J. Appl. Probab. **60,** 223–240 (2023)
*doi:*10.1017/jpr.2022.33

MULTIVARIATE POISSON AND POISSON PROCESS APPROXIMATIONS WITH APPLICATIONS TO BERNOULLI SUMS AND *U*-STATISTICS

FEDERICO PIANOFORTE* ** AND
RICCARDO TURIN ,* *** *University of Bern*

Abstract

This article derives quantitative limit theorems for multivariate Poisson and Poisson process approximations. Employing the solution of the Stein 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, stronger than the total variation distance, defined as 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.

Keywords: Chen–Stein method; Wasserstein distance; multinomial distribution; m-dependent random vectors; Papangelou intensity

2020 Mathematics Subject Classification: Primary 60F05
Secondary 60G55; 62E17

1. Introduction and main results

In this paper we use the Stein method to estimate the Wasserstein distance between a non-negative integer-valued random vector and a Poisson random vector. This problem has been studied by several authors, mostly in terms of the total variation distance; among others we mention [1, 3, 4, 6, 13, 28, 29]. Furthermore, we use our abstract result on multivariate Poisson approximation to derive a limit theorem for the Poisson process approximation.

More precisely, let $\mathbf{X} = (X_1, \dots, X_d)$ be an integrable random vector taking values in $\mathbb{N}_0^d$, $d \in \mathbb{N}$, where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, 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 contribution of this paper is an upper bound on the Wasserstein distance

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

Received 31 May 2021; revision received 1 March 2022.
* Postal address: Institute of Mathematical Statistics and Actuarial Science, University of Bern, Alpeneggstrasse 22, 3012 Bern, Switzerland.
** Email address: federico.pianoforte@stat.unibe.ch
*** Email address: riccardo.turin@stat.unibe.ch

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 $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. 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 [25] for the Poisson approximation of random variables to the multivariate case.

Throughout the paper, for any $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and index $1 \le j \le d$, we denote by $x_{1:j}$ and $x_{j:d}$ the subvectors $(x_1, \ldots, x_j)$ and $(x_j, \ldots, x_d)$, respectively.

Theorem 1.1. *Let $\mathbf{X} = (X_1, \ldots, X_d)$ be an integrable random vector with values in $\mathbb{N}_0^d$, $d \in \mathbb{N}$, and let $\mathbf{P} = (P_1, \ldots, P_d)$ be a Poisson random vector with $\mathbb{E}[\mathbf{P}] = (\lambda_1, \ldots, \lambda_d) \in [0, \infty)^d$. For $1 \le i \le d$, consider any random vector $\mathbf{Z}^{(i)} = \left(Z_1^{(i)}, \ldots, Z_i^{(i)}\right)$ in $\mathbb{Z}^i$ defined on the same probability space as $\mathbf{X}$, and define*

$$q_{m_{1:i}} = m_i \mathbb{P}\big(X_{1:i} = m_{1:i}\big) - \lambda_i \mathbb{P}\big(X_{1:i} + \mathbf{Z}^{(i)} = (m_{1:i-1}, m_i - 1)\big) \tag{1.1}$$

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

$$d_W(\mathbf{X}, \mathbf{P}) \le \sum_{i=1}^d \left(\lambda_i \mathbb{E}\big|Z_i^{(i)}\big| + 2\lambda_i \sum_{j=1}^{i-1} \mathbb{E}\big|Z_j^{(i)}\big| + \sum_{\substack{m_{1:i} \in \mathbb{N}_0^i \\ m_i \neq 0}} \big|q_{m_{1:i}}\big| \right). \tag{1.2}$$

It should be noted that a bound that slightly improves (1.2) can easily be obtained as shown in the following section in Remark 2.1, which corresponds to (1.8) in [25, Theorem 1.3] when $d = 1$.

In order to give an interpretation of Equation (1.1), let us consider the random vectors

$$\mathbf{Y}^{(i)} = (X_{1:i-1}, X_i + 1) + \mathbf{Z}^{(i)}, \quad i = 1, \ldots, d, \tag{1.3}$$

with $\mathbf{X}$ and $\mathbf{Z}^{(i)}$ defined as in Theorem 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 $q_{m_{1:i}}$, $m_{1:i} \in \mathbb{N}_0^i$ with $m_i \neq 0$ satisfies Equation (1.1) if and only if

$$\mathbb{E}[X_i f(X_{1:i})] = \lambda_i \mathbb{E}[f(\mathbf{Y}^{(i)})] + \sum_{m_{1:i} \in \mathbb{N}_0^i,\, m_i \neq 0} q_{m_{1:i}} f(m_{1:i}) \tag{1.4}$$

for all functions $f : \mathbb{N}_0^i \to \mathbb{R}$ such that $\mathbb{E}\,|X_i f(X_{1:i})| < \infty$, where to prove that (1.4) implies (1.1) it is enough to consider f to be the function with value 1 at $m_{1:i}$ and 0 elsewhere. When the $q_{m_{1:i}}$ are all zeros and $\mathbb{E}[X_i] = \lambda_i$, the condition $\mathbb{P}(X_{1:i} + \mathbf{Z}^{(i)} \in \mathbb{N}_0^i) = 1$ is satisfied, as can be seen by taking the sum over $m_{1:i} \in \mathbb{N}_0^i$ with $m_i \neq 0$ in (1.1). In this case, (1.4) becomes

$$\mathbb{E}[X_i f(X_{1:i})] = \mathbb{E}[X_i]\mathbb{E}[f(\mathbf{Y}^{(i)})]. \tag{1.5}$$

Recall that, for a random variable $X \ge 0$ with mean $\mathbb{E}[X] > 0$, a random variable X^s has the size bias distribution of X if it satisfies

$$\mathbb{E}[Xf(X)] = \mathbb{E}[X]\mathbb{E}[f(X^s)] \tag{1.6}$$

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press

for all measurable $f : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}\,|Xf(X)| < \infty$. Therefore, if for some $1 \le i \le d$ the $q_{m_{1:i}}$ are all zeros and $\mathbb{E}[X_i] = \lambda_i$, the distribution of the random vector $\mathbf{Y}^{(i)}$ can be seen as the size bias distribution of $X_{1:i}$, as it satisfies (1.5), which corresponds to (1.6) in the one-dimensional case. Note that this definition is similar to that of $\mathbf{X}$-size bias distribution in the ith coordinate introduced in [15].

Following this interpretation, when $\mathbb{E}[\mathbf{X}] = (\lambda_1, \ldots, \lambda_d)$ and the random vectors $\mathbf{Z}^{(i)}$ are chosen to be such that the $q^{(i)}_{m_{1:i}}$ are not zero, we can think of the distribution of $\mathbf{Y}^{(i)}$ defined by (1.3) as an approximate size bias distribution of $X_{1:i}$, where instead of assuming that $\mathbf{Y}^{(i)}$ satisfies (1.5) exactly, we allow error terms $q_{m_{1:i}}$. This is an important advantage of Theorem 1.1, since one does not need to find random vectors with an exact size bias distribution (in the sense of (1.5)); 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 vectors with high probability.

The second main contribution of our work concerns Poisson process approximation of point processes with finite intensity measure. For a point process ξ and a Poisson process η on a measurable space $\mathbb{X}$ with finite intensity measure, Theorem 1.1 provides bounds on the Wasserstein distance

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

where $A_1, \ldots, A_d$ are measurable subsets of $\mathbb{X}$. This allows for a way to compare the distributions of ξ and η, by taking the supremum of the Wasserstein distances between the point processes evaluated on arbitrary collections $(A_1, \ldots, A_d)$ of disjoint sets. More precisely, let $(\mathbb{X}, \mathcal{X})$ be a measurable space and define $\mathsf{N}_{\mathbb{X}}$ as the collection of all σ-finite counting measures. The set $\mathsf{N}_{\mathbb{X}}$ is equipped with the σ-field $\mathcal{N}_{\mathbb{X}}$ generated by the collection of all subsets of $\mathsf{N}_{\mathbb{X}}$ of the form

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

This means that $\mathcal{N}_{\mathbb{X}}$ is the smallest σ-field on $\mathsf{N}_{\mathbb{X}}$ that makes the map $\nu \mapsto \nu(B)$ measurable for all $B \in \mathcal{X}$. A point process ξ on $\mathbb{X}$ is a random element in $(\mathsf{N}_{\mathbb{X}}, \mathcal{N}_{\mathbb{X}})$. The intensity of ξ is the measure λ on $(\mathbb{X}, \mathcal{X})$ defined by $\lambda(B) = \mathbb{E}[\xi(B)]$, $B \in \mathcal{X}$. When a point process ξ has finite intensity measure λ, for any choice of subsets $A_1, \ldots, A_d \in \mathcal{X}$, the random vector $(\xi(A_1), \ldots, \xi(A_d))$ takes values in $\mathbb{N}_0^d$ (almost surely). Thus, we define a metric in the space of point processes with finite intensity measure in the following way.

Definition 1.1. Let ξ and ζ be point processes on $\mathbb{X}$ with finite intensity measure. The distance d_π between the distributions of ξ and ζ is defined as

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

where

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

The function d_π is a probability distance between the distributions of point processes, which follows immediately from its definition and, e.g., [19, Proposition 2.10]. To the best of our knowledge, this is the first time the distance d_π has been 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 $\mathcal{X}^d_{\mathrm{disj}}$ to suitable families of sets. For example, for locally finite point processes on a locally compact second-countable Hausdorff space (lcscH), we may define the distance d_π by replacing $\mathcal{X}^d_{\mathrm{disj}}$ with the family of d-tuples of disjoint and relatively compact Borel sets. However, this falls outside the scope of this paper, and it will be treated elsewhere. Let us now state our main theoretical result on Poisson process approximation.

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press

Theorem 1.2. *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 $(A_1, \dots, A_i) \in \mathcal{X}^i_{\mathrm{disj}}$ with $i \in \mathbb{N}$, consider a random vector $\mathbf{Z}^{A_{1:i}} = \left(Z_1^{A_{1:i}}, \dots, Z_i^{A_{1:i}}\right)$ defined on the same probability space as ξ with values in $\mathbb{Z}^i$, and define*

$$\begin{aligned} q_{m_{1:i}}^{A_{1:i}} = m_i \mathbb{P}\big((\xi(A_1), \dots, \xi(A_i)) = m_{1:i}\big) \\ - \lambda(A_i)\mathbb{P}\big((\xi(A_1), \dots, \xi(A_i)) + \mathbf{Z}^{A_{1:i}} = (m_{1:i-1}, m_i - 1)\big) \end{aligned} \tag{1.7}$$

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

$$d_\pi(\xi, \eta) \leq \sup_{(A_1,\dots,A_d)\in\mathcal{X}^d_{\mathrm{disj}}, d\in\mathbb{N}} \sum_{i=1}^{d} \left(\sum_{\substack{m_{1:i}\in\mathbb{N}_0^i \\ m_i\neq 0}} \left| q_{m_{1:i}}^{A_{1:i}} \right| + 2\lambda(A_i) \sum_{j=1}^{i} \mathbb{E}\left| Z_j^{A_{1:i}} \right| \right). \tag{1.8}$$

Note that a bound slightly sharper than (1.8) can be derived, as expressed in Remark 2.2.

The Poisson process approximation has mostly been studied in terms of the total variation distance in the literature; see e.g. [2, 3, 5, 8, 9, 30, 31] and references therein. In contrast, [10, 11] deal with Poisson process approximation using the Kantorovich–Rubinstein distance. Recall that the total variation distance between two point processes ξ and ζ on $\mathbb{X}$ is

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

We prove 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.

Proposition 1.1. *Let ξ and ζ be two point processes on $\mathbb{X}$ with finite intensity measure. Then*

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

Note that, since $d_\pi(\xi, \zeta) \geq |\mathbb{E}[\xi(\mathbb{X})] - \mathbb{E}[\zeta(\mathbb{X})]|$, Example 2.2 in [10] provides a sequence of point processes $(\zeta_n)_{n\geq 1}$ 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.

The Kantorovich–Rubinstein distance between two point processes ξ and ζ with finite intensity measure is defined as the optimal transportation cost between their distributions, when the cost function is the total variation distance between measures; that is,

$$d_{KR}(\xi, \zeta) = \inf_{(\varrho_1,\varrho_2)\in\Sigma(\xi,\zeta)} \mathbb{E} \sup_{A\in\mathcal{X}} |\varrho_1(A) - \varrho_2(A)|,$$

where $\Sigma(\xi, \zeta)$ denotes the set of all pairs of point processes ϱ_1, ϱ_2 on $\mathbb{X}$ defined on the same probability space such that ϱ_1 and ϱ_2 follow the distributions of ξ and ζ, respectively. We prove that, under suitable assumptions on the space, d_π is dominated by $2d_{KR}$, while it remains an open problem whether the two distances are equivalent or not.

Proposition 1.2. *Let ξ and ζ be two point processes with finite intensity measure on an lcscH space $\mathbb{X}$ with Borel σ-field $\mathcal{X}$. Then*

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

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press

The factor 2 in Proposition 1.2 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 that

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

On the other hand, d_{KR} is bounded by the expected total variation distance between the two counting measures; thus $d_{KR}(\delta_a, \delta_b) \leq 1$. Hence, in this case $d_\pi(\delta_a, \delta_b) = 2d_{KR}(\delta_a, \delta_b)$. It is worth mentioning that our general result, Theorem 1.2, permits the approximation of point processes by Poisson processes on any measurable space. Hence, Theorem 1.2 can be used to obtain approximation results for point processes also when the notion of weak convergence is not defined. Moreover, when $\mathbb{X}$ is lcscH, convergence with respect to d_π implies convergence in distribution, as easily follows from [16, Theorem 16.16(iii)].

To demonstrate the versatility of our general main results, we apply them to several examples. In Subsection 3.1, we approximate the sum of Bernoulli random vectors by a Poisson random vector. By a Bernoulli random vector, we mean a random vector with values in the set composed of the canonical vectors of $\mathbb{R}^d$ and the null 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. [27]). We derive an explicit approximation result in the Wasserstein distance for the more general case of m-dependent Bernoulli random vectors.

In Subsections 3.2 and 3.3, we apply Theorem 1.2 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 [11, Theorem 3.7] and [10, 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 1.2 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 (1.7) in Theorem 1.2 is satisfied with $q^{A_{1:i}}_{m_{1:i}} \equiv 0$ for all collections of disjoint sets.

The proof of Theorem 1.1 is based on the Chen–Stein method applied to each component of the random vectors and the coupling in (1.1). In the proof of Theorem 1.2 we mimic the approach used in [1] to prove Theorem 2, as we derive the process bound as a consequence of the d-dimensional bound.

Before we discuss the applications in Section 3, we prove our main results in the next section.

2. Proofs of the main results

Throughout this section, $\mathbf{X} = (X_1, \ldots, X_d)$ is an integrable random vector with values in $\mathbb{N}_0^d$ and $\mathbf{P} = (P_1, \ldots, P_d)$ is a Poisson random vector with mean $\mathbb{E}[\mathbf{P}] = (\lambda_1, \ldots, \lambda_d) \in [0, \infty)^d$. Without loss of generality we assume that $\mathbf{X}$ and $\mathbf{P}$ are independent and defined on the same probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. 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,

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

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press

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

For any fixed $g \in \mathrm{Lip}(1)$, a solution of Stein's equation for the Poisson distribution is a real-valued function $\widehat{g}^{(\lambda)} : \mathbb{N}_0 \to \mathbb{R}$ that satisfies

$$\lambda \widehat{g}^{(\lambda)}(i+1) - i\widehat{g}^{(\lambda)}(i) = g(i) - \mathbb{E}[g(P_\lambda)], \quad i \in \mathbb{N}_0, \tag{2.1}$$

where P_λ is a Poisson random variable with mean $\lambda \geq 0$. For convenience, we fix the initial condition $\widehat{g}^{(\lambda)}(0) = 0$. With this assumption, the function $\widehat{g}^{(\lambda)}$ is unique and may be obtained by solving (2.1) recursively on i. An explicit expression for this solution is given in [14, Theorem 1.2]. The following lemma is a direct consequence of [7, Theorem 1.1] (note that the case $\lambda = 0$ is trivial).

Lemma 2.1. *For any $\lambda \geq 0$ and $g \in \mathrm{Lip}(1)$, let $\widehat{g}^{(\lambda)}$ be the solution of the Stein equation (2.1) with initial condition $\widehat{g}^{(\lambda)}(0) = 0$. Then*

$$\sup_{i \in \mathbb{N}_0} \left| \widehat{g}^{(\lambda)}(i) \right| \leq 1 \quad \text{and} \quad \sup_{i \in \mathbb{N}_0} \left| \widehat{g}^{(\lambda)}(i+1) - \widehat{g}^{(\lambda)}(i) \right| \leq 1. \tag{2.2}$$

Recall that, for any $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and some index $1 \leq j \leq d$, we write $x_{1:j}$ and $x_{j:d}$ for the subvectors $(x_1, \ldots, x_j)$ and $(x_j, \ldots, x_d)$, respectively. For $g \in \mathrm{Lip}^d(1)$, let $\widehat{g}^{(\lambda)}_{x_{1:i-1}|x_{i+1:d}}$ denote the solution to (2.1) 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 1.1.

Proposition 2.1. *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 2.1. First, observe that

$$\mathbb{E}\big[g(\mathbf{P}) - g(\mathbf{X})\big] = \sum_{i=1}^{d} \mathbb{E}\left[g(X_{1:i-1}, P_{i:d}) - g(X_{1:i}, P_{i+1:d}) \right], \tag{2.3}$$

with the conventions $(X_{1:0}, P_{1:d}) = \mathbf{P}$ and $(X_{1:d}, P_{d+1:d}) = \mathbf{X}$. 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 that

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

for all $i = 1, \ldots, d$. Together with (2.3), this leads to the desired conclusion. □

Proof of Theorem 1.1. In view of Proposition 2.1, it suffices to bound

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

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press

For the remainder of the proof, the index i is fixed and we omit the superscript (i) in $Z_{1:i}^{(i)}$. Define the function $h\colon \mathbb{N}_0^i \to \mathbb{R}$ so that

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

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 $m_{1:d} \notin \mathbb{N}_0^d$, it follows from (1.1) 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 $|h(X_{1:i})| \leq 1$ by (2.2), the triangle inequality establishes

$$\left|\mathbb{E}\left[X_i \widehat{g}^{(\lambda_i)}_{X_{1:i-1}|P_{i+1:d}}(X_i) - \lambda_i \widehat{g}^{(\lambda_i)}_{X_{1:i-1}|P_{i+1:d}}(X_i + 1)\right]\right| \leq \sum_{\substack{m_{1:i}\in\mathbb{N}_0^i \\ m_i \neq 0}} \left|q_{m_{1:i}}\right| + \lambda_i(H_1 + H_2), \tag{2.4}$$

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 (2.2) guarantee

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

Combining (2.4) with the bounds for H_1 and H_2 and summing over $i = 1, \ldots, d$ concludes the proof. □

Remark 2.1. It follows directly from the previous proof that the term $\sum_{j=1}^{i-1} \mathbb{E}|Z_j|$ in (1.2) could be replaced by $\mathbb{P}(Z_{1:i-1} \neq 0)$. Moreover, applying (1.4) from [7, Theorem 1.1] yields

$$H_1 \leq \min\left\{1, \frac{8}{3\sqrt{2e\lambda_i}}\right\} \mathbb{E}|Z_i|\,.$$

These two observations together lead to the improved bound for Theorem 1.1:

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press

$$d_W(\mathbf{X},\mathbf{P}) \le \sum_{i=1}^{d} \left(\min\left\{ \lambda_i, \frac{8\sqrt{\lambda_i}}{3\sqrt{2e}} \right\} \mathbb{E}\left|Z_i^{(i)}\right| + 2\lambda_i \mathbb{P}\left(Z_{1:i-1}^{(i)} \neq 0\right) + \sum_{\substack{m_{1:i}\in\mathbb{N}_0^i \\ m_i\neq 0}} \left|q_{m_{1:i}}\right| \right).$$

Next, we derive Theorem 1.2 from Theorem 1.1.

Proof of Theorem 1.2. Let $d \in \mathbb{N}$ and $\mathbf{A} = (A_1, \dots, A_d) \in \mathcal{X}_{\mathrm{disj}}^d$. 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 1.1 with $\mathbf{Z}^{(i)} = \mathbf{Z}^{A_{1:i}}$, we obtain

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

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

Remark 2.2. By taking into account Remark 2.1, one immediately obtains

$$d_\pi(\xi,\eta) \le \sup_{(A_1,\dots,A_d)\in\mathcal{X}_{\mathrm{disj}}^d, d\in\mathbb{N}} \sum_{i=1}^{d} \left(\min\left\{ \lambda(A_i), \frac{8\sqrt{\lambda(A_i)}}{3\sqrt{2e}} \right\} \mathbb{E}\left|Z_i^{A_{1:i}}\right| + 2\lambda(A_i)\mathbb{P}\left(Z_{1:i-1}^{A_{1:i}} \neq 0\right) + \sum_{\substack{m_{1:i}\in\mathbb{N}_0^i \\ m_i\neq 0}} \left|q_{m_{1:i}}^{A_{1:i}}\right| \right).$$

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

$$d_{TV}(\xi,\zeta) = \sup_{B\in\mathcal{N}_{\mathbb{X}}} \left|\mathbb{P}(\xi\in B) - \mathbb{P}(\zeta\in B)\right|. \tag{2.5}$$

The result is obtained by a monotone class theorem, [21, Theorem 1.3], which is stated below 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 $A_n \downarrow A$, we have $A \in \mathcal{A}$.

Lemma 2.2. *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 $\mathcal{U}$.*

Proof of Proposition 1.1. Let us first introduce the set of finite counting measures

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

with the trace σ-field

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

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press

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 (2.5):

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

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

$$d_\pi(\xi, \zeta) \geq \sup_{U \in \mathcal{U}} |\mathbb{P}(\xi \in U) - \mathbb{P}(\zeta \in U)|\,, \tag{2.6}$$

with

$$\mathcal{U} = \left\{ \left\{ \nu \in \mathsf{N}_{\mathbb{X}}^{<\infty} : (\nu(A_1), \ldots, \nu(A_d)) \in M \right\} : d \in \mathbb{N},\ (A_1, \ldots, A_d) \in \mathcal{X}_{\mathrm{disj}}^d,\ M \in \mathcal{P}(\mathbb{N}_0^d) \right\}.$$

It can easily be verified that $\mathcal{U}$ is an algebra, $\mathcal{U} \subset \mathcal{N}_{\mathbb{X}}^{<\infty}$, and $\sigma(\mathcal{U}) = \mathcal{N}_{\mathbb{X}}^{<\infty}$. Moreover, by (2.6), $\mathcal{U}$ is a subset of the monotone class

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

Lemma 2.2 concludes the proof. □

In the last part of this section, we show that d_π is dominated by $2d_{KR}$ when the underlying space is lcscH and $\mathcal{X}$ is the Borel σ-field. 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. Recall that the Kantorovich–Rubinstein distance between two point processes ξ and ζ, with finite intensity measure on a measurable space $\mathbb{X}$, is given by

$$d_{KR}(\xi, \zeta) = \inf_{(\varrho_1, \varrho_2) \in \Sigma(\xi, \zeta)} \mathbb{E} \sup_{A \in \mathcal{X}} |\varrho_1(A) - \varrho_2(A)|,$$

where $\Sigma(\xi, \zeta)$ denotes the set of all pairs of point processes ϱ_1, ϱ_2 on $\mathbb{X}$ defined on the same probability space such that ϱ_1 and ϱ_2 follow the distributions of ξ and ζ, respectively. When the configuration space $\mathbb{X}$ is lcscH, the Kantorovich duality theorem [33, Theorem 5.10] yields an equivalent definition for this metric:

$$d_{KR}(\xi, \zeta) = \sup_{h \in \mathcal{L}(1)} |\mathbb{E}[h(\xi)] - \mathbb{E}[h(\zeta)]|\,, \tag{2.7}$$

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

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

with Lipschitz constant bounded by 1. Since ξ and ζ take values in $\mathsf{N}_{\mathbb{X}}^{<\infty}$, by [22, Theorem 1] we may assume that h is defined on $\mathsf{N}_{\mathbb{X}}^{<\infty}$.

Proof of Proposition 1.2. For $g \in \mathrm{Lip}^d(1)$ and disjoint sets $A_1, \ldots, A_d \in \mathcal{X}$, $d \in \mathbb{N}$, define $h : \mathsf{N}_{\mathbb{X}}^{<\infty} \to \mathbb{R}$ by $h(\nu) = g(\nu(A_1), \ldots, \nu(A_d))$. For finite point configurations ν_1 and ν_2, we obtain

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press

$$|h(\nu_1) - h(\nu_2)| \le |g(\nu_1(A_1), \dots, \nu_1(A_d)) - g(\nu_2(A_1), \dots, \nu_2(A_d))|$$

$$\le \sum_{i=1}^{d} |\nu_1(A_i) - \nu_2(A_i)| \le 2d_{TV,\mathsf{N}_{\mathbb{X}}}(\nu_1, \nu_2).$$

Therefore, we have $h/2 \in \mathcal{L}(1)$. Together with (2.7), this implies $|\mathbb{E}[h(\xi)] - \mathbb{E}[h(\zeta)]| \le 2d_{KR}(\xi, \zeta)$ and concludes the proof. □

3. Applications

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. The distributions of $\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(n)}$ are given by

$$\begin{aligned} &\mathbb{P}(\mathbf{Y}^{(r)} = \mathbf{e}_j) = p_{r,j} \in [0, 1], \quad r = 1, \dots, n, \quad j = 1, \dots, d, \\ &\mathbb{P}(\mathbf{Y}^{(r)} = \mathbf{0}) = 1 - \sum_{j=1}^{d} p_{r,j} \in [0, 1], \quad r = 1, \dots, n, \end{aligned} \tag{3.1}$$

where $\mathbf{e}_j$ denotes the vector with entry 1 at position j and entry 0 otherwise. If the Bernoulli random vectors are independent and identically distributed (i.i.d.), $\mathbf{X}$ 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 [4, 12, 27, 29] and the survey [23]. Unlike the abovementioned 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 [13] using metrics that are weaker than the total variation distance. To the best of our knowledge, this is the first paper where the Poisson approximation of the sum of m-dependent Bernoulli random vectors is investigated in terms of 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 (3.1), and assume that for a given fixed $m \in \mathbb{N}_0$ and any two subsets S and T of $\{1, \dots, n\}$ such that $\min(S) - \max(T) > m$, the collections $\left(\mathbf{Y}^{(s)}\right)_{s \in S}$ and $\left(\mathbf{Y}^{(t)}\right)_{t \in T}$ are independent. Define the random vector $\mathbf{X} = (X_1, \dots, X_d)$ as

$$\mathbf{X} = \sum_{r=1}^{n} \mathbf{Y}^{(r)}. \tag{3.2}$$

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

$$\lambda_j = \sum_{r=1}^{n} p_{r,j}, \quad j = 1, \dots, d. \tag{3.3}$$

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

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

We now state the main result of this subsection.

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press

Theorem 3.1. *Let* $\mathbf{X}$ *be as in* (3.2)*, and let* $\mathbf{P} = (P_1, \ldots, P_d)$ *be a Poisson random vector with mean* $\mathbb{E}[\mathbf{P}] = (\lambda_1, \ldots, \lambda_d)$ *given by* (3.3)*. Then*

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

The proof of Theorem 3.1 is obtained by applying Theorem 1.1. When $d = 1$, Equation (1.1) corresponds to the condition required in [25, Theorem 1.2], which establishes sharper Poisson approximation results than the one obtained in the univariate case from Theorem 1.1. Therefore, for the sum of dependent Bernoulli random variables, a sharper bound for the Wasserstein distance can be derived from [25, Theorem 1.2], while for the total variation distance a bound may be deduced from [1, Theorem 1], [25, Theorem 1.2], or [32, Theorem 1].

As a consequence of Theorem 3.1, we obtain the following result for the sum of independent Bernoulli random vectors.

Corollary 3.1. *For* $n \in \mathbb{N}$*, let* $\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(n)}$ *be independent Bernoulli random vectors with distribution given by* (3.1)*, and let* $\mathbf{X}$ *be the random vector defined by* (3.2)*. Let* $\mathbf{P} = (P_1, \ldots, P_d)$ *be a Poisson random vector with mean* $\mathbb{E}[\mathbf{P}] = (\lambda_1, \ldots, \lambda_d)$ *given by* (3.3)*. Then*

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

In [27, Theorem 1], a sharper bound for the total variation distance than the one obtained by Corollary 3.1 is proven. When the vectors are identically distributed and $\sum_{j=1}^{d} p_{1,j} \leq \alpha/n$ for some constant $\alpha > 0$, our bound for the Wasserstein distance and the one in [27, Theorem 1] for the total variation distance differ only by a constant that does not depend on n, d, or the probabilities $p_{i,j}$.

Proof of Theorem 3.1. Without loss of generality we may assume that $\lambda_1, \ldots, \lambda_d > 0$. Define the random vectors

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

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

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

$$\begin{aligned} \ell_i \mathbb{P}(X_{1:i} = \ell_{1:i}) &= \mathbb{E} \sum_{k=1}^{n} \mathbf{1}\{\mathbf{Y}^{(k)} = \mathbf{e}_i\} \mathbf{1}\{X_{1:i} = \ell_{1:i}\} \qquad (3.4) \\ &= \mathbb{E} \sum_{k=1}^{n} \mathbf{1}\{\mathbf{Y}^{(k)} = \mathbf{e}_i\} \mathbf{1}\big\{X_{1:i}^{(k)} + W_{1:i}^{(k)} = (\ell_{1:i-1}, \ell_i - 1)\big\}. \end{aligned}$$

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press

Let $H_{\ell_{1:i}}$ and $q_{\ell_{1:i}}$ be the quantities given by

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

$$q_{\ell_{1:i}} = \ell_i \mathbb{P}(X_{1:i} = \ell_{1:i}) - H_{\ell_{1:i}}.$$

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

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

Since $\mathbf{Y}^{(r)}$, $r = 1, \dots, n$, are m-dependent, the random vectors $\mathbf{Y}^{(k)} = \left(Y_1^{(k)}, \dots, Y_d^{(k)}\right)$ and $\mathbf{X}^{(k)}$ are independent for all $k = 1, \dots, n$. Therefore

$$\begin{aligned} H_{\ell_{1:i}} &= \sum_{k=1}^{n} p_{k,i}\mathbb{P}\big(X_{1:i}^{(k)} = (\ell_{1:i-1}, \ell_i - 1)\big) \\ &= \sum_{k=1}^{n} p_{k,i}\mathbb{P}\big(X_{1:i} - W_{1:i}^{(k)} - Y_{1:i}^{(k)} = (\ell_{1:i-1}, \ell_i - 1)\big) \\ &= \lambda_i \mathbb{P}\big(X_{1:i} - W_{1:i}^{(\tau_i)} - Y_{1:i}^{(\tau_i)} = (\ell_{1:i-1}, \ell_i - 1)\big). \end{aligned}$$

Then, by Theorem 1.1 we obtain

$$d_W(\mathbf{X}, \mathbf{P}) \le \sum_{i=1}^{d}\left(\lambda_i \mathbb{E}\left[W_i^{(\tau_i)} + Y_i^{(\tau_i)}\right] + 2\lambda_i \sum_{j=1}^{i-1}\mathbb{E}\left[W_j^{(\tau_i)} + Y_j^{(\tau_i)}\right] + \sum_{\substack{\ell_{1:i}\in\mathbb{N}_0^d \\ \ell_i \neq 0}} \left|q_{\ell_{1:i}}\right|\right). \tag{3.5}$$

From (3.4) and the definition of $q_{\ell_{1:i}}$ it follows that

$$\begin{aligned} |q_{\ell_{1:i}}| &\le \mathbb{E}\sum_{k=1}^{n} \mathbf{1}\{\mathbf{Y}^{(k)} = \mathbf{e}_i\}\left|\mathbf{1}\big\{X_{1:i}^{(k)} + W_{1:i}^{(k)} = (\ell_{1:i-1}, \ell_i - 1)\big\} - \mathbf{1}\big\{X_{1:i}^{(k)} = (\ell_{1:i-1}, \ell_i - 1)\big\}\right| \\ &\le \mathbb{E}\sum_{k=1}^{n} \mathbf{1}\{\mathbf{Y}^{(k)} = \mathbf{e}_i\}\mathbf{1}\{W_{1:i}^{(k)} \neq 0\}\mathbf{1}\big\{X_{1:i}^{(k)} + W_{1:i}^{(k)} = (\ell_{1:i-1}, \ell_i - 1)\big\} \\ &\quad + \mathbb{E}\sum_{k=1}^{n} \mathbf{1}\{\mathbf{Y}^{(k)} = \mathbf{e}_i\}\mathbf{1}\{W_{1:i}^{(k)} \neq 0\}\mathbf{1}\big\{X_{1:i}^{(k)} = (\ell_{1:i-1}, \ell_i - 1)\big\}. \end{aligned}$$

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

$$\begin{aligned} \sum_{\substack{\ell_{1:i}\in\mathbb{N}_0^i \\ \ell_i \neq 0}} \left|q_{\ell_{1:i}}\right| &\le 2\mathbb{E}\sum_{k=1}^{n} \mathbf{1}\{\mathbf{Y}^{(k)} = \mathbf{e}_i\}\mathbf{1}\{W_{1:i}^{(k)} \neq 0\} \\ &\le 2\mathbb{E}\sum_{k=1}^{n}\sum_{j=1}^{i} \mathbf{1}\{\mathbf{Y}^{(k)} = \mathbf{e}_i\}W_j^{(k)} \le 4mi\sum_{k=1}^{n} Q(k). \end{aligned} \tag{3.6}$$

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press

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 (3.5) and (3.6), this leads to

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

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

which completes the proof. □

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 $\xi = \delta_{X_1} + \cdots + \delta_{X_\tau}$, for some random elements $X_1, X_2, \dots$ in $\mathbb{X}$ and a random variable $\tau \in \mathbb{N}_0 \cup \{\infty\}$. Note that any Poisson process can be seen as a proper point process, and that all locally finite point processes are proper if $(\mathbb{X}, \mathcal{X})$ is a Borel space; see e.g. [19, Corollaries 3.7 and 6.5]. The so-called reduced Campbell measure $\mathcal{C}$ of ξ is defined on the product space $(\mathbb{X} \times \mathsf{N}_{\mathbb{X}}, \mathcal{X} \otimes \mathcal{N}_{\mathbb{X}})$ by

$$\mathcal{C}(A) = \mathbb{E} \int_{\mathbb{X}} \mathbf{1}_A(x, \xi \setminus x)\, \xi(dx), \quad A \in \mathcal{X} \otimes \mathcal{N}_{\mathbb{X}},$$

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 $(\mathsf{N}_{\mathbb{X}}, \mathcal{N}_{\mathbb{X}})$. If $\mathcal{C}$ is absolutely continuous with respect to $\nu \otimes \mathbb{P}_\xi$, any density c of $\mathcal{C}$ with respect to $\nu \otimes \mathbb{P}_\xi$ is called (a version of) the Papangelou intensity of ξ. This notion was originally introduced by Papangelou in [24]. In other words, c is a Papangelou intensity of ξ relative to the measure ν if the Georgii–Nguyen–Zessin equation

$$\mathbb{E} \int_{\mathbb{X}} u(x, \xi \setminus x)\, \xi(dx) = \int_{\mathbb{X}} \mathbb{E}[c(x, \xi) u(x, \xi)] \nu(dx) \tag{3.7}$$

is satisfied for all measurable functions $u : \mathbb{X} \times \mathsf{N}_{\mathbb{X}} \to [0, \infty)$. Intuitively, $c(x, \xi)$ is a random variable that measures the interaction between x and ξ; as a reinforcement of this idea, 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 $\lambda(A) = \int_A f(x) \nu(dx)$, $A \in \mathcal{X}$; see e.g. [19, Theorem 4.1]. For more details on this interpretation we refer the reader to [11, Section 4]; see also [18] and [31] for connections between the Papangelou intensity and Gibbs point processes.

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press

In the next theorem we prove a bound for the d_π distance between a point process ξ that admits Papangelou intensity relative to a measure ν, and a Poisson process η with intensity measure λ absolutely continuous with respect to ν. For a locally compact metric space, Theorem 3.2 yields the same bound as [11, Theorem 3.7], but for the metric d_π instead of the Kantorovich–Rubinstein distance.

Theorem 3.2. *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 $\int_{\mathbb{X}} \mathbb{E}|c(x,\xi)|\nu(dx) < \infty$. Let η be a Poisson process on $\mathbb{X}$ with finite intensity measure λ having density f with respect to ν. Then*

$$d_\pi(\xi,\eta) \le \int_{\mathbb{X}} \mathbb{E}\,|c(x,\xi) - f(x)|\,\nu(dx).$$

Proof of Theorem 3.2. The condition $\int_{\mathbb{X}} \mathbb{E}|c(x,\xi)|\nu(dx) < \infty$ and Equation (3.7) ensure that ξ has finite intensity measure. Consider $i \in \mathbb{N}$ and $(A_1,\dots,A_i) \in \mathcal{X}^i_{\mathrm{disj}}$. 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 1.2 with the random vectors $\mathbf{Z}^{A_{1:i}}$ assumed to be $\mathbf{0}$. In this case,

$$\begin{aligned} q^{A_{1:i}}_{m_{1:i}} &= m_i \mathbb{P}\big(\xi(A_{1:i}) = m_{1:i}\big) - \lambda(A_i)\mathbb{P}\big(\xi(A_{1:i}) = (m_{1:i-1}, m_i - 1)\big) \\ &= m_i \mathbb{P}\big(\xi(A_{1:i}) = m_{1:i}\big) - \int_{\mathbb{X}} \mathbb{E}\big[f(x)\mathbf{1}_{A_i}(x)\mathbf{1}\{\xi(A_{1:i}) = (m_{1:i-1}, m_i - 1)\}\big]\nu(dx) \end{aligned}$$

for $m_{1:i} \in \mathbb{N}_0^i$ with $m_i \neq 0$, $i = 1,\dots,d$. It follows from (3.7) that

$$\begin{aligned} m_i \mathbb{P}\big(\xi(A_{1:i}) = m_{1:i}\big) &= \mathbb{E}\int_{\mathbb{X}} \mathbf{1}_{A_i}(x)\mathbf{1}\{\xi \setminus x(A_{1:i}) = (m_{1:i-1}, m_i - 1)\}\,\xi(dx) \\ &= \int_{\mathbb{X}} \mathbb{E}\big[c(x,\xi)\mathbf{1}_{A_i}(x)\mathbf{1}\{\xi(A_{1:i}) = (m_{1:i-1}, m_i - 1)\}\big]\nu(dx); \end{aligned}$$

hence

$$q^{A_{1:i}}_{m_{1:i}} = \int_{\mathbb{X}} \mathbb{E}\big[(c(x,\xi) - f(x))\mathbf{1}_{A_i}(x)\mathbf{1}\{\xi(A_{1:i}) = (m_{1:i-1}, m_i - 1)\}\big]\nu(dx).$$

Theorem 1.2 yields

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

The inequalities

$$\begin{aligned} \sum_{\substack{m_{1:i}\in\mathbb{N}_0^i \\ m_i\neq 0}} \left| q^{A_{1:i}}_{m_{1:i}} \right| &\le \sum_{\substack{m_{1:i}\in\mathbb{N}_0^i, \\ m_i\neq 0}} \int_{\mathbb{X}} \mathbb{E}\big[|c(x,\xi) - f(x)|\mathbf{1}_{A_i}(x)\mathbf{1}\{\xi(A_{1:i}) = (m_{1:i-1}, m_i - 1)\}\big]\nu(dx) \\ &\le \int_{\mathbb{X}} \mathbb{E}\Big[|c(x,\xi) - f(x)|\mathbf{1}_{A_i}(x) \sum_{\substack{m_{1:i}\in\mathbb{N}_0^i \\ m_i\neq 0}} \mathbf{1}\{\xi(A_{1:i}) = (m_{1:i-1}, m_i - 1)\}\Big]\nu(dx) \\ &\le \int_{\mathbb{X}} \mathbb{E}\big[|c(x,\xi) - f(x)|\mathbf{1}_{A_i}(x)\big]\nu(dx) \end{aligned}$$

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press

imply that

$$\sum_{i=1}^{d} \sum_{\substack{m_{1:i} \in \mathbb{N}_0^i \\ m_i \neq 0}} \left| q_{m_{1:i}}^{A_{1:i}} \right| \leq \int_{\mathbb{X}} \mathbb{E} \left| c(x, \xi) - f(x) \right| \nu(dx)$$

for any $A_{1:d} \in \mathcal{X}_{\mathrm{disj}}^d$ with $d \in \mathbb{N}$. Thus, we obtain the assertion. □

3.3. Point processes of Poisson *U*-statistic structure

Let $(\mathbb{X}, \mathcal{X})$ and $(\mathbb{Y}, \mathcal{Y})$ be measurable spaces. For $k \in \mathbb{N}$ and a symmetric domain $D \in \mathcal{X}^k$, let $g : D \to \mathbb{Y}$ be a symmetric measurable function; i.e., for any $(x_1, \ldots, x_k) \in D$ and index permutation σ, $(x_{\sigma(1)}, \ldots, x_{\sigma(k)}) \in D$ and $g(x_1, \ldots, x_k) = g(x_{\sigma(1)}, \ldots, x_{\sigma(k)})$. Let η be a Poisson process on $\mathbb{X}$ with finite intensity measure μ. We are interested in the point process on $\mathbb{Y}$ given by

$$\xi = \frac{1}{k!} \sum_{(x_1, \ldots, x_k) \in \eta_{\neq}^k \cap D} \delta_{g(x_1, \ldots, x_k)}, \tag{3.8}$$

where $\eta_{\neq}^k$ denotes the collection of all k-tuples $(x_1, \ldots, x_k)$ of points from η with pairwise distinct indices. 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 of order k. We refer to the monographs [17, 20] for more details on U-statistics and their applications to statistics. Hereafter we discuss the Poisson process approximation in the metric d_π for the point process ξ. We prove the exact analogue of [10, Theorem 3.1], with the Kantorovich–Rubinstein distance replaced by d_π. Several applications of this result are presented in [10], alongside the case of underlying binomial point processes. It is worth mentioning that [10] relies on a slightly less general setup: $\mathbb{X}$ is assumed to be an lcscH space, while in the present work any measurable space is allowed.

Let λ denote the intensity measure of ξ, and note that, since μ is a finite measure on $\mathbb{X}$, by the multivariate Mecke formula $\lambda(\mathbb{Y}) < \infty$. Define

$$R = \max_{1 \leq i \leq k-1} \int_{\mathbb{X}^i} \left(\int_{\mathbb{X}^{k-i}} \mathbf{1}\{(x_1, \ldots, x_k) \in D\} \, \mu^{k-i}(d(x_{i+1}, \ldots, x_k)) \right)^2 \mu^i(d(x_1, \ldots, x_i))$$

for $k \geq 2$, and put $R = 0$ for $k = 1$.

Theorem 3.3. *Let ξ, λ, and R be as above, and let γ be a Poisson process on $\mathbb{Y}$ with intensity measure λ. Then*

$$d_\pi(\xi, \gamma) \leq \frac{2^{k+1}}{k!} R.$$

If the intensity measure λ of ξ is the zero measure, then the proof of Theorem 3.3 is trivial. From now on, we assume $0 < \lambda(\mathbb{Y}) < \infty$. The multivariate Mecke formula yields for every $A \in \mathcal{Y}$ that

$$\lambda(A) = \mathbb{E}[\xi(A)] = \frac{1}{k!} \mathbb{E} \sum_{\mathbf{x} \in \eta_{\neq}^k \cap D} \mathbf{1}\{g(\mathbf{x}) \in A\} = \frac{1}{k!} \int_D \mathbf{1}\{g(\mathbf{x}) \in A\} \, \mu^k(d\mathbf{x}).$$

Define the random element $\mathbf{X}^A = (X_1^A, \ldots, X_k^A)$ in $\mathbb{X}^k$ independent of η and distributed according to

$$\mathbb{P}\Big(\mathbf{X}^A \in B\Big) = \frac{1}{k! \lambda(A)} \int_D \mathbf{1}\{g(\mathbf{x}) \in A\} \mathbf{1}\{\mathbf{x} \in B\} \, \mu^k(d\mathbf{x})$$

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press

for all B in the product σ-field of $\mathbb{X}^k$ when $\lambda(A) > 0$, and set $\mathbf{X}^A = \mathbf{x}_0$ for some $\mathbf{x}_0 \in \mathbb{X}^k$ when $\lambda(A) = 0$. For any vector $\mathbf{x} = (x_1, \dots, x_k) \in \mathbb{X}^k$, denote by $\Delta(\mathbf{x})$ the sum of k Dirac measures located at the vector components; that is,

$$\Delta(\mathbf{x}) = \Delta(x_1, \dots, x_k) = \sum_{i=1}^{k} \delta_{x_i} .$$

In what follows, for any point process ζ on $\mathbb{X}$, $\xi(\zeta)$ is the point process defined as in (3.8) with η replaced by ζ. Furthermore, as in Section 3.2, $\xi(A_{1:i})$ denotes the random vector $(\xi(A_1), \dots, \xi(A_i))$, for any $A_1, \dots, A_i \in \mathcal{Y}$, $i \in \mathbb{N}$.

Proof of Theorem 3.3. For $k = 1$, Theorem 3.3 is a direct consequence of [19, Theorem 5.1]. We therefore assume $k \geq 2$. 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

$$\begin{aligned}
m_i \mathbb{P}(\xi(A_{1:i}) = m_{1:i}) &= \frac{1}{k!} \mathbb{E} \sum_{\mathbf{x} \in \eta_{\neq}^k \cap D} \mathbf{1}\{g(\mathbf{x}) \in A_i\} \mathbf{1}\{\xi(A_{1:i}) = m_{1:i}\} \\
&= \frac{1}{k!} \int_D \mathbf{1}\{g(\mathbf{x}) \in A_i\} \mathbb{P}(\xi(\eta + \Delta(\mathbf{x}))(A_{1:i}) = m_{1:i}) \, \mu^k(d\mathbf{x}) \\
&= \frac{1}{k!} \int_D \mathbf{1}\{g(\mathbf{x}) \in A_i\} \mathbb{P}\big(\xi(\eta + \Delta(\mathbf{x}))(A_{1:i}) - \delta_{g(\mathbf{x})}(A_{1:i}) = (m_{1:i-1}, m_i - 1)\big) \, \mu^k(d\mathbf{x}) \\
&= \lambda(A_i) \mathbb{P}\left(\xi\left(\eta + \Delta\left(\mathbf{X}^{A_i}\right)\right)(A_{1:i}) - \delta_{g\left(\mathbf{X}^{A_i}\right)}(A_{1:i}) = (m_{1:i-1}, m_i - 1) \right),
\end{aligned} \tag{3.9}$$

where the second-to-last equality 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 also satisfied 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 (1.7) in Theorem 1.2 is zero. Note that $\mathbf{Z}^{A_{1:i}}$ has nonnegative components. Hence, for any $d \in \mathbb{N}$ and $(A_1, \dots, A_d) \in \mathcal{X}_{\mathrm{disj}}^d$,

$$\begin{aligned}
\sum_{i=1}^{d} \lambda(A_i) \sum_{j=1}^{i} \mathbb{E}\left|\mathbf{Z}_j^{A_{1:i}}\right| &= \sum_{i=1}^{d} \lambda(A_i) \sum_{j=1}^{i} \mathbb{E}\left[\xi\left(\eta + \Delta\left(\mathbf{X}^{A_i}\right)\right)(A_j) - \xi(A_j) - \delta_{g\left(\mathbf{X}^{A_i}\right)}(A_j) \right] \\
&\leq \sum_{i=1}^{d} \lambda(A_i) \mathbb{E}\left[\xi\left(\eta + \Delta\left(\mathbf{X}^{A_i}\right)\right)(\mathbb{Y}) - \xi(\mathbb{Y}) - 1 \right] \\
&= \frac{1}{k!} \sum_{i=1}^{d} \int_D \mathbf{1}\{g(\mathbf{x}) \in A_i\} \mathbb{E}\left[\xi(\eta + \Delta(\mathbf{x}))(\mathbb{Y}) - \xi(\mathbb{Y}) - 1\right] \mu^k(d\mathbf{x}) \\
&\leq \lambda(\mathbb{Y}) \mathbb{E}\left[\xi\left(\eta + \Delta\left(\mathbf{X}^{\mathbb{Y}}\right)\right)(\mathbb{Y}) - \xi(\mathbb{Y}) - 1 \right].
\end{aligned}$$

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press

Thus, Theorem 1.2 gives

$$d_\pi(\xi, \gamma) \le 2\lambda(\mathbb{Y})\mathbb{E}\left[\xi\left(\eta + \Delta\left(\mathbf{X}^{\mathbb{Y}}\right)\right)(\mathbb{Y}) - \xi(\mathbb{Y}) - 1\right]. \tag{3.10}$$

From (3.9) with $i = 1$ and $A_1 = \mathbb{Y}$, it follows that the random variable $\xi\left(\eta + \Delta\left(\mathbf{X}^{\mathbb{Y}}\right)\right)(\mathbb{Y})$ has the size bias distribution of $\xi(\mathbb{Y})$. The property (1.6) with f being the identity function and simple algebraic computations yield

$$\begin{aligned}\mathbb{E}\left[\xi\left(\eta + \Delta\left(\mathbf{X}^{\mathbb{Y}}\right)\right)(\mathbb{Y}) - \xi(\mathbb{Y}) - 1\right] &= \lambda(\mathbb{Y})^{-1}\left\{\mathbb{E}\left[\xi(\mathbb{Y})^2\right] - \lambda(\mathbb{Y})^2 - \lambda(\mathbb{Y})\right\} \\ &= \lambda(\mathbb{Y})^{-1}\left\{\mathrm{Var}(\xi(\mathbb{Y})) - \lambda(\mathbb{Y})\right\}.\end{aligned} \tag{3.11}$$

Moreover, [26, Lemma 3.5] gives

$$\mathrm{Var}(\xi(\mathbb{Y})) - \lambda(\mathbb{Y}) \le \sum_{i=1}^{k-1} \frac{1}{k!}\binom{k}{i} R \le \frac{2^k - 1}{k!} R\,.$$

These inequalities combined with (3.10) and (3.11) deliver the assertion. □

Acknowledgements

The authors would like to thank Chinmoy Bhattacharjee, Ilya Molchanov, and Matthias Schulte for valuable comments. The authors are also thankful to one anonymous referee for precise comments.

Funding information

This research was supported by the Swiss National Science Foundation, grant number 200021_175584.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] Arratia, R., Goldstein, L. and Gordon, L. (1989). Two moments suffice for Poisson approximations: the Chen–Stein method. *Ann. Prob.* **17**, 9–25.

[2] Arratia, R., Goldstein, L. and Gordon, L. (1990). Poisson approximation and the Chen–Stein method. *Statist. Sci.* **5**, 403–434.

[3] Barbour, A. D. (1988). Stein's method and Poisson process convergence. *J. Appl. Prob.* **25A**, 175–184.

[4] Barbour, A. D. (2005). Multivariate Poisson-binomial approximation using Stein's method. In *Stein's Method and Applications*, Singapore University Press, pp. 131–142.

[5] Barbour, A. D. and Brown, T. C. (1992). Stein's method and point process approximation. *Stoch. Process. Appl.* **43**, 9–31.

[6] Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation*. Oxford University Press.

[7] Barbour, A. D. and Xia, A. (2006). On Stein's factors for Poisson approximation in Wasserstein distance. *Bernoulli* **12**, 943–954.

[8] Brown, T. C. and Xia, A. (2001). Stein's method and birth–death processes. *Ann. Prob.* **29**, 1373–1403.

[9] Chen, L. H. Y. and Xia, A. (2004). Stein's method, Palm theory and Poisson process approximation. *Ann. Prob.* **32**, 2545–2569.

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press

[10] DECREUSEFOND, L., SCHULTE, M. AND THÄLE, C. (2016). Functional Poisson approximation in Kantorovich–Rubinstein distance with applications to U-statistics and stochastic geometry. *Ann. Prob.* **44**, 2147–2197.
[11] DECREUSEFOND, L. AND VASSEUR, A. (2018). Stein's method and Papangelou intensity for Poisson or Cox process approximation. Preprint. Available at https://arxiv.org/abs/1807.02453.
[12] DEHEUVELS, P. AND PFEIFER, D. (1988). Poisson approximations of multinomial distributions and point processes. *J. Multivariate Anal.* **25**, 65–89.
[13] ČEKANAVIČIUS, V. AND VELLAISAMY, P. (2020). Compound Poisson approximations in ℓ_p-norm for sums of weakly dependent vectors. *J. Theoret. Prob.* **34**, 2241–2264.
[14] ERHARDSSON, T. (2005). Stein's method for Poisson and compound Poisson approximation. In *An Introduction to Stein's Method*, Singapore University Press, pp. 61–113.
[15] GOLDSTEIN, L. AND RINOTT, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Prob.* **33**, 1–17.
[16] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd edn. Springer, New York.
[17] KOROLJUK, V. S. AND BOROVSKICH, Y. V. (1994). *Theory of U-Statistics*. Kluwer Academic Publishers Group, Dordrecht.
[18] LAST, G. AND OTTO, M. (2021). Disagreement coupling of Gibbs processes with an application to Poisson approximation. Preprint. Available at https://arxiv.org/abs/2104.00737.
[19] LAST, G. AND PENROSE, M. (2018). Lectures on the Poisson Process. Cambridge University Press.
[20] LEE, A. J. (1990). *U-Statistics*. Marcel Dekker, New York.
[21] LIEB, E. H. AND LOSS, M. (2001). *Analysis*, 2nd edn. American Mathematical Society, Providence, RI.
[22] MCSHANE, E. J. (1934). Extension of range of functions. *Bull. Amer. Math. Soc.* **40**, 837–842.
[23] NOVAK, S. Y. (2019). Poisson approximation. *Prob. Surveys* **16**, 228–276.
[24] PAPANGELOU, F. (1973/74). The conditional intensity of general point processes and an application to line processes. *Z. Wahrscheinlichkeitsth.* **28**, 207–226.
[25] PIANOFORTE, F. AND SCHULTE, M. (2021). Poisson approximation with applications to stochastic geometry. Preprint. Available at https://arxiv.org/abs/2104.02528.
[26] REITZNER, M. AND SCHULTE, M. (2013). Central limit theorems for U-statistics of Poisson point processes. *Ann. Prob.* **41**, 3879–3909.
[27] ROOS, B. (1999). On the rate of multivariate Poisson convergence. *J. Multivariate Anal.* **69**, 120–134.
[28] ROOS, B. (2003). Poisson approximation of multivariate Poisson mixtures. *J. Appl. Prob.* **40**, 376–390.
[29] ROOS, B. (2017). Refined total variation bounds in the multivariate and compound Poisson approximation. *ALEA Latin Amer. J. Prob. Math. Statist.* **14**, 337–360.
[30] SCHUHMACHER, D. (2009). Stein's method and Poisson process approximation for a class of Wasserstein metrics. *Bernoulli* **15**, 550–568.
[31] SCHUHMACHER, D. AND STUCKI, K. (2014). Gibbs point process approximation: total variation bounds using Stein's method. *Ann. Prob.* **42**, 1911–1951.
[32] SMITH, R. L. (1988). Extreme value theory for dependent sequences via the Stein–Chen method of Poisson approximation. *Stochastic Process. Appl.* **30**, 317–327.
[33] VILLANI, C. (2009). *Optimal Transport*. Springer, Berlin.

https://doi.org/10.1017/jpr.2022.33 Published online by Cambridge University Press