respectively. Since the function $g:(x_1,x_2)\mapsto x_1-x_2$ is 1-Lipschitz, it follows

$$d_\pi(\delta_a,\delta_b)\ge|\,g(\delta_a(\{a\}),\delta_a(\{b\}))-g(\delta_b(\{a\}),\delta_b(\{b\}))\,|=2.$$

On the other hand, d_{KR} is bounded by the expected total variation distance between the two counting measures, thus $d_{KR}(\delta_a,\delta_b)\le 1$. Hence, in this case $d_\pi(\delta_a,\delta_b)=2d_{KR}(\delta_a,\delta_b)$.

It remains an open problem whether the distances d_π and d_{KR} are equivalent or not. It is worth mentioning that our general result, Theorem 4.1.3, permits to study the

Poisson process approximation in the metric d_π of point processes on any measurable space. Then, Theorem 4.1.3 can be used to obtain approximation results for point processes also when the notion of weak convergence is not defined.

To demonstrate the versatility of our general main results, we apply them to several examples. In Subsection 4.3.1, we approximate the sum of Bernoulli random vectors by a Poisson random vector. This problem has mainly been studied in terms of the total variation distance and under the assumption that the Bernoulli random vectors are independent (see e.g. [57]). We derive an explicit approximation result in the Wasserstein distance for the more general case of m-dependent Bernoulli random vectors.

In Subsections 4.3.2 and 4.3.3, we apply Theorem 4.1.3 to obtain explicit Poisson process approximation results for point processes with Papangelou intensity and point processes of Poisson U-statistic structure. The latter are point processes that, once evaluated on a measurable set, become Poisson U-statistics. Analogous results were already proven for the Kantorovich-Rubinstein distance in [23, Theorem 3.7] and [22, Theorem 3.1], under the additional condition that the configuration space $\mathbb{X}$ is lcscH. It is interesting to note that the proof of our result for point processes with Papangelou intensity employs Theorem 4.1.3 with $\mathbf{Z}^{A_{1:i}}$ set to zero for all i, while for point processes of U-statistic structure, we find $\mathbf{Z}^{A_{1:i}}$ such that Equation (4.7) in Theorem 4.1.3 is satisfied with $q_{m_{1:i}}^{A_{1:i}} \equiv 0$ for all collections of disjoint sets.

The proof of Theorem 4.1.1 is based on the Chen-Stein method applied to each component of the random vectors and the coupling in (4.1). For the proof of Theorem 4.1.3, we mimic the approach used to prove [1, Theorem 2], as we derive the process bound as a consequence of the d-dimensional bound.

Before we discuss the applications in Section 4.3, we prove our main results in the next section.

4.2 Proofs of the results of Section 4.1

Throughout this section, $\mathbf{X} = (X_1, \dots, X_d)$ is an integrable random vector with values in $\mathbb{N}_0^d$ and $\mathbf{P} = (P_1, \dots, P_d)$ is a Poisson random vector with mean $\mathbb{E}[\mathbf{P}] = (\lambda_1, \dots, \lambda_d) \in [0, \infty)^d$. Without loss of generality we assume that $\mathbf{X}$ and $\mathbf{P}$ are independent and defined on the same probability space. We denote by $\mathrm{Lip}^d(1)$ the collection of Lipschitz functions $g : \mathbb{N}_0^d \to \mathbb{R}$ with respect to the metric induced by the 1-norm and Lipschitz constant bounded by 1, that is

$$|g(\mathbf{x}) - g(\mathbf{y})| \leq |\mathbf{x} - \mathbf{y}|_1 = \sum_{i=1}^{d} |x_i - y_i|, \quad \mathbf{x}, \mathbf{y} \in \mathbb{N}_0^d.$$

Clearly, this family of functions contains the 1-Lipschitz functions with respect to the metric induced by the Euclidean norm. For $d = 1$, we use the convention $\mathrm{Lip}(1) = \mathrm{Lip}^1(1)$.

From now on, for any $g \in \mathrm{Lip}(1)$, we denote by $\widehat{g}^{(\lambda)}$ the solution of the Stein equation

$$\lambda \widehat{g}^{(\lambda)}(i+1) - i\widehat{g}^{(\lambda)}(i) = g(i) - \mathbb{E}[g(P_\lambda)], \quad i \in \mathbb{N}_0, \tag{4.8}$$

such that $\widehat{g}^{(\lambda)}(0) = 0$, where P_λ is a Poisson random variable with mean $\lambda \geq 0$. From the inequalities (2.4) in Section 2.3, we obtain the following result for the Stein factors (Note that the case $\lambda = 0$ is trivial).

Lemma 4.2.1. *For any $\lambda \geq 0$ and $g \in \mathrm{Lip}(1)$, let $\widehat{g}^{(\lambda)}$ be the solution of the Stein equation* (4.8) *with initial condition $\widehat{g}^{(\lambda)}(0) = 0$. Then,*

$$\sup_{i\in\mathbb{N}_0} \left|\widehat{g}^{(\lambda)}(i)\right| \leq 1 \quad \text{and} \quad \sup_{i\in\mathbb{N}_0} \left|\widehat{g}^{(\lambda)}(i+1) - \widehat{g}^{(\lambda)}(i)\right| \leq 1. \tag{4.9}$$

Recall that, for any $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ and index $1 \leq j \leq d$, we write $x_{1:j}$ and $x_{j:d}$ for the subvectors $(x_1, \dots, x_j)$ and $(x_j, \dots, x_d)$, respectively. For $g \in \mathrm{Lip}^d(1)$, we denote by $\widehat{g}^{(\lambda)}_{x_{1:i-1}|x_{i+1:d}}$ the solution of (4.8) for the Lipschitz function $g(x_{1:i-1}, \cdot\,, x_{i+1:d})$ with fixed $x_{1:i-1} \in \mathbb{N}_0^{i-1}$ and $x_{i+1:d} \in \mathbb{N}_0^{d-i}$. Since $\widehat{g}^{(\lambda)}$ takes vectors from $\mathbb{N}_0^d$ as input, we do not need to worry about measurability issues. The following proposition is the first building block for the proof of Theorem 4.1.1.

Proposition 4.2.2. *For any $g \in \mathrm{Lip}^d(1)$,*

$$\mathbb{E}[g(\mathbf{P}) - g(\mathbf{X})] = \sum_{i=1}^{d} \mathbb{E}\left[X_i \widehat{g}^{(\lambda_i)}_{X_{1:i-1}|P_{i+1:d}}(X_i) - \lambda_i \widehat{g}^{(\lambda_i)}_{X_{1:i-1}|P_{i+1:d}}(X_i + 1)\right].$$

Proof of Proposition 4.2.2. First, observe that

$$\mathbb{E}\left[g(\mathbf{P}) - g(\mathbf{X})\right] = \sum_{i=1}^{d} \mathbb{E}\left[g(X_{1:i-1}, P_{i:d}) - g(X_{1:i}, P_{i+1:d})\right]. \tag{4.10}$$

The independence of P_i from $P_{i+1:d}$ and $X_{1:i}$ implies

$$\mathbb{E}\big[g(X_{1:i-1}, P_{i:d}) - g(X_{1:i}, P_{i+1:d})\big] = \mathbb{E}\big[\mathbb{E}^{P_i}[g(X_{1:i-1}, P_{i:d})] - g(X_{1:i}, P_{i+1:d})\big],$$

where $\mathbb{E}^{P_i}$ denotes the expectation with respect to the random variable P_i. From the definition of $\widehat{g}^{(\lambda_i)}_{x_{1:i-1}|x_{i+1:d}}$ with $x_{1:i-1} = X_{i:i-1}$ and $x_{i+1:d} = P_{i+1:d}$, it follows

$$\mathbb{E}^{P_i}[g(X_{1:i-1}, P_{i:d})] - g(X_{1:i}, P_{i+1:d}) = X_i \widehat{g}^{(\lambda_i)}_{X_{1:i-1}|P_{i+1:d}}(X_i) - \lambda_i \widehat{g}^{(\lambda_i)}_{X_{1:i-1}|P_{i+1:d}}(X_i + 1)$$

for all $i = 1, \dots, d$. Together with (4.10), this leads to the desired conclusion. □

Proof of Theorem 4.1.1. In view of Proposition 4.2.2, it suffices to bound

$$\left|\mathbb{E}\left[X_i \widehat{g}^{(\lambda_i)}_{X_{1:i-1}|P_{i+1:d}}(X_i) - \lambda_i \widehat{g}^{(\lambda_i)}_{X_{1:i-1}|P_{i+1:d}}(X_i + 1)\right]\right|, \quad i = 1, \dots, d.$$

For the remaining of the proof, the index i is fixed and we omit the superscript (i) in $Z^{(i)}_{i:d}$ and $q^{(i)}_{m_{1:i}}$. Define the function

$$h(X_{1:i}) = \mathbb{E}\left[\widehat{g}^{(\lambda_i)}_{X_{1:i-1}|P_{i+1:d}}(X_i) \,\middle|\, X_{1:i}\right],$$

where $\mathbb{E}[\,\cdot\,|\,Y]$ denotes the conditional expectation with respect to a random element Y. With the convention $\widehat{g}^{(\lambda_i)}_{m_{1:i-1}|m_{i+1:d}}(m_i) = 0$ if $m_{1:d} \notin \mathbb{N}_0^d$ or $m_i = 0$, it follows from (4.1) that

$$\begin{aligned}
\mathbb{E}\left[X_i \widehat{g}^{(\lambda_i)}_{X_{1:i-1}|P_{i+1:d}}(X_i)\right] &= \mathbb{E}[X_i h(X_{1:i})] = \sum_{m_{1:i}\in\mathbb{N}_0^i} m_i h(m_{1:i}) \mathbb{P}(X_{1:i} = m_{1:i}) \\
&= \sum_{\substack{m_{1:i}\in\mathbb{N}_0^i \\ m_i\neq 0}} h(m_{1:i}) q_{m_{1:i}} + \lambda_i \sum_{\substack{m_{1:i}\in\mathbb{N}_0^i \\ m_i\neq 0}} h(m_{1:i}) \mathbb{P}\left(X_{1:i} + Z_{1:i} = (m_{1:i-1}, m_i - 1)\right) \\
&= \sum_{\substack{m_{1:i}\in\mathbb{N}_0^i \\ m_i\neq 0}} h(m_{1:i}) q_{m_{1:i}} + \lambda_i \mathbb{E}\left[\widehat{g}^{(\lambda_i)}_{X_{1:i-1}+Z_{1:i-1}|P_{i+1:d}}(X_i + Z_i + 1)\right].
\end{aligned}$$

Since $|h(X_{1:i})| \le 1$ by (4.9), the triangle inequality establishes

$$\left|\mathbb{E}\left[X_i \widehat{g}^{(\lambda_i)}_{X_{1:i-1}|P_{i+1:d}}(X_i) - \lambda_i \widehat{g}^{(\lambda_i)}_{X_{1:i-1}|P_{i+1:d}}(X_i+1)\right]\right| \le \sum_{\substack{m_{1:i}\in\mathbb{N}_0^i \\ m_i\neq 0}} |q_{m_{1:i}}| + \lambda_i(H_1+H_2), \tag{4.11}$$

with

$$H_1 = \left|\mathbb{E}\left[\widehat{g}^{(\lambda_i)}_{X_{1:i-1}+Z_{1:i-1}|P_{i+1:d}}(X_i+Z_i+1) - \widehat{g}^{(\lambda_i)}_{X_{1:i-1}+Z_{1:i-1}|P_{i+1:d}}(X_i+1)\right]\right|$$

and

$$H_2 = \left|\mathbb{E}\left[\widehat{g}^{(\lambda_i)}_{X_{1:i-1}+Z_{1:i-1}|P_{i+1:d}}(X_i+1) - \widehat{g}^{(\lambda_i)}_{X_{1:i-1}|P_{i+1:d}}(X_i+1)\right]\right|.$$

The inequalities in (4.9) guarantee

$$H_1 \le \mathbb{E}|Z_i| \quad \text{and} \quad H_2 \le 2\mathbb{P}(Z_{1:i-1}\neq 0) \le \sum_{j=1}^{i-1} 2\mathbb{P}(Z_j\neq 0) \le 2\sum_{j=1}^{i-1}\mathbb{E}|Z_j|.$$

Combining (4.11) with the bounds for H_1 and H_2, and summing over $i = 1, \dots, d$ concludes the proof. □

Remark 4.2.3. *It follows directly from the previous proof that the bound* (4.2) *in Theorem 4.1.1 can be improved in the following way:*

$$d_W(\mathbf{X},\mathbf{P}) \le \sum_{i=1}^{d}\left(\lambda_i \mathbb{E}\big|Z_i^{(i)}\big| + 2\lambda_i \mathbb{P}\big(Z_{1:i-1}^{(i)}\neq 0\big) + \sum_{\substack{m_{1:i}\in\mathbb{N}_0^i \\ m_i\neq 0}} \left|q^{(i)}_{m_{1:i}}\right|\right).$$

Next, we derive Theorem 4.1.3 from Theorem 4.1.1.

Proof of Theorem 4.1.3. Let $d\in\mathbb{N}$ and $\mathbf{A} = (A_1,\dots,A_d)\in\mathcal{X}^d_{\text{disj}}$. Define

$$\mathbf{X}^{\mathbf{A}} = (\xi(A_1),\dots,\xi(A_d)) \quad \text{and} \quad \mathbf{P}^{\mathbf{A}} = (\eta(A_1),\dots,\eta(A_d)),$$

where $\mathbf{P}^{\mathbf{A}}$ is a Poisson random vector with mean $\mathbb{E}[\mathbf{P}^{\mathbf{A}}] = (\lambda(A_1),\dots,\lambda(A_d))$. By Theorem 4.1.1 with $\mathbf{Z}^{(i)} = \mathbf{Z}^{A_{1:i}}$, we obtain

$$d_W(\mathbf{X}^{\mathbf{A}},\mathbf{P}^{\mathbf{A}}) \le \sum_{i=1}^{d}\left(\sum_{\substack{m_{1:i}\in\mathbb{N}_0^i \\ m_i\neq 0}} \left|q^{A_{1:i}}_{m_{1:i}}\right| + 2\lambda(A_i)\sum_{j=1}^{i}\mathbb{E}|Z_j^{A_{1:i}}|\right).$$

Taking the supremum over all d-tuples of disjoint measurable sets concludes the proof. □

Let us now prove that the total variation distance is dominated by d_π. Recall that the total variation distance between two point processes ξ and ζ on $\mathbb{X}$ is defined as

$$d_{TV}(\xi,\zeta) = \sup_{B\in\mathcal{N}_{\mathbb{X}}} |\mathbb{P}(\xi\in B) - \mathbb{P}(\zeta\in B)|. \tag{4.12}$$

The result is obtained by a monotone class Theorem, [41, Theorem 1.3], which is stated hereafter as a lemma. A monotone class $\mathcal{A}$ is a collection of sets closed under monotone limits, that is, for any $A_1, A_2, \dots \in \mathcal{A}$ with $A_n \uparrow A$ or $A_n \downarrow A$, then $A\in\mathcal{A}$.

Lemma 4.2.4. *Let U be a set and let $\mathcal{U}$ be an algebra of subsets of U. Then, the monotone class generated by $\mathcal{U}$ coincides with the σ-field generated by $\mathcal{U}$.*

Proposition 4.2.5. *Let ξ and ζ be two point processes on $\mathbb{X}$ with finite intensity measure. Then,*

$$d_{TV}(\xi, \zeta) \leq d_\pi(\xi, \zeta).$$

Proof. Let us first introduce the set of finite counting measures

$$\mathbf{N}_{\mathbb{X}}^{<\infty} = \{\nu \in \mathbf{N}_{\mathbb{X}} : \nu(\mathbb{X}) < \infty\},$$

with the trace σ-field

$$\mathcal{N}_{\mathbb{X}}^{<\infty} = \{B \cap \mathbf{N}_{\mathbb{X}}^{<\infty} : B \in \mathcal{N}_{\mathbb{X}}\}.$$

As we are dealing with finite point processes, the total variation distance is equivalently obtained if $\mathcal{N}_X$ is replaced by $\mathcal{N}_{\mathbb{X}}^{<\infty}$ in (4.12):

$$d_{TV}(\xi, \zeta) = \sup_{B \in \mathcal{N}_{\mathbb{X}}^{<\infty}} |\mathbb{P}(\xi \in B) - \mathbb{P}(\zeta \in B)|.$$

Let $\mathcal{P}(\mathbb{N}_0^d)$ denote the power set of $\mathbb{N}_0^d$, that is, the collection of all subsets of $\mathbb{N}_0^d$. For any $d \in \mathbb{N}$ and $M \in \mathcal{P}(\mathbb{N}_0^d)$ note that $\mathbf{1}_M(\cdot) \in \mathrm{Lip}^d(1)$, therefore

$$d_\pi(\xi, \zeta) \geq \sup_{U \in \mathcal{U}} |\mathbb{P}(\xi \in U) - \mathbb{P}(\zeta \in U)|, \tag{4.13}$$

with

$$\mathcal{U} = \Big\{ \big\{ \nu \in \mathbf{N}_{\mathbb{X}}^{<\infty} : (\nu(A_1), \dots, \nu(A_d)) \in M \big\} : d \in \mathbb{N},\, A \in \mathcal{X}_{\mathrm{disj}}^d,\, M \in \mathcal{P}(\mathbb{N}_0^d) \Big\} \subset \mathcal{N}_{\mathbb{X}}^{<\infty}.$$

It can be easily verified that $\mathcal{U}$ is an algebra and $\sigma(\mathcal{U}) = \mathcal{N}_{\mathbb{X}}^{<\infty}$. Moreover, by (4.13), $\mathcal{U}$ is a subset of the monotone class

$$\big\{ U \in \mathcal{N}_{\mathbb{X}}^{<\infty} : |\mathbb{P}(\xi \in U) - \mathbb{P}(\zeta \in U)| \leq d_\pi(\xi, \zeta) \big\}.$$

Lemma 4.2.4 concludes the proof. □

In the last part of this section, we show that the distance d_π is dominated by $2d_{KR}$ when the underlying space is locally compact second countable Hausdorff (lcscH). Recall that, a topological space is second countable if its topology has a countable basis, and it is locally compact if every point has an open neighborhood whose topological closure is compact. Suppose that $\mathbb{X}$ is lcscH with Borel σ-field $\mathcal{X}$. Recall that the Kantorovich-Rubinstein distance between two point processes ξ and ζ on $\mathbb{X}$ with finite intensity measure is given by

$$d_{KR}(\xi, \zeta) = \sup_{h \in \mathcal{L}(1)} |\mathbb{E}[h(\xi)] - \mathbb{E}[h(\zeta)]|,$$

where $\mathcal{L}(1)$ is the set of all measurable functions $h : \mathbf{N}_{\mathbb{X}} \to \mathbb{R}$ that are Lipschitz continuous with respect to the total variation distance between measures

$$d_{TV,\mathbf{N}_{\mathbb{X}}}(\mu, \nu) = \sup_{\substack{A \in \mathcal{X}, \\ \mu(A), \nu(A) < \infty}} |\mu(A) - \nu(A)|, \quad \mu, \nu \in \mathbf{N}_{\mathbb{X}},$$

and with Lipschitz constant bounded by 1. Since ξ and ζ take values in $\mathbf{N}_{\mathbb{X}}^{<\infty}$, we can consider h to be defined on $\mathbf{N}_{\mathbb{X}}^{<\infty}$ by [43, Theorem 1].

Lemma 4.2.6. *Let ξ and ζ be two point processes with finite intensity measure on a lcscH space $\mathbb{X}$ with Borel σ-field $\mathcal{X}$. Then,*

$$d_\pi(\xi,\zeta) \le 2d_{KR}(\xi,\zeta).$$

Proof. For $g \in \mathrm{Lip}^d(1)$ and disjoint sets $A_1,\dots,A_d \in \mathcal{X}, d \in \mathbb{N}$, define $h : \mathbf{N}_{\mathbb{X}}^{<\infty} \to \mathbb{R}$ by $h(\nu) = g(\nu(A_1),\dots,\nu(A_d))$. For finite point configurations ν_1 and ν_2, we obtain

$$\begin{aligned}|h(\nu_1) - h(\nu_2)| &\le |g(\nu_1(A_1),\dots,\nu_1(A_d)) - g(\nu_2(A_1),\dots,\nu_2(A_d))| \\ &\le \sum_{i=1}^{d} |\nu_1(A_i) - \nu_2(A_i)| \le 2d_{TV,\mathbf{N}_{\mathbb{X}}}(\nu_1,\nu_2).\end{aligned}$$

This implies $h/2 \in \mathcal{L}(1)$. Hence $|\mathbb{E}[h(\xi)] - \mathbb{E}[h(\zeta)]| \le 2d_{KR}(\xi,\zeta)$. □

4.3 Applications

In this section, we discuss some applications of Theorem 4.1.1 and Theorem 4.1.3. We study the multivariate Poisson approximation of the sum of m-dependent Bernoulli random vectors, and we prove the analogues of [23, Theorem 3.7] and [22, Theorem 3.1] for the metric d_π in a slightly more general set up.

4.3.1 Sum of m-dependent Bernoulli random vectors

In this subsection, we consider a finite family of Bernoulli random vectors $\mathbf{Y}^{(1)},\dots,\mathbf{Y}^{(n)}$ and investigate the multivariate Poisson approximation of $\mathbf{X} = \sum_{r=1}^{n} \mathbf{Y}^{(r)}$ in the Wasserstein distance. If the Bernoulli random vectors are i.i.d., then $\mathbf{X}$ has the so called multinomial distribution. The multivariate Poisson approximation of the multinomial distribution, and more generally of the sum of independent Bernoulli random vectors, has already been tackled by many authors in terms of the total variation distance. Among others, we refer the reader to [6, 24, 57, 59] and the survey [46]. Unlike the mentioned papers, we assume that $\mathbf{Y}^{(1)},\dots,\mathbf{Y}^{(n)}$ are m-dependent. Note that the case of sums of 1-dependent random vectors has recently been treated in [25] using metrics that are weaker than the total variation distance. To the best of our knowledge, this is the first time the Poisson approximation of the sum of m-dependent Bernoulli random vectors is investigated using the Wasserstein distance.

More precisely, for $n \in \mathbb{N}$, let $\mathbf{Y}^{(1)},\dots,\mathbf{Y}^{(n)}$ be Bernoulli random vectors with distributions given by

$$\begin{aligned}&\mathbb{P}(\mathbf{Y}^{(r)} = \mathbf{e}_j) = p_{r,j} \in [0,1], \quad r = 1,\dots,n, \quad j = 1,\dots,d, \\ &\mathbb{P}(\mathbf{Y}^{(r)} = \mathbf{0}) = 1 - \sum_{j=1}^{d} p_{r,j} \in [0,1], \quad r = 1,\dots,n,\end{aligned} \tag{4.14}$$

where $\mathbf{e}_j$ denotes the vector with entry 1 at position j and entry 0 otherwise. Assume that $\mathbf{Y}^{(1)},\dots,\mathbf{Y}^{(n)}$ are m-dependent for a given fixed $m \in \mathbb{N}_0$. This means that for any two subsets S and T of $\{1,\dots,n\}$ such that $\min(S) - \max(T) > m$, the collections $(\mathbf{Y}^{(s)})_{s\in S}$ and $(\mathbf{Y}^{(t)})_{t\in T}$ are independent. Define the random vector $\mathbf{X} = (X_1,\dots,X_d)$ as

$$\mathbf{X} = \sum_{r=1}^{n} \mathbf{Y}^{(r)}. \tag{4.15}$$

Note that if $\mathbf{Y}^{(r)}, r = 1, \dots, n$, are i.i.d., then $m = 0$ and $\mathbf{X}$ has the multinomial distribution. The mean vector of $\mathbf{X}$ is $\mathbb{E}[\mathbf{X}] = (\lambda_1, \dots, \lambda_d)$ with

$$\lambda_j = \sum_{r=1}^{n} p_{r,j}, \quad j = 1, \dots, d. \tag{4.16}$$

For $k = 1, \dots, n$, let $Q(k)$ be the quantity given by

$$Q(k) = \max_{\substack{r \in \{1,\dots,n\} : 1 \le |k-r| \le m \\ i,j=1,\dots,d}} \mathbb{E}\big[\mathbf{1}\{\mathbf{Y}^{(k)} = \mathbf{e}_i\}\mathbf{1}\{\mathbf{Y}^{(r)} = \mathbf{e}_j\}\big].$$

We now state the main result of this subsection.

Theorem 4.3.1. *Let* $\mathbf{X}$ *be as in* (4.15), *and let* $\mathbf{P} = (P_1, \dots, P_d)$ *be a Poisson random vector with mean* $\mathbb{E}[\mathbf{P}] = (\lambda_1, \dots, \lambda_d)$ *given by* (4.16). *Then,*

$$d_W(\mathbf{X}, \mathbf{P}) \le \sum_{k=1}^{n} \sum_{i=1}^{d} \Bigg[\sum_{\substack{r=1,\dots,n, \\ |r-k| \le m}} p_{r,i} + 2 \sum_{j=1}^{i-1} \sum_{\substack{r=1,\dots,n, \\ |r-k| \le m}} p_{r,j} \Bigg] p_{k,i} + 2d(d+1)m \sum_{k=1}^{n} Q(k).$$

The proof of Theorem 4.3.1 is obtained by applying Theorem 4.1.1. In the one-dimensional case, Equation (4.1) corresponds to the condition required in Theorem 3.1.3, which establishes better Poisson approximation results than Theorem 4.1.1. Then, for the sum of dependent Bernoulli random variables, a sharper bound for the Wasserstein distance than one in Theorem 4.3.1 can be derived from the inequality (3.11) in Theorem 3.1.3, while for the total variation distance, a better bound can be deduced from the inequality (3.9) in Theorem 3.1.3, [1, Theorem 1] and [68, Theorem 1]. As a consequence of Theorem 4.3.1, we obtain the following result for the sum of independent Bernoulli random vectors.

Corollary 4.3.2. *For* $n \in \mathbb{N}$, *let* $\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(n)}$ *be independent Bernoulli random vectors with distribution given by* (4.14), *and let* $\mathbf{X}$ *be the random vector defined by* (4.15). *Let* $\mathbf{P} = (P_1, \dots, P_d)$ *be a Poisson random vector with mean* $\mathbb{E}[\mathbf{P}] = (\lambda_1, \dots, \lambda_d)$ *given by* (4.16). *Then,*

$$d_W(\mathbf{X}, \mathbf{P}) \le \sum_{k=1}^{n} \Bigg[\sum_{i=1}^{d} p_{k,i} \Bigg]^2.$$

A sharper bound for the total variation distance than the one obtained by Corollary 4.3.2 is established in e.g. [57, Theorem 1].

Proof of Theorem 4.3.1. Without loss of generality we may assume that $\lambda_1, \dots, \lambda_d > 0$. Define the random vectors

$$\mathbf{W}^{(k)} = \big(W_1^{(k)}, \dots, W_d^{(k)}\big) = \sum_{\substack{r=1,\dots,n, \\ 1 \le |r-k| \le m}} \mathbf{Y}^{(r)},$$

$$\mathbf{X}^{(k)} = \big(X_1^{(k)}, \dots, X_d^{(k)}\big) = \mathbf{X} - \mathbf{Y}^{(k)} - \mathbf{W}^{(k)},$$

for $k = 1, \dots, n$. Let us fix $i = 1, \dots, d$ and $\ell_{1:i} \in \mathbb{N}_0^i$ with $\ell_i \neq 0$. From straightforward calculations it follows that

$$\begin{aligned} \ell_i \mathbb{P}(X_{1:i} = \ell_{1:i}) &= \mathbb{E} \sum_{k=1}^{n} \mathbf{1}\{\mathbf{Y}^{(k)} = \mathbf{e}_i\} \mathbf{1}\{X_{1:i} = \ell_{1:i}\} \qquad (4.17) \\ &= \mathbb{E} \sum_{k=1}^{n} \mathbf{1}\{\mathbf{Y}^{(k)} = \mathbf{e}_i\} \mathbf{1}\big\{X_{1:i}^{(k)} + W_{1:i}^{(k)} = (\ell_{1:i-1}, \ell_i - 1)\big\}. \end{aligned}$$

Let $H^{(i)}_{\ell_{1:i}}$ and $q^{(i)}_{\ell_{1:i}}$ be the quantities given by

$$H^{(i)}_{\ell_{1:i}} = \mathbb{E}\sum_{k=1}^{n} \mathbf{1}\{\mathbf{Y}^{(k)} = \mathbf{e}_i\}\mathbf{1}\big\{X^{(k)}_{1:i} = (\ell_{1:i-1}, \ell_i - 1)\big\},$$
$$q^{(i)}_{\ell_{1:i}} = \ell_i \mathbb{P}(X_{1:i} = \ell_{1:i}) - H^{(i)}_{\ell_{1:i}}$$

so that

$$\ell_i \mathbb{P}(X_{1:i} = \ell_{1:i}) = H^{(i)}_{\ell_{1:i}} + q^{(i)}_{\ell_{1:i}}.$$

For $i = 1, \dots, d$, let τ_i be a random variable independent of $(\mathbf{Y}^{(r)})_{r=1}^n$ with distribution

$$\mathbb{P}(\tau_i = k) = \frac{p_{k,i}}{\lambda_i}, \quad k = 1, \dots, n\,.$$

Since $\mathbf{Y}^{(r)}$, $r = 1, \dots, n$, are m-dependent, the random vectors $\mathbf{Y}^{(k)} = (Y^{(k)}_1, \dots, Y^{(k)}_d)$ and $\mathbf{X}^{(k)}$ are independent for all $k = 1, \dots, n$. Therefore

$$\begin{aligned} H^{(i)}_{\ell_{1:i}} &= \sum_{k=1}^{n} p_{k,i} \mathbb{P}\big(X^{(k)}_{1:i} = (\ell_{1:i-1}, \ell_i - 1)\big) \\ &= \sum_{k=1}^{n} p_{k,i} \mathbb{P}\big(X_{1:i} - W^{(k)}_{1:i} - Y^{(k)}_{1:i} = (\ell_{1:i-1}, \ell_i - i)\big) \\ &= \lambda_i \mathbb{P}\big(X_{1:i} - W^{(\tau_i)}_{1:i} - Y^{(\tau_i)}_{1:i} = (\ell_{1:i-1}, \ell_i - i)\big). \end{aligned}$$

Then, by Theorem 4.1.1 we obtain

$$d_W(\mathbf{X}, \mathbf{P}) \le \sum_{i=1}^{d} \Bigg(\lambda_i \mathbb{E}\left[W^{(\tau_i)}_i + Y^{(\tau_i)}_i\right] + 2\lambda_i \sum_{j=1}^{i-1} \mathbb{E}\left[W^{(\tau_i)}_j + Y^{(\tau_i)}_j\right] + \sum_{\substack{\ell_{1:i} \in \mathbb{N}_0^d \\ \ell_i \neq 0}} \big|q^{(i)}_{\ell_{1:i}}\big| \Bigg). \quad (4.18)$$

From (4.17) and the definition of $q^{(i)}_{\ell_{1:i}}$ it follows that

$$\begin{aligned} |q^{(i)}_{\ell_{1:i}}| &\le \mathbb{E}\sum_{k=1}^{n} \mathbf{1}\{\mathbf{Y}^{(k)} = \mathbf{e}_i\} \Big| \mathbf{1}\big\{X^{(k)}_{1:i} + W^{(k)}_{1:i} = (\ell_{1:i-1}, \ell_i - 1)\big\} - \mathbf{1}\big\{X^{(k)}_{1:i} = (\ell_{1:i-1}, \ell_i - 1)\big\} \Big| \\ &\le \mathbb{E}\sum_{k=1}^{n} \mathbf{1}\{\mathbf{Y}^{(k)} = \mathbf{e}_i\}\mathbf{1}\{W^{(k)}_{1:i} \neq 0\}\mathbf{1}\big\{X^{(k)}_{1:i} + W^{(k)}_{1:i} = (\ell_{1:i-1}, \ell_i - 1)\big\} \\ &\quad + \mathbb{E}\sum_{k=1}^{n} \mathbf{1}\{\mathbf{Y}^{(k)} = \mathbf{e}_i\}\mathbf{1}\{W^{(k)}_{1:i} \neq 0\}\mathbf{1}\big\{X^{(k)}_{1:i} = (\ell_{1:i-1}, \ell_i - 1)\big\}. \end{aligned}$$

Thus, by the inequality $\mathbf{1}\{W^{(k)}_{1:i} \neq 0\} \le \sum_{j=1}^{i} W^{(k)}_j$, we obtain

$$\begin{aligned} \sum_{\substack{\ell_{1:i} \in \mathbb{N}_0^i \\ \ell_i \neq 0}} \big|q^{(i)}_{\ell_{1:i}}\big| &\le 2\mathbb{E}\sum_{k=1}^{n} \mathbf{1}\{\mathbf{Y}^{(k)} = \mathbf{e}_i\}\mathbf{1}\{W^{(k)}_{1:i} \neq 0\} \\ &\le 2\mathbb{E}\sum_{k=1}^{n}\sum_{j=1}^{i} \mathbf{1}\{\mathbf{Y}^{(k)} = \mathbf{e}_i\} W^{(k)}_j \le 4mi \sum_{k=1}^{n} Q(k). \end{aligned} \quad (4.19)$$

Moreover, for any $i, j = 1, \dots, d$, we have

$$\lambda_i \mathbb{E}\Big[W_j^{(\tau_i)} + Y_j^{(\tau_i)}\Big] = \lambda_i \mathbb{E} \sum_{\substack{r=1,\dots,n,\\ |r-\tau_i|\le m}} \mathbf{1}\{\mathbf{Y}^{(r)} = \mathbf{e}_j\}$$
$$= \sum_{k=1}^{n} p_{k,i}\, \mathbb{E} \sum_{\substack{r=1,\dots,n,\\ |r-k|\le m}} \mathbf{1}\{\mathbf{Y}^{(r)} = \mathbf{e}_j\} = \sum_{\substack{k,r=1,\dots,n,\\ |r-k|\le m}} p_{k,i} p_{r,j}.$$

Together with (4.18) and (4.19), this leads to

$$d_W(\mathbf{X}, \mathbf{P}) \le \sum_{i=1}^{d} \sum_{\substack{k,r=1,\dots,n,\\ |r-k|\le m}} p_{k,i} p_{r,i} + 2 \sum_{i=1}^{d} \sum_{j=1}^{i-1} \sum_{\substack{k,r=1,\dots,n,\\ |r-k|\le m}} p_{k,i} p_{r,j} + 2d(d+1)m \sum_{k=1}^{n} Q(k)$$
$$\le \sum_{k=1}^{n} \sum_{i=1}^{d} \Bigg[\sum_{\substack{r=1,\dots,n,\\ |r-k|\le m}} p_{r,i} + 2 \sum_{j=1}^{i-1} \sum_{\substack{r=1,\dots,n,\\ |r-k|\le m}} p_{r,j} \Bigg] p_{k,i} + 2d(d+1)m \sum_{k=1}^{n} Q(k),$$

which completes the proof. □

4.3.2 Point processes with Papangelou intensity

Let ξ be a proper point process on a measurable space $(\mathbb{X}, \mathcal{X})$, that is, a point process that can be written as $\xi = \sum_{i=1}^{\tau} \delta_{X_i}$, for some random elements $X_1, X_2, \dots$ in $\mathbb{X}$ and a random variable $\tau \in \mathbb{N}_0 \cup \{\infty\}$. Recall that any Poisson process can be seen as a proper point process, and that, by [38, Corollary 6.5], all locally finite point processes are proper if $(\mathbb{X}, \mathcal{X})$ is a Borel space. The so-called reduced Campbell measure $\mathcal{C}$ of ξ is defined on the product space $\mathbb{X} \times \mathbf{N}_{\mathbb{X}}$ by

$$\mathcal{C}(A) = \mathbb{E} \int_{\mathbb{X}} \mathbf{1}_A(x, \xi \setminus \{x\})\, d\xi(x)$$

for all A from the product σ-field generated by $\mathcal{X} \times \mathcal{N}_{\mathbb{X}}$, where $\xi \setminus \{x\}$ denotes the point process $\xi - \delta_x$ if $x \in \xi$, and ξ otherwise. Let ν be a σ-finite measure on $(\mathbb{X}, \mathcal{X})$ and let $\mathbb{P}_\xi$ be the distribution of ξ on $(\mathbf{N}_{\mathbb{X}}, \mathcal{N}_{\mathbb{X}})$. If $\mathcal{C}$ is absolutely continuous with respect to $\nu \times \mathbb{P}_\xi$, any density c of $\mathcal{C}$ with respect to $\nu \times \mathbb{P}_\xi$ is called (a version of) the Papangelou intensity of ξ. This notion was originally introduced by Papangelou in [48]. In other words, c is a Papangelou intensity of ξ relative to the measure ν if the Georgii–Nguyen–Zessin equation

$$\mathbb{E} \int_{\mathbb{X}} u(x, \xi \setminus \{x\})\, d\xi(x) = \int_{\mathbb{X}} \mathbb{E}[c(x, \xi) u(x, \xi)]\, d\nu(x) \tag{4.20}$$

is satisfied for all measurable functions $u : \mathbb{X} \times \mathbf{N}_{\mathbb{X}} \to [0, \infty)$. Intuitively $c(x, \xi)$ is a random variable that measures the interaction between x and ξ; as a reinforcement of this exposition, it is well-known that if c is deterministic, that is, $c(x, \xi) = f(x)$ for some positive and measurable function f, then ξ is a Poisson process with intensity measure $\lambda(A) = \int_A f(x) d\nu(x)$, $A \in \mathcal{X}$, (see e.g. [38, Theorem 4.1]). For more details on this interpretation, we refer to [23, Section 4]. See also [37] and [64] for connections between Papangelou intensity and Gibbs point processes. We show that for any proper point process ξ that admits Papangelou intensity c relative to a measure ν, the d_π distance between ξ and a Poisson process with finite intensity measure λ, which is absolutely continuous with respect to ν, can be bounded by the distance in $L^1(\nu \times \mathbb{P}_\xi)$ between c and the density of λ. For a locally compact metric space, Theorem 4.3.3 yields the same bound as [23, Theorem 3.7], but for the metric d_π instead of the Kantorovich-Rubinstein distance. Observe that the inequality in Theorem 4.3.3 follows with an additional factor 2 immediately from [23, Theorem 3.7] and Lemma 4.2.6 when the underlying space is a locally compact metric space.

Theorem 4.3.3. *Let ξ be a proper point process on $\mathbb{X}$ that admits Papangelou intensity c with respect to a σ-finite measure ν such that $\int_{\mathbb{X}} \mathbb{E}\,|c(x,\xi)|\,d\nu(x) < \infty$. Let η be a Poisson process on $\mathbb{X}$ with finite intensity measure λ having density f with respect to ν. Then,*

$$d_\pi(\xi,\eta) \leq \int_{\mathbb{X}} \mathbb{E}\,|c(x,\xi) - f(x)|\,d\nu(x).$$

Proof of Theorem 4.3.3. The condition $\int_{\mathbb{X}} \mathbb{E}\,|c(x,\xi)|\,d\nu(x) < \infty$ and (4.20) ensure that ξ has finite intensity measure. Consider $i \in \mathbb{N}$ and $(A_1,\dots,A_i) \in \mathcal{X}^i_{\text{disj}}$. Hereafter, $\xi(A_{1:i})$ is shorthand notation for $(\xi(A_1),\dots,\xi(A_i))$. The idea of the proof is to apply Theorem 4.1.3 with the random vectors $\mathbf{Z}^{A_{1:i}}$ assumed to be $\mathbf{0}$. In this case,

$$\begin{aligned} q^{A_{1:i}}_{m_{1:i}} &= m_i \mathbb{P}\big(\xi(A_{1:i}) = m_{1:i}\big) - \lambda(A_i)\mathbb{P}\big(\xi(A_{1:i}) = (m_{1:i-1}, m_i - 1)\big) \\ &= m_i \mathbb{P}\big(\xi(A_{1:i}) = m_{1:i}\big) - \int_{\mathbb{X}} \mathbb{E}\big[f(x)\mathbf{1}_{A_i}(x)\mathbf{1}\{\xi(A_{1:i}) = (m_{1:i-1}, m_i - 1)\}\big]\,d\nu(x) \end{aligned}$$

for $m_{1:i} \in \mathbb{N}_0^i$ with $m_i \neq 0$, $i = 1,\dots,d$. It follows from (4.20) that

$$\begin{aligned} m_i \mathbb{P}\big(\xi(A_{1:i}) = m_{1:i}\big) &= \mathbb{E}\int_{\mathbb{X}} \mathbf{1}_{A_i}(x)\mathbf{1}\{\xi\setminus\{x\}(A_{1:i}) = (m_{1:i-1}, m_i - 1)\}\,d\xi(x) \\ &= \int_{\mathbb{X}} \mathbb{E}\big[c(x,\xi)\mathbf{1}_{A_i}(x)\mathbf{1}\{\xi(A_{1:i}) = (m_{1:i-1}, m_i - 1)\}\big]\,d\nu(x). \end{aligned}$$

Hence

$$q^{A_{1:i}}_{m_{1:i}} = \int_{\mathbb{X}} \mathbb{E}\big[(c(x,\xi) - f(x))\mathbf{1}_{A_i}(x)\mathbf{1}\{\xi(A_{1:i}) = (m_{1:i-1}, m_i - 1)\}\big]\,d\nu(x).$$

By Theorem 4.1.3, we obtain

$$d_\pi(\xi,\eta) \leq \sup_{(A_1,\dots,A_d)\in\mathcal{X}^d_{\text{disj}}, d\in\mathbb{N}} \sum_{i=1}^{d} \sum_{\substack{m_{1:i}\in\mathbb{N}_0^i \\ m_i\neq 0}} \big|q^{A_{1:i}}_{m_{1:i}}\big|\,.$$

Furthermore, the inequalities

$$\begin{aligned} \sum_{\substack{m_{1:i}\in\mathbb{N}_0^i \\ m_i\neq 0}} \big|q^{A_{1:i}}_{m_{1:i}}\big| &\leq \sum_{\substack{m_{1:i}\in\mathbb{N}_0^i, \\ m_i\neq 0}} \int_{\mathbb{X}} \mathbb{E}\big[|c(x,\xi) - f(x)|\mathbf{1}_{A_i}(x)\mathbf{1}\{\xi(A_{1:i}) = (m_{1:i-1}, m_i - 1)\}\big]\,d\nu(x) \\ &\leq \int_{\mathbb{X}} \mathbb{E}\Big[|c(x,\xi) - f(x)|\mathbf{1}_{A_i}(x) \sum_{\substack{m_{1:i}\in\mathbb{N}_0^i \\ m_i\neq 0}} \mathbf{1}\{\xi(A_{1:i}) = (m_{1:i-1}, m_i - 1)\}\Big]\,d\nu(x) \\ &\leq \int_{\mathbb{X}} \mathbb{E}\big[|c(x,\xi) - f(x)|\mathbf{1}_{A_i}(x)\big]\,d\nu(x) \end{aligned}$$

imply that

$$\sum_{i=1}^{d} \sum_{\substack{m_{1:i}\in\mathbb{N}_0^i \\ m_i\neq 0}} \left|q_{m_{1:i}}^{A_{1:i}}\right| \le \int_{\mathbb{X}} \mathbb{E}\,|c(x,\xi)-f(x)|\,d\nu(x)$$

for any $A_{1:d} \in \mathcal{X}_{\mathrm{disj}}^d$ with $d \in \mathbb{N}$. Thus, we obtain the assertion. □

4.3.3 Point processes of Poisson U-statistic structure

Let $(\mathbb{X}, \mathcal{X})$ and $(\mathbb{Y}, \mathcal{Y})$ be measurable spaces. For $\ell \in \mathbb{N}$ and a symmetric domain $D \in \mathcal{X}^\ell$, let $g : D \to \mathbb{Y}$ be a symmetric measurable function, i.e., for any $(x_1, \dots, x_\ell) \in D$ and any index permutation σ, $g(x_1, \dots, x_\ell) = g(x_{\sigma(1)}, \dots, x_{\sigma(\ell)})$. Let η be a Poisson process on $\mathbb{X}$ with finite intensity measure L. We are interested in the point process on $\mathbb{Y}$ given by

$$\xi = \frac{1}{\ell!} \sum_{(x_1,\dots,x_\ell)\in\eta_{\neq}^{\ell}\cap D} \delta_{g(x_1,\dots,x_\ell)}, \tag{4.21}$$

where for any $\zeta = \sum_{i\in I} \delta_{x_i} \in \mathbf{N}_{\mathbb{X}}$ with I at most countable, $\zeta_{\neq}^{\ell}$ denotes the collection of all ℓ-tuples $(x_1, \dots, x_\ell)$ of points from ζ with pairwise distinct indexes. The point process ξ has a Poisson U-statistic structure in the sense that, for any $B \in \mathcal{Y}$, $\xi(B)$ is a Poisson U-statistic. In Subsections 3.3.1 and 3.3.2, we studied the Poisson approximation of U-statistics. Hereafter we discuss the Poisson process approximation in the metric d_π for the point process ξ. We prove the exact analogue of [22, Theorem 3.1], with the Kantorovich–Rubinstein distance replaced by d_π. Several applications of this result are presented in [22], alongside with the case of underlying binomial point processes. It is worth mentioning that [22] relies on a slightly less general setup: $\mathbb{X}$ is assumed to be a locally compact second countable Hausdorff space (lcscH), while in the present work any measurable space is allowed. Observe that, when $\mathbb{X}$ is lcscH, the inequality in Theorem 4.3.4 follows with an additional factor 2 immediately from [22, Theorem 3.1] and Lemma 4.2.6.

Let λ denote the intensity measure of ξ, and note that, since L is a finite measure on $\mathbb{X}$, then $\lambda(\mathbb{Y}) < \infty$ by the multivariate Mecke formula. Define

$$R = \max_{1\le i\le \ell-1} \int_{\mathbb{X}^i} \left(\int_{\mathbb{X}^{\ell-i}} \mathbf{1}\{(x_1,\dots,x_\ell)\in D\}\, dL^{\ell-i}(x_{i+1},\dots,x_\ell) \right)^2 dL^i(x_1,\dots,x_i)$$

for $\ell \ge 2$, and put $R = 0$ for $\ell = 1$. The expression R is used to quantify the accuracy of the Poisson process approximation of ξ, and corresponds to the quantity r given in Subsection 3.3.2 for $h = \mathbf{1}\{(x_1, \dots, x_\ell) \in D\}$, which is used to study the Poisson approximation of Poisson U-statistics.

Theorem 4.3.4. *Let ξ, λ and R be as above, and let γ be a Poisson process on $\mathbb{Y}$ with intensity measure λ. Then,*

$$d_\pi(\xi, \gamma) \le \frac{2^{\ell+1}}{\ell!} R.$$

If the intensity measure λ of ξ is the zero measure, then the proof of Theorem 4.3.4 is trivial. From now on, we assume $0 < \lambda(\mathbb{Y}) < \infty$. The multivariate Mecke formula yields for every $A \in \mathcal{Y}$ that

$$\lambda(A) = \mathbb{E}[\xi(A)] = \frac{1}{\ell!}\mathbb{E} \sum_{\mathbf{x}\in\eta_{\neq}^{\ell}\cap D} \mathbf{1}\{g(\mathbf{x})\in A\} = \frac{1}{\ell!} \int_D \mathbf{1}\{g(\mathbf{x})\in A\}\, dL^\ell(\mathbf{x}).$$

Define for $\lambda(A) > 0$ the random element $\mathbf{X}^A = (X_1^A, \dots, X_\ell^A)$ in $\mathbb{X}^\ell$ independent of η and distributed according to

$$\mathbb{P}\left(\mathbf{X}^A \in B\right) = \frac{1}{\ell!\lambda(A)} \int_D \mathbf{1}\{g(\mathbf{x}) \in A\} \mathbf{1}\{\mathbf{x} \in B\} \, dL^\ell(\mathbf{x})$$

for B from the product σ-field generated by $\mathcal{X}^\ell$, and put $\mathbf{X}^A = x_0 \in \mathbb{X}^\ell$ for $\lambda(A) = 0$. For any vector $\mathbf{x} = (x_1, \dots, x_\ell) \in \mathbb{X}^\ell$, denote by $\Delta(\mathbf{x})$ the sum of ℓ Dirac measures located by the vector components, that is

$$\Delta(\mathbf{x}) = \Delta(x_1, \dots, x_\ell) = \sum_{i=1}^{\ell} \delta_{x_i} .$$

In what follows, for any point process ζ on $\mathbb{X}$, $\xi(\zeta)$ is the point process defined as in (4.21) with η replaced by ζ. As in Section 4.3.2, $\xi(A_{1:i})$ denotes the random vector $(\xi(A_1), \dots, \xi(A_i))$ for all $A_1, \dots, A_i \in \mathcal{Y}$, $i \in \mathbb{N}$.

Proof of Theorem 4.3.4. For $\ell = 1$, Theorem 4.3.4 is a direct consequence of [38, Theorem 5.1]. Whence, we assume $\ell \geq 2$. Let $A_1, \dots, A_i \in \mathcal{Y}$ with $i \in \mathbb{N}$ be disjoint sets and let $m_{1:i} \in \mathbb{N}_0^i$ with $m_i \neq 0$. Suppose $\lambda(A_i) > 0$. The multivariate Mecke formula implies that

$$\begin{aligned}
m_i \mathbb{P}(\xi(A_{1:i}) = m_{1:i}) &= \frac{1}{\ell!} \mathbb{E} \sum_{\mathbf{x} \in \eta^\ell_{\neq} \cap D} \mathbf{1}\{g(\mathbf{x}) \in A_i\} \mathbf{1}\{\xi(A_{1:i}) = m_{1:i}\} \\
&= \frac{1}{\ell!} \int_D \mathbf{1}\{g(\mathbf{x}) \in A_i\} \mathbb{P}(\xi(\eta + \Delta(\mathbf{x}))(A_{1:i}) = m_{1:i}) \, dL^\ell(\mathbf{x}) \\
&= \frac{1}{\ell!} \int_D \mathbf{1}\{g(\mathbf{x}) \in A_i\} \mathbb{P}\left(\xi(\eta + \Delta(\mathbf{x}))(A_{1:i}) - \delta_{g(\mathbf{x})}(A_{1:i}) = (m_{1:i-1}, m_i - 1)\right) dL^\ell(\mathbf{x}) \\
&= \lambda(A_i) \mathbb{P}\left(\xi\left(\eta + \Delta\left(\mathbf{X}^{A_i}\right)\right)(A_{1:i}) - \delta_{g\left(\mathbf{X}^{A_i}\right)}(A_{1:i}) = (m_{1:i-1}, m_i - 1)\right),
\end{aligned} \tag{4.22}$$

where the second last inequality holds true because $\delta_{g(\mathbf{x})}(A_{1:i})$ is the vector $(0, \dots, 0, 1) \in \mathbb{N}_0^i$ when $g(\mathbf{x}) \in A_i$. The previous identity is verified also if $\lambda(A_i) = 0$. Hence, for

$$\mathbf{Z}^{A_{1:i}} = \xi\left(\eta + \Delta\left(\mathbf{X}^{A_i}\right)\right)(A_{1:i}) - \xi(A_{1:i}) - \delta_{g\left(\mathbf{X}^{A_i}\right)}(A_{1:i}) ,$$

the quantity $q_{m_{1:i}}^{A_{1:i}}$ defined by Equation (4.7) in Theorem 4.1.3 is zero. Note that $\mathbf{Z}^{A_{1:i}}$ has non-negative components. Therefore for any $(A_1, \dots, A_d) \in \mathcal{X}^d_{\text{disj}}$ with $d \in \mathbb{N}$,

$$\begin{aligned}
\sum_{i=1}^{d} \lambda(A_i) \sum_{j=1}^{i} \mathbb{E}\left|\mathbf{Z}_j^{A_{1:i}}\right| &= \sum_{i=1}^{d} \lambda(A_i) \sum_{j=1}^{i} \mathbb{E}\left[\xi\left(\eta + \Delta\left(\mathbf{X}^{A_i}\right)\right)(A_j) - \xi(A_j) - \delta_{g\left(\mathbf{X}^{A_i}\right)}(A_j)\right] \\
&\leq \sum_{i=1}^{d} \lambda(A_i) \mathbb{E}\left[\xi\left(\eta + \Delta\left(\mathbf{X}^{A_i}\right)\right)(\mathbb{Y}) - \xi(\mathbb{Y}) - 1\right] \\
&= \frac{1}{\ell!} \sum_{i=1}^{d} \int_D \mathbf{1}\{g(\mathbf{x}) \in A_i\} \mathbb{E}\left[\xi(\eta + \Delta(\mathbf{x}))(\mathbb{Y}) - \xi(\mathbb{Y}) - 1\right] dL^\ell(\mathbf{x}) \\
&\leq \lambda(\mathbb{Y}) \mathbb{E}\left[\xi\left(\eta + \Delta\left(\mathbf{X}^{\mathbb{Y}}\right)\right)(\mathbb{Y}) - \xi(\mathbb{Y}) - 1\right].
\end{aligned}$$

Thus, Theorem 4.1.3 establishes

$$d_\pi(\xi,\gamma) \le 2\lambda(\mathbb{Y})\mathbb{E}\left[\xi\left(\eta+\Delta\left(\mathbf{X}^{\mathbb{Y}}\right)\right)(\mathbb{Y}) - \xi(\mathbb{Y}) - 1\right]. \tag{4.23}$$

From (4.22) with $i = 1$ and $A_1 = \mathbb{Y}$, it follows that the random variable $\xi\big(\eta+\Delta\left(\mathbf{X}^{\mathbb{Y}}\right)\big)(\mathbb{Y})$ is the size bias coupling of $\xi(\mathbb{Y})$. Property (4.6) with f being the identity function and simple algebraic computations yield

$$\begin{aligned}\mathbb{E}\left[\xi\left(\eta+\Delta\left(\mathbf{X}^{\mathbb{Y}}\right)\right)(\mathbb{Y}) - \xi(\mathbb{Y}) - 1\right] &= \lambda(\mathbb{Y})^{-1}\left\{\mathbb{E}\big[\xi(\mathbb{Y})^2\big] - \lambda(\mathbb{Y})^2 - \lambda(\mathbb{Y})\right\} \\ &= \lambda(\mathbb{Y})^{-1}\left\{\mathrm{Var}(\xi(\mathbb{Y})) - \lambda(\mathbb{Y})\right\}.\end{aligned} \tag{4.24}$$

Moreover, [55, Lemma 3.5] gives

$$\mathrm{Var}(\xi(\mathbb{Y})) - \lambda(\mathbb{Y}) \le \sum_{i=1}^{\ell-1}\frac{1}{\ell!}\binom{\ell}{i}R \le \frac{2^\ell - 1}{\ell!}R.$$

Combining these inequalities with (4.23) and (4.24) concludes the proof. □

Chapter 5

Criteria for Poisson process convergence

This chapter is a slightly modified and adjusted version of the following preprint article jointly written with Matthias Schulte:

F. Pianoforte and M. Schulte. Criteria for Poisson process convergence with applications to inhomogeneous Poisson-Voronoi tessellations. arXiv:2101.07739, 2021.

Abstract. In this chapter, we employ the relation between probabilities of two consecutive values of a Poisson random variable to derive conditions for the weak convergence of locally finite point processes to a Poisson process. As applications, we consider the starting points of k-runs in a sequence of Bernoulli random variables and point processes constructed using inradii and circumscribed radii of inhomogeneous Poisson-Voronoi tessellations.

Acknowledgments. The research was supported by the Swiss National Science Foundation Grant No. 200021_175584. We would like to thank two anonymous referees for valuable hints and comments.

5.1 Introduction and main results

Let X be a random variable taking values in $\mathbb{N}_0$ and let $\lambda > 0$. It is well-known that

$$k\mathbb{P}(X = k) = \lambda\mathbb{P}(X = k - 1), \quad k \in \mathbb{N}, \tag{5.1}$$

if and only if X follows a Poisson distribution with parameter λ. We use this observation to establish weak convergence to a Poisson process. Indeed, we will prove that a tight sequence of locally finite point processes ξ_n, $n \in \mathbb{N}$, satisfies

$$\lim_{n\to\infty} k\mathbb{P}(\xi_n(B) = k) - \lambda(B)\mathbb{P}(\xi_n(B) = k - 1) = 0, \quad k \in \mathbb{N},$$

for any B in a certain family of sets and some locally finite measure λ, if and only if ξ_n converges in distribution to a Poisson process with intensity measure λ. Many different methods to investigate Poisson process convergence are available in the literature; we refer to surveys and classical results [33, 45, 46]. Using Stein's method, one can even derive quantitative bounds for the Poisson process approximation; see Chapter 4 and e.g. [2, 5, 7, 10, 17, 18, 22, 47, 62, 63, 71] and the references therein. In contrast to these results, the findings in this chapter are purely qualitative and do not provide rates of convergence, but they have the advantage that the underlying conditions are easy to verify. This is demonstrated in Sections 5.3.2 and 5.3.3, where weak convergence of locally point processes constructed using inradii and circumscribed radii of inhomogeneous Poisson-Voronoi tessellations is established.

The proof of our abstract criterion for Poisson process convergence relies on characterizations of locally finite point process convergence from Subsection 2.2.2 and the characterizing equation (5.1) for the Poisson distribution.

Let us now give a precise formulation of our results. Let S be a locally compact second countable Hausdorff space (lcscH space) with Borel σ-field $\mathcal{S}$. Recall that, a non-empty class $\mathcal{U}$ of subsets of S is called a ring if it is closed under finite unions and intersections, and under proper differences. Let $\widehat{\mathcal{S}}$ denote the class of relatively compact Borel sets of S. We say that a measure λ on S is non-atomic if $\lambda(\{x\}) = 0$ for all $x \in S$, and we define

$$\widehat{\mathcal{S}}_\lambda = \{B \in \widehat{\mathcal{S}} \,:\, \lambda(\partial B) = 0\},$$

where ∂B indicates the boundary of B.

Our first main result provides a characterization of weak convergence to a Poisson process.

Theorem 5.1.1. *Let ξ_n, $n \in \mathbb{N}$, be a sequence of locally finite point processes, and let λ be a non-atomic locally finite measure on S. Let $\mathcal{U} \subseteq \widehat{\mathcal{S}}_\lambda$ be a ring containing a countable topological basis of S. Then the following statements are equivalent:*

(i) For all open sets $B \in \mathcal{U}$ and $k \in \mathbb{N}$, $\xi_n(B)$, $n \in \mathbb{N}$, is tight and

$$\lim_{n\to\infty} k\mathbb{P}(\xi_n(B) = k) - \lambda(B)\mathbb{P}(\xi_n(B) = k-1) = 0. \tag{5.2}$$

(ii) ξ_n, $n \in \mathbb{N}$, converges in distribution to a Poisson process with intensity measure λ.

Remark 5.1.2. *Note that the sequence $\xi_n(B)$, $n \in \mathbb{N}$, in Theorem 5.1.1 is tight by the Markov inequality if $\mathbb{E}[\xi_n(B)] \to \lambda(B)$.*

Remark 5.1.3. *For a point process ϱ, the function $f : S \times \mathcal{N}(S) \to [0, \infty)$ defined as*

$$f(x, \mu) = \mathbf{1}_B(x)\mathbf{1}\{\mu(B) = k\} \tag{5.3}$$

with $k \in \mathbb{N}$ and $B \in \mathcal{U}$ satisfies

$$\mathbb{E}\sum_{x\in\varrho} f(x, \varrho) - \int_S \mathbb{E}\big[f(x, \varrho + \delta_x)\big]d\lambda(x) = k\mathbb{P}(\varrho(B) = k) - \lambda(B)\mathbb{P}(\varrho(B) = k-1). \tag{5.4}$$

By the Mecke formula, the left-hand side of (5.4) *equals zero for all integrable functions $f : S \times \mathcal{N}(S) \to \mathbb{R}$, if and only if ϱ is a Poisson process with intensity measure λ (see Proposition 2.2.5). Theorem 5.1.1 shows that, if we replace ϱ by ξ_n, $n \in \mathbb{N}$, satisfying a tightness assumption, then the left-hand side of* (5.4) *vanishes as $n \to \infty$ for all f of the form* (5.3) *if and only if ξ_n, $n \in \mathbb{N}$, converges weakly to a Poisson process with intensity measure λ.*

Next we apply Theorem 5.1.1 to investigate point processes on S that are constructed from an underlying Poisson or binomial point process on a measurable space $(Y, \mathcal{Y})$. For

$t \geq 1$ let η_t be a Poisson process on Y with a σ-finite intensity measure P_t, while β_n is a binomial point process of $n \in \mathbb{N}$ independent points in Y which are distributed according to a probability measure Q_n. For a family of measurable functions $h_t : V_t \times \mathbf{N}_Y \to S$ with $V_t \in \mathcal{Y}$, $t \geq 1$, we are interested in the point processes

$$\sum_{x \in \eta_t \cap V_t} \delta_{h_t(x,\eta_t)}, \quad t \geq 1, \quad \text{and} \quad \sum_{x \in \beta_n \cap V_n} \delta_{h_n(x,\beta_n)}, \quad n \in \mathbb{N}.$$

In order to deal with both situations simultaneously, we introduce a joint notation. In the sequel, we study the point processes

$$\xi_t = \sum_{x \in \zeta_t \cap U_t} \delta_{g_t(x,\zeta_t)}, \quad t \geq 1, \tag{5.5}$$

where $\zeta_t = \eta_t$, $g_t = h_t$ and $U_t = V_t$ in the Poisson case, while $\zeta_t = \beta_{\lfloor t \rfloor}$, $g_t = h_{\lfloor t \rfloor}$ and $U_t = V_{\lfloor t \rfloor}$ in the binomial case. We assume

$$\mathbb{P}(\xi_t(B) < \infty) = 1 \quad \text{for all } B \in \widehat{\mathcal{S}}$$

so that ξ_t is locally finite. Let M_t be the intensity measure of ξ_t. By K_t we denote the intensity measure of ζ_t, i.e. $K_t = P_t$ if $\zeta_t = \eta_t$ and $K_t = \lfloor t \rfloor Q_{\lfloor t \rfloor}$ if $\zeta_t = \beta_{\lfloor t \rfloor}$. Moreover, we define $\hat{\zeta}_t = \eta_t$ in the Poisson case and $\hat{\zeta}_t = \beta_{\lfloor t \rfloor - 1}$ in the binomial case. From Theorem 5.1.1 we derive the following criterion for convergence of ξ_t, $t \geq 1$, to a Poisson process.

Theorem 5.1.4. *Let $\xi_t, t \geq 1$, be a family of locally finite point processes on S given by* (5.5) *and let M be a non-atomic locally finite measure on S. Fix any ring $\mathcal{U} \subset \widehat{\mathcal{S}}_M$ containing a countable topological basis, and assume that*

$$\lim_{t \to \infty} M_t(B) = M(B) \tag{5.6}$$

for all open sets $B \in \mathcal{U}$. Then,

$$\begin{aligned} \lim_{t \to \infty} \int_{U_t} \mathbb{E}\Big[\mathbf{1}\{g_t(x, \hat{\zeta}_t + \delta_x) \in B\} \mathbf{1}\Big\{ \sum_{y \in \hat{\zeta}_t \cap U_t} \delta_{g_t(y, \hat{\zeta}_t + \delta_x)}(B) = m \Big\}\Big] dK_t(x) \\ - M(B)\mathbb{P}(\xi_t(B) = m) = 0 \end{aligned} \tag{5.7}$$

for all open sets $B \in \mathcal{U}$ and $m \in \mathbb{N}_0$, if and only if $\xi_t, t \geq 1$, converges weakly to a Poisson process with intensity measure M.

Remark 5.1.5. *One is often interested in Poisson process convergence for $S = \mathbb{R}^d$, $d \geq 1$, and for the situation that the intensity measure of the Poisson process is absolutely continuous (with respect to the Lebesgue measure). In this case, we can apply Theorem 5.1.1 and Theorem 5.1.4 in the following way. The family $\mathcal{R}^d$ of sets in $\mathbb{R}^d$ that are finite unions of Cartesian products of bounded intervals is a ring contained in the relatively compact sets of $\mathbb{R}^d$. For any absolutely continuous measure the boundaries of sets from $\mathcal{R}^d$ have zero measure. By $\mathcal{I}^d$ we denote the subset of open sets of $\mathcal{R}^d$, which contains a countable topological basis of $\mathbb{R}^d$. Note that the sets of $\mathcal{I}^d$ are finite unions of Cartesian products of bounded open intervals. Thus, we prove weak convergence for sequences of locally finite point processes on $\mathbb{R}^d$ to Poisson processes with absolutely continuous locally finite intensity measures by showing* (5.2) *or* (5.6) *and* (5.7) *for all sets from $\mathcal{I}^d$. For $d = 1$ we use the convention $\mathcal{I} = \mathcal{I}^1$.*

Theorem 5.1.4 says that in order to establish Poisson process convergence for point processes of the form (5.5), one has to deal with the dependence between

$$\mathbf{1}\{g_t(x,\hat{\zeta}_t+\delta_x)\in B\} \quad \text{and} \quad \mathbf{1}\Big\{\sum_{y\in\hat{\zeta}_t\cap U_t}\mathbf{1}\{g_t(y,\hat{\zeta}_t+\delta_x)\in B\}=m\Big\}.$$

We say that a statistic is locally dependent if its value at a given point depends only on a local and deterministic neighborhood. That is, for any fixed $x\in Y$ and $B\in\mathcal{U}$, there exists a set $A_{t,x}\in\mathcal{Y}$ with $x\in A_{t,x}$ such that

$$\mathbf{1}\{g_t(x,\hat{\zeta}_t+\delta_x)\in B\}=\mathbf{1}\{g_t(x,\hat{\zeta}_t|_{A_{t,x}}+\delta_x)\in B\}, \tag{5.8}$$

where $\mu|_A$ denotes the restriction of a measure μ to a set A. For further notions of local dependence in the context of point processes we refer to [7, 17, 18]. Next we describe heuristically how (5.8) can be applied to show (5.7) in Theorem 5.1.4 for $\zeta_t=\eta_t$ if

$$\mathbf{1}\Big\{\sum_{y\in\eta_t\cap U_t}\delta_{g_t(y,\eta_t+\delta_x)}(B)=m\Big\}\approx\mathbf{1}\Big\{\sum_{y\in\eta_t\cap A^c_{t,x}\cap U_t}\delta_{g_t(y,\eta_t|_{A^c_{t,x}})}(B)=m\Big\} \tag{5.9}$$

for $x\in Y$, where A^c denotes the complement of $A\subset Y$, and where by $\approx$ we mean that, as t increases to infinity, the two indicator functions have the same limit behavior. Under the assumption (5.8), the integral in (5.7) coincides with

$$\int_{U_t}\mathbb{E}\Big[\mathbf{1}\{g_t(x,\eta_t|_{A_{t,x}}+\delta_x)\in B\}\mathbf{1}\Big\{\sum_{y\in\eta_t\cap U_t}\delta_{g_t(y,\eta_t+\delta_x)}(B)=m\Big\}\Big]dK_t(x). \tag{5.10}$$

By (5.9), the last expression can be approximated by

$$\int_{U_t}\mathbb{E}\Big[\mathbf{1}\{g_t(x,\eta_t|_{A_{t,x}}+\delta_x)\in B\}\mathbf{1}\Big\{\sum_{y\in\eta_t\cap A^c_{t,x}\cap U_t}\delta_{g_t(y,\eta_t|_{A^c_{t,x}})}(B)=m\Big\}\Big]dK_t(x). \tag{5.11}$$

Due to the independence of $\eta_t|_{A_{t,x}}$ and $\eta_t|_{A^c_{t,x}}$, this can be rewritten as

$$\int_{U_t}\mathbb{P}\Big(\sum_{y\in\eta_t\cap A^c_{t,x}\cap U_t}\delta_{g_t(y,\eta_t|_{A^c_{t,x}})}(B)=m\Big)\mathbb{E}\big[\mathbf{1}\{g_t(x,\eta_t|_{A_{t,x}}+\delta_x)\in B\}\big]dK_t(x). \tag{5.12}$$

Using once more (5.8) and (5.9), the previous term can be approximated by

$$\mathbb{P}(\xi_t(B)=m)\int_{U_t}\mathbb{E}\big[\mathbf{1}\{g_t(x,\eta_t+\delta_x)\in B\}\big]dK_t(x)=\mathbb{P}(\xi_t(B)=m)M_t(B), \tag{5.13}$$

where the last equality follows from the Mecke formula. Consequently, the expression on the left-hand side of (5.7) becomes small if the approximation in (5.9) is good.

We believe that, under the assumption (5.8), condition (5.9) is similar to the ones from [62, Theorem 2.1]. For example, suppose $S=\mathbb{R}, Y=\mathbb{R}^d$ with $d\in\mathbb{N}$, and assume for any fixed $B=(u,v)\subset\mathbb{R}$ with $u<v\in\mathbb{R}$ that there exists $r(t)>0$ such that $r(t)\to 0$ as $t\to\infty$, and

$$A_{t,x}\subset x+r(t)[-1,1]^d, \quad x\in U_t\subset\mathbb{R}^d. \tag{5.14}$$

Then, verifying that the approximation in (5.9) is good (in the sense that the steps above are correct) is almost equivalent to consider

$$\xi_t|_{(u,v)} = \sum_{i=1}^{N(t)} \sum_{x\in\eta_t\cap U_i^t} \delta_{g_t(x,\eta_t)} \mathbf{1}\{g_t(x,\eta_t)\in(u,v)\} =: \sum_{i=1}^{N} \xi_i^t,$$

where $P_t = \{U_i^t, i = 1,\dots,N(t)\}$ is a partition of U_t of d-dimensional cubes of side length $r(t)$ (here we are not considering technical issues related to the existence of such partition), and to prove that

$$\begin{aligned} H_1 + H_2 + H_3 &:= \sum_{i=1}^{N(t)} \mathbb{P}\big(\xi_i^t(U_t)\geq 2\big) + \sum_{i=1}^{N(t)}\sum_{j\in\Gamma_i^s} \mathbb{P}\big(\xi_i^t(U_t)\geq 1\big)\mathbb{P}\big(\xi_j^t(U_t)\geq 1\big) \\ &\quad + \sum_{i=1}^{N(t)}\sum_{j\in\Gamma_i^s} \mathbb{P}\big(\xi_i^t(U_t)\geq 1, \xi_j^t(U_t)\geq 1\big) \end{aligned}$$

vanishes as $t\to\infty$, where for each i, Γ_i^s consists of the indexes of the sets in P_t surrounding U_i^t. Indeed, if H_1 is small,

$$\mathbf{1}\{\xi_i^t(B)\geq 1\} \approx \xi_i^t(B)$$

for all $i = 1,\dots,N(t)$, and hence, by the Mecke formula, $\mathbb{P}\big(\xi_i^t(B)\geq 1, \xi_j^t(B)\geq 1\big)$ can be approximated by

$$\int_{U_i^t} \mathbb{E}\big[\mathbf{1}\{g_t(x,\eta_t|_{A_{t,x}}+\delta_x)\in B\}\mathbf{1}\{\xi_j^t(\eta_t+\delta_x)(B)\geq 1\}\big]dK_t(x)$$

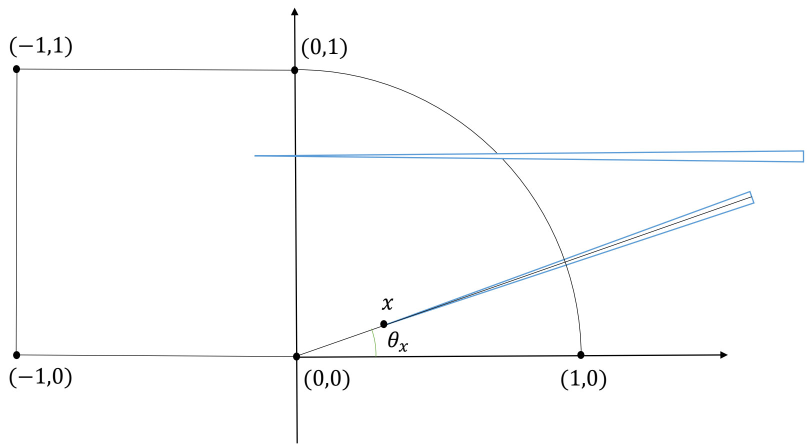
for $i\neq j$, where $\xi_j^t(\eta_t+\delta_x)$ is defined as ξ_j^t with η_t replaced by $\eta_t+\delta_x$. Now, since for each $x\in U_i^t$, $g_t(x,\eta_t)\in(u,v)$ depends only on the points of η_t in $U_i^t + r(t)[-1,1]^d$, ξ_i^t is independent of all ξ_j^t except of the ones corresponding to the sets U_j^t surrounding U_i^t. Therefore, we obtain that the difference between (5.10) and (5.11) can be estimated by

$$\begin{aligned} &\sum_{i=1}^{N(t)} \int_{U_i^t} \mathbb{E}\big[\mathbf{1}\{g_t(x,\eta_t|_{A_{t,x}}+\delta_x)\in B\}\mathbf{1}\{\xi_i^t(\eta_t+\delta_x)(U_t)\geq 2\}\big]dK_t(x) \\ &+ \sum_{i=1}^{N(t)}\sum_{j\in\Gamma_i^s} \int_{U_i^t} \mathbb{E}\big[\mathbf{1}\{g_t(x,\eta_t|_{A_{t,x}}+\delta_x)\in B\}\mathbf{1}\{\xi_j^t(\eta_t+\delta_x)(U_t)\geq 1\}\big]dK_t(x) \end{aligned}$$

which corresponds approximately to $H_1 + H_3$. By similar arguments, one can use H_1 and H_2 to estimate the difference between (5.12) and (5.13).

The previous quantities, H_1, H_2 and H_3, correspond to the first and third sums on the right-hand side of the inequality in [62, Theorem 2.1]. Since ξ_i^t is independent of all ξ_j^t except of the ones corresponding to the sets U_j^t surrounding U_i^t, the third and the fourth quantities on the right-hand side of the inequality in [62, Theorem 2.1] are zero. Then, under the assumptions (5.8) and (5.14), one may try to apply [62, Theorem 2.1] to prove Poisson process convergence of ξ_t, which unlike Theorem 5.1.4, provides also an estimate for the accuracy of the approximation. However, the reader should notice that the discussion below Remark 5.1.5, concerning possible situations for which Theorem 5.1.4 can be applied, also holds when the sets $A_{t,x}, x\in\mathbb{X}$, do not satisfy (5.14), while,

Figure 5.1: Triangles orientation.



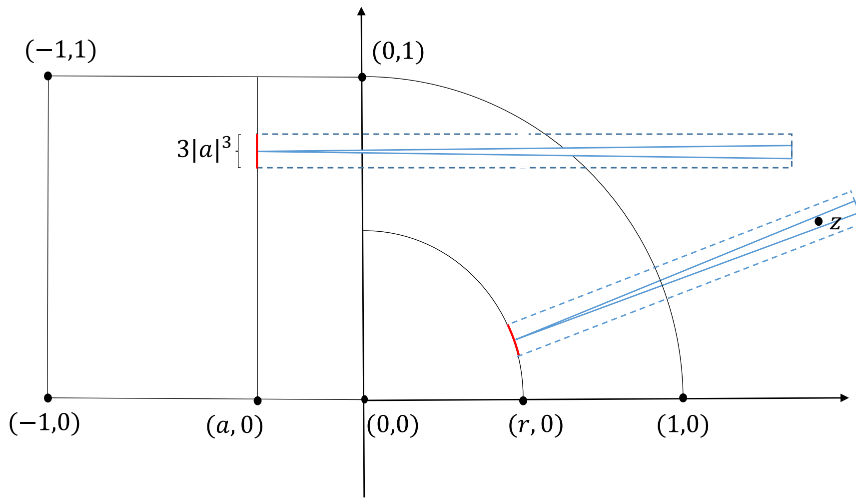
since [62, Theorem 2.1] is based on a discretization argument, it may not be applicable without this hypothesis. For instance, if $\xi_t, t \geq 1$, is a family of point processes on S with Papangelou intensity $c_t(x, \xi_t)$ (see Subsection 4.3.2 for the definition), one can obtain from Theorem 5.1.4 and a simple modification of the proof of Theorem 4.3.3 that ξ_t converges weakly to a Poisson process, if $c_t(x, \xi_t)$ converges in L^1 to a deterministic function as $t \to \infty$. Hence, Theorem 5.1.4 permits the derivation of the qualitative version of Theorem 4.3.3, while showing this result using [62, Theorem 2.1] may not be possible. Furthermore, Theorem 5.1.4 can also be applied when the point processes $\xi_t, t \geq 1$, defined by (5.5) have an underlying binomial point process, and for these, the fourth and fifth expressions on the right-hand side of the inequality in [62, Theorem 2.1] are not zero. Let us now consider an example where Theorem 5.1.4 can be easily applied to prove Poisson convergence, whereas it is unclear how to employ [62, Theorem 2.1] to establish the same result.

Example 5.1.6. *Let $D \subset \mathbb{R}^2$ be the open disk centered at $(0,0)$ with radius 1, and let $W = \{D \cap (0,1)^2\} \cup \{(-1,0) \times (0,1)\}$. For any fixed $x \in W$, let Δ_x be an isosceles triangle with x as the vertex that connects two sides of equal length. We denote by b_x the side of Δ_x that does not have x as vertex. For $x \in W \cap (0,1)^2$, we assume that b_x has length $\|x\|^5$ and height $\pi^2/(2\|x\|)$, where $\|\cdot\|$ denotes the Euclidean norm, and we consider Δ_x such that the angle between the vector $e_1 = (1,0)$ and*

$$w(x) := \operatorname{argmin}\{\|x-y\| \,:\, y \in b_x\} - x$$

is equal to the angle between e_1 and x, which we denote by $\theta_x \in (0, \pi/2)$. For $x = (a,b) \in (-1,0) \times (0,1)$, we assume that b_x has length $|a|^3$ and height $4/|a|$, and that the angle between e_1 and $w(x)$ is 0. Under the previous assumptions, Δ_x is uniquely determined for all $x \in W$ (see Figure 5.1). Let λ_2 be the Lebesgue measure on $\mathbb{R}^2$. Since $\lambda_2(\Delta_x)$

Figure 5.2: Geometrical interpretation of Example 5.1.6.



equals $\pi^2\|x\|^4/4$ *for* $x \in D \cap (0,1)^2$*, and* $4|a|^2$ *for* $x = (a,b) \in (-1,0) \times (0,1)$*, it can be easily verified that*

$$\lambda_2\big(\{x \in W : \sqrt{\lambda_2(\Delta_x)} \in (0,\alpha)\}\big) = \alpha, \quad \alpha \in [0,1]. \tag{5.15}$$

Let $z \in W$ *be fixed. For each circle* C *centered at the origin with radius* $r < 1$*, the set of all* $x \in C \cap \{W \cap (0,1)^2\}$ *with* $z \in \Delta_x$ *is an arc of length bounded by* $s\|x\|^5$ *for some* $s > 0$ *independent of* C *(see Figure 5.2). Moreover, for any* $a \in (-1,0)$*, the length of* $\{b \in (0,1) : z \in \Delta_{(a,b)}\}$ *is bounded by* $3|a|^3$*. Therefore*

$$\begin{aligned}
&\lim_{t\to\infty} t\lambda_2\big(\{x \in W : \sqrt{\lambda_2(\Delta_x)} \in (0,\alpha/t), z \in \Delta_x\}\big) \\
&= \lim_{t\to\infty} t\Big[\lambda_2\big(\{x \in W \cap (0,1)^2 : \pi\|x\|^2/2 \le \alpha/t, z \in \Delta_x\}\big) \\
&\qquad + \lambda_2\big(\{x = (a,b) \in W \cap \{(-1,0)\times(0,1)\} : 2|a| \le \alpha/t, z \in \Delta_x\}\big)\Big] \\
&\le \lim_{t\to\infty} t\Big[\int_0^{\sqrt{2\alpha/(\pi t)}} sr^6 dr + \int_{-\alpha/(2t)}^0 3|a|^3 da\Big] = 0
\end{aligned} \tag{5.16}$$

for all $\alpha > 0$ *and* $z \in W$*. Let* η_t *be a Poisson process on* $\mathbb{R}^2$ *with intensity measure* $t\lambda_2, t > 0$*. Define* $g_t(x,\eta_t) = \sqrt{\lambda_2(\Delta_x)}$ *if* $\eta_t(\Delta_x) = 0$*, and* 0 *otherwise. Consider the point process on* $(0,\infty)$ *given by*

$$\xi_t = \sum_{x \in \eta_t \cap W} \delta_{tg_t(x,\eta_t)} \mathbf{1}\{tg_t(x,\eta_t) > 0\},$$

and observe that

$$\mathbf{1}\{g_t(x,\eta_t+\delta_x)\in B\}=\mathbf{1}\{g_t(x,\eta_t|_{\Delta_x}+\delta_x)\in B\}=\mathbf{1}\big\{\eta_t(\Delta_x)=0,t\sqrt{\lambda_2(\Delta_x)}\in B\big\} \tag{5.17}$$

for any $B\subset(0,\infty)$ *and* $x\in W$. *In this example it is not possible to define a square* $r(t)[-1,1]^2$ *for some* $r(t)>0, r(t)\to 0$ *as* $t\to\infty$, *such that* (5.14) *is satisfied for all* $x\in W$, *because the triangles* Δ_x *become thinner and longer as we approach the origin, and rotate along the circles for* $x\in D\cap(0,1)^2$, *while they are horizontal for* $x\in(-1,0)\times(0,1)$. *For this reason, showing that* ξ_t *converges weakly to a Poisson process as* $t\to\infty$ *by applying [62, Theorem 2.1] may be complicated. On the other hand, one can derive this result from Theorem 5.1.4 in a simple way. Indeed, from the Mecke formula and the assumptions on* Δ_x, *we have for all finite unions of open and bounded intervals* $B\subset(0,\infty)$,

$$\begin{aligned}\lim_{t\to\infty}\mathbb{E}[\xi_t(B)]&=\lim_{t\to\infty}t\int_W\mathbb{P}(\eta_t(\Delta_x)=0)\mathbf{1}\big\{t\sqrt{\lambda_2(\Delta_x)}\in B\big\}dx\\&=\lim_{t\to\infty}t\int_W\mathbf{1}\big\{t\sqrt{\lambda_2(\Delta_x)}\in B\big\}dx,\end{aligned}$$

where we used that $\lambda_2(\Delta_x)\le(\sup(B)/t)^2$ *for* $t\sqrt{\lambda_2(\Delta_x)}\in B$ *in the last step. From* (5.15) *it follows that the previous limit equals the Lebesgue measure of* B. *Moreover, the independence of* $\eta_t|_{\Delta_x}$ *and* $\eta_t|_{\Delta_x^c}$ *and* (5.17) *imply for all* $m\in\mathbb{N}_0$ *that*

$$\begin{aligned}&\lim_{t\to\infty}t\int_W\mathbb{E}\Big[\mathbf{1}\{g_t(x,\eta_t+\delta_x)\in B\}\mathbf{1}\Big\{\sum_{y\in\eta_t\cap W}\delta_{g_t(y,\eta_t+\delta_x)}(B)=m\Big\}\Big]dx\\&\qquad-\lambda_2(B)\mathbb{P}(\xi_t(B)=m)\\&=\lim_{t\to\infty}t\int_W\mathbf{1}\big\{t\sqrt{\lambda_2(\Delta_x)}\in B\big\}\mathbb{P}\Big(\eta_t(\Delta_x)=0,\sum_{y\in\eta_t\cap W\cap\Delta_x^c}\delta_{g_t(y,\eta_t|_{\Delta_x^c}+\delta_x)}(B)=m\Big)dx\\&\qquad-\lambda_2(B)\mathbb{P}(\xi_t(B)=m)\\&=\lim_{t\to\infty}t\int_W\mathbf{1}\big\{t\sqrt{\lambda_2(\Delta_x)}\in B\big\}\mathbb{P}(\eta_t(\Delta_x)=0)\\&\qquad\times\mathbb{P}\Big(\sum_{y\in\eta_t\cap W\cap\Delta_x^c}\delta_{g_t(y,\eta_t|_{\Delta_x^c}+\delta_x)}(B)=m\Big)dx-\lambda_2(B)\mathbb{P}(\xi_t(B)=m)\\&=\lim_{t\to\infty}t\int_W\mathbf{1}\big\{t\sqrt{\lambda_2(\Delta_x)}\in B\big\}\\&\qquad\times\Big[\mathbb{P}\Big(\sum_{y\in\eta_t\cap W\cap\Delta_x^c}\delta_{g_t(y,\eta_t|_{\Delta_x^c}+\delta_x)}(B)=m\Big)-\mathbb{P}(\xi_t(B)=m)\Big]dx,\end{aligned}$$

where we used that $\lambda_2(\Delta_x)\le(\sup(B)/t)^2$ *for* $t\sqrt{\lambda_2(\Delta_x)}\in B$ *in the last step. The right-hand side is bounded by*

$$\lim_{t\to\infty}t\int_W\mathbf{1}\{t\sqrt{\lambda_2(\Delta_x)}\}\Big[\mathbb{P}(\eta_t(\Delta_x)>0)+\mathbb{P}\big(\eta_t\big(\big\{y\in W\colon t\sqrt{\lambda_2(\Delta_y)}\in B,x\in\Delta_y\big\}\big)>0\big)\Big]dx.$$

By applying the Markov inequality and the Mecke formula, and using the bound for $\lambda_2(\Delta_x)$ *and* (5.16), *we obtain that the previous limit equals* 0. *Therefore, from Theorem 5.1.5 and Remark 5.1.4 it follows that* ξ_t *converges weakly to a Poisson process with the restriction of the Lebesgue measure to* $(0,\infty)$ *as the intensity measure.*

In Section 5.3, we provide further examples for applying our abstract main results Theorem 5.1.1 and Theorem 5.1.4. Our first example in Subsection 5.3.1 are k-runs, i.e. at least k successes in a row in a sequence of Bernoulli random variables. For the situation that the success probabilities converge to zero, we show that the rescaled starting points of the k-runs behave like a Poisson process if some independence assumptions on the underlying Bernoulli random variables are satisfied.

As the second and third example, we consider statistics related to inradii and circumscribed radii of inhomogeneous Poisson-Voronoi tessellations. We study the Voronoi tessellation generated by a Poisson process η_t, $t > 0$, on $\mathbb{R}^d$ with intensity measure $t\mu$, where μ is a locally finite and absolutely continuous measure with density f. In Subsection 5.3.2, for any cell with the nucleus in a compact set, we take the μ-measure of the ball centered at the nucleus and with twice the inradius as the radius. We prove that the point process formed by these statistics converges in distribution after a transformation depending on t to a Poisson process as $t \to \infty$ under some minor assumptions on the density f. Our transformation allows us to describe the behavior of the balls with large μ-measures. In Subsection 5.3.3, we consider for each cell with the nucleus in a compact convex set the μ-measure of the ball around the nucleus with the circumscribed radius as radius and establish, after rescaling with a power of t, convergence in distribution to a Poisson process for $t \to \infty$. This result requires continuity of f, but under weaker assumptions on f, we provide lower and upper bounds for the tail distribution of the minimal μ-measure of these balls having the circumscribed radii as radii.

In [15], the limiting distributions of the maximal inradius and the minimal circumscribed radius of a stationary Poisson-Voronoi tessellation were derived. In our work, we extend these results in two directions. First, our findings imply Poisson process convergence of the transformed inradii and circumscribed radii for the stationary case. This implies the mentioned results from [15] and allows to deal with the m-th largest (or smallest) value or combinations of several order statistics. Second, we deal with inhomogeneous Poisson-Voronoi tessellations. In Subsections 3.3.5 and 3.3.6 Poisson approximation results for the minimal circumscribed radius and the maximal inradius for the stationary case are established, while in [19] some general results for the extremes of stationary tessellations were deduced. For stationary Poisson-Voronoi tessellations the convergence of the nuclei of extreme cells to a compound Poisson process was studied in [20].

As our Theorem 5.1.4 deals with underlying Poisson and binomial point processes, we expect that one can extend our results on inradii and circumscribed radii of Poisson-Voronoi tessellations to Voronoi tessellations constructed from an underlying binomial point process.

Before we discuss our applications in Section 5.3, we prove our main results in the next section.

5.2 Proofs of the results of Section 5.1

Let S be a locally compact second countable Hausdorff space, which is abbreviated as lcscH space. Recall that a family of sets $\mathcal{C} \subset \widehat{\mathcal{S}}$ is called dissecting if

(i) every open set $G \subset S$ can be written as a countable union of sets in $\mathcal{C}$,

(ii) every relatively compact set $B \in \widehat{\mathcal{S}}$ is covered by finitely many sets in $\mathcal{C}$.

Lemma 5.2.1. *A countable topological basis $\mathcal{T}$ of S is dissecting.*

Proof. By the definition of a countable topological basis $\mathcal{T}$ has property (i) of a dissecting family of sets. Since, for any $B \in \widehat{\mathcal{S}}$, $\cup_{T\in\mathcal{T}} T = S \supset \overline{B}$, the compactness of $\overline{B}$ implies that (ii) is satisfied. □

Let us now state a consequence of Proposition 2.2.10 and Proposition 2.2.11. This result will be used in the proof of Theorem 5.1.1. We write $\xrightarrow{d}$ to denote convergence in distribution.

Lemma 5.2.2. *Let $\xi_n, n \in \mathbb{N}$, be a sequence of locally finite point processes on S, and let γ be a Poisson process on S with a non-atomic locally finite intensity measure λ. Let $\mathcal{U} \subset \widehat{\mathcal{S}}_\lambda$ be a ring containing a countable topological basis. Then the following statements are equivalent:*

(i) $\xi_n \xrightarrow{d} \gamma$.

(ii) $\xi_n(B) \xrightarrow{d} \gamma(B)$ *for all open sets* $B \in \mathcal{U}$.

Proof. Observe that [38, Theorem 3.6] ensures the existence of a Poisson process γ with intensity measure λ. Since λ has no atoms, from Lemma 2.2.8 it follows that γ is a simple point process (i.e. $\mathbb{P}(\gamma(\{x\}) \le 1$ for all $x \in S) = 1$). Elementary arguments also yield $\widehat{\mathcal{S}}_\lambda = \{B \in \widehat{\mathcal{S}} : \gamma(\partial B) = 0 \text{ a.s.}\} = \widehat{\mathcal{S}}_\gamma$, and it follows from Lemma 5.2.1 that $\mathcal{U}$ is dissecting.

By Proposition 2.2.10, we obtain that (i) implies (ii).

Conversely, if $\xi_n(U) \xrightarrow{d} \xi(U)$ for all $U \in \mathcal{U}$, the desired result follows from Proposition 2.2.11, whose conditions are satisfied with $\mathcal{U}$ as dissecting ring. Thus, it is enough to show that (ii) implies $\xi_n(U) \xrightarrow{d} \xi(U)$ for all $U \in \mathcal{U}$.

For any $U \in \mathcal{U}$ there exists a sequence of open sets $A_j, j \in \mathbb{N}$, such that

$$U \subset A_j, \quad A_{j+1} \subset A_j \quad \text{and} \quad \overline{U} = \cap_{j\in\mathbb{N}} A_j.$$

Since $\mathcal{U}$ contains a countable topological basis, for any A_j one can find a countable family of open sets $B_\ell^{(j)}, \ell \in \mathbb{N}$, in $\mathcal{U}$ such that $\cup_{\ell\in\mathbb{N}} B_\ell^{(j)} = A_j$. In particular, they cover the compact set $\overline{U}$. So there exists a finite subcover of elements from $B_\ell^{(j)}, \ell \in \mathbb{N}$, that covers $\overline{U}$. Since $\mathcal{U}$ is a ring, the union of the elements belonging to this subcover of $\overline{U}$ is in $\mathcal{U}$ for each $j \in \mathbb{N}$. Because $\mathcal{U}$ is closed under finite intersections, we can make this family of sets from $\mathcal{U}$ that contain $\overline{U}$ monotonously decreasing in j. Thus, without loss of generality, we may assume $A_j \in \mathcal{U}$ for all $j \in \mathbb{N}$.

Since $\mathcal{U}$ is a ring and contains a countable topological basis, for the interior $\text{int}(U)$ of U there exists a sequence of open sets $B_j \in \mathcal{U}$, $j \in \mathbb{N}$, such that

$$B_j \subset U, \quad B_j \subset B_{j+1} \quad \text{and} \quad \text{int}(U) = \cup_{j\in\mathbb{N}} B_j.$$

For a fixed $m \in \mathbb{N}$, we have that

$$\mathbb{P}(\xi_n(B_j) \ge m) \le \mathbb{P}(\xi_n(U) \ge m) \le \mathbb{P}(\xi_n(A_j) \ge m)$$

for all $n \in \mathbb{N}$. By $\xi_n(U') \xrightarrow{d} \gamma(U')$ for all open sets $U' \in \mathcal{U}$, we obtain

$$\mathbb{P}(\gamma(B_j) \ge m) \le \liminf_{n\to\infty} \mathbb{P}(\xi_n(U) \ge m) \le \limsup_{n\to\infty} \mathbb{P}(\xi_n(U) \ge m) \le \mathbb{P}(\gamma(A_j) \ge m). \tag{5.18}$$

Moreover, from $U \in \widehat{\mathcal{S}}_\lambda$, whence $\lambda(\partial U) = 0$, it follows that $\lambda(B_j) \to \lambda(\mathrm{int}(U)) = \lambda(U)$ and $\lambda(A_j) \to \lambda(\overline{U}) = \lambda(U)$ as $j \to \infty$. Thus, letting $j \to \infty$ in (5.18) and using that γ is a Poisson process lead to

$$\lim_{n\to\infty} \mathbb{P}(\xi_n(U) \geq m) = \mathbb{P}(\gamma(U) \geq m).$$

This establishes $\xi_n(U) \xrightarrow{d} \xi(U)$ and concludes the proof. □

We are now in the position to prove the first main result of Section 5.1.

Proof of Theorem 5.1.1. Let us show (i) implies (ii). By Lemma 5.2.2 it is enough to prove that $\xi_n(B) \xrightarrow{d} \gamma(B)$ for all open sets $B \in \mathcal{U}$. Since $\mathbb{P}(\xi_n(B) = 0)$, $n \in \mathbb{N}$, is a bounded sequence in $[0, 1]$, there exists a subsequence such that $\lim\limits_{j\to\infty} \mathbb{P}(\xi_{n_j}(B) = 0)$ exists; then repeated applications of (5.2) yield for $k \in \mathbb{N}$ that

$$\lim_{j\to\infty} \mathbb{P}(\xi_{n_j}(B) = k) = \frac{\lambda(B)^k}{k!} \lim_{j\to\infty} \mathbb{P}(\xi_{n_j}(B) = 0). \tag{5.19}$$

Consequently we have for any $N \in \mathbb{N}$,

$$\begin{aligned}\sum_{k=0}^{N} \lim_{j\to\infty} \mathbb{P}(\xi_{n_j}(B) = k) &= \lim_{j\to\infty} \mathbb{P}(\xi_{n_j}(B) \in \{0, \dots, N\}) \\ &= 1 - \lim_{j\to\infty} \mathbb{P}(\xi_{n_j}(B) \in \{N+1, N+2, \dots\}).\end{aligned}$$

By tightness of $\xi_{n_j}(B)$, $j \in \mathbb{N}$, the right-hand side of the equation converges to 1 as $N \to \infty$ so that

$$\sum_{k\in\mathbb{N}_0} \lim_{j\to\infty} \mathbb{P}(\xi_{n_j}(B) = k) = 1.$$

Thus, from (5.19) we deduce $\lim\limits_{j\to\infty} \mathbb{P}(\xi_{n_j}(B) = 0) = e^{-\lambda(B)}$. Together with (5.19), this proves that

$$\lim_{j\to\infty} \mathbb{P}(\xi_{n_j}(B) = k) = \frac{\lambda(B)}{k!}\, e^{-\lambda(B)}$$

for all $k \in \mathbb{N}_0$. In conclusion, since for any subsequence $(n_\ell)_{\ell\in\mathbb{N}}$ there exists a further subsequence $(n_{\ell_i})_{i\in\mathbb{N}}$ such that $\mathbb{P}(\xi_{n_{\ell_i}}(B) = 0)$, $i \in \mathbb{N}$, converges to $e^{-\lambda(B)}$, we obtain

$$\lim_{n\to\infty} \mathbb{P}(\xi_n(B) = k) = \frac{\lambda(B)}{k!}\, e^{-\lambda(B)}$$

for all $k \in \mathbb{N}_0$. The result follows by applying Lemma 5.2.2.

Conversely, let us assume $\xi_n \xrightarrow{d} \gamma$ for some Poisson process γ with intensity measure λ. It follows from Lemma 5.2.2 that, for any open set $B \in \mathcal{U}$, $\xi_n(B) \xrightarrow{d} \gamma(B)$ so that $\xi_n(B), n \in \mathbb{N}$, is tight and

$$\begin{aligned}0 &= k\,\mathbb{P}(\gamma(B) = k) - \lambda(B)\mathbb{P}(\gamma(B) = k-1) \\ &= \lim_{n\to\infty} k\,\mathbb{P}(\xi_n(B) = k) - \lambda(B)\mathbb{P}(\xi_n(B) = k-1)\end{aligned}$$

for $k \in \mathbb{N}$, which shows (i). □

Finally, we derive Theorem 5.1.4 from Theorem 5.1.1.

Proof of Theorem 5.1.4. By (5.6) and the Markov inequality we deduce that $\xi_t(B), t \geq 1$, is tight for all open $B \in \mathcal{U}$. Let $f : S \times \mathcal{N}(S) \to [0, \infty)$ be the function given by

$$f(x, \mu) = \mathbf{1}_B(x)\mathbf{1}\{\mu(B) = k\}$$

for $k \in \mathbb{N}$ and $B \in \mathcal{U}$. Then, by applying the Mecke equation (if $\zeta_t = \eta_t$) and the identity

$$\mathbb{E} \sum_{x \in \beta_n} u(x, \beta_n) = n \int_Y \mathbb{E}[u(x, \beta_{n-1} + \delta_x)] dQ_n(x)$$

for any measurable function $u : Y \times \mathbf{N}_Y \to [0, \infty)$ (if $\zeta_t = \beta_{\lfloor t \rfloor}$), we obtain

$$\begin{aligned} k\mathbb{P}(\xi_t(B) = k) &= \mathbb{E} \sum_{z \in \xi_t} f(z, \xi_t) = \mathbb{E} \sum_{x \in \zeta_t \cap U_t} f(g_t(x, \zeta_t), \xi_t(\zeta_t)) \\ &= \int_{U_t} \mathbb{E}\Big[\mathbf{1}\{g_t(x, \hat{\zeta}_t + \delta_x) \in B\}\mathbf{1}\Big\{ \sum_{y \in \hat{\zeta}_t \cap U_t} \delta_{g_t(y, \hat{\zeta}_t + \delta_x)}(B) = k - 1\Big\}\Big] dK_t(x). \end{aligned}$$

Thus, Theorem 5.1.1 yields the equivalence between (5.7) and the convergence in distribution of $\xi_t, t \geq 1$, to a Poisson process with intensity measure M. □

5.3 Applications

All our examples throughout this section concern locally finite point processes on $\mathbb{R}$. These are constructed from the starting points of the k-runs in a sequence of Bernoulli random variables, and from circumscribed radii or inradii of an inhomogeneous Poisson-Voronoi tessellation. We show the Poisson convergence of these point processes using our general criteria Theorem 5.1.1 and Theorem 5.1.4. By Remark 5.1.5, it is sufficient for the convergence of such point processes to a Poisson process on $\mathbb{R}$ with absolutely continuous locally finite intensity measure to show (5.2) or (5.6) and (5.7) for all sets from $\mathcal{I}$, i.e. for all finite unions of open and bounded intervals.

5.3.1 k-runs in a sequence of Bernoulli random variables

Consider a sequence of Bernoulli random variables. A k-head run is defined as an uninterrupted sequence of k successes, where k is a positive integer. In Subsection 3.3.4 we derived Poisson approximation results for the number of k-runs in a sequence of n i.i.d. Bernoulli random variables. Here, we are interested on the limiting behavior of the point process constructed from the starting points of the k-runs in a sequence of Bernoulli random variables. Let the starting point of a k-head run be the index of its first success. Our goal is to find explicit conditions under which the point process of rescaled starting points of the k-head runs converges weakly to a Poisson process. Our investigation relies on two assumptions: the probability of having a k-head run is the same for all k consecutive elements of the sequence, and the Bernoulli random variables are independent if *far away*. We will see that if these conditions are satisfied and if the probability of having a k-head run goes to 0 slower than the probability of having a k-head run with at least another k-head run *nearby*, then the aforementioned point process converges in distribution to a Poisson process.

Let us now give a precise formulation of our result. Let $X_i^{(n)}$, $i, n \in \mathbb{N}$, be an array of Bernoulli distributed random variables and let $k \in \mathbb{N}$. Assume that there exists a function $f : \mathbb{N} \to \mathbb{N}$ such that for all $q, n \in \mathbb{N}$ the random variable $X_q^{(n)}$ is independent of $\{X_\ell^{(n)} : |q-\ell| \geq f(n),\, \ell \in \mathbb{N}\}$ and that

$$y_n := \mathbb{P}\big(X_q^{(n)} = 1, \dots, X_{q+k-1}^{(n)} = 1\big) > 0$$

does not depend on q. If $X_i^{(n)}$, $i \in \mathbb{N}$, are i.i.d. for $n \in \mathbb{N}$, then $y_n = p_n^k$ with $p_n := \mathbb{P}\big(X_1^{(n)} = 1\big)$. Define

$$I_i^{(n)} = \mathbf{1}\big\{X_i^{(n)} = 1, \dots, X_{i+k-1}^{(n)} = 1\big\}, \quad i \in \mathbb{N}.$$

Let ξ_n be the point process of the k-head runs for $X_i^{(n)}, i \in \mathbb{N}$, that is

$$\xi_n = \sum_{i=1}^{\infty} I_i^{(n)} \delta_{iy_n}. \tag{5.20}$$

For any $i_0 \in \mathbb{N}$, let

$$W_{i_0}^{(n)} = \sum_{j \in \mathbb{N}\,:\, 1 \leq |j-i_0| \leq f(n)+k-2} I_j^{(n)}.$$

We denote by λ_1 the restriction of the Lebesgue measure to $[0, \infty)$.

Theorem 5.3.1. *Let ξ_n, $n \in \mathbb{N}$, be the sequence of point processes given by* (5.20)*. Assume that $f(n)y_n \to 0$ and that*

$$\lim_{n \to \infty} \sup_{i \in \mathbb{N}} y_n^{-1} \mathbb{E}\big[I_i^{(n)} \mathbf{1}\{W_i^{(n)} > 0\}\big] = 0. \tag{5.21}$$

Then ξ_n converges weakly to a Poisson process with intensity measure λ_1.

For underlying independent Bernoulli random variables, the Poisson approximation of the random variable $\xi_n((0, u))$, $u > 0$, is considered in Subsection 3.3.4 and [2, 10, 27, 35], and the Poisson process convergence follows from the results of [2]. Quantitative bounds for the Poisson process approximation of 2-runs in the i.i.d. case were derived in [63, Proposition 3.C] and [71, Theorem 6.3]; see also [17, Subsection 3.5], where the Poisson process approximation for the more general problem of counting rare words is considered.

As a consequence of Theorem 5.3.1, we can study the limiting distribution of

$$T_n = \min\{i \in \mathbb{N} \,:\, I_i^{(n)} = 1\},$$

which gives the first arrival time of a k-head run for a sequence of Bernoulli random variables.

Corollary 5.3.2. *If the assumptions of Theorem 5.3.1 are satisfied, then $y_n T_n$ converges in distribution to an exponentially distributed random variable with parameter* 1*.*

Clearly, in the case when the Bernoulli random variables $(X_i^{(n)})_{i \in \mathbb{N}}$ are i.i.d. with parameter $p_n > 0$, if p_n converges to 0, the assumptions of Theorem 5.3.1 are fulfilled with $f(n) \equiv 1$, and so ξ_n converges in distribution to a Poisson process. Other conditions for weak convergence are given in the following corollary. This result can also be shown by applying [62, Theorem 2.1] to the restriction of ξ_n on $(0, u)$ for each $u > 0$.

Corollary 5.3.3. *Let ξ_n, $n \in \mathbb{N}$, be the sequence of point processes given by* (5.20)*. Let us assume that $f(n)y_n \to 0$ and*

$$\lim_{n\to\infty} \sup_{i\in\mathbb{N}} y_n^{-1} \sum_{j\in\mathbb{N}\,:\,1\le|i-j|\le f(n)+k-2} \mathbb{E}[I_i^{(n)} I_j^{(n)}] = 0.$$

Then ξ_n converges weakly to a Poisson process with intensity measure λ_1.

Let us now prove the main result of this subsection, Theorem 5.3.1.

Proof of Theorem 5.3.1. For any bounded interval $A \subset [0,\infty)$, the assumptions on $X_i^{(n)}$, $i \in \mathbb{N}$, imply that

$$\mathbb{E}[\xi_n(A)] = y_n \sum_{i=1}^{\infty} \delta_{iy_n}(A) = (\sup(A)y_n^{-1} + b_n)y_n - (\inf(A)y_n^{-1} + a_n)y_n$$

for some $a_n, b_n \in [-1,1]$. By $y_n \to 0$, we have $\mathbb{E}[\xi_n(A)] \to \lambda_1(A)$ and, consequently, $\mathbb{E}[\xi_n(B)] \to \lambda_1(B)$ for all $B \in \mathcal{I}$. Moreover, $\xi_n(B)$, $n \in \mathbb{N}$, is tight (see Remark 5.1.2). Then, we can write $\xi_n(B)$ as

$$\xi_n(B) = \sum_{i\in\mathcal{A}_n} I_i^{(n)}$$

with $\mathcal{A}_n = \{i \in \mathbb{N} : iy_n \in B\}$.

For $i_0 \in \mathcal{A}_n$, we have for any $m \in \mathbb{N}$ that

$$\begin{aligned}&\big|\mathbb{E}\big[I_{i_0}^{(n)}\mathbf{1}\big\{\xi_n(B) - I_{i_0}^{(n)} = m-1\big\}\big] - \mathbb{E}\big[I_{i_0}^{(n)}\mathbf{1}\big\{\xi_n(B) - W_{i_0}^{(n)} - I_{i_0}^{(n)} = m-1\big\}\big]\big|\\ &\le \mathbb{E}\big[I_{i_0}^{(n)}\mathbf{1}\{W_{i_0}^{(n)} > 0\}\big].\end{aligned}$$

Together with $\mathbb{E}\big[\xi_n(B)\big] = |\mathcal{A}_n|y_n$, this yields

$$\begin{aligned}H_n := \Big| &\sum_{i\in\mathcal{A}_n} \mathbb{E}\big[I_i^{(n)}\mathbf{1}\big\{\xi_n(B) - I_i^{(n)} = m-1\big\}\big]\\ &- \sum_{i\in\mathcal{A}_n} \mathbb{E}\big[I_i^{(n)}\mathbf{1}\big\{\xi_n(B) - W_i^{(n)} - I_i^{(n)} = m-1\big\}\big]\Big|\\ \le &\sum_{i\in\mathcal{A}_n} \mathbb{E}\big[I_i^{(n)}\mathbf{1}\{W_i^{(n)} > 0\}\big] \le \Big(\sup_{i\in\mathbb{N}} y_n^{-1}\mathbb{E}\big[I_i^{(n)}\mathbf{1}\{W_i^{(n)} > 0\}\big]\Big)\mathbb{E}[\xi_n(B)].\end{aligned}$$

Therefore from (5.21), we obtain $H_n \to 0$. From the independence of $I_{i_0}^{(n)}$ and $\xi_n(B) - W_{i_0}^{(n)} - I_{i_0}^{(n)}$ for $i_0 \in \mathcal{A}_n$, it follows that

$$\mathbb{E}\big[I_{i_0}^{(n)}\mathbf{1}\big\{\xi_n(B) - W_{i_0}^{(n)} - I_{i_0}^{(n)} = m-1\big\}\big] = \mathbb{E}\big[I_{i_0}^{(n)}\big]\mathbb{P}\big(\xi_n(B) - W_{i_0}^{(n)} - I_{i_0}^{(n)} = m-1\big).$$

Combining the previous arguments implies for $m \in \mathbb{N}$ that

$$\begin{aligned}&\limsup_{n\to\infty} \big|m\mathbb{P}(\xi_n(B) = m) - \lambda_1(B)\mathbb{P}(\xi_n(B) = m-1)\big|\\ &= \limsup_{n\to\infty} \Big|\sum_{i\in\mathcal{A}_n} \mathbb{E}\big[I_i^{(n)}\mathbf{1}\big\{\xi_n(B) - I_i^{(n)} = m-1\big\}\big] - \lambda_1(B)\mathbb{P}(\xi_n(B) = m-1)\Big|\\ &= \limsup_{n\to\infty} \Big|\sum_{i\in\mathcal{A}_n} \mathbb{E}\big[I_i^{(n)}\mathbf{1}\big\{\xi_n(B) - W_i^{(n)} - I_i^{(n)} = m-1\big\}\big] - \mathbb{E}[\xi_n(B)]\mathbb{P}(\xi_n(B) = m-1)\Big|\\ &= \limsup_{n\to\infty} \Big|\sum_{i\in\mathcal{A}_n} \mathbb{E}[I_i^{(n)}]\mathbb{P}\big(\xi_n(B) - W_i^{(n)} - I_i^{(n)} = m-1\big) - \sum_{i\in\mathcal{A}_n} \mathbb{E}[I_i^{(n)}]\mathbb{P}(\xi_n(B) = m-1)\Big|\\ &\le \limsup_{n\to\infty} \sum_{i\in\mathcal{A}_n} \mathbb{E}[I_i^{(n)}]\mathbb{P}\big(W_i^{(n)} + I_i^{(n)} > 0\big) \le \lambda_1(B) \limsup_{n\to\infty} \sup_{i\in\mathbb{N}} \mathbb{P}\big(W_i^{(n)} + I_i^{(n)} > 0\big).\end{aligned}$$

Finally, the inequality

$$\mathbb{P}\big(W_i^{(n)} + I_i^{(n)} > 0\big) \leq (2k + 2f(n) - 3)y_n, \quad i \in \mathbb{N},$$

and the assumption $f(n)y_n \to 0$ lead to

$$\lim_{n\to\infty} \big|m\mathbb{P}(\xi_n(B) = m) - \lambda_1(B)\mathbb{P}(\xi_n(B) = m - 1)\big| = 0.$$

The result follows by applying Theorem 5.1.1. □

Proof of Corollary 5.3.3. This follows directly from Theorem 5.3.1 and

$$\mathbb{E}\big[I_i \mathbf{1}\{W_i^{(n)} > 0\}\big] \leq \mathbb{E}\big[I_i^{(n)} W_i^{(n)}\big] = \sum_{j\in\mathbb{N}\,:\,1\leq|i-j|\leq f(n)+k-2} \mathbb{E}\big[I_i^{(n)} I_j^{(n)}\big]$$

for any $i \in \mathbb{N}$. □

5.3.2 Inradii of an inhomogeneous Poisson-Voronoi tessellation

In this subsection, we consider the inradii of an inhomogeneous Voronoi tessellation generated by a Poisson process with a certain intensity measure $t\mu, t > 0$; recall that the inradius of a cell is the largest radius for which the ball centered at the nucleus is contained in the cell. We study the point process on $\mathbb{R}$ constructed by taking for any cell with the nucleus in a compact set, a transform of the μ-measure of the ball centered at the nucleus and with twice the inradius as the radius. In Subsection 3.3.6, we proved that, for the stationary case, the Kolmogorov distance between a transform of the largest inradius and a Gumbel random variable converges to 0 at a rate of $\log(t)/\sqrt{t}$ as the intensity t of the underlying Poisson process goes to infinity. Now, we aim to continue the work in [15] by extending the result on the largest inradius to inhomogeneous Poisson-Voronoi tessellations and proving weak convergence of the aforementioned point process to a Poisson process. More details on Poisson-Voronoi tessellations are given in Subsection 3.3.5. Recall that, for any locally finite counting measure ν on $\mathbb{R}^d$, we denote by $N(x,\nu)$ the Voronoi cell with nucleus $x \in \mathbb{R}^d$ generated by $\nu + \delta_x$, that is

$$N(x,\nu) = \Big\{y \in \mathbb{R}^d \,:\, \|x - y\| \leq \|y - x'\|, x \neq x' \in \nu\Big\},$$

where $\|\cdot\|$ denotes the Euclidean norm. The inradius of the Voronoi cell $N(x,\nu)$ is given by

$$c(x,\nu) = \sup\{R \geq 0 \,:\mathbf{B}(x,R) \subset N(x,\nu)\},$$

where $\mathbf{B}(x,r)$ denotes the open ball centered at $x \in \mathbb{R}^d$ with radius $r > 0$.

Let η_t, $t > 0$, be a Poisson process on $\mathbb{R}^d$ with intensity measure $t\mu$, where μ is a locally finite measure on $\mathbb{R}^d$ with density $f : \mathbb{R}^d \to [0,\infty)$. Consider a compact set $W \subset \mathbb{R}^d$ with $\mu(W) = 1$, and assume that there exists a bounded open set $A \subset \mathbb{R}^d$ with $W \subset A$ such that $f_{min} := \inf_{x\in A} f(x) > 0$ and $f_{max} := \sup_{x\in A} f(x) < \infty$. For any Voronoi cell $N(x,\eta_t)$ with $x \in \eta_t$, we take the μ-measure of the ball around x with twice the inradius as radius, and we define the point process ξ_t on $\mathbb{R}$ as

$$\xi_t = \xi_t(\eta_t) = \sum_{x\in\eta_t\cap W} \delta_{t\mu(\mathbf{B}(x,2c(x,\eta_t)))-\log(t)}. \tag{5.22}$$

Let M be the measure on $\mathbb{R}$ given by $M([u,\infty)) = e^{-u}$ for $u \in \mathbb{R}$.

Theorem 5.3.4. *Let ξ_t, $t > 0$, be the family of point processes on $\mathbb{R}$ given by* (5.22). *Then ξ_t converges in distribution to a Poisson process with intensity measure M.*

This result is obtained by applying Theorem 5.1.4. We believe that an alternative proof can be deduced from [62, Theorem 2.1]. The next theorem shows that if the density function f is Hölder continuous, it is possible to take out the factor 2 from $\mu(\mathbf{B}(x, 2c(x, \eta_t)))$ and to consider $2^d\mu(\mathbf{B}(x, c(x, \eta_t)))$. Recall that a function $h : \mathbb{R}^d \to \mathbb{R}$ is Hölder continuous with exponent $b > 0$ if there exists a constant $C > 0$ such that

$$|h(x) - h(y)| \leq C\|x - y\|^b$$

for all $x, y \in \mathbb{R}^d$. We define the point process $\widehat{\xi}_t$ as

$$\widehat{\xi}_t = \widehat{\xi}_t(\eta_t) = \sum_{x \in \eta_t \cap W} \delta_{2^d t\mu(\mathbf{B}(x,c(x,\eta_t)))-\log(t)}.$$

Theorem 5.3.5. *Let f be Hölder continuous. Then $\widehat{\xi}_t, t > 0$, converges in distribution to a Poisson process with intensity measure M.*

As a corollary of the previous theorems, we have the following generalization to the inhomogeneous case of the result obtained in [15, Theorem 1, Equation (2a)] for the stationary case; see also [19, Section 5] for the maximal inradius of a stationary Poisson-Voronoi tessellation and of a stationary Gauss-Poisson-Voronoi tessellation.

Corollary 5.3.6. *For $u \in \mathbb{R}$,*

$$\lim_{t\to\infty} \mathbb{P}\big(\max_{x\in\eta_t\cap W} t\mu(\mathbf{B}(x, 2c(x, \eta_t))) - \log(t) \leq u\big) = e^{-e^{-u}}. \tag{5.23}$$

Moreover, if f is Hölder continuous,

$$\lim_{t\to\infty} \mathbb{P}\big(\max_{x\in\eta_t\cap W} 2^d t\mu(\mathbf{B}(x, c(x, \eta_t))) - \log(t) \leq u\big) = e^{-e^{-u}}. \tag{5.24}$$

For an underlying binomial point process, (5.23) was shown under similar assumptions in [29]. The related problem of maximal weighted r-th nearest neighbor distances for the points of a binomial point process was studied in [30]; see also [31].

For the proofs of Theorem 5.3.4 and Theorem 5.3.5, we will use the quantities $v_t(x, u)$ and $q_t(x, u)$, which are introduced in the next lemma.

Lemma 5.3.7. *For any $u \in \mathbb{R}$ there exists $t_0 > 0$ such that for all $x \in W$ and $t > t_0$ the equations*

$$t\mu(\mathbf{B}(x, 2v_t(x, u))) = u + \log(t) \quad \text{and} \quad 2^d t\mu(\mathbf{B}(x, q_t(x, u))) = u + \log(t) \tag{5.25}$$

have unique solutions $v_t(x, u)$ and $q_t(x, u)$, respectively, which satisfy

$$\max\{v_t(x, u), q_t(x, u)\} \leq \Big(\frac{u + \log(t)}{2^d f_{min} k_d t}\Big)^{1/d}, \tag{5.26}$$

where k_d is the volume of the d-dimensional unit ball.

Proof. Let $u \in \mathbb{R}$ be fixed and set $m = \inf\{\|x - y\| : x \in \partial W, y \in \partial A\} \in (0, \infty)$. Note that $B(x, m) \subset A$ for all $x \in W$. Choose $t_0 > 0$ such that

$$\frac{u + \log(t)}{t} < f_{min} k_d m^d$$

for $t > t_0$. For $x \in W$ and $t > t_0$ this implies that

$$2^d t\mu(\mathbf{B}(x,m)) \geq t\mu(\mathbf{B}(x,m)) \geq t f_{min} k_d m^d > u + \log(t)$$

and, obviously, $t\mu(\mathbf{B}(x,0)) = 0$. Since the function $[0,m] \ni a \to \mu(\mathbf{B}(x,a))$ is continuous and strictly increasing because of $f_{min} > 0$, by the intermediate value theorem, the equations in (5.25) have unique solutions $v_t(x,u)$ and $q_t(x,u)$. Since $\max\{2v_t(x,u), q_t(x,u)\} < m$, we obtain

$$\frac{u + \log(t)}{t} = \mu(\mathbf{B}(x, 2v_t(x,u))) \geq 2^d f_{min} k_d v_t(x,u)^d$$

and

$$\frac{u + \log(t)}{t} = 2^d \mu(\mathbf{B}(x, q_t(x,u))) \geq 2^d f_{min} k_d q_t(x,u)^d,$$

which prove (5.26). □

Let M_t be the intensity measure of ξ_t. Then from the Mecke formula and Lemma 5.3.7 it follows that for any $u \in \mathbb{R}$ there exists $t_0 > 0$ such that for $t > t_0$,

$$\begin{aligned}
M_t([u,\infty)) &= t \int_W \mathbb{P}\big(t\mu(\mathbf{B}(x, 2c(x, \eta_t + \delta_x))) \geq u + \log(t)\big) f(x) dx \\
&= t \int_W \mathbb{P}(c(x, \eta_t + \delta_x) \geq v_t(x,u)) f(x) dx \\
&= t \int_W \mathbb{P}\big(\eta_t(\mathbf{B}(x, 2v_t(x,u))) = 0\big) f(x) dx \\
&= t \int_W e^{-t\mu(\mathbf{B}(x, 2v_t(x,u)))} f(x) dx = t e^{-u - \log(t)} \mu(W) = e^{-u} = M([u,\infty)),
\end{aligned} \tag{5.27}$$

where we used (5.25) and $\mu(W) = 1$ in the last steps. For any $y \in \mathbb{R}$ and point configuration ν on $\mathbb{R}^d$ with $y \in \nu$, we denote by $h_t(y,\nu)$ the quantity

$$h_t(y,\nu) = t\mu(\mathbf{B}(y, 2c(y,\nu))) - \log(t), \tag{5.28}$$

where $c(y,\nu)$ is the inradius of the Voronoi cell with nucleus y generated by ν. So we can rewrite ξ_t as

$$\xi_t = \sum_{x \in \eta_t \cap W} \delta_{h_t(x,\eta_t)}.$$

Proof of Theorem 5.3.4. From Theorem 5.1.4 and (5.27) it follows that it is enough to show that

$$\begin{aligned}
\lim_{t\to\infty} t \int_W \mathbb{E}\Big[\mathbf{1}\{h_t(x, \eta_t + \delta_x) \in B\} \mathbf{1}\Big\{ \sum_{y \in \eta_t \cap W} \delta_{h_t(y, \eta_t + \delta_x)}(B) = m \Big\}\Big] f(x) dx \\
- M(B) \mathbb{P}(\xi_t(B) = m) = 0
\end{aligned} \tag{5.29}$$

for any $m \in \mathbb{N}_0$ and $B \in \mathcal{I}$. Let $B = \bigcup_{j=1}^{\ell} (u_{2j-1}, u_{2j})$ with $u_1 < u_2 < \cdots < u_{2\ell}$ and $\ell \in \mathbb{N}$. By Lemma 5.3.7 there is a $t_0 > 0$ such that $v_t(x, u_k)$ exists for all $k = 1, \ldots, 2\ell$, $x \in W$ and $t > t_0$. Assume $t > t_0$ in the following. Elementary arguments imply that

$$\begin{aligned}
& t \int_W \mathbb{E}\Big[\mathbf{1}\{h_t(x, \eta_t + \delta_x) \in B\} \mathbf{1}\Big\{ \sum_{y \in \eta_t \cap W} \delta_{h_t(y, \eta_t + \delta_x)}(B) = m \Big\}\Big] f(x) dx \\
&= \sum_{j=1}^{\ell} t \int_W \mathbb{E}\Big[\mathbf{1}\{h_t(x, \eta_t + \delta_x) \in (u_{2j-1}, u_{2j})\} \mathbf{1}\Big\{ \sum_{y \in \eta_t \cap W} \delta_{h_t(y, \eta_t + \delta_x)}(B) = m \Big\}\Big] f(x) dx.
\end{aligned} \tag{5.30}$$

For each $k = 1, \dots, 2\ell$, set $w_{k,t,x} = 2v_t(x, u_k)$. Since $h_t(x, \eta_t + \delta_x) \in (u_{2j-1}, u_{2j})$ if and only if $c(x, \eta_t + \delta_x) \in (v_t(x, u_{2j-1}), v_t(x, u_{2j}))$, or equivalently, $\eta_t(\mathbf{B}(x, w_{2j-1,t,x})) = 0$ and $\eta_t(\mathbf{B}(x, w_{2j,t,x})) > 0$, we obtain that

$$\begin{aligned} S_j :&= t \int_W \mathbb{E}\Big[\mathbf{1}\{h_t(x, \eta_t + \delta_x) \in (u_{2j-1}, u_{2j})\}\mathbf{1}\Big\{ \sum_{y \in \eta_t \cap W} \delta_{h_t(y, \eta_t + \delta_x)}(B) = m\Big\}\Big] f(x) dx \\ &= t \int_W \mathbb{E}\Big[\mathbf{1}\left\{\eta_t\big(\mathbf{B}(x, w_{2j-1,t,x})\big) = 0\right\} \\ &\qquad \times \mathbf{1}\Big\{\eta_t\big(\mathbf{B}(x, w_{2j,t,x})\big) > 0, \sum_{y \in \eta_t \cap W} \delta_{h_t(y, \eta_t + \delta_x)}(B) = m\Big\}\Big] f(x) dx. \end{aligned}$$

For any point configuration ν on $\mathbb{R}^d$ and $x \in W$, let $\xi_{t,x}(\nu)$ be the counting measure given by $\xi_{t,x}(\nu) = \sum_{y \in \nu \cap W} \delta_{h_t(y, \nu + \delta_x)}$ so that

$$\begin{aligned} S_j &= t \int_W \mathbb{E}\big[\mathbf{1}\left\{\eta_t\big(\mathbf{B}(x, w_{2j-1,t,x})\big) = 0\right\} \\ &\qquad \times \mathbf{1}\Big\{\eta_t\big(\mathbf{B}(x, w_{2j,t,x})\big) > 0, \xi_{t,x}\big(\eta_t|_{\mathbf{B}(x, w_{2j-1,t,x})^c}\big)(B) = m\Big\}\big] f(x) dx \\ &= t \int_W \mathbb{P}\left(\eta_t\big(\mathbf{B}(x, w_{2j-1,t,x})\big) = 0\right) \\ &\qquad \times \mathbb{P}\big(\eta_t\big(\mathbf{B}(x, w_{2j,t,x}) \setminus \mathbf{B}(x, w_{2j-1,t,x})\big) > 0, \xi_{t,x}\big(\eta_t|_{\mathbf{B}(x, w_{2j-1,t,x})^c}\big)(B) = m\big) f(x) dx. \end{aligned}$$

Similar arguments as used to compute $M_t([u, \infty))$ for $u \in \mathbb{R}$ imply for $x \in W$ that

$$t\mathbb{P}\left(\eta_t\big(\mathbf{B}(x, w_{2j-1,t,x})\big) = 0\right) = e^{-u_{2j-1}},$$

and so we deduce that

$$\begin{aligned} S_j &= e^{-u_{2j-1}} \int_W \mathbb{P}\left(\xi_{t,x}\big(\eta_t|_{\mathbf{B}(x, w_{2j-1,t,x})^c}\big)(B) = m\right) f(x) dx \\ &\quad - e^{-u_{2j-1}} \int_W \mathbb{P}\big(\eta_t\big(\mathbf{B}(x, w_{2j,t,x}) \setminus \mathbf{B}(x, w_{2j-1,t,x})\big) = 0, \\ &\qquad\qquad \xi_{t,x}\big(\eta_t|_{\mathbf{B}(x, w_{2j-1,t,x})^c}\big)(B) = m\big) f(x) dx. \end{aligned} \tag{5.31}$$

Furthermore, we can rewrite the second integral as

$$\begin{aligned} &\int_W \mathbb{P}\left(\eta_t\big(\mathbf{B}(x, w_{2j,t,x}) \setminus \mathbf{B}(x, w_{2j-1,t,x})\big) = 0\right) \mathbb{P}\left(\xi_{t,x}\big(\eta_t|_{\mathbf{B}(x, w_{2j,t,x})^c}\big)(B) = m\right) f(x) dx \\ &= \int_W e^{-t\mu(\mathbf{B}(x, w_{2j,t,x})) + t\mu(\mathbf{B}(x, w_{2j-1,t,x}))} \mathbb{P}\left(\xi_{t,x}\big(\eta_t|_{\mathbf{B}(x, w_{2j,t,x})^c}\big)(B) = m\right) f(x) dx \\ &= e^{-u_{2j} + u_{2j-1}} \int_W \mathbb{P}\left(\xi_{t,x}\big(\eta_t|_{\mathbf{B}(x, w_{2j,t,x})^c}\big)(B) = m\right) f(x) dx. \end{aligned}$$

Combining this and (5.31) yields

$$\begin{aligned} S_j &= e^{-u_{2j-1}} \int_W \mathbb{P}\left(\xi_{t,x}\big(\eta_t|_{\mathbf{B}(x, w_{2j-1,t,x})^c}\big)(B) = m\right) f(x) dx \\ &\quad - e^{-u_{2j}} \int_W \mathbb{P}\left(\xi_{t,x}\big(\eta_t|_{\mathbf{B}(x, w_{2j,t,x})^c}\big)(B) = m\right) f(x) dx. \end{aligned}$$

Substituting this into (5.30) implies that to prove (5.29) and to complete the proof, it is enough to show for all $x \in W$ and $k = 1, \dots, 2\ell$ that

$$\lim_{t\to\infty} \mathbb{P}\left(\xi_{t,x}\big(\eta_t|_{\mathbf{B}(x,w_{k,t,x})^c}\big)(B) = m\right) - \mathbb{P}(\xi_t(B) = m) = 0. \tag{5.32}$$

Let $x \in W$, $k \in \{1, \dots, 2\ell\}$ and $\varepsilon > 0$ be fixed. Set

$$a_t = 2\Big(\frac{u_{2\ell} + \log(t)}{2^d f_{min} k_d t}\Big)^{1/d}.$$

From the application of Lemma 5.3.7 at the beginning of the proof it follows that $w_{k,t,y} \leq w_{2\ell,t,y} \leq a_t$ for all $y \in W$ and $t > t_0$. Without loss of generality we may assume that $2a_t < \min\{\|z_1 - z_2\| : z_1 \in \partial W, z_2 \in \partial A\}$. Therefore the observation

$$h_t(y, \nu) \in B \quad \text{if and only if} \quad h_t(y, \nu|_{\mathbf{B}(y,w_{2\ell,t,y})}) \in B$$

for any point configuration ν on $\mathbb{R}^d$ and $y \in \nu \cap W$ leads to

$$\begin{aligned}
&\Big|\mathbb{P}\left(\xi_{t,x}\big(\eta_t|_{\mathbf{B}(x,w_{k,t,x})^c}\big)(B) = m\right) - \mathbb{P}\left(\xi_t(B) = m\right)\Big| \leq \mathbb{E}\big[|\xi_{t,x}\big(\eta_t|_{\mathbf{B}(x,w_{k,t,x})^c}\big)(B) - \xi_t(B)|\big] \\
&\leq \mathbb{E} \sum_{y\in\eta_t\cap\mathbf{B}(x,2a_t)\cap\mathbf{B}(x,w_{k,t,x})^c\cap W} \mathbf{1}\{h_t(y, \eta_t|_{\mathbf{B}(x,w_{k,t,x})^c} + \delta_x) \in B\} \\
&\quad + \mathbb{E} \sum_{y\in\eta_t\cap\mathbf{B}(x,2a_t)\cap W} \mathbf{1}\{h_t(y, \eta_t) \in B\} \\
&\leq \mathbb{E} \sum_{y\in\eta_t\cap\mathbf{B}(x,2a_t)\cap\mathbf{B}(x,w_{k,t,x})^c\cap W} \mathbf{1}\{h_t(y, \eta_t|_{\mathbf{B}(x,w_{k,t,x})^c} + \delta_x) > u_1\} \\
&\quad + \mathbb{E} \sum_{y\in\eta_t\cap\mathbf{B}(x,2a_t)\cap W} \mathbf{1}\{h_t(y, \eta_t) > u_1\}.
\end{aligned}$$

Then, the Mecke formula and (5.28) imply that

$$\begin{aligned}
&\Big|\mathbb{P}\left(\xi_{t,x}\big(\eta_t|_{\mathbf{B}(x,w_{k,t,x})^c}\big)(B) = m\right) - \mathbb{P}\left(\xi_t(B) = m\right)\Big| \\
&\leq t \int_{\mathbf{B}(x,2a_t)\cap\mathbf{B}(x,w_{k,t,x})^c\cap W} \mathbb{P}(h_t(y, \eta_t|_{\mathbf{B}(x,w_{k,t,x})^c} + \delta_x + \delta_y) > u_1) f(y) dy \\
&\quad + t \int_{\mathbf{B}(x,2a_t)\cap W} \mathbb{P}(h_t(y, \eta_t + \delta_y) > u_1) f(y) dy \\
&= t \int_{\mathbf{B}(x,2a_t)\cap\mathbf{B}(x,w_{k,t,x})^c\cap W} \mathbb{P}\big(t\mu(\mathbf{B}(y, 2c(y, \eta_t|_{\mathbf{B}(x,w_{k,t,x})^c} + \delta_x + \delta_y))) > u_1 + \log(t)\big) f(y) dy \\
&\quad + t \int_{\mathbf{B}(x,2a_t)\cap W} \mathbb{P}\big(t\mu(\mathbf{B}(y, 2c(y, \eta_t + \delta_y))) > u_1 + \log(t)\big) f(y) dy.
\end{aligned}$$

Since $c(y, \nu+\delta_y+\delta_x) > v_t(y, u_1)$ only if $c(y, \nu+\delta_y) > v_t(y, u_1)$ for any point configuration ν on $\mathbb{R}^d$ and $x, y \in W$, it follows for $x \in W$ and $y \in \mathbf{B}(x, 2a_t) \cap \mathbf{B}(x, w_{k,t,x})^c \cap W$ that

$$\begin{aligned}
&\mathbb{P}\big(t\mu(\mathbf{B}(y, 2c(y, \eta_t|_{\mathbf{B}(x,w_{k,t,x})^c} + \delta_x + \delta_y))) > u_1 + \log(t)\big) \\
&= \mathbb{P}\big(c(y, \eta_t|_{\mathbf{B}(x,w_{k,t,x})^c} + \delta_x + \delta_y) > v_t(y, u_1)\big) \\
&\leq \mathbb{P}\big(c(y, \eta_t|_{\mathbf{B}(x,w_{k,t,x})^c} + \delta_y) > v_t(y, u_1)\big) \\
&= \mathbb{P}\big(\eta_t|_{\mathbf{B}(x,w_{k,t,x})^c}(\mathbf{B}(y, 2v_t(y, u_1))) = 0\big) = \exp(-t\mu(\mathbf{B}(y, 2v_t(y, u_1)) \cap \mathbf{B}(x, w_{k,t,x})^c)).
\end{aligned}$$

Let λ_d denote the Lebesgue measure on $\mathbb{R}^d$. For $A_1, A_2 \in \mathcal{B}(\mathbb{R}^d)$ with $A_1, A_2 \subset A$ and $\lambda_d(A_2) > 0$ we obtain

$$\frac{\mu(A_1)}{\mu(A_2)} \geq \frac{f_{min}}{f_{max}} \frac{\lambda_d(A_1)}{\lambda_d(A_2)}.$$

With $\tau := f_{min}/f_{max} \in (0, 1]$,

$$A_1 = \mathbf{B}(y, 2v_t(y, u_1)) \cap \mathbf{B}(x, w_{k,t,x})^c \quad \textit{and} \quad A_2 = \mathbf{B}(y, 2v_t(y, u_1)),$$

this implies for $x \in W$ and $y \in \mathbf{B}(x, 2a_t) \cap \mathbf{B}(x, w_{k,t,x})^c \cap W$ that

$$t\mu(\mathbf{B}(y, 2v_t(y, u_1)) \cap \mathbf{B}(x, w_{k,t,x})^c) \geq \frac{\tau}{2} t\mu(\mathbf{B}(y, 2v_t(y, u_1))) = \frac{\tau}{2}(u_1 + \log(t)).$$

Moreover, we have that

$$\mathbb{P}\big(t\mu(\mathbf{B}(y, 2c(y, \eta_t + \delta_y))) > u_1 + \log(t)\big) = \mathbb{P}\big(\eta_t(\mathbf{B}(y, 2v_t(y, u_1))) = 0\big) = e^{-u_1 - \log(t)}.$$

In conclusion, combining the previous bounds leads to

$$\begin{aligned}
&\left|\mathbb{P}\left(\xi_{t,x}\big(\eta_t|_{\mathbf{B}(x,w_{k,t,x})^c}\big)(B) = m\right) - \mathbb{P}\left(\xi_t(B) = m\right)\right| \\
&\leq t^{1-\tau/2} e^{-\tau u_1/2} \mu(\mathbf{B}(x, 2a_t)) + e^{-u_1} \mu(\mathbf{B}(x, 2a_t)) \leq (2a_t)^d k_d f_{max} (t^{1-\tau/2} e^{-\tau u_1/2} + e^{-u_1}),
\end{aligned}$$

where in the last step we used the fact that f is bounded by f_{max} in A and, by the choice of a_t, $\mathbf{B}(x, 2a_t) \subset A$. Again, from the definition of a_t it follows that the right-hand side converges to 0 as $t \to \infty$. This shows (5.32) and concludes the proof. □

Next, we derive Theorem 5.3.5 from Theorem 5.3.4.

Proof of Theorem 5.3.5. Assume that f is Hölder continuous with exponent $b > 0$. From Lemma 5.2.2, Theorem 5.3.4 and Remark 5.1.5 we obtain that it is enough to show that $\mathbb{E}[|\xi_t(B) - \widehat{\xi}_t(B)|] \to 0$ as $t \to \infty$ for all $B \in \mathcal{I}$. By Lemma 5.3.7, for any $u \in \mathbb{R}$ there exists $t_0 > 0$ such that

$$\begin{aligned}
&\mu(\mathbf{B}(x, 2v_t(x, u))) = 2^d \mu(\mathbf{B}(x, q_t(x, u))) \\
&= 2^d k_d f(x) q_t(x, u)^d + 2^d \int_{\mathbf{B}(x, q_t(x,u))} (f(y) - f(x)) dy \\
&= \mu(\mathbf{B}(x, 2q_t(x, u))) - \int_{\mathbf{B}(x, 2q_t(x,u))} (f(y) - f(x)) dy + 2^d \int_{\mathbf{B}(x, q_t(x,u))} (f(y) - f(x)) dy
\end{aligned}$$

for all $x \in W, t > t_0$, where

$$\max\{v_t(x, u), q_t(x, u)\} \leq \Big(\frac{u + \log(t)}{2^d f_{min} k_d t}\Big)^{1/d}.$$

Thus, the Hölder continuity of f and elementary arguments establish that

$$|\mu(\mathbf{B}(x, 2v_t(x, u))) - \mu(\mathbf{B}(x, 2q_t(x, u)))| \leq C \Big(\frac{u + \log(t)}{t}\Big)^{1+b/d}, \quad x \in W, t > t_0, \tag{5.33}$$

for some $C > 0$. In particular, from the definition of $v_t(x, u)$ it follows that

$$\begin{aligned}
\mu(\mathbf{B}(x, 2v_t(x, u))) &= \frac{u + \log(t)}{t} \\
\mu(\mathbf{B}(x, 2q_t(x, u))) &\geq \frac{u + \log(t)}{t} - C\Big(\frac{u + \log(t)}{t}\Big)^{1+b/d}
\end{aligned} \tag{5.34}$$

for $t > t_0$. Next, we write $B = \bigcup_{j=1}^{\ell}(u_{2j-1}, u_{2j})$ for some $\ell \in \mathbb{N}$ and $u_1 < \cdots < u_{2\ell}$. The triangle inequality yields

$$\begin{aligned} \mathbb{E}[|\xi_t(B) - \widehat{\xi}_t(B)|] &\leq \sum_{j=1}^{\ell} \mathbb{E}[|\xi_t((u_{2j-1}, u_{2j})) - \widehat{\xi}_t((u_{2j-1}, u_{2j}))|] \\ &\leq \sum_{j=1}^{\ell} \mathbb{E}[|\xi_t((u_{2j-1}, \infty)) - \widehat{\xi}_t((u_{2j-1}, \infty))|] + \mathbb{E}[|\xi_t([u_{2j}, \infty)) - \widehat{\xi}_t([u_{2j}, \infty))|]. \end{aligned} \tag{5.35}$$

Moreover, the Mecke formula establishes for $u \in \mathbb{R}$ that

$$\begin{aligned} &\mathbb{E}[|\xi_t((u, \infty)) - \widehat{\xi}_t((u, \infty))|] \\ &\leq \mathbb{E} \sum_{x \in \eta_t \cap W} |\mathbf{1}\{c(x, \eta_t + \delta_x) > v_t(x, u)\} - \mathbf{1}\{c(x, \eta_t + \delta_x) > q_t(x, u)\}| \\ &= t \int_W \mathbb{P}\big(\eta_t(\mathbf{B}(x, 2v_t(x, u))) = 0, \eta_t(\mathbf{B}(x, 2q_t(x, u))) > 0\big) f(x) dx \\ &\quad + t \int_W \mathbb{P}\big(\eta_t(\mathbf{B}(x, 2v_t(x, u))) > 0, \eta_t(\mathbf{B}(x, 2q_t(x, u))) = 0\big) f(x) dx \\ &\leq f_{max} t \int_W \big[\exp\big(-t\mu(\mathbf{B}(x, 2v_t(x, u)))\big) + \exp\big(-t\mu(\mathbf{B}(x, 2q_t(x, u)))\big)\big] \\ &\qquad\qquad \times \big[1 - \exp\big(-t|\mu(\mathbf{B}(x, 2q_t(x, u))) - \mu(\mathbf{B}(x, 2v_t(x, u)))|\big)\big] dx. \end{aligned}$$

Therefore, from (5.33) and (5.34), it follows that

$$\lim_{t \to \infty} \mathbb{E}[|\xi_t((u, \infty)) - \widehat{\xi}_t((u, \infty))|] = 0. \tag{5.36}$$

Together with (5.35) and a similar computation for the half-closed intervals on the right-hand side of (5.35), this concludes the proof. □

Proof of Corollary 5.3.6. Let $u \in \mathbb{R}$ be fixed. By Markov's inequality we have for $u_0 > u$ that

$$\begin{aligned} \mathbb{P}(\xi_t((u, u_0)) > 0) \leq \mathbb{P}(\xi_t((u, \infty)) > 0) &= \mathbb{P}\big(\max_{x \in \eta_t \cap W} t\mu(\mathbf{B}(x, 2c(x, \eta_t))) - \log(t) > u\big) \\ &\leq \mathbb{P}(\xi_t((u, u_0)) > 0) + \mathbb{E}[\xi_t([u_0, \infty))]. \end{aligned}$$

Thus, Theorem 5.3.4 and (5.27) yield

$$\limsup_{t \to \infty} |\mathbb{P}\big(\max_{x \in \eta_t \cap W} t\mu(\mathbf{B}(x, 2c(x, \eta_t))) - \log(t) > u\big) - 1 + e^{-M((u, u_0))}| \leq e^{-u_0}.$$

Then, letting $u_0 \to \infty$ leads to (5.23). Since, for $u > 0$,

$$\big|\mathbb{P}(\xi_t((u, \infty)) > 0) - \mathbb{P}(\widehat{\xi}_t((u, \infty)) > 0)\big| \leq \mathbb{E}[|\xi_t((u, \infty)) - \widehat{\xi}_t((u, \infty))|],$$

(5.23) and (5.36) imply (5.24). □

5.3.3 Circumscribed radii of an inhomogeneous Poisson-Voronoi tessellation

In this last subsection, we consider the circumscribed radii of an inhomogeneous Voronoi tessellation generated by a Poisson process with a certain intensity measure $t\mu, t > 0$;

recall that the circumscribed radius of a cell is the smallest radius for which the ball centered at the nucleus contains the cell. We study the point process on the non-negative real line constructed by taking for any cell with the nucleus in a compact convex set, a transform of the μ-measure of the ball centered at the nucleus and with the circumscribed radius as the radius. In Subsection 3.3.5, we proved that, for the stationary case, the Kolmogorov distance between a transform of the minimal circumscribed radius and a Weibull random variable converges to 0 at a rate of $1/t^{1/(d+1)}$ when the intensity t of the underlying Poisson process goes to infinity. In this subsection, we continue the work started in [15] by extending the result on the smallest circumscribed radius to inhomogeneous Poisson-Voronoi tessellations and by proving weak convergence of the aforementioned point process to a Poisson process.

More precisely, let μ be an absolutely continuous measure on $\mathbb{R}^d$ with continuous density $f : \mathbb{R}^d \to [0, \infty)$. Consider a Poisson process η_t with intensity measure $t\mu$, $t > 0$. The circumscribed radius of the Voronoi cell $N(x, \eta_t)$ with $x \in \eta_t$ is given by

$$C(x, \eta_t) = \inf \{R \geq 0 \, : \, \mathbf{B}(x, R) \supset N(x, \eta_t)\},$$

with the convention $\inf \emptyset = \infty$; see Subsection 3.3.5 for more details on Voronoi tessellations.

Let $W \subset \mathbb{R}^d$ be a compact convex set with $f > 0$ on W. We consider the point process

$$\xi_t = \sum_{x \in \eta_t \cap W} \delta_{\alpha_2 t^{(d+2)/(d+1)} \mu(\mathbf{B}(x, C(x,\eta_t)))}. \tag{5.37}$$

Here the positive constant α_2 is given by

$$\alpha_2 = \left(\frac{2^{d(d+1)}}{(d+1)!} p_{d+1} \right)^{1/(d+1)}$$

with

$$p_{d+1} := \mathbb{P}\Big(N\Big(0, \sum_{j=1}^{d+1} \delta_{Y_j}\Big) \subseteq \mathbf{B}(0,1) \Big),$$

where $Y_1, \dots, Y_{d+1}$ are independent and uniformly distributed random points in $\mathbf{B}(0, 2)$. We write M for the measure on $[0, \infty)$ satisfying $M([0, u]) = \mu(W) u^{d+1}$ for $u \geq 0$.

Theorem 5.3.8. *Let ξ_t, $t > 0$, be the family of point processes on $[0, \infty)$ given by (5.37). Then ξ_t converges in distribution to a Poisson process with intensity measure M.*

This result is obtained by applying Theorem 5.1.4. We believe that an alternative proof can be deduced from [62, Theorem 2.1]. An immediate consequence of this theorem is that a transform of the minimal μ-measure of the balls, having circumscribed radii and nuclei of the Voronoi cells as radii and centers respectively, converges to a Weibull distributed random variable. This generalizes [15, Theorem 1, Equation (2d)]. For the situation that, in contrast to Theorem 5.3.8, the density of the intensity measure of the underlying Poisson process is not continuous, we can still derive some upper and lower bounds.

Theorem 5.3.9. *Let ζ_t be a Poisson process with intensity measure $t\vartheta$, where $t > 0$ and ϑ is an absolutely continuous measure on $\mathbb{R}^d$ with density ϕ. Let $f_1, f_2 : \mathbb{R}^d \to [0, \infty)$ be continuous and $f_1, f_2 > 0$ on W.*

(i) If there exists $s \in (0,1]$ such that $s\phi \le f_1 \le \phi$, then

$$\limsup_{t\to\infty} \mathbb{P}\Big(s\alpha_2 t^{(d+2)/(d+1)} \min_{x\in\zeta_t\cap W} \vartheta(\mathbf{B}(x, C(x,\zeta_t))) > u\Big) \le \exp\big(-s\vartheta(W)u^{d+1}\big)$$

for $u \ge 0$.

(ii) If there exists $r \ge 1$ such that $\phi \le f_2 \le r\phi$, then

$$\liminf_{t\to\infty} \mathbb{P}\Big(r\alpha_2 t^{(d+2)/(d+1)} \min_{x\in\zeta_t\cap W} \vartheta(\mathbf{B}(x, C(x,\zeta_t))) > u\Big) \ge \exp\big(-r\vartheta(W)u^{d+1}\big)$$

for $u \ge 0$.

Let us now prepare the proof of Theorem 5.3.8. We first have to study the distribution of $C(x, \eta_t + \delta_x)$, which is defined as the circumscribed radius of the Voronoi cell with nucleus $x \in \mathbb{R}^d$ generated by $\eta_t + \delta_x$. To this end, we define $g : W \times T \to [0,\infty)$ by the equation

$$\mu(\mathbf{B}(x, g(x,u))) = u \tag{5.38}$$

for $T := [0, \mu(W)]$. Since W is compact and convex and $f > 0$ on W, we have that (5.38) admits a unique solution $g(x,u)$ for all $(x,u) \in W \times T$. As this is the only place where we use the convexity of W, we believe that one can omit this assumption. However, we refrained from doing so in order to not further increase the complexity of the proof. Set

$$s_t = \alpha_2 t^{(d+2)/(d+1)}.$$

Thus, we may write

$$\mathbb{P}(s_t\mu(\mathbf{B}(x, C(x, \eta_t + \delta_x))) \le u) = \mathbb{P}(C(x, \eta_t + \delta_x) \le g(x, u/s_t)), \quad u/s_t \in T. \tag{5.39}$$

Lemma 5.3.10. *For any $u \in T$, $g(\cdot, u) : W \to \mathbb{R}$ is continuous and*

$$\lim_{u\to 0^+} \sup_{x\in W} |g(x,u)| = 0.$$

Proof. First we show that $g(\cdot, u)$ is continuous for any fixed $u \in T$. For $u = 0$, we obtain $g(x,u) = 0$ for all $x \in W$. Assume $u > 0$ and let $x_0 \in W$ and $\varepsilon > 0$. Then for all $x \in \mathbf{B}(x_0, \varepsilon')$ with $\varepsilon' := \min\{g(x_0,u)/2, \varepsilon\}$, we have

$$\mathbf{B}(x_0, g(x_0,u)) \subset \mathbf{B}(x, g(x_0,u) + \varepsilon') \quad \text{and} \quad \mathbf{B}(x, g(x_0,u) - \varepsilon') \subset \mathbf{B}(x_0, g(x_0,u)).$$

Together with (5.38), this leads to

$$\mu(\mathbf{B}(x, g(x_0,u) + \varepsilon')) \ge u \quad \text{and} \quad \mu(\mathbf{B}(x, g(x_0,u) - \varepsilon')) \le u.$$

Now it follows from the definition of g that

$$g(x,u) \le g(x_0,u) + \varepsilon' \quad \text{and} \quad g(x,u) \ge g(x_0,u) - \varepsilon'.$$

This yields

$$|g(x,u) - g(x_0,u)| \le \varepsilon' \le \varepsilon$$

for all $x \in \mathbf{B}(x_0, \varepsilon')$ so that g is continuous at x_0. In conclusion since

$$\lim_{u\to 0^+} g(x,u) = 0$$

and $g(x,u_1) < g(x,u_2)$ for all $x \in W$ and $0 \le u_1 < u_2$, Dini's theorem implies that $\sup_{x\in W} |g(x,u)| \to 0$ as $u \to 0$. □

We define

$$\beta = \min_{x \in W} f(x) > 0.$$

Lemma 5.3.11. *There exists $u_0 \in T$ such that*

$$g(x,u) \le \Big(\frac{2u}{\beta k_d}\Big)^{1/d} \tag{5.40}$$

for all $u \in [0,u_0]$ and $x \in W$.

Proof. Since f is continuous and $f > 0$ on W, it follows that

$$\min_{x \in \overline{W + \mathbf{B}(0,\delta)}} f(x) > \frac{\beta}{2}$$

for some $\delta > 0$. Furthermore, by Lemma 5.3.10 we obtain that there exists $u_0 \in T$ such that $g(x,u) \le \delta$ for all $u \in [0,u_0]$ and $x \in W$. Then, we obtain

$$u = \mu(\mathbf{B}(x, g(x,u))) = \int_{\mathbf{B}(x,g(x,u))} f(y)dy \ge \frac{\beta}{2} k_d g(x,u)^d$$

for all $x \in W$ and $u \in [0,u_0]$, which shows (5.40). □

For $x \in W$ and $u \ge 0$, we consider a sequence of independent and identically distributed random points $(X_i^{(x,u)})_{i \in \mathbb{N}}$ in $\mathbb{R}^d$ with distribution

$$\mathbb{P}\big(X_i^{(x,u)} \in E\big) = \frac{\mu(\mathbf{B}(x,2u) \cap E)}{\mu(\mathbf{B}(x,2u))}, \quad i \in \mathbb{N}, E \in \mathcal{B}(\mathbb{R}^d).$$

Recall that, for $k \ge d+1$, $N\Big(x, \sum_{j=1}^k \delta_{X_j^{(x,u)}}\Big)$ denotes the Voronoi cell with nucleus x generated by $X_1^{(x,u)}, \dots, X_k^{(x,u)}$ and x. Then the distribution function of $C(x, \eta_t + \delta_x)$ is equal to

$$\mathbb{P}\big(C(x, \eta_t + \delta_x) \le u\big) = \sum_{k=d+1}^{\infty} \mathbb{P}\big(\eta_t(\mathbf{B}(x,2u)) = k\big) p_k(x,u) \tag{5.41}$$

for $u \ge 0$, with $p_k(x,u)$ defined as

$$p_k(x,u) = \mathbb{P}\Big(N\Big(x, \sum_{j=1}^k \delta_{X_j^{(x,u)}}\Big) \subseteq \mathbf{B}(x,u)\Big).$$

Combining (5.39) and (5.41) establishes for $u/s_t \in T$ that

$$\begin{aligned} &\mathbb{P}(s_t \mu(\mathbf{B}(x, C(x, \eta_t + \delta_x))) \le u) \\ &= \sum_{k=d+1}^{\infty} \mathbb{P}\big(\eta_t(\mathbf{B}(x, 2g(x,u/s_t))) = k\big) p_k(x, g(x,u/s_t)). \end{aligned} \tag{5.42}$$

For $k \in \mathbb{N}$ with $k \ge d+1$, we define the probability

$$p_k = \mathbb{P}\Big(N\Big(0, \sum_{j=1}^k \delta_{Y_j}\Big) \subseteq \mathbf{B}(0,1)\Big),$$

where $Y_1, \dots, Y_k$ are independent and uniformly distributed random points in $\mathbf{B}(0,2)$. As discussed in [14, Section 5.2.3] and [15, Section 3], one can reinterpret p_k as the probability to cover the unit sphere with k independent spherical caps with random radii. In the next lemma, we prove that $p_k(x,r) \to p_k$ as $r \to 0$ for all $x \in W$ and $k \ge d+1$, which together with Lemma 5.3.11 yields $p_k(x, g(u/s_t)) \to p_k$ as $t \to \infty$.

Lemma 5.3.12. *For any $k \geq d+1$ and $x \in W$,*

$$\lim_{r \to 0^+} p_k(x,r) = p_k.$$

Proof. In this proof, to simplify the notation, for any $x \in W$, $k \geq d+1$ and $y_1, \dots, y_k \in \mathbb{R}^d$, we denote by $K_k^{(x)}(y_1, \dots, y_k)$ the Voronoi cell $N\big(x, \sum_{j=1}^k \delta_{y_j}\big)$ with nucleus x generated by $y_1, \dots, y_k$ and x. Thus, we may write

$$p_k(x,r) = \mathbb{P}\big(K_k^{(x)}\big(X_1^{(x,r)}, \dots, X_k^{(x,r)}\big) \subseteq \mathbf{B}(x,r)\big),$$

and so from the independence of $X_1^{(x,r)}, \dots, X_k^{(x,r)}$ it follows that

$$\begin{aligned} p_k(x,r) &= \frac{1}{\mu(\mathbf{B}(x,2r))^k} \int_{\mathbf{B}(x,2r)^k} \mathbf{1}\big\{K_k^{(x)}(z_1, \dots, z_k) \subseteq \mathbf{B}(x,r)\big\} \prod_{i=1}^k f(z_i)\, dz_1 \dots dz_k \\ &= \frac{(2r)^{kd}}{\mu(\mathbf{B}(x,2r))^k} \int_{\mathbf{B}(0,1)^k} \mathbf{1}\big\{K_k^{(x)}(x+2rz_1, \dots, x+2rz_k) \subseteq \mathbf{B}(x,r)\big\} \\ &\qquad\qquad \times \prod_{i=1}^k f(x+2rz_i)\, dz_1 \dots dz_k. \end{aligned}$$

Furthermore, by the definition of $K_k^{(x)}$ we deduce that

$$\mathbf{1}\big\{K_k^{(x)}(x+2rz_1, \dots, x+2rz_k) \subseteq \mathbf{B}(x,r)\big\} = \mathbf{1}\big\{K_k^{(0)}(2z_1, \dots, 2z_k) \subseteq \mathbf{B}(0,1)\big\}$$

for all $z_1, \dots, z_k \in \mathbf{B}(0,1)$, whence

$$\begin{aligned} p_k(x,r) = \frac{(2r)^{kd}}{\mu(\mathbf{B}(x,2r))^k} \int_{\mathbf{B}(0,1)^k} \mathbf{1}\big\{K_k^{(0)}(2z_1, \dots, 2z_k) \subseteq \mathbf{B}(0,1)\big\} \\ \times \prod_{i=1}^k f(x+2rz_i)\, dz_1 \dots dz_k. \end{aligned}$$

Using the dominated convergence theorem for the integral, the continuity of f and

$$\lim_{r \to 0^+} \frac{(2r)^{kd}}{\mu(\mathbf{B}(x,2r))^k} = \frac{1}{k_d^k f(x)^k},$$

we obtain

$$\begin{aligned} \lim_{r \to 0^+} p_k(x,r) &= \frac{1}{k_d^k} \int_{\mathbf{B}(0,1)^k} \mathbf{1}\big\{K_k^{(0)}(2z_1, \dots, 2z_k) \subseteq \mathbf{B}(0,1)\big\}\, dz_1 \dots dz_k \\ &= \frac{1}{(2^d k_d)^k} \int_{\mathbf{B}(0,2)^k} \mathbf{1}\big\{K_k^{(0)}(z_1, \dots, z_k) \subseteq \mathbf{B}(0,1)\big\}\, dz_1 \dots dz_k = p_k, \end{aligned}$$

which concludes the proof. □

Let M_t be the intensity measure of ξ_t and let

$$\begin{aligned} \widehat{M_t}([0,u]) = t \int_W \mathbb{E}\Big[&\mathbf{1}\big\{s_t \mu(\mathbf{B}(x, C(x, \eta_t + \delta_x))) \in [0,u]\big\} \\ &\times \mathbf{1}\Big\{\eta_t\Big(\mathbf{B}\Big(x, 4\Big(\frac{2u}{\beta s_t k_d}\Big)^{1/d}\Big)\Big) = d+1\Big\}\Big] f(x) dx \end{aligned}$$

and

$$\theta_t([0,u]) = t\int_W \mathbb{E}\Big[\mathbf{1}\{s_t\mu(\mathbf{B}(x,C(x,\eta_t+\delta_x)))\in[0,u]\}\\ \times \mathbf{1}\Big\{\eta_t\Big(\mathbf{B}\Big(x,4\Big(\frac{2u}{\beta s_t k_d}\Big)^{1/d}\Big)\Big) > d+1\Big\}\Big] f(x)dx$$

for $u \geq 0$. Observe that

$$M_t([0,u]) = \widehat{M_t}([0,u]) + \theta_t([0,u]), \quad u \geq 0. \tag{5.43}$$

Lemma 5.3.13. *For any* $u \geq 0$,

$$\lim_{t\to\infty} \widehat{M_t}([0,u]) = \mu(W)u^{d+1}$$

and

$$\theta_t([0,u]) \leq t\int_W \mathbb{P}\Big(\eta_t\Big(\mathbf{B}\Big(x,4\Big(\frac{2u}{\beta s_t k_d}\Big)^{1/d}\Big)\Big) > d+1\Big) f(x)dx \to 0 \quad \textit{as} \quad t\to\infty.$$

Proof. Let $u \geq 0$ be fixed and $u_t := u/s_t$. Without loss of generality we may assume $u_t \in T$. For $x \in W$ we deduce from (5.38), $g(x,u_t) \to 0$ as $t \to \infty$ and the continuity of f that

$$\lim_{t\to\infty} \frac{\mu(\mathbf{B}(x,2g(x,u_t)))}{u_t} = \lim_{t\to\infty} \frac{\mu(\mathbf{B}(x,2g(x,u_t)))}{2^d k_d g(x,u_t)^d}\frac{2^d k_d g(x,u_t)^d}{\mu(\mathbf{B}(x,g(x,u_t)))} = \frac{2^d f(x)}{f(x)} = 2^d.$$

Together with $u_t = u/s_t$ and $s_t = \alpha_2 t^{(d+2)/(d+1)}$, this leads to

$$\lim_{t\to\infty} t^{d+2}\mu\big(\mathbf{B}(x,2g(x,u_t))\big)^{d+1} = (2^d u/\alpha_2)^{d+1}. \tag{5.44}$$

Similarly, we obtain from Lemma 5.3.11 that for t sufficiently large,

$$\begin{aligned}\sup_{x\in W} t^{d+2}\mu\big(\mathbf{B}(x,2g(x,u_t))\big)^{d+1} &\leq t^{d+2}(2^{d+1}u_t/\beta)^{d+1}\sup_{x\in W}\sup_{y\in\mathbf{B}(x,2g(x,u_t))} f(y)^{d+1}\\ &\leq (2^{d+1}u/(\alpha_2\beta))^{d+1}\sup_{y\in W+\mathbf{B}(0,1)} f(y)^{d+1}.\end{aligned} \tag{5.45}$$

Let us now compute the limit of $\widehat{M_t}([0,u])$. By Lemma 5.3.11 we obtain for $\ell_t := 4\big(\frac{2u_t}{\beta k_d}\big)^{1/d}$ that there exists $t_0 > 0$ such that $2g(x,u_t) \leq \ell_t$ for all $t > t_0$ and $x \in W$. From (5.42) we deduce for $x \in W$ that $s_t\mu(\mathbf{B}(x,C(x,\eta_t+\delta_x))) \in [0,u]$ only if there are at least $d+1$ points of η_t in $\mathbf{B}\big(x,2g(x,u_t)\big)$. Then for $t > t_0$, we have

$$\begin{aligned}\widehat{M_t}([0,u]) &= t\int_W \mathbb{P}\big(\eta_t(\mathbf{B}(x,2g(x,u_t))) = d+1\big)p_{d+1}(x,g(x,u_t))\\ &\qquad \times \mathbb{P}\big(\eta_t\big(\mathbf{B}(x,\ell_t)\setminus\mathbf{B}(x,2g(x,u_t))\big) = 0\big) f(x)dx\\ &= \int_W \frac{t^{d+2}\mu\big(\mathbf{B}(x,2g(x,u_t))\big)^{d+1}}{(d+1)!}e^{-t\mu(\mathbf{B}(x,\ell_t))}p_{d+1}(x,g(x,u_t))f(x)dx.\end{aligned}$$

Elementary arguments imply that

$$\lim_{t\to\infty} t\mu(\mathbf{B}(x,\ell_t)) = 0.$$

Therefore combining (5.44) and Lemma 5.3.12 yields

$$\begin{aligned}
&\lim_{t\to\infty} \frac{t^{d+2}\mu\big(\mathbf{B}(x,2g(x,u_t))\big)^{d+1}}{(d+1)!} e^{-t\mu(\mathbf{B}(x,\ell_t))} p_{d+1}(x,g(x,u_t)) f(x)\\
&= \Big(\frac{2^d u}{\alpha_2}\Big)^{d+1} \frac{p_{d+1}}{(d+1)!} f(x) = u^{d+1} f(x),
\end{aligned}$$

where we used $\alpha_2 = \big(\frac{2^{d(d+1)}}{(d+1)!} p_{d+1}\big)^{1/(d+1)}$ in the last step. Thus, by (5.45) and the dominated convergence theorem we obtain

$$\lim_{t\to\infty} \widehat{M}_t([0,u]) = u^{d+1} \int_W f(x)dx = \mu(W)u^{d+1}.$$

Finally, let us compute the limit of $\theta_t([0,u])$. For a Poisson distributed random variable Z with parameter $v > 0$ we have

$$\mathbb{P}(Z \geq d+2) = \sum_{k=d+2}^{\infty} \frac{v^k}{k!} e^{-v} \leq v^{d+2} \sum_{k=0}^{\infty} \frac{v^k}{k!} e^{-v} = v^{d+2}.$$

This implies that

$$\begin{aligned}
\theta_t([0,u]) &\leq t \int_W \mathbb{P}\Big(\eta_t\Big(\mathbf{B}\Big(x, 4\Big(\frac{2u_t}{\beta k_d}\Big)^{1/d}\Big)\Big) > d+1\Big) f(x)dx\\
&\leq t^{d+3} \int_W \mu\Big(\mathbf{B}\Big(x, 4\Big(\frac{2u_t}{\beta k_d}\Big)^{1/d}\Big)\Big)^{d+2} f(x)dx\\
&\leq \sup_{y\in W+\mathbf{B}\left(0,4\left(\frac{2u_t}{\beta k_d}\right)^{1/d}\right)} f(y) \int_W f(x)dx \frac{2^{2d^2+5d+2}}{\beta^{d+2}} t^{d+3} u_t^{d+2}\\
&= \sup_{y\in W+\mathbf{B}\left(0,4\left(\frac{2u_t}{\beta k_d}\right)^{1/d}\right)} f(y)\mu(W) \frac{2^{2d^2+5d+2}}{\beta^{d+2}} \frac{1}{\alpha_2^{d+2}} t^{-\frac{1}{d+1}} u^{d+2}.
\end{aligned}$$

Here, the supremum converges to a constant as $t \to \infty$ so that the second inequality in the assertion is proven. □

In the next lemma, we show a technical result, which will be needed in the proof of Theorem 5.3.8. For $A \subset \mathbb{R}^d$, let $\operatorname{conv}(A)$ denote the convex hull of A.

Lemma 5.3.14. *Let $x_0, \dots, x_{d+1} \in \mathbb{R}^d$ be in general position (i.e. no k-dimensional affine subspace of $\mathbb{R}^d$ with $k \in \{0, \dots, d-1\}$ contains more than $k+1$ of the points) and assume that $N(x_0, \sum_{j=0}^{d+1} \delta_{x_j})$ is bounded. Then,*

a) $x_0 \in \operatorname{int}(\operatorname{conv}(\{x_1, \dots, x_{d+1}\}))$;

b) $N(x_i, \sum_{j=0}^{d+1} \delta_{x_j})$ is unbounded for any $i \in \{1, \dots, d+1\}$.

Proof. Assume that $x_0 \notin \operatorname{int}(\operatorname{conv}(\{x_1, \dots, x_{d+1}\}))$. By the hyperplane separation theorem for convex sets there exists a hyperplane through x_0 with a normal vector $u \in \mathbb{R}^d$ such that $\langle u, x_i\rangle \leq \langle u, x_0\rangle$ for all $i \in \{1, \dots, d+1\}$, where $\langle\cdot,\cdot\rangle$ denotes the standard scalar product on $\mathbb{R}^d$. Define the set $R = \{x_0 + ru : r \in [0,\infty)\}$. For any $y \in R$, x_0 is the closest point to y out of $\{x_0,\dots,x_{d+1}\}$, whence $R \subset N(x_0, \sum_{j=0}^{d+1}\delta_{x_j})$ and $N(x_0, \sum_{j=0}^{d+1}\delta_{x_j})$ is unbounded. This gives us a contradiction and, thus, proves part a).

Let $i \in \{1,\dots,d+1\}$ and assume that $N(x_i, \sum_{j=0}^{d+1}\delta_{x_j})$ is bounded. It follows from part a) that $x_i \in \operatorname{int}(\operatorname{conv}(\{x_0,\dots,x_{i-1},x_{i+1},\dots,x_{d+1}\}))$. On the other hand, again by part a), we have that $x_0 \in \operatorname{int}(\operatorname{conv}(\{x_1,\dots,x_{d+1}\}))$. This implies that

$$\operatorname{conv}(\{x_0,\dots,x_{i-1},x_{i+1},\dots,x_{d+1}\}) = \operatorname{conv}(\{x_0,\dots,x_{d+1}\}) = \operatorname{conv}(\{x_1,\dots,x_{d+1}\}),$$

and, thus, either $x_i, x_0 \in \operatorname{conv}(\{x_1,\dots,x_{i-1},x_{i+1},\dots,x_{d+1}\})$ or $x_i = x_0$. This gives us a contradiction and concludes the proof of part b). □

Finally, we are in position to prove the main result of this subsection.

Proof of Theorem 5.3.8. From Lemma 5.3.13 and (5.43) we deduce that $M_t(B) \to M(B)$ as $t \to \infty$ for all $B \in \mathcal{I}$. Then, by Theorem 5.1.4 it is sufficient to show

$$\lim_{t\to\infty} t \int_W \mathbb{E}\Big[\mathbf{1}\{s_t\mu(\mathbf{B}(x, C(x,\eta_t+\delta_x))) \in B\} \\ \times \mathbf{1}\Big\{\sum_{y\in\eta_t\cap W} \delta_{s_t\mu(\mathbf{B}(y,C(y,\eta_t+\delta_x)))}(B) = m\Big\}\Big] f(x)dx - M(B)\mathbb{P}(\xi_t(B)=m) = 0$$

for $m \in \mathbb{N}_0$ and $B \in \mathcal{I}$. Put $\overline{u} = \sup(B)$ and let $\ell_t = 4\big(\frac{2\overline{u}}{\beta s_t k_d}\big)^{1/d}$. We write

$$t\int_W \mathbb{E}\Big[\mathbf{1}\big\{s_t\mu(\mathbf{B}(x, C(x,\eta_t+\delta_x))) \in B\big\}\mathbf{1}\Big\{\sum_{y\in\eta_t\cap W} \delta_{s_t\mu(\mathbf{B}(y,C(y,\eta_t+\delta_x)))}(B) = m\Big\}\Big] f(x)dx$$
$$= t\int_W \mathbb{E}\Big[\mathbf{1}\big\{s_t\mu(\mathbf{B}(x, C(x,\eta_t+\delta_x))) \in B\big\}\mathbf{1}\big\{\eta_t(\mathbf{B}(x,\ell_t)) = d+1\big\} \\ \times \mathbf{1}\Big\{\sum_{y\in\eta_t\cap W} \delta_{s_t\mu(\mathbf{B}(y,C(y,\eta_t+\delta_x)))}(B) = m\Big\}\Big] f(x)dx$$
$$+ t\int_W \mathbb{E}\Big[\mathbf{1}\big\{s_t\mu(\mathbf{B}(x, C(x,\eta_t+\delta_x))) \in B\big\}\mathbf{1}\big\{\eta_t(\mathbf{B}(x,\ell_t)) > d+1\big\} \\ \times \mathbf{1}\Big\{\sum_{y\in\eta_t\cap W} \delta_{s_t\mu(\mathbf{B}(y,C(y,\eta_t+\delta_x)))}(B) = m\Big\}\Big] f(x)dx$$
$$=: A_t + R_t.$$

By Lemma 5.3.13, we obtain $R_t \to 0$ as $t \to \infty$. Let us study A_t. From Lemma 5.3.11 it follows that there exists $t_0 > 0$ such that $\overline{u}/s_t \in T$ and $\ell_t \geq 4g(y, \overline{u}/s_t)$ for all $y \in W$ and $t > t_0$. Assume $t > t_0$. In case there are only $d+1$ points of η_t in $\mathbf{B}(x,\ell_t)$, we deduce that $s_t\mu(\mathbf{B}(x, C(x,\eta_t+\delta_x))) \in B$ only if the $d+1$ points belong to $\mathbf{B}(x, 2g(x,\overline{u}/s_t))$. Then, by $\ell_t \geq 4g(x,\overline{u}/s_t)$ we obtain

$$A_t = t\int_W \mathbb{E}\Big[\mathbf{1}\big\{s_t\mu(\mathbf{B}(x, C(x,\eta_t+\delta_x))) \in B\big\} \\ \times \mathbf{1}\big\{\eta_t(\mathbf{B}(x,\ell_t)\setminus\mathbf{B}(x,\ell_t/2)) = 0, \eta_t(\mathbf{B}(x,\ell_t/2)) = d+1\big\} \\ \times \mathbf{1}\Big\{\sum_{y\in\eta_t\cap W} \delta_{s_t\mu(\mathbf{B}(y,C(y,\eta_t+\delta_x)))}(B) = m\Big\}\Big] f(x)dx. \tag{5.46}$$

Furthermore, since $\ell_t \geq 4g(y,\overline{u}/s_t)$ for all $y \in W$, we have that

$$\mathbf{B}(y, 2g(y,\overline{u}/s_t)) \cap \mathbf{B}(x,\ell_t/2) = \emptyset, \quad y \in \mathbf{B}(x,\ell_t)^c \cap W.$$

Now the observation that

$$s_t\mu(\mathbf{B}(y, C(y, \eta_t + \delta_x))) \in B \quad \text{if and only if} \quad s_t\mu(\mathbf{B}(y, C(y, (\eta_t + \delta_x)|_{\mathbf{B}(y,2g(y,\overline{u}/s_t))}))) \in B$$

for $y \in \eta_t$ establishes that

$$\begin{aligned}A_t = t\int_W \mathbb{E}\Big[&\mathbf{1}\big\{s_t\mu(\mathbf{B}(x, C(x, \eta_t + \delta_x))) \in B\big\} \\ &\times \mathbf{1}\big\{\eta_t(\mathbf{B}(x,\ell_t) \setminus \mathbf{B}(x,\ell_t/2)) = 0, \eta_t(\mathbf{B}(x,\ell_t/2)) = d+1\big\} \\ &\times \mathbf{1}\Big\{\xi_t(\eta_t|_{\mathbf{B}(x,\ell_t)^c})(B) + \sum_{y\in\eta_t\cap\mathbf{B}(x,\ell_t/2)\cap W} \delta_{s_t\mu(\mathbf{B}(y,C(y,\eta_t+\delta_x)))}(B) = m\Big\}\Big] f(x)dx.\end{aligned}$$

Suppose that $s_t\mu(\mathbf{B}(x, C(x, \eta_t + \delta_x))) \in B$ and that there are exactly $d+1$ points $y_1, \dots, y_{d+1}$ of η_t in $\mathbf{B}(x, \ell_t/2)$ and $\eta_t \cap \mathbf{B}(x,\ell_t) \cap \mathbf{B}(x,\ell_t/2)^c = \emptyset$. From Lemma 5.3.14 it follows that $x \in \mathrm{int}(\mathrm{conv}(\{y_1, \dots, y_{d+1}\}))$ and that the Voronoi cells $N(y_i, \eta_t|_{\mathbf{B}(x,\ell_t)} + \delta_x), i = 1, \dots, d+1$, are unbounded. In particular, we have

$$C(y_i, \eta_t + \delta_x) > \ell_t/4 > g(y_i, \overline{u}/s_t), \quad i = 1, \dots, d+1.$$

Together with the same arguments used to show (5.46), this implies that

$$\begin{aligned}A_t &= t\int_W \mathbb{E}\big[\mathbf{1}\big\{s_t\mu(\mathbf{B}(x, C(x, \eta_t + \delta_x))) \in B\big\} \\ &\qquad\qquad \times \mathbf{1}\big\{\eta_t(\mathbf{B}(x,\ell_t) \setminus \mathbf{B}(x,\ell_t/2)) = 0, \eta_t(\mathbf{B}(x,\ell_t/2)) = d+1\big\} \\ &\qquad\qquad \times \mathbf{1}\big\{\xi_t(\eta_t|_{\mathbf{B}(x,\ell_t)^c})(B) = m\big\}\big] f(x)dx \\ &= t\int_W \mathbb{E}\big[\mathbf{1}\big\{s_t\mu(\mathbf{B}(x, C(x, \eta_t + \delta_x))) \in B\big\}\mathbf{1}\big\{\eta_t(\mathbf{B}(x,\ell_t)) = d+1\big\}\big] \\ &\qquad\qquad \times \mathbb{P}\big(\xi_t(\eta_t|_{\mathbf{B}(x,\ell_t)^c})(B) = m\big) f(x)dx.\end{aligned}$$

Furthermore, we obtain

$$\big|\mathbb{P}\big(\xi_t(\eta_t|_{\mathbf{B}(x,\ell_t)^c})(B) = m\big) - \mathbb{P}\big(\xi_t(B) = m\big)\big| \le \mathbb{P}(\eta_t(\mathbf{B}(x,\ell_t)) > 0) \le t\mu(\mathbf{B}(x,\ell_t))$$

for any $x \in W$, where we used the Markov inequality in the last step. Combining the previous formulas leads to

$$\begin{aligned}&\big|A_t - M(B)\mathbb{P}(\xi_t(B) = m)\big| \\ &\le |M_t(B) - M(B)| \\ &\quad + t\int_W \mathbb{E}\big[\mathbf{1}\big\{s_t\mu(\mathbf{B}(x, C(x, \eta_t + \delta_x))) \in B\big\}\mathbf{1}\big\{\eta_t(\mathbf{B}(x,\ell_t)) > d+1\big\}\big] f(x)dx \\ &\quad + t\int_W \mathbb{E}\big[\mathbf{1}\big\{s_t\mu(\mathbf{B}(x, C(x, \eta_t + \delta_x))) \in B\big\}\mathbf{1}\big\{\eta_t(\mathbf{B}(x,\ell_t)) = d+1\big\}\big] \\ &\qquad\qquad \times \big|\mathbb{P}\big(\xi_t(\eta_t|_{\mathbf{B}(x,\ell_t)^c})(B) = m\big) - \mathbb{P}(\xi_t(B) = m)\big| f(x)dx \\ &\le |M_t(B) - M(B)| + t\int_W \mathbb{P}\big(\eta_t(\mathbf{B}(x,\ell_t)) > d+1\big) f(x)dx + \widehat{M}_t([0,\overline{u}]) \sup_{x\in W} t\mu(\mathbf{B}(x,\ell_t)).\end{aligned}$$

It follows from Lemma 5.3.13 that, as $t \to \infty$, $\widehat{M}_t([0,\overline{u}]) \to M([0,\overline{u}])$, $M_t(B) \to M(B)$ and the integral on the right-hand side vanishes. Without loss of generality we may assume $\ell_t \le 1$, and thus the continuity of f on $W + \overline{\mathbf{B}(0,1)}$ implies that

$$t\mu(\mathbf{B}(x,\ell_t)) \le k_d \max_{z\in W+\overline{\mathbf{B}(0,1)}} f(z) t\ell_t^d$$

for all $x \in W$. Now $\ell_t = 4\big(\frac{2\overline{u}}{\beta s_t k_d}\big)^{1/d}$ and $s_t = \alpha_2 t^{(d+2)/(d+1)}$ yield that the right-hand side vanishes as $t \to \infty$. Thus, we obtain

$$\lim_{t\to\infty} A_t - M(B)\mathbb{P}(\xi_t(B) = m) = 0,$$

which together with $R_t \to 0$ as $t \to \infty$ concludes the proof. □

Proof of Theorem 5.3.9. Let γ be a Poisson process on $\mathbb{R}^d \times [0,\infty)$ with the restriction of the Lebesgue measure as intensity measure. Let μ_1 and μ_2 denote the absolutely continuous measures with densities f_1 and f_2, respectively. Then, [38, Corollary 5.9 and Proposition 6.16] imply that

$$\varrho_t^{(1)} = \sum_{(x,y)\in\gamma} \mathbf{1}\{y \le tf_1(x)\}\delta_x, \quad \varrho_t^{(2)} = \sum_{(x,y)\in\gamma} \mathbf{1}\{y \le tf_2(x)\}\delta_x$$

and

$$\varrho_t = \sum_{(x,y)\in\gamma} \mathbf{1}\{y \le t\phi(x)\}\delta_x$$

are Poisson processes on $\mathbb{R}^d$ with intensity measures $t\mu_1, t\mu_2$ and $t\vartheta$, respectively. They satisfy

$$\varrho_t^{(1)}(A) \le \varrho_t(A) \le \varrho_t^{(2)}(A) \text{ a.s.} \quad \text{and} \quad \varrho_t \overset{d}{=} \zeta_t, \qquad A \subset \mathbb{R}^d,\, t > 0.$$

Therefore for any $v \ge 0$, we obtain

$$\mathbb{P}\big(\min_{x\in\zeta_t\cap W} \mu_1(\mathbf{B}(x, C(x,\zeta_t))) > v\big) \le \mathbb{P}\big(\min_{x\in\varrho_t^{(1)}\cap W} \mu_1(\mathbf{B}(x, C(x,\varrho_t^{(1)}))) > v\big) \tag{5.47}$$

and similarly

$$\mathbb{P}\big(\min_{x\in\zeta_t\cap W} \mu_2(\mathbf{B}(x, C(x,\zeta_t))) > v\big) \ge \mathbb{P}\big(\min_{x\in\varrho_t^{(2)}\cap W} \mu_2(\mathbf{B}(x, C(x,\varrho_t^{(2)}))) > v\big). \tag{5.48}$$

From Theorem 5.3.8, it follows for $j = 1,2$ and $\nu(t) = u(\alpha_2 t^{(d+2)/(d+1)})^{-1}$ with $u \ge 0$, that

$$\lim_{t\to\infty} \mathbb{P}\big(\min_{x\in\varrho_t^{(j)}\cap W} \mu_j(\mathbf{B}(x, C(x,\varrho_t^{(j)}))) > \nu(t)\big) = e^{-\mu_j(W)u^{d+1}}.$$

If $s\phi \le f_1 \le \phi$ for some $s \in (0,1]$, combining (5.47), the previous limit with $j=1$, and the inequality

$$\mathbb{P}\big(\min_{x\in\zeta_t\cap W} s\vartheta(\mathbf{B}(x, C(x,\zeta_t))) > \nu(t)\big) \le \mathbb{P}\big(\min_{x\in\zeta_t\cap W} \mu_1(\mathbf{B}(x, C(x,\zeta_t))) > \nu(t)\big)$$

implies that

$$\limsup_{t\to\infty} \mathbb{P}\big(\min_{x\in\zeta_t\cap W} s\vartheta(\mathbf{B}(x, C(x,\zeta_t))) > \nu(t)\big) \le e^{-\mu_1(W)u^{d+1}}.$$

Then, $s\vartheta(W) \le \mu_1(W)$ concludes the proof of (i). Analogously, if $\phi \le f_2 \le r\phi$ for some $r \ge 1$, combining (5.48), the limit above with $j=2$, the inequality

$$\mathbb{P}\big(\min_{x\in\zeta_t\cap W} r\vartheta(\mathbf{B}(x, C(x,\zeta_t))) > \nu(t)\big) \ge \mathbb{P}\big(\min_{x\in\zeta_t\cap W} \mu_2(\mathbf{B}(x, C(x,\zeta_t))) > \nu(t)\big)$$

and $\mu_2(W) \le r\vartheta(W)$ for $u \ge 0$ shows (ii). □

Bibliography

[1] R. Arratia, L. Goldstein, and L. Gordon. Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.*, 17(1):9–25, 1989.

[2] R. Arratia, L. Goldstein, and L. Gordon. Poisson approximation and the Chen-Stein method. *Statist. Sci.*, 5(4):403–434, 1990.

[3] R. Arratia, L. Goldstein, and F. Kochman. Size bias for one and all. *Probab. Surv.*, 16:1–61, 2019.

[4] N. Balakrishnan and M. V. Koutras. *Runs and scans with applications.* Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], New York, 2002.

[5] A. D. Barbour. Stein's method and Poisson process convergence. *J. Appl. Probab.*, (25A):175–184, 1988.

[6] A. D. Barbour. Multivariate Poisson-binomial approximation using Stein's method. In *Stein's method and applications*, volume 5 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 131–142. Singapore Univ. Press, Singapore, 2005.

[7] A. D. Barbour and T. C. Brown. Stein's method and point process approximation. *Stochastic Process. Appl.*, 43(1):9–31, 1992.

[8] A. D. Barbour and G. K. Eagleson. Poisson convergence for dissociated statistics. *J. Roy. Statist. Soc. Ser. B*, 46(3):397–402, 1984.

[9] A. D. Barbour and L. Holst. Some applications of the Stein-Chen method for proving Poisson convergence. *Adv. in Appl. Probab.*, 21(1):74–90, 1989.

[10] A. D. Barbour, L. Holst, and S. Janson. *Poisson approximation*, volume 2 of *Oxford Studies in Probability.* The Clarendon Press, Oxford University Press, New York, 1992.

[11] A. D. Barbour and A. Xia. On Stein's factors for Poisson approximation in Wasserstein distance. *Bernoulli*, 12(6):943–954, 2006.

[12] B. Błaszczyszyn and R. Schott. Approximations of functionals of some modulated-Poisson Voronoi tessellations with applications to modeling of communication networks. *Japan J. Indust. Appl. Math.*, 22(2):179–204, 2005.

[13] T. C. Brown and A. Xia. Stein's method and birth-death processes. *Ann. Probab.*, 29(3):1373–1403, 2001.

[14] P. Calka. Tessellations. In *New perspectives in stochastic geometry*, pages 145–169. Oxford Univ. Press, Oxford, 2010.

[15] P. Calka and N. Chenavier. Extreme values for characteristic radii of a Poisson-Voronoi tessellation. *Extremes*, 17(3):359–385, 2014.

[16] L. H. Y. Chen. Stein's method: some perspectives with applications. In *Probability towards 2000 (New York, 1995)*, volume 128 of *Lect. Notes Stat.*, pages 97–122. Springer, New York, 1998.

[17] L. H. Y. Chen and A. Xia. Stein's method, Palm theory and Poisson process approximation. *Ann. Probab.*, 32(3B):2545–2569, 2004.

[18] L. H. Y. Chen and A. Xia. Poisson process approximation: from Palm theory to Stein's method. In *Time series and related topics*, volume 52 of *IMS Lecture Notes Monogr. Ser.*, pages 236–244. Inst. Math. Statist., Beachwood, OH, 2006.

[19] N. Chenavier. A general study of extremes of stationary tessellations with examples. *Stochastic Process. Appl.*, 124(9):2917–2953, 2014.

[20] N. Chenavier and C. Y. Robert. Cluster size distributions of extreme values for the Poisson-Voronoi tessellation. *Ann. Appl. Probab.*, 28(6):3291–3323, 2018.

[21] F. Daly and O. Johnson. Relaxation of monotone coupling conditions: Poisson approximation and beyond. *J. Appl. Probab.*, 55(3):742–759, 2018.

[22] L. Decreusefond, M. Schulte, and C. Thäle. Functional Poisson approximation in Kantorovich-Rubinstein distance with applications to U-statistics and stochastic geometry. *Ann. Probab.*, 44(3):2147–2197, 2016.

[23] L. Decreusefond and A. Vasseur. Stein's method and Papangelou intensity for Poisson or Cox process approximation. arXiv:1807.02453, 2018.

[24] P. Deheuvels and D. Pfeifer. Poisson approximations of multinomial distributions and point processes. *J. Multivariate Anal.*, 25(1):65–89, 1988.

[25] V. Čekanavičius and P. Vellaisamy. Compound Poisson approximations in ℓ_p-norm for sums of weakly dependent vectors. *J. Theor. Probab.*, 2020. DOI: 10.1007/s10959-020-01042-9.

[26] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997.

[27] A. P. Godbole. Poisson approximations for runs and patterns of rare events. *Adv. in Appl. Probab.*, 23(4):851–865, 1991.

[28] L. Goldstein and Y. Rinott. Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.*, 33(1):1–17, 1996.

[29] L. Györfi, N. Henze, and H. Walk. The limit distribution of the maximum probability nearest-neighbour ball. *J. Appl. Probab.*, 56(2):574–589, 2019.

[30] N. Henze. The limit distribution for maxima of "weighted" rth-nearest-neighbour distances. *J. Appl. Probab.*, 19(2):344–354, 1982.

[31] N. Henze. Ein asymptotischer Satz über den maximalen Minimalabstand von unabhängigen Zufallsvektoren mit Anwendung auf einen Anpassungstest im $\mathbf{R}^p$ und auf der Kugel. *Metrika*, 30(4):245–259, 1983.

[32] O. Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[33] O. Kallenberg. *Random measures, theory and applications*, volume 77 of *Probability Theory and Stochastic Modelling.* Springer, Cham, 2017.

[34] V. S. Koroljuk and Y. V. Borovskich. *Theory of U-statistics*, volume 273 of *Mathematics and its Applications.* Kluwer Academic Publishers Group, Dordrecht, 1994. Translated from the 1989 Russian original by P. V. Malyshev and D. V. Malyshev and revised by the authors.

[35] M. V. Koutras, G. K. Papadopoulos, and S. G. Papastavridis. Runs on a circle. *J. Appl. Probab.*, 32(2):396–404, 1995.

[36] W. Lao and M. Mayer. U-max-statistics. *J. Multivariate Anal.*, 99(9):2039–2052, 2008.

[37] G. Last and M. Otto. Disagreement coupling of Gibbs processes with an application to Poisson approximation. arXiv:2104.00737, 2021.

[38] G. Last and M. Penrose. *Lectures on the Poisson process*, volume 7 of *Institute of Mathematical Statistics Textbooks.* Cambridge University Press, Cambridge, 2018.

[39] A. J. Lee. *U-statistics*, volume 110 of *Statistics: Textbooks and Monographs.* Marcel Dekker, Inc., New York, 1990.

[40] C. Ley, G. Reinert, and Y. Swan. Stein's method for comparison of univariate distributions. *Probab. Surv.*, 14:1–52, 2017.

[41] E. H. Lieb and M. Loss. *Analysis*, volume 14 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, second edition, 2001.

[42] M. Mayer and I. Molchanov. Limit theorems for the diameter of a random sample in the unit ball. *Extremes*, 10(3):129–150, 2007.

[43] E. J. McShane. Extension of range of functions. *Bull. Amer. Math. Soc.*, 40(12):837–842, 1934.

[44] J. Møller. *Lectures on random Voronoĭ tessellations*, volume 87 of *Lecture Notes in Statistics.* Springer-Verlag, New York, 1994.

[45] S. Y. Novak. *Extreme value methods with applications to finance*, volume 122 of *Monographs on Statistics and Applied Probability.* CRC Press, Boca Raton, FL, 2012.

[46] S. Y. Novak. Poisson approximation. *Probab. Surv.*, 16:228–276, 2019.

[47] M. Otto. Poisson approximation of Poisson-driven point processes and extreme values in stochastic geometry. arXiv:2005.10116, 2020.

[48] F. Papangelou. The conditional intensity of general point processes and an application to line processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 28:207–226, 1973/74.

[49] M. Penrose. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003.

[50] M. D. Penrose. The longest edge of the random minimal spanning tree. *Ann. Appl. Probab.*, 7(2):340–361, 1997.

[51] M. D. Penrose. Inhomogeneous random graphs, isolated vertices, and Poisson approximation. *J. Appl. Probab.*, 55(1):112–136, 2018.

[52] F. Pianoforte and M. Schulte. Poisson approximation with applications to stochastic geometry. arXiv:2104.02528, 2021.

[53] A. Poupon. Voronoi and Voronoi-related tessellations in studies of protein structure and interaction. *Current Opinion in Structural Biology*, 14(2):233–241, 2004.

[54] M. Ramella, W. Boschin, D. Fadda, and M. Nonino. Finding galaxy clusters using Voronoi tessellations. *A&A*, 368(3):776–786, 2001.

[55] M. Reitzner and M. Schulte. Central limit theorems for U-statistics of Poisson point processes. *Ann. Probab.*, 41(6):3879–3909, 2013.

[56] A. Röllin. Translated Poisson approximation using exchangeable pair couplings. *Ann. Appl. Probab.*, 17(5-6):1596–1614, 2007.

[57] B. Roos. On the rate of multivariate Poisson convergence. *J. Multivariate Anal.*, 69(1):120–134, 1999.

[58] B. Roos. Poisson approximation of multivariate Poisson mixtures. *J. Appl. Probab.*, 40(2):376–390, 2003.

[59] B. Roos. Refined total variation bounds in the multivariate and compound Poisson approximation. *ALEA Lat. Am. J. Probab. Math. Stat.*, 14(1):337–360, 2017.

[60] N. Ross. Fundamentals of Stein's method. *Probab. Surv.*, 8:210–293, 2011.

[61] R. Schneider and W. Weil. *Stochastic and integral geometry*. Probability and its Applications (New York). Springer-Verlag, Berlin, 2008.

[62] D. Schuhmacher. Distance estimates for dependent superpositions of point processes. *Stochastic Process. Appl.*, 115(11):1819–1837, 2005.

[63] D. Schuhmacher. Stein's method and Poisson process approximation for a class of Wasserstein metrics. *Bernoulli*, 15(2):550–568, 2009.

[64] D. Schuhmacher and K. Stucki. Gibbs point process approximation: total variation bounds using Stein's method. *Ann. Probab.*, 42(5):1911–1951, 2014.

[65] M. Schulte and C. Thäle. The scaling limit of Poisson-driven order statistics with applications in geometric probability. *Stochastic Process. Appl.*, 122(12):4096–4120, 2012.

[66] M. Schulte and C. Thäle. Poisson point process convergence and extreme values in stochastic geometry. In *Stochastic analysis for Poisson point processes*, volume 7 of *Bocconi Springer Ser.*, pages 255–294. Bocconi Univ. Press, 2016.

[67] B. Silverman and T. Brown. Short distances, flat triangles and Poisson limits. *J. Appl. Probab.*, 15(4):815–825, 1978.

[68] R. L. Smith. Extreme value theory for dependent sequences via the Stein-Chen method of Poisson approximation. *Stochastic Process. Appl.*, 30(2):317–327, 1988.

[69] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602, 1972.

[70] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften.* Springer-Verlag, Berlin, 2009.

[71] A. Xia. Stein's method and Poisson process approximation. In *An introduction to Stein's method*, volume 4 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 115–181. Singapore Univ. Press, Singapore, 2005.

Declaration of consent

on the basis of Article 18 of the PromR Phil.-nat. 19

Name/First Name: Pianoforte Federico

Registration Number: 17-143-983

Study program: PhD in Statistics

Bachelor ☐ Master ☐ Dissertation ☑

Title of the thesis: Poisson convergence in stochastic geometry via generalized size-bias coupling

Supervisor: Prof. Dr. Matthias Schulte

I declare herewith that this thesis is my own work and that I have not used any sources other than those stated. I have indicated the adoption of quotations as well as thoughts taken from other authors as such in the thesis. I am aware that the Senate pursuant to Article 36 paragraph 1 litera r of the University Act of September 5th, 1996 and Article 69 of the University Statute of June 7th, 2011 is authorized to revoke the doctoral degree awarded on the basis of this thesis.
For the purposes of evaluation and verification of compliance with the declaration of originality and the regulations governing plagiarism, I hereby grant the University of Bern the right to process my personal data and to perform the acts of use this requires, in particular, to reproduce the written thesis and to store it permanently in a database, and to use said database, or to make said database available, to enable comparison with theses submitted by others.

Bern, 23/06/2021

Place/Date

Signature

Federico Pianoforte