The Journal of Open Source Software

emcee v3: A Python ensemble sampling toolkit for affine-invariant MCMC

Daniel Foreman-Mackey[1], Will M. Farr[1, 2], Manodeep Sinha[3, 4], Anne M. Archibald[5], David W. Hogg[1, 6], Jeremy S. Sanders[7], Joe Zuntz[8], Peter K. G. Williams[9, 10], Andrew R. J. Nelson[11], Miguel de Val-Borro[12], Tobias Erhardt[13], Ilya Pashchenko[14], and Oriol Abril Pla[15]

1 Center for Computational Astrophysics, Flatiron Institute **2** Department of Physics and Astronomy, Stony Brook University, United States **3** Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Australia **4** ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D) **5** University of Newcastle **6** Center for Cosmology and Particle Physics, Department of Physics, New York University **7** Max Planck Institute for Extraterrestrial Physics **8** Institute for Astronomy, University of Edinburgh, Edinburgh, EH9 3HJ, UK **9** Center for Astrophysics | Harvard & Smithsonian **10** American Astronomical Society **11** Australian Nuclear Science and Technology Organisation, NSW, Australia **12** Planetary Science Institute, 1700 East Fort Lowell Rd., Suite 106, Tucson, AZ 85719, USA **13** Climate and Environmental Physics and Oeschger Center for Climate Change Research, University of Bern, Bern, Switzerland **14** P.N. Lebedev Physical Institute of the Russian Academy of Sciences, Moscow, Russia **15** Universitat Pompeu Fabra, Barcelona

DOI: 10.21105/joss.01864

Software

- Review
- Repository
- Archive

Editor: Juanjo Bazán

Reviewers:

- @benjaminrose
- @mattpitkin

Submitted: 28 October 2019
Published: 17 November 2019

Summary

emcee is a Python library implementing a class of affine-invariant ensemble samplers for Markov chain Monte Carlo (MCMC). This package has been widely applied to probabilistic modeling problems in astrophysics where it was originally published (Foreman-Mackey, Hogg, Lang, & Goodman, 2013), with some applications in other fields. When it was first released in 2012, the interface implemented in emcee was fundamentally different from the MCMC libraries that were popular at the time, such as PyMC, because it was specifically designed to work with "black box" models instead of structured graphical models. This has been a popular interface for applications in astrophysics because it is often non-trivial to implement realistic physics within the modeling frameworks required by other libraries. Since emcee's release, other libraries have been developed with similar interfaces, such as dynesty (Speagle, 2019). The version 3.0 release of emcee is the first major release of the library in about 6 years and it includes a full re-write of the computational backend, several commonly requested features, and a set of new "move" implementations.

This new release includes both small quality of life improvements—like a progress bar using tqdm—and larger features. For example, the new backends interface implements real time serialization of sampling results. By default emcee saves its results in memory (as in the original implementation), but it now also includes a HDFBackend class that serializes the chain to disk using h5py.

The most important new feature included in the version 3.0 release of emcee is the new moves interface. Originally, emcee implemented the affine-invariant "stretch move" proposed by Goodman & Weare (2010), but there are other ensemble proposals that can get better performance for certain applications. emcee now includes implementations of several other ensemble moves and an interface for defining custom proposals. The implemented moves include:

- The "stretch move" proposed by Goodman & Weare (2010),
- The "differential evolution" and "differential evolution snooker update" moves (ter Braak, 2006; ter Braak & Vrugt, 2008), and
- A "kernel density proposal" based on the implementation in the `kombine` library (Farr & Farr, 2015).

`emcee` has been widely used and the original paper has been highly cited, but there have been many contributions from members of the community. This paper is meant to highlight these contributions and provide citation credit to the academic contributors. A full up-to-date list of contributors can always be found on GitHub.

References

Farr, B., & Farr, W. M. (2015). Kombine: A kernel-density-based, embarrassingly parallel ensemble sampler. Retrieved from https://github.com/bfarr/kombine

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2013). emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*, *125*(925), 306. doi:10.1086/670067

Goodman, J., & Weare, J. (2010). Ensemble samplers with affine invariance. *Communications in applied mathematics and computational science*, *5*(1), 65–80. doi:10.2140/camcos.2010.5.65

Speagle, J. S. (2019). dynesty: A Dynamic Nested Sampling Package for Estimating Bayesian Posteriors and Evidences. *arXiv e-prints*, arXiv:1904.02180. Retrieved from http://arxiv.org/abs/1904.02180

ter Braak, C. J. F. (2006). A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing*, *16*(3), 239–249. doi:10.1007/s11222-006-8769-1

ter Braak, C. J. F., & Vrugt, J. A. (2008). Differential evolution Markov chain with snooker updater and fewer chains. *Statistics and Computing*, *18*(4), 435–446. doi:10.1007/s11222-008-9104-9