Offre n°2022-04648

PhD Position F/M Models and algorithms for statistical learning of spatio-temporal marked point processes. Application: cosmological data analysis and characterization.

Le descriptif de l'offre ci-dessous est en Anglais

Type de contrat : CDD

Niveau de diplôme exigé : Bac + 5 ou équivalent

Autre diplôme apprécié : Master or Engineering degree

Fonction : Doctorant

Contexte et atouts du poste

The thesis will take place at the Institut Elie Cartan de Lorraine (IECL), the research mathematical institute of the University of Lorraine, within the Probability and Statistics team. The candidate will be also a member of the PASTA Inria team. The thesis advisor, Radu Stoica is a member of both these structures.

The thesis will take place in the IAEM Doctoral School (Computer Science, Automation, Electronics-Electrotechnics, Mathematics, ED 77) of the University of Lorraine, in Nancy.

The student will benefit from the IECL's computer equipment, an in-house mathematics library and access to scientific publications via subscriptions from the IECL, the University of Lorraine and Inria. The IECL also has a baby-cluster that can be used for the computational needs of the thesis.

This thesis has a strong multidisciplinary character. Close interactions with cosmologists from IAS (France) but also from other european countries (Estonia, Finland, Spain, The Netherlands) are expected and strongly encouraged.

Mission confiée

Galaxies are not distributed uniformly in our Universe. Their positions exhibit different structures such as filaments, clusters and sheets. These objects form an intricate network enclosing immense void regions [5]. Mapping these structures is maybe one of the most challenging problems of the beginning of this century. The complexity of these patterns and the huge quantity of available data recommend probabilistic approaches as guiding lines whenever tackling this problem. During the last decade, significant research was developed following this direction [6, 8, 11, 10, 2]. The philosophy at the basis of this work is to consider these patterns as an entity made of complex objects that interact, while driven by a probabilistic model. Conditionally to the observed galaxies, the hidden pattern is detected as the configuration maximizing the law governing the considered model.

Once these structures are mapped, important characteristics of our Universe can be derived, by fitting a probabilistic model to the galaxy positions. Recent algorithms proposed by [9, 7] allow the derivation of law of the parameters driving the considered models, that is the posterior distribution. These algorithms were already tested by fitting to the observed cosmological data different types of models, the area-interaction, the connected components and the Geyer point processes [4].

1.1 Complete data
Let us consider that the a given region of the Universe is entirely observed, that is all the galaxies positions within the considered region are known. Clearly, the filament network can be detected and hence, conditionally on this pattern the following challenges are to be integrated within the following program:

· propose new mixture models composed by the superposition of several point processes that take into account a priori cosmological information regarding the galaxies distribution: clustering, repulsion, distance to the filaments pattern
· derive the significance of each component of the previous model
· characterise and validate the spatial distribution of galaxies belonging to particular clusters

1.2 Incomplete data
There exists regions in our universe where galaxies are not observed. Since galaxies are not observed, patterns of filaments cannot be detected or if they are detected they suffer from large positional

uncertainties. Hence, the following research program proposes to model galaxy distributions, while overcoming this drawback. The idea is to work in two steps. First, generate filament patterns within the un-observed regions. Second, use the simulated filament networks and the previously studied models in order to in-paint with galaxies the unobserved regions of our Universe. The first step can be achieved following the program:

· detect filaments in the observed regions
· estimate the filament pattern parameters in the domain given by the observed and un-observed regions
· check and validate the obtained results
· simulate filaments patterns within the un-observed regions condition- ally on the detected filaments in the observed regions

The parameter estimation of the filament pattern model can be achieved via EM algorithm. The new sampling algorithms [9, 7] adapted to the EM inference framework should improve the convergence properties of the newresulting method with respect to the classical ones. The check and validate steps are to be achieved using an adapted residual based strategies [1].

This step may be further developed by considering the estimation of the density field controlling the galaxy distribution in hidden regions with generative adversarial neural networks [12]. Within the same scope recent statistical developments should be also considered [3].

The second step builds galaxy models within the observed and the un-observed Universe regions following the points:

· consider one simulated filament pattern; if several simulations are to be considered the whole procedure can be iterated in order to obtain robust statistics
· conditionally on this pattern, the observed galaxies consider the models derived from the complete data case in order to estimate their parameters within the observed and un-observed regions
· check and validate the obtained results
· conditionally on the observed galaxies, simulate new galaxies catalogues in the unobserved regions in order to produce cosmological inference

The parameter estimation and model choice validation are to be done as performed previously. Here, the final result is to be validated using cosmological criteria.

2. Scientific direction

Radu S. Stoica
Full Professor
Inria/IECL - Université de Lorraine
mail: radu.stoica@inria.fr
web: https://sites.google.com/site/radustefanstoica/

2.1 Cross disciplinary collaboration

The project aim is to develop new mathematical modelling and inference procedures to be applied to cosmological data analysis and characterisation. The current proposal is built in close cooperation with cosmologists from IAS (Institut d'Astrophysique Spatiale - Paris Saclay) (Nabila Aghanim and Jenny Sorce) and Complutense University of Madrid (Aurélien Decelle). Interactions with cosmologists from Tartu Observatory (Estonia) and Valencia Observatory (Spain) are also considered. These collaborations will help with appropriate data and with expertise for the analysis of the obtained results.

References:

[1] A. J. Baddeley, E. Rubak, and R. Turner. Spatial Point Patterns: Methodology and Applications with R. Chapman and Hall/CRC Press, London, 2016.

[2] T. Bonnaire, N. Aghanim, A. Decelle, and M. Douspis. T-ReX: a graph-based filament detection method. Astronomy and Astrophysics, 637(A18):1–15, 2020.

[3] E. Gabriel, F. J. Rodrı guez-Cortes, J. Coville, J. Mateu, and J. Chadoeuf. Mapping the intensity function of a non-stationary point process in unobserved areas. ArXiv Preprint, arXiv:2111.14403, 2021.

[4] L. Hurtado-Gil, R. S. Stoica, V. J. Martı nez, and P. Arnalte-Mur. Morpho-statistical characterisation of the spatial galaxy distri-
bution through gibbs point processes. Monthly Notices of the Royal Astronomical Society, 507(2):1710–1722, 2021.

[5] V. J. Martinez and E. Saar. Statistics of the galaxy distribution. Chapman and Hall, 2002.

[6] R. S. Stoica. Marked point processes for statistical and morphological analysis of astronomical data. The European Physical Journal Special Topics, 186:123–165, 2010.

[7] R. S. Stoica, M. Deaconu, A. Philippe, and L. Hurtado. Shadow Simulated Annealing: a new algorithm for approximate Bayesian inference of Gibbs point processes. Spatial Statistics, 43, 2021.

[8] R. S. Stoica, V. J. Martinez, and E. Saar. Filaments in observed and mock galaxy catalogues. Astronomy and Astrophysics, 510(A38):1–12, 2010.

[9] R. S. Stoica, A. Philippe, P. Gregori, and J. Mateu. Abc shadow algorithm: a tool for statistical analysis of spatial patterns. Statistics and Computing, 27:1225–1238, 2017.

[10] E. Tempel, M. Kruuse, R. Kipper, T. Tuvikene, J. G. Sorce, and R. S. Stoica. Bayesian group finder based on marked point processes. method and application to the 2mrs data set. Astronomy and Astrophysics, 618:1–18, 2018.

[11] E. Tempel, R. S. Stoica, E. Saar, V. J. Martinez, L. J. Liivamägi, and G. Castellan. Detecting filamentary pattern in the cosmic web: a catalogue of filaments for the SDSS. Monthly Notices of the Royal Astronomical Society, 438(4):3465–3482, 2014.

[12] M. Ullmo, A. Decelle, and N. Aghanim. Encoding large-scale cosmo- logical structure with generative adversarial networks. Astronomy and Astrophysics, 651(A46):1–15, 2021.

Principales activités

The sucessful candidate will dedicate to accomplish the research program described before.

For this purpose, the main activities are:

- bibliography study
- build and program new models and algorithms for the problem on-hand
- write scientific papers to communicate the obtained results
- collaborate with the mathematicians and the cosmologists involved in the project
- participate to scientific manifestations to present the obtained results

Compétences

Technical skills and level required : applied mathematics (probability and statistics), programming (C++, R, Matlab/Scilab)

Languages : English (compulsory), French (optional)

Relational skills : sense of humour, creativity

Other valued appreciated : general culture

Avantages

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Rémunération

Salary: 1982€ gross/month for 1st and 2nd year. 2085€ gross/month for 3rd year.

Monthly salary after taxes : around 1596,05€ for 1st and 2nd year. 1678,99€ for 3rd year.

Informations générales

- **Thème/Domaine :** Approches stochastiques Statistiques (Big data) (BAP E)
- **Ville :** Villers lès Nancy
- **Centre Inria :** Centre Inria de l'Université de Lorraine
- **Date de prise de fonction souhaitée :** 2022-10-03
- **Durée de contrat :** 3 ans
- **Date limite pour postuler :** 2022-05-31

Contacts

- **Équipe Inria :** PASTA
- **Directeur de thèse :** Stoica Radu / radu.stoica@loria.fr

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires

académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 200 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

L'essentiel pour réussir

The ideal candidate should possess excellent skills in applied mathematics, especially in probability and statistics, while being highly motivated by practical applications. The rigorous knowledge of an object oriented programming language such C++ is also required. Being familiar with mathematical software such as Matlab, Scilab or R, it is also recommended.

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

Consignes pour postuler

Sécurité défense :
Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :
Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.