

## PROBABILISTIC RESULTS FOR A MOBILE SERVICE SCENARIO

JESPER MØLLER,\* *Aalborg University*

MAN LUNG YIU,\*\* *Hong Kong Polytechnic University*

### Abstract

We consider the following stochastic model for a mobile service scenario. Consider a stationary Poisson process in  $\mathbb{R}^d$ , with its points radially ordered with respect to the origin (the anchor); if  $d = 2$ , the points may correspond to locations of e.g. restaurants. A user, with a location different from the origin, asks for the location of the first Poisson point and keeps asking for the location of the next Poisson point until the first time that he can be completely certain that he knows which Poisson point is his nearest neighbour. This waiting time is the communication cost, while the inferred privacy region is a random set obtained by an adversary who only knows the anchor and the points received from the server, where the adversary ‘does the best’ to infer the possible locations of the user. Probabilistic results related to the communication cost and the inferred privacy region are established for any dimension  $d \geq 1$ . Furthermore, special results when  $d = 1$  and particularly when  $d = 2$  are derived.

*Keywords:* communication cost; nearest-neighbour search; Poisson process; privacy region; radial simulation algorithm; Voronoi tessellation

2010 Mathematics Subject Classification: Primary 60D05;60G55;62M30

Secondary 68U20

---

\* Postal address: Department of Mathematical Sciences, Aalborg University Fredrik Bajers Vej 7G, DK-9220 Aalborg, Denmark. Email address: jm@math.auc.dk

\*\* Postal address: Department of Computing Hong Kong Polytechnic University Hung Hom, Kowloon Hong Kong. Email address: csmlyiu@comp.polyu.edu.hk

## 1. Introduction

The mobile Internet offers services that e.g. receive the location of the nearest point of interest such as a store, restaurant, or tourist attraction, see [9] and the references therein. This paper demonstrates that tools from applied probability and in particular stochastic geometry can be useful when analysing the performance of such services.

The paper considers a setting for a mobile service protocol proposed by [9], where a user is located at a point  $q \in \mathbb{R}^d$  and a stationary Poisson point process  $\Phi = \{X_1, X_2, \dots\} \subset \mathbb{R}^d$  is given; for the problem setting in [9],  $d = 2$  and the points in  $\Phi$  may e.g. correspond to the locations of stores. In order to preserve some privacy, the user queries a server for nearby points in  $\Phi$  but he reports not his correct location  $q$  but another location  $q' \in \mathbb{R}^d$  referred to as the anchor. An incremental query processing on the server is used so that the points  $X_1, X_2, \dots$  are ordered in increasing distance to the anchor. The user then stops to query the server as soon as possible, i.e., when the nearest point in  $\Phi$  with respect to  $q$  can be determined. The waiting time for this to happen is called the communication cost and is denoted  $M$ . Another object of interest is the inferred privacy region, which is a random set  $\mathcal{R} \subset \mathbb{R}^d$  obtained by an adversary who only knows the location of the anchor and the points received from the server, where the adversary ‘does the best’ to infer the possible locations of the user. The precise definitions of  $M$  and  $\mathcal{R}$  are given in Sections 2 and 3, respectively.

The assumption that  $\Phi$  is a stationary Poisson process is motivated by that this is the most fundamental spatial point process model in stochastic geometry, and it often serves as a natural starting point for statistical analysis, see, e.g., [1, 4, 8]. Our objective is to analyse the distribution of  $M$  and various properties of  $\mathcal{R}$ , where we exploit the independence properties of the Poisson process to derive analytical results; for other point process models Monte Carlo simulations will probably be needed. In Section 2, the distribution and moments for  $M$  are derived in detail. Section 3 describes first the geometric properties of  $\mathcal{R}$ , and second establishes results related to the probability that  $\mathcal{R}$  contains a given point in  $\mathbb{R}^d$  and the expected value of  $V$ , where  $V = |\mathcal{R}|$  is the  $d$ -dimensional volume of the inferred privacy region.

## 2. The communication cost

### 2.1. Preliminaries

2.1.1. *Assumptions:* Though in the sequel it may be assumed without loss of generality that  $q' = o$ , the origin in  $\mathbb{R}^d$ , we will keep using the notation  $q'$  in order to remind the reader that this refers to the anchor. Denote  $l = \|q - q'\|$  the distance between the anchor and the user location, and  $R_i = \|X_i - q'\|$  the distance of  $X_i$  to the anchor. The case where  $l = 0$  turns out to be trivial since  $\Phi$  is a stationary Poisson process, so we assume that  $l > 0$ . Any point  $X_i$  in  $\Phi$  is a random variable, and we order the points in  $\Phi$  such that  $0 \leq R_1 \leq R_2 \leq \dots$ . Note that these inequalities are strict almost surely. Denote  $Z$  the nearest neighbour to  $q$  among  $X_1, X_2, \dots$ . For  $i = 1, 2, \dots$ , let  $Q_i$  be the nearest neighbour to  $q$  among the  $i$  first points  $X_1, \dots, X_i$ , and set  $D_i = \|q - Q_i\|$  (so  $Z$  and  $Q_i$  are almost surely uniquely defined). Denote  $B(x, r) = \{y \in \mathbb{R}^d : \|y - x\| \leq r\}$  the closed ball in  $\mathbb{R}^d$  with centre  $x \in \mathbb{R}^d$  and radius  $r \geq 0$ . Let  $|\cdot|$  denote volume (Lebesgue measure) in  $\mathbb{R}^d$ , and

$$\omega_d = |B(0, 1)| = \pi^{d/2} / \Gamma(1 + d/2)$$

the volume of the  $d$ -dimensional unit ball, where  $\Gamma$  is the gamma function. Finally, denote  $\rho > 0$  the intensity of  $\Phi$  and define

$$\alpha = (\rho |B(0, l)|)^{1/d} = (\omega_d \rho)^{1/d} l.$$

2.1.2. *Definition of the communication cost:* Denote  $\mathbb{N} = \{1, 2, \dots\}$  the set of positive integers. For  $i \in \mathbb{N}$ , define the *demand space* by  $\mathcal{D}_i = B(q, D_i)$  and the *supply space* by  $\mathcal{S}_i = B(q', R_i)$ . Then  $(\mathcal{D}_i)_{i \in \mathbb{N}}$  is decreasing,  $(\mathcal{S}_i)_{i \in \mathbb{N}}$  is increasing, and we define the *communication cost* as the discrete random variable  $M$  given by the first time the demand space is included in the supply space, that is,

$$M = \inf\{i \in \mathbb{N} : \mathcal{D}_i \subseteq \mathcal{S}_i\} \quad (1)$$

(setting  $\inf \emptyset = \infty$ ). See Figure 1. In other words, for any  $m \in \mathbb{N}$ ,  $M = m$  if  $m$  is the first time the user can be completely ensured that  $Z = Q_m$  when he has only received  $X_1, \dots, X_m$  from the server, i.e., no matter where the points  $X_{m+1}, X_{m+2}, \dots$  potentially could be located in  $\mathbb{R}^d \setminus B(q', R_m)$ . Note that  $M$  is almost surely finite

(this is verified in Lemma 1 below), in which case

$$Z = Q_M \tag{2}$$

is returned as the nearest neighbour to  $q$ .

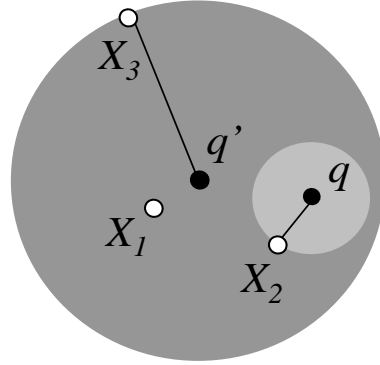


FIGURE 1: A planar example ( $d = 2$ ) with  $M = 3$  and showing the corresponding demand space  $\mathcal{D}_3$  and supply space  $\mathcal{S}_3$ . Note that  $\mathcal{D}_2 = \mathcal{D}_3$  and  $Z = Q_2 = Q_3$ .

## 2.2. Results

For  $d \geq 2$ , we can exclude certain events of zero probability and thereby simplify the meaning of  $M$  and  $Z$  as stated in the following lemma.

**Lemma 1.** *For  $d \geq 2$ , with probability one,  $M$  is finite and given by the smallest integer  $m \geq 2$  such that  $D_{m-1}$  is a strict subset of  $S_m$ , and*

$$Z = Q_M = Q_{M-1}, \quad D_M = D_{M-1}, \quad R_M \geq l. \tag{3}$$

*Proof.* Clearly, by (1),  $R_M \geq l$ . Since  $M = 1$  implies that  $X_1$  lies on the halfline  $H$  with endpoint  $q$  and direction  $q - q'$ , and this happens with probability zero, we have almost surely that  $M \geq 2$ . Further, with probability one, for  $m \in \{2, 3, \dots\}$ ,  $M = m$  implies that  $Q_m \neq X_m$  because  $Q_m = X_m$  would imply that  $X_m \in H$  which happens with probability zero. Moreover, with probability one, the sequence  $R_1, R_2, \dots$  is strictly increasing to infinity, and so the sequence of supply spaces  $\mathcal{S}_1 \subset \mathcal{S}_2 \subset \dots$  tends to  $\mathbb{R}^d$ . On the other hand, the sequence of demand spaces decreases. Combining these facts with (1) and (2), the assertions of the lemma are seen to be true.

Let  $T = \|Z - q\|$  be the distance from the user to its nearest point in  $\Phi$ , and let  $\Psi = \Phi \cap [B(q', T+l) \setminus B(q, T)]$  be the restriction of  $\Phi$  to the random set  $B(q', T+l) \setminus B(q, T)$ . Let  $N$  denote the number of points in  $\Psi$ , and set  $S = \omega_d \rho T^d$  and

$$\Lambda = \rho |B(q', T+l) \setminus B(q, T)| = (\alpha + S^{1/d})^d - S = \sum_{i=0}^{d-1} \binom{d}{i} \alpha^{d-i} S^{i/d}. \quad (4)$$

Since  $\Phi$  is a Poisson process, we obtain that

- (i)  $S$  is exponentially distributed with parameter one;
- (ii) conditional on  $Z$ ,  $\Phi \setminus B(q, T)$  is a homogeneous Poisson process on  $\mathbb{R}^d \setminus B(q, T)$  with intensity  $\rho$ , and  $\Psi$  is a homogeneous Poisson process on  $B(q', T+l) \setminus B(q, T)$  with its mean number of points equal to  $\Lambda$ ;
- (iii) in the special case  $d = 1$ , the event  $Z = X_M$  is equivalent to that  $Z - q$  has the same sign as  $q - q'$ , so  $P(Z = X_M) = 1/2$ , and if  $Z = X_M$  then  $N = M - 1$  and  $\Psi = \{X_1, \dots, X_{M-1}\}$ , while if  $Z \neq X_M$  then  $N = M - 2$  and  $\Psi = \{X_1, \dots, X_{M-1}\} \setminus \{Z\}$ ;
- (iv) for  $d \geq 2$ , with probability one,  $N = M - 2$  and  $\Psi = \{X_1, \dots, X_{M-1}\} \setminus \{Z\}$ , cf. Lemma 1.

These results are now used to obtain the distribution of  $N$  (or equivalently  $M$ ), where  $\text{po}(\alpha)$  denotes the Poisson distribution with parameter  $\alpha$ .

**Theorem 1.** (a) For  $d = 1$ ,  $N$  is independent of  $Z$  and follows  $\text{po}(\alpha)$ ,

$$P(M = 1) = \frac{e^{-\alpha}}{2}, \quad P(M = m) = \left( \frac{\alpha^{m-1}}{(m-1)!} + \frac{\alpha^{m-2}}{(m-2)!} \right) \frac{e^{-\alpha}}{2}, \quad m = 2, 3, \dots, \quad (5)$$

and  $M$  has mean and variance

$$E(M) = \alpha + 3/2, \quad V(M) = \alpha + 1/4, \quad (6)$$

with  $\alpha = 2\rho l$ .

(b) For  $d = 2$ ,

$$P(M = m) = \int_0^\infty \frac{(\alpha^2 + 2\alpha\sqrt{s})^{m-2}}{(m-2)!} e^{-(\alpha^2 + 2\alpha\sqrt{s} + s)} ds, \quad m = 2, 3, \dots, \quad (7)$$

and

$$E(M) = \alpha^2 + \sqrt{\pi}\alpha + 2, \quad V(M) = (5 - \pi)\alpha^2 + \sqrt{\pi}\alpha, \quad (8)$$

with  $\alpha = \sqrt{\pi\rho}l$ .

(c) For  $d \geq 2$ ,  $M - 2$  conditional on  $Z$  follows  $\text{po}(\Lambda)$ , and

$$\mathbf{E}(M) = 2 + \sum_{i=0}^{d-1} \binom{d}{i} \alpha^{d-i} \Gamma\left(1 + \frac{i}{d}\right), \quad (9)$$

$$\begin{aligned} \mathbf{V}(M) &= \sum_{i=0}^{d-1} \binom{d}{i} \alpha^{d-i} \Gamma\left(1 + \frac{i}{d}\right) \\ &\quad + \sum_{i=1}^{d-1} \sum_{j=1}^{d-1} \binom{d}{i} \binom{d}{j} \alpha^{2d-i-j} \left[ \Gamma\left(1 + \frac{i+j}{d}\right) - \Gamma\left(1 + \frac{i}{d}\right) \Gamma\left(1 + \frac{j}{d}\right) \right]. \end{aligned} \quad (10)$$

(d) For  $d \geq 1$  and any number  $\beta$ ,  $\mathbf{E}(M^\beta) < \infty$ .

*Proof.* If  $d = 1$ , since  $\Lambda = \alpha$  is then deterministic, (ii) implies that  $N$  is independent of  $Z$  and follows  $\text{po}(\alpha)$ . Hence (iii) easily implies (5) and (6), and so (a) follows.

If  $d \geq 2$ , then by (ii),  $N$  conditional on  $Z$  follows  $\text{po}(\Lambda)$ , and by (iv),  $M = N + 2$ , so  $\mathbf{E}(M) = 2 + \mathbf{E}(\Lambda)$  and

$$\mathbf{V}(M) = \mathbf{V}(N) = \mathbf{E}(\mathbf{V}(N|\Lambda)) + \mathbf{V}(\mathbf{E}(N|\Lambda)) = \mathbf{E}(\Lambda) + \mathbf{V}(\Lambda).$$

Combining this with (4) and that by (i),  $\mathbf{E}(S^\beta) = \Gamma(\beta + 1)$  for  $\beta > -1$ , we obtain after a straightforward calculation that (9) and (10) hold for  $d \geq 2$ , where (8) is the case with  $d = 2$ . Then (7) immediately follows by combining (4), (i), and the fact that  $N$  conditional on  $S$  follows  $\text{po}(\Lambda)$ . Thereby (b)-(c) are verified.

For  $\beta \leq 0$ ,  $M^\beta \leq 1$ , and so (d) clearly holds. For  $\beta > 0$ , (d) follows immediately from (a) if  $d = 1$ , and from (c) and (4) if  $d \geq 2$ , using again that  $\mathbf{E}(S^\beta) = \Gamma(\beta + 1)$ . Hence (d) is verified.

Suppose  $d = 2$  and let  $\text{erf}(\alpha) = (2/\sqrt{\pi}) \int_0^\alpha \exp(-t^2) dt$  be the ‘error function’. Then (7) gives

$$\mathbf{P}(M = 2) = \exp(-\alpha^2) + \alpha\sqrt{\pi}(\text{erf}(\alpha) - 1),$$

which strictly decreases from one to zero as  $\alpha$  decreases from zero to infinity. We have also evaluated the integral in (7) for  $m = 3, 4, \dots$  using the computational software program Maple, but since the number of terms increases fast as  $m$  increases we omit the results here. By (4) and (c),  $M - 2$  conditional on  $S$  follows  $\text{po}(\alpha^2 + 2\alpha\sqrt{S})$ . Hence, as  $\alpha \rightarrow \infty$ ,  $(M/\alpha) - \alpha$  converges in distribution to a mixture of normal distributions with mean  $2\sqrt{S}$  and unit variance.

### 3. The inferred privacy region

#### 3.1. Preliminaries

3.1.1. *Definition of the inferred privacy region* Suppose that an adversary knows the location  $q'$  of the anchor, the termination conditions (1)-(2), the termination time  $M$ , and the points  $X_1, \dots, X_M$  received from the server, while the location  $q$  of the user is unknown to him. If the adversary then wants to infer the possible locations of  $q$ , the best the adversary can do is to estimate  $q$  to be contained in the inferred privacy region which is a random set  $\mathcal{R}$  specified below.

Consider the Voronoi tessellation of  $\mathbb{R}^d$  generated by  $\{X_1, \dots, X_M\}$ , with cells

$$C_i = \{x \in \mathbb{R}^d : \|x - X_i\| \leq \|x - X_j\|, j = 1, \dots, M\}, \quad i = 1, \dots, M.$$

The Voronoi cells have disjoint interiors with boundaries of zero volume (with respect to Lebesgue measure in  $\mathbb{R}^d$ ), and with probability one they are  $d$ -dimensional sets [3, 5]. Note that  $B(x, \|x - X_i\|) \subseteq B(q', r)$  if and only if  $\|x - q'\| + \|x - X_i\| \leq r$ . Consequently, if  $M \geq 2$  and  $i \in \{1, \dots, M-1\}$ , the set

$$E_i = \{x \in C_i : R_{M-1} < \|x - q'\| + \|x - X_i\| \leq R_M\} \quad (11)$$

consists ‘essentially’ of all possible locations  $x$  of the user such that  $X_i$  is returned under the termination conditions as the nearest neighbour to  $x$ . By ‘essentially’ we mean that if  $x$  is on the boundary of  $C_i$  so that  $x \in C_j$  for some  $j < i$ , then  $X_i$  would not had been returned, but the set of such points  $x$  has zero volume and as argued in comment (E) below, it plays no important rule but is just convenient that we have included such points in  $E_i$ . Moreover, the set of all possible locations  $x \in C_M \setminus \cup_{i=1}^{M-1} C_i$  of the user is given by

$$E_M = [q', X_M] \setminus \bigcup_{i=1}^{M-1} C_i, \quad (12)$$

where  $[q', X_M]$  is the closed line segment with end points  $q'$  and  $X_M$ , and we set  $\cup_{i=1}^{M-1} C_i = \emptyset$  if  $M = 1$ . The *inferred privacy region* is therefore given by

$$\mathcal{R} = \bigcup_{i=1}^M E_i. \quad (13)$$

Figure 2 shows an example of  $\mathcal{R}$  when  $d = 2$  and the Poisson points are generated by the conventional radial simulation algorithm due to Quine and Watson [6].

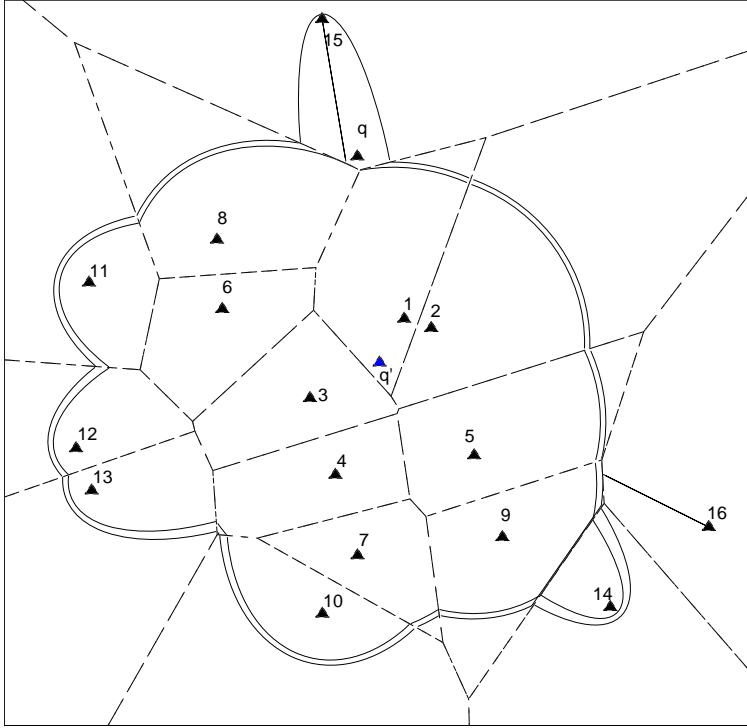


FIGURE 2: A planar inferred privacy region  $\mathcal{R}$  (the solid lines except those on the boundary of the square) defined by  $M = 16$  points (the triangles marked  $1, \dots, 16$ ) and the anchor  $q'$ . The Voronoi tessellation with nuclei  $X_1, \dots, X_M$  is shown as dotted lines. Also the user location  $q$  is shown, but the inferred privacy region is unchanged if any other point in  $\mathcal{R}$  had been the location of the user.

3.1.2. *Comments:* Some remarks are in order.

- (A) If  $M = 1$  then simply  $\mathcal{R} = E_M = [q', X_M]$ .
- (B) If  $1 \leq i \leq M - 1$  then by (11),  $E_i = C_i \cap (F_i \setminus G_i)$ , where  $F_i$  and  $G_i$  are the ellipsoidal regions with foci  $q'$  and  $X_i$ , such that any point on the boundary has its sum of distances to the foci equal to  $R_M$  and  $R_{M-1}$ , respectively (see cells  $1, \dots, 15$  in Figure 2). As illustrated in Figure 2,  $E_i$  can be a connected set (see e.g. cell 1) or a disconnected set (see cell 9) or the empty set (see e.g. cell 3),  $G_{M-1} = [q', X_{M-1}]$  is a closed line segment (see cell 15), while for  $1 \leq i \leq M - 2$ ,  $F_i$  and  $G_i$  are almost surely of dimension  $d$ .
- (C) If  $M \geq 2$  then  $E_M$  is the line segment given by the intersection of  $[q', X_M]$  and



the interior of  $C_M$ , cf. (12) (see cell 16 in Figure 2).

- (D) If  $d \geq 2$ ,  $E_M$  has zero  $d$ -dimensional volume, and the adversary could exclude both the possibility that  $M = 1$  and the possibility that  $X_M$  is the nearest point in  $\Phi$  to the user, since the event that  $M = 1$  or  $x \in [q, X_M]$  has probability zero.
- (E) Recalling the considerations in connection to (11) concerning the boundary points of the Voronoi cells, note that  $\mathcal{R} = \cup_{i=1}^M E'_i$ , where  $E'_i = E_i \setminus \cup_{j=1}^{i-1} C_j$  is exactly the set of all possible locations  $x$  of the user such that  $X_i$  is returned under the termination conditions as the nearest neighbour to  $x$  (setting  $\cup_{j=1}^{i-1} C_j = \emptyset$  if  $i = 1$ ). Clearly,  $E_i \setminus E'_i$  has zero volume.

### 3.2. Representation formula

The volume of the inferred privacy region,  $V = |\mathcal{R}|$ , is a ‘measure of privacy’ with mean value

$$\mathbb{E}(V) = \int p(x) \, dx, \quad (14)$$

where for any location  $x \in \mathbb{R}^d$ ,

$$p(x) = \mathbb{P}(x \in \mathcal{R})$$

is the probability that  $x$  is in the inferred privacy region. Writing  $Z = q + TU$ , then  $U$  is a uniformly distributed unit vector in  $\mathbb{R}^d$ ,  $T$  and  $U$  are independent, and  $T^d$  is exponentially distributed with rate  $\rho\omega_d$ , so  $Z$  has density function

$$f(z) = \rho \exp(-\rho\omega_d \|z - q\|^d). \quad (15)$$

Our strategy is first to determine the conditional probability

$$p(x|z) = \mathbb{P}(x \in \mathcal{R} | Z = z),$$

second to calculate

$$p(x) = \int p(x|z) f(z) \, dz, \quad (16)$$

and finally to obtain  $\mathbb{E}(V)$  from (14).

In the remainder of this paper, we focus on the case  $d \geq 2$ . Theorem 2 below establishes an expression of  $p(x|z)$ ; similar techniques apply in the special case  $d = 1$ , but the details are then somewhat more complicated since we have to account for each of the cases where  $Z = X_M$  and  $Z \neq X_M$ , cf. (iii) (above Theorem 1).

Let  $\mathbf{1}[\cdot]$  denote the indicator function. For  $x, y \in \mathbb{R}^d$  and  $r, s, t \geq 0$ , define

$$c(r, s, t) = |B(x, s) \cap B(y, t)| \quad \text{if } r = \|x - y\|$$

and

$$\begin{aligned} p(x, y, t) = & \mathbf{1}[\|x - q'\| + \|y - x\| \geq t + l] \exp\left(-\rho\{\omega_d[(\|x - q'\| + \|y - x\|)^d - (t + l)^d] \right. \\ & \left. - c(\|x - q'\|, t + l, \|y - x\|) + c(\|x - q\|, t, \|y - x\|)\right) \\ & + \mathbf{1}[\|x - q'\| + \|y - x\| < t + l] \exp\left(-\rho\{\omega_d[(t + l)^d - t^d - (\|x - q'\| + \|y - x\|)^d \right. \\ & \left. + \|y - x\|^d] + c(l, t, \|x - q'\| + \|y - x\|) - c(\|x - q\|, t, \|y - x\|)\right). \end{aligned} \quad (17)$$

Note that  $c(r, s, t) = 0$  if  $r \geq s + t$  or  $s = 0$  or  $t = 0$ , and  $c(r, s, t) = \omega_d \min\{s^d, t^d\}$  if  $r = 0$ . If  $d = 2$ ,  $r < s + t$ , and  $r, s, t > 0$ , then

$$\begin{aligned} c(r, s, t) = & s^2 \arccos\left(\frac{r^2 + s^2 - t^2}{2rs}\right) - \frac{r^2 + s^2 - t^2}{4r^2} \left[4r^2 s^2 - (r^2 + s^2 - t^2)^2\right]^{1/2} \\ & + t^2 \arccos\left(\frac{r^2 + t^2 - s^2}{2rt}\right) - \frac{r^2 + t^2 - s^2}{4r^2} \left[4r^2 t^2 - (r^2 + t^2 - s^2)^2\right]^{1/2}. \end{aligned}$$

**Theorem 2.** For  $d \geq 2$  and  $x, z \in \mathbb{R}^d$ , letting  $t = \|z - q\|$ , then

$$\begin{aligned} p(x|z) = & p(x, z, t) + \\ & \rho \int \mathbf{1}[y \in B(q', t + l) \setminus B(q, t), z \in B(q', \|x - q'\| + \|y - x\|) \setminus B(x, \|y - x\|)] \\ & p(x, y, t) \, dy. \end{aligned} \quad (18)$$

*Proof.* We start by verifying that

$$p(x, y, t) = p_1(x, y, t)p_2(x, y, t), \quad (19)$$

where

$$p_1(x, y, t) = \exp\left(-\rho\omega_d \max\{0, (\|x - q'\| + \|y - x\|)^d - (t + l)^d\}\right)$$

and

$$\begin{aligned} p_2(x, y, t) = & \\ & \exp\left(-\rho|\{B(q', t + l) \setminus B(q, t)\} \setminus \{B(q', \|x - q'\| + \|y - x\|) \setminus B(x, \|y - x\|)\}|\right). \end{aligned}$$

Note that

$$\begin{aligned} p_1(x, y, t)p_2(x, y, t) &= \mathbf{1}[\|x - q'\| + \|y - x\| \geq t + l]p_1(x, y, t)p_2(x, y, t) \\ &\quad + \mathbf{1}[\|x - q'\| + \|y - x\| < t + l]p_2(x, y, t). \end{aligned} \quad (20)$$

If  $\|x - q'\| + \|y - x\| \geq t + l$ , then  $B(q, t) \subseteq B(q', t + l) \subseteq B(q', \|x - q'\| + \|y - x\|)$ , and so

$$p_2(x, y, t) = \exp(-\rho[c(\|x - q'\|, t + l, \|y - x\|) - c(\|x - q\|, t, \|y - x\|)]). \quad (21)$$

If  $\|x - q'\| + \|y - x\| < t + l$ , then  $B(x, \|y - x\|) \subseteq B(q', \|x - q'\| + \|y - x\|) \subseteq B(q', t + l)$  and  $B(q, t) \subseteq B(q', t + l)$ , and so if  $A^c = \mathbb{R}^d \setminus A$  denotes the complement of a set  $A \subseteq \mathbb{R}^d$ ,

$$\begin{aligned} &|\{B(q', t + l) \setminus B(q, t)\} \setminus \{B(q', \|x - q'\| + \|y - x\|) \setminus B(x, \|y - x\|)\}| \\ &= |B(q', t + l) \cap B(q, t)^c \cap [B(q', \|x - q'\| + \|y - x\|)^c \cup B(x, \|y - x\|)]| \\ &= |B(q', t + l) \cap B(q, t)^c \cap B(q', \|x - q'\| + \|y - x\|)^c| \\ &\quad + |B(q', t + l) \cap B(q, t)^c \cap B(x, \|y - x\|)| \\ &= [|B(q', t + l)| - |B(q, t)| - |B(q', \|x - q'\| + \|y - x\|)| \\ &\quad + |B(q, t) \cap B(q', \|x - q'\| + \|y - x\|)|] + [|B(x, \|y - x\|)| - |B(q, t) \cap B(x, \|y - x\|)|] \end{aligned} \quad (22)$$

Now, (19) follows from (17) and (20)-(22).

Denote  $Z(x)$  the almost surely unique nearest point to  $x$  in  $\{X_1, \dots, X_M\} = \Psi \cup \{Z, X_M\}$  (so  $Z(q) = Z$ ). By (11)-(13) and (D) in Section 3.1.2,

$$\begin{aligned} P(x \in \mathcal{R}|Z) &= P(Z = Z(x), R_{M-1} < \|x - q'\| + \|Z - x\| \leq R_M|Z) \\ &\quad + E\left(\sum_{i=1}^N \mathbf{1}[X_i = Z(x), R_{M-1} < \|x - q'\| + \|X_i - x\| \leq R_M] \middle| Z\right). \end{aligned} \quad (23)$$

By (iv) (above Theorem 1), with probability one,

$$Z = Z(x) \Leftrightarrow [\Psi \cup \{X_M\}] \cap B(x, \|Z - x\|) = \emptyset.$$

Conditional on  $Z$ ,  $R_M^d - (T + l)^d$  is exponentially distributed and independent of  $\Psi$ , cf. (ii) (above Theorem 1). Therefore, ignoring the null set where  $\|Z - q'\| =$

$\|x - q'\| + \|Z - x\|$  (i.e.,  $\|Z - q'\| < \|x - q'\| + \|Z - x\|$  almost surely),

$$\begin{aligned}
& \mathbb{P}(Z = Z(x), R_{M-1} < \|x - q'\| + \|Z - x\| \leq R_M | Z) \\
& = \mathbb{P}(\|x - q'\| + \|Z - x\| \leq R_M, X_M \notin B(x, \|Z - x\|) | Z) \\
& \quad \mathbb{P}(\Psi \subset B(q', \|x - q'\| + \|Z - x\|) \setminus B(x, \|Z - x\|) | Z) \\
& = p_1(x, Z, T) p_2(x, Z, T),
\end{aligned} \tag{24}$$

using in the last equality that  $\|x - q'\| + \|Z - x\| \leq R_M$  implies that  $X_M \notin B(x, \|Z - x\|)$ , since  $B(x, \|Z - x\|) \subseteq B(q', \|x - q'\| + \|Z - x\|)$ . Moreover,

$$\begin{aligned}
& \mathbb{E} \left( \sum_{i=1}^N \mathbf{1} [X_i = Z(x), R_{M-1} < \|x - q'\| + \|X_i - x\| \leq R_M] \middle| Z \right) \\
& = \mathbb{E} \left( \sum_{i=1}^N \mathbf{1} [(\Psi \setminus \{X_i\}) \cup \{Z, X_M\}] \cap B(x, \|X_i - x\|) = \emptyset, \right. \\
& \quad \left. (\Psi \setminus \{X_i\}) \cup \{Z\} \subset B(q', \|x - q'\| + \|X_i - x\|), \|x - q'\| + \|X_i - x\| \leq R_M \middle| Z \right) \\
& = \rho \int \mathbf{1} [y \in B(q', T + l) \setminus B(q, T), Z \in B(q', \|x - q'\| + \|y - x\|) \setminus B(x, \|y - x\|)] \\
& \quad \mathbb{P}(\|x - q'\| + \|y - x\| \leq R_M, X_M \notin B(x, \|y - x\|) | Z) \\
& \quad \mathbb{P}(\Psi \subset B(q', \|x - q'\| + \|y - x\|) \setminus B(x, \|y - x\|) | Z) \, dy \\
& = \rho \int \mathbf{1} [y \in B(q', T + l) \setminus B(q, T), Z \in B(q', \|x - q'\| + \|y - x\|) \setminus B(x, \|y - x\|)] \\
& \quad p_1(x, y | Z) p_2(x, y | Z) \, dy,
\end{aligned} \tag{25}$$

using in the second equality the Slivnyak-Mecke formula for the Poisson process  $\Phi \setminus \{Z\}$  conditional on  $Z$  [2, 7] (or see Theorem 3.2 in [4]), and using in the last equality a similar argument as when we obtained the last equality in (24). Finally, from (23)-(25) we obtain (18).

### 3.3. Equivariance and scaling properties

The following equivariance property (26) and scaling property (27) simplify things when considering the distribution of  $\mathcal{R}$  and in particular the distribution of  $V$ .

For any number  $s$ ,  $d \times d$  matrix  $\mathcal{O}$ , and set  $A \subseteq \mathbb{R}^d$ , define  $sA = \{sa : a \in A\}$ ,  $A + q' = \{a + q' : a \in A\}$ , and  $\mathcal{O}A = \{\mathcal{O}a : a \in A\}$ . To stress the dependence of  $(\Phi, q, q')$ , write  $\mathcal{R} = \mathcal{R}(\Phi, q, q')$  and  $V = V(\Phi, q, q')$ . Let  $\mathcal{R}(\Phi, l)$  and  $V(\Phi, l)$  denote

the cases of  $\mathcal{R}$  and  $V$  when  $q' = o$  and  $q = u$ , where the first coordinate of  $u$  is  $l > 0$  and the remaining coordinates are zero. Further, let  $\sim$  mean ‘is distributed as’. Then, if  $q - q' = \mathcal{O}u$  and  $\mathcal{O}$  is an orthonormal matrix,

$$\mathcal{R}(\Phi, q, q') \sim \mathcal{O}\mathcal{R}(\Phi, l) + q', \quad V(\Phi, q, q') \sim V(\Phi, l), \quad (26)$$

since  $\mathcal{O}^t(\Phi - q') \sim \Phi$  where  $\mathcal{O}^t$  is the transpose of  $\mathcal{O}$ . Therefore, without any loss of generality we can assume that  $q' = o$  and  $q = u$ .

Moreover, writing  $\Phi = \Phi_\rho$  for the Poisson process on  $\mathbb{R}^d$  with intensity  $\rho > 0$ , we can make the coupling  $\Phi_\rho = \rho^{-1/d}\Phi_1$ . Then  $\mathcal{R}(\Phi_\rho, l) = l\mathcal{R}(\Phi_{\rho l^d}, 1) = \rho^{-1/d}\mathcal{R}(\Phi_1, \rho^{1/d}l)$  and

$$V(\Phi_\rho, l) = l^d V(\Phi_{\rho l^d}, 1) = \rho^{-1} V(\Phi_1, \rho^{1/d}l), \quad (27)$$

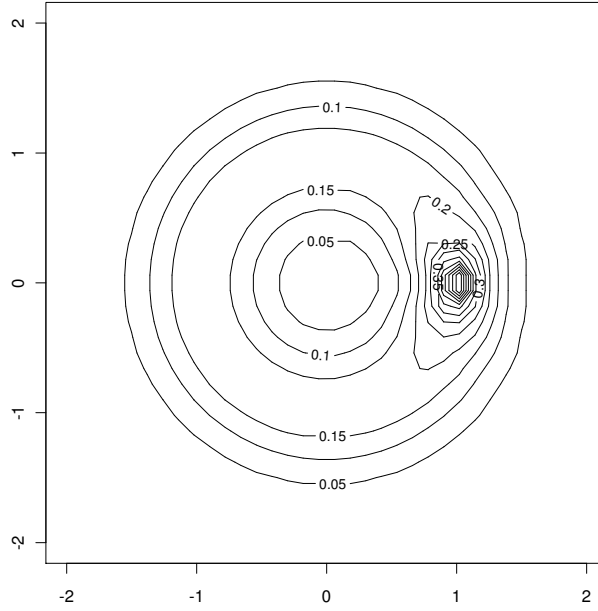
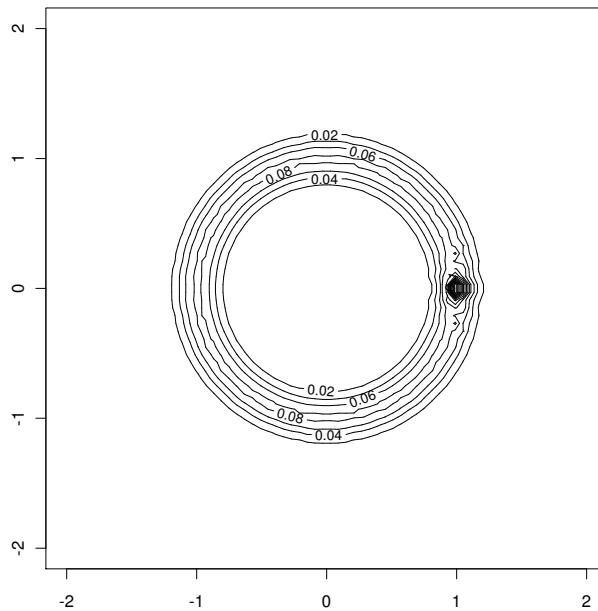
so for the volume of the privacy region it boils down to consider the distribution of either  $V(\Phi_\rho, 1)$  for  $\rho > 0$  (the case with  $l = 1$ ) or  $V(\Phi_1, l)$  for  $l > 0$  (the case with  $\rho = 1$ ).

### 3.4. Numerical results

This section discusses some numerical results for  $p(x)$  and  $E(V)$ .

Recall that  $p(x)$  is the probability that  $x$  belongs to the inferred privacy region. Clearly,  $p(q) = 1$  since  $q \in \mathcal{R}$ . For  $x \neq q$  and  $l$  fixed, both  $p(x)$  and  $E(V)$  approach 0 as  $\rho$  tends to  $\infty$ . This follows by combining (14)–(16) and (17)–(18), and it is accordance with the fact that as  $\rho \rightarrow \infty$ , then  $\|Z - q\| \rightarrow 0$  almost surely, and so  $\mathcal{R}$  tends to the empty set. It is interesting to study how  $p(x)$  varies as a function of  $x$  when  $d = 2$ . Figure 3 plots the contours of  $p(x)$  when  $\rho = 1$  and  $l = 1$ . Observe that contours with high  $p(x)$  are located close to the point  $q = (1, 0)$ . Contours with low  $p(x)$  appear as circle-like shapes, with centre  $q' = (0, 0)$  and in pairs, with radii below 1.0 and above 1.0, respectively. Figure 4 shows the contours of  $p(x)$  when  $\rho = 10$  and  $l = 1$ . This function resembles that of Figure 3, except that the region with high  $p(x)$  is shrunked significantly.

Two methods are employed to evaluate  $E(V)$  when  $d = 2$ . Again we take  $q' = (0, 0)$  and  $q = (l, 0)$ . The first method is *Monte Carlo*, which executes 10,000 instances of the conventional radial simulation algorithm from [6] for generating a stationary Poisson process until the termination time  $M$  is determined. For each simulation, we estimate

FIGURE 3: Contours of  $p(x)$ , when  $\rho = 1$ ,  $l = 1$ , and  $d = 2$ .FIGURE 4: Contours of  $p(x)$ , when  $\rho = 10$ ,  $l = 1$ , and  $d = 2$ .

the area of  $V$  by using an  $100 \times 100$  square grid over the domain  $[-R_M, R_M]^2$  which contains  $\mathcal{R}$ . The value of  $E(V)$  is then estimated by the average area obtained from all

simulations. The second method is *numerical integration*, using the following setting when  $l = 1$  and  $\rho = 1, 2, 5, 10, 20, 50, 100$ .

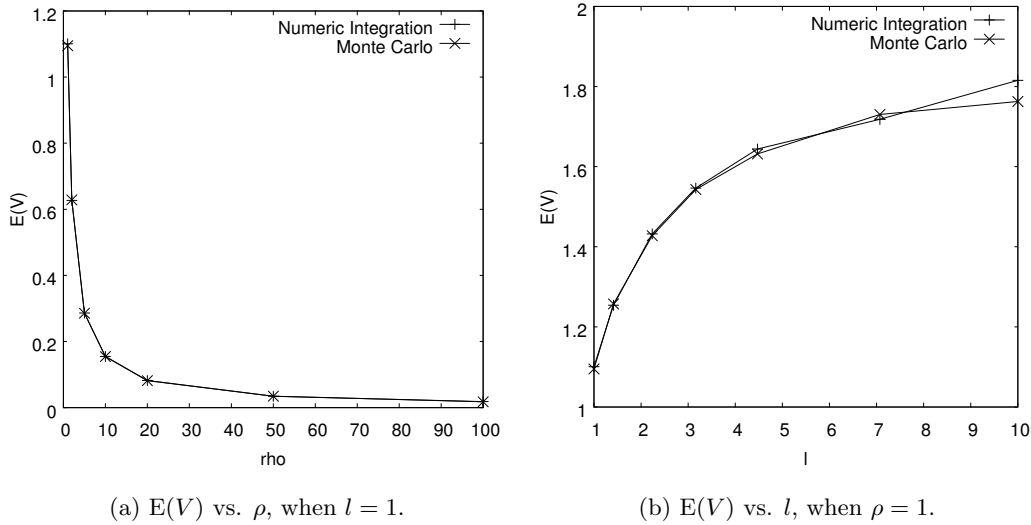
- The integral in (18) is computed by using a  $100 \times 100$  square grid in the region  $[-(t+l), (t+l)]^2$ . All possible points  $y$  such that the indicator function in (18) is equal to one must be located inside such a region.
- The integral in  $p(x) = \int p(x|z)f(z) dz$  is computed by using a  $100 \times 100$  square grid in the region  $[-3l, 3l]^2$ . In fact  $p(x)$  is effectively zero outside this region and seemingly only limited value is lost even though the full space  $\mathbb{R}^2$  is not used as the domain for numeric integration.
- For  $\rho = 1, 2, 5, 10, 20, 50$ , the integral in  $E(V) = \int p(x) dx$  is computed by using a  $100 \times 100$  square grid in the region  $[-3l, 3l]^2$ . Like above, only limited value is lost when  $[-3l, 3l]^2$  is used as the bounding region. For  $\rho = 100$ , after first using the same method and comparing with the Monte Carlo estimate, we found it appropriate to use polar coordinates  $(\theta, r)$  for  $x$  and a  $100 \times 100$  square grid for  $(\theta, r)$  in the domain  $[0, 2\pi) \times [0.75, 1.25)$ . Outside this domain  $p(x)$  is almost zero.

Figure 5a shows  $E(V)$  as a function of  $\rho$  when  $l = 1$  and  $d = 2$ . Observe that the value obtained by numerical integration is close to the corresponding value obtained by Monte Carlo. For  $\rho = 100$ , the difference between the values obtained by numerical integration and Monte Carlo becomes more visible (the estimates of  $E(V)$  when  $(\rho, l) = (100, 1)$  and obtained using a square grid for respective  $x$  and its polar coordinates are 0.0160 and 0.0182 as compared to the Monte Carlo estimate 0.0176). Figure 5a suggests that  $E(V)$  is a strictly decreasing convex function of  $\rho$ .

Figure 5b plots  $E(V)$  as a function of  $l$  when  $\rho = 1$  and  $d = 2$ . This plot is just obtained from the results in Figure 5a using (27), i.e., the fact that  $V(\Phi_1, l) \sim l^d V(\Phi_{l^d}, 1)$ . Note that  $E(V)$  appears to be an increasing concave function of  $l$ .

### Acknowledgements

An anonymous referee and the coordinating editor of SGSA are acknowledged for helpful comments. Jesper Møller was supported by the Danish Natural Science Re-

FIGURE 5: Evaluation of  $E(V)$ .

search Council, grants 272-06-0442 and 09-072331, "Point process modelling and statistical inference", and by the Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation. Man Lung Yiu was supported by ICRG grants A-PJ79 and G-U807 from the Hong Kong Polytechnic University.

### References

- [1] KINGMAN, J. F. C. (1993). *Poisson Processes*. Oxford University Press.
- [2] MECKE, J. (1967). Stationäre zufällige Maße auf lokalkompakten Abelschen Gruppen. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **9**, 36–58.
- [3] MØLLER, J. (1994). *Lectures on Random Voronoi Tessellations*. Lecture Notes in Statistics 87, Springer-Verlag, New York.
- [4] MØLLER, J. AND WAAGEPETERSEN, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton.
- [5] OKABE, A., BOOTS, B., SUGIHARA, K. AND CHIU, S. (2000). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams* second ed. Wiley.



- [6] QUINE, M. P. AND WATSON, D. F. (1984). Radial Generation of  $n$ -Dimensional Poisson Processes. *Journal of Applied Probability* **21**, 548–557.
- [7] SLIVNYAK, I. M. (1962). Some properties of stationary flows of homogeneous random events. *Theory Prob. Appl.* **7**, 336–341.
- [8] STOYAN, D., KENDALL, W. S. AND MECKE, J. (1995). *Stochastic Geometry and Its Applications* second ed. J. Wiley & Sons, Chichester.
- [9] YIU, M. L., JENSEN, C. S., HUANG, X. AND LU, H. (2008). Spacetwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services. In *ICDE*. pp. 366–375.