植物個体群へ応用できる繰り返し可能な非定常クラスター点過程

島谷健一郎・統計数理研究所

植物集団には集中分布を示すものが多い。集中分布を表現するモデルの基本のひとつに、ネイマン・スコット過程がある。そのひとつである Thomas 過程では、

1．　母親はランダムに分布する（定常ポアソン過程に従う）。
2．　各母親はポアソン分布に従う個数の娘を生産する。
3．　娘は母親から正規分布に従って近傍に散布される。
4．　母親は死んで、娘たちが集中分布を示す。

というアルゴリズムで点分布が生成される（図 1）。

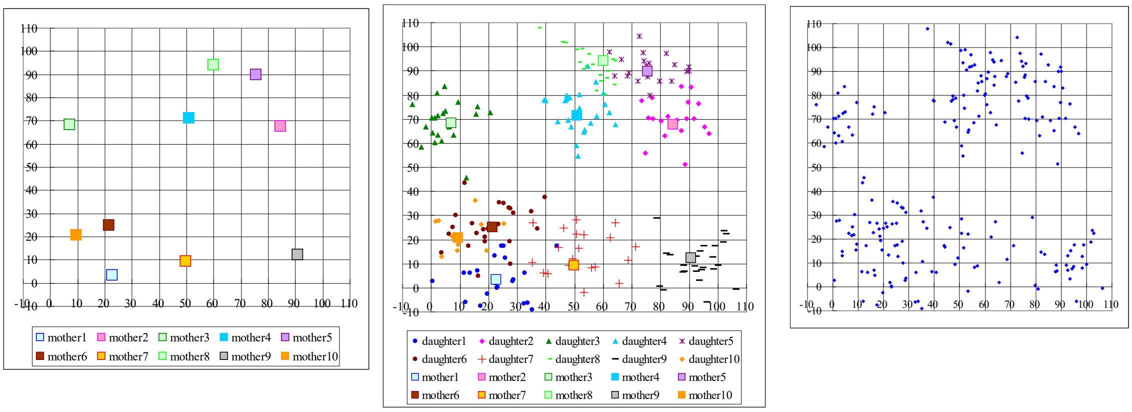


図 1. Thomas 過程のアルゴリズム

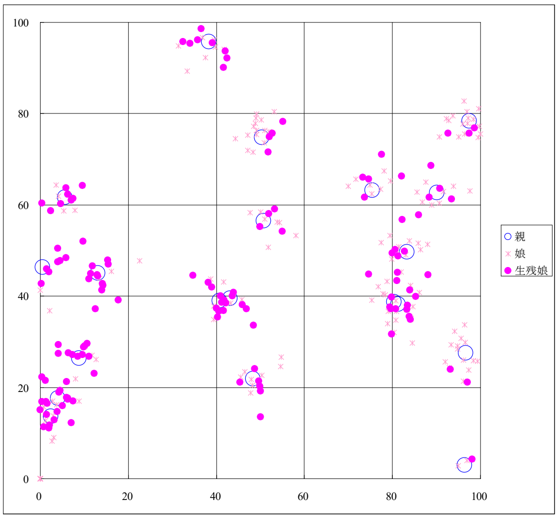
実際の植物個体群に適用するには、いくつか問題がある。

1．　ネイマン・スコット過程は定常性を仮定する。しかし、現実には生残率は環境に依存する。すなわち、非定常な点過程モデルにする必要がある。
2．　植物は繁殖を繰り返す。だから、親集団も集中分布だったと仮定するほうが自然である。その親集団も環境に依存した非定常な分布と考えるほうがさらに自然である。要するに、繰り返し可能な非定常点過程である必要がある。
3．　多くの種において種子散布は正規分布より裾野が重い分布を示す。また、親のごく近傍に娘は定着しない場合が多い。

実際のところ、非定常で集中分布を生成するアルゴリズムは容易に作ることができる。しかし、1 次モーメント、2 次モーメントなどを数学として導出しないと、データからの未知パラメータ推定など、必要な統計的推定を行えない。な

お、3 については、混合正規分布を用いて Thomas 過程を拡張することで解決できる。

非定常性を加えたネイマン・スコット過程は 2000 年代に入り、数多く提唱されており、未知パラメータの推定法も数多く試されている。しかし、その多くが娘集団への非定常性しか考慮していない（図 2 左）。そして、2 次モーメントを初等関数で書き下せないと、何らかの代替法でパラメータ推定を行うが、計算量が膨大になりがちである。



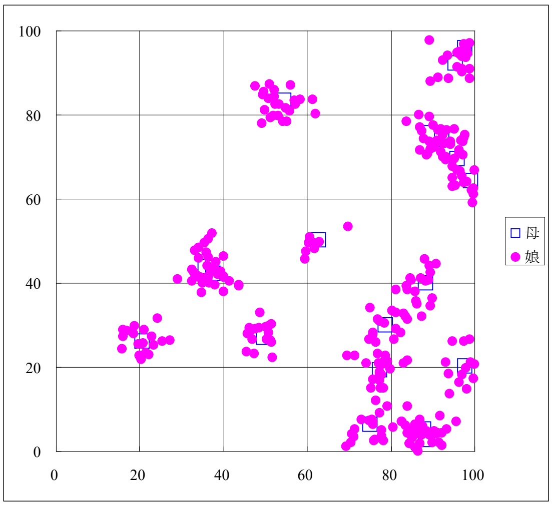


図 2 左：娘にのみ非定常な生残が働く。右：親集団が集中分布を示す.

ここでは、集中分布する非定常な親集団から集中分布する非定常な娘集団を生成する点過程モデルで、2 次モーメントまで初等関数で書き表せるものを提唱する。

娘集団の 1 次モーメント $d_D^{(1)}(\mathbf{x})$、2 次モーメント $d_D^{(2)}(\mathbf{x}, \mathbf{y})$は、母親集団の 1 次モーメント $d_M^{(1)}(\mathbf{x})$、2 次モーメント $d_M^{(2)}(\mathbf{x}, \mathbf{y})$を用いて、以下の漸化式で表される。

$$d_D^{(1)}(x) = \int_A d_M^{(1)}(z)\mathbf{E}(u)Nl(x;z,s^2)dz \cdot s(x)$$

$$d_D^{(2)}(x,y) = \int_A \int_A d_M^{(2)}(z,w)\mathbf{E}(u(u-1))Nl(x;z,s^2)Nl(y;w,s^2)dzdw \cdot s(x)s(y)$$

ここで、$Nl(x;m,\sigma^2)$ は正規分布の確率密度関数で母親から娘への散布に対応する。u は娘数という確率変数、$s(\mathbf{x})$は $\mathbf{x}$ での生残率で、多くの場合、m 個の環境条件 $l_i(\mathbf{x})$の回帰式で表される。

$$s(\mathbf{x}; \boldsymbol{\theta}) = \exp(\theta_0 + \theta_1 l_1(\mathbf{x}) + \cdots + \theta_m l_m(\mathbf{x}))/\mathrm{C}$$

なお、この定式化では、$\mathbf{E}(u)$, $\mathbf{E}(u(u\text{-}1))$と $s(\mathbf{x})$は、定数倍を区別できない。
環境データはサンプリング地点で計測し、カーネル平滑化などで任意の点 $\mathbf{x}$ における環境条件として用いることが多い。母親集団の 1 次モーメント、2 次モーメントが非定常な生残関数 $s(\mathbf{x})$を含む形なので、それに正規分布を乗じた漸化式の積分は当然、初等関数では表せない。
ところが、平滑化を回帰式の後で行うという近似を行うと、生残率は混合正規分布の形で表わされ、上の漸化式は正規分布の convolution の公式で explicit に解けてしまう。すると、2 次モーメントを explicit に書けるため、composite likelihood などによるパラメータ推定が可能となる。ここでは Tanaka-Ogata の Palm likelihood 法を適用するための数式を導出し、実践例を紹介する。