An analytical calculation of neighbourhood order probabilities for high dimensional Poissonian processes and mean field models

2007 J. Phys. A: Math. Theor. 40 1981

(http://iopscience.iop.org/1751-8121/40/9/005)

J. Phys. A: Math. Theor. **40** (2007) 1981–1989 doi:10.1088/1751-8113/40/9/005

An analytical calculation of neighbourhood order probabilities for high dimensional Poissonian processes and mean field models

César Augusto Sangaletti Terçariol[1,2], **Felipe de Moura Kiipper**[1] **and Alexandre Souto Martinez**[1]

[1] Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Avenida Bandeirantes, 3900, 14040-901, Ribeirãao Preto, SP, Brazil
[2] Centro Universitário Barão de Mauá, Rua Ramos de Azevedo, 423, 14090-180, Ribeirão Preto, SP, Brazil

E-mail: cesartercariol@gmail.com, fekiipper@yahoo.com and asmartinez@ffclrp.usp.br

Received 8 September 2006, in final form 3 January 2007
Published 14 February 2007
Online at stacks.iop.org/JPhysA/40/1981

Abstract
Consider that the coordinates of N points are randomly generated along the edges of a d-dimensional hypercube (random point problem). The probability $P_{m,n}^{(d,N)}$ that an arbitrary point is the mth nearest neighbour to its own nth nearest neighbour (Cox probabilities) plays an important role in spatial statistics. Also, it has been useful in the description of physical processes in disordered media. Here we propose a simpler derivation of Cox probabilities, where we stress the role played by the system dimensionality d. In the limit $d \to \infty$, the distances between pair of points become independent (random link model) and closed analytical forms for the neighbourhood probabilities are obtained both for the thermodynamic limit and finite-size system. Breaking the distance symmetry constraint drives us to the random map model, for which the Cox probabilities are obtained for two cases: whether a point is its own nearest neighbour or not.

PACS numbers: 05.90.+m, 02.50.Ey

1. Introduction

Consider N points, whose coordinates are independently and uniformly distributed along the edges of a d-dimensional hypercube. The determination of the distance and neighbourhood statistics between any pair of points is known as the *random point problem* (RPP). This is a standard approach to construct disordered (random) media.

Due to boundary effects and triangular restrictions, the distances between any pair of points are not all independent random variables. For fixed N in the RPP, as the system dimensionality d increases, the boundary effects become more and more pronounced and the distances between pair of points become less and less correlated. One can minimize boundary effects considering periodic boundary condition, and in the limit $d \to \infty$ all the two–point distances are independent and identically distributed (i.i.d.) random variables. This is the *random link (distance) model* (RLM) [1], which is a mean field description of the RPP. Intuitively one think of the distance D_{ij} between the arbitrary points i and j as $D_{ij}^2 = \sum_{k=1}^{d} \left[D_k^{(ij)}\right]^2$, with $D_k^{(ij)} = x_k^{(i)} - x_k^{(j)}$. If $D_k^{(ij)}$ were independent and gaussian distributed (zero mean and unitary variance), the distribution of D_{ij}^2 would be χ^2 with $d-1$ degrees of freedom, which converges to a gaussian in the limit $d \to \infty$. The intriguing point is that the coordinates are quenched, so that the $D_k^{(ij)}$ are not all independent. Nevertheless, the central limit theorem is expected to be valid even with finite range correlations (scales are renormalized to the correlation length). This leads to a gaussian random variable. The dependence among the D_{ij} is assured by the triangular inequality, which is somehow weakened as d increases.

In the RLM, there exist two Euclidean constraints: (i) the distance from a point to itself is always null ($D_{ii} = 0$, for all i) and (ii) the forward and backward distances are equal ($D_{ij} = D_{ji}$, for all i, j). The second constraint limits numerical simulations to small systems, as long as the distance matrix requires a memory consumption of order $O(N^2)$ in a conventional implementation (note that in the RPP the coordinates of points are at disposal and one can calculate all the distances related to only one point at once, leading to a memory consumption of order $O(N)$). The algorithm presented in [2] implements RLM with memory consumption of order $O(N)$, as in RPP. If the distance symmetry constraint is broken, the model becomes the *random map model* (RMM) [3, 4]. In this latter model, a point can be whether its own nearest neighbour ($D_{ii} = 0$) or not ($D_{ii} \neq 0$), which is the mean field approximation for Kauffman automata [5].

Both, the RPP and RLM have been very fruitful in the determination of numerical and analytical results in several interesting systems. Applications range from statistics on the optimal trajectories in the context of travelling salesman problem on a random set of cities [6–10], passing by frustrated dimerization optimization modelled by the minimum matching problem [11, 12] (or equivalently spin-glasses [11]), and going to partial self-avoiding deterministic tourist walk [13–16] and its stochastic version [17, 18]. Partial self-avoiding walks have been our main motivation to address the RPP and its mean field models. Although the distance distribution as a function of the dimensionality d plays an important role in the stochastic tourist version, in the deterministic case one is mainly interested on the neighbourhood ranking of random points.

As pointed above, boundary effects are important as the dimensionality of the system increases. The points get closer to the surface and to capture the bulk effect, one must increase N. In certain systems it may be difficult to have such large N values and it would be suitable to have analytical expressions for finite N, for instance, to test reliability of numerical codes or to develop new statistical tests.

Here we focus on the distribution of neighbourhood ranks. The probability $P_{m,n}^{(d,N)}$ that an arbitrary point is the mth nearest neighbour of its own nth nearest neighbour in the RPP has attracted attention of researchers since the seminal studies of Clark and Evans [19] and Clark [20] on some aspects of spatial pattern in biological populations. They devised the term reflexive neighbours for the case $m = n$ and their calculated reflexive neighbourhood probability ranking has been corrected by Dacey [21] ($m > 1$) in the context of geographical analysis and then generalized (for $m \neq n$) by Cox [22], which we call the *Cox probabilities*.

In this paper, in section 2 we obtain the Cox probabilities using only Poisson distribution instead of the various distinct distributions used in the original paper [22]. As in Cox calculation, we write the probabilities in the thermodynamic limit $N \to \infty$. Unlike Cox,

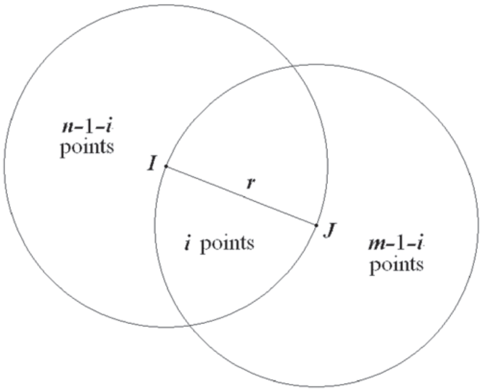

Figure 1. Two-dimensional Poissonian process. The circles centred in the points I and J have surface $V_2 = \pi r^2$ and the intersection has an area $V_{\cap,2} = V_2(1 - p_2) = (2\pi/3 - \sqrt{3}/2)r^2$. There are i points in the intersection of the V_2 surfaces and in the I and J crescents there are $n - 1 - i$ and $m - 1 - i$ points, respectively.

we write them in terms of known functions (rather than in terms of an integral) and known distributions (multinomial, binomial and hypergeometric). In section 3, the use of known special functions allows us to take the high dimensionality limit, which leads to the RLM neighbourhood probability. Using similar arguments to obtain Cox probabilities, we are able to obtain neighbourhood probability for finite-size RLM systems. Finally, in section 4 we explicitly write the Cox probabilities for the two considered case of the RMM and finish with the concluding remarks (section 5). All analytical results have been compared and validated by numerical Monte Carlo simulations. These results have been important in the validation of numerical codes and in the derivation of analytical results in the study of partially self avoiding deterministic and stochastic walks presented in [13–18].

2. Alternative derivation of Cox probabilities

This alternative derivation of Cox formula is simpler than the original paper, since it uses only the Poisson distribution, rather then the Poisson, binomial and gamma distributions as in the original paper.

In a d-dimensional Poissonian medium with a mean density of λ_d points per unitary volume, the probability that a volume V_d (with an arbitrary shape, even with disconnected parts) contains k points is given by the Poisson distribution $\text{Pois}(\mu; k) = \mu^k \mathrm{e}^{-\mu}/k!$, where $k = 0, 1, 2, \ldots, \infty$ and $\mu = \langle k \rangle = \lambda_d V_d$ is the expected number of points inside the volume V_d. Note that the thermodynamic limit is taken letting k freely vary and that the only parameter of this distribution is μ (the medium dimensionality d is not a relevant quantity).

Let I and J be two points of a d-dimensional space separated by a distance r. The volume $V_d(r)$ of the hypersphere of radius r centred in I (thus, which pass through J) is $V_d(r) = \pi^{d/2} r^d / \Gamma(d/2 + 1) = \pi^{(d-1)/2} r^d B[1/2, (d + 1)/2]/\Gamma[(d + 1)/2]$, where $\Gamma(z) = \int_0^\infty \mathrm{d}t\, t^{z-1} \mathrm{e}^{-t} = (z - 1)!$ is the *gamma function* [23] and $B(a, b) = B(b, a) = \int_0^1 \mathrm{d}t\, t^{a-1}(1 - t)^{b-1} = \Gamma(a)\Gamma(b)/\Gamma(a + b)$ is the *beta function* [23]. While the former generalizes the factorial the latter is a generalization of the inverse of Newton binomial. Obviously the volume of the hypersphere centred in J passing through I is also $V_d(r)$. Figure 1 shows the case $d = 2$.

The volume $V_{\cap,d}(r)$ of the intersection of these two hyperspheres is $V_{\cap,d}(r) = \{\pi^{(d-1)/2} r^d / \Gamma[(d+1)/2]\} \int_{1/4}^{1} \mathrm{d}t\, t^{-1/2}(1-t)^{(d-1)/2}$. The relative volume of a crescent (compared to one hypersphere) is $p_d = [V_d(r) - V_{\cap,d}(r)]/V_d(r) = \int_0^{1/4} \mathrm{d}t\, t^{-1/2}(1-t)^{(d-1)/2}/B[1/2, (d+1)/2]$ or

$$p_d = I_{1/4}\left(\frac{1}{2}, \frac{d+1}{2}\right), \tag{1}$$

where $I_z(a,b) = \int_0^z \mathrm{d}t\, t^{a-1}(1-t)^{b-1}/B(a,b)$ with $\mathrm{Re}(a) > 0$, $\mathrm{Re}(b) > 0$ is the *normalized incomplete beta function* [23]. Note that p_d depends exclusively on the dimensionality d and does not depend on the hypersphere radius r.

It is interesting to mention that p_d plays an important role in the parametrization of the deterministic tourist walk problem [15, 16]. It can be generalized to an arbitrary distance $D_{IJ} = rx$ between the points I and J, with x ranging from 0 to 2 (from concentric hyperspheres to disjoined ones). In this case, one has $p_d(x) = I_{(x/2)^2}[1/2, (d+1)/2]$, with $p_d(1) = p_d$, which has allowed us to tackle analytically the stochastic tourist walk problems [17, 18]. Further, the mean overlap calculation has been done by Dall and Christensen in [24] in the random graph context.

The following conditions must hold for I be the mth nearest neighbour of J and J be the nth nearest neighbour of I:

(i) there must exist i points inside the intersection of the hyperspheres, with i ranging from 0 to $\min(m-1, n-1)$, the expected number of points is $\mu(1-p_d)$;
(ii) there must exist $m-1-i$ points inside the crescent of J, the expected number of points is μp_d;
(iii) there must exist $n-1-i$ points inside the crescent of I, the expected number of points is also μp_d;
(iv) the distance r between I and J may assume any value in the interval $[0, \infty)$, allowing the volume $V_d(r)$ and expected number of points $\mu = \lambda_d V_d$ inside it also vary from 0 to ∞ (continuous value).

Taking these conditions altogether, one obtains the following expression for the probability $P_{m,n}^{(d)} = P_{n,m}^{(d)}$:

$$P_{m,n}^{(d)} = \int_0^\infty \mathrm{d}\mu \sum_{i=0}^{\min(m-1,n-1)} \frac{[\mu(1-p_d)]^i\, \mathrm{e}^{-\mu(1-p_d)}}{i!} \cdot \frac{(\mu p_d)^{m-1-i}\, \mathrm{e}^{-\mu p_d}}{(m-1-i)!} \cdot \frac{(\mu p_d)^{n-1-i}\, \mathrm{e}^{-\mu p_d}}{(n-1-i)!}.$$

Collecting the factors which do not depend on μ, the remaining integral can be written in terms of the gamma function: $\int_0^\infty \mathrm{d}\mu\, \mu^{m+n-2-i}\, \mathrm{e}^{-\mu(1+p_d)} = \Gamma(m+n-1-i)/(1+p_d)^{m+n-1-i}$ and one obtains the original form of Cox probabilities:

$$P_{m,n}^{(d)} = \sum_{i=0}^{\min(m-1,n-1)} \frac{(m+n-2-i)!}{i!(m-1-i)!(n-1-i)!} \frac{(1-p_d)^i\, p_d^{m+n-2-2i}}{(1+p_d)^{m+n-1-i}} \tag{2}$$

with $m = 1, 2, \ldots, \infty$ and $n = 1, 2, \ldots, \infty$. Letting i vary from 1 to $\min(m, n)$ and rearranging the terms, one has

$$\frac{P_{m,n}^{(d)}}{P_{1,1}^{(d)}} = \sum_{i=1}^{\min(m,n)} \mathrm{Mult}\left(i-1, m-i, n-i; \frac{1-p_d}{1+p_d}, \frac{p_d}{1+p_d}, \frac{p_d}{1+p_d}\right) \tag{3}$$

$$P_{1,1}^{(d)} = \frac{1}{1+p_d}, \tag{4}$$

Table 1. Some values of neighbourhood probability. For low dimensionalities, one uses equation (2). An interesting limiting case is $d = 0$, which yields $p_0 = \int_0^{1/4} \mathrm{d}t/[\pi\sqrt{t(1-t)}] = 1/3$. For $d \gg 1$, one uses equation (6) and for the random link model $d \to \infty$, one uses equation (9).

d	p_d	$P_{1,1}^{(d)}$	$P_{1,2}^{(d)}$	$P_{2,2}^{(d)}$
0	1/3	3/4	3/16	15/32
1	1/2	2/3	2/9	10/27
2	$\frac{2\pi+3\sqrt{3}}{6\pi}$	$\frac{6\pi}{8\pi+3\sqrt{3}}$	$\frac{6\pi(2\pi+3\sqrt{3})}{(8\pi+3\sqrt{3})^2}$	$\frac{6\pi(40\pi^2+12\sqrt{3}\pi+27)}{(8\pi+3\sqrt{3})^3}$
3	11/16	16/27	176/729	6032/19683
$\vdots$		$\vdots$	$\vdots$	$\vdots$
$\gg$1	$1-\alpha_d$	$(1+p_d)^{-1}$	$p_d(1+p_d)^{-2}$	$(1+p_d^2)(1+p_d)^{-3}$
∞ (rl)	1	1/2	1/4	1/4

where $P_{1,1}^{(d)}$ is the couple density (mutually nearest neighbours) and $\mathrm{Mult}(n_a, n_b, n_c; \pi_a, \pi_b, \pi_c) = (n_a + n_b + n_c)!\pi_a^{n_a}\pi_b^{n_b}\pi_c^{n_c}\big/(n_a!n_b!n_c!)$ is the multinomial distribution.

Numerical values of equations (1) and (2) are shown in table 1. Note that the Cox probability distribution is not a joint distribution. The summation $\sum_{m,n=1}^{\infty} P_{m,n}^{(d)}$ diverges since for each neighbourhood degree m it must be normalized $\sum_{n=1}^{\infty} P_{m,n}^{(d)} = 1$ and one obtains the mean $\langle n\rangle = m + p_d$ and the variance $\langle n^2\rangle - \langle n\rangle^2 = (2m + p_d - 1)p_d$. The system dimensionality d is the bare parameter that emerges from the medium while the considered neighbourhood order m is fixed according to the convenience.

3. Random link model and high dimensionality probabilities

The high dimensionality can be obtained directly from Cox probabilities. In this procedure, one can easily obtain the first-order correction from the random link model neighbourhood probabilities. Next we recall that we are considering the thermodynamic limit and give a geometrical interpretation for the random link expression, which corresponds to all the points being on the surface surrounding the volume V_d. In the following we correct the random link model neighbourhood probabilities to finite-size systems.

3.1. Thermodynamic limit

Let us consider the high dimensionality situation ($d \gg 1$). This corresponds to take $b = (d+1)/2 \gg a = 1/2$ in equation (1). Since $b \gg a$, the approximation $B(a, b) \approx \Gamma(a)/b^a$ can be used for $I_z(a, b) \approx b^a/\Gamma(a)\int_0^z \mathrm{d}t\, t^{a-1}(1-t)^b$. Once $t \leqslant z = 1/4$ implies $t \ll 1$, the approximation $(1-t)^b = \mathrm{e}^{b\ln(1-t)} \approx \mathrm{e}^{-bt}$ yields $I_z(a, b) \approx \gamma(a, bz)/\Gamma(a)$, where $\gamma(a, b) = \int_0^b \mathrm{d}t\, t^{a-1}\,\mathrm{e}^{-t}$ is the *non-normalized incomplete gamma function* [23], which presents the following property $\gamma(1/2, x) = 2\int_0^{\sqrt{x}} \mathrm{d}t\, \mathrm{e}^{-t^2} = \sqrt{\pi}\mathrm{erf}(\sqrt{x})$ with the *error function* [23] defined by $\mathrm{erf}(z) = (2/\sqrt{\pi})\int_0^z \mathrm{d}t\, \mathrm{e}^{-t^2}$ which monotonically increases from $\mathrm{erf}(0) = 0$ to $\mathrm{erf}(\infty) = 1$. Since $a = 1/2$, the following property [23] can be used: $I_z(a, b) \approx \gamma(1/2, bz)/\Gamma(1/2) = \mathrm{erf}(\sqrt{bz})$ and equation (1) can be re-written as

$$p_d \approx \mathrm{erf}\left(\sqrt{\frac{d}{8}}\right), \tag{5}$$

where a characteristic dimensionality $d_0 = 8$ naturally emerges from the analysis.

The *complementary error function* is defined by $\mathrm{erfc}(z) = (2/\sqrt{\pi})\int_z^\infty \mathrm{d}t\, \mathrm{e}^{-t^2} = 1-\mathrm{erf}(z)$. For $|z| \gg 1$: $\mathrm{erfc}(z) = \mathrm{e}^{-z^2}/(z\sqrt{\pi})(1-z^2/2+\cdots)$ [23]. A further approximation can be performed noticing that $\mathrm{erfc}(z) = 1-\mathrm{erf}(z)$, so that for $|z| \gg 1$, it can be written as

$$p_d \approx 1 - \mathrm{erfc}\left(\sqrt{\frac{d}{8}}\right) = 1 - \alpha_d, \tag{6}$$

where

$$\alpha_d = \frac{1}{\sqrt{\pi}} \frac{\mathrm{e}^{-d/8}}{\sqrt{d/8}} \left(1 - \frac{4}{d} + \cdots\right) \tag{7}$$

is an approximation for the two-hypersphere intersection relative volume.

Using $1 - p_d \approx \alpha_d$ and $p_d \approx 1$, the summation of equation (2) gives Cox probabilities as a power series in α_d for high dimensional systems:

$$P_{m,n}^{(d\gg 1)} = P_{m,n}^{(rl)} + \frac{2^{2-(m+n)}}{B(m-1,n-1)}\alpha_d + \cdots, \tag{8}$$

where in the random link approximation ($d \to \infty$) this probability is

$$\frac{P_{m,n}^{(rl)}}{P_{1,1}^{(rl)}} = \frac{1}{2^{m+n-2}} \frac{\Gamma(m+n-1)}{\Gamma(m)\Gamma(n)} = \mathrm{Bin}\left(m-1, n-1, \frac{1}{2}, \frac{1}{2}\right) \tag{9}$$

$$P_{1,1}^{(rl)} = \frac{1}{2}, \tag{10}$$

with $P_{1,1}^{(rl)}$ being the couple density and $\mathrm{Bin}(n_a, n_b; \pi_a, \pi_b) = (n_a+n_b)!\pi_a^{n_a}\pi_b^{n_b}/(n_a!n_b!)$ is the binomial distribution. Simple expressions can be obtained such as $P_{1,n}^{(rl)} = 1/2^n$, $P_{2,n}^{(rl)} = n/2^{n+1}$.

In the high dimensionality limit $d \to \infty$, the relative volume of the crescent (equation (1)) tends to 1 ($p_d \to 1$) and the expected number of points $\mu(1-p_d)$ inside the intersection vanishes. Since $\lim_{p_d\to 1}[(1-p_d)/(1+p_d)]^i = \delta_{i,0}$, where $\delta_{i,j}$ is the Kronecker delta, the multinomial distribution in equation (3) becomes the binomial distribution of equation (9). This is easily seen if one considers a hypersphere of radius r inside in a hypercube of edge $2r$, as the dimensionality increases the hypersphere volume decreases relatively to the hypercube and difference of volumes increases meaning that all the points lie on the external volume to the hypersphere [25].

The numerical values related to the high dimensionality cases are shown in table 1.

3.2. Finite size system

The RPP high dimensional limit $d \to \infty$ corresponds to the RLM, where all distances become i.i.d. random variables. Since Euclidean distances are only a means to obtain the ranking neighbourhood probabilities, it is independent of particular choice for the distance probability distribution function (pdf) [16]. For simplicity, we will consider uniform deviates in the interval [0, 1] for the distances among the N points.

As before, let I be the mth nearest neighbour of J and J be the nth nearest neighbour of I. Thus, the following conditions hold:

(i) the distance x from I to J may assume any value in the interval [0, 1],
(ii) the distances from J to each of its $m-1$ nearest neighbours must be less than x and
(iii) the distances from J to each of its $N-m-1$ farthest neighbours must be greater than x, as well as

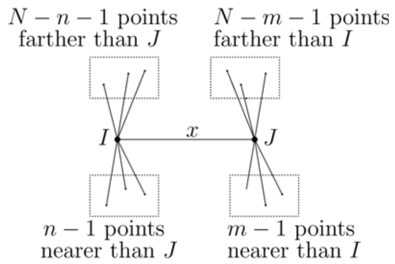

Figure 2. Schematic illustration of the points I and J and their neighbours in a N-point random link model.

(iv) the distances from I to each of its $n-1$ nearest neighbours must be less than x and
(v) the distances from I to each of its $N-n-1$ farthest neighbours must be greater than x.

Figure 2 illustrates the situation.

It also must be noticed that:

(i) choosing an arbitrary point I, its mth nearest neighbour is automatically set, and there is $N-1$ possibilities for this,
(ii) it must be counted all possible combinations in distributing the $N-2$ neighbours of J between the $m-1$ nearest and the $N-m-1$ farthest than J,
(iii) the same counting must be done for the $N-2$ neighbours of I.

Combining these three countings and those five distance restrictions, one has

$$P_{m,n}^{(rl,N)} = \frac{(N-1)[(N-2)!]^2}{(m-1)!(N-m-1)!(n-1)!(N-n-1)!} \times \int_0^1 \mathrm{d}x \left[\int_0^x \mathrm{d}y\right]^{m+n-2} \cdot \left[\int_x^1 \mathrm{d}y\right]^{2N-m-n-2}$$

Since $\int_0^1 \mathrm{d}x\, x^{m+n-2}(1-x)^{2N-m-n-2} = B(m+n-1, 2N-m-n-1) = (m+n-2)!(2N-m-n-2)!/[(2N-3)(2N-4)!]$ then

$$\frac{P_{m,n}^{(rl,N)}}{P_{1,1}^{(rl,N)}} = \mathrm{Hypg}(N-2, N-2; m-1, n-1) \tag{11}$$

$$P_{1,1}^{(rl,N)} = \frac{N-1}{2N-3}, \tag{12}$$

with $m = 1, 2, 3, \ldots, N-1$ and $n = 1, 2, 3, \ldots, N-1$, where $\mathrm{Hypg}(N_a, N_b; n_a, n_b) = \binom{N_a}{n_a}\binom{N_b}{n_b}/\binom{N_a+N_b}{n_a+n_b}$ is the hypergeometric distribution and $P_{1,1}^{(rl,N)}$ is the couple density. These equations (equations (11) and (12)) reduce to equations (9) and (10) as $N \gg 1$.

Figure 3 shows $P_{m,n}^{(rl,N)}$ as a function of n in a ten-point RLM. Note that each curve reaches its maximum at the reflexive case $m = n$ and that they are symmetric in pairs with respect to $N/2$.

4. Random map model

Breaking the distance symmetry constraint $D_{ij} = D_{ji}$ in the RLM leads to the RMM. The RMM is the mean field approximation to several problems and analytical results may be obtained. Also, Cox probabilities can be obtained for the RMM.

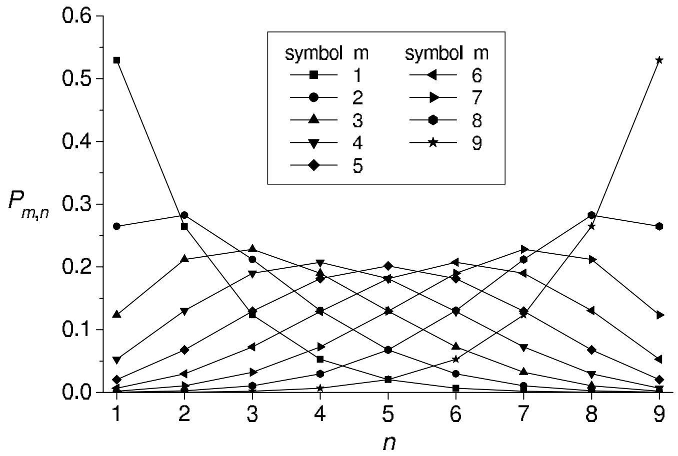

Figure 3. Neighbourhood probabilities in a ten-point RLM. The distributions are discrete and the lines are only a guide to the eyes.

In the case which the constraint $D_{ii} = 0, \forall i$ is preserved, if an arbitrary point *I* is chosen, its *m*th neighbour J is automatically set, but the *n*th neighbour of J is equally probable to be anyone of the other $N - 1$ points, since the distances are totally independent. Thus, the probability $P^{(rm)}_{m,n}$ that the point *I* is the *n*th neighbour of its *m*th neighbour is simply:

$$P^{(rm)}_{m,n} = \frac{1}{N-1}, \tag{13}$$

where $m = 1, 2, \ldots, N-1$ and $n = 1, 2, \ldots, N-1$.

On the other hand, in the case which $D_{ii} \neq 0$ is allowed, the probability $P^{(rm)}_{m,n}$ is twice as large for reflexive neighbours than for non-reflexive ones, because now one must consider that every point is always its own *m*th nearest neighbour, for some *m*. Therefore

$$P^{(rm)}_{m,n} = \frac{1 + \delta_{m,n}}{N+1}, \tag{14}$$

where $\delta_{m,n}$ is the Kronecker delta, $m = 1, 2, \ldots, N$ and $n = 1, 2, \ldots, N$.

Note that in the thermodynamic limit $N \gg 1$, these cases are still distinguishable due to the presence of the factor 2 for the reflexive neighbours.

5. Conclusion

Using only Poisson distribution, Cox probabilities have been obtained through a simple derivation and they have been identified with the multinomial distribution. Writing the dimensionality parameter p_d in terms of the normalized incomplete beta function allowed us to obtain the high dimensional approximation for the neighbourhood probabilities in Poissonian processes (RPP, for instance) as the binomial distribution.

Using the same line of reasoning, the neighbourhood probabilities have been obtained for RLM finite-size systems. In this case the probabilities have been identified with the hypergeometric distribution. Also, simple expressions have been obtained for the RMM.

Up to now, we are devoting efforts to try to obtain the neighbourhood probabilities for finite-size and low-dimensionality systems.

Acknowledgments

The authors thank the discussions with R S González and U P C Neves. ASM acknowledges the support from the Brazilian agencies CNPq (305527/2004-5) and FAPESP (2005/02408-0).

References

[1] Mézard M and Parisi G 1986 Mean-field equations for the matching and travelling salesman problem *Europhys. Lett.* **2** 913–8
[2] Terçariol C A S and Martinez A S 2006 An efficient algorithm to generate random uncorrelated Euclidean distances: the random link model *Braz. J. Phys.* **36** 232–6
[3] Harris B 1960 Probability-distributions related to random mappings *Ann. Math. Stat.* **31** 1045–62
[4] Derrida B and Flyvbjerg H 1987 The random map model—a disordered model with deterministic dynamics *J. Phys. (Paris)* **48** 971–8
[5] Kauffman S A 1969 Metabolic stability and epigenesis in randomly constructed genetic nets *J. Theor. Biol.* **22** 437
[6] Percus A G and Martin O C 1996 Finite size and dimensional dependence in the Euclidean travelling salesman problem *Phys. Rev. Lett.* **76** 1188–91
[7] Cerf N J, de Monvel J H B, Bohigas O, Martin O C and Percus A G 1997 The random link approximation for the Euclidean travelling salesman problem *J. Phys. I (France)* **7** 117–36
[8] Percus A G and Martin O C 1999 The stochastic traveling salesman problem: finite size scaling and the cavity prediction *J. Stat. Phys.* **94** 739–58
[9] Aldous D and Percus A G 2003 Scaling and universality in continous length combinatorial optimization *Proc. Nat. Acad. Sci. USA* **100** 11211–5
[10] Aldous D J 2005 Percolation-like scaling exponents for minimal paths and trees in stochastic mean field model *Proc. R. Soc.* A **461** 825–38
[11] de Monvel J H B and Martin O C 1997 Mean field and corrections for the Euclidean minimum matching problem *Phys. Rev. Lett.* **79** 167–70
[12] Houdayer J, de Monvel J H B and Martin O C 1998 Comparing mean field and Euclidean matching problems *Eur. Phys. J.* B **6** 383–93
[13] Lima G F, Martinez A S and Kinouchi O 2001 Deterministic walks in random media *Phys. Rev. Lett.* **87** 010603
[14] Stanley H E and Buldyrev S V 2001 Statistical physics—the salesman and the tourist *Nature* **413** 373–4
[15] Kinouchi O, Martinez A S, Lima G F, Lourenço G M and Risau-Gusman S 2002 Deterministic walks in random networks: an application to thesaurus graphs *Physica* A **315** 665–76
[16] Terçariol C A S and Martinez A S 2005 Analytical results for the statistical distribution related to a memoryless deterministic walk: dimensionality effect and mean-field models *Phys. Rev.* E **72** 021103
[17] Risau-Gusman S, Martinez A S and Kinouchi O 2003 Escaping from cycles through a glass transition *Phys. Rev.* E **68** 016104 Part 2
[18] Martinez A S, Kinouchi O and Risau-Gusman S 2004 Exploratory behavior, trap models and glass transitions *Phys. Rev.* E **69** 017101 Part 2
[19] Clark P J and Evans F C 1955 On some aspects of spatial pattern in biological populations *Science* **121** 397–8
[20] Clark P J 1956 Grouping in spatial distributions *Science* **123** 373–4
[21] Dacey M F 1969 Proportion of reflexive *n*th-order neighbours in a spatial distribution *Geogr. Anal.* **1** 385–8
[22] Cox T F 1981 Reflexive nearest neighbours *Biometrics* **37** 367–9
[23] Abramowitz M and Stegun I A (ed) 1972 *Handbook of Mathematical Functions* (New York: Dover)
[24] Dall J and Christensen M 2002 Random geometric graphs *Phys. Rev.* E **66** 016121
[25] González R S 2006 Difusão anômala: Transição entre os regimes localizado e estendido na caminhada do turista unidimensional *Master's Thesis* Universidade de São Paulo, Ribeirão Preto