Estimating the pollen backward dispersion function using genetic markers

The backward dispersion function ?

- Forward dispersion function:

  ”where does the pollen go when flying from a father tree at 0 ?”

- Backward dispersion function:

  ”where does the pollen grain observed on the seed of a mother tree at 0 come from?

Model

- trees = stationary point process X
- genotypes of trees = marks M
- genotype of seeds attached to trees = marks G

Assumptions

- genotypes are randomly independently distributed among trees
- the loci are independent
- the two alleles at a given locus are independent
- a seed receives an allele from the father, the other from the mother
- the allele received from a given parent (father or mother) is chosen randomly among the two alleles of the parent

Data

- sample some trees and genotype them
- sample some seeds on these trees

locations of other trees unknown

What do researcher in pollen dispersion do ?

X Poisson

- $f(x - y)$ = probability density to find the father of a flower at y at position x
- copaternity $Q(A, B)$ = probability that two flowers on trees at A and B have the same father
- adjust parameters by least square between $Q(A, B)/Q(A, A)$ and its estimator

Computing $Q(A, B)$

$$Q(A,B) = \int_x f(x-A) \qquad f(x-B) \qquad \frac{1}{\lambda} dx$$

tree pollinating A is at x	tree pollinating B falls within the same region	around x there are λdx trees

- self-pollination not taken into account
- dependance between the two pollination events is not taken into account
- Poisson necessary to drop λ in $Q(A, B)/Q(A, A)$

What do we want to estimate ?

$$1_{\{\text{pollinator of } A \text{ is at } x\}} = 1_{\{x \in X\}} 1_{\{x \text{ pollinates } A\}}$$

- $p(\text{the pollinator of } A \text{ is at } x) = \lambda h(x; A) r(x - A)$, $x \neq A$
- $p(\text{the pollinator of } A \text{ is at } x) = h(x; A)$, $\qquad x = A$

- λ intensity of X
- $r(u), u \in \mathrm{R}$ pair correlation function
- $h(y; 0)$ probability that a tree at y pollinates 0

==> $f(x) = 1_{\{x \neq 0\}} \lambda h(x; 0) r(x), \qquad F(0) = h(0; 0)$

Looking at the copaternity

$$1_{\{\text{pollinator of } A \text{ is at } x\}} 1_{\{\text{pollinator of } B \text{ is at } x\}} = 1_{\{x \in X\}} 1_{\{x \text{ pollinates } A\}} 1_{\{x \text{ pollinates } B\}}$$

- $p(\text{the pollinator of } A \text{ and } B \text{ is at } x) = \lambda h^{(2)}(x; A, B) r(x; A, B)$, $x \neq A$,$x \neq B$,
- $p(\text{the pollinator of } A \text{ is at } x) = h^{(2)}(x; A, B)$, $x = A$ or $x = B$

- $$\begin{aligned} h^{(2)}(x;A,B) &= \mathrm{E}(1_{\{x \text{ pollinates } A\}} 1_{\{x \text{ pollinates } B\}} \mid x \in X) \\ &\neq \mathrm{E}(1_{\{x \text{ pollinates } A\}} \mid x \in X) \mathrm{E}(1_{\{x \text{ pollinates } B\}} \mid x \in X) \\ &= h(x;A) h(x;B) \end{aligned}$$

- $\lambda r(x; A, B)$ intensity of the Palm measure with respect to A and B.

$$Q(A,B) = 2h^{(2)}(x;A,B) + \lambda \int_x h^{(2)}(x;A,B) r(x;A,B) dx$$

Conclusion

- No self-pollination taken into account
- what is estimated is not what is wanted

Estimating $f(x)$

Assumptions

- X stationary isotropic point process
- marks M are independent
- No self-pollination

- λ and $r(x)$ classically
- estimate $h(x; 0)$ through observations of the genotypes of x, 0 and seeds on 0

Non parametric estimation

- (m_1, m_2) the alleles of the tree at 0
- (k_1, k_2) the alleles of the tree at B
- (g_1, g_2) the alleles of a seed on the tree at O
- α_g the allele frequency of g in the tree population
- $F = \{$the father gives the first allele$\}$
- $C = \{O$ carries $(m_1, m_2), B$ carries ...$\}$

$$P(g_1 \mid C \cap F) = \frac{1}{2} h(x)(1_{\{g_1=k_1\}} + 1_{\{g_1=k_2\}}) + (1 - h(x))\alpha_{g_1}.$$

$$P((g_1,g_2)\mid C)=a_1\alpha_{g_1}h(x)+a_2\alpha_{g_2}h(x)+bh(x)+c_1\alpha_{g_1}+c_2\alpha_{g_2} \qquad (1)$$

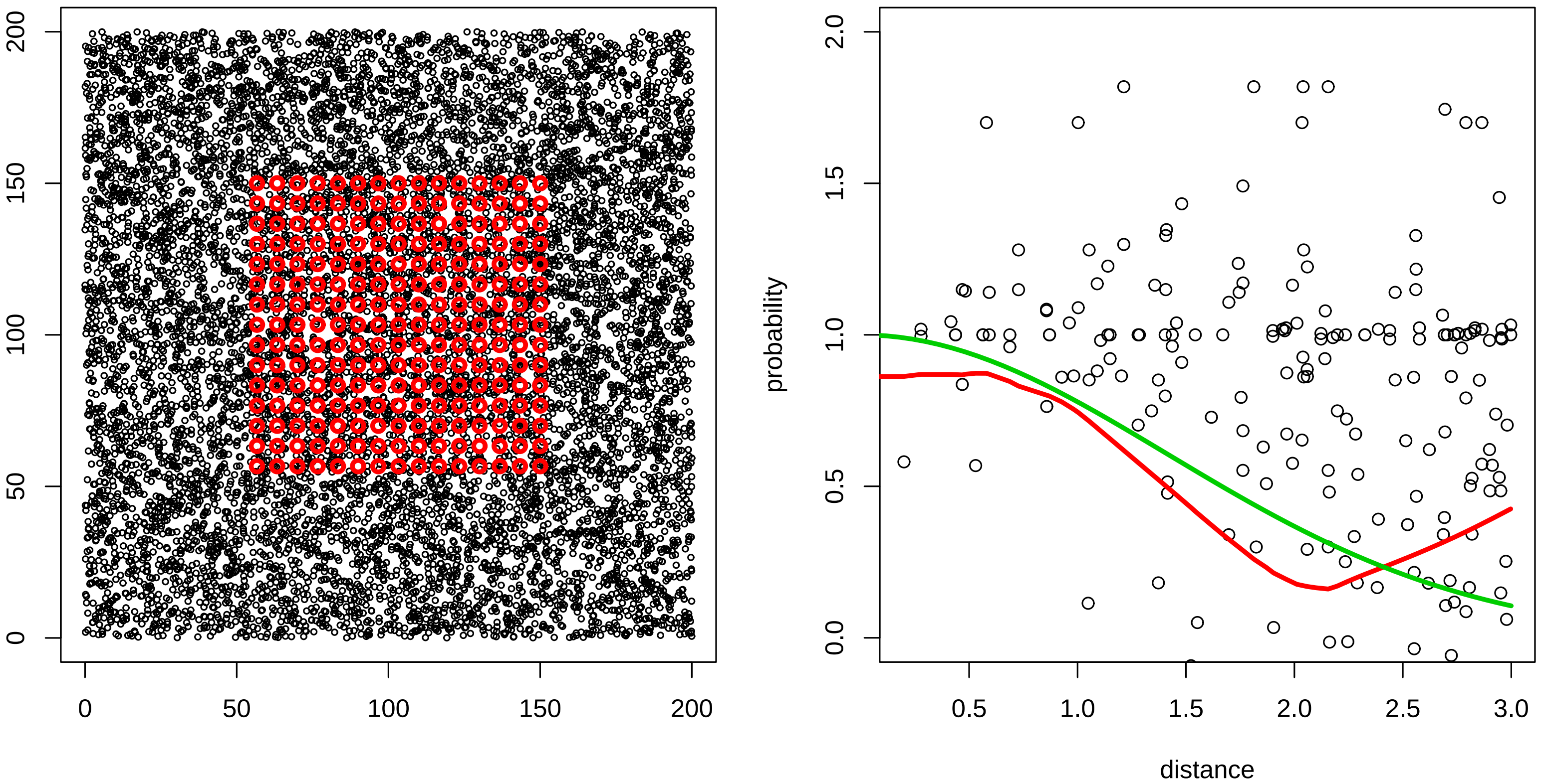
with

$$\begin{aligned}
a_1 &= -\tfrac{1}{4}(1_{\{g_2=m_1\}}+1_{\{g_2=m_2\}})\\
a_2 &= -\tfrac{1}{4}(1_{\{g_1=m_1\}}+1_{\{g_1=m_2\}})\\
b &= \tfrac{1}{8}(1_{\{g_1=k_1\}}+1_{\{g_1=k_2\}})(1_{\{g_2=m_1\}}+1_{\{g_2=m_2\}})\\
&\quad +\tfrac{1}{8}(1_{\{g_2=k_1\}}+1_{\{g_2=k_2\}})(1_{\{g_1=m_1\}}+1_{\{g_1=m_2\}})\\
c_1 &= \tfrac{1}{4}(1_{\{g_2=m_1\}}+1_{\{g_2=m_2\}})=-a_1\\
c_2 &= \tfrac{1}{4}(1_{\{g_1=m_1\}}+1_{\{g_1=m_2\}})=-a_2
\end{aligned}$$

$$\hat{h}(x)=\frac{\sum_i\sum_k\sum_{g_1}\sum_{g_2}K_u(x-\mid x_i-y_i\mid)U_i(g_1,g_2)}{2G^2\sum_i\sum_k K_u(x-\mid x_i-y_i\mid)}$$

with

$$\begin{aligned}
U_i(g_1,g_2)= \quad & \frac{\hat{P}^{(A,i,k)}(g_1,g_2)-c_1^{(A,i,k)}\alpha_{g_1}-c_2^{(A,i,k)}\alpha_{g_2}}{a_1^{(A,B,i,k)}\alpha_{g_1}+a_2^{(A,B,i,k)}\alpha_{g_2}+b^{(A,B,i,k)}}\\
&+\frac{\hat{P}^{(B,i,k)}(g_1,g_2)-c_1^{(B,i,k)}\alpha_{g_1}-c_2^{(B,i,k)}\alpha_{g_2}}{a_1^{(B,A,i,k)}\alpha_{g_1}+a_2^{(B,A,i,k)}\alpha_{g_2}+b^{(B,A,i,k)}}
\end{aligned}$$



$\lambda = 0.25$

$\alpha = (0.1, 0.1, 0.4, 0.4)$

father = nearest tree

15 x 15 couples

Parametric estimation

$$P((g_1^{(A,i,k)}, g_2^{(A,i;k)}) \mid C_{A,B,i}) = P_{2,i,k} h_\theta(d(A_i, B_i)) + P_{2,i,k}(1 - h_\theta(d(A_i, B_i)))$$

B is not the father

$$\begin{aligned} P_{1,i,k} = \quad & (\tfrac{1}{2})^L \prod_{l \le L} \Big\{ \alpha_{g_{2,l}^{(A,i,k)}} \left(1_{g_{1,l}^{(A,i,k)} = m_{1,l}^{(A,i)}} + 1_{g_{1,l}^{(A,i,k)} = m_{2,l}^{(A,i)}} \right) \\ & + \alpha_{g_{1,l}^{(A,i,k)}} \left(1_{g_{2,l}^{(A,i,k)} = m_{1,l}^{(A,i)}} + 1_{g_{2,l}^{(A,i,k)} = m_{2,l}^{(A,i)}} \right) \Big\} \end{aligned}$$

B is the father

$$P_{2,i,k} = (\tfrac{1}{2})^L \prod_{l \le L}$$

$$\Big\{ \left(1_{g_{1,l}^{(A,i,k)} = m_{1,l}^{(A,i)}} + 1_{g_{1,l}^{(A,i,k)} = m_{2,l}^{(A,i)}} \right) \left(1_{g_{2,l}^{(A,i,k)} = m_{1,l}^{(B,i)}} + 1_{g_{2,l}^{(A,i,k)} = m_{2,l}^{(B,i)}} \right)$$

$$+ \left(1_{g_{1,l}^{(A,i,k)} = m_{1,l}^{(B,i)}} + 1_{g_{1,l}^{(A,i,k)} = m_{2,l}^{(B,i)}} \right) \left(1_{g_{2,l}^{(A,i,k)} = m_{1,l}^{(A,i)}} + 1_{g_{2,l}^{(A,i,k)} = m_{2,l}^{(A,i)}} \right) \Big\}$$

Conclusion

- parametric and non-parametric estimations can be performed
- no need to use second order statistics
- interest in confronting this estimation with one based on second order stat?
- convergence... strong mixing
- confidence intervals block-bootstrap

self-pollination

$i(x) = P(\text{the tree at 0 is the father of its flower} \mid x \neq 0)$
$\epsilon(x) = P(\text{the father of the flower at } 0 \notin \{0, x\})$

$1 = h(x) + i(x) + \epsilon(x)$
$P((g_1, g_2) \mid C) = ah(h) + bi(x) + c$

- $h(x)$ and $i(x)$ estimated in the same way
- self-pollination estimated as $1 - \lambda \int_x h(x)r(x)dx$
- $i(x)$ can be used to choose the self-pollination model

genetic dependance of trees

- no self-pollination
- $\lambda r_x(y)$ intensity of the Palm measure with respect to 0 and x.
- $h_x(y)$ the probability that a tree a y pollinates a given seed at 0 knowing that trees are present at 0 ant x,
- $P_x((l_1(y), l_2(y)) \mid C)$ the probability that the tree at y carries genotype $(l_1(y), l_2(y))$ knowing the genotypes of the trees at 0 and x.

$$P(g_1 \mid C \cap F) = \quad \tfrac{h(x)}{2}(1_{\{g_1=k_1\}} + 1_{\{g_1=k_2\}})$$
$$+\lambda \textstyle\int_y r_x(y) h_x(y) \sum_{l_1,l_2} P_x((l_1(y), l_2(y)) = (l_1, l_2) \mid C) \frac{1}{2}(1\{g_1 = l_1\} + 1_{\{g_2=l_2\}})dy$$

Conclusion(1)

- estimate effectively $f(x)$
- no need to impose Poisson assumption
- possible to estimate furthermore with self-pollination, spatial genetic dependance(?)
- focus on $h(x)$ instead of $f(x)$

- link between $i(x)$ (self-pollination knowing a tree at x) and self-pollination ?
- estimate $h_x(y)$ (opening toward classical forward dispersion function) knowing $h(x)$?

Conclusion(2)

- use of Campbell theorems

- $0, A \in X$

- $1 = P(\text{flower at 0 pollinated})$

  $= 1_{\{A \text{ pollinates}\}} + \sum_{x \in X \setminus \{0,A\}} 1_{\{x \text{ pollinates}\}}$

- Austerlitz and al focus on the second term and forget the first term

- we focus on the first term and consider the second one as noise

  ==> How to combine ?