Replicated point processes with application to population dynamics models

Marco Favretti

Dipartimento di Matematica "Tullio Levi-Civita", Università di Padova, Padova, Italy

Abstract

In this paper we study spatially clustered distribution of individuals using point process theory. In particular we discuss the spatially explicit model of population dynamics of Shimatani (2010) which extend previous works on Malécot theory of isolation by distance. We reformulate Shimatani model of replicated Neyman-Scott process to allow for a general dispersal kernel function and we show that the random immigration hypothesis can be substituted by the long dispersal distance property of the kernel. Moreover, the extended framework presented here is fit to handle spatially explicit statistical estimators of genetic variability like Moran autocorrelation index, Sørensen similarity index, average kinship coefficient. We discuss the pivotal role of the choice of dispersal kernel for the above estimators in a toy model of dynamic population genetics theory.

Keywords: Spatial point process, dispersal kernel, kinship coefficient, spatial autocorrelation

1. Introduction

Spatial point processes [1, 2] are a valuable tool for quantitative analysis and model design in population genetics and community ecology studies. They are useful for discussing central issues such as isolation by distance, clustering, distance decay of similarity and more generally in all cases when the spatial distribution of individuals and spatial genetic variability matters. In this paper we focus on the specific class of spatial processes called Neyman-Scott processes [20], see Section 2 below. In these, each point of a given random distribution (called parent points) generates offspring points that are dispersed away from the parent location with a prescribed probability distribution, called dispersal kernel. In a *replicated* Neyman-Scott process, offspring become new parent points while the previous generation

ones are discarded. In this way we build a spatially explicit dynamic genetic model of non-overlapping generations clustered around parent points. Of course the choice of the dispersal kernel function is a crucial issue to get a realistic description of the spatial pattern and in effect there is a large literature dealing with the theoretical choice and experimental validation of this key tool (see [13] and the review [12]).

The original idea behind this model can be traced back to Malécot theory of isolation by distance [4, 5, 6]. Felsenstein [3] criticized the use of two-dimensional Gaussian probability density as the dispersal kernel. This choice seems the most natural one in view of the nature of the computation but it produces an unbounded clustering. Therefore the model was rejected by Felsenstein as physically unsound (see however the comment in [7] to Felsenstein paper). Recently the model has been reformulated by Shimatani [8] in the spatial point process framework still considering a Gaussian dispersal kernel but adding the hypothesis that at each generation a fixed fraction of offsprings migrates randomly away from their parents points. In this way the resulting process reaches a finite clustering even if an infinite number of generations is considered.

In this paper we repeat the computation along the lines of [8] for a general dispersal kernel function. We thus obtain a model of greater generality in which we are able to discuss the role of the choice of dispersal kernel on the issue of divergent clustering. We can show that the use of a leptokurtic dispersal kernel (i.e. with a decay slower than the gaussian or exponential function) do produce a finite clustering without the need of the random migration hypothesis, which is thus replaced by a *long distance dispersal* hypothesis. Gaussian and exponential kernels were commonly considered until the turning of the century, but nowadays are outperformed by various fat tailed dispersal kernels (see Ch.15 in [12]).

Replicated point process as pioneered by Shimatani [8, 9, 10] allow to consider a dynamic and spatially explicit population genetics model. In this way a coalescent theory is recovered. A powerful feature of replicated point process is that they allow to study distance-related statistical properties of spatial *and* genetic patterns. If the population is described by a marked point process $(x, m(x))$ where the qualitative or quantitative mark $m(x)$ encodes the genetic information of the individual sitting at x , genetic variability on a given spatial pattern can be measured by the so called spatial correlation Moran Index [17] while the degree of relatedness between a couple of offsprings can be measured by their average kinship coefficient G , see Section 3 below. In Section 3 we consider a replicated Neyman-Scott process where genetic variability is described by a qualitative mark which

represents the genotype of the individual. We assume that the probability that a parent of genotype i produces an offspring of genotype j is given by a Markov matrix Q_{ij} and that the vector of genotype frequencies q is a stationary distribution for Q . In this way genotype frequencies are constant from one generation to the other. We compute the pair correlation function between offsprings of genotype (i, j) for each generation. As a measure of genetic variability we compute for our model the point process formulation of Moran spatial autocorrelation index (see [9]) and Sørensen similarity index and their infinite generation limit in the case of an ergodic Markov chain. In a specific example where the qualitative marks represents the allele at a single locus, we compute explicitly the Moran index for three different dispersal kernels (Gaussian, exponential, Cauchy) and we show that a long distance dispersal kernel has a long range of spatial genetic correlation.

2. Replicated Neyman-Scott process with a general dispersal kernel

We briefly recall the basic facts of Neyman-Scott point processes (see [1] and [2] for an introduction). They are stationary and isotropic point processes in the plane defined by the following steps

- a) parents points are distributed on the plane with an homogeneous Poisson point process of intensity λ , therefore they represent a collection of points distributed at random uniformly and independently
- b) each parent points produces a number u of daughters sampled from a given probability distribution (usually a Poisson distribution) of given expected value $U = E(u)$ and variance V .
- c) daughter seeds are dispersed from parents location using a two-dimensional radially symmetric probability $h(r)$ (the so called *dispersal kernel*). The probability that a daughter of a parent located at the origin O lands on an infinitesimal area dx centered around point x in the plane is

$$P(O \rightarrow x)dx = h(|x|)dx, \quad \int_0^{+\infty} 2\pi r h(r)dr = 1 \quad (1)$$

Parent point are removed, therefore the intensity (i.e expected spatial density of individuals) of the first generation is λU .

In a replicated Neyman-Scott process, the first generation daughters become parent points and the steps b) and c) are repeated n times to form the spatial

distribution of individuals after n non overlapping generations. Note that only the initial $t = 0$ generation points are randomly distributed whereas the subsequent $t = 1, \dots, n$ generations points forms clusters of individuals. A statistical description of the spatial distribution of individuals can be given along the following lines. Let $n(x)$ and $n(y)$ be the random variables representing the number of individuals (points) falling in infinitesimal disjoint circular regions dx and dy around points x and y . Let λ be the average spatial density of points of the process (first order intensity) and $\rho(x, y)$ the so called second order product density. The following interpretations are standard (hereafter $P(x)$ denotes probability density with respect to the area measure dx)

$$P(x)dx = dP(n(x) = 1) = \lambda dx \quad (2)$$

$$P(x, y)dxdy = dP(n(x) = 1, n(y) = 1) = \rho(x, y)dxdy \quad (3)$$

The pair correlation function g measures the ratio of the probability of finding two individuals at x and y in a clustered distribution and in a completely random one, therefore it gives a measure of the degree of spatial clustering

$$g(x, y) = \frac{P(x, y)}{P(x)P(y)} = \frac{\rho(x, y)}{\lambda^2} \quad (4)$$

For the $t = 0$ generation, which is described by an homogeneous isotropic Poisson process we have $g \equiv 1$, while for the subsequent ones we may have $g > 1$ but the distribution is still stationary and isotropic (i.e rotation and translation invariant), therefore we may always assume that $P(x) = P(|x|)$ and $P(x, y) = P(y - x) = P(|y - x|)$ throughout. By requirement *c*) above, we have that after t generations, the expected density of individuals is (see (2) and *c*) above)

$$P_t(x)dx = \lambda U^t dx \quad (5)$$

Our main aim is to compute the joint presence probability $P_t(x, y)$ for $t \geq 1$. The following computation can be considered the generalization of the one contained in Shimatani paper [8] which is the primary source of inspiration for this work. The technical part of the computation is contained in the Supplementary material, Sect. 5.

Let x_1 and x_2 be the location points of two daughters at the t generation. They may be offsprings of different mothers (dm) located at x and y or be offsprings of the same mother (sm) at x . These are exhaustive and mutually incompatible alternatives, therefore we can write

$$P_t(x_2 - x_1) = P_t(x_2 - x_1, dm) + P_t(x_2 - x_1, sm) \quad (6)$$

For the (sm) case we have, writing $P_{t-1}(x) = \lambda U^{t-1}$, see (5),

$$dP_t(x_2 - x_1, sm) = P_{t-1}(x)dxP(x \rightarrow x_1)P(x \rightarrow x_2)dx_1dx_2 \quad (7)$$

$$= \lambda U^{t-1}dxh(x - x_1)h(x - x_2)dx_1dx_2 \quad (8)$$

to be integrated in dx over the plane and multiplied for the average number of couple of daughters produced at x $E(u(u - 1)) = U^2W$ where $W = (U^2 - U + V)/U^2$ to give

$$dP_t(x_2 - x_1, sm) = \lambda U^{t+1}Wdx_1dx_2 \int_{\mathbb{R}^2} h(x - x_1)h(x - x_2)dx \quad (9)$$

By writing $x - x_2 = x_2 - x_1 - (x - x_1)$ and $dx = d(x - x_1)$ the above integral represents the convolution of h with itself

$$h_{\star}^2(z) = h(z) \star h(z) = \int_{\mathbb{R}^2} h(x)h(z - x)dx$$

For the sake of simplicity, we consider the dx_1dx_2 density of the above function (9)

$$P_t(x_2 - x_1, sm) = \lambda U^{t+1}Wh_{\star}^2(x_2 - x_1). \quad (10)$$

For the (dm) case we have

$$\begin{aligned} dP_t(x_2 - x_1, dm) &= P_{t-1}(y - x)dxdyP(x \rightarrow x_1)P(y \rightarrow x_2)dx_1dx_2 \\ &= P_{t-1}(y - x)dxdyh(x - x_1)h(y - x_2)dx_1dx_2 \end{aligned} \quad (11)$$

to be integrated in the area measure $dxdy$ over all the possible parent locations and multiplied for the average number of daughters produced by a parent at x and y , i.e. $E(u)^2 = U^2$. Passing to the dx_1dx_2 density as above, we have

$$P_t(x_2 - x_1, dm) = U^2 \int_{\mathbb{R}^2 \times \mathbb{R}^2} P_{t-1}(y - x)h(x - x_1)h(y - x_2)dxdy \quad (12)$$

While for the (sm) case we have obtained the closed formula (10), for the (dm) case the probability $P_{t-1}(y - x)$ in (12) has to be computed recursively from (6) and (10). Let us do the computation for a few generations. Since the initial distribution of individuals is random $P_0(y - x) = \rho(x, y) = \lambda^2$ we have easily (using Fubini' Theorem and (1)) that for the $t = 1$ generation $P_1(x_2 - x_1, dm) = \lambda^2 U^2$ hence from (6)

$$P_1(x_2 - x_1) = \lambda^2 U^2 + \lambda U^2 Wh_{\star}^2(x_2 - x_1) \quad (13)$$

For the sake of simplicity, from now on we set $r = x_2 - x_1$. For the $t = 2$ generation, we have from (10) $P_2(r, sm) = \lambda U^3 W h_\star^2(r)$ and, inserting (13) in (12) we get

$$P_2(r, dm) = P_2(r, dm/dm) + P_2(r, sm/dm) = \lambda^2 U^4 + \lambda U^4 W h_\star^4(x_2 - x_1) \quad (14)$$

For the $t = 1$ generation we had two alternatives: (sm) corresponding to half-siblings, and (dm) corresponding to independent lineages; for the $t = 2$ generation we have three alternatives: (sm) , (sm/dm) corresponding to first cousins with a common grandmother or (dm/dm) for independent lineages. By iterating the same computation (see Supplementary Material) we find that at generation t we have the following $t + 1$ exhaustive and mutually incompatible alternatives:

i) $(dm)^t$ corresponding to independent lineages (no common ancestor)

$$P_t(r, dm^t) = P_t(\bar{ca}) = \lambda^2 U^{2t},$$

ii) $sm/(dm)^k$ corresponding to a common ancestor k generations before for $k = 1, \dots, t$, with probability

$$P(r, sm/(dm)^k) = P(r, ca(k)) = \lambda W U^{n+k} h_\star^{2k}(x_2 - x_1) \quad k = 1, \dots, t.$$

Therefore, the probability of having a common ancestor (ca) for two individuals of the generation t at distance r is

$$P_t(r, ca) = \sum_{k=1}^t P_t(sm/dm^k) = \sum_{k=1}^t \lambda W U^{n+k} h_\star^{2k}(x_2 - x_1) \quad (15)$$

while the probability that they represent independent lineages ($\bar{ca}$) is $P_t(r, \bar{ca}) = \lambda^2 U^{2t}$. Their sum gives the joint probability of finding two individuals at distance r

$$P_t(r) = P_t(r, \bar{ca}) + \sum_{k=1}^t P_t(r, ca(k)) = \lambda^2 U^{2t} + \sum_{k=1}^t \lambda W U^{n+k} h_\star^{2k}(r) \quad (16)$$

If we divide the above equality by $P_t(r)$, using the well known relation $P(B|A) = P(A, B)/P(A)$

$$1 = P_t(\bar{ca}|r) + \sum_{k=1}^t P_t(ca(k)|r)$$

we can compute the conditional probability that two individuals belongs to different lineages given that they are at distance r

$$P_t(\bar{c}a|r) = \frac{\lambda^2 U^{2t}}{P_t(r)} \quad (17)$$

or the conditional probability that they share a common ancestor k generations ago given that they are at distance r

$$P_t(ca(k)|r) = \frac{\lambda W U^{n+k} h_{\star}^{2k}(r)}{P_t(r)} \quad (18)$$

All these formulas are useful for spatial genetic applications and they show that the replicated Neyman-Scott process model is an instance of a spatially explicit coalescent theory, as pointed out by [8].

2.1. The long distance dispersal hypothesis

Here we expose the main contribution of this paper. The t generation pair correlation function computed from its definition (4) is

$$g_t(r) = \frac{P_t(r)}{(\lambda U^t)^2} = 1 + \frac{W}{\lambda U^t} \sum_{k=1}^t U^k h_{\star}^{2k}(r). \quad (19)$$

We see that the behavior of the pair correlation function (19) depends on the convergence of the series of function with positive terms $h_{\star}^{2k}(r)$.

Before considering the possible choices of the dispersal kernel h , let us note that in this model the spatial density of individuals after t generations is λU^t , therefore in absence of any regulatory mechanism the density diverges if the fecundity is $U > 1$ while the population goes extinct if $U < 1$. In the forests the density of plants is nearly constant, hence it is reasonable to assume $U = 1$ throughout in our model. In this case W coincides with the variance V and the pair correlation function depends on the ratio V/λ

$$g_t(r) = 1 + \frac{V}{\lambda} \sum_{k=1}^t U^k h_{\star}^{2k}(r) \quad (20)$$

In the previously cited papers, the dispersal kernel was assumed to be a 2-dimensional Gaussian probability density

$$h(r) = \frac{e^{-\frac{r^2}{2\sigma^2}}}{2\pi\sigma^2} \quad (21)$$

whose $2k$ -fold convolution as it is well known is still a Gaussian with variance $k\sigma^2$ that is

$$h_{\star}^{2k}(r) = \frac{e^{-\frac{r^2}{2k\sigma^2}}}{2\pi k\sigma^2} \quad (22)$$

It is easy to see that, in the limit $t \rightarrow \infty$ the series of function in (19) diverges since for every $r \geq 0$

$$\sum_{k=1}^t h_{\star}^{2k}(r) \geq \frac{e^{-\frac{r^2}{2\sigma^2}}}{2\pi\sigma^2} \sum_{k=1}^t \frac{1}{k}.$$

Therefore, in absence of other mechanisms, the degree of clustering diverges for $t \rightarrow \infty$. See however Figure 1 below: after 1000 generations the value of the pcf at $r = 0$ is only 2.6 times the value at $t = 0$, therefore the divergence phenomena could not be visible for realistic values of the parameters.

This feature of Malécot model was pointed out by Felsenstein which on the basis of this result dismissed the model as biologically non relevant. In [8] Shimatani introduces an interesting variant of the model, by supposing that at every generation a constant fraction of the offsprings migrate and are redistributed at random on the plane. As a consequence, the degree of clustering remains finite. In this paper, we select a few kernel functions with various features to test the relevancy of our model (see Nathan review paper on kernels [12] and [13]) and we show that the initial model is capable to show finite clustering if a long distance dispersal (LDD) kernel is used, without need of introducing the at random migration hypothesis.

Figure 1: Pair correlation function $g_t(r)$ for a Gaussian kernel with parameters $\sigma = 1$, $\lambda = 0.05$, $V = 0.1$, $(V/\lambda = 2)$ computed for $t = 1, 5, 10, 20, 100, 1000$ generations. The pcf grows unbounded with the number of generations.

2.1.1. Cauchy dispersal kernel

Figure 2: Pair correlation function $g_t(r)$ for a Cauchy kernel with parameter $b = 1$ for a process with $\lambda = 0.05$, $V = 0.1$ ($V/\lambda = 2$) computed for $t = 1, 5, 10, 20, 100$ generations. The pair correlation function increases with the number of generations but the series converges. Horizontal line is $y = g_\infty(0) = 1 + V\pi/48b^2\lambda$

This kernel is used for seed dispersal studies [15, 12, 19]. It can be obtained as a continuous mixture of Gaussian kernels with variance parameters distributed as the inverse of a Gamma distribution, see [14]. It is defined by

$$h(r) = \frac{b}{2\pi [b^2 + r^2]^{\frac{3}{2}}} \quad (23)$$

where b is a scale parameter. Contrarily to the Gaussian kernel, which has a finite average cluster radius given by

$$r_c = \int_{\mathbb{R}^2} h(x)|x|dx = 2\pi \int_0^{+\infty} r^2 h(r)dr = \sqrt{\pi} \frac{b}{2} \quad (24)$$

for a Cauchy kernel the above integral diverges. Therefore it is well suited for representing long distance dispersal of offsprings. In the Supplementary material we compute its $2k$ -fold convolution to be

$$h_\star^{2k}(r) = \frac{bk}{\pi [4(bk)^2 + r^2]^{\frac{3}{2}}} \quad (25)$$

Since for $r \geq 0$

$$0 < h_\star^{2k}(r) \leq h_\star^{2k}(0) = \frac{1}{8\pi b^2 k^2} \quad \text{with} \quad \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$$

the series of functions $h_{\star}^{2k}(r)$ is uniformly convergent by Weierstrass' criterion. Even if a closed form for the sum of the series is not available, we can compute the value of $g_t(r)$ in (19) for $r = 0$ and $r \rightarrow +\infty$ (with $U = 1$)

$$g_t(r) = \frac{P_t(r)}{\lambda^2} = 1 + \frac{V}{\lambda} \sum_{k=1}^t h_{\star}^{2k}(r) \quad (26)$$

hence

$$g_t(0) = 1 + \frac{V}{\lambda} \sum_{k=1}^t h_{\star}^{2k}(0) = 1 + \frac{V}{8\pi b^2 \lambda} \sum_{k=1}^t \frac{1}{k^2} \quad (27)$$

We remark that, even if the clustering is finite for a Cauchy kernel, we see that the clustering at $r = 0$ *increases* with the number of generations up to a certain limit value. In the limit $t \rightarrow \infty$ we have $g_{\infty}(0) = 1 + V\pi/48b^2\lambda$. Moreover since $h_{\star}^{2k}(r)$ is vanishing for $r \rightarrow \infty$ we have $\lim_{r \rightarrow +\infty} g_t(r) = 1$.

2.1.2. Exponential kernel

It is used as a reference against more fat-tailed kernels. When used to describe pollen or seedlings dispersion it outperforms the Gaussian kernel (which is the kernel used in mechanistic diffusion or random walk models). It is defined by ([12, 13])

$$h(r) = \frac{e^{-\frac{r}{b}}}{2\pi b^2} \quad (28)$$

where b is a scale parameter. The average cluster radius is $r_c = 2b$ and its $2k$ -fold convolution (see Supplementary material) is

$$h_{\star}^{2k}(r) = \frac{r^{3k-1} K_{3k-1}(\frac{r}{b})}{\pi b^{3k+1} 2^{3k} \Gamma(3k)} \quad (29)$$

where $K_n(r)$ is the modified Bessel function of the second kind. To study qualitatively the convergence of the series of functions we adopt the asymptotic approximation *a*) for $r \rightarrow +\infty$ and *b*) for $r \rightarrow 0$ below (see [18] Ch. 9)

$$a) \quad K_n(r) = \sqrt{\frac{\pi}{2r}} e^{-r} [1 + \mathcal{O}(1/r)], \quad b) \quad K_n(r) = \frac{\Gamma(n)}{2} \left(\frac{2}{r}\right)^n$$

so that for large r

$$h_{\star}(r) = \frac{\sqrt{\pi} e^{-\frac{r}{b}}}{2^{3k+\frac{1}{2}} b^{3k-\frac{1}{2}}} \frac{r^{3(k-\frac{1}{2})}}{(3k-1)!}$$

and the series of function obtained neglecting $\mathcal{O}(1/r)$ terms converges and has closed form

$$\sum_{k=1}^{\infty} h_{\star}^{2k}(r) = \frac{\sqrt{\frac{\pi}{2}} e^{-\frac{5r}{4b}} \left(e^{\frac{3r}{4b}} - 2 \sin \left(\frac{2\pi b + 3\sqrt{3}r}{12b} \right) \right)}{6\sqrt{b}\sqrt{r}}$$

Hence for large r

$$g_{\infty}(r) \approx 1 + \frac{V}{\lambda} \frac{e^{-\frac{r}{2b}}}{\sqrt{r}}, \quad \lim_{r \rightarrow +\infty} g_t(r) = 1 \quad (30)$$

For $r \rightarrow 0$ we have

$$h_{\star}^{2k}(r) = \frac{1}{4b^2(3k-1)}$$

therefore

$$g_t(0) = 1 + \frac{V}{\lambda} \sum_{k=1}^t \frac{1}{4b^2(3k-1)}$$

and the pair correlation function at $r = 0$ diverges for $t \rightarrow \infty$ contrarily to the Cauchy kernel.

Figure 3: Pair correlation function $g_t(r)$ for an Exponential kernel with parameter $b = 1$ for a process with $\lambda = 0.05$, $V = 0.1$ ($V/\lambda = 2$) computed for $t = 1, 5, 10, 20, 100$ generations. The pair correlation function increases with the number of generations, diverges for $r \rightarrow 0$ but converges for large r

3. Marked point processes and spatial genetic applications

In this section we consider a replicate marked point process $(x, m(x))$ where the qualitative mark $m(x)$ corresponds to the genotype of the individual at x and we compute the pair correlation function between individuals

of given genotype. This model is related to the one in [9]. We suppose that the parent population is the sum of individuals having m different genotypes. For example, in the case of a single locus having l possible alleles, there are $m = l$ different genotypes for an haploid individual, while there are $m = l^2$ genotypes for a diploid individual. Let λ_i be the intensity of genotype i in the population, $\lambda = \sum_i \lambda_i$ and $q_i = \lambda_i/\lambda$ be the frequency of genotype i . As before, we state that

$$dP(n_i(x) = 1) = \lambda_i dx = \lambda q_i dx \quad (31)$$

A parent of genotype i may have a daughter of genotype j with conditional probability $P(j|i) = Q_{ij}$, where $\sum_j Q_{ij} = 1$, so that the probability of a $i \rightarrow j$ filiation is $Q_{ij}q_i$ and the frequency of genotype j in the first offspring generation is

$$q_j^1 = \sum_i Q_{ij}q_i = (q^T Q)_j \quad (32)$$

where the superscript T denotes transposition. If we stipulate that our reproductive mechanism is described by the time-independent Markov chain Q and that the genotype frequencies q are invariant from one generation to another, that is

$$q^1 = q^T Q = q \quad (33)$$

then the population is at the Hardy-Weinberg equilibrium. Moreover we assume that the dispersal mechanism described by the dispersal kernel h is independent of the genotype i . We want to compute the probability of finding two daughters of genotype i and j at distance $r = x_2 - x_1$. As before,

$$P_{ij}(r) = P_{ij}(r, sm) + P_{ij}(r, dm)$$

For the (sm) case, let x be the mother location. The sought probability $P_{ij}(r, sm)$ is the product of the probability $M_{ij}(sm)$ of obtaining a couple of daughters of genotype i, j from a mother of arbitrary genotype

$$M_{ij}(sm) = K_{ij} = \sum_k q_k Q_{ki} Q_{kj} \quad (34)$$

times the probability $P(r, sm)$ in (10) since the dispersal probability kernel h is independent of the genotype. Hence

$$P_{ij}(r, sm) = M_{ij}(sm)P(r, sm) = K_{ij}\lambda V h_\star^2(r) \quad (35)$$

For the (dm) case, let x, y be the parent locations. We have that

$$M_{ij}(dm) = \sum_{k,l} q_k q_l Q_{ki} Q_{lj} = q_i q_j = (q \otimes q)_{ij} \quad (36)$$

due to the stationarity of the chain (33) (Hardy-Weinberg equilibrium hypothesis). Moreover, see (12)

$$P_{ij}(r, dm) = \lambda^2 M_{ij}(dm) \int_{\mathbb{R}^4} h(x - x_1) h(y - x_2) dx dy = \lambda^2 q_i q_j \quad (37)$$

Note that in both (sm) and (dm) cases, $P_{ij} = P_{ji}$. Summing up the above results, for the first offspring generation we have

$$P_{ij}^1(r) = \lambda^2 q_i q_j + K_{ij} \lambda V h_\star^2(r) \quad (38)$$

while for the subsequent generations, in the (sm) case, since the genotype frequency is invariant, the computation gives the same result of (10)

$$P_{ij}^{(t)}(r, sm) = K_{ij} \lambda V h_\star^2(r) \quad (39)$$

For the (dm) case at the t generation, we have

$$M_{ij}^{(t)}(r, dm) = \sum_{kl} P_{kl}^{t-1}(r) Q_{ki} Q_{lj} \quad (40)$$

and

$$P_{ij}^{(t)}(r, dm) = \int M_{ij}^{(t)}(x - y, dm) h(x - x_1) h(y - x_2) dx dy. \quad (41)$$

Let $Q^k = Q \times \dots \times Q$ be the k -times product of Q . It is easy to see that Q_{mi}^k gives the probability that an ancestor of genotype m has a genotype i daughter k generations later. We thus denote with

$$K_{ij}^{(k)} = \sum_m q_m Q_{mi}^k Q_{mj}^k \quad (42)$$

the probability of obtaining a couple of daughters of genotype i, j having a common ancestor k generations before. Let us do the computation for the $t = 2$ generation:

$$\begin{aligned} M_{ij}^{(2)}(r, dm) &= \sum_{kl} P_{kl}^1(r) Q_{ki} Q_{lj} = \sum_{kl} [\lambda^2 q_i q_j + \lambda V K_{kl} h_\star^2(r)] Q_{ki} Q_{lj} \\ &= \lambda^2 q_i q_j + \lambda V h_\star^2(r) \sum_m q_m Q_{mi}^2 Q_{mj}^2 \\ &= \lambda^2 q_i q_j + \lambda V h_\star^2(r) K_{ij}^{(2)} \end{aligned} \quad (43)$$

and

$$\begin{aligned} P_{ij}^{(2)}(r, dm) &= \int M_{ij}^{(2)}(x - y, dm) h(x - x_1) h(y - x_2) dx dy \\ &= \lambda^2 q_i q_j + \lambda V K_{ij}^{(2)} h_\star^4(r) \end{aligned} \quad (44)$$

We recognize in the first term the probability of the event $(dm)^2$ and in the second of (sm/dm) . Therefore, by iterating the computation above, we get

$$P_{ij}^{(t)}(r, sm/dm^k) = P^{(t)}(r, ca(k)) = \lambda V K_{ij}^{(k)} h_{\star}^{2k}(r)$$

and

$$P_{ij}^{(t)}(r) = P_{ij}^{(t)}(r, \bar{ca}) + \sum_{k=1}^t P_{ij}^{(t)}(r, ca(k)) = \lambda^2 q_i q_j + \lambda V \sum_{k=1}^t K_{ij}^{(k)} h_{\star}^{2k}(r)$$

The associated pair correlation function between i, j genotypes is

$$g_{ij}^{(t)}(r) = \frac{P_{ij}^{(t)}(r)}{\lambda^2 q_i q_j} = 1 + \frac{V}{\lambda q_i q_j} \sum_{k=1}^t K_{ij}^{(k)} h_{\star}^{2k}(r) \quad (45)$$

3.1. Ergodic case

We have already assumed that the genotype frequencies q are the stationary (Hardy-Weinberg) equilibrium distribution of the Markov chain Q . If we suppose that $(Q^{\nu})_{ij} > 0$ for some integer ν , then by the Ergodic Theorem for Markov Chains (see e.g. [16]), we have that

$$\lim_{k \rightarrow \infty} (Q^k)_{ij} = q_j, \quad j = 1, \dots, m \quad (46)$$

hence $K_{ij}^{(k)}$ in (42) becomes $K_{ij}^{(\infty)} = q_i q_j$ in the infinite generations limit, and all the pair correlation functions tend to

$$g_{ij}^{(\infty)}(r) = g(r) = 1 + \frac{V}{\lambda} \sum_{k=1}^{\infty} h_{\star}^{2k}(r) \quad (47)$$

3.2. Distance decay of genetic similarity

Isolation by distance is the phenomenon of increasing genetic diversity between two localized subpopulations that are geographically distant or separated by a barrier. If we consider species composition in a given area instead of genetic diversity, the same concept is termed species turnover or β -diversity. Many drivers of diversity such as distance, ecological niches or barriers may act at the same time. Point process models are particularly effective to describe the role of clustering on the decay of similarity. A commonly used statistical estimator is Sørensen similarity index [22] between two regions A and B

$$\chi(A, B) = 2 \frac{s(A, B)}{s(A) + s(B)} \quad (48)$$

where $s(A, B)$ is the number of species present in A and B while $s(A)$ is the number of species in A . This formula has been rewritten in the point process framework in [19] in case A and B are infinitesimal regions dx at distance r as

$$\chi(r) = \lambda \sum_{i=1}^m g_i(r) q_i^2 \quad (49)$$

where m is the species number, $g_i(r)$ is the pair correlation function of species i and q_i is the frequency of species i . This formula can be used to measure decay of genetic similarity with distance if m is the number of genotypes and $g_i(r)$ is given by $g_{ij}^{(t)}(r)$ in (45) for $i = j$. Hence

$$\chi^{(t)}(r) = \lambda \sum_{i=1}^m g_{ii}^{(t)}(r) q_i^2 \quad (50)$$

or equivalently

$$\chi^{(t)}(r) = \lambda \sum_{i=1}^m q_i^2 + V \sum_{i=1}^m \sum_{k=1}^t K_{ii}^{(k)} h_{\star}^{2k}(r) \quad (51)$$

In the infinite distance limit $r \rightarrow \infty$ the second term in (51) vanishes and the above $\chi^{(t)}(r)$ tends to $\lambda \sum_{i=1}^m q_i^2 = \lambda(1 - D)$ where $D = 1 - \sum_{i=1}^m q_i^2$ is the Gini-Simpson index. It represents the similarity at a distance where the clustering has no effect and the spatial distribution of individual can be considered completely random. Interestingly, in the ergodic case and in the infinite generations limit $K_{ii}^{(\infty)} = q_i^2$ and the above $\chi^{(t)}(r)$ tends to

$$\chi^{(\infty)}(r) = \lambda \sum_{i=1}^m q_i^2 \left[1 + \frac{V}{\lambda} \sum_{k=1}^{\infty} h_{\star}^{2k}(r) \right] \quad (52)$$

showing that there is a residual contribution to the similarity due to the clustering.

3.3. Spatial autocorrelation of genotypes

To make the paper self-contained, we briefly recall the notion of Moran' index of spatial correlation [17]. We consider a lattice of n sites and let a_i be the value of a numerical variable at site i . The average and variance of a are

$$\bar{a} = \frac{1}{n} \sum_{i=1}^n a_i, \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^n (a_i - \bar{a})^2$$

Let $W = w_{ij}$, $i, j = 1, \dots, n$ be a matrix with non negative entries giving a weight w_{ij} to every couple of sites, $|W| = \sum_{ij} w_{ij}$ the sum of weights and

$$Z_{ij} = z_i z_j = (z \otimes z)_{ij} \quad \text{where} \quad z_i = \frac{(a_i - \bar{a})}{\sigma} \quad (53)$$

the score matrix. Moran index is

$$I = \frac{1}{|W|} \sum_{ij=1}^n w_{ij} Z_{ij} = \frac{1}{|W|} W \cdot Z$$

If we consider only sites with distance $(i, j) = |i - j| = r$ we obtain the spatial auto-correlation of variable a at distance r

$$I(r) = \frac{1}{|W|} \sum_{|i-j|=r} w_{ij} Z_{ij}$$

Moran index has been rewritten in the language of point processes by Shimatani in [9] by substituting i) regular lattice sites i with a distribution x of points in the plane (spatial point process), ii) the variable a with the quantitative mark $m(x)$ of point x and iii) w_{ij} with $P(x, y)$. To establish a direct link with the genetic model exposed in the section above, we suppose that to every genotype $i = 1, \dots, m$ there corresponds a number a_i , which can be considered a *quantitative* mark $a(m(x))$. Moreover we consider the average and variance of a with respect to the frequency vector q

$$\bar{a} = E_q(a) = \sum_{i=1}^m a_i q_i, \quad \sigma^2 = E_q[(a - \bar{a})^2] = \sum_{i=1}^m (a_i - \bar{a})^2 q_i \quad (54)$$

and $w_{ij} = P_{ij}(r) = \lambda_i \lambda_j g_{ij}(r)$ given by (45) with total weight

$$|P(r)| = \sum_{ij=1}^m P_{ij}(r) = \lambda^2 \sum_{ij=1}^m q_i q_j g_{ij}(r) = \lambda^2 (q \otimes q) \cdot g(r)$$

If we want to compute the value of the index I at generation t , we have to plug into the above formula $g_{ij}^{(t)}(r)$ in (45). Hence, the point process version of Moran Index (see [9], [2]) is

$$I^{(t)}(r) = \frac{1}{|P(r)|} \sum_{i,j=1}^n P_{ij}(r) Z_{ij} = \frac{\sum_{i,j=1}^n q_i q_j g_{ij}^{(t)}(r) Z_{ij}}{\sum_{i,j=1}^n q_i q_j g_{ij}^{(t)}(r)} \quad (55)$$

While the score matrix depends only on the average and variance of a with respect to the equilibrium frequencies q , the pair correlation matrix $g_{ij}^{(t)}(r)$ depends on distance and generations. For r going to ∞ the correlation between individuals vanishes ($P_{ij} = q_i q_j$, $g = 1$) hence Moran index tends to ($\sum_{i,j} q_i q_j = 1$)

$$I^{(t)}(\infty) = \frac{\sum_{i,j=1}^n q_i q_j Z_{ij}}{\sum_{i,j=1}^n q_i q_j} = q \otimes q \cdot Z \quad (56)$$

On the other hand, if the ergodic hypothesis (46) holds, then $g_{ij}^{(t)}(r)$ tends to $g(r)$ in (47) for $t \rightarrow +\infty$ hence again Moran index become independent of the clustering of the process

$$I^{(\infty)}(r) = I^{(t)}(\infty) = q \otimes q \cdot Z \quad \forall r, t.$$

It is instructive to examine in detail the form of the correlation index $I^{(t)}(r)$ for finite distance r . By plugging into (55) the form of the pair correlation function $g_{ij}^{(t)}(r)$ in (45), we get, after simple manipulations and using the fact that the power Q^k of a Markov matrix Q is Markov for all k

$$I^{(t)}(r) = \frac{\frac{\lambda}{V} q \otimes q \cdot Z + \left(\sum_{k=1}^t h_{\star}^{2k}(r) K^{(k)} \right) \cdot Z}{\frac{\lambda}{V} + \sum_{k=1}^t h_{\star}^{2k}(r)} \quad (57)$$

The above limits can be recovered from the last formula since $h_{\star}^{2k}(r) \rightarrow 0$ for $r \rightarrow \infty$ and $K^{(k)} \rightarrow q \otimes q$ for $k \rightarrow \infty$. By rewriting (57) as

$$I^{(t)}(r) = \frac{\left(\frac{\lambda}{V} q \otimes q + h_{\star}^2(r) K^{(1)} + h_{\star}^4(r) K^{(2)} + h_{\star}^6(r) K^{(3)} + \dots \right) \cdot Z}{\frac{\lambda}{V} + h_{\star}^2(r) + h_{\star}^4(r) + h_{\star}^6(r) + \dots} \quad (58)$$

we see that the spatial correlation at very small distances $r \approx 0$ depends on the ratio λ/V and on the value of the dispersal kernel $h_{\star}^{2k}(0)$. For a Cauchy kernel we have

$$h_{\star}^{2k}(0) = \frac{1}{8\pi b^2 k^2} \sim \frac{1}{k^2}$$

while for an exponential kernel we have

$$h_{\star}^{2k}(0) = \frac{1}{4b^2(3k-1)} \sim \frac{1}{k}$$

At very high spatial density of individuals λ the contribution of past generations is weak (see example) and the value of the index tends to the value

of the uncorrelated case $q \otimes q \cdot Z$. At very low spatial density of individuals the effect of past generations can be strong depending on the type of kernel. For a Cauchy kernel representative of a situation where offsprings migrate at very large distances from the parents, only the first few generations affect the value of the index, while for an exponential kernel past generations have a greater weight. See the following example based on a similar one in [9]

3.4. Example

Let us consider a marked point process where the qualitative mark at a single locus of a diploid individual is defined by the following genotypes: $1 = AA$, $2 = A\star$ and $3 = \star\star$. We transform the qualitative mark into a quantitative one by the following one-to-one map $a(1) = 1$, $a(2) = 1/2$, $a(3) = 0$. Given a value $a \in (0, 1)$ representing the frequency of allele A in the population, we suppose that the genotype frequencies are

$$q = (q_1, q_2, q_3) = (a^2, 2a(1-a), (1-a)^2)$$

The average and variance of a , defined by (54) are

$$\bar{a} = E_q(a) = a, \quad \sigma^2 = E_q(a - \bar{a})^2 = \frac{a(1-a)}{2} \quad (59)$$

Let the mutation probabilities be given by the following Markov transition matrix

$$Q = \begin{pmatrix} a & 1-a & 0 \\ \frac{a}{2} & \frac{1}{2} & \frac{(1-a)}{2} \\ 0 & a & 1-a \end{pmatrix}$$

One can check that q is the equilibrium distribution for Q i.e. $q = q^T Q$ see (33), hence, genotypes frequencies and their variance are invariant from one generation to the other. Moreover, since $Q_{ij}^2 > 0$, the hypotheses of the Ergodic Theorem for Markov chains (46) are fulfilled and it holds that $\lim_{t \rightarrow \infty} Q_{ij}^t = q_j$. We may now compute the score matrix Z using (53) and the matrices $K^{(k)}$ using (42). We get

$$Z = \begin{pmatrix} \frac{2(1-a)}{2} & \frac{1-2a}{a} & -2 \\ \frac{1-2a}{a} & \frac{2(1/2-a)^2}{a(1-a)} & \frac{(2a-1)}{1-a} \\ -2 & \frac{(2a-1)}{1-a} & \frac{2a}{1-a} \end{pmatrix}$$

A straightforward computation shows that for this system $q \otimes q \cdot Z = 0$ meaning that in absence of clustering $P_{ij}(r) = q_i q_j$ and the genotypes show

no spatial correlation. By computing Q^k and $K^{(k)}$ in (42) for this system we find that $K^{(k)} \cdot Z = 1/4^k$ (independent of a) hence Moran index in (57) become in this case

$$I^{(t)}(r) = \frac{\sum_{k=1}^t \frac{1}{4^k} h_{\star}^{2k}(r)}{\frac{\lambda}{V} + \sum_{k=1}^t h_{\star}^{2k}(r)} \quad (60)$$

3.5. Comparison of Moran index for different kernels

Figure 4: Left: Moran Index for a Gaussian kernel with parameter $\sigma = 1$ for a process with $\lambda/V = 0.01$ and $t = 1, 5, 10, 20$. Right: Moran Index for an Exponential kernel with parameters $b = \sqrt{\pi}\sigma/4$, $\sigma = 1$ for a process with $\lambda/V = 0.01$ and $t = 1, 5, 10, 20$. Curves are decreasing with the number of generations t . For both kernels there is a steep decay in the genetic clustering for $r > 5r_C$

Figure 5: Moran Index for a Cauchy kernel with parameter $b = 1$ (left) and $b = 4$ (right) for a process with $\lambda/V = 0.01$ and $t = 1, 5, 10, 20$. Curves are decreasing with t . The fat tail in the curves indicates that there is slow decay of spatial genetic correlation with the interdistance

In this Section we compare the plot of Moran Index for Gaussian, exponential and Cauchy kernels. We set the kernel parameters in order to have the same average cluster radius, when possible. For a Gaussian kernel the average cluster radius is $r_C(Gauss) = \sqrt{\pi}\sigma/2$, while for an exponential

kernel it is $r_C(Exp) = 2b$. We therefore set $b = \sqrt{\pi}\sigma/4$. For a Cauchy kernel $r_C(Cauchy) = +\infty$, therefore we set $b = 1$ and $b = 4$ for comparison.

Gaussian kernel: We see (Figure 4) that the genetic clustering decreases rapidly with the number of generations in a range $0 \leq r \leq 6r_C(Gauss)$ and it vanishes for couple of individuals at larger interdistance.

Exponential kernel (Figure 4): the behavior is similar to the Gaussian kernel case, but the genetic clustering at $r \approx 0$ is higher and the index shows little variation for $t > 10$, hence the Moran index can be considered close to the limit value for $t \geq 20$

Cauchy kernel (Figure 5): the curves have a longer tail, hence the genetic relatedness between individuals is nonzero even for larger interdistances. Moreover, the index is clearly close to its limit value for $t \geq 20$.

As a general feature, spatial correlation between individuals at certain interdistance r as measured by the pair correlation function $g_t(r)$ *increases* with the number of generations (see e.g. Figure 3) while the genetic relatedness as measured by Moran spatial auto correlation index $I^t(r)$ or by the average kinship coefficient $G_t(r)$ *decreases* with the number of generation. This is in accord with the results in [9], and at contrast with previous theoretical studies on lattice models (see [9]). However, at a finer scale the behaviour of these statistical descriptors do depend on the form of the dispersal kernel function as shown in this work. Hence, the joint use of pattern and genetic descriptor allows a better design and validation of spatially explicit models to describe population genetics and spatial ecology models (see e.g. [11]).

3.6. Kinship coefficient

A commonly used measure of genetic relatedness is the so called kinship coefficient ϕ between two arbitrarily chosen individuals: it gives the probability that two randomly chosen alleles from the same autosomal locus are identical by descent, i.e they are identical and they come from the same ancestor, see [21]. The kinship coefficient can be computed given the pedigree. ϕ is 1/4 for a couple of full siblings or for a couple parent-offspring, it is 1/8 for a couple of half-sib (only the mother is in common). In our model we have only maternal half-sib pairs. ϕ is zero in the (dm) case because the two individuals are not genetically related, it is 1/8 in the (sm) case and it is

$$\phi(sm/(dm)^k) = \frac{1}{2} \frac{1}{4^k}$$

for two individuals having a common ancestor k generations ago. In [8] Shimatani computes the average kinship coefficient for a couple of individuals

at distance r along t generations as

$$G_t(r) = \sum_{k=1}^t \frac{1}{2} \frac{1}{4^k} \frac{P(sm/(dm)^k)}{P_t(r)} \quad (61)$$

By substituting (15) and (16) in the above formula we get

$$G_t(r) = \frac{\frac{\lambda V}{2} \sum_{k=1}^t \frac{1}{4^k} h_{\star}^{2k}(r)}{\lambda^2 + \lambda V \sum_{k=1}^t h_{\star}^{2k}(r)} = \frac{\sum_{k=1}^t \frac{1}{4^k} h_{\star}^{2k}(r)}{\frac{\lambda}{V} + \sum_{k=1}^t h_{\star}^{2k}(r)} \quad (62)$$

Note that $G_t(r)$ in the above formula coincides with Moran Index $I^{(t)}(r)$ in (60). Therefore the above plots in Figures 4 and Figure 5 can be used to compare the average kinship coefficient for different kernels type.

3.7. Founder effect

The founder effect [23, 24] is the loss of genetic variation that occurs when a new population is established by a very small number of individuals segregated from the original larger population. Being small in size the new population is not a representative sample of the ancestral population, moreover it will undergo considerable genetic differentiation due to the accumulation of inbreeding caused by the isolation.

We show how our model can be tuned to model the founder effect. We mimick the segregation effect by using a short range dispersal kernel and the effect of inbreeding by considering high population density (intensity). Looking at short interdistance (small size community), we want to see if there is still a high value of Moran index after a few generations. From Figure 4 we already know that genetic clustering (spatial correlation) will eventually decrease when the number of generations t become large.

On the one hand, from Figure 6 we see that for Gaussian and Exponential kernel, which are short range dispersal kernel, Moran index remains considerably high also when population density is increased 100 times in a interdistance range between 0 and 5 times the average cluster radius. On the other hand, (see Figure 7) using the Cauchy long dispersal range kernel there is a strong decrease of genetic clustering in the same interdistance range at higher population density. The large dispersal distance property of this kernel prevents the segregation effect and enhance the genetic mixing therefore it prevents the founder effect to display.

In Figure 8 we show the flexibility of the model with respect to the dispersal kernel parameters: using a Gaussian and exponential kernel with an average cluster radius 6 times higher of the one used in Figure 6 and a scale parameter $b = 1$ for Cauchy kernel half the value of the one used in Figure 7 we obtain a comparable behavior for the three kernels.

Figure 6: Moran index for different kernels and increasing intensity. Left: Gaussian kernel with $\sigma = 1$ and intensity $\lambda_0 = 10^{-3}, \lambda_1 = 10^{-2}, \lambda_2 = 10^{-1}$, Right: Exponential kernel with $b = \sqrt{\pi}\sigma/4$ and intensity $\lambda_0, \lambda_1, \lambda_2$. Curves are decreasing with λ

Figure 7: Moran Index for a Cauchy kernel with $b = 2$ and intensity $\lambda_0 = 10^{-3}, \lambda_1 = 10^{-2}, \lambda_2 = 10^{-1}$. The curves is decreasing with λ and show a strong decay in the genetic clustering at short interdistance.

4. Conclusions

The starting point for this work is the replicated Neyman-Scott point process model in [8]. The main weak point of the model in [8] and of previous papers is the use of a short dispersal range Gaussian kernel. In this way the clustering diverges with the number of generations. Here we have computed the pair correlation function of the replicated point process for a general dispersal kernel and given sufficient conditions for the convergence of the clustering. Clustering is finite for leptokurtic (fat tailed) dispersal kernels, which in many studies outperform the Gaussian or exponential kernels at fitting seed dispersal data. The result in this paper has an immediate

Figure 8: Cumulative illustration of the founder effect. Moran Index computed for a Cauchy kernel ($b = 1$, orange), Gauss kernel ($\sigma = 6$, gray) and Exponential kernel ($b = 3\sqrt{\pi}/2$, blue). Intensity is increased with the number of generations: $\lambda/V = 0.001$ for $t = 1$, $\lambda/V = 0.01$ for $t = 5$, $\lambda/V = 0.1$ for $t = 10$ and $t = 20$. For all kernels curves are decreasing with t and λ . Using for the Gaussian and exponential kernel an average cluster radius 6 times the one used in Figure 6, and a scale parameter $b = 1$ half the value of the one used in Figure 7 we obtain a comparable behavior for the three kernels.

application e.g. to the decay of similarity model introduced in [19], which is based on the pair correlation functions of the various species.

Replicated marked point process models considered in [8] are important because they allow to discuss simultaneously spatial and genetic variability patterns. In this way spatial distribution and population genetic information can be used jointly to pinpoint the model and select the form of the dispersal kernel, a task which is not easily performed for fat tailed kernels if only spatial data are available.

In Section 3 we have discussed a simple model of replicated point process of genotyped individuals where the genotype frequencies at equilibrium are expressed by a discrete time, finite state Markov chain. We have computed pair correlation function for genotyped individuals (called mark correlation function in [2]). We have discussed with examples the role of the dispersal kernel in this model for describing classical statistical estimators as Moran Index of spatial correlation of genotypes or the average kinship coefficient. The possibility of using different kernels gives a great flexibility to the model and allows to discuss classical notions as the founder effect. In particular, the divergence of the clustering for short range kernels in the infinite generations limit seems not to be a serious limitation when modeling actual systems of community ecology. We think that the techniques discussed here have a

great potential and a wide range of applicability in population genetic and community ecology studies. A next step would be to relax the assumption of constant intensity (density of individuals) at every generation, in order to include in the model important effect such as bottlenecks or extinctions.

Competing Interests

The author declares no competing financial interests.

References

- [1] Stoyan, Dietrich, and Helga Stoyan. Fractals, random shapes, and point fields: methods of geometrical statistics. Vol. 302. John Wiley & Sons Inc, 1994.
- [2] Illian, Janine, et al. Statistical analysis and modelling of spatial point patterns. Vol. 70. John Wiley & Sons, 2008
- [3] Felsenstein, Joseph. "A pain in the torus: some difficulties with models of isolation by distance." *The American Naturalist* 109.967 (1975): 359-368
- [4] Malécot, Gustave. "The Mathematics of Heredity (Revised, edited and translated by Yermanos, DM)." (1969).
- [5] Malécot, Gustave. "Identical loci and relationship." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Biology and Problems of Health*. The Regents of the University of California, 1967.
- [6] Malécot, G. "The decrease of relationship with distance." *Cold Spring Harbor Symp. Quant. Biol.* Vol. 20. 1955.
- [7] Lalouel, J. M. "The conceptual framework of Malcot's model of isolation by distance." *Annals of human genetics* 40.3 (1977): 355-360.
- [8] Shimatani, Ichiro K. "Spatially explicit neutral models for population genetics and community ecology: Extensions of the NeymanScott clustering process." *Theoretical population biology* 77.1 (2010): 32-41
- [9] Shimatani, Kenichiro. "Point processes for fine scale spatial genetics and molecular ecology." *Journal of Mathematical Methods in Biosciences* 44.3 (2002): 325-352

- [10] Shimatani, K., and M. Takahashi. "On methods of spatial analysis for genotyped individuals." *Heredity* 91.2 (2003): 173.
- [11] Furstenau, Tara N., and Reed A. Cartwright. "The effect of the dispersal kernel on isolation-by-distance in a continuous population." *PeerJ* 4 (2016): e1848.
- [12] Clobert, Jean, et al., eds. *Dispersal ecology and evolution*. Oxford University Press, 2012.
- [13] Bullock, James M., et al. "A synthesis of empirical plant dispersal kernels." *Journal of Ecology* 105.1 (2017): 6-19.
- [14] Clark, James S., et al. "Seed dispersal near and far: patterns across temperate and tropical forests." *Ecology* 80.5 (1999): 1475-1494.
- [15] Leigh Jr, Egbert G. "A Spatially Explicit Neutral Model of β -diversity in Tropical Forests." *Theoretical Population Biology* 62 (2002): 153-168.
- [16] Shiriyayev, Albert Nicolaevich. "Probability" *Graduate texts in mathematics* , volume 95 (1984)
- [17] Moran, Patrick AP. "Notes on continuous stochastic phenomena." *Biometrika* 37.1/2 (1950): 17-23.
- [18] Abramowitz, Milton, and Irene A. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Vol. 55. New York: Dover publications, 1972.
- [19] Tovo, A., and M. Favretti. "The distance decay of similarity in tropical rainforests. A spatial point processes analytical formulation." *Theoretical population biology* 120 (2018): 78.
- [20] Neyman, Jerzy, and Elizabeth L. Scott. "Statistical approach to problems of cosmology." *Journal of the Royal Statistical Society. Series B (Methodological)* (1958): 1-43.
- [21] Crow, James Franklin, and Motoo Kimura. "An introduction to population genetics theory." *An introduction to population genetics theory*. (1970).
- [22] Sørensen, Thorvald. "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application

to analyses of the vegetation on Danish commons.” Biol. Skr. 5 (1948): 1-34.

[23] Wright, Sewall. ”Statistical genetics and evolution.” Bulletin of the American Mathematical Society 48.4 (1942): 223-246.

[24] Wright, Sewall. ”Isolation by distance.” Genetics 28.2 (1943): 114.

[25] Birkinshaw, Mark. ”Radially-symmetric Fourier transforms.” Astronomical Data Analysis Software and Systems III. Vol. 61. 1994.

5. Supplementary Material

5.1. Replicated Neyman Scott process with a radial dispersal kernel

The two-dimensional stationary and isotropic Neyman-Scott process is defined by conditions i) -iii) below:

- i) parents are distributed over the plane following a homogeneous Poisson process with density λ
- ii) each parent produces u daughters where u is a random variable with known distribution, Let $U = \mathbb{E}(u)$ be the expectation and V the variance
- iii) daughters are dispersed around the mother according to a radial dispersal kernel $h(r)$ with $2\pi \int_0^{+\infty} rh(r)dr = 1$

Parents are removed from the plane at each generation produced by steps ii) and iii). The main object to describe a clustered distributions of individuals in the plane is the product density

$$dP(x, y) = P(x, y)dxdy \quad (63)$$

which is the probability of finding two individuals in the infinitesimal areas dx and dy located by the vectors $x = x - 0$ and $y = y - 0$ in the plane, where 0 is an arbitrarily fixed origin in the plane. As usual $dP(x) = \lambda dx$ is the probability of finding an individual located at x . We denote with x the point of the plane and with $x - 0$ the vector from the origin to the point if there is risk of confusing them. The pair correlation function is defined as

$$g(x, y) = P(x, y)/\lambda^2 \quad (64)$$

where λ is the density of individuals. Both can be computed given the dispersal kernel function which gives the probability that a daughter of a mother at x can be found (lands) at u

$$dP(x \rightarrow u) = h(|x - u|)du$$

where we have assumed that $h(|x - y|)$ is a radial function, hence $h(r) = h(-r)$.

Given two points x_1 and x_2 we want to compute the probability $P_1(x_1, x_2)$ of finding two first generation daughters at these points. We have to consider two cases: *dm*) daughters have two different mothers located at x and y respectively and *sm*) daughters have the same mother located at x . Introduce for the case *dm*) the vectors

$$r_1 = x_1 - x, \quad r_2 = x_2 - y, \quad z = y - x, \quad r = x_2 - x_1, \quad \text{so that} \quad r_1 + r = r_2 + z \quad (65)$$

and for the case *sm*)

$$r_1 = x_1 - x, \quad r_2 = x_2 - x, \quad r = x_2 - x_1, \quad \text{so that} \quad r_1 + r = r_2 \quad (66)$$

The sought probability $P(x_1, x_2)$ can be seen as the sum of two mutually incompatible and exhaustive alternatives: *dm* (different mother) and *sm* (same mother) thus

$$P_1(x_1, x_2) = P_1(x_1, x_2, sm) + P_1(x_1, x_2, dm) \quad (67)$$

We have

$$dP_1(x_1, x_2, dm) = dP_0(x)dP_0(y)dP(x \rightarrow x_1)dP(y \rightarrow x_2) = \lambda^2 dx dy h(r_1)h(r_2)dx_1 dx_2$$

Since on the average a mother produces U daughters and the two mothers can be located everywhere in the plane we have to multiply by U^2 and integrate with respect to the area measures dx and dy . Set for simplicity sake

$$\langle f(x), dx \rangle = \int_{\mathbb{R}^2} f(x) dx, \quad \langle f(x, y), dxdy \rangle = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} f(x, y) dxdy$$

and suppose that Fubini Theorem holds everywhere so that

$$\langle f(x, y), dxdy \rangle = \langle \langle f(x, y), dx \rangle, dy \rangle$$

Moreover recall that for a dispersal kernel $\langle h(r), dr \rangle = 1$. Using the change of variables in (65) $x_1 = r_1 + x$ and $x_2 = r_2 + y$ we can integrate with respect to $dr_1 = -dx_1$ and $dr_2 = -dx_2$ so that in the *dm* case we have

$$P_1(x_1, x_2, dm) = \lambda^2 U^2 \langle h(r_1)h(r_2), dr_1 dr_2 \rangle = \lambda^2 U^2 \langle h(r), dr \rangle^2 = \lambda^2 U^2 \quad (68)$$

independent of x_1, x_2 . For the sm case we have

$$dP_1(x_1, x_2, sm) = dP_0(x)dP(x \rightarrow x_1)dP(x \rightarrow x_2) = \lambda dx h(r_1)h(r_2)dx_1dx_2 \quad (69)$$

We have to multiply by $\mathbb{E}(u(u-1)) = WU^2$ where $W = (V + U^2 - U)/U^2$ and use the change of variables $x_1 = x + r_1$ in (66) to integrate in $-dr_1$ hence

$$P_1(x_1, x_2, sm) = \lambda U^2 W \langle h(-r_1)h(r - (-r_1)), h(-r_1) \rangle = \lambda U^2 W d_\star^2(r)$$

where we have introduced the convolution of two functions on the plane $f(x)$ and $g(x)$

$$f \star g(x) = \langle f(y)g(x-y), dy \rangle, \quad f_\star^2(x) = f \star f(x) \quad (70)$$

Note that the convolution of two radial functions is a radial function. Collecting the previous results we have

$$P_1(x_1, x_2) = P_1(x_1 - x_2) = P_1(r) = P_1(r, dm) + P_1(r, sm) = \lambda^2 U^2 + \lambda U^2 W h_\star^2(r) \quad (71)$$

Note also that since $P_1(r) = P_1(r, dm) + P_1(r, sm)$ then

$$1 = \frac{P_1(r, dm)}{P_1(r)} + \frac{P_1(r, sm)}{P_1(r)} = P_1(dm|r) + P_1(sm|r)$$

Now we consider the generation $n = 2$. Note that at each generation the kernel terms like e.g $dP(x \rightarrow x_1)$ do not change, the terms $P_n(x)dx$ have to be updated with the new parent density λU^n and the terms $P_n(x, y)dx dy$ are the result of the $n - 1$ step calculation. Hence

$$dP_n(x_1, x_2, sm) = dP_{n-1}(x)d(r_1)d(r_2)dx_1dx_2 = \lambda U^{n-1} dx h(r_1)h(r_2)dx_1dx_2$$

and the same computation in (69) gives the result

$$dP_n(x_1, x_2, sm) = \lambda W U^{n+1} h_\star^2(r). \quad (71)$$

For the dm case, we use the result (68). Hence

$$\begin{aligned} dP_2(x_1, x_2, dm) &= dP_1(x, y)dP(x \rightarrow x_1)dP(y \rightarrow x_2) \\ &= dP_1(y - x)dP(x \rightarrow x_1)dP(y \rightarrow x_2) \\ &= \left[\lambda^2 U^2 + \lambda U^2 W d_\star^2(z) \right] dx dy h(r_1)h(r_2)dx_1dx_2 \end{aligned}$$

we have to multiply by U^2 and integrate in dx and dy . The first term is as before

$$U^2 \langle \lambda^2 U^2 h(r_1) h(r_2), dr_1 dr_2 \rangle = \lambda^2 U^4$$

while for the second since $z = r_1 + r - r_2$ and $dx = -dr_1$, $dy = -dr_2$

$$\begin{aligned} \langle d_\star^2(z) h(r_1) h(r_2), dx dy \rangle &= \langle h_\star^2(r + r_1 - r_2) h(r_2) h(r_1), dr_1 dr_2 \rangle \\ &= \langle \langle h_\star^2(r + r_1 - r_2) h(r_2), dr_2 \rangle h(r_1), dr_1 \rangle \\ &= \langle h_\star^3(r + r_1) h(r_1), dr_1 \rangle \\ &= \langle h_\star^3(-r - r_1) h(r_1), d(r_1) \rangle = h_\star^4(-r) = h_\star^4(r) \end{aligned}$$

hence

$$P_2(r, sm) = \lambda^2 U^4 + \lambda^2 U^4 h_\star^4(r) \quad (70)$$

and

$$P_2(r) = \lambda W U^3 h_\star^2(r) + \lambda^2 U^4 + \lambda^2 U^4 h_\star^4(r) \quad (70)$$

In the general case, we have the recursive formula with $z = r_1 + r - r_2$

$$P_n(r) = P_n(r, sm) + P_n(r, dm) = \lambda W U^{n+1} h_\star^2(r) + U^2 [\langle P_{n-1}(z) h(r_1) h(r_2), dr_1 dr_2 \rangle]$$

One can check that

$$P_n(r) = \lambda^2 U^n + \lambda W U^n \sum_{i=1}^n U^i h_\star^{2i}(r) \quad (70)$$

and, using the n generation density λU^n , that

$$g_n(r) = \frac{P_n(r)}{(\lambda U^n)^2} = 1 + \frac{W}{\lambda U^n} \sum_{i=1}^n U^i h_\star^{2i}(r) \quad (70)$$

These formulae are the generalization to an arbitrary radial kernel of the results obtained by Shimatani in [8], formulae (9) and (10), for a Gaussian kernel.

5.2. Convolution of 2-dimensional radial functions

We briefly report here the theory necessary for computing the convolution function f given the dispersal kernel h [25]. Given a function $f(x)$ on the plane, $x \in \mathbb{R}^2$, let us denote with $F(f) = \tilde{f}$ its Fourier transform. We recall also the Convolution Theorem:

$$\text{if } h(x) = f(x) \star g(x) \text{ then } \tilde{h} = F(h) = 2\pi \tilde{f} \tilde{g}. \quad (70)$$

For radial functions $f(r) = f(|x|)$ the Fourier transform can be computed¹ using the Bessel function $J_0(z)$

$$J_0(z) = \frac{1}{\pi} \int_0^\pi \cos(z \cos \theta) d\theta \quad (70)$$

as

$$\begin{aligned} \tilde{f}(k) &= \int_0^{+\infty} r f(r) J_0(kr) dr, \\ F^{-1}(\tilde{f}) &= f(r) = \int_0^{+\infty} k \tilde{f}(k) J_0(kr) dk \end{aligned}$$

which shows that the Fourier Transform of a radial function is a radial function and vice versa.

We need to compute $h \star h$ where h is a radial function. Since

$$h \star h = F^{-1}(F(h \star h)) = F^{-1}(2\pi(\tilde{h})^2) \quad (70)$$

we have that $h \star h$ is a *radial* function that can be computed as

$$h \star h(r) = 2\pi \int_0^{+\infty} k (\tilde{h}(k))^2 J_0(kr) dk \quad (70)$$

where

$$\tilde{h}(k) = \int_0^{+\infty} r h(r) J_0(kr) dr. \quad (70)$$

In the general case, writing $h_\star^k = h \star \dots \star h$ to denote the k -fold convolution product, we have

$$h_\star^{2k} = (2\pi)^{2k-1} F^{-1}(F(h)^{2k}) \quad (70)$$

and

$$h_\star^{2k}(r) = (2\pi)^{2k-1} \int_0^{+\infty} k \left[\int_0^{+\infty} r h(r) J_0(kr) dr \right]^{2k} J_0(kr) dk \quad (70)$$

5.2.1. Cauchy kernel

It is defined by

$$h(r) = \frac{b}{2\pi [b^2 + r^2]^{\frac{3}{2}}}$$

where b is a scale parameter. Using the above formula (5.2) we compute its $2k$ -fold convolution

$$h_\star^{2k}(r) = \frac{bk}{\pi [4(bk)^2 + r^2]^{\frac{3}{2}}}$$

¹Note that the Fourier transform exists only if $\int_0^{+\infty} \sqrt{r} |f(r)| dr < +\infty$.

5.2.2. Exponential kernel

It is defined by

$$h(r) = \frac{e^{-\frac{r}{b}}}{2\pi b^2}$$

where b is a scale parameter. The average cluster radius is $r_c = 2b$ and its $2k$ -fold convolution is

$$h_{\star}^{2k}(r) = \frac{r^{3k-1} K_{3k-1}\left(\frac{r}{b}\right)}{\pi b^{3k+1} 2^{3k} \Gamma(3k)}$$

where $K_n(r)$ is the modified Bessel function of the second kind.