Linköping, December 11, 2025

PhD student in Statistics
(Ref IDA–2025–00412)

Krzysztof Bartoszek
Division of Statistics and Machine Learning
Department of Computer and Information Science
Linköping University, Linköping, Sweden
Contact: krzysztof.bartoszek@liu.se

The Division of Statistics and Machine Learning (STIMA) at Linköping University is expanding. In this call we are looking for a new PhD student who wants to join us in our effort of strengthening our research in the direction of developing and mathematically studying statistical inference methods for stochastic processes (in particular continuous–time–continuous–space) in phylogenetics. This document briefly presents the research at the division and the background of the PI (Section 1). This is followed by a few concrete suggestions for research topics for the PhD position (Section 2). Some administrative details are:

- **Application deadline:** March 3, 2026
- **Application procedure:** All applications are to be submitted via Linköping University's online application system, available via this link:
  `https://liu.se/en/work-at-liu/vacancies/28230` (English)
  `https://liu.se/jobba-pa-liu/lediga-jobb/28210` (Swedish)
- **Questions:** If you have any questions do not hesitate to contact Krzysztof via e–mail.

1 The Division of Statistics and Machine Learning

The position is formally based at the Division of Statistics and Machine Learning (STIMA) within the Department of Computer and Information Science. At STIMA we conduct research and education in both statistics and machine learning, at the undergraduate, advanced and PhD levels. We regularly publish solid contributions to top statistical, probabilistic and applied journals and at the best machine learning conferences. STIMA is characterized by a modern view of the statistical subject, where probabilistic models are combined with computational algorithms to solve challenging complex problems, as well as a statistical view of machine learning which clearly integrates the two subject areas within the division. We use our expertise to advance applied fields, e.g., evolutionary biology, medical image analysis, bioinformatics or meteorology. For more information about STIMA, please see `https://liu.se/en/organisation/liu/ida/stima`

Krzysztof Bartoszek, who will be the main supervisor for the PhD student recruited in this call, is a biträdande professor (Senior Associate Professor) in Statistics at STIMA. His research background is in stochastic processes with a special focus on phylogenetic comparative methods and other applications of probabilistic modelling in life science problems. In particular, his research has focused on probabilistic models and the development of computational algorithms for learning and reasoning about these models. Using probability theory, probabilistic models are able to systematically represent and cope with the uncertainty that is inherent to most data. This is of central importance in many applications of machine learning.

The group at STIMA has a wide network of strong international collaborators all around the world. The intended project work will give the student direct possibilities of working with scientists from the Heinrich Heine University in Düsseldorf, Florida International University, Łódź University, University of the Balearic Islands, University of Gdańsk, University of Oslo, and University of Toruń. Other members of the Division have joint projects with researchers for example at Stanford University. As for all our PhD students, we strive also for you to get a solid international experience during their PhD studies.

2 Potential Research Topics

The research projects for the advertised position will be in the area of "Advancing phylogenetic comparative methods—limit results, constrained evolution, tests of appropriateness, confidence intervals and fast inference techniques ", Krzysztof Bartoszek's Swedish Research Council Project Grant that partially funds the PhD position. A few examples of potential research topics are briefly outlined below. In brief the high–level description is that one has a tree structure and top of this tree, i.e., along the tree's branches some stochastic processes evolves. The values of the process are observed only at the tips of the tree and based on such a sample (and the tree) one is to make inference about the data generating stochastic process. The approach has as its goal to model the evolution of traits, e.g. body size, on a between–species level. In the scope of the project the focus will be on methods for modelling evolution of both univariate and multivariate interacting traits.

As an applicant you are not required to specify a specific research topic in your application, but you are of course welcome to do so if you want. The topics below are provided mainly to make the advertised positions more concrete. We do welcome your own initiatives and the precise research topic of each PhD student will be decided in a dialog between the student and the supervisor after a successful appointment. Depending on the interests of the student the topics can be approached from either a mathematical perspective, a software development/computational one or a mixture of the two. Apart from the main methods development work the PhD student will have the opportunity to interact with biological researchers from various backgrounds and institutions, testing the methods and contributing to the analyses of real data.

Markov processes on trees: The vast majority of modelling and inference approaches for phylogenetic comparative methods require an explicit form of the transition density along a given length of time and also the ability to integrate out analytically the unobserved internal node stats. This allows for effective estimation schemes but severely restricts the family of permissible models of evolution. From a practical point of view, only Gaussian transitions seem possible today.

A potential PhD project is to change the modelling approach and to focus on a different description of a Markov process. A process defined by a stochastic differential equation (SDE) can be described through its generator. Solving the associated partial differential equation on top of a tree will allow one to obtain the likelihood and hence do parameter inference for all SDE models of evolution. In particular one will be able to approach in a computationally effective way models of bounded evolution. These can be used for working with traits that are constrained. Every trait has some natural upper and lower limit, but it is also needed to work with compositional data or proportions. In relation to this direction the student will have opportunity to work with the evolution of genes' positions in bacterial species. This is a naturally bounded trait—as the genes position is a particular fraction of the genome away from the point of origin of the bacterial chromosome.

Evolution with interactions: Development of models with after speciation, between lineage interactions (e.g., hybridization or migration) is only now starting to take place due to recent computational developments. A possible PhD project is to study the mathematical properties of and develop inference software for branching Markov processes with interactions. Another possible interaction between species, are so–called co–phylogeny systems. These are most evident in host–parasite systems, where the hosts have one phylogeny and the parasites another. However, the traits of the hosts and parasites are at a constant arms race between each other and hence interacting. This project can be done in close collaboration with biological groups who have well curated data sets and clear hypotheses between which species did mixing events take place, for example figs and wasps.

DANE

Properties of branching stochastic processes: For a theoretically minded student there is the possibility to explore the mathematical properties of the above described branching stochastic processes. Possible problems include central limit theorems, uncertainty of estimates, law of the iterated logarithm, the maximum processes, or properties of models for evolution with jumps. One can couple this with branching process models for the tree, including trait dependent speciation. Despite being formulated in an abstract way such results are of utmost importance for statistical estimation method development. They provide theoretically justified initial states for iterative inference methods and indicate what is estimable, i.e., what are our possibilities and limitations. Furthermore, they can lead to improvements in inference software development and simulation methods. Development of computational approaches for the described framework can be an important component of the thesis work.