Workshop at the 7th Symposium on Advances in Approximate Bayesian Inference (non-archival), 2025 1–10

Learning Likelihood-Free Reference Priors

Nicholas Bishop NICHOLAS.BISHOP@CS.OX.AC.UK
Joel Dyer JOEL.DYER@CS.OX.AC.UK
Daniel Jarne Ornia DANIEL.JARNEORNIA@CS.OX.AC.UK
Anisoara Calinescu
Michael Wooldridge
Department of Computer Science, University of Oxford

Abstract

Simulation modeling offers a flexible approach to constructing high-fidelity synthetic representations of complex real-world systems. However, the increased complexity of such models introduces additional complications when carrying out statistical inference procedures. This has motivated a large and growing literature on *likelihood-free* or *simulation-based* inference methods, which approximate (e.g., Bayesian) inference without assuming access to the simulator's intractable likelihood function. A hitherto neglected problem in the simulation-based Bayesian inference literature is the challenge of constructing uninformative *reference priors* for complex simulation models. Such priors maximise an expected Kullback-Leibler divergence from the prior to the posterior, thereby influencing posterior inferences minimally and enabling an "objective" approach to Bayesian inference that do not necessitate the incorporation of strong subjective prior beliefs. In this paper, we propose and test a selection of likelihood-free methods for learning reference priors for simulation models, using variational approximations and a variety of mutual information estimators. Our experiments demonstrate that good approximations to reference priors for simulation models are in this way attainable, providing a first step towards the development of likelihood-free objective Bayesian inference procedures.

1. Introduction

Simulation models have played a crucial role across a range of scientific disciplines including epidemiology (Kerr et al., 2021), economics (Dyer et al., 2024; Wiese et al., 2024) and robotics (Todorov et al., 2012). More generally, simulation models can be used to explore intricate dynamics, test hypotheses, and make predictions about real-world phenomena. However, simulation models *often lack an analytically tractable likelihood function*, preventing the direct application of classical statistical inference methods to learn simulator parameters. In response to this challenge, a variety of inference procedures have been developed by the simulation-based inference (SBI) community, ranging from classical approaches such as Approximate Bayesian Computation (Beaumont, 2019), to modern methods based in density (and density ratio) estimation (Thomas et al., 2016; Papamakarios and Murray, 2016; Hermans et al., 2020; Greenberg et al., 2019). Each of these procedures are united by the common assumption that the likelihood function cannot be evaluated, but only sampled from via forward-simulation.

More formally, consider a simulator which, given parameters $\theta \in \Theta \subseteq \mathbb{R}^d$, produces a random output supported on an output space $\mathcal{X}$. That is, the simulator implicitly defines

a density p_θ over outputs for each parameter value $\theta \in \Theta$. Given a prior distribution π over model parameters as well as observed iid data, $x_{1:n} = (x_1, \ldots, x_n)$, generated by the simulator's real-world counterpart, the goal of Bayesian simulation-based inference methods is to approximate the posterior distribution given by Bayes' Theorem:

$$\pi_{x_{1:n}}(\theta) \propto p_\theta(x_{1:n})\pi(\theta). \tag{1}$$

As discussed above, such inference must be performed under the assumption that p_θ cannot be evaluated, but instead readily sampled from via forward-simulation.

The prior, π, is intended to encapsulate pre-existing knowledge or beliefs about the plausible parameter values, whilst the posterior $\pi_{x_{1:n}}$ encapsulates updated beliefs about plausible values for θ having observed the real-world data $x_{1:n}$. However, in many practical scenarios, strong prior information regarding likely values for model parameters may be unavailable or unreliable (Jarne Ornia et al., 2024). Even when prior beliefs exist and can be readily expressed in the form of a prior distribution, they may not be confidently held. As a result, the modeller may wish to carry out Bayesian inference in a way that minimises the influence of their own prior beliefs, and determine the posterior that emerges when minimal prior information is encoded into the inference procedure.

Such considerations have motivated the development of *objective* Bayesian methods, which offer a principled approach to constructing priors that are minimally informative and, correspondingly, posteriors that are maximally data-driven. In particular, reference priors (Bernardo, 1979; Berger and Yang, 1994; Bernardo, 1997; Berger et al., 2009) have emerged as a prominent choice of uninformed prior. In short, a reference prior maximizes the expected information gain from observed data and minimizes the influence of the prior on the posterior in an information-theoretic sense. More formally, the n-reference prior π_n^* associated with a simulator and a class of plausible prior distributions is Π is given by:

$$\pi_n^* = \arg\max_{\pi \in \Pi} \mathbb{E}_{\substack{\theta \sim \pi \\ x_{1:n} \sim p_\theta}} \left[\log \frac{\pi_{x_{1:n}}(\theta)}{\pi(\theta)} \right]. \tag{2}$$

The objective on right-hand side of Equation (2) is equivalent to the mutual information $I_\pi(x_{1:n}, \theta)$ between $x_{1:n}$ and $\theta \sim \pi$. Reference priors have several desirable properties, including invariance under reparameterization and good frequentist coverage (Consonni et al., 2018). In addition, reference priors asymptotically achieve the minimax entropy risk when Π is the class of continuous priors on Θ (Clarke and Barron, 1994). Unfortunately, deriving reference priors analytically is often difficult and completely intractable for anything aside from simple models. This issue is further exacerbated in the context of SBI, since unavailability of p_θ precludes direct evaluation of the expectation in Equation (2).

In this paper, we address this problem by proposing and testing two novel likelihood-free approaches for learning reference priors. The first approach, based on the entropy decomposition of the mutual information (Kozachenko and Leonenko, 1987; Kraskov et al., 2004), makes no further assumptions on simulator behavior. Meanwhile our second approach assumes differentiability, and exploits this to maximize variational lower bounds on the mutual information (Oord et al., 2018; Song and Ermon, 2020; Letizia et al., 2023). We demonstrate our approach on several simple use cases where the asymptotic reference prior is known.

2. Methods

Before presenting each of our methods in detail, we first outline the general structure shared by all methods. Recall that Equation (2) may be expressed as an optimisation problem with objective corresponding to the mutual information $I_\pi(x_{1:n}, \theta)$ between θ and a set of model outputs $x_{1:n}$ sampled from p_θ. Each of our methods involve computing an estimate $\hat{I}$ of the mutual information, which is in turn used as an optimization objective to train a variational prior. This general procedure is outlined by Figure 1.

In words, a batch of B parameters $\{\theta^{(b)}\}_{b=1}^B$ is first sampled from a variational prior with tractable density π_ϕ and parameters ϕ. A corresponding set of model outputs $x_{1:n}^{(b)}$ is generated for each parameter by forward-simulation. The outputs are then encoded into a low-dimensional approximate sufficient statistic using an encoder s_φ with parameters φ. The encoded outputs are paired with their corresponding parameter value and passed to an estimator which returns an approximation $\hat{I}$ of the mutual information that is used to update the respective parameters of the variational prior π_ϕ and the encoder s_φ. The nature of this update depends on the structure of the mutual information estimator.

In what follows, we present two different classes of mutual information estimators which correspond to two different schemes for learning reference priors. As Figure 1 indicates, we assume that the variational prior is parameterised by a normalizing flow. For the sake of brevity, we use $h_\pi(x_{1:n}, \theta) = p_\theta(x_{1:n})\pi(\theta)$ to denote the joint distribution of $(x_{1:n}, \theta)$ under the prior π. Likewise, we use $m_\pi(x_{1:n})$ to define the corresponding marginal distribution over model outputs.

2.1. Generative Difference of Entropy Estimators

The first of our methods relies on the classical entropy decomposition of the mutual information:

$$I_\pi(x_{1:n}, \theta) = \mathbb{H}_\pi[\theta] - \mathbb{E}_{x_{1:n}\sim m_\pi}\mathbb{H}_{\pi_{x_{1:n}}}[\theta] \tag{3}$$

By estimating each entropy term in Equation (3), we may construct an estimate of the mutual information. Given our assumption that the variational prior has a tractable density $\pi_\phi(\theta)$, $\mathbb{H}_{\pi_\phi}[\theta]$ can be unbiasedly estimated via Monte-Carlo:

$$\mathbb{H}_{\pi_\phi}[\theta] \approx -\frac{1}{B}\sum_{b=1}^{B} \log \pi_\phi(\theta^{(b)}). \tag{4}$$

Thus, we shift our focus to estimating the second term of Equation (3). Taking inspiration from Pichler et al. (2022), we construct a conditional density estimator $\pi^\psi_{x_{1:n}}$ for the posterior $\pi_{x_{1:n}}$. We parameterize $\pi^\psi_{x_{1:n}}$ with a conditional normalizing flow, which is conditioned on $s_\varphi(x_{1:n})$. Using $\pi^\psi_{x_{1:n}}$, we may estimate the second term of Equation (3) as follows:

$$\mathbb{E}_{x_{1:n}\sim m_\pi}\mathbb{H}_{\pi_{x_{1:n}}}[\theta] \approx -\frac{1}{B}\sum_{b=1}^{B} \log \pi^\psi_{x_{1:n}^{(b)}}(\theta^{(b)}). \tag{5}$$

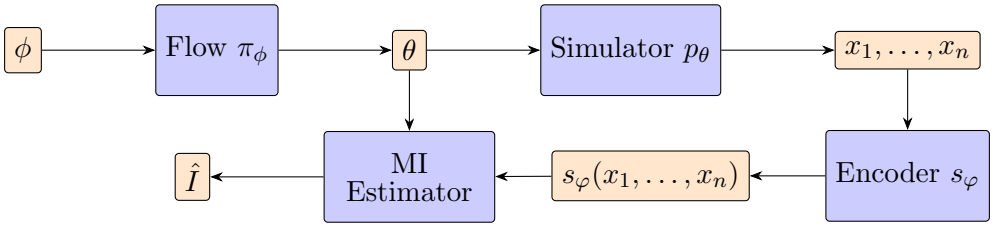


Figure 1: A diagram depicting the overall pipeline for the methods described in Section 2.

Combining Equations (4) and (5) we arrive at the following estimator for the mutual information:

$$\hat{I} = -\frac{1}{B}\sum_{b=1}^{B} \log \pi_\phi(\theta^{(b)}) + \frac{1}{B}\sum_{b=1}^{B} \log \pi^{\psi}_{x^{(b)}_{1:n}}(\theta^{(b)}). \tag{6}$$

To learn a reference prior, we may simultaneously update ϕ, φ and ψ via stochastic gradient ascent on $\hat{I}$. We refer to this method of learning a reference prior as the Generalized Entropy Difference Estimator (GED). Note that GED does not require the simulator to be differentiable.

2.2. Variational Lower Bound Methods

When the simulator is differentiable, we may also exploit variational lower bounds on the mutual information to learn an approximate reference prior. A wide range of variational lower bounds have been proposed in the literature (see Poole et al. (2019) for a thorough overview). Many such bounds rely on the following variational characterization of the KL-divergence (Donsker and Varadhan, 1975) between two distributions P and Q:

$$D_{\mathrm{KL}}(P\|Q) = \sup_{T \in L^\infty} \mathbb{E}_P[T] - \log \mathbb{E}_Q[e^T], \tag{7}$$

where L_∞ is the space of essentially bounded measurable functions. In particular, Belghazi et al. (2018) exploit Equation (7) by proposing MINE, which parameterises T with a neural network (commonly referred to as a *critic*) so that a reasonably tight lower bound on the mutual information can be learned via stochastic gradient ascent. Due to the second term of Equation (7), which corresponds to the log-partition function of Q, MINE suffers from high variance, especially when the mutual information is large (Song and Ermon, 2020; McAllester and Stratos, 2020). To mitigate this issue, Song and Ermon (2020) propose SMILE, which clips empirical approximations of the log-parition function to lie in the range $[-\tau, \tau]$, where $\tau > 0$ is a hyperparameter controlling the bias-variance trade-off associated with truncating the log-partition function.

In the context of contrastive predictive coding, the InfoNCE objective was proposed by Oord et al. (2018) for the purpose of estimating the density ratio of a joint distribution $P(X, Y)$ of random variables X and Y:

$$\sup_T \mathbb{E}_{P^n(X,Y)}\left[\frac{1}{n}\sum_{i=1}^{n} \log \frac{T(x_i, y_i)}{\frac{1}{n}\sum_{j\neq i} T(x_i, y_j)}\right], \tag{8}$$

where P^n denotes the n-fold product distribution associated with P. As observed by the original authors, InfoNCE implicitly maximises a variational lower bound on the mutual information between X and Y. In essence, the goal of the critic T is to accurately distinguish between samples drawn jointly from $P(X, Y)$ and samples drawn from the marginal product $P(X)P(Y)$.

Given a variational lower bound on the mutual information, we may jointly train a critic T and a variational prior π_ϕ to learn a reference prior. The variational prior is responsible for inducing high mutual information between $x_{1:n}$ and θ by maximising the variational lower bound produced by the critic. Meanwhile, the critic is tasked with keeping the variational lower bound tight so that the variational prior has a good approximation of the mutual information to benchmark against. Both networks may be simultaneously updated via stochastic gradient ascent. For instance, using InfoNCE, a critic T_μ with parameters μ, and a flow π_ϕ, we may proceed as follows:

1. Sample $\{(x_{1:n}^{(b)}, \theta^{(b)})\}_{b=1}^{B}$ from h_{π_ϕ}.

2. Compute the variational lower bound using critic T_μ:

$$\hat{I} = \frac{1}{B}\sum_{b=1}^{B} \log \frac{T_\mu(s_\varphi(x_{1:n}^{(b)}), \theta^{(b)})}{\frac{1}{B}\sum_{a\neq b} T_\mu(s_\varphi(x_{1:n}^{(b)}), \theta^{(a)})}.$$

3. Simultaneously update the parameters ϕ, φ and μ via stochastic gradient ascent on $\hat{I}$.

Note that there is a different version of the training scheme above for each variational approximation of the mutual information. During our experiments, we adopt both SMILE and InfoNCE as variational objectives, for several reasons. Many real-world simulators produce high dimensional outputs such as time series, leading to the possibility of high variance outputs. SMILE is naturally suited to this setting due to the hyperparameter τ that enables explicit management of the bias-variance trade-off. Meanwhile, InfoNCE typically exhibits low variance, since the optimal critic does not depend on batch size (Poole et al., 2019). This property is especially important for expensive simulators, since only smaller batch sizes can be used under limited simulation budgets.

3. Performing Simulation-based Inference

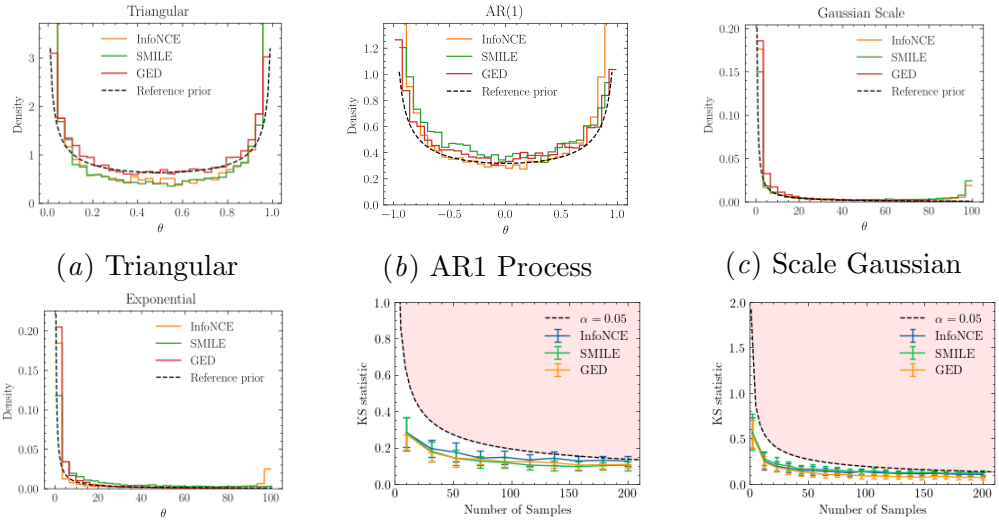
It is worth highlighting that, for each of the methods described in the Section 2, the ability to perform SBI using the learned reference prior π_ϕ comes at no further training cost. For instance, GED entails the construction of an amortised (in $x_{1:n}$) estimator $\pi_{x_{1:n}}^{\psi}(\theta)$ for the posterior density that results from the use of the learned prior π_ϕ, which can be immediately reused to generate posterior samples. Similarly, for the approaches outlined in Section 2.2, which are based on optimizing a variational lower bound of the mutual information between $x_{1:n}$ and θ, the learned discriminators T_μ estimate a function w of the density ratio $h_{\pi_\phi}(x_{1:n}, \theta)/m_{\pi_\phi}(x_{1:n})\pi_\phi(\theta)$. For example, in the case of SMILE, w is the identity. Since we have assumed that the variational reference prior density π_ϕ is tractable to evaluate,

estimates of the log-posterior density $\log \pi_{x_{1:n}}(\theta)$ resulting from the use of the prior π_ϕ may be obtained as

$$\log \pi_{x_{1:n}}(\theta) \approx \log w^{-1}(T_\mu(x_{1:n}, \theta)) + \log \pi_\phi(\theta). \tag{9}$$

Equation (9) can be used in, for example, an MCMC procedure to generate samples from the posterior $\pi_{x_{1:n}}$. In both cases, Bayesian SBI can be immediately performed with no further training of density ratio or posterior estimators.

4. Experiments



(a) Triangular (b) AR1 Process (c) Scale Gaussian

(d) Exponential Rate (e) KST (AR1) (f) KST (Triangular)

Figure 2: Comparison of proposed methods for learning reference priors on tractable models.

In this section, we present a series of experiments to assess the ability of the methods described in Section 2. More specifically, we evaluate each method on four simulators where the asymptotic reference prior is known. We consider a scale Gaussian model, an exponential rate model, a triangular model and an AR1 process model. Full details about each model and their corresponding asymptotic reference priors can be found in the Appendix.

Results are shown in Figures 2(a)-(d). The asymptotic reference prior is marked by a dashed line in each plot. Note that all of our methods closely match the asymptotic reference prior. To provide a quantitative analysis we compare the learned reference priors for both the triangular and AR1 models against the corresponding true asymptotic reference priors via a two-sample Komolgorov-Smirnov test (KS). Results are shown in Figures 2(e)-(f). The x-axis of each plot indicates the number of samples used in the KST whilst the y-axis denotes the value of the KS test statistic. In both cases all of our methods consistently pass the KST for up to 200 samples at a significance level of $\alpha = 0.05$, avoiding the failure regions shaded in red.

5. Conclusion

In this paper, we investigate several novel approaches to learning reference priors for arbitrary simulation models. Through experiments on tractable examples, we have shown that these methods can learn good reference priors for simple simulators. Testing our methods on more complex simulators, where the reference prior is not analytically available, forms an interesting direction for future work. Finally, reference priors are only one possible approach to conducting "objective" Bayesian inference, and other considerations can lead to alternative objective priors in Bayesian analysis (Consonni et al., 2018). The present work is a first step towards enabling objective Bayesian inference for complex simulation models, and developing likelihood-free methods for estimating other classes of objective priors constitutes an interesting direction for future work.

Acknowledgments

NB, JD, DJO, AC, and MW acknowledge funding from a UKRI AI World Leading Researcher Fellowship awarded to Wooldridge (grant EP/W002949/1). MW and AC also acknowledge funding from Trustworthy AI - Integrating Learning, Optimisation and Reasoning (TAILOR), a project funded by the European Union Horizon2020 research and innovation program under Grant Agreement 952215.

References

Mark A Beaumont. Approximate Bayesian computation. *Annual review of statistics and its application*, 6:379–403, 2019.

Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and Devon Hjelm. Mutual information neural estimation. In *Proceedings of the 35th International Conference on Machine Learning*, 2018.

Yoshua Bengio, Nicholas Léonard, and Aaron Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432*, 2013.

James O Berger and Ruo-Yong Yang. Noninformative priors and Bayesian testing for the AR (1) model. *Econometric Theory*, 10(3-4):461–482, 1994.

James O Berger, José M Bernardo, and Dongchu Sun. The formal definition of reference priors. *The Annals of Statistics*, 37(2):905–938, 2009.

Jose M Bernardo. Reference posterior distributions for bayesian inference. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 41(2):113–128, 1979.

Jose M Bernardo. Non-informative priors do not exist. *Journal of Statistical Planning and Inference*, 65(1):159–189, 1997.

Bertrand S Clarke and Andrew R Barron. Jeffreys' prior is asymptotically least favorable under entropy risk. *Journal of Statistical planning and Inference*, 41(1):37–60, 1994.

Guido Consonni, Dimitris Fouskakis, Brunero Liseo, and Ioannis Ntzoufras. Prior Distributions for Objective Bayesian Analysis. *Bayesian Analysis*, 13(2):627 – 679, 2018.

M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain markov process expectations for large time, i. *Communications on Pure and Applied Mathematics*, 28(1): 1–47, 1975.

Joel Dyer, Patrick Cannon, J Doyne Farmer, and Sebastian M Schmon. Black-box bayesian inference for agent-based models. *Journal of Economic Dynamics and Control*, 161: 104827, 2024.

David S. Greenberg, Marcel Nonnenmacher, and Jakob H. Macke. Automatic posterior transformation for likelihood-free inference. *36th International Conference on Machine Learning*, 2019.

Joeri Hermans, Volodimir Begy, and Gilles Louppe. Likelihood-free MCMC with amortized approximate ratio estimators. In *International Conference on Machine Learning*, 2020.

Daniel Jarne Ornia, Joel Dyer, Nicholas George Bishop, Ani Calinescu, and Michael J. Wooldridge. Model exploration through marginal likelihood entropy maximisation. In *NeurIPS 2024 Workshop on Data-driven and Differentiable Simulations, Surrogates, and Solvers*, 2024.

Cliff C Kerr, Robyn M Stuart, Dina Mistry, Romesh G Abeysuriya, Katherine Rosenfeld, Gregory R Hart, Rafael C Núñez, Jamie A Cohen, Prashanth Selvaraj, Brittany Hagedorn, et al. Covasim: an agent-based model of covid-19 dynamics and interventions. *PLOS Computational Biology*, 17(7):e1009149, 2021.

Lyudmyla F Kozachenko and Nikolai N Leonenko. Sample estimate of the entropy of a random vector. *Problemy Peredachi Informatsii*, 23(2):9–16, 1987.

Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 69(6):066138, 2004.

Nunzio A Letizia, Nicola Novello, and Andrea M Tonello. Variational f-divergence and derangements for discriminative mutual information estimation. *arXiv preprint arXiv:2305.20025*, 2023.

David McAllester and Karl Stratos. Formal limitations on the measurement of mutual information. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, 2020.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

George Papamakarios and Iain Murray. Fast ε-free inference of simulation models with bayesian conditional density estimation. *Advances in neural information processing systems*, 29, 2016.

Georg Pichler, Pierre Jean A Colombo, Malik Boudiaf, Günther Koliander, and Pablo Piantanida. A differential entropy estimator for training neural networks. In *International Conference on Machine Learning*, 2022.

Ben Poole, Sherjil Ozair, Aaron Van Den Oord, Alex Alemi, and George Tucker. On variational bounds of mutual information. In *Proceedings of the 36th International Conference on Machine Learning*, 2019.

Jiaming Song and Stefano Ermon. Understanding the limitations of variational mutual information estimators. In *International Conference on Learning Representations*, 2020.

Owen Thomas, Ritabrata Dutta, Jukka Corander, Samuel Kaski, and Michael U Gutmann. Likelihood-free inference by ratio estimation. *arXiv preprint arXiv:1611.10242*, 2016.

Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012.

Samuel Wiese, Jagoda Kaszowska-Mojsa, Joel Dyer, Jose Moran, Marco Pangallo, Francois Lafond, John Muellbauer, Anisoara Calinescu, and J Doyne Farmer. Forecasting macroeconomic dynamics using a calibrated data-driven agent-based model. *arXiv preprint arXiv:2409.18760*, 2024.

Ruoyong Yang and James O Berger. *A catalog of noninformative priors*, volume 1018. Institute of Statistics and Decision Sciences, Duke University Durham, NC, USA, 1996.

Appendix A. Benchmark Models

Here we provide further details about each of the models used in our experiments.

Gaussian Scale Model. The first tractable example we consider simulates Gaussian random variables. For this model, n samples are generated iid from $\mathcal{N}(\mu, \sigma^2)$ as $x_t = \mu + \theta u_t$, with $u_t \sim \mathcal{N}(0,1)$, and where $\mu \in \mathbb{R}$ is known, $\theta > 0$ is a free parameter, and $\mathcal{N}(a, b^2)$ is a Normal distribution with mean a and variance b^2. Here, the reference prior (Yang and Berger, 1996) for θ is $\pi^*(\theta) \propto 1/\theta$.

Exponential Rate Model. We next consider an exponential model, whose generative process is as follows: for $t = 1, \ldots, n$, we generate random variables x_t from an $\mathrm{Exp}(\theta)$ density as $x_t = -\log(1-u_t)/\theta$, where $u_t \sim U(0,1)$, $\theta > 0$ is a parameter, and $U(a,b)$ is a uniform distribution on $[a,b]$. As with the Gaussian scale model, the reference prior for θ is known (Yang and Berger, 1996) to be $\pi^*(\theta) \propto 1/\theta$.

Triangular Model. In this example, we generate random variables from a triangular distribution on $[0,1]$. Here, iid data is generated for $t = 1, \ldots, n$ as

$$x_t = \mathbb{I}[u_t \leq \theta] \sqrt{\theta \cdot u_t} + (1 - \mathbb{I}[u_t \leq \theta]) \left(1 - \sqrt{(1-u_t)(1-\theta)}\right), \tag{10}$$

where $\theta \in (0,1)$ is a free parameter and $u_t \sim U(0,1)$ is a random variable distributed uniformly on $[0,1]$. The reference prior is known (Berger et al., 2009) to be well-approximated

by a Beta$(1/2, 1/2)$ distribution. Though the derivative of x_t with respect to θ for fixed u_t is 0 almost everywhere when defined in the usual sense, we may nonetheless define an approximate surrogate gradient through, e.g., the straight-through gradient trick (Bengio et al., 2013) in order to backpropagate through x_t. As such, we may continue to apply the methods described in Section 2.2, which require a differentiable simulator.

Autoregressive Time-series Model. Finally, we consider the standard autoregressive time-series model of order 1 (AR(1)). Using $u_t \sim \mathcal{N}(0, 1)$, $t = 1, \ldots, n$, this model generates a time-series $x_1, \ldots, x_n$ as

$$x_1 = \sigma u_1, \quad \text{and} \quad x_t = \theta x_{t-1} + \sigma u_t \ \text{ for } \ t = 2, \ldots, n, \tag{11}$$

where $\sigma > 0$ is fixed and $\theta \in [-1, 1]$ is a free parameter. It can be shown (Berger and Yang, 1994) that the corresponding reference prior for $\theta \in [-1, 1]$ is $\pi^*(\theta) \propto \left(1 - \theta^2\right)^{-1/2}$.