How much do galaxies weigh{new Bayesian state space modelling of missing data

Dalia Chakrabarty
Department of Mathematics, University of Leicester, Leicester, U.K. and
Department of Statistics, University of Warwick, Coventry, U.K.

The pursuit of the density function $\rho(\boldsymbol{X})$ of the mass of all gravitating matter in a real galaxy, given noisy data, on some–of all the phase space coordinates–of individual galactic particles, is a difficult inverse problem, albeit crucially relevant to astronomy. Here $\boldsymbol{X} \in \mathbb{R}^3$ is the particle location vector. In the Bayesian framework, the posterior of the unknown $\rho(\boldsymbol{X})$ given such missing data is sought; however, the relation between the data and the unknown $\rho(\boldsymbol{X})$ is itself not known. It may appear possible to learn this unknown functional relationship by modelling it with a Gaussian Process and training the parametrisation of the covariance structure of this GP using training data. However, for real galaxies, training data is unavailable. In lieu of training data, the first step is to construct such training data that comprises a set of values of the observable, at chosen values of the unknown, $\rho(\boldsymbol{X})$. This is performed in a paradigm in which the support of the sought $\rho(\boldsymbol{X})$ is discretised, reducing the problem to the estimation of a very large number of independent parameters, each of which represents the value of this unknown function over a chosen interval of the corresponding domain variable. We present a method in which the unknown $\rho(\boldsymbol{X})$ is embedded within the definition of the support of the system's state space pdf $f(\boldsymbol{X}, \boldsymbol{V})$ and the likelihood is written in terms of this pdf; here $\boldsymbol{V} \in \mathbb{R}^3$ is the particle velocity vector. Measurement uncertainties are accounted for by convolving the resulting likelihood with the density of the error in measurement. As motivated above, the support of the unknown $\rho(\boldsymbol{X})$ as well as that of the unknown $f(\boldsymbol{X}, \boldsymbol{V})$, are discretised, and values of these functions over chosen intervals of $\boldsymbol{x}$ and $\boldsymbol{v}$ are estimated. The problem of missing data, as manifest in some components of $\boldsymbol{X}$ and $\boldsymbol{V}$ being unobservable, is addressed by projecting the phase space pdf onto the space of the observed variables; this is demonstrated to be more easily achieved if the phase space pdf is considered an isotropic scalar-valued function of $\boldsymbol{X}$ and $\boldsymbol{V}$, than otherwise. Upon invoking suitable priors, the posterior density of these unknown parameters given the data is written and sampled from using adaptive Metropolis-Hastings. Applications to real and simulated data will be presented.

Keywords: Bayesian Inverse Problems; Training Data; State Space Modelling; Galactic Gravitational Mass Density.