Nearest-Neighbor and Contact Distance Distributions for Thomas Cluster Process

Mehrnaz Afshang, Chiranjib Saha, and Harpreet S. Dhillon

Abstract—We characterize the statistics of nearest-neighbor and contact distance distributions for Thomas cluster process (TCP), which is a special case of Poisson cluster process. In particular, we derive the cumulative distribution function (CDF) of the distance to the nearest point of TCP from a reference point for three different cases: (i) reference point is not a part of the point process, (ii) it is chosen uniformly at random from the TCP, and (iii) it is a randomly chosen point from a cluster chosen uniformly at random from the TCP. While the first corresponds to the contact distance distribution, the other two provide two different viewpoints for the nearest-neighbor distance distribution. Closed-form bounds are also provided for the first two cases.

Index Terms—Stochastic geometry, Thomas cluster process, contact distance, and nearest-neighbor distance.

I. INTRODUCTION

Poisson cluster processes have numerous applications in diverse branches of science (such as geodesy and ecology) since it captures the attraction (clustering) between nearby points, which is a common occurrence in many point patterns [1]. In wireless networks, the locations of users (network subscribers) tend to be clustered. For instance, users are typically concentrated in some specific areas, e.g., residential and/or commercial complexes, called user hotspots. The clustering nature of user distribution is often considered in the modeling of real-world wireless networks [2]. For example, 3GPP customarily models the users forming clusters over a circular disc in their simulations [3]. Moreover, subscriber-owned small cell base stations (SBSs) are typically deployed on the scale of one per household/business to serve a group (cluster) of users. Thus, the spatial distribution of the SBSs also exhibit clustering [4]. This makes Poisson cluster process a natural choice for modeling user as well as SBS locations in wireless networks. As evident from the prior art on the analysis of wireless networks using tools from stochastic geometry [5], nearest-neighbor and contact distance distributions play a crucial role in the tractable characterization of key performance metrics. However, the derivation of explicit expressions for these distributions for general Poisson cluster processes are not well explored. Thus, despite Poisson cluster process being a suitable model for SBS and user distributions, its applicability in the analysis of wireless networks is significantly restricted.

In this letter, we focus on characterizing the nearest-neighbor and contact distance distributions of TCP, which is a popular special case of Poisson cluster process. The sparsely existing prior art of relevance can be classified in two main directions. The first one focuses on formulating likelihood function of a multinomial Poisson cluster process which yields the desired distributions for a TCP as a special case [6]. Though the results derived are general and applicable for a wide family of Poisson cluster processes, the exposition of the analysis requires substantial proficiency in measure theory and is hence somewhat pedantic to attract general interest in the wireless community. As a result, the second direction, which is a common practice in the literature of wireless networks, is to approximate the first order statistic of the Poisson cluster process with that of homogenous Poisson point process (PPP), and then simply use the well-known contact distribution of PPP; see [7], [8] for a small subset. The substantial complexity of the first direction and inaccuracy of the second direction motivate us to derive the explicit expressions of nearest-neighbor and contact distance distributions of TCP.

In contrast to a homogenous PPP, where the density functions of contact distance and nearest-neighbor distance are identical [5], the choice of a reference point is crucial to characterize such distance distributions in TCP. In this letter, we derive the CDF of distance from a reference point to its nearest point of TCP in three cases: i) the reference point is extraneous to the TCP, ii) the reference point is chosen uniformly at random from the TCP, and iii) a representative cluster is first chosen uniformly at random, and then the reference point is sampled from the representative cluster. We also derive closed-form upper bounds on the CDFs in the first two cases.

II. SYSTEM MODEL

Thomas cluster process is a stationary and isotropic Poisson cluster process generated by a set of offspring points independently and identically distributed (i.i.d.) around each point of a parent PPP [9]. In particular, the locations of parent points are modeled as a homogenous PPP \( \{x\} \equiv \Phi_p \) with density \( \lambda_p \) around which offspring points are distributed according to a symmetric normal distribution with variance \( \sigma^2 \). Thus, the probability density function (PDF) of an offspring point location \( y \in \mathbb{R}^2 \) relative to its parent point is:

\[
    f_y(y) = \frac{1}{2\pi\sigma^2} \exp \left( -\frac{|y|^2}{2\sigma^2} \right), \quad y \in \mathbb{R}^2.
\]

Denote by \( B^y \) the set of offspring points for the cluster centered at \( x \in \Phi_p \). The cluster process can be expressed as:

\[
    \Psi = \bigcup_{x \in \Phi_p} B^x.
\]
where the number of points per cluster \(|B^x|\) is Poisson distributed random variable with mean \(\bar{m}\). Let \(D^x = \{u : u = \|x + y\|, \forall y \in B^x\}\) be the sequence of distances from the reference point (assumed to be located at the origin) to the points located at the cluster centered at \(x \in \Phi_p\). The elements in the sequence \(D^x\) are correlated due to the common factor \(x\). This correlation can be handled by conditioning on \(x\) because the locations of offspring points around cluster center are i.i.d. by assumption. For any choice of the reference point, the conditional distribution of \(U\) is Rician with PDF\[10\], Lemma 1):

\[
f_U(u\|x\|) = \text{Ricepdf}(u, \nu = \|x\|; \sigma^2) = \frac{u}{\sigma^2} \exp\left(-\frac{u^2 + \nu^2}{2\sigma^2}\right) I_0\left(\frac{uv}{\sigma^2}\right),
\]

where \(u\) is a realization of \(U\) and \(I_0(\cdot)\) is the modified Bessel function with order zero.

### III. Distance Distributions

#### A. Contact distance distribution

In this subsection, we derive the statistical distribution of the contact distance, which is formally defined next.

**Definition 1** (Contact distance distribution). The contact distance distribution or the empty space distribution function is:

\[
F_{RC}(r) \triangleq \mathbb{P}(\|\Psi\| \leq r) = \mathbb{P}(\Psi(\mathbf{b}(o, r) > 0), \tag{4}
\]

where \(RC\) denotes the contact distance, \(\Psi(\mathbf{b}(o, r))\) denotes the number of points within a ball of radius \(r\) centered at \(o \equiv (0, 0)\). Due to the stationarity of TCP, any arbitrary point in \(\mathbb{R}^2\) can be treated as reference point or origin.

Note that the reference point, i.e., the origin, is not a part of the original point process \(\Psi (o \notin \Psi)\). The CDF of the contact distance is derived in the next Theorem.

**Theorem 1** (Contact distance distribution). The CDF of the contact distance is:

\[
F_{RC}(r) = 1 - \exp\left(-2\pi \lambda_p \int_{0}^{\infty} \left(1 - \exp\left(1 - Q_1(\frac{r}{\sigma}, \frac{\nu}{\sigma})\right)vdv\right)\right), \tag{5}
\]

where \(Q_1(\alpha, \beta)\) is the Marcum Q-function defined as \(Q_1(\alpha, \beta) = \int_{\beta}^{\infty} y e^{-\frac{x^2+y^2}{2}} I_0(\alpha y)dy\).

**Proof:** See Appendix \[A\].

After characterizing the CDF of the contact distance, we now derive a bound on this CDF in the next Corollary.

**Corollary 1.** The CDF \(F_{RC}(r)\) can be upper bounded as:

\[
F_{RC}(r) \leq 1 - \exp(-\pi \lambda_p \bar{m} r^2). \tag{6}
\]

**Proof:** Using Taylor expansion of exponential function, we get

\[
F_{RC}(r) \overset{(a)}{=} 1 - \exp\left(-2\pi \lambda_p \bar{m} \int_{0}^{r} \int_{0}^{\infty} f_U(u|v)vdvdu\right)
\]

\[
\overset{(b)}{=} 1 - \exp(-\pi \lambda_p \bar{m} r^2),
\]

where \((a)\) follows form \(1 - \exp(-\rho x) \leq \rho x, \rho > 0\) and \((b)\) follows from \(\int_{0}^{\infty} f_U(u|v)vdv = u\).

**Remark 1.** The bound presented in Corollary\[7\] can be interpreted as the CDF of the contact distance for a homogeneous PPP with the same density \(\bar{m}\lambda_P\) as that of the TCP. In fact, it is not uncommon to see this result being used as an “approximation” for the true contact distribution of a TCP, e.g., see \[7\]. However, to the best of our understanding, this expression has never been claimed to be a bound on the exact CDF. Nevertheless, as will be evident from the numerical results, the bound given by \(6\) is rather loose, which highlights the importance of the exact distribution derived in Theorem\[7\].

**Remark 2.** As per the interpretation discussed in the above remark, the upper bound presented in Corollary 1 is expected to get tighter when TCP itself starts converging towards a homogeneous PPP. This happens when we increase scattering variance \(\sigma^2\) or decrease the number of points per cluster \(\bar{m}\).

Therefore, this bound will be relatively more useful when \(\sigma^2\) is large and/or \(\bar{m}\) is small.

#### B. Nearest-neighbor distance distribution

The contact distance becomes the nearest-neighbor distance when a reference point is a part of the original point process. For any stationary point process, the nearest-neighbor distance is defined as:

\[
R_N = \|\arg \min_{z \in \Psi \setminus \o} (\|z\|)\|, \quad o \in \Psi, \tag{7}
\]

where the reference point is assumed to be located at the origin. The formal definition of the nearest-neighbor distance distribution is provided next.

**Definition 2** (Nearest-neighbor distance distribution). In the stationary point process, the distribution of nearest-neighbor distance \(R_N\) is defined as:

\[
F_{R_N}(r) = 1 - \mathbb{P}(\Psi(\mathbf{b}(o, r)) = 1|o \in \Psi), \tag{8}
\]

For derivation of the nearest-neighbor distance distribution, we focus on the following two different cases.

- **Case 1:** The reference point is chosen uniformly at random amongst all offspring points. In this case, the number of offspring points within the reference point’s own cluster \(N_0^{(1)}\) is number weighted Poisson distribution, with probability mass function (PMF)\[11\] Sec. 5.3:

\[
\mathbb{P}(N_0^{(1)} = \ell) = \frac{\ell \bar{m}^{\ell} e^{-\bar{m}}}{\ell!} \text{ for } \ell \in \mathbb{Z}^+, \tag{9}
\]

where \(\mathbb{Z}^+\) is set of positive integer. This is because the probabilities of different clusters being chosen are proportional to their size (i.e., number of offspring points per cluster). This is a spatial incarnation of the length-biased sampling that is the fundamental reason behind the waiting-bus paradox\[11\]. In other words, a point chosen uniformly at random amongst all offspring points is more likely to be from a larger cluster.

- **Case 2:** In this case, we choose the representative cluster (denoted by \(E_0^{(2)}\)) uniformly at random from the TCP, and
then, given $B_1^{(2)}$, the reference point is chosen uniformly at random amongst all offspring points within the representative cluster. It should be noted that the representative cluster that contains the reference point cannot be empty. Denote by $N_0^{(2)}$ the number of offspring points within the reference point’s own cluster. The distribution of $N_0^{(2)}$ is Poisson with mean $\bar{m}$ conditioned on $N_0^{(2)}$ being greater than one, with PMF

$$ \mathbb{P}(N_0^{(2)} = \ell) = \frac{\bar{m}^\ell e^{-\bar{m}}}{\ell!(1-e^{-\bar{m}})} \text{ for } \ell \in \mathbb{Z}^+. \quad (10) $$

It is to be noted that with some work, the likelihood function of poisson cluster process presented in [6] can be used for the derivation of the contact distance and nearest-neighbor distance distributions under Case 1. However, developing sufficient understanding of likelihood functions in order to use them for these derivations require substantial background in measure theory. Therefore, for the wireless network community to comprehend and appreciate these results, we provide an alternate method which circumvents the need to use measure theoretic notions. Moreover, the CDF of nearest-neighbor distance under Case 2 cannot be derived using likelihood function of [6], which made it necessary to develop an alternate approach. Next, we focus on the derivation of the nearest-neighbor distance distribution for the Case 1.

**Theorem 2** (Case 1: nearest-neighbor distance distribution).

The CDF of the nearest-neighbor distance distribution is:

$$ F_{R_1}^{(1)}(r) = 1 - (1 - F_{R_C}(r)) \int_0^\infty \exp \left( -\bar{m} \left( 1 - Q_1 \left( \frac{v_0}{\sigma}, \frac{r}{\sigma} \right) \right) \right) f_{V_0}(v_0) dv_0, \quad (11) $$

where $F_{R_C}(\cdot)$ is given by (5) and $f_{V_0}(v_0) = \frac{v_0}{\sigma^2} \exp \left( -\frac{v_0^2}{2\sigma^2} \right)$.

**Proof:** See Appendix B.

We further derive a closed-form bound on the CDF of the nearest-neighbor distance in the next Corollary.

**Corollary 2** (Case 1: nearest-neighbor distance distribution).

The CDF of nearest-neighbor distance is upper bounded by

$$ F_{R_1}^{(1)}(r) \leq 1 - \exp \left( -\pi \bar{m} \bar{r}^2 \right) \times \exp \left( -\bar{m} \left( 1 - \exp(-r^2/4\sigma^2) \right) \right). \quad (12) $$

**Proof:** See Appendix C.

As will be evident from our numerical comparisons, the upper bound presented in Corollary 2 tightly approximates the statistics of the nearest-neighbor distance for Case 1.

We derive the CDF of the nearest-neighbor distance for Case 2 in the next Theorem.

**Theorem 3** (Case 2: nearest-neighbor distance distribution).

The CDF of the nearest-neighbor distribution is:

$$ F_{R_2}^{(1)}(r) = 1 - (1 - F_{R_C}(r)) \int_0^\infty \exp \left( -\bar{m} Q_1 \left( \frac{v_0}{\sigma}, \frac{r}{\sigma} \right) \right) - \frac{e^{-\bar{m}}} {1-e^{-\bar{m}}} f_{V_0}(v_0) dv_0, \quad (13) $$

where $F_{R_C}(\cdot)$ is given by (5) and $f_{V_0}(v_0) = \frac{v_0}{\sigma^2} \exp \left( -\frac{v_0^2}{2\sigma^2} \right)$.

**Proof:** See Appendix D.

We now comment on the accuracy of the analysis and the tightness of the bounds. As shown in Fig. 1, Theorems 1-3 are perfectly matched with simulation, which corroborates the accuracy of our analysis. In addition, Fig. 1 shows that the upper bound given by Corollary 1 is somewhat loose while Corollary 2 provides a tight upper bound on Theorem 2.

**IV. Conclusion**

In this paper, we derived the CDF of contact and nearest-neighbor distance distributions for a TCP. For the latter, we considered two different approaches of sampling the reference point from the point process. This work has numerous extensions. From communication perspective, it will enable the characterization of the key performance metrics, e.g., coverage and rate, when the users and/or SBSs are modeled as TCP. From stochastic geometry perspective, it is important to extend the framework to characterize the distribution of the distance of the reference point from its $k^{th}$ closest neighbor.

**APPENDIX**

A. **Proof of Theorem 3**

Let us start with the derivation of the probability generating function (PGF) of the number of point within $b(o, r)$. For $\theta \in \mathbb{R}$, we define

$$ G_C(\theta) = \mathbb{E}[b(\theta)] = \sum_{x+y \in B^\times} \frac{1}{\pi \bar{m}} \left( 1 - \theta^1(\|x+y\| < r) \right) = \mathbb{E} \left[ \prod_{x \in \Phi} \prod_{y \in B^\times} \theta^1(\|x+y\| < r) \right] $$

$$ = \mathbb{E} \left[ \prod_{x \in \Phi} \int_{\mathbb{R}^2} \exp(-\bar{m} \int_{\mathbb{R}^2} (1 - \theta^1(\|x+y\| < r)) f_Y(y) dy \right] $$

$$ \xrightarrow{(a)} \exp \left( -\lambda_p \int_{\mathbb{R}^2} \left( 1 - \exp \left( -\bar{m} \int_{\mathbb{R}^2} (1 - \theta^1(\|z\| < r)) f_Y(z) dz \right) \right) \right) \times f_Y(z-x) dz dx \right) $$

$$ \xrightarrow{(b)} \exp \left( -\lambda_p \int_{\mathbb{R}^2} \left( 1 - \exp \left( -\bar{m} \int_{\mathbb{R}^2} (1 - \theta^1(\|z\| < r)) f_Y(z) dz \right) \right) \right) \times f_Y(z-x) dz dx \right) $$
\[ \begin{align*}
&\exp \left( -2\pi\lambda_p \int_0^\infty \left( 1 - \exp \left( -\bar{m} \int_0^r (1 - \theta) \right) \right) f_U(u|v)du \right) vdv) \right), \\
& \times \int_0^r f_U(u|v_0) f_{V_0}(v_0)dv_0, \\
& \times \int_0^r f_U(u|v_0) f_{V_0}(v_0)dv_0, \\
& \leq 1 - \exp(-\pi\lambda_p \bar{m} r^2) \exp \left( -\bar{m} \right) \int_0^r f_U(u|v_0) f_{V_0}(v_0)dv_0. \\
& \leq 1 - \exp(-\pi\lambda_p \bar{m} r^2) \exp \left( -\bar{m} \right) \int_0^r f_U(u|v_0) f_{V_0}(v_0)dv_0. \\
\end{align*} \]

where (a) follows from the PGF of Poisson random variable along with the fact that the points in \( B^x \) are i.i.d. with PDF \( f_Y(y) \) given by (1). (b) follows from the probability generating functional (PGFL) of PPP and change of variable \( z = x + y, \) and (c) follows from converting Cartesian to polar coordinates, where \( f_U(u|v) \) is given by (3). Now, the CDF of contact distance can be derived as follows:

\[ F_{Rc}(r) = 1 - P(N = 0) = 1 - \exp \left( -2\pi\lambda_p \int_0^\infty \left( 1 - \exp \left( -\bar{m} \int_0^r (1 - \theta) \right) \right) f_U(u|v)du \right) vdv, \]

where the final result is obtained by using the definition of Marcum Q-function.

**B. Proof of Theorem 2**

Denote by \( B_0^{(1)} \) the set of offspring points within the reference point’s own cluster for Case 1. The PGF of the number of points within \( b(o, r) \) is \( G_N^{(1)}(\theta) \)

\[ \begin{align*}
&= \mathbb{E} \left[ \sum_{x \in \Psi} \sum_{y \in \Psi^x} 1 \{ |x+y| < r \} + \sum_{y \in \Theta} 1 \{ |x_0+y| < r \} \right] \\
&= \mathbb{E} \left[ \prod_{x \in \Psi} \prod_{y \in \Psi^x} \theta^1 \{ |x+y| < r \} \right] \\
&= G_N(\theta) \int_0^\infty \sum_{\ell=1}^\infty \left( \int_0^r f_U(u|v_0) du + \theta \int_0^r f_U(u|v_0) du \right) \}

where (a) follows from the same lines as the proof of Theorem 2 using the PMF of number of points within reference point’s own cluster given by (10). Now, the CDF of nearest-neighbor distance for Case 2 is \( F_{Rc}^{(2)}(r) = 1 - G_N^{(2)}(0) \).

**REFERENCES**

[1] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*, 3rd ed. New York: John Wiley and Sons, 2013.
[2] C. Saha, M. Afshang, and H. S. Dhillon, “Enriched K-tier HetNet model to enable the analysis of user-centric small cell deployments,” submitted to IEEE Trans. on Wireless Commun., 2016, available online:arxiv.org/abs/1606.06223.
[3] 3GPP. “Consideration of UE Cluster Position and Penetration in Heterogeneous Deployment Configuration 4,” Discussion/ Decision R1-100477, Jan. 2010. 8.2.1 Relevant scenarios of Heterogeneous Networks.
[4] J. G. Andrews, R. K. Ganti, M. Haenggi, N. Jindal, and S. Weber, “A primer on spatial modeling and analysis in wireless networks,” IEEE Commun. Magazine, vol. 48, no. 11, pp. 156–163, Nov. 2010.
[5] M. Haenggi, *Stochastic Geometry for Wireless Networks*. Cambridge University Press, 2012.
[6] M. Baudin, “Likelihood and nearest-neighbor distance properties of multidimensional poison cluster processes,” *Journal of Applied Probability*, pp. 879–888, 1991.
[7] Y. J. Chun, M. O. Hasna, and A. Ghrayeb, “Modeling heterogeneous cellular networks interference using poison cluster processes,” IEEE Journal on Sel. Areas in Commun., vol. 33, no. 10, pp. 2182–2195, Oct. 2015.
[8] Y. Wang and Q. Zhu, “Modeling and analysis of small cells based on clustered stochastic geometry,” IEEE Commun. Letters, 2016, to appear.
[9] K. G. Ganti, M. Haenggi, “Interference and outage in clustered wireless ad hoc networks,” IEEE Trans. on Info. Theory, vol. 55, no. 9, pp. 4067–4086, Sep. 2009.
[10] M. Afshang, H. S. Dhillon, and P. H. J. Chong, “Modeling and performance analysis of clustered device-to-device networks,” IEEE Trans. on Wireless Commun., vol. 15, no. 7, pp. 4957–4972, Jul. 2016.
[11] W. Feller, *An introduction to probability theory and its applications*. John Wiley & Sons London-New York-Sydney-Toronto, 1968, vol. 3.