ABSTRACTS

11th Seminar D^2 Seminar Series

Florence Center for Data Science 'Double' Seminar Series

Fabio Corradi - Department of Statistics, Computer Science, Applications "G. Parenti", University of Florence

Title: *Learning the two parameters of the Poisson-Dirichlet distribution with a forensic application*

Abstract: This contribution is motivated by the rare type match problem, a relevant forensic issue. There, difficulties arise to evaluate the likelihood ratio comparing the defense and the prosecution hypotheses since the specific matching characteristic from the suspect and the crime scene is not in the reference database. A recently proposed solution approximates the likelihood ratio by plugging in the parameters MLE of a Poisson Dirichlet distribution, a Bayesian nonparametric prior modeling probability mass function showing a power-law behavior in the infinite-dimensional space. We instead consider how to learn the parameters of a Posson-Dirichlet and we propose two sampling schemes: Monte Carlo Markov Chain and Approximate Bayesian Computation. We demonstrate that the previously employed plug-in solution produces a systematic bias that Bayesian inference avoids entirely. Finally, we evaluate the method using a database of Y-chromosome haplotypes.

Michela Baccini - Department of Statistics, Computer Science, Applications "G. Parenti", University of Florence

Title: *Combining and comparing regional epidemic dynamics in Italy: Bayesian meta-analysis of compartmental models and model assessment via Global Sensitivity Analysis*

Abstract: During autumn 2020, Italy faced a second important SARS-CoV-2 epidemic wave. We explored the time pattern of the instantaneous reproductive number R0(t), and estimated the prevalence of infections by region from August to December calibrating SIRD models on COVID19-related deaths, fixing at values from literature Infection Fatality Rate (IFR) and infection duration. A Global Sensitivity Analysis (GSA) was performed on the regional SIRD models. Then, we used Bayesian meta-analysis and meta-regression to combine and compare the regional results and investigate their heterogeneity. The meta-analytic R0(t) curves were similar in the Northern and Central regions, while a less peaked curve was estimated for the South. The maximum R0(t) ranged from 2.61 (North) to 2.15 (South) with an increase following school reopening and a decline at the end of October. Average temperature, urbanization, characteristics of family medicine and health care system, economic dynamism, and use of public transport could partly explain the regional heterogeneity. The GSA indicated the robustness of the regional R0(t) curves to different assumptions on IFR. The infectious period turned out to have a key role in determining the model results, but without compromising between-region comparisons.