The Journal of Open Source Software

fgivenx: A Python package for functional posterior plotting

Will Handley[1, 2, 3]

1 Astrophysics Group, Cavendish Laboratory, J.J.Thomson Avenue, Cambridge, CB3 0HE, UK **2** Kavli Institute for Cosmology, Madingley Road, Cambridge, CB3 0HA, UK **3** Gonville & Caius College, Trinity Street, Cambridge, CB2 1TA, UK

DOI: 10.21105/joss.00849

Software

- Review
- Repository
- Archive

Submitted: 18 July 2018
Published: 28 August 2018

Summary

Researchers are often concerned with numerical values of parameters in numerical models. Our knowledge of such things can be quantified and presented using probability distributions as demonstrated in Figure 1.

Contour plots such as Figure 1 can be created using two-dimensional kernel density estimation using packages such as scipy (Jones, Oliphant, & Peterson, 2001), getdist (Lewis, 2015), corner (Foreman-Mackey, 2016) and pygtc (Bocquet & Carter, 2016), where the samples provided as inputs to such programs are typically created by a Markov Chain Monte Carlo (MCMC) analysis. For further information on MCMC and Bayesian analysis in general, "Information Theory, Inference and Learning Algorithms" is highly recommended (MacKay, 2002), which is available freely online.

As well as quantifying the uncertainty of real-valued parameters, scientists may also be interested in producing a probability distribution for the predictive posterior of a function `f(x)`. Take as a universally-relatable case the equation of a line `y = m*x + c`. Given posterior probability distributions for the gradient `m` and intercept `c`, then the ability to predict `y` knowing `x` given their linear relationship would also be characterized by some uncertainty. This is depicted as `P(y|x)` in the bottom right panel of Figure 2.

`fgivenx` is a Python package for showing the relationships as depicted in Figure 2, including the conditional Kullback-Leibler divergence (Kullback & Leibler, 1951). This `y=m*x+c` example provides a simple illustration, but the code has been used in recent Planck studies to quantify our knowledge of the primordial power spectrum of curvature perturbations (Planck Collaboration, 2016)(Planck Collaboration, 2018a)(Planck Collaboration, 2018b), in examining the dark energy equation of state (Hee, Handley, Hobson, & Lasenby, 2016) (Hee, Vázquez, Handley, Hobson, & Lasenby, 2017) for measuring errors in parameter estimation (Higson, Handley, Hobson, & Lasenby, 2017), for providing diagnostic tests for nested sampling (Higson, Handley, Hobson, & Lasenby, 2018a) and for Bayesian compressive sensing (Higson, Handley, Hobson, & Lasenby, 2018b).

`fgivenx` is a Python package for functional posterior plotting, currently used in astronomy, but will be of use to scientists performing any Bayesian analysis which has predictive posteriors that are functions. The source code for `fgivenx` is available on GitHub and has been archived as `v2.1.17` to Zenodo with the linked DOI: (Handley, 2018).

Acknowledgements

Contributions and bug-testing were provided by Ed Higson and Sonke Hee.

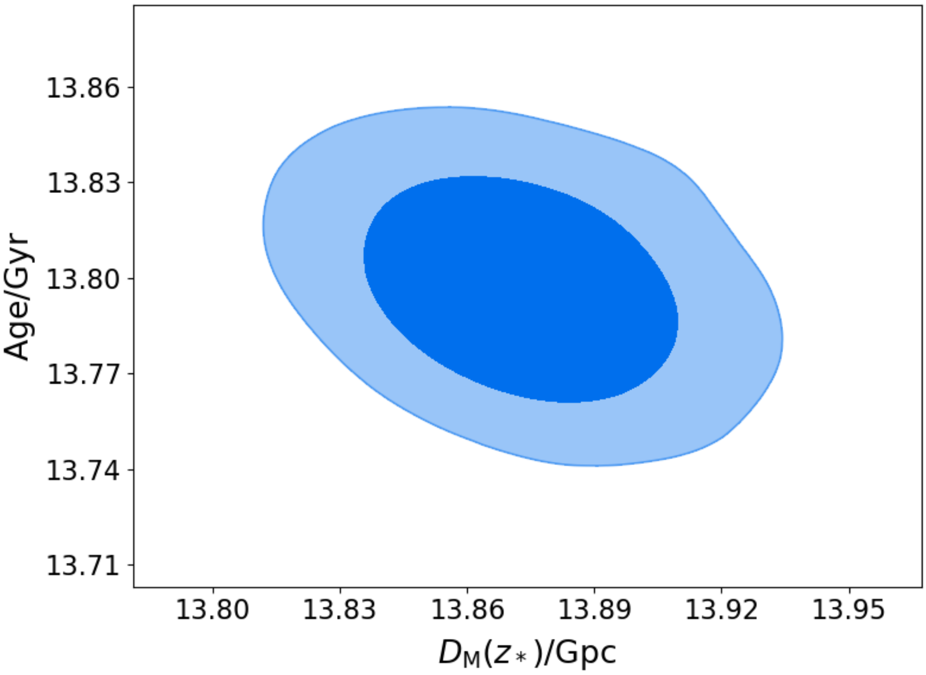

Figure 1: The age and size of the universe, as measured using Planck 2018 data. (non-Astro)Physicists may note that 14 Gigaparsecs is roughly 46 billion light years. The fact that the observable universe is roughly three times larger in light years in comparison with its age is explained by the expansion of space over cosmic history. Contours indicate 67% and 95% marginalised iso-probability credibility regions.

References

Bocquet, S., & Carter, F. W. (2016). Pygtc: Beautiful parameter covariance plots (aka. Giant triangle confusograms). *The Journal of Open Source Software*, *1*(6). doi:10.21105/joss.00046

Foreman-Mackey, D. (2016). Corner.py: Scatterplot matrices in python. *The Journal of Open Source Software*, *24*. doi:10.21105/joss.00024

Handley, W. (2018). Fgivenx: V2.1.17. doi:10.5281/zenodo.1404584

Hee, S., Handley, W. J., Hobson, M. P., & Lasenby, A. N. (2016). Bayesian model selection without evidences: application to the dark energy equation-of-state. *MNRAS*, *455*, 2461–2473. doi:10.1093/mnras/stv2217

Hee, S., Vázquez, J. A., Handley, W. J., Hobson, M. P., & Lasenby, A. N. (2017). Constraining the dark energy equation of state using Bayes theorem and the Kullback-Leibler divergence. *MNRAS*, *466*, 369–377. doi:10.1093/mnras/stw3102

Higson, E., Handley, W., Hobson, M., & Lasenby, A. (2017). Sampling Errors in Nested Sampling Parameter Estimation. *ArXiv e-prints*.

Higson, E., Handley, W., Hobson, M., & Lasenby, A. (2018a). Diagnostic Tests for Nested Sampling Calculations. *ArXiv e-prints*.

Higson, E., Handley, W., Hobson, M., & Lasenby, A. (2018b). *Bayesian sparse reconstruction: A brute-force approach to astronomical imaging and machine learning.*

Jones, E., Oliphant, T., & Peterson, P. (2001). SciPy: Open source scientific tools for Python. Retrieved from http://www.scipy.org/

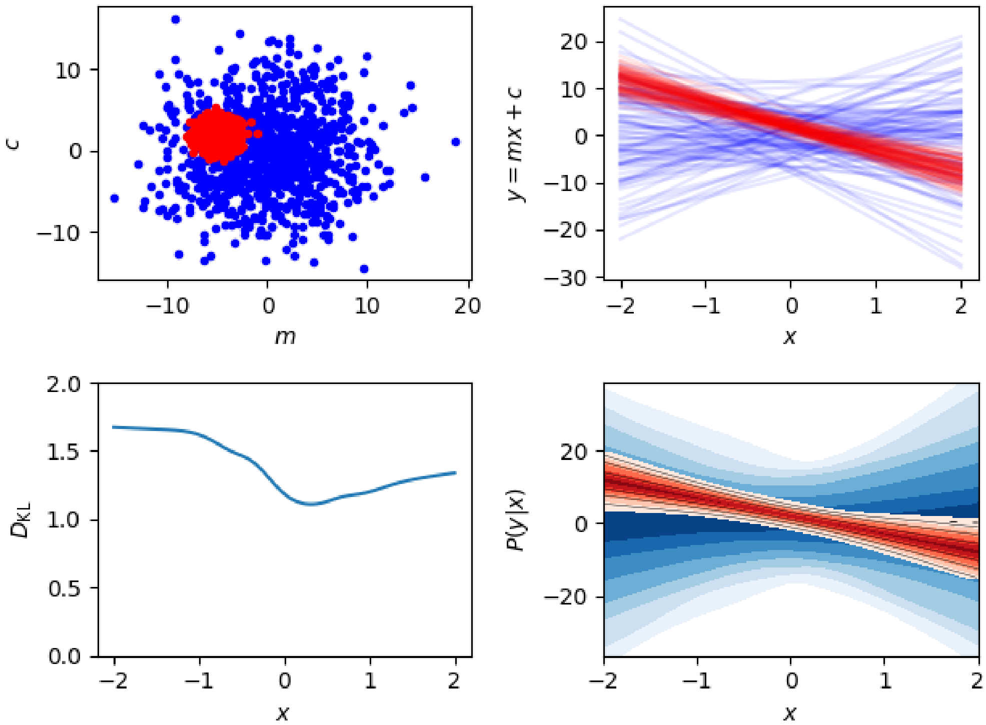

Figure 2: Example output of fgivenx provided some prior and posterior samples and a linear test function. Top-left: underlying parameter covariances between `m` and `c` for realizations from the prior (blue) and from the posterior (red). Top-right realisations function `y=m*x+c`. Bottom-left: The conditional Kullback-Leibler divergence. Bottom-right: The probability of measuring y for a given x, essential a contour version of the panel directly above, where contours indicate 67%, 95% and 99% iso-probability credibility regions.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.*, *22*(1), 79–86. doi:10.1214/aoms/1177729694

Lewis, A. (2015). Getdist github repository. Retrieved from https://github.com/cmbant/getdist

MacKay, D. J. C. (2002). *Information theory, inference & learning algorithms*. New York, NY, USA: Cambridge University Press.

Planck Collaboration. (2016). Planck 2015 results. XX. Constraints on inflation. *A&A*, *594*, A20. doi:10.1051/0004-6361/201525898

Planck Collaboration. (2018a). Planck 2018 results. I. Overview and the cosmological legacy of Planck. *ArXiv e-prints*.

Planck Collaboration. (2018b). Planck 2018 results. X. Constraints on inflation. *ArXiv e-prints*.