Study the galaxy distribution characterisation via Bayesian statistical learning of spatial marked point processes

Nathan Gillot, Radu S. Stoica, Didier Gemmerlé

Study the galaxy distribution characterisation via Bayesian statistical learning of spatial marked point processes

Nathan Gillot[1], Radu S. Stoica[1], Didier Gemmerlé[2]

[1] Université de Lorraine, CNRS, IECL, Inria, F-54000 Nancy, France

[2] Université de Lorraine, CNRS, IECL, F-54000 Nancy, France

{nathan.gillot,radu-stefan.stoica,didier.gemmerle}@univ-lorraine.fr

Context

Galaxies are not uniformly distributed in the observable Universe. Their positions induce structures such as filaments, void zones or even clusters of galaxies. The complexity of these structures and the amount of data available on the subject led to the idea of a probabilistic approach to explain the characteristics of these structures, based on point process models ([7] , [5]). An important part of this probabilistic framework is to use algorithms able to estimate the parameters of the models proposed to fit the observed data such as Approximate Bayesian Computation (ABC) algorithms ([3, 8, 10]).

Point processes models

The points $\mathbf{x}$ are situated in W a compact region of $\mathbb{R}^d$. We assume that the data we observe have the following properties :

• The Universe can be seen as the representation of a stochastic process where galaxies are randomly located points in space.

• Two such points cannot share the same position: for a given point $\xi \in W$, no other point has the same coordinates in W.

Poisson point process: completely random patterns (independence)
The Poisson point process probability density is proportional to

$$f(\mathbf{x}|\rho) \propto \exp\left(w(\mathbf{x}) \log \rho\right) \tag{1}$$

where $w(\mathbf{x}) = \sum_{i=1}^{n(\mathbf{x})} v(x_i)$ is the potential associated to each point in $\mathbf{x}$. If $w(\mathbf{x})$ is a fixed constant, the point process will be called homogeneous.

Strauss point process: repulsive patterns Its probability density is proportional to

$$f(\mathbf{x}|\gamma_s) \propto \exp\left(s_r(\mathbf{x}) \log(\gamma_s)\right) \tag{2}$$

where $s_r(\mathbf{x})$ represent the number of pairs of points closer than the distance r, $\gamma_s \in]0, 1]$ the model parameter.

Area interaction process: repulsive or clustered patterns Its probability density is proportional to

$$f(\mathbf{x}|\gamma_a) \propto \exp\left(a_R(\mathbf{x}) \log(\gamma_a)\right) \tag{3}$$

where $a_R(\mathbf{x}) = -|\cup_{\xi \in \mathbf{x}} b(\xi, R)|$ represent the d-volume (area if $d = 2$) of the union of balls of radius R attached to the points, $\gamma_a \geq 0$ is the model parameter.

Parametric inference : posterior sampling

Problem: sampling the posterior distribution to estimate model parameters is difficult → normalising constant need to be evaluated.

Adopted solution: ABC Shadow algorithm [8, 10].

Key ideas: approximate the behaviour of Markov chain that has the equilibrium distribution the posterior of interest : the outputs are approximate samples from the posterior.

Algorithm description: fix δ a perturbation parameter, m number of iterations and θ_0 an initial condition. Assume the observed pattern is $\mathbf{x}$ and the current state of the parameters is θ_0.

1. With the Metropolis Hastings algorithm, generate the auxiliary pattern $\mathbf{y}$ according to $f(\mathbf{y}|\boldsymbol{\theta_0})$
2. For $k = 1$ to m :
   . Propose a new parameter ψ according to the density $U_\delta(\theta_{k-1} \to \psi)$ defined by $U_\delta(\theta \to \psi) = \frac{1}{|b(\theta, \delta/2)|} \mathbf{1}_{b(\theta,\delta/2}\{\psi\}$.
   . The new state $\theta_k = \psi$ is accepted with probability $\alpha_s(\theta_{k-1} \to \psi) = \min\{1, \frac{f(\mathbf{x}|\theta_k)p(\theta_k)}{f(\mathbf{x}|\theta_{k-1})p(\theta_{k-1})} \times \frac{f(\mathbf{y}|\theta_{k-1})}{f(\mathbf{y}|\theta_k)}\}$ otherwise $\theta_k = \theta_{k-1}$.
3. Return θ_m.
4. If more samples are needed, go to step 1 and set $\theta_0 = \theta_m$

Point pattern simulation

Problem: the normalising constant of the previous models is not available in analytical closed form.

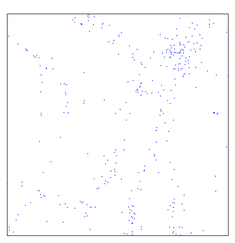
Solution : use MCMC methods which consists in simulating a Markov chain whose unique equilibrium distribution is the distribution of the point process of interest.

Algorithms: spatial birth-and-death processes, Metropolis-Hastings dynamics.

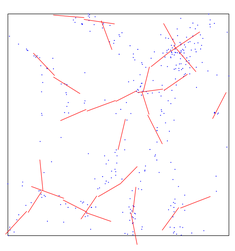
Key ideas: add or remove a point from the current configuration till equilibrium is reached. The construction of the acceptance probability for the proposed transition guarantees convergence properties of the simulation algorithms [4, 7]

Data and Modelling

Cosmological simulation used to set up the first filaments pattern detector based on marked point process [9].



Galaxies pattern



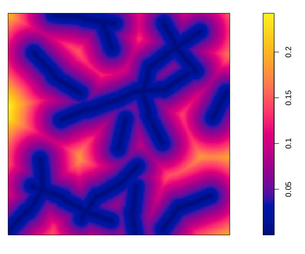
Corresponding detected filaments

The **ABC Shadow** algorithm was used to fit a superposition of models with the following components :
• Poisson component : in-homogeneity that takes into account $d(\xi, F)$, the shortest distance from a point $\xi \in W$ to the the given filament network. This distance is presented in the Figure below. The sufficient statistic attached to this component is: $w(\mathbf{x}) = \sum_{i=1}^{n(\mathbf{x})} \mathbf{1}_{d(\xi_i,F)\leq 0.05}(\xi) \times \frac{1}{1+d(\xi_i,F)}$.
• Strauss component : the same as the interaction part in (2)
• Area-Interaction component : the same as the interaction part in (3)

Observing the galaxy pattern $\mathbf{x}$, the posterior distribution is

$$p(\log\rho, \log\gamma_s, \log\gamma_a|\mathbf{x}) \propto \exp\left(w(\mathbf{x})\log\rho + s_r(\mathbf{x})\log(\gamma_s) + a_R(\mathbf{x})\log(\gamma_a)\right) p(\rho, \gamma_s, \gamma_a) \tag{4}$$

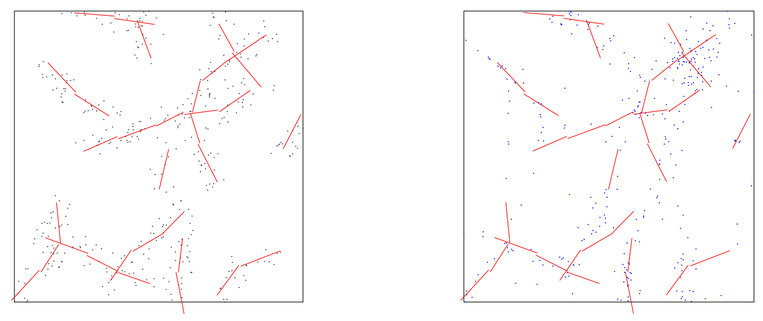
with $p(\log\rho, \log\gamma_s, \log\gamma_a)$ the prior knowledge regarding the model parameters.



Shortest distance between any point in the domain to the given filament network

Results

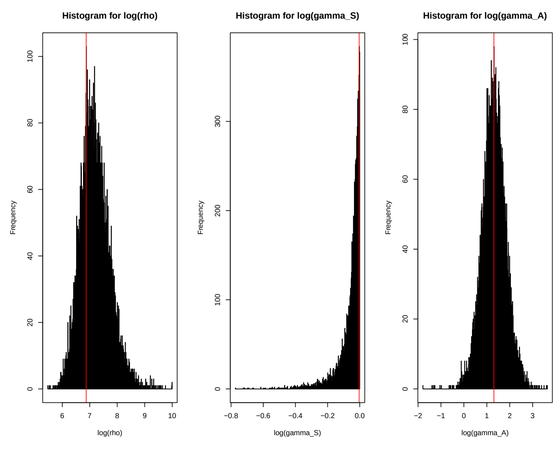
• For each radius tuple (r_S, r_A) among $(0.01, 0.01)$; $(0.01, 0.03)$; $(0.01, 0.05)$; $(0.03, 0.01)$; $(0.05, 0.01)$, the ABC Shadow algorithm was initialised with the observed pattern's sufficient statistics.
• Prior density : uniform distribution on the interval $[0, 10] \times [-10, 0] \times [-10, 10]$. At every step, the auxiliary variable was sampled with 250 iterations of the Metropolis-Hastings algorithm.
• δ was set to $(0.01, 0.01, 0.01)$, m to 100 and θ_0 was set randomly inside the prior density interval. This procedure was run 10^4 times, giving us a sample of size 10^4 of the estimated parameters.



Simulated galaxies distribution using the estimated parameters (left) and Observed galaxies distribution (right)

Below, the table summarises the parameter estimation for the different fixed radius with their asymptotic standard errors ([1, 2, 6]) and an illustration of the outputs of the algorithm, giving the posterior approximation used for the parameter with $(r_S, r_A) = (0.01, 0.03)$.

Radius	Estimates of $\log(\rho)$, $\log(\gamma_S)$ and $\log(\gamma_A)$		
(r_S, r_A)	$\log(\rho)$	$\log(\gamma_S)$	$\log(\gamma_A)$
$(0.01, 0.01)$	9.04 ± 0.24	-0.52 ± 0.16	2.55 ± 0.28
$(0.01, 0.03)$	7.19 ± 0.08	-0.05 ± 0.12	1.31 ± 0.32
$(0.01, 0.05)$	6.83 ± 0.09	-0.03 ± 0.17	-1.57 ± 0.93
$(0.03, 0.01)$	8.36 ± 0.20	-0.02 ± 0.03	1.84 ± 0.21
$(0.05, 0.01)$	8.33 ± 0.21	-0.009 ± 0.02	1.8 ± 0.20



Bibliography

1. C. J. Geyer, *Chapter 1 Likelihood Inference for Spatial Point Processes*, 1999.
2. C. J. Geyer, *Journal of the Royal Statistical Society. Series B (Methodological)* **56**, 261–274, ISSN: 00359246 (1994).
3. L. Hurtado-Gil, R. S. Stoica, V. J. Martínez, P. Arnalte-Mur, *Monthly Notices of the Royal Astronomical Society* **507**, 1710–1722 (2021).
4. M. N. M. van Lieshout, *Markov Point Processes and their Applications* (Imperial College Press, London, 2000).
5. M. N. M. van Lieshout, *Theory of Spatial Statistics : A concise Introduction* (Chapman & Hall, 2019).
6. M. N. M. van Lieshout, R. S. Stoica, *Statistica Neerlandica* **57**, 177–206 (2003).
7. J. Møller, R. P. Waagepetersen, *Statistical Inference and Simulation for Spatial Point Processes* (Chapman & Hall/CRC, 2004).
8. R. S. Stoica, M. Deaconu, A. Philippe, L. Hurtado-Gil, *Spatial Statistics* **43**, 1–21 (2021).
9. R. S. Stoica, V. J. Martinez, J. Mateu, E. Saar, *Astronomy and Astrophysics - A&A* **434**, 423–432 (2005).
10. R. S. Stoica, A. Philippe, P. Gregori, J. Mateu, *Statistics and Computing* **27(5)**, 1225–1238 (2017).