

Asymptotic Distributions for the Performance Analysis of Hypothesis Testing of Isolated-Point-Penalization Point Processes

Majeed M. Hayat, *Member, IEEE*, John A. Gubner, *Member, IEEE*, and Sajjad Abdullah, *Student Member, IEEE*

Abstract— The performance of the likelihood ratio test is considered for a many-point interaction point process featuring a reduced number of isolated points. Limit theorems are proved that establish the Poissonian asymptotic distribution of the log-likelihood function for point processes with the isolated-point-penalization joint probability density function. The asymptotic distribution is used to approximate the detection probability associated with the likelihood ratio test. The approximation is compared to empirical results generated using Markov-chain Monte Carlo simulation. The reported results provide an efficient alternative method to simulation in assessing the performance of hypothesis testing for the point-process model considered.

Index Terms— Detection probability, hypothesis testing, isolated points, many-point interaction, point process, Poisson approximation.

I. INTRODUCTION

Spatial point processes are natural models for signals arising in many applications such as forestry, seismology, image analysis, and statistical minefield detection, where information is represented by the number and the location of points in Euclidean space [4], [11]. The simplest of all point processes is the Poisson process, for which, conditional on the number of points, the constituent points are independently and uniformly distributed in a given region. The points, thus, exhibit no interaction, and the point pattern is often referred to as *totally random*. An important feature that may distinguish one point process from another is whether the points exhibit mutual interaction. For example, certain models for the spatial distribution of trees in a forest or the distribution of mines in a minefield assume that the presence of a point results in an inhibition effect reducing the likelihood of the presence of other points in its vicinity. The Strauss process is a simple example of such a point process for which the likelihood of realizations containing many pairs of points that are within a certain fixed range is reduced [12]. The Strauss process is

Manuscript received February 6, 1998; revised June 26, 1998. This work was supported by ONR under Grant N00014-94-1-0366. The material in this paper was presented in part at SPIE AEROSENSE '97, Orlando, FL, April 1997.

M. M. Hayat and S. Abdullah are with the Department of Electrical and Computer Engineering, University of Dayton, Dayton, OH 45469-0245 USA (e-mail: mhayat@engr.udayton.edu).

J. A. Gubner is with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706-1691 USA (e-mail: gubner@engr.wisc.edu).

Communicated by P. Moulin, Associate Editor for Nonparametric Estimation, Classification, and Neural Networks.

Publisher Item Identifier S 0018-9448(99)00054-1.

a special case of the more general pairwise interaction point processes for which the joint probability density function (pdf) of the points is *only* a function of the distances between pairs of points [1]. In certain applications, the interaction mechanism between points is more complex than pairwise interaction. The joint pdf in such cases not only depends on the distances between points in pairs of points but also on the distances between points in larger groups of points (e.g., groups of three or even all the points). Such point processes are often referred to as many-point interaction point processes. For example, certain minefield models [5], [6], assume that the likelihood of finding an isolated mine is low. The corresponding joint pdf for this isolated-point-penalization (IPP) point process is a function of distances between points in all groups of points [6].

It is often of interest to determine whether the points in an observed point pattern are totally random. In minefield detection, for example, totally random patterns are potentially classified as clutter. If an alternative model (e.g., pairwise or many-point-interaction point processes) to the total randomness hypothesis is available, then statistical hypothesis testing can be performed. Due to the complex nature of the joint pdf of interaction point processes, exact performance analysis of hypothesis testing is generally intractable even in the simple case of the Strauss process. Monte Carlo simulation and approximation theory have been extensively used in estimating the performance of hypothesis testing problems involving interaction point processes [3], [9].

In this paper, we present an approximation to the detection probability for the IPP point process. In Section II, we briefly describe the IPP process and indicate its applications in minefield modeling. In Section III, we present and prove a new theorem that establishes the Poisson approximation of the number of isolated points when the points are totally random. This result is then generalized to the IPP process which is used to establish that the distribution of the log-likelihood function of an IPP point process can be approximated by a Poisson distribution. The performance analysis of the likelihood ratio test using the Poisson approximation is compared to empirical results generated using Markov-chain Monte Carlo simulation.

II. PRELIMINARIES

Let B be a bounded subset of the plane. For any integer n , let $X = (X_1, X_2, \dots, X_n)$ be a B^n -valued random vector with joint pdf f . The vector X represents the location of n points in B distributed based on the pdf f . The random

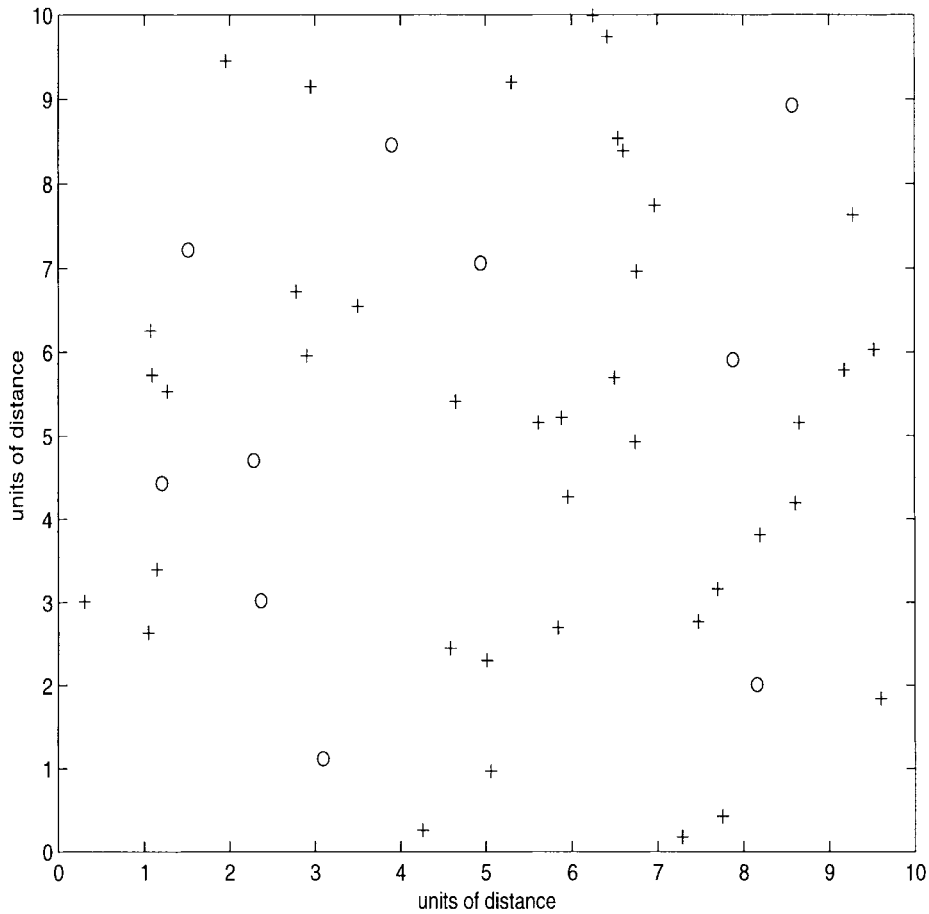


Fig. 1. A 50-point realization of a totally random X in the region $B = [0, 10] \times [0, 10]$. Isolated points are designated by circles. There are ten isolated points ($d = 1.0$) in this realization of X .

vector X can be thought of as a conditional point process given that each realization of the point process has exactly n points. By selecting the form of f , various interaction schemes between points can be generated. The pdf's considered in this paper correspond to IPP many-point interaction point processes. Throughout the examples, we will assume that $B = [0, 10] \times [0, 10]$. All simulations are generated using the Markov-chain Monte Carlo algorithm reported in [8]. We will now give a brief description of the IPP process.

A. Isolated-Point-Penalization Point Processes

We first give a precise definition of the number of isolated points. For a given $d > 0$ and $x = (x_1, \dots, x_n) \in B^n$, the number of *isolated points* in x is defined by

$$I(x) := \{i: \|x_i - x_j\| > d, \forall j \neq i, x_i \in A\} \quad (1)$$

where $A = \{\xi \in B: D(\xi, d) \subset B\}$, $D(\xi, d)$ is the closed disc of radius d centered at ξ , $|E|$ denotes the cardinality of the set E , and $\|\cdot\|$ is the Euclidean norm. In words, $I(x)$ represents the number of isolated points in x that are at least within a distance d away from the boundary of B . The IPP pdf f on B^n is defined by

$$f(x) = \frac{1}{z} \gamma^{I(x)}, \quad x \in B^n. \quad (2)$$

The *penalization strength* $0 < \gamma \leq 1$ is responsible for penalizing realizations with many isolated points, and z is a normalizing constant making f a pdf. Clearly, when $c = 1$, the random variables $X_i, i = 1, 2, \dots, n$, are independent and uniformly distributed in B , and the resulting point pattern is totally random. The parameter d is called the *interaction distance* and it plays an important role in the hypotheses of the approximation theorems of Section III.

Example 1: Fig. 1 shows a 50-point realization of a totally random pattern (i.e., drawn from the IPP pdf with $c = 1$), and Fig. 2 shows a 50-point realization of X sampled from an IPP pdf with $c = 0.3$ and $d = 1.0$. Note that the number of isolated points is reduced from ten in the case of the totally random pattern to only one in the case of the IPP pattern. It is interesting to note that, as a result of the reduction of isolated points, there is a tendency for points to cluster. The IPP model has therefore the potential for use in situations when clustering is desirable. Note that it is always possible to define a function on the space of all point configurations whose restriction to the space of n -point configurations ($n \geq 1$) is the IPP pdf. A point process with a random number of points can then be defined by defining its distribution as the product of the above function and a Poisson measure. Note that for $c \leq 1$, a distribution defined in the above sense is always dominated by the Poisson measure. Therefore, an IPP point

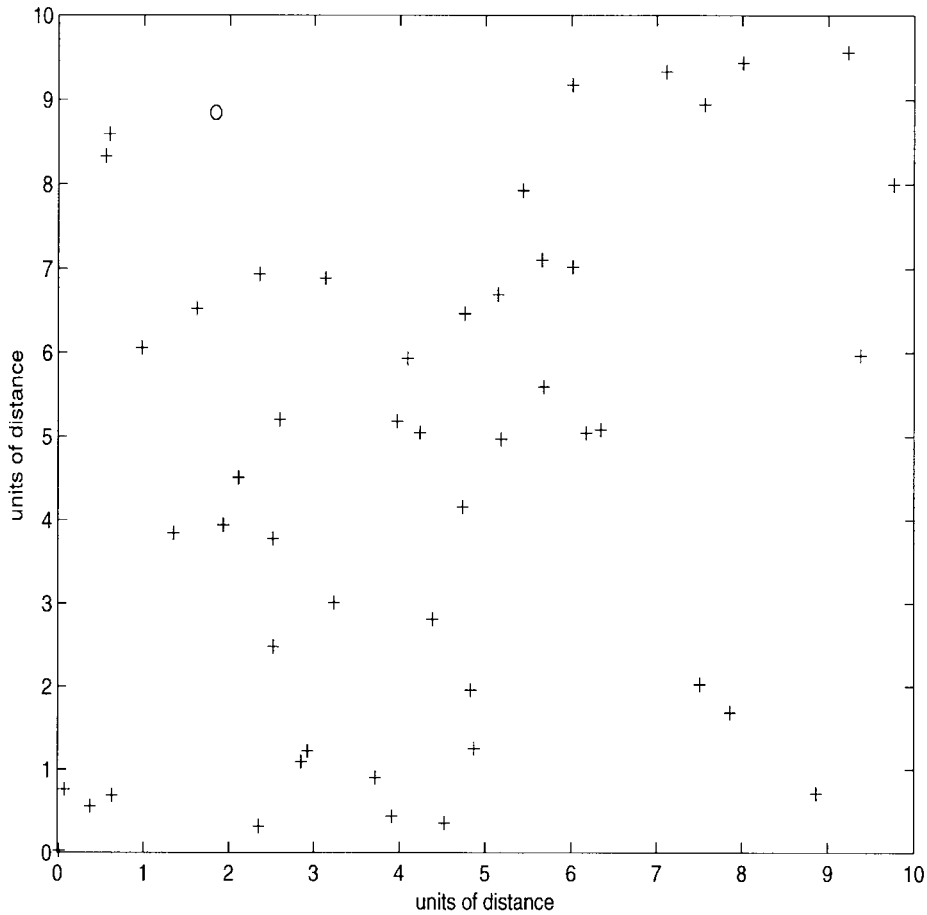


Fig. 2. A 50-point realization of an IPP X with $c = 0.3$ and $d = 1.0$. The process X again takes values in $B = [0, 10] \times [0, 10]$. Isolated points are designated by circles. Note that the number of isolated points is reduced from ten in Fig. 1 to one.

process with a random number of points is always well-defined in a sense that its density with respect to the Poisson measure is integrable. This integrability feature is not inherent in many pairwise-interaction clustering densities [1], [7].

III. HYPOTHESIS TESTING AND PERFORMANCE ANALYSIS

In this section we consider the problem of binary hypothesis testing when the observed signal is a point process drawn from the IPP probability density function. We present two limit theorems which are used to approximate the detection probability under certain asymptotic conditions. Since the asymptotic results are established in the limit when the size of the region and the number of points both increases to infinity, it is necessary to replace the observation region B with a sequence of bounded sets $B_n, n = 1, 2, \dots$. Each B_n is assumed to contain n points. Furthermore, the set A is replaced by the sequence

$$A_n = \{\xi \in B_n: D(\xi, d) \subset B_n\}. \tag{3}$$

Now let $X^n = (X_1^n, X_2^n, \dots, X_n^n)$ be a random vector in B_n^n , and let H_0 denote the hypothesis that X^n is totally random. Let H_1 denote the hypothesis that X^n is an IPP point process. Note that the pdf in (2) along with the associated normalizing constant are dependent on n . However, to simplify notation, this dependence is not explicitly expressed. The

likelihood ratio test for the hypothesis testing problem can be simplified to

$$I(x^n) \underset{H_0}{\overset{H_1}{\geq}} y \tag{4}$$

where y is a specified detection threshold. For performance evaluation, we need to compute probabilities of the form

$$P\{I(X^n) > y\} \tag{5}$$

when X^n is either an IPP process or a totally random process. The difficulty of this problem is due to the fact that evaluating the distribution of $I(X^n)$ is generally intractable. This difficulty can be alleviated by determining the asymptotic distribution of $I(X^n)$ and using it to approximate the detection probability (5). We now proceed to establish the Poisson approximation of $I(X^n)$.

A. Poisson Approximation of $I(X^n)$

It is intuitive to suspect that the asymptotic behavior of $I(X^n)$, as $n \rightarrow \infty$, is ultimately governed by the way the area of B_n increases as a function of the number of points n it contains. In the case of a totally random process, for example, if the area grows at a much faster rate than n , then it is clear that $I(X^n)$ will diverge to ∞ since almost every point will eventually become isolated. On the other hand, if

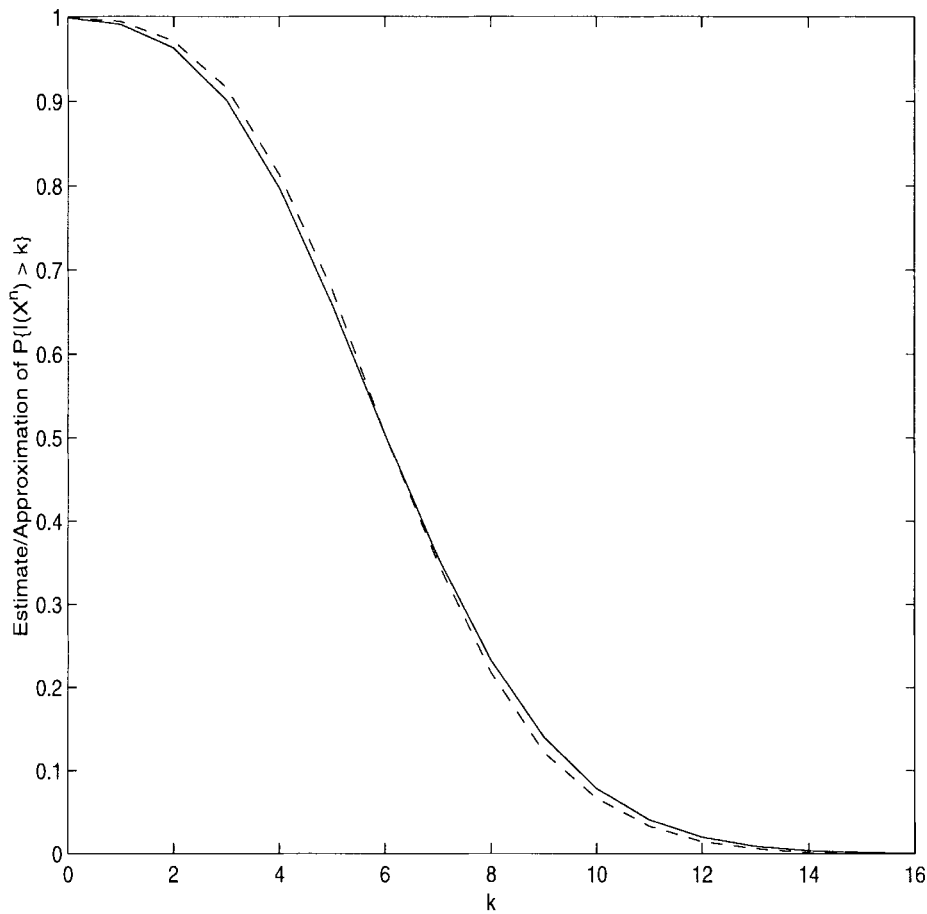


Fig. 3. The empirical estimate (dashed line) and the Poisson approximation (solid line) of the probability in (5) for $n = 50$ and $d = 1$ when X^n is totally random, i.e., $c = 1$.

the area grows at a much slower rate than n , then we suspect that $I(X^n)$ will converge to zero since the points will be become tightly packed in the region. Interestingly, there is an intermediate growth of the area for which the mean of $I(X^n)$ converges to a constant. We refer to the latter form of growth as the *denseness condition* to be formally defined next. Let μ be Lebesgue measure in the plane.

Definition: For a fixed $d > 0$, we say that the sequence of subsets B_1, B_2, \dots , satisfies the denseness condition with parameter λ if

- 1) $\lim_{n \rightarrow \infty} \log(n/\lambda)\mu(B_n)/n = \pi d^2$;
- 2) $\lim_{n \rightarrow \infty} \mu(A_n)/\mu(B_n) = 1$, where A_n is defined in (3); and
- 3) as $n \rightarrow \infty$, the number of disjoint discs of radius d contained in B_n diverges to infinity.

Remark: The denseness condition is satisfied if B_n is a square with area $\pi d^2 n / \log(n/\lambda)$.

Remark: The denseness condition is analogous to the sparseness condition introduced by Saunders and Funk [10] in the context of pairwise interaction (PI) point processes. Under the sparseness assumption, the mean number of pairs of points (from a PI process) that are within a certain fixed distance converges to a constant. Furthermore, the asymptotic distribution of the number of pairs is Poisson. The sparseness

and denseness conditions cannot be simultaneously satisfied in practical cases.

The following Theorem (see Appendix for proof) asserts that if X^n is totally randomly distributed and the denseness condition is satisfied, then the distribution of the random variable $I(X^n)$ is asymptotically Poisson.

Theorem 1: Suppose that the random variables X_1^n, \dots, X_n^n are independent and uniformly distributed in B_n and that the sequence $B_n, n = 1, 2, \dots$, satisfies the denseness condition with parameter λ . Then, the random variable $I(X^n)$ converges in distribution, as $n \rightarrow \infty$, to a Poisson random variable with mean λ .

In the examples to follow, B_n is taken to be the $[0, 10] \times [0, 10]$ square. The parameter λ can be evaluated using $\lambda = n \exp(-\pi d^2 n / 100) \cdot (1 - d/5)^2$. The factor $(1 - d/5)^2$ accounts for the reduced area of A_n in comparison to B_n .

Example 2: Figs. 3 and 4 compare the Poisson approximation of the probability in (5) to the empirical estimate using 5000 samples of the totally random process for $n = 50$ and $n = 75$, respectively. The interaction distance $d = 1$. The values for λ for the cases $n = 50$ and $n = 75$ are 6.65 and 4.55, respectively. These values of n are found to be large enough to show good approximation.

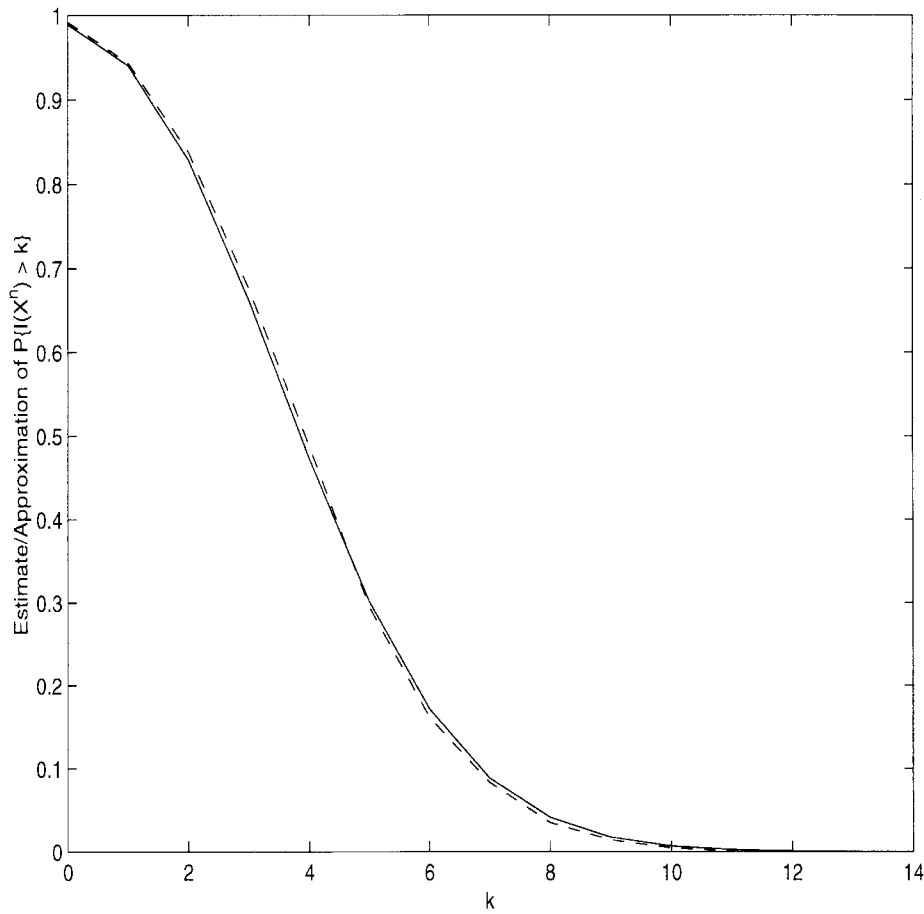


Fig. 4. Same as Fig. 3 but with $n = 75$.

The next theorem extends Theorem 1 to the case when X^n has an IPP pdf.

Theorem 2: Suppose that $X^n = (X_1^n, \dots, X_n^n)$ has the IPP pdf given in (2) and that the sequence $B_n, n = 1, 2, \dots$, satisfies the denseness condition with parameter λ . Then, $I(X^n)$ converges in distribution to a Poisson random variable with mean $c\lambda$.

Proof: For each n , let $Q_n^0(s) := E[e^{sI(X^n)} | H_0]$ denote the moment-generating function (mgf) of the random variable $I(X^n)$ under the hypothesis H_0 , i.e.,

$$Q_n^0(s) = \mu(B_n)^{-n} \int_{B_n^n} e^{sI(x)} dx.$$

Similarly, let $Q_n(s) := E[e^{sI(X^n)} | H_1]$ denote the mgf of the random variable $I(X^n)$ under hypothesis H_1 . Observe that

$$\begin{aligned} Q_n(s) &= \frac{1}{z_n} \int_{B_n^n} e^{sI(x)} c^{I(x)} dx \\ &= \frac{1}{z_n} \int_{B_n^n} e^{I(x)[s+\log c]} dx \\ &= \frac{\mu(B_n)^n}{\int_{B_n^n} c^{I(x)} dx} Q_n^0(s + \log c) \\ &= \frac{Q_n^0(s + \log c)}{Q_n^0(\log c)}. \end{aligned}$$

As $n \rightarrow \infty$, Theorem 1 dictates that

$$Q_n^0(s) \rightarrow \exp\{-\lambda(1 - e^s)\}$$

the mgf of a Poisson random variable with mean λ . Therefore,

$$\begin{aligned} \lim_{n \rightarrow \infty} Q_n(s) &= \lim_{n \rightarrow \infty} \frac{Q_n^0(s + \log c)}{Q_n^0(\log c)} \\ &= \exp\{-c\lambda(1 - e^s)\} \end{aligned}$$

which is the mgf of a Poisson random variable with mean $c\lambda$. \square

To illustrate the above result, the probability in (5) when the alternative hypothesis is an IPP pdf is computed using the above approximations and compared to empirical results obtained using Markov-chain Monte Carlo simulation.

Example 3: Consider the IPP model with parameters $c = 0.61$, and $d = 1.0$. Figs. 5 and 6 compare the empirical estimate of the probability in (5) using 5000 samples of the IPP process with the estimate obtained using the Poisson approximation for the cases $n = 50$ and $n = 75$, respectively. The values of the parameter λ are 4.06 and 2.78 for the cases $n = 50$ and $n = 75$, respectively.

IV. SUMMARY

Since direct evaluation of the performance of hypothesis testing for the majority of interaction point processes is not feasible, this paper was devoted to developing limit theorems

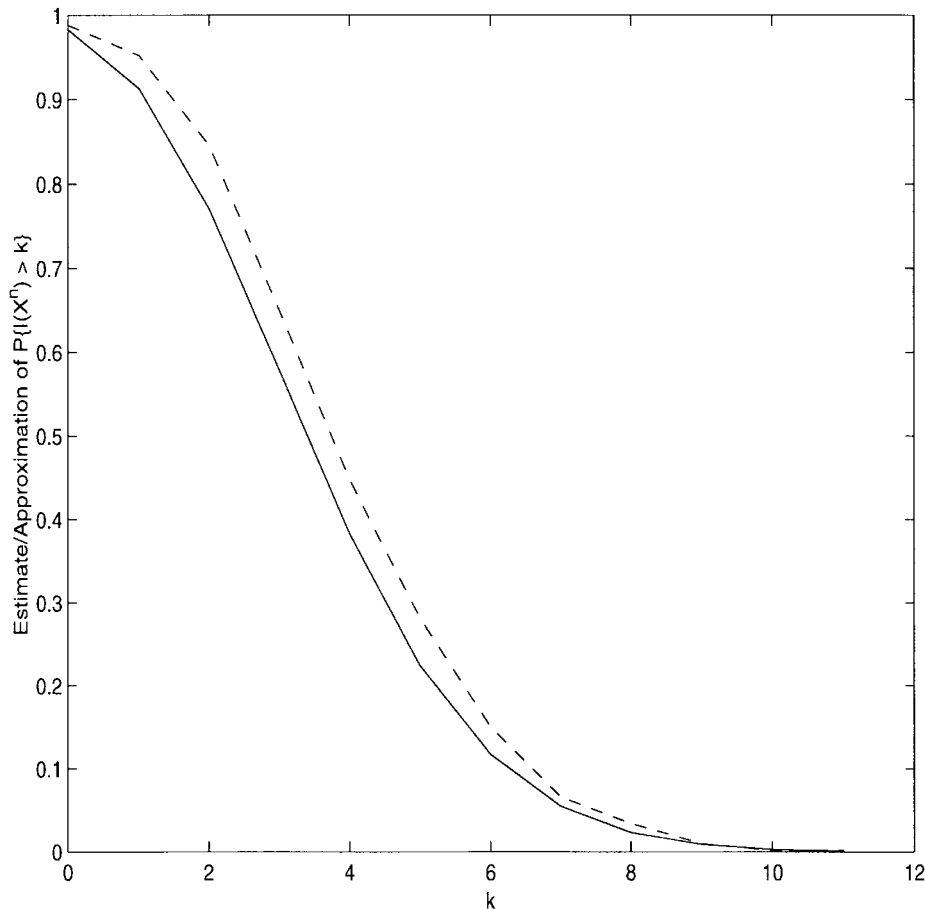


Fig. 5. The empirical estimate (dashed line) and the Poisson approximation (solid line) of the probability in (5) when the alternative hypothesis is an IPP pdf with $c = 0.61$, $n = 50$, and $d = 1.0$.

that are used to approximate probabilities that arise in the performance analysis for hypothesis testing of the isolated-point-penalization many-point-interaction point process. The asymptotic distribution function of the log-likelihood function is shown to be a Poisson distribution under certain asymptotic requirements, called the denseness condition. The denseness condition relates the growth of the area of the observation region and the number of points contained. The reported approximation performs well and uses only a fraction of the computing time required by Markov-chain Monte Carlo simulation. For example, running the simulation for Example 3 required approximately 150 min on a SUN SPARC-20 Workstation while the Poisson approximation was executed in a few seconds.

APPENDIX PROOF OF THEOREM 1

We begin with some preliminary observations. We will need the fact that for any positive integer i , the first denseness condition implies that

$$\lim_{n \rightarrow \infty} n^i \left(1 - \frac{i\pi d^2}{\mu(B_n)}\right)^{n-i} = \lambda^i. \quad (6)$$

This can be derived by taking logarithms and noting that

$$|\log(1+x) - x| \leq 2x^2, \quad |x| \leq 1/2$$

which follows from the Taylor series for $\log(1+x)$ with

differential error bound. Also note that the first limit of the denseness condition implies that $\mu(B_n)/n \rightarrow 0$ and that $n/\mu(B_n)^2 \rightarrow 0$. (This second fact follows because for $0 < \epsilon < \pi d^2$ and for large n , $\mu(B_n) > (\pi d^2 - \epsilon)n/\log(n/\lambda)$.)

Next, using the identity

$$\lambda^{n-1} e^{-\lambda} = \frac{1}{n} \left[\frac{d}{d\lambda} (\lambda^n e^{-\lambda}) + \lambda^n e^{-\lambda} \right]$$

a simple calculation shows that for a Poisson random variable Y with mean λ , the $(k+1)$ th moment $m_{k+1}(\lambda) := E[Y^{k+1}]$ satisfies the recursive relation

$$m_{k+1}(\lambda) = \lambda(m_k(\lambda) + m'_k(\lambda)). \quad (7)$$

In particular, it follows from (7) that $m_{k+1}(\lambda)$ is a polynomial in λ of degree $(k+1)$

$$m_{k+1}(\lambda) = \lambda + a_{k+1,2}\lambda^2 + \dots + a_{k+1,k}\lambda^k + \lambda^{k+1}. \quad (8)$$

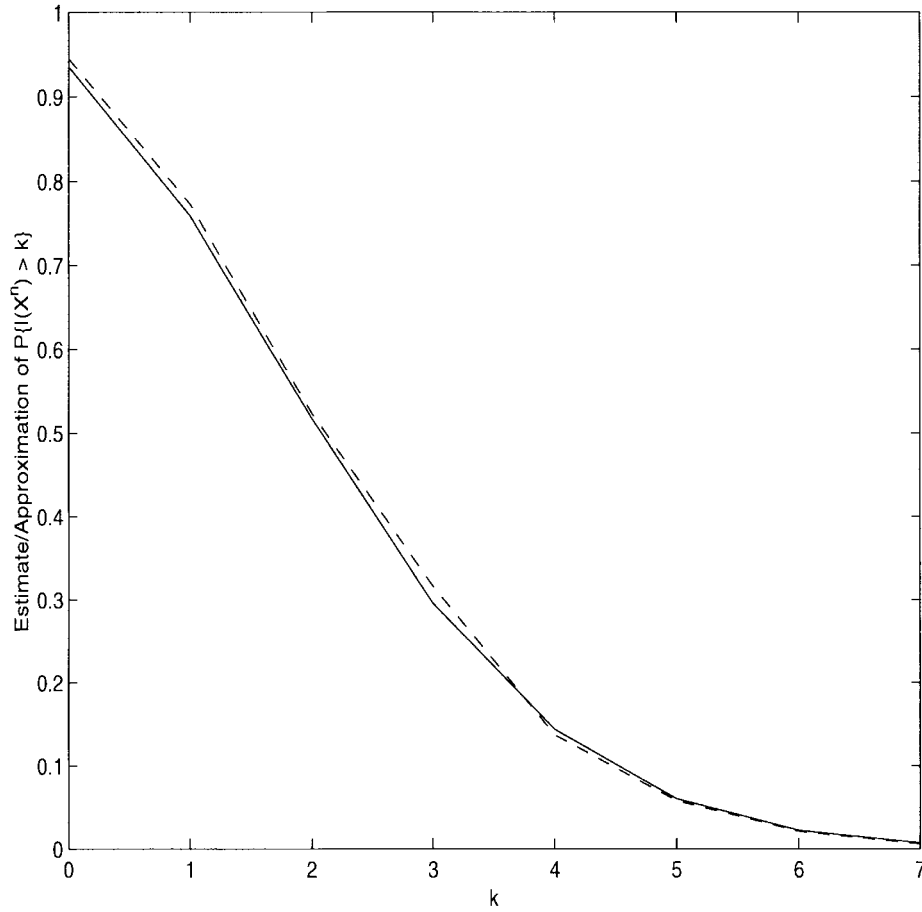
It follows from (7) and (8) that

$$a_{k+1,i} = a_{k,i-1} + ia_{k,i}, \quad i = 2, 3, \dots, k. \quad (9)$$

Note that $a_{k,1} = a_{k,k} = 1$ for all $k \geq 1$. The key step in proving the Theorem is to show that for each $k \geq 1$

$$\lim_{n \rightarrow \infty} E[I^k(X^n)] = m_k(\lambda).$$

We then invoke the standard method-of-moments result [2, p. 390, Theorem 30.2] which establishes the convergence of


 Fig. 6. Same as Fig. 5 but with $n = 75$.

$I(X^n)$ to Y in distribution, as $n \rightarrow \infty$. To this end, observe that

$$I(X^n) = \sum_{i=1}^n \mathbf{1}_{A_n}(X_i^n) \prod_{j=1, j \neq i}^n \chi(\|X_j^n - X_i^n\|/d) \quad (10)$$

where $\mathbf{1}_A(\cdot)$ is the indicator function of a set A , and $\chi(\cdot) = \mathbf{1}_{(1, \infty)}(\cdot)$. Then

$$\begin{aligned} \mathbb{E}[I^k(X^n)] &= \mathbb{E} \left[\sum_{i_1=1}^n \cdots \sum_{i_k=1}^n \mathbf{1}_{A_n}(X_{i_1}^n) \cdots \mathbf{1}_{A_n}(X_{i_k}^n) \right. \\ &\quad \cdot \prod_{j_1 \neq i_1}^n \chi(\|X_{j_1}^n - X_{i_1}^n\|/d) \\ &\quad \left. \cdots \prod_{j_k \neq i_k}^n \chi(\|X_{j_k}^n - X_{i_k}^n\|/d) \right] \\ &= \sum_{i_1=1}^n \cdots \sum_{i_k=1}^n \mathbb{E} \left[\mathbf{1}_{A_n}(X_{i_1}^n) \cdots \mathbf{1}_{A_n}(X_{i_k}^n) \right. \\ &\quad \cdot \prod_{j_1 \neq i_1}^n \chi(\|X_{j_1}^n - X_{i_1}^n\|/d) \\ &\quad \left. \cdots \prod_{j_k \neq i_k}^n \chi(\|X_{j_k}^n - X_{i_k}^n\|/d) \right]. \end{aligned} \quad (11)$$

For each $1 \leq k \leq n$ and $0 \leq i \leq k$, let $t_i^{(k)}(n)$ denote the number of terms in $(x_1 + \cdots + x_n)^k$ that consist of exactly i distinct factors (e.g., $x_2^{k-4}x_5x_6^3$ is a term consisting of three factors). Using a simple combinatorial argument, it can be shown that

$$t_i^{(k+1)}(n) = it_i^{(k)}(n) + (n-i+1)t_{i-1}^{(k)}(n) \quad (12)$$

where

$$\begin{aligned} t_0^{(k)}(n) &= 0 \\ t_i^{(i)}(n) &= n(n-1)\cdots(n-i+1) \\ t_1^{(k)}(n) &= n \end{aligned}$$

for all $k \geq 1$, and $t_j^{(i)}(n) = 0$ whenever $j > i$. (The idea behind (12) is that each term with exactly i distinct factors in $(x_1 + \cdots + x_n)^k$ will result in i terms each with i distinct factors once $(x_1 + \cdots + x_n)^k$ is multiplied by $(x_1 + \cdots + x_n)$; on the other hand, each term with exactly $i-1$ distinct factors in $(x_1 + \cdots + x_n)^k$ will generate $n-i+1$ terms each with i distinct factors as a result of the multiplication.) It is also true that for each $1 \leq k \leq n$ and $i \leq k$, $t_i^{(k)}(n)$ is a polynomial of degree i in n . (This can be proven inductively (in i) using (12).) Using (12) and (9), and the fact that $t_i^{(k)}(n)$ is a polynomial in n of degree i , it can be easily shown inductively that

$$t_i^{(k)}(n) = a_{k,i}n^i + l_{i-1}^{(k)}(n) \quad (13)$$

where $l_{i-1}^{(k)}$ is a polynomial of degree at most $i-1$ in n .

Using the independent and identical distribution (i.i.d.) assumption of the sequence $\{X_i^n\}_{i=1}^n$ along with the above notation, (11) can be simplified to

$$\mathbb{E}[I^k(X^n)] = \sum_{i=1}^k t_i^{(k)}(n) \mathbb{E} \left[\mathbf{1}_{A_n}(X_1^n) \cdots \mathbf{1}_{A_n}(X_i^n) \prod_{j_1 \neq 1}^n \chi(\|X_{j_1}^n - X_1^n\|/d) \cdots \prod_{j_i \neq i}^n \chi(\|X_{j_i}^n - X_i^n\|/d) \right]. \quad (14)$$

We now introduce some events to be used in the calculation of (14). The event that the first i points are in A_n is

$$\mathcal{T}_i := \bigcap_{k=1}^i \{X_k^n \in A_n\}.$$

The event that X_i^n is isolated from X_k^n for $k \leq j$, $k \neq i$ is denoted by

$$F_i^j := \bigcap_{k \leq j, k \neq i} \{\|X_k^n - X_i^n\| > d\}.$$

Note that \mathcal{T}_i and F_i^j depend on n . This dependence is not indicated in the notation here and below. The event that each of the first i points is isolated with respect to all of the other $n-1$ points is written as

$$\mathcal{F}_i^n := F_1^n \cap \cdots \cap F_i^n$$

which can be simplified to

$$\mathcal{F}_i^n = \bigcap_{\ell=1}^i \left(\bigcap_{k=\ell+1}^n \{\|X_k^n - X_\ell^n\| > d\} \right). \quad (15)$$

In any case, the i th expectation in right-hand side of (14) is simply $\mathbb{P}(\mathcal{T}_i \cap \mathcal{F}_i^n)$, i.e.,

$$\mathbb{E}[I^k(X^n)] = \sum_{i=1}^k t_i^{(k)}(n) \mathbb{P}(\mathcal{T}_i \cap \mathcal{F}_i^n). \quad (16)$$

The $i=1$ term is

$$t_1^{(k)}(n) \mathbb{P}(\mathcal{T}_1 \cap \mathcal{F}_1^n) = n \mathbb{P}(\mathcal{T}_1 \cap \mathcal{F}_1^n).$$

Since $\mathcal{T}_1 = \{X_1^n \in A_n\}$ is an event depending only on X_1^n , we can write

$$\mathbb{P}(\mathcal{T}_1 \cap \mathcal{F}_1^n) = \int_{\mathcal{T}_1} \mathbb{P}(\mathcal{F}_1^n | X_1^n) d\mathbb{P}.$$

Next, since $\mathcal{F}_1^n = F_1^n = \bigcap_{k=2}^n \{\|X_k^n - X_1^n\| > d\}$

$$\begin{aligned} \mathbb{P}(\mathcal{F}_1^n | X_1^n = x) &= \mathbb{P} \left(\bigcap_{k=2}^n \{\|X_k^n - X_1^n\| > d\} | X_1^n = x \right) \\ &= \prod_{k=2}^n \mathbb{P}(\|X_k^n - x\| > d). \end{aligned}$$

Now, since the above integral is over $\mathcal{T}_1 = \{X_1^n \in A_n\}$, we only need to consider $x \in A_n$. For such x we have

$$\mathbb{P}(\mathcal{F}_1^n | X_1^n = x) = \left(1 - \frac{\pi d^2}{\mu(B_n)} \right)^{n-1}$$

and thus

$$\mathbb{P}(\mathcal{T}_1 \cap \mathcal{F}_1^n) = \left(1 - \frac{\pi d^2}{\mu(B_n)} \right)^{n-1} \cdot \frac{\mu(A_n)}{\mu(B_n)}.$$

From this last expression, it follows that

$$\lim_{n \rightarrow \infty} n \mathbb{P}(\mathcal{T}_1 \cap \mathcal{F}_1^n) = \lambda = m_1(k).$$

To generalize the above limit to the case $i \geq 2$, we partition the set $\mathcal{T}_i \cap \mathcal{F}_i^n$, and write

$$\mathbb{P}(\mathcal{T}_i \cap \mathcal{F}_i^n) = \mathbb{P}(\mathcal{T}_i \cap \mathcal{F}_i^n \cap \tilde{\mathcal{F}}^i) + \mathbb{P}(\mathcal{T}_i \cap \mathcal{F}_i^n \cap (\tilde{\mathcal{F}}^i)^c)$$

where $\tilde{\mathcal{F}}^i$ is defined below. We also rewrite the first term on the right as

$$\mathbb{P}(\mathcal{T}_i \cap \mathcal{F}_i^n \cap \tilde{\mathcal{F}}^i) = \mathbb{P}(\mathcal{F}_i^n | \mathcal{T}_i \cap \tilde{\mathcal{F}}^i) \mathbb{P}(\tilde{\mathcal{F}}^i | \mathcal{T}_i) \mathbb{P}(\mathcal{T}_i).$$

In particular, the above partition and conditioning becomes useful in estimating the probability $\mathbb{P}(\mathcal{T}_i \cap \mathcal{F}_i^n)$ if we let \tilde{F}_i^j be defined as F_i^j except that d is replaced by $2d$, i.e.,

$$\tilde{F}_i^j := \bigcap_{k \leq j, k \neq i} \{\|X_k^n - X_i^n\| > 2d\}.$$

Then for $i = 2, \dots, k$, put

$$\tilde{\mathcal{F}}^i := \tilde{F}_1^i \cap \cdots \cap \tilde{F}_i^i.$$

Recall that the event \mathcal{F}_i^n is the event that each of the first i points is isolated (distance d) with respect to the remaining $n-1$ points, whereas $\tilde{\mathcal{F}}^i$ is the event that each of the first i points is isolated (distance $2d$) with respect to the other $i-1$ points. (The idea is that the conditional probability $\mathbb{P}(\mathcal{F}_i^n | \mathcal{T}_i \cap \tilde{\mathcal{F}}^i)$ is easy to calculate since it is, roughly speaking, the probability of randomly scattering $n-i$ point in the region A_n with i disjoint discs, each with radius d , removed from it.)

Of course, $\mathbb{P}(\mathcal{T}_i) = [\mu(A_n)/\mu(B_n)]^i \rightarrow 1$ as $n \rightarrow \infty$. For $i \geq 2$, we show a little later that as $n \rightarrow \infty$

$$\begin{aligned} \mathbb{P}(\tilde{\mathcal{F}}^i | \mathcal{T}_i) &\rightarrow 1 \\ n^i \mathbb{P}(\mathcal{F}_i^n | \mathcal{T}_i \cap \tilde{\mathcal{F}}^i) &\rightarrow \lambda^i \\ n^i \mathbb{P}(\mathcal{F}_i^n \cap \mathcal{T}_i \cap (\tilde{\mathcal{F}}^i)^c) &\rightarrow 0. \end{aligned} \quad (17)$$

Using these limits, it further follows that for $j < i$

$$\lim_{n \rightarrow \infty} n^j \mathbb{P}(\mathcal{T}_i \cap \mathcal{F}_i^n) = \lim_{n \rightarrow \infty} \frac{n^i \mathbb{P}(\mathcal{T}_i \cap \mathcal{F}_i^n)}{n^{i-j}} = 0. \quad (18)$$

Upon substituting (13) into (16), and applying the limit results in (17) and (18), we conclude that

$$\lim_{n \rightarrow \infty} \mathbb{E}[I^k(X^n)] = \sum_{i=1}^k a_{k,i} \lambda^i = m_k(\lambda)$$

as desired.

The final step in the proof of the Theorem is to establish the limits in (17). This is quite lengthy, and it requires the remainder of this appendix. The convergence of $\mathbb{P}(\tilde{\mathcal{F}}^i | \mathcal{T}_i)$ to 1, for $i=2$, can be proven by conditioning on the location of X_1 , and the convergence for $i \geq 3$ is proven by induction and by conditioning on the number of points that have no neighbors within a distance $2d$.

With respect to $\mathbb{P}(\mathcal{F}_i^n | \mathcal{I}_i \cap \tilde{\mathcal{F}}^i)$, observe that for n large enough so that B_n can contain i disjoint discs each of radius d (i.e., $\tilde{\mathcal{F}}^i \neq \emptyset$), the independence and uniformity of the distributions of X_{i+1}, \dots, X_n imply that

$$\begin{aligned} & \mathbb{P}(\mathcal{F}_i^n | \mathcal{I}_i \cap \tilde{\mathcal{F}}^i) \\ &= \mathbb{P}\left(\bigcap_{\ell=1}^i \left(\bigcap_{k=i+1}^n \{\|X_\ell^n - X_k^n\| > d\}\right) \middle| \mathcal{I}_i \cap \tilde{\mathcal{F}}^i\right) \\ &= \left(1 - \frac{i\pi d^2}{\mu(B_n)}\right)^{n-i}. \end{aligned}$$

(Note how the lower limit of the inner intersection has changed compared with (15). This simplification follows because the omitted sets have conditional probability one; note that $\{\|X_\ell^n - X_k^n\| > 2d\} \subset \{\|X_\ell^n - X_k^n\| > d\}$.) Using (6), we obtain

$$\lim_{n \rightarrow \infty} n^i \mathbb{P}(\mathcal{F}_i^n | \mathcal{I}_i \cap \tilde{\mathcal{F}}^i) = \lambda^i.$$

We now show that for $i \geq 2$

$$\lim_{n \rightarrow \infty} n^i \mathbb{P}(\mathcal{F}_i^n \cap \mathcal{I}_i \cap (\tilde{\mathcal{F}}^i)^c) = 0. \quad (19)$$

For a given set of $i \geq 2$ points $\{x_1, x_2, \dots, x_i\}$, we say that the points $x_\ell, x_m, 1 \leq \ell, m \leq i$, are related by the relation \sim if either $\|x_\ell - x_m\| \leq 2d$, or there exist distinct points

$$x_{j_1}, \dots, x_{j_r} \in \{x_1, \dots, x_i\} \setminus \{x_\ell, x_m\}$$

such that $\|x_\ell - x_{j_1}\| \leq 2d$

$$\|x_{j_1} - x_{j_2}\| \leq 2d, \dots, \|x_{j_{r-1}} - x_{j_r}\| \leq 2d$$

and $\|x_{j_r} - x_m\| \leq 2d$. We write $x_\ell \sim x_m$ to express the above relation. Clearly, \sim defines an equivalence relation on the set $\{x_1, x_2, \dots, x_i\}$, and hence, it induces a partition of the set. Let \mathcal{C} be the collection of equivalence classes induced by \sim (i.e., if $C \in \mathcal{C}$, then for any $x_\ell, x_m \in C, x_\ell \sim x_m$, and furthermore, $x_\ell \not\sim x_j$ and $x_m \not\sim x_j$ for all $x_j \notin C$). It may be helpful to think of an equivalence class C as a ‘‘cluster’’ of points. Note that $1 \leq |C| \leq i$. Now for $1 \leq j \leq i$ let

$$N_j := |\{C \in \mathcal{C} : |C| = j\}|$$

i.e., N_j is the number of equivalence classes that contain exactly j points. Clearly

$$\sum_{j=1}^i j N_j = i. \quad (20)$$

Note that if we now repeat the partitioning process with the set $\{x_1, x_2, \dots, x_i\}$ replaced with the set of random variables $\{X_1^n, X_2^n, \dots, X_i^n\}$, then the quantity $N_j, 1 \leq j \leq i$, becomes a random variable. By recalling that the event $\tilde{\mathcal{F}}^i$ is the event that each of the first i points are $2d$ -isolated with respect to the other $i - 1$ points, we observe that $\tilde{\mathcal{F}}^i = \{N_1 = i\}$ and $(\tilde{\mathcal{F}}^i)^c = \{N_1 \leq i - 1\}$.

We now introduce some events which will be useful in the calculation of $\mathbb{P}(\mathcal{F}_i^n \cap \mathcal{I}_i \cap (\tilde{\mathcal{F}}^i)^c)$. For $1 \leq \ell \leq i$ and $1 \leq j_\ell \leq i$ let

$$V_{j_1 \dots j_i} := \{N_1 = j_1, \dots, N_i = j_i\}.$$

We can now write $(\tilde{\mathcal{F}}^i)^c$ as the disjoint union

$$(\tilde{\mathcal{F}}^i)^c = \bigcup_{j_1, \dots, j_i, j_1 \leq i-1} V_{j_1 \dots j_i}.$$

Note that it is not possible to have $N_1 = i - 1$, since then there would be $i - 1$ clusters, each containing one point. The remaining point would belong to some other cluster, which of course, can contain at most one point, contradicting $N_1 = i - 1$. Hence, the above union can be simplified to

$$(\tilde{\mathcal{F}}^i)^c = \bigcup_{j_1, \dots, j_i, j_1 \leq i-2} V_{j_1 \dots j_i}.$$

We can now write

$$\mathbb{P}(\mathcal{I}_i \cap \mathcal{F}_i^n \cap (\tilde{\mathcal{F}}^i)^c) = \sum_{j_1, \dots, j_i, j_1 \leq i-2} \mathbb{P}(\mathcal{I}_i \cap \mathcal{F}_i^n \cap V_{j_1 \dots j_i}). \quad (21)$$

We now turn to the calculation of

$$\mathbb{P}(\mathcal{I}_i \cap \mathcal{F}_i^n \cap V_{j_1 \dots j_i}). \quad (22)$$

Write

$$\mathbb{P}(\mathcal{I}_i \cap \mathcal{F}_i^n \cap V_{j_1 \dots j_i}) = \int_{\mathcal{I}_i \cap V_{j_1 \dots j_i}} \mathbb{P}(\mathcal{F}_i^n | X_1^n, \dots, X_i^n) d\mathbb{P}.$$

Consider $\mathbb{P}(\mathcal{F}_i^n | X_1^n = x_1, \dots, X_i^n = x_i)$, where for the purposes of bounding the above integral, we can restrict attention to the case where x_1, \dots, x_i each belong to $A_n, N_r = j_r, 1 \leq r \leq i, N_1 \leq i - 2, \sum_{r=1}^i r j_r = i$, and $\|x_\ell - x_m\| > d$ for $1 \leq \ell, m \leq i$. It now follows that $\mathbb{P}(\mathcal{F}_i^n | X_1^n = x_1, \dots, X_i^n = x_i)$ is upper-bounded by

$$\left[1 - \mathbb{P}\left(\bigcup_{\ell=1}^i \{X_{i+1}^n \in D(x_\ell, d)\}\right)\right]^{n-i}.$$

To get an upper bound on this last expression, we need a lower bound on the probability of the union. By partitioning $\{X_1^n, \dots, X_i^n\}$ according to the equivalence classes introduced previously, we can write the above probability of the union as

$$\mathbb{P}\left(\left\{X_{i+1}^n \in \bigcup_{\ell=1}^i \bigcup_{C \in \mathcal{C}, |C|=\ell} \left(\bigcup_{x \in C} D(x, d)\right)\right\}\right).$$

Note that if x and x' belong to different equivalence classes, then $D(x, d)$ and $D(x', d)$ are necessarily disjoint. Hence, we can write

$$\begin{aligned} & \mathbb{P}\left(\bigcup_{\ell=1}^i \{X_{i+1}^n \in D(x_\ell, d)\}\right) \\ &= \sum_{\ell=1}^i j_\ell \mathbb{P}\left(\left\{X_{i+1}^n \in \bigcup_{x \in C_\ell} D(x, d)\right\}\right) \end{aligned}$$

where for $\ell = 1, \dots, i, C_\ell$ is some equivalence class containing ℓ points. Note that in the case where there is more than one equivalence class for a given ℓ , it is not important which of the equivalence classes (each with ℓ points) actually C_ℓ is. (Also note that since all the discs have centers in A_n ,

the discs lie entirely in B_n .) We can now write the probability of the union as

$$\sum_{\ell=1}^i j_\ell \frac{\mu(\cup_{x \in C_\ell} D(x, d))}{\mu(B_n)}.$$

We now lower-bound the area $\mu(\cup_{x \in C_\ell} D(x, d))$. For $\ell = 1$, this area is simply πd^2 . For $\ell = 2$, we use the fact the centers of the two discs are at least d units apart to obtain a lower bound which is strictly greater than $1.5\pi d^2$. For the case $\ell = 3$, the area of the union of the three discs is minimum when the centers of the discs are the vertices of an equilateral triangle of side length d . This triangle can be circumscribed by a larger equilateral triangle of side length $2d$, where the midpoints of the sides of the larger triangle are the centers of the three discs. The area of the three discs is then the area of the larger triangle plus $3 \cdot \frac{1}{2} \pi d^2$; i.e., the area is $(\sqrt{3} + 1.5\pi) d^2$. For the case $4 \leq \ell \leq i$, observe that $\cup_{x \in C_\ell} D(x, d/2)$ is a proper subset of $\cup_{x \in C_\ell} D(x, d)$, and hence

$$\mu\left(\bigcup_{x \in C_\ell} D(x, d)\right) \geq (\ell + \delta_\ell) \pi d^2 / 4$$

where for $\ell \geq 4$, $\delta_\ell > 0$. In fact, the above lower bound also holds for $\ell = 1, 2$, and 3 if we put $\delta_1 = 3, \delta_2 = \epsilon(d)$ where $\epsilon(d) > 4$, and $\delta_3 = 3 + 4\sqrt{3}/\pi$. Hence

$$\mathbb{P}\left(\bigcup_{\ell=1}^i \{X_{i+1}^n \in D(x_\ell, d)\}\right) \geq \frac{\sum_{\ell=1}^i j_\ell (\ell + \delta_\ell) \pi d^2 / 4}{\mu(B_n)}.$$

We can now write

$$\begin{aligned} & \mathbb{P}(\mathcal{I}_i \cap \mathcal{F}_i^n \cap V_{j_1, \dots, j_i}) \\ & \leq \left[1 - \left(\sum_{\ell=1}^i j_\ell (\ell + \delta_\ell) / 4 \right) \frac{\pi d^2}{\mu(B_n)} \right]^{n-i} \mathbb{P}(\mathcal{I}_i \cap V_{j_1, \dots, j_i}). \end{aligned}$$

Combining this result with (21), we see that to establish the last limit in (17), it suffices to prove that

$$n^i \left[1 - \left(\sum_{\ell=1}^i j_\ell (\ell + \delta_\ell) / 4 \right) \frac{\pi d^2}{\mu(B_n)} \right]^{n-i} \mathbb{P}(\mathcal{I}_i \cap V_{j_1, \dots, j_i}) \rightarrow 0.$$

To begin, write the expression on the left as

$$\begin{aligned} & n^{\sum_{\ell=1}^i j_\ell (\ell + \delta_\ell) / 4} \left[1 - \left(\sum_{\ell=1}^i j_\ell (\ell + \delta_\ell) / 4 \right) \frac{\pi d^2}{\mu(B_n)} \right]^{n-i} \\ & \cdot n^{i - \sum_{\ell=1}^i j_\ell (\ell + \delta_\ell) / 4} \mathbb{P}(\mathcal{I}_i \cap V_{j_1, \dots, j_i}). \quad (23) \end{aligned}$$

The product on the left is bounded (it converges to $\lambda^{\sum_{\ell=1}^i j_\ell (\ell + \delta_\ell) / 4}$). Furthermore, we will show that

$$\mathbb{P}(V_{j_1, \dots, j_i}) \leq \text{const}_1 \left(\frac{\log n}{n} \right)^{\sum_{\ell=2}^i (\ell-1) j_\ell}. \quad (24)$$

Note that since $\sum_{\ell=1}^i \ell j_\ell = i$, the above exponent is equal to $i - \sum_{\ell=1}^i j_\ell$. Using this fact and by substituting (24) in (23) and using (20), we can upper-bound (23) by

$$\text{const}_2 (\log n)^{\sum_{\ell=2}^i (\ell-1) j_\ell} \cdot n^{\sum_{\ell=1}^i [1 - 0.25(\ell + \delta_\ell)] j_\ell}. \quad (25)$$

Using the fact that $\delta_1 = 3, \delta_2 > 4, \delta_3 > 5, \delta_\ell > 0$ for $1 \leq \ell \leq i$, and that $j_1 < i$, we conclude that $\sum_{\ell=1}^i [1 - 0.25(\ell + \delta_\ell)] j_\ell < 0$, which guarantees the convergence of (25) to zero, as desired.

We now establish (24). If $W \subset \{1, \dots, i\}$, we let \mathcal{E}_W denote the event that $\{X_r^n, r \in W\}$ is a cluster; i.e., for $r, s \in W, X_r^n \sim X_s^n$, and for $s \notin W, X_r^n \not\sim X_s^n$. Now fix j_1, \dots, j_i . We say that a partition of $\{1, \dots, i\}$ is j -admissible if the partition contains exactly j_r sets of cardinality r for $r = 1, \dots, i$. Let \mathcal{S} denote the set of all j -admissible partitions. Then

$$V_{j_1, \dots, j_i} = \bigcup_{\mathcal{K} \in \mathcal{S}} \bigcap_{W \in \mathcal{K}} \mathcal{E}_W$$

and

$$\mathbb{P}(V_{j_1, \dots, j_i}) \leq \sum_{\mathcal{K} \in \mathcal{S}} \mathbb{P}\left(\bigcap_{W \in \mathcal{K}} \mathcal{E}_W\right).$$

The next step is to observe that $\mathcal{E}_W \subset \mathcal{E}'_W$, where \mathcal{E}'_W denotes the event that for all $r, s \in W$, either $\|X_r^n - X_s^n\| \leq 2d$ or there exist indices $r_1, \dots, r_p \in W, p \leq |W| - 2$, with $\|X_r^n - X_{r_1}^n\| \leq 2d, \|X_{r_1}^n - X_{r_2}^n\| \leq 2d, \dots, \|X_{r_{p-1}}^n - X_{r_p}^n\| \leq 2d$, and $\|X_{r_p}^n - X_s^n\| \leq 2d$. Now we can further write $\mathcal{E}'_W \subset \mathcal{E}''_W$, where

$$\mathcal{E}''_W := \bigcap_{r, s \in W} \{\|X_r^n - X_s^n\| \leq (|W| - 1) \cdot 2d\}.$$

With this notation, we can write

$$\mathbb{P}\left(\bigcap_{W \in \mathcal{K}} \mathcal{E}_W\right) \leq \mathbb{P}\left(\bigcap_{W \in \mathcal{K}} \mathcal{E}''_W\right).$$

Since the W in \mathcal{K} are disjoint, it is easy to see that the \mathcal{E}''_W are independent. Hence, the probability on the right is equal to a product of probabilities of the form $\mathbb{P}(\mathcal{E}''_W)$. We claim that $\mathbb{P}(\mathcal{E}''_W) \leq \text{const}_3 [(\log n)/n]^{|W|-1}$. The claim then implies (24). Now suppose $|W| = r$ (of course, $r \leq i$). Without loss of generality, take $W = \{1, \dots, r\}$, and observe that

$$\begin{aligned} \mathbb{P}(\mathcal{E}''_W) & \leq \mathbb{P}\left(\bigcap_{s=1}^{r-1} \{\|X_s^n - X_r^n\| \leq 2d(r-1)\}\right) \\ & \leq \left[\frac{\pi [2d(r-1)]^2}{\mu(B_n)} \right]^{r-1}. \end{aligned}$$

As noted at the beginning of the appendix, for large n

$$\mu(B_n) > n(\pi d^2 - \epsilon) / \log(n/\lambda).$$

This establishes the claim. \square

ACKNOWLEDGMENT

The authors thank both reviewers for evaluating this lengthy manuscript so quickly (about five months)! In particular, the authors are grateful to the second reviewer for an unusually careful and meticulous reading of the manuscript, which revealed a number of typos and places where some clarification was in order.

REFERENCES

- [1] A. Baddeley and J. Møller, "Nearest-neighbour Markov point processes and random sets," *Int. Statist. Rev.*, vol. 57, pp. 89–121, 1989.
- [2] P. Billingsley, *Probability and Measure*, 3rd ed. New York: Wiley, 1995.
- [3] W.-B. Chang and J. A. Gubner, "Poisson limits and nonparametric estimation for pairwise interaction point processes," in review.
- [4] N. A. C. Cressie, *Statistics for Spatial Data*. New York: Wiley, 1993.
- [5] M. M. Hayat and J. A. Gubner, "A two-step Markov point process," in *Proc. 1995 IEEE Int. Symp. Information Theory* (Whistler, BC, Canada, Sept. 1995).
- [6] ———, "A two-step Markov point process," Tech. Rep. ECE-96-2, Dept. Elect. Comp. Eng., Univ. Wisconsin–Madison, 1996.
- [7] F. P. Kelly and B. D. Ripley, "A note on Strauss's model for clustering," *Biometrika*, vol. 63, pp. 357–360, 1976.
- [8] B. D. Ripley, "Simulating spatial patterns: Dependent samples from a multivariate density," *Appl. Statist.*, vol. 28, pp. 109–112, 1979.
- [9] ———, *Statistical Inference for Spatial Point Processes*. New York: Cambridge Univ. Press, 1988.
- [10] R. Saunders and G. M. Funk, "Poisson limits for a clustering model of Strauss," *J. Appl. Probab.*, vol. 14, pp. 776–784, 1977.
- [11] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*. New York: Springer-Verlag, 1991.
- [12] D. J. Strauss, "A model for clustering," *Biometrika*, vol. 62, pp. 467–475, 1975.