Implementation of the EM method for parameter estimation in kappa distributions within the superstatistics framework

Leonardo Herrera[1]

[1]Universidad Andres Bello, Santiago, Chile

In space plasmas, the kappa distribution has been used for several decades to model empirical observations, and more recently its observation under laboratory conditions has been suggested. From a physical standpoint, observed energy distributions exhibit suprathermal tails that the classical Maxwellian fails to reproduce. The family of kappa distributions, which introduces a single additional parameter κ, has proven capable of capturing both the thermal component and the energetic population. However, its use remains largely phenomenological, where we generally identify 3 areas for improvement: (i) the value of the parameter κ is chosen "by hand" or by simple curve fitting to match the observed histogram, (ii) the uncertainty in κ is not propagated rigorously and (iii) each group adopts different criteria for accepting or rejecting the kappa distribution.

From the superstatistics perspective, the kappa distribution emerges naturally as a continuous superposition of Maxwellians weighted by fluctuations of the inverse temperature parameter β. This probabilistic framework allows the parameter estimation problem to be addressed as a Bayesian inference problem with latent variables, opening the possibility of applying modern numerical analysis tools to estimate the value of κ and its uncertainty in the analysis of experimental or simulated data.

In this work we present the implementation of an Expectation–Maximization (EM) algorithm specifically designed to estimate the parameter κ in the context of superstatistics. The procedure alternates an E-step that computes the likelihood of the joint distribution of the observed data and the latent variable, and an M-step that updates the parameter values of the latent distribution by maximizing the complete likelihood, achieving robust convergence even in the presence of noisy or incomplete data.