A space of parameter spaces in the space sciences: parametric Bayesian inference in astronomy, cosmology and particle physics

Johannes Buchner

Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, 85741 Garching, Germany

10th July 2023

ABSTRACT

A sample of parametric Bayesian inference applications from astronomy, cosmology and particle physics is studied, augmented by mock data sets and toy problems. The parameter spaces and posterior distributions are characterized by (1) the number of model parameters, (2) whether the posterior shape is similar to a gaussian, (3) whether the posterior has light or heavy tails, (4) how small the posterior is compared to the prior, i.e., how informative the data are, (5) whether some parameters remain unconstrained while others are highly constrained, (6) whether the posterior has multiple, disconnected modes. These axis define a parameter space of inference problems. We characterize each of the inference problems and observe that inference in astrophysics spans the entire parameter space, from low to high dimensionality, mono- to multi-modal, and a variety of complex distributions that range from uninformative to highly informative. Furthermore, the computational cost of the physical models can range from milliseconds to dozens of seconds. The collated sample of inference problems is proposed as a standard test bed for new samplers. For reproducibility and ease of use, a Docker compute image is provided.

Key words. Bayesian inference; parametric models; astrophysics

1. Introduction

Fitting parametric models to experimental data is one of the key methods to infer physical parameters. In physics, investigation of distant processes is possible by modelling the measurement process accurately. Here, we focus on problems where the model has continuous parameters with predefined prior ranges, and where a likelihood function has been defined to compare the model prediction to data. Some examples include fitting the power spectrum of the Cosmic Microwave Background (CMB) with Dark Energy and Cold Dark Matter (λCDM) cosmologies, fitting time series of the radial velocity of a host star gravitationally pulled by its exoplanets, dissecting multiple components in spectra and population inference from uncertain measurements of many individual objects, such as luminosity or mass functions. The plausible ranges of model parameters that match the data are typically tested in a Bayesian framework once prior and likelihood are specified with Monte Carlo samplers.

Monte Carlo sampling methods of varying complexity have been developed over the last decades. This includes variations of Markov Chain Monte Carlo (MCMC), Particle Monte Carlo (PMC), Importance Samplers (IS) and Nested Samplers (NS). Specific implementations specify the initialisation, exploration strategy (e.g., proposal function) and termination criterion. These are often tuned for the application. Reliable parameter recovery of a method can be tested by Monte Carlo simulating new datasets, and analysing them. An alternative are toy inference problems that approximate features of the real problem. These can be more easily understood and more rapidly analysed. Given the diversity of algorithms and inference problems, it is interesting to consider whether a different algorithm can perform well on the same problem, and whether the currently used algorithm can be transferred to another problem. This work is focusing on the applicability of Monte Carlo Samplers over different types of inference problems.

Inference problems differ substantially by the posterior distribution that a Monte Carlo sampler has to explore. The main characteristics of problems include (1) the number of model parameters, (2) whether the posterior shape is similar to a gaussian, (3) whether the posterior has light or heavy tails, (4) how small the posterior is compared to the prior (i.e., how informative the data are), (5) whether some parameters remain unconstrained while others are highly constrained, (6) whether the posterior has multiple, disconnected peaks. Besides a systematic classification of inference problems based on five characteristics, this work presents a diverse set of real and toy inference problems that cover the entire classification space.

2. Data: Model survey

To cover most of the problem space, we collected inference problems published in the literature and available in open-source physics packages. Appendix A introduces real-world problems, which form the main sample in this work. Simplified mock problems with generated data sets that approximate real-world problems are presented in Appendix B. Appendix C introduces artificial toy problems.

For each problem, the likelihood function $L(\theta)$ is defined together with prior distribution $\pi(\theta)$ over the parameter space.

3. Methodology: A classification space for parametric inference

Various difficulties are encountered by different sub-disciplines. Here we specify six characteristics and give a mathematical definition for each.

3.1. Dimensionality

In astrophysics, fitting problem dimensionalities range from 1 to millions parameters. Examples of extremely high-dimensional problems include pixel reconstructions (e.g.,) and . With more parameters, the possible combinations of parameter values rises exponentially (the curse of dimensionality). This makes the problem complex to explore and distances between parameter space points meaningless.

Here we defined three common sub-groups: low-dimensional ($d = 2 - 9$), mid-dimensional ($d = 10 - 29$) and high-dimension ($d \geq 30$). The boundaries are set near where simple and more sophisticated ideas of geometric sampling start failing. Extremely high dimensions ($d \gg 100$) are not the focus of this work. We note that these virtually always require the derivatives of likelihood functions to effectively navigate the parameter space. The availability of likelihood derivatives could be considered an additional classification category.

3.2. Information gain

Depending on the data quality of the experiment, the posterior may be a tiny region of the prior, or be identical with the prior. This can be quantified by the Kullback-Leibler divergence, or surprise, between the two probability distributions:

$$D_{\mathrm{KL}} = \int \pi(\theta) \ln \frac{\pi(\theta)}{P(\theta)} d\theta$$

Here, π and P give the prior and posterior over the parameter vector space θ. In the case of a base-e logarithm, the unit of D_{KL} is nats, and approximately means how many e-foldings it takes to cut the prior until the posterior is reached. Finding that small region can be a challenge for sampling algorithms (and maximum likelihood minimizers).

In practice, the information gain is already computed by nested sampling algorithms internally for error estimation, and we adopt that method as a measurement. Also, the information gain is related to the number of iterations of the nested sampling algorithms needs to zoom in until the likelihood appears flat.

3.3. Multi-modality

When data can be explained with similar quality by different combinations of processes, the posterior exhibits multiple peaks. This is common in fits of multiple components with (nearly-)interchangable predictions, paired with poor descrimination power of the data. Algorithms based on local jumps can find it difficult to navigate between modes,

because a proposal tuned to a single mode may reach another distant mode with vanishing proposal probability. The bottom row of Figure 1 shows examples of multimodal distributions.

To mathematically define multi-modality, a threshold criterion is needed to define disconnectedness. In principle any clustering algorithm can be used. For simplicity and reproducibility, we adopt a simple approach. First, histograms of the marginal posteriors are histogrammed into 20 bins. Bins with density less than $1/5$ of the peak density are considered "empty". Gaps are identified, and thresholds that bracket the peaks extracted. This is repeated for every dimension. Then, all combinations of brackets are computed. These are the clusters. If posterior samples are members of multiple clusters, the clusters are merged. The number of remaining, non-empty clusters is the number of modes of the problem.

3.4. Non-Gaussianity

The Bernstein–von Mises theorem states that when the data are highly informative, the posterior is shaped like a multi-variate Gaussian. Then, Laplace's approximation of the posterior with a log-quadratic density function is justified. This can furthermore occur if the model is linear in its parameters, or a first-order Taylor expansion of the model provides a reasonable approximation at the maximum a posteriori. For this reason, some algorithms are constructed to behave optimally when the posterior is Gaussian (see e.g., Laplace approximation).

The gaussianity of a posterior can be easily measured through the surprise from a best-fitting multi-variate Gaussian to the posterior. We obtain a suitable gaussian from the mean and covariance of the posterior samples. In multimodal cases, the posterior is highly non-gaussian.

3.5. Tail weight

While the posterior may have ellipsoidal contours like a Gaussian, the posterior density may decline steeper or shallower than a square-exponential, i.e., have thin or heavy tails. For example, when outliers are allowed (e.g., in student-t distribution or explicit outlier modelling), the wings of the posterior can be wide. Some algorithms may be optimized for square-exponential declines. Another cause of heavy tails is when most data points are fitted well by one component, and a minor component relevant for a small data subset improves the fit slightly (for example, in blind spectral line searches on top of a continuum). This leads to a phase transition, where the parameter space to be explored changes rapidly (with small likelihood change) from a wide volume to a narrow volume. This is difficult for many samplers.

To quantify how the weight is distributed over the prior volume, we compute the prior volume above a given likelihood threshold. This is illustrated in Figure 2. We compute the 5% and 95% quantiles of the volume ranges where most of the probability mass resides, and compute the tail weight as:

$$\mathrm{TW} = \log \frac{V_{5\%}}{V_{95\%}}$$

In practice, the mapping of volume and likelihood, as well as the normalising constant, the marginal likelihood in-

Fig. 1. Selected pair-wise posterior distributions from some of our problems. These illustrate non-linear degeneracies (e.g., middle panels), unequal axes (e.g., left-most and right-most top panels), multi-modality (bottom panels). The loggamma problem (third panel, bottom row) also has heavy tails towards the left. Some parameters are uninformative (e.g., param1 in top right panel) or at the prior parameter edge (middle left panel, ratio $q \geq 1$, top right panel, $0 \leq$ param2 ≤ 1).
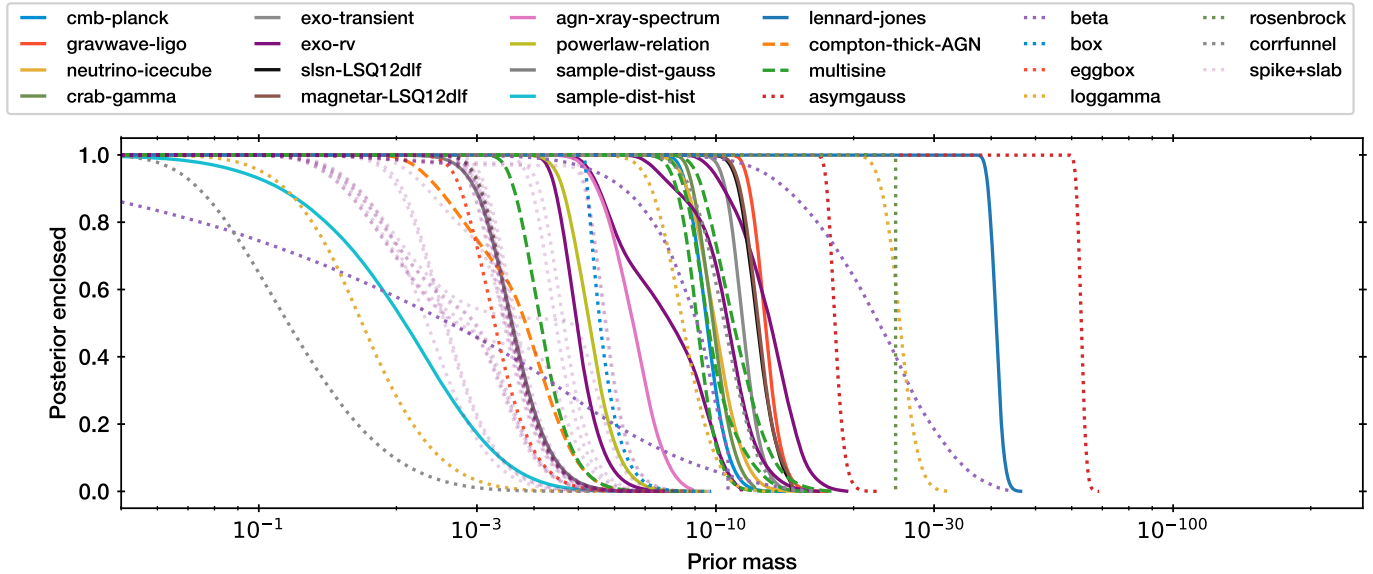
Fig. 2. Probability fraction enclosed as a function of prior volume. From the highest likelihood regions outwards, the volume is increased from left to right until the entire prior space ($V = 1$) is enclosed. The median (cross) indicates how large the posterior volume is relative to the prior, and is related to the *information gain*. Quantiles at 5% and 95% indicated as circles indicate the shell where most probability mass is enclosed. This is related to the *tail weight*. Some problems show wide transitions (e.g., green dashed) relative to a gaussian (blue dashed).

tegrated over volume shrinkages, is already computed internally in nested sampling. We cap the value at TW = $\log 10 \times d$.

3.6. Parameter Inequality

Some model parameters may alter the model prediction strongly, while others have more minute implications. Because of this, the posterior of some parameters can be consistent with the prior with no information learnt. For other parameters, its plausible range may have diminished by several orders of magnitude. Proposals that are isotropic over the parameter space directions then may perform poorly. The top row of Figure 1 shows such examples in pairs of parameters.

We quantify the inequality of parameters in problems by marginal posterior standard deviations σ_i:

$$\text{IE} = \log \frac{\max \sigma_i}{\min \sigma_i}$$

If all parameters have similar posterior uncertainties, IE $\approx$ 0. We cap this value to at most IE = 10.

4. Results

4.1. Problem Space

Figure 3 presents the location of each inference problem in
.

5. Discussion

This work has assembled a list of parametric Bayesian inference applications from astronomy, cosmology and particle physics. This is augmented by mock applications and toy problems. The parameter spaces and posterior distributions have been characterized and cast into a space of parametric inference spaces. The applications are diverse, spanning from low to high dimensionality, mono to multimodal, and a variety of complex distributions that range from uninformative to highly informative. Furthermore, the computational cost of the physical models can range from milliseconds to dozens of seconds.

In recent years, a variety of Bayesian inference sampling packages have been published. MCMC-based algorithms include emcee Foreman-Mackey et al. (2013), Stan Carpenter et al. (2017), zeus Karamanis & Beutler (2020) and pocoMC Karamanis et al. (2022)). Nested sampling-based algorithms include multinest Feroz et al. (2009), polychord Handley et al. (2015), DNest4 Brewer & Foreman-Mackey (2018), dynesty Speagle (2020), ultranest Buchner (2020), nautilus Lange (2023) and i-nessai Williams et al. (2023)). For a simple reference with Laplace's approximation[1] is available with snowline[2], which also provides variational Bayes and importance sampling [3]. It is thus a logical next

---

[1] based on `https://iminuit.readthedocs.io/`, see Dembinski & et al. (2020); James & Roos (1975)
[2] `https://johannesbuchner.github.io/snowline/`
[3] based on `https://github.com/pypmc/pypmc`, see Beaujean & Caldwell (2013)

field	name	dim	cost	depth	width	modes	asym	!gauss	phase
cosmology	cmb-planck	6	659	1.59	2.4	1	1.34	0.01	0.003
gravitational waves	gravwave-ligo	7	4	1.83	2.6	1	2.14	0.05	0.004
astroparticle	neutrino-icecube	16	146	0.62	3.2	1	29.68	0.12	0.004
supernova remnants	crab-gamma	6	76	1.64	2.6	1	3.27	0.05	0.006
exoplanets	exo-transient	9	4	1.27	2.5	1	33.09	0.08	0.003
	exo-rv-0	2	3	2.48	1.8	1	1.19	0.00	0.000
	exo-rv-1	7	3	1.14	5.0	3	33.09	0.44	0.225
	exo-rv-2	12	2	0.88	5.9	6	33.07	0.30	0.015
	exo-rv-3	17	3	0.77	6.3	9	6.16	0.06	0.031
transients	slsn-LSQ12dlf	12	16	1.03	2.8	6	14.22	0.15	0.023
	magnetar-LSQ12dlf	12	14	1.03	2.8	7	17.31	0.17	0.020
extragalactic	agn-xray-spectrum	4	4	1.65	2.7	1	3.86	0.20	0.007
	powerlaw-relation	3	3	1.76	2.0	1	1.18	0.01	0.001
	sample-dist-gauss	2	1	1.79	2.1	1	1.03	0.08	0.001
	sample-dist-hist	11	3	0.19	2.4	1	11.51	0.08	0.005
materials	lennard-jones-6	12	0	3.42	4.2	1	1.00	0.11	0.003
mock	compton-thick-AGN	5	1	0.77	2.7	2	7.42	0.02	0.037
	multisine-0	2	1	2.07	1.8	1	1.51	0.00	0.000
	multisine-1	5	0	1.82	2.5	2	2.00	0.02	0.002
	multisine-2	8	0	1.20	4.3	20	32.43	0.50	0.002
	multisine-3	11	0	0.99	5.2	66	33.04	0.54	0.006
toy	asymgauss-4d	4	0	4.55	2.1	1	3.07	0.02	0.001
	asymgauss-16d	16	0	3.93	3.8	1	3.79	0.08	0.002
	beta-2d	2	0	1.31	11.9	1	11.78	12.88	0.010
	beta-10d	10	0	0.91	4.8	8	4.50	0.10	0.014
	beta-30d	30	0	0.76	26.4	1024	12.19	1.04	0.039
	box-5d	5	0	1.12	1.2	1	1.00	0.35	0.000
	eggbox-2d	2	0	1.64	1.8	18	1.00	0.01	0.001
	loggamma-2d	2	0	0.84	1.9	4	1.15	0.01	0.001
	loggamma-10d	10	0	0.85	3.3	4	1.76	0.01	0.005
	loggamma-30d	30	6	0.84	5.9	4	1.82	0.12	0.007
	rosenbrock-2d	2	0	1.78	1.9	1	1.16	0.02	0.001
	rosenbrock-20d	20	0	1.24	-1.3	1	1.00	1.27	0.000
	corrfunnel-2d	2	0	0.58	1.7	1	4.82	0.02	0.001
	corrfunnel-10d	10	14	1.04	4.5	1	1.65	0.28	0.004
	spikeslab-1000-2d-4000	2	2	2.87	1.9	1	1.00	0.22	0.003
	spikeslab-1000-2d-400	2	2	2.23	1.9	1	1.00	0.07	0.001
	spikeslab-1000-2d-40-off0.1	2	2	1.79	1.9	1	1.01	0.04	0.001
	spikeslab-1000-2d-40-off0.2	2	2	1.76	1.9	1	1.01	0.02	0.001
	spikeslab-1000-2d-40-off0.4	2	2	1.79	1.9	1	1.00	0.00	0.001
	spikeslab-1000-2d-40-off10	2	2	1.76	1.9	1	1.01	0.15	0.001
	spikeslab-1000-2d-40-off1	2	2	1.78	1.9	1	1.01	0.02	0.001
	spikeslab-1000-2d-40-off2	2	2	1.72	1.9	1	1.01	0.02	0.001
	spikeslab-1000-2d-40-off4	2	2	1.77	1.9	1	1.01	0.11	0.001
	spikeslab-1000-2d-40	2	2	1.83	1.9	1	1.00	0.00	0.002
	spikeslab-1000-2d-4	2	2	1.29	1.9	1	1.00	0.01	0.001
	spikeslab-1-2d-4000	2	1	2.41	5.0	1	1.01	2.03	0.510
	spikeslab-1-2d-400	2	1	1.34	4.2	1	1.03	1.71	0.358
	spikeslab-1-2d-40-off10	2	2	1.74	2.7	1	1.01	0.41	0.028
	spikeslab-1-2d-40-off1	2	2	1.48	3.0	1	1.01	0.82	0.173
	spikeslab-1-2d-40-off2	2	2	1.36	3.1	2	1.00	0.93	0.187
	spikeslab-1-2d-40-off4	2	2	1.43	3.0	2	1.00	0.88	0.193
	spikeslab-1-2d-40	2	1	1.49	3.1	1	1.01	0.84	0.190
	spikeslab-1-2d-4	2	1	1.17	2.2	1	1.01	0.19	0.001
	spikeslab-40-2d-4000	2	2	2.86	4.4	1	1.02	0.29	0.053
	spikeslab-40-2d-400	2	2	2.32	2.2	1	1.02	0.10	0.024
	spikeslab-40-2d-40-off10	2	2	1.73	1.9	1	1.00	0.13	0.001
	spikeslab-40-2d-40-off1	2	2	1.78	2.2	1	1.01	0.06	0.006
	spikeslab-40-2d-40-off2	2	2	1.81	2.0	1	1.00	0.03	0.007
	spikeslab-40-2d-40-off4	2	2	1.76	2.1	1	1.01	0.11	0.006
	spikeslab-40-2d-40	2	2	1.79	2.0	1	1.00	0.04	0.005
	spikeslab-40-2d-4	2	2	1.27	1.9	1	1.01	0.01	0.001

Table 1. List of problems analysed. The columns describe (1) research field, (2) name, (3) dimensionality, (4) model evaluation cost in milliseconds, (5) information gain, (6) tail weight, (7) parameter inequality, (8) Gaussian approximation information loss, (9) phase transitions.
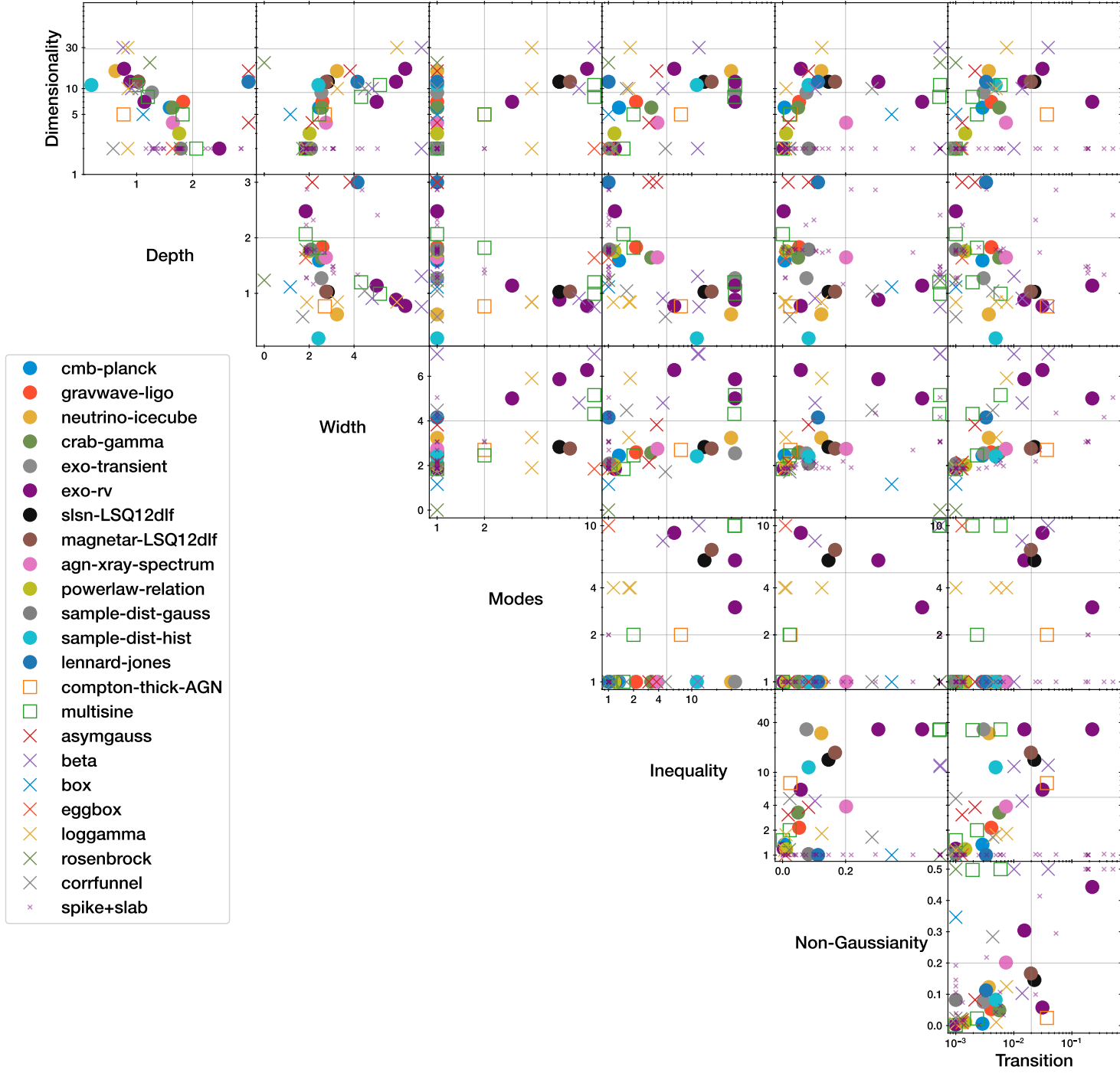
Fig. 3. Parameter space of the problems over the defined six characteristics.

step to compare these samplers. However, there are additional steps to consider.

We point out a few guidelines for comparing Bayesian samplers. Initially, the full sample of problems presented here should be considered. If only a subset is of interest, this should be clearly and transparently stated (e.g., focus on low-dimensional inference problems, on mono-modal inference problems, etc.). Secondly, it is a trivial statement that with more compute budget, a better accuracy can be achieved. Therefore, placing bias or accuracy against computational cost for each algorithm and runtime is interesting. A basic computational cost quantity is the number of (likelihood) model evaluations or wall-clock time. The

latter is relevant for computationally cheap models paired with algorithms that require costly training (such as deep neural networks). Thirdly, for how to quantify posterior fidelity, see for example the probability-probability plots and Jensen-Shannon divergence in Romero-Shaw et al. (2020). Finally, where the true marginal likelihood Z is known, it is of little interest to recover $\ln Z$ to an accuracy better than 0.3, as it only mildly affects the interpretation in the context of Bayes factors.

6. Future work

Breakthrough progress in machine learning is typically driven by (1) open, large, high-quality data sets and (2) a

clear formulation of a meaningful objective. Examples span from MNIST's challenge of digitizing hand-drawn numbers to the Critical Assessment of Structure Prediction (CASP) Moult et al. (1995) protein-folding challenge, recently met by AlphaFold Jumper et al. (2021). The "Learning to learn by gradient descent by gradient descent" paper Andrychowicz et al. (2016) demonstrated that optimization algorithms can be derived by machine learning. Perhaps a similar breakthrough can be accomplished for optimal Bayesian inference sampling procedures, by providing a open, large data base and a clear objective. To this end, the survey of inference problems encountered across cosmology, particle physics, astrophysics and astronomy is presented. The representative database includes fully specified likelihood and priors in a runnable docker image with python interfaces. While similar previous work has focusd on providing simplified models that serve as unit-tests (e.g., Sountsov et al. 2020; Magnusson et al. 2021), this work uses real-world inference with the software pipelines employed by researchers.

As a first step, existing and novel algorithms can be judged across the parameter space of problems for their empirical behavior and robustness, and to make well-founded recommendation for specific applications. For more rapid testing and a unified interface, model emulators for the provided real-world examples may be helpful. This is left for future work.

A step further in the future is comparable to Atari computer game playing AIs Bellemare et al. (2012) that learn optimal game playing strategies with reinforcement learning: A playground for learning optimal Bayesian inference algorithms.

References

Aartsen, M. G., Ackermann, M., Adams, J., et al. 2018, Phys. Rev. Lett., 120, 071801
Aartsen, M. G., Ackermann, M., Adams, J., et al. 2020, European Physical Journal C, 80, 9
Aartsen, M. G., Ackermann, M., Adams, J., et al. 2019, Phys. Rev. D, 99, 032007
Andrychowicz, M., Denil, M., Gomez, S., et al. 2016, arXiv e-prints, arXiv:1606.04474
Baronchelli, L., Nandra, K., & Buchner, J. 2018, ArXiv e-prints
Beaujean, F. & Caldwell, A. 2013, ArXiv e-prints
Bellemare, M. G., Naddaf, Y., Veness, J., & Bowling, M. 2012, arXiv e-prints, arXiv:1207.4708
Betancourt, M. J. & Girolami, M. 2013, arXiv e-prints, arXiv:1312.0906
Brahm, R., Hartman, J. D., Jordán, A., et al. 2018, AJ, 155, 112
Brewer, B. J. 2016, A Rosenbrock challenge for MCMC folks...
Brewer, B. J. & Foreman-Mackey, D. 2018, Journal of Statistical Software, Articles, 86, 1
Buchner, J. 2020, UltraNest v2.2.1, https://johannesbuchner.github.io/UltraNest/
Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125
Carpenter, B., Gelman, A., Hoffman, M. D., et al. 2017, Journal of statistical software, 76, 1
Dembinski, H. & et al., P. O. 2020
Feroz, F. & Hobson, M. P. 2008, MNRAS, 384, 449
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015, MNRAS, 453, 4384

James, F. & Roos, M. 1975, Comput. Phys. Commun., 10, 343
Jumper, J., Evans, R., Pritzel, A., et al. 2021, Nature, 596, 583
Karamanis, M. & Beutler, F. 2020, arXiv e-prints, arXiv:2002.06212
Karamanis, M., Nabergoj, D., Beutler, F., Peacock, J., & Seljak, U. 2022, The Journal of Open Source Software, 7, 4634
Kormendy, J. & Ho, L. C. 2013, ARA&A, 51, 511
Lange, J. U. 2023, arXiv e-prints, arXiv:2306.16923
Magnusson, M., Bürkner, P., & Vehtari, A. 2021
Moult, J., Pedersen, J. T., Judson, R., & Fidelis, K. 1995, A large-scale experiment to assess protein structure prediction methods
Nelson, B. E., Ford, E. B., Buchner, J., et al. 2020, AJ, 159, 73
Nicholl, M., Guillochon, J., & Berger, E. 2017, ApJ, 850, 55
Romero-Shaw, I. M., Talbot, C., Biscoveanu, S., et al. 2020, MNRAS, 499, 3295
Sountsov, P., Radul, A., & contributors. 2020, Inference Gym
Speagle, J. S. 2020, MNRAS, 493, 3132
van Dokkum, P., Danieli, S., Cohen, Y., et al. 2018, Nature, 555, 629
Williams, M. J., Veitch, J., & Messenger, C. 2023, arXiv e-prints, arXiv:2302.08526

Appendix A: Real inference problems

Appendix A.1: Exoplanet detection from Radial Velocity data

One of the most efficient ways to detect exoplanets is through changes in the line-of-sight (radial) velocity of individual stars, as the tug of planets gravitationally accelerates them (Doppler shift). A planetary system leads to a complex overlay of periodic velocity changes. Additionally, the measurement of velocities is uncertain, in part because of the instrument accuracy and precision, and in part because the spectral emission lines used to measure Doppler shifts can be unstable due to stellar activity. That latter process has been modeled with Gaussian processes in recent years.

Here we adopt the problem setup of the Extreme Precision Radial Velocity III challenge Nelson et al. (2020). They simulated several artificial exoplanet systems containing two planets, with a Gaussian process and realistic observation time sampling. The challenge participants were asked to compute the Bayesian marginal likelihoods of each data set assuming that the true number of planets was 0, 1, 2 or 3. Participants did not know the true simulation input, but they were told the exact Gaussian noise properties. Here we repeat this exercise, for their dataset 5 (shown in Figure A.1). The exact specification of this problem is defined in Nelson et al. (2020). Essentially, it is similar to the sine time series problem above, except that the periodic signal can be asymmetric due to ellipticity, giving 5 parameters per planet (signal amplitude, period, pericenter time, eccentricity and mean anomaly) that describe a Keplerian orbit. Additionally, the white noise amplitude σ_j and the systematic velocity C are free parameters, giving $2 + N_{\text{planets}} \times 5$ free parameters for $N_{\text{planets}} = 0, 1, 2, 3$.

Appendix A.2: Cosmology with the Cosmic Microwave Background

The Cosmic Microwave Background (CMB) is the oldest electromagnetic signal observable. It originated when the Universe underwent a transition from being so dense that photons would be constantly scattered to the current state where photons can travel freely. These photons allow us to measure the temperature of their regions of origin approximately 379,000 years after the Big Bang. A map of their emission over the sky gives information about the temperature correlation. This carries information of how the Big Bang inflation proceeded, which is dependent on important
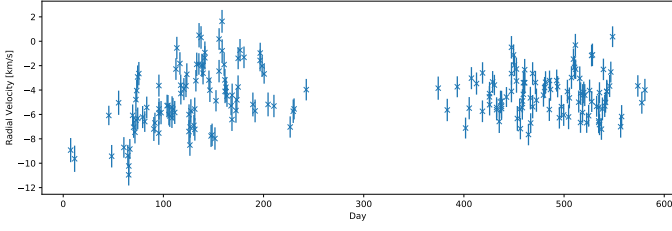
Fig. A.1. Exoplanet Doppler shift time series data (blue).

constituents of the Universe, such as the baryonic matter and dark matter content of the Universe. The CMB remains one of the most important experiments to measure cosmological parameters in the dark energy and cold dark matter cosmology framework (ΛCDM). One difficulty in the fitting of cosmological models like ΛCDM to the CMB is that the prediction of angular correlations is computationally expensive (of the order of a few seconds per likelihood evaluation).

Here we adopt a tutorial example of the MontePython cosmology fitting package. We use the fake_planck_bluebook likelihood which emulates a Planck measurement. The free parameters are ΛCDM cosmological parameters, namely the baryonic density Ω_b, the dark matter density Ω_{cdm}, the scalar spectral index n_s, A_s, the hubble parameter value, relative to $100\,\mathrm{km/s/Mpc}$, h, and the time of reionisation τ_{reion}.

Appendix A.3: X-ray spectral analysis of Active Galactic Nuclei

Active Galactic Nuclei (AGN) are regions in the centres of massive galaxies where super-massive black holes grow. As gas swirls into the black hole, enormous amounts of radiation are produced by release of gravitational energy, sometimes shining brighter than all host galaxy stars together. Close to the black hole, X-rays are also produced and are an important tracer of the mass inflow into the black hole. They also allow identifying AGN in the sky, even when the black hole is surrounded by thick gas and dust, as most of the energetic X-rays penetrate through any obscurers. X-ray focusing instruments on-board satellites allow measurements of the X-ray spectra, which carry information on the AGN luminosity (and thus black hole mass accretion rate), obscuring column density and properties of the X-ray emitter (photon index, energy turn-over). The detection of X-ray radiation is performed by counting photon events and capturing an estimate of their energy, time of arrival and location on the sky. However, the energy response and energy-dependent sensitivity of the instrument adds some analysis complexity. Additionally, when few counts are detected, as is commonly the case, the process needs to be modelled with a Poisson likelihood.

Here, we present a example which emulates problems commonly encountered in X-ray analyses. The emission spectrum at low energies can be described as:

$$F = B + A_{LE} \times E^{-\Gamma} \times \left[\exp\left(-N_{\mathrm{H}} \times E^{-3} \right) + f_{\mathrm{scat}} \right]$$

with background amplitude B, signal amplitude A, spectral photon index Γ, obscuring column density N_H, and fraction of the powerlaw that escapes unobscured f_{scat}. This model shows a degeneracy when A is low and comparable to B,

because high N_H and high f_{scat} look like $N_H = 0$ (a simple powerlaw). The full emission model is illustrated in Figure A.2. At high energies, a smooth break is included (blue curve in right panel). However, the signal is low-count, so that the width of the break, Δ_{brk}, is difficult to estimate. Additionally, the break location, E_{brk}, and photon index above the break, Γ_{HE}, are degenerate. We define the priors as:

|	
$\log A \sim$	$\mathrm{Uniform}(-5, +5)$
$\Gamma \sim$	$\mathrm{Normal}(2, 0.2^2)$
$\log N_{\mathrm{H}} \sim$	$\mathrm{Uniform}(-3, +3)$
$\log f_{\mathrm{scat}} \sim$	$\mathrm{Uniform}(-7, -1)$
$\ln B \sim$	$\mathrm{Normal}(0.2, 0.1^2)$
$\log A_{HE} \sim$	$\mathrm{Normal}(\log A, 0.5^2)$
$\log E_{\mathrm{brk}} \sim$	$\mathrm{Uniform}(1, 3)$
$\log \Delta_{\mathrm{brk}} \sim$	$\mathrm{Uniform}(-2, 1)$
$\Gamma_{HE} \sim$	$\mathrm{Uniform}(-5, 5)$

Appendix A.4: Gravitational wave analysis

Gravitational waves originate from distortions of space-time by compact objects. Recently, the development of multiple, extremely sensitive instruments have allowed the observation of two black holes merging [ref]. The gravitational event GW170817 was the first event detected by three detectors and allowed for the first time localisation on the sky, albeit with substantial parameter degeneracies.

Here we adopt a tutorial example of the PyCBC Gravitational wave analysis package[4]. The merging system GW170817 is described by the mass ratio q, the chirp mass m_{chirp}, which is a combination of the two masses that influences the signal amplitude, the inclination i of the system relative to the observer, the time of coalescence t_c in seconds, the distance d and position on the sky (RA, DEC). The priors adopted are:

$q \sim$	$\mathrm{Uniform}(1, 2)$
$m_{\mathrm{chirp}} \sim$	$\mathrm{Uniform}(1, 2)$
$\sin i \sim$	$\mathrm{Uniform}(0, 1)$
$t_c - t_m \sim$	$\mathrm{Uniform}(0.02, 0.05)$
$d \sim$	$\mathrm{Uniform}(10, 100)$
$RA \sim$	$\mathrm{Uniform}(0, 2\pi)$
$\cos DEC \sim$	$\mathrm{Uniform}(-1, 1)$

---

[4] based on https://github.com/gwastro/PyCBC-Tutorials/blob/7f5ff8fdd40c5dce2237b6082e1755b4aba9b989/tutorial/inference_1_ModelsAndPEByHand.ipynb
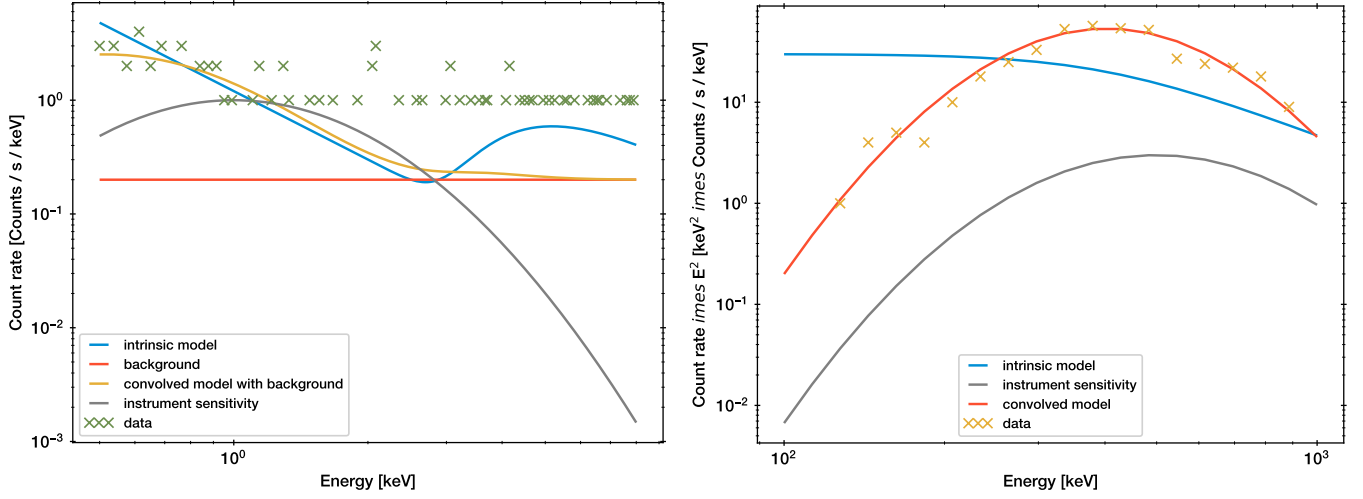
Fig. A.2. X-ray spectra simultaneously fitted. *Left panel*: Low-energy spectrum. The model (blue curve) is a sum of two powerlaws, one altered by a exponential truncation towards the low-energy side. That model is multiplied by the instrument sensitivty (gray curve), and background contamination (orange) is added. The final model is shown in green, from which poisson data are drawn (red crosses). *Right panel*: High-energy spectrum. The intrinsic model is the same powerlaw, with a smooth powerlaw turnover (blue). The instrument sensitivity (gray) leads to an effective model (orange), from which poisson data (green crosses) are sampled.

where t_m is the time automatically associated by an automated pipeline searching for candidate events in the noise.

Appendix A.5: Atmospheric Neutrino Oscillations from IceCube

IceCube is a neutrino detector near the south pole, which has collected atmospheric neutrino data Aartsen et al. (2018,?, 2019, 2020) that can be studied for neutrino oscillation. Here we adopt a tutorial example of the PISA IceCube analysis package[5] applied to three-year IceCube data, where the muon and neutrinos model predicts a 2d-histogrammed signal, which is compared to collected data. The parameters are adopted from the PISA defaults:

nue_umu_ratio $\sim$	Gauss(1, 0.05)
Barr_uphor_ratio $\sim$	Gauss(0, 1)
Barr_nu_nubar_ratio $\sim$	Gauss(0, 1)
delta_index $\sim$	Gauss(0, 0.1)
theta13 $\sim$	Gauss(8.5°, 0.205°)
theta23 $\sim$	Uniform(31°, 59°)
deltam31 $\sim$	Uniform(0.001 eV, 0.007 eV)
aeff_scale $\sim$	Uniform(0, 3)
nutau_norm $\sim$	Uniform(−1, 8.5)
nu_nc_norm $\sim$	Gauss(1, 0.2)
opt_eff_overall $\sim$	Gauss(1, 0.1)
opt_eff_lateral $\sim$	Gauss(25, 10)
opt_eff_headon $\sim$	Uniform(−5, 2)
ice_scattering $\sim$	Gauss(0, 10)
ice_absorption $\sim$	Gauss(0, 10)
atm_muon_scale $\sim$	Uniform(0, 5)

Appendix A.6: Gamma-rays from the Crab pulsar-wind nebula

The Crab nebula is the brightest astrophysical source in the sky at high energies. The Large-Area Telescope (LAT) on-board Fermi has observed emission from the region, which is a super-position of the Crab Pulsar (PSR J0534+2200), and synchrotron and Inverse Compton emission from the Crab Nebular. We follow the tutorial of the 3ML multi-messenger package [6].

The parameters include the normalisation (uniform from 0 to $1.4 \times 10^{10} \mathrm{keV}^{-1}\mathrm{s}^{-1}\mathrm{cm}^{-2}$) and spectral index (uniform from -10 to 10) of the super_cutoff_powerlaw

---

[5] based on https://github.com/icecube/pisa/blob/
f91224b58360ee9ecefe4bdb232249263c9eee17/pisa_
examples/IceCube_3y_oscillations_example.ipynb

[6] based on https://threeml.readthedocs.io/en/stable/
notebooks/Fermipy_LAT.html

model for PSR J0534+2200, the normalisation for the power law spectrum of NVSS J052622+224801 and 4FGL J0544.4+2238, and the isotropic background normalization and the galactic background factor. All of these parameters are assigned uninformative (uniform or log-uniform) priors:

uniform: Parameter K = 1.4140014052984935e-13 [1 / (keV s cm2)]
(min_value = 1.414001405298436e-16, max_value = 1.4140014052984935e-10, delta = 0.1, free = True)
uniform: Parameter index = -1.9322177752420284 []
(min_value = -10.0, max_value = 10.0, delta = 0.2, free = True)
log-uniform: Parameter K = 1.1098000000000038e-17 [1 / (keV s cm2)]
(min_value = 1.1098000000000037e-20, max_value = 1.1098000000000038e-14, delta = 0.1, free = True)
log-uniform: Parameter K = 1.3915000000000025e-16 [1 / (keV s cm2)]
(min_value = 1.3915000000000024e-19, max_value = 1.3915000000000025e-13, delta = 0.1, free = True)

Appendix A.7: Exoplanet transit observations

Photometric light curves of stars can show small, periodic dips in brightness, which are interpreted as transits of exoplanets in front of the star. We follow the tutorial of the juliet package[7] to analyse HATS-46 light curve observed with the Transiting Exoplanet Survey Satellite Brahm et al. (2018). The model is essentially a constant light curve modified by a U-shaped dip that repeats with some period.

The model parameters include the properties of the planet (period P in days, time-of-transit center t0, planet-to-star radius ratio p, eccentricity ecc, argument of periastron passage omega), as well as nuisance parameters (dilution factor mdilution, offset relative flux mflux, jitter sigma, Limb-darkening parameters q1 and q2):

P_p1 $\sim$	Gauss(4.7, 0.1)
t0_p1 $\sim$	Gauss(1358.4, 0.1)
r1_p1 $\sim$	Uniform(0, 1)
r2_p1 $\sim$	Uniform(0, 1)
q1_TESS $\sim$	Uniform(0, 1)
q2_TESS $\sim$	Uniform(0, 1)
ecc_p1 $=$	0
omega_p1 $=$	90
rho $\sim$	LogUniform(100, 10000)
mdilution_TESS $=$	1.0
mflux_TESS $\sim$	Gauss(0, 0.1)
sigma_w_TESS $\sim$	LogUniform(0.1, 1000)

Appendix A.8: Superluminous supernova

Super-Luminous Supernovae (SLSN) are extreme explosions at the end of stellar evolution. By fitting photometric

observations over time, the explosion mechanism and its physical parameters can be determined (and distinguished from other transients). We follow the tutorial of MOS-FiT [8], to analyse LSQ12dlf, with two models described in Nicholl et al. (2017): a magnetar engine with a simple spectral energy distribution ('magnetar' model) and a magnetar with a modified spectrum and additional constraints ('slsn' model). The parameters are:

nhhost $\sim$	LogUniform(10^{16}, 10^{23})
Pspin $\sim$	Uniform(1, 10)
Bfield $\sim$	LogUniform(0.1, 10)
Mns $\sim$	Uniform(1, 2)
thetaPB $\sim$	Uniform(0, 1.5708)
texplosion $\sim$	Uniform(-500, 0)
kappa $\sim$	Uniform(0.05, 0.2)
kappagamma $\sim$	LogUniform(0.1, 10000)
mejecta $\sim$	LogUniform(0.001, 100)
vejecta $\sim$	Uniform(5000, 20000)
temperature $\sim$	LogUniform(1000, 100000)
variance $\sim$	LogUniform(0.001, 100)
codeltatime $\sim$	LogUniform(0.001, 100)
codeltalambda $\sim$	LogUniform(0.1, 10000)

In case of 'slsn', the following parameters differ:

Bfield $\sim$	Uniform(0.1, 10)
mejecta $\sim$	LogUniform(0.1, 100)
vejecta $\sim$	Uniform(5000, 20000)
temperature $\sim$	Uniform(3000, 10000)

Appendix A.9: Lennard-Jones potential

In material science, the group behaviour of atoms gives rise to structures and phases of matter (solids, liquids, gases, crystals, metals, etc). A standard example is the Lennard-Jones potential, which we adopt here with 6 particles in a box. Pairs of particles feel two forces, a repulsive force at short distances (Pauli repulsion) and an attractive force at longer distances (van der Waals force). This are formulated as:

$$\mathcal{L} = \prod_i \prod_{j>i} \left(\frac{\sigma}{r_{ij}}\right)^6 - \left(\frac{\sigma}{r_{ij}}\right)^{12}$$

where r_{ij} is the Euclidean distance between particles with indices i and j, but at least σ. We set $\sigma = 10^{-3}$. To avoid identical modes, we set the likelihood to zero when $|z_1| < |z_2| < ... < |z_n|$ is not satisfied.

Each particle's three-dimensional location (x,y,z) is a free parameter between -1 and +1. Due to translation symmetry, we assume the first particle is placed in the coordinate centre (0,0,0). From rotation symmetry, we assume the second particle is placed along the positive z direction (0,0,z2). From a second rotation symmetry, the third particle is placed on the (0,y3,z3) plane. The remaining particles have all (x,y,z) coordinates as free parameters.

---

[7] based on https://juliet.readthedocs.io/en/latest/tutorials/transitfits.html#transit-fits

[8] based on https://mosfit.readthedocs.io/en/latest/fitting.html#public-data

Stars in the centres of galaxies move akin to a ideal gas, so that the velocity dispersion of galactic central elliptical components (bulges) correlates with the bulge mass. Kormendy & Ho (2013) presented a compilation of measurements of velocity dispersions and masses, both annotated with asymmetric (and heteroscedastic) error bars. Following the "Fitting a line" tutorial[9] of UltraNest, we assume these correspond to Gaussian tails, scaled according to the error bar size. The measurements are fit with a powerlaw, with the intrinsic scatter along the powerlaw accounted for by a log-Normal distribution. The parameters are

$$
\begin{aligned}
\text{slope} &\sim \quad \text{Uniform}(-3, 3) \\
\text{offset} &\sim \quad \text{LogUniform}(10, 1000) \\
\text{scatter} &\sim \quad \text{LogUniform}(0.001, 10)
\end{aligned}
$$

The likelihood is integrating the Log-Normal distribution over the data point distribution, i.e., it is a hierarchical Bayesian model. This integration is performed numerically.

Appendix A.11: Sample distributions: A galaxy without dark matter

Appendix A.11.1: Gaussian Sample distribution

van Dokkum et al. (2018) observed the velocity of globular clusters in a low-mass, ultra-diffuse galaxy. The width of the (Gaussian) distribution of velocities, i.e., the velocity dispersion, is directly related to the total galaxy mass. From comparing the stellar light to the total light, they inferred that the dark matter mass in this galaxy is negligible. Following [10], we analyse their velocity measurements with Gaussian error bars, assuming a Gaussian sample distribution (see e.g., Baronchelli et al. (2018) for such a Gaussian hierarchical model in astronomy). The parameters are the mean, for which we assume a uniform distribution between -100 and 100 km/s, and the scatter, for which we assume a log-uniform distribution between 1 and 1000 km/s. The parameter for each data point's true value is integrated out, as implemented in [11].

Appendix A.11.2: Dirichlet sample distribution

The above example is repeated, but with a more flexible sample distribution. Following the PosteriorStacker tutorial[12], a uniform Dirichlet distribution is adopted, with 11 uniformly spaced bins between -80 km/s and +80 km/s.

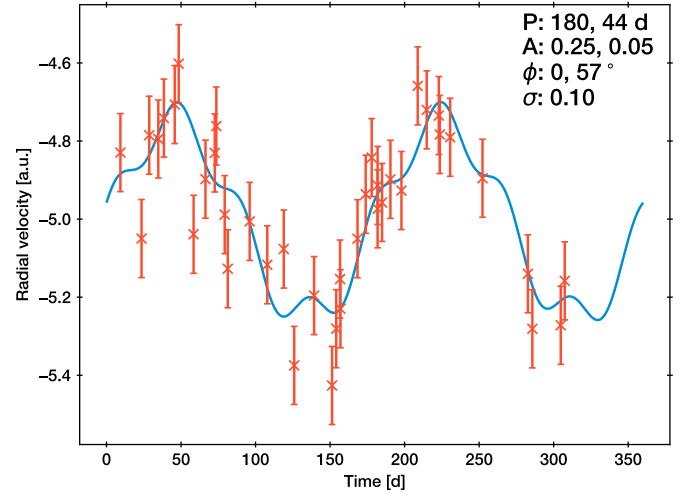## Appendix B: Mock problems

Appendix B.1: Sine time series

Analyses of inhomogeneously sampled light curves is a common problem in astrophysics, sometimes with complex noise processes and semi-periodic signals. Here we present

---

[9] https://johannesbuchner.github.io/UltraNest/example-line.html
[10] https://johannesbuchner.github.io/UltraNest/example-intrinsic-distribution.html
[11] https://github.com/JohannesBuchner/PosteriorStacker
[12] https://github.com/JohannesBuchner/PosteriorStacker

Fig. B.1. Sine time series data (orange) with generating two-component model.

a simple multi-component sine signal as a toy problem, and uniformly randomly sample observing times. The problem is defined as:

$$
\begin{aligned}
\log L &= \prod_{i=1}^{M} \text{Normal}(m(t_i, \theta) - d_i, \sigma^2) \\
m(t, \theta) &= m_0 + \sum_{j=1}^{n_{\text{comp}}} A_c \times \sin\left(\frac{2\pi t}{P_j} + \phi_j\right) \\
m_0 &\sim \quad \text{Uniform}(-50, 50) \\
\log \sigma &\sim \quad \text{Uniform}(-1.5, -0.5) \\
\log A &\sim \quad \text{Uniform}(-2, 2) \\
\phi &\sim \quad \text{Uniform}(0, 2\pi) \\
\log P &\sim \quad \text{Uniform}(-1, 3)
\end{aligned}
$$

Figure B.1 presents a graph (t_i, d_i) of $M = 40$ data points, generated by sampling $n_{\text{comp}} = 2$ components with offset $m_0 = -5$, measurement uncertainty $\sigma = 0.1$, amplitudes $A_1 = 0.25$, $A_2 = 0.05$, periods $P_1 = 180$, $P_2 = 44$ and phases $\phi_1 = 0$, $\phi_2 = 1$ rad. The same data are analysed with $n_{\text{comp}} = 0, 1, 2, 3$, producing two to eleven dimensional problems with completely exchangable components.

Appendix B.2: Low X-ray count observations of a Compton-thick AGN

This mock problem describes a 5-dimensional degenerate physical parameter space discussed in Buchner et al. (2014). TODO

Appendix C: Toy problems

In this section, the prior is the unit hypercube ($0 < \theta_i < 1$ for $1 \leq i \leq d$), unless specified otherwise. If analytically known, the marginal parameter posterior distribution $p(\theta_i|D)$ and marginal likelihood Z are given.

Appendix C.1: Asymmetric Gaussian

Integration of a Gaussian distribution is a standard test problem. The variation here introduces some parameter inequality and spreads the means in a sine pattern, so that the posterior is not centred on the prior range.

$$L = \prod_{i=1}^{d} \mathrm{Normal}(\mu_i, \sigma_i^2)$$

$$\sigma_i = 0.1 \times 10^{-\left(9 - \frac{\sqrt{d}}{2}\right) \times \frac{i-1}{d-1}}$$

$$\mu_i = \frac{1}{2} + \frac{1 - 5\sigma_i}{2} \times \sin \frac{i-1}{2d}$$

This problem is evaluated using uniform priors on $d = 4$ (making σ_i range from 10^{-9} to 10^{-1}), 16 ($10^{-8} < \sigma_i < 10^{-1}$) and 100 dimensions ($10^{-5} < \sigma_i < 10^{-1}$). The four-dimensional case is shown in the top left panel of Figure 1. Note the different axes ranges. The true posterior is $p(\theta_i|D) = \mathrm{Normal}(\mu_i, \sigma_i^2)$, and the marginal likelihood $Z \approx 1$.

Appendix C.2: Correlated Funnel

Neil's funnel is a standard test problem that represents features of hierarchical Bayesian models. It is a normal distribution with the standard deviation also a free parameter. This causes a funnel shape (see middle panel of Figure 1) involving all parameters. Such non-affine correlations can sometimes be eased significantly by reparametrizations which scale the unknown mean parameters by the standard deviation parameter Betancourt & Girolami (2013). Here we use the correlated version of Karamanis & Beutler (2020):

$$L = \prod_{i} \mathrm{Normal}(\mu_i - \gamma \times \mu_{i-1}, \Sigma^2)$$

$$\ln \sigma \sim \mathrm{Normal}(0, 1)$$

$$\mu_i \sim \mathrm{Uniform}(-100, 100)$$

$$\Sigma_{ij} = \begin{cases} \sigma & \text{if } i = j \\ \gamma \times \sigma & \text{otherwise} \end{cases}$$

This problem is tested in 2, 10 and 50 dimensions with correlation strength $\gamma = 0.95$. The true marginal posteriors are $p(\mu_i|D) = \mathrm{Normal}(0, 1)$, $p(\ln \mu_i|D) = \mathrm{Normal}(0, 1)$, and the marginal likelihood $Z \approx 1$.

Appendix C.3: Rosenbrock function

The Rosenbrock function is a standard test problem in optimization. It exhibits a non-linear, narrow degeneracy that can be difficult to navigate (right middle panel of Figure 1). We adopt a probabilistic formulation suggested by Brewer (2016) (see a similar version in ?):

$$\log L = -2 \times \sum_{i=1}^{d-1} 100 \times \left(\theta_{i+1} - x_i^2\right)^2 + (1 - \theta_i)^2$$

$$\theta_i \sim \mathrm{Uniform}(-10, 10)$$

We test this problem in 2, 20 and 50 dimensions.

Appendix C.4: Eggbox

The eggbox function is a two-dimensional extremely multi-modal function (bottom left panel of Figure 1) proposed by Feroz & Hobson (2008), defined as:

$$\log L = \left(2 + \cos(5\pi \cdot \theta_1) \cdot \cos(5\pi \cdot \theta_2)\right)^5$$

$$\theta_i \sim \mathrm{Uniform}(0, 10\pi)$$

Appendix C.5: Beta product

Diverse test problems can be generated by combining standard one-dimensional probability distributions. Here, the Beta distribution is used to represent diverse posterior shapes in each parameter:

$$L = \prod_{i=1}^{d} \mathrm{Beta}(a_i, b_i)$$

with fixed, known a,b, randomly generated as:

$$\log a_i \sim \mathrm{Uniform}(-1, 1)$$

$$\log b_i \sim \mathrm{Uniform}(-1, 1)$$

This distribution can produce multiple modes (where $a_i < 1$ and $b_i < 1$), non-Gaussian tails. The likelihood is relatively uninformative on each parameter (completely non-informative when $a_i = b_i = 0$). We test this problem in 2, 10 and 30 dimensions. The true marginal posteriors are given by $P(\theta_i|D) = \mathrm{Beta}(a_i, b_i)$, and the marginal likelihood is $Z = 1$.

Appendix C.6: LogGamma

The LogGamma problem Beaujean & Caldwell (2013) exhibits multi-modality and heavy tails, which lead to non-elliptical contours.

$$g_a \sim \mathrm{LogGamma}\left(1, \frac{1}{3}, \frac{1}{30}\right)$$

$$g_b \sim \mathrm{LogGamma}\left(1, \frac{2}{3}, \frac{1}{30}\right)$$

$$n_c \sim \mathrm{Normal}\left(\frac{1}{3}, \frac{1}{30}\right)$$

$$n_d \sim \mathrm{Normal}\left(\frac{2}{3}, \frac{1}{30}\right)$$

$$d_i \sim \mathrm{LogGamma}\left(1, \frac{2}{3}, \frac{1}{30}\right) \quad \text{if } 3 \leq i \leq \frac{d+2}{2}$$

$$d_i \sim \mathrm{Normal}\left(\frac{2}{3}, \frac{1}{30}\right) \quad \text{if } \frac{d+2}{2} < i$$

$$L_1 = \frac{1}{2}\left(g_a(x_1) + g_b(x_1)\right)$$

$$L_2 = \frac{1}{2}\left(n_c(x_2) + n_d(x_2)\right)$$

$$L = L_1 \times L_2 \times \prod_{i=3}^{d} d_i(x_i)$$

We test this problem in 2, 10 and 30 dimensions. The true marginal posteriors are given by $P(x_1|D) = L_1(x_1)$, and the marginal likelihood is $Z = 1$.