Amortized Variational Inference for Type Ia Supernova Light Curves

Alexis Sánchez
Department of Computer Science, University of Concepción
Millenium Institute of Astrophysics
alexisanchez@inf.udec.cl

Guillermo Cabrera
Department of Computer Science, University of Concepción
Millenium Institute of Astrophysics

Pablo Huijse
Informatics Institute, Universidad Austral de Chile
Millenium Institute of Astrophysics

Francisco Förster
Data & Artificial Intelligence Initiative / Center for Mathematical Modeling, University of Chile
Millenium Institute of Astrophysics

Abstract

Markov Chain Monte Carlo (MCMC) methods are widely used for Bayesian inference in astronomy. However, when applied to datasets coming from next-generation telescopes, inference becomes computationally expensive. We propose using amortized variational inference to estimate the posterior of a supernova light curve parametric model. We show that amortization with a recurrent neural network is significantly faster than MCMC while providing competitive estimates of the predictive distribution. To the best of our knowledge, this is the first time this fast amortized framework is applied to supernova light curves. This approach will be essential when estimating the posterior of astrophysical parameters for terabytes of data per night that next-generation telescopes will produce.

1 Introduction

Supernovae (SNe) are highly energetic explosions that occur at the end of the life cycle of stars. These cataclysmic events can be used as extragalactic distance probes and hence are key to further advance cosmological theories [22, 20]. Parametric models are often used to aid the distance determination [10], classification [24], and the physical characterization of SNe [9]. One desired property when performing inference with these models is dealing with the uncertainty present in the data and predictions. Having uncertainties of our estimates is critical for the scientific interpretation of the data, e.g., for the selection of well characterized samples for detailed analysis, for understanding the level of significance of any derived conclusions, or for the reliable identification of extreme or outlier events based on their physical parameters.

Through the use of Bayesian methods we can obtain uncertainty estimates both for the parameters of the model and the predictions obtained by it. However, for complex models the posterior of the parameters may not be analytically tractable. Two possible solutions are the use of Markov Chain

Fourth Workshop on Machine Learning and the Physical Sciences (NeurIPS 2021).

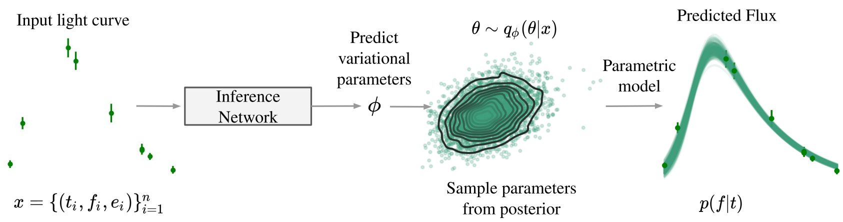

Figure 1: Amortized variational inference for supernova light curves. The inference networks uses light curves as input, consisting of the time since the first observation t_i, normalized flux f_i and measured error e_i and outputs the variational parameters ϕ for the approximate posterior $q_\phi(\theta|x)$. Using samples from this posterior we can sample realizations using the parametric model.

Monte Carlo (MCMC) [18, 5] methods to sample from the posterior and Variational Inference (VI) [15, 4] which optimizes the parameters of an approximate posterior. Both methods present difficulties in scaling to large datasets because applying them for individual light curves is computationally expensive for the entire dataset.

Instead, amortized variational inference (AVI) [17, 21], leverages powerful function approximators, such as deep neural networks, to map the data to the variational parameters directly. This means that after an offline optimization step for the approximator, amortization allows for fast online parameter inference by using the trained approximator. Furthermore, through techniques such as stochastic variational inference (SVI) [12] this approach can scale efficiently to large datasets and even provide inference in batches.

In this paper, we aim at performing fast Bayesian inference on the observed optical flux evolution of astronomical objects, or *light curves*, in preparation for large astronomical surveys such as the Vera C. Rubin Observatory and its Legacy Survey of Space and Time (LSST) [14]. Specifically, we learn a recurrent neural network to perform amortized variational inference over the posterior distribution of a parametric model for supernova light curves.We evaluate our approach on synthetic and real alert light curves from the Zwicky Transient Facility (ZTF) [2], a precursor to the LSST, and compare it against MCMC, both in terms of the estimated posteriors and the computational cost required.

2 Amortized Variational Inference

Consider a parametric model $f(x; \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ represents the set of adjustable parameters. We consider performing fast approximate Bayesian inference for $f(x; \boldsymbol{\theta})$. Variational Bayesian methods allow us to approximate the posterior of this model by introducing a variational distribution $q_\phi(\boldsymbol{\theta})$ with parameters ϕ. The parameters ϕ are found by maximizing the evidence lower bound (ELBO):

$$ELBO = E_q[\log p(x|\theta)] - D_{KL}(q_\phi(\theta)||p(\theta)), \tag{1}$$

which, when using standard variational inference would require performing optimization for each light curve independently . Instead in AVI we introduce a function $g(\boldsymbol{x})$ that maps data x to the variational parameters ϕ. This means that instead of learning local parameters for each data point we learn the global parameters of our function g. The introduction of the amortization procedure means we spend an upfront computational cost in learning $g(\boldsymbol{x})$. However, once this function is learned, performing Bayesian inference is reduced to evaluating this function with new observations to obtain the variational posterior $q_\phi(\boldsymbol{\theta}|\boldsymbol{x})$. For modeling g, neural networks are commonly used, since they are flexible function approximators [7] which are capable of modeling the unknown relationship between data and variational parameters. We apply AVI to infer the parameters of a Sne model from ZTF alert light curves. The general procedure is illustrated in Figure 1, where the inference network maps the light curve data to the variational parameters ϕ of the approximate posterior. After obtaining $q_\phi(\theta|x)$ we sample θ and reconstruct the input using a fixed parametric SNe model. The inference network is trained by maximizing the ELBO using SVI [12] which enables scaling to larger datasets.

3 Methodology

3.1 Generative model

We consider the widely used parametric model introduced by [1]:

$$f(t) = A \frac{e^{\frac{-(t-t_0)}{\tau_{fall}}}}{1 + e^{\frac{-(t-t_0)}{\tau_{rise}}}}, \tag{2}$$

where $\theta = (A, t_0, \tau_{rise}, \tau_{fall})$ correspond to the parameters of the model. The A parameter is a scaling factor to accommodate the maximum brightness observed. The parameter t_0 corresponds to the time at which the brightness equals $A/2$. Finally, the parameters τ_{rise} and τ_{fall} represent the rise and fall characteristic timescales of the light curve, respectively. We represent τ_{rise} as a fraction of τ_{fall} with a parameter γ such that $\tau_{rise} = \gamma\tau_{fall}$. All parameters are estimated after the light curve is divided by the maximum flux and time is measured with respect to the time of the first detection.

We assume a normal distribution for the likelihood $p(x|\boldsymbol{\theta})$, with mean given by Eq. 2 and standard deviation known from the error measurements of the light curves. We assume independent priors based on astronomical knowledge for $p(\theta)$:

$$\begin{aligned} \log(A) &\sim \mathcal{N}(0.397, 0.133) & \quad \operatorname{logit}(\gamma) &\sim \mathcal{N}(-3, 1) \\ t_0 &\sim \mathcal{N}(0, 10) & \quad \log\tau_{fall} &\sim \mathcal{N}(3.47, 0.41) \end{aligned} \tag{3}$$

This parametrization forces the non-negativity of A and τ_{fall}, $0 \leq \gamma \leq 1$, and allows for t_0 to have negative values (no observations previous to peak). After exponentiating τ_{fall} we add three to constraint its minimum value. Note that we use the same priors on both MCMC and AVI.

3.2 Inference model

For our variational approximation we use a multivariate normal distribution over $\log(A)$, t_0, $\log(\tau_{fall})$ and $\operatorname{logit}(\gamma)$: $q_\phi(\boldsymbol{\theta}|\boldsymbol{x}) = \mathrm{MvN}(\mu, LL^T)$, where $\phi = (\mu, L)$, $\mu \in \mathbb{R}^4$ and L is a 4x4 lower triangular matrix. The parameters ϕ are predicted by the inference network, for which we use a Long Short Term Memory Network (LSTM) [11]. By using a recurrent neural network we can take into account the variable length and irregular sampling of light curves.

3.3 Experimental Setup

We use both real and simulated light curves for our experiments. The real data consists of a set of 709 supernova light curves from the alert stream of ZTF, downloaded using the API of the ALeRCE [8, 23] astronomical broker[1]. We use only the 'g' filter and light curves with at least 6 reported alerts and at least 30 days between the first and last observation. The simulated light curves corresponds to a set of $10,000$ simulations based on the priors used and the real data. Both real and simulated light curves are normalized before inference by dividing the fluxes and errors by the maximum flux. We compare our amortized approach with the No U-Turn MCMC sampler (NUTS) [13] from the Stan [6] package. For each light curve in the dataset we run 4 parallel chains for $6,000$ iterations discarding the first $2,000$ samples.

The neural network for the amortization is implemented in PyTorch [19] and we use SVI [12] with Pyro [3]. The network is trained for $1,000$ epochs using the Adam [16] optimizer with a learning rate of 5×10^{-4} and a batch size of 64, additionally we clip the norm of the gradient to 10. We employ early stopping during our training, where we set the number of epochs with no validation improvement before stopping to 50. Experiments were conducted both on CPU and GPU to measure execution time. The CPU experiments were conducted on a 32 core AMD EPYC processors and the GPU used for the other experiments corresponds to a NVIDIA V100.

[1] `https://alerceapi.readthedocs.io/en/latest/ztf_db.html`

	Synthetic data		Real data	
	MCMC	Amortized	MCMC	Amortized
mean	20.92	18.53	−3.05	−31.45
std	11.42	11.77	150.80	226.63
5%	9.41	8.12	−74.51	−178.84
50%	18.26	16.55	16.16	10.92
95%	41.02	36.99	60.52	42.51

Table 1: Statistics from the distribution of average log-likelihoods.

Method	Training	Inference
MCMC	–	$1,902.74 \pm 183.00$
AVI (CPU)	$1,005.10 \pm 37.35$	29.04 ± 0.84
AVI (GPU)	699.25 ± 29.12	3.79 ± 0.04

Table 2: Time in seconds to perform inference for $10,000$ simulated light curves. The reported time corresponds to the mean and standard deviation of 5 runs.

Parameter	MCMC	Amortized
A	0.11	0.14
t_0	1.25	1.73
τ_{fall}	2.70	4.63
τ_{rise}	0.98	1.16

Table 3: Median absolute deviation between the median marginal posterior and the parameter used for the simulation.

4 Results

We evaluate the average log-likelihood for both synthetic and real light curves when the parameters are sampled from the posterior distribution. The distribution of the average likelihood for the data is shown in table 1. We consistently see that the amortized posterior presents lower likelihood than MCMC. However, we can observe that for both synthetic and real data, the median log-likelihood between both approaches remains close.

Table 2 shows the average time required to train the model and infer the parameters in the $10,000$ synthetic light curves set using MCMC and AVI. Although training is slower, our approach is ~ 65 times faster when performing inference on CPU and ~ 500 times faster on GPU. Experiments for MCMC are performed using the CPU. This enables the use of approximate inference for future massive surveys, where new alerts arrive on a night by night basis.

For the simulated light curves, we compute the median absolute deviation (MAD) between the known parameter used in the simulation and the median marginal posterior. Table 3 shows that the amortized posterior presents higher MADs compared to MCMC as expected, notably, this error is higher in the t_{fall} parameter, indicating possibly worse fits for the decline of the light curve. This indicates that the decline of the light curves present worse fits than MCMC. On the other hand, for the parameters A and t_0 both approaches exhibit similar errors, with only a difference of 0.5 days for t_0.

Finally we compare samples from the predictive distribution for real light curves. This is shown in Figure 2, where the top and bottom rows correspond to draws from MCMC and AVI for five selected light curves, respectively. We observe that both methods provide visually similar fits, although AVI predictions are slightly more biased than MCMC. At the same time MCMC provides predictions with higher uncertainties than AVI.

5 Conclusions

With the increase in size of current and future astronomical datasets, classical techniques for Bayesian inference are presented with difficulties in scaling. In this work we used AVI to estimate the posterior of the physical parameters from supernova light curves. The amortized posterior can be obtained significantly faster than MCMC once the inference network is trained, enabling fast Bayesian inference for incoming astronomical surveys like the LSST [14]. As future work we will test this methodology using physical models of other types of transient astronomical objects.

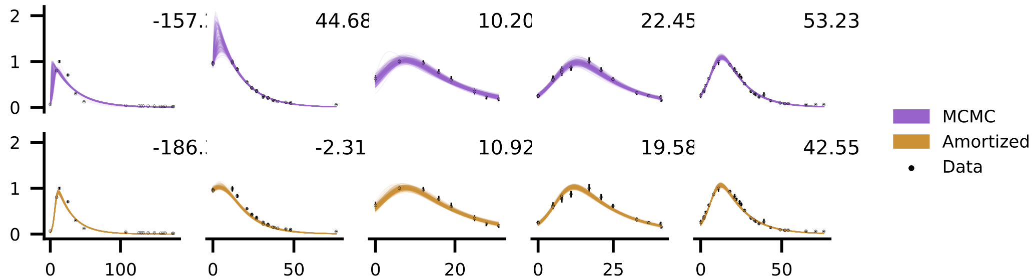

Figure 2: Random samples from the predictive distribution when sampling from the posterior for real data. Top row corresponds to MCMC posterior, bottom row are samples from the amortized posterior. Each example has the log-likelihood shown in the top right corner.

Broader Impact

We believe that the use of amortized inference serves as a first step towards performing Bayesian inference on demand for next generation surveys such as the LSST. For the particular case of Supernovae, having reliable physical parameters and uncertainties will help advance current studies on the physics of these objects and on the large-scale structure in the Universe. The proposed methodology can also benefit studies on other astronomical objects simply by changing the physical model that maps parameters to light curves and the priors.

The posterior distributions obtained using the proposed method can be used as an input to design follow-up rules, i.e. to trigger third-party telescopes to follow novel and/or interesting targets. For the particular case of fast transients follow-up should ideally start as soon as the alerts from the main survey arrive. Quick methods such as the ones proposed in this paper would avoid missing such follow-up targets in the massive streams of near future surveys.

Acknowledgments and Disclosure of Funding

We acknowledge support from CONICYT/ANID through the grants CONICYT-PFCHA/Magíster Nacional/2018-22181100 (A.S); FONDECYT Initiation Nº 11191130 (G.C.V.); FONDECYT Regular Nº 1211374 (P.H.); BASAL Center of Mathematical Modelling (AFB-170001, ACE210010 and FB210005) and FONDECYT Regular Nº 1200710 (F.F). This work has been partially funded by ANID -Millennium Science Initiative Program - ICN12_009 awarded to the Millennium Institute of Astrophysics (MAS). Powered@NLHPC: This research was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02). This work has been possible thanks to the use of AWS credits managed by the NLHPC.

References

[1] Bazin, G., Palanque-Delabrouille, N., Rich, J., Ruhlmann-Kleider, V., Aubourg, E., Le Guillou, L., Astier, P., Balland, C., Basa, S., Carlberg, R. G., Conley, A., Fouchez, D., Guy, J., Hardin, D., Hook, I. M., Howell, D. A., Pain, R., Perrett, K., Pritchet, C. J.G, Regnault, N., Sullivan, M., Antilogus, P., Arsenijevic, V., Baumont, S., Fabbro, S., Le Du, J., Lidman, C., Mouchet, M., Mourão, A., and Walker, E. S. The core-collapse rate from the supernova legacy survey. *A&A*, 499(3):653–660, 2009.

[2] E. C. Bellm, S. R. Kulkarni, M. J. Graham, R. Dekany, R. M. Smith, R. Riddle, F. J. Masci, G. Helou, T. A. Prince, S. M. Adams, et al. The zwicky transient facility: system overview, performance, and first results. *Publications of the Astronomical Society of the Pacific*, 131(995):018002, 2018.

[3] E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. A. Szerlip, P. Horsfall, and N. D. Goodman. Pyro: Deep universal probabilistic programming. *J. Mach. Learn. Res.*, 20:28:1–28:6, 2019.

[4] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

[5] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. *Handbook of markov chain monte carlo*. CRC press, 2011.

[6] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76(1), 2017.

[7] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

[8] F. Förster, G. Cabrera-Vives, E. Castillo-Navarrete, P. Estévez, P. Sánchez-Sáez, J. Arredondo, F. Bauer, R. Carrasco-Davis, M. Catelan, F. Elorrieta, et al. The automatic learning for the rapid classification of events (alerce) alert broker. *The Astronomical Journal*, 161(5):242, 2021.

[9] F. Förster, T. J. Moriya, J. C. Maureira, J. P. Anderson, S. Blinnikov, F. Bufano, G. Cabrera-Vives, A. Clocchiatti, T. de Jaeger, P. A. Estévez, L. Galbany, S. González-Gaitán, G. Gräfener, M. Hamuy, E. Y. Hsiao, P. Huentelemu, P. Huijse, H. Kuncarayakti, J. Martínez, G. Medina, F. Olivares E., G. Pignata, A. Razza, I. Reyes, J. San Martín, R. C. Smith, E. Vera, A. K. Vivas, A. de Ugarte Postigo, S. C. Yoon, C. Ashall, M. Fraser, A. Gal-Yam, E. Kankare, L. Le Guillou, P. A. Mazzali, N. A. Walton, and D. R. Young. The delay of shock breakout due to circumstellar material evident in most type II supernovae. *Nature Astronomy*, 2:808, Sept. 2018.

[10] J. Guy, P. Astier, S. Baumont, D. Hardin, R. Pain, N. Regnault, S. Basa, R. G. Carlberg, A. Conley, S. Fabbro, D. Fouchez, I. M. Hook, D. A. Howell, K. Perrett, C. J. Pritchet, J. Rich, M. Sullivan, P. Antilogus, E. Aubourg, G. Bazin, J. Bronder, M. Filiol, N. Palanque-Delabrouille, P. Ripoche, and V. Ruhlmann-Kleider. SALT2: using distant supernovae to improve the use of type Ia supernovae as distance indicators. *Astronomy & Astrophysics*, 466(1):11–21, Apr. 2007.

[11] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[12] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley. Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347, 2013.

[13] M. D. Hoffman and A. Gelman. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.

[14] Ž. Ivezić, S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, R. Allsman, D. Alonso, Y. AlSayyad, S. F. Anderson, J. Andrew, and et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *The Astrophysical Journal*, 873:111, Mar. 2019.

[15] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233, 1999.

[16] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[17] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[18] R. M. Neal. *Probabilistic inference using Markov chain Monte Carlo methods*. Department of Computer Science, University of Toronto Toronto, Ontario, Canada, 1993.

[19] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[20] S. Perlmutter, G. Aldering, G. Goldhaber, R. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, et al. Measurements of ω and λ from 42 high-redshift supernovae. *The Astrophysical Journal*, 517(2):565, 1999.

[21] D. J. Rezende, S. Mohamed, and D. Wierstra. Stochastic backpropagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*, 2014.

[22] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3):1009, 1998.

[23] P. Sánchez-Sáez, I. Reyes, C. Valenzuela, F. Förster, S. Eyheramendy, F. Elorrieta, F. E. Bauer, G. Cabrera-Vives, P. A. Estévez, M. Catelan, G. Pignata, P. Huijse, D. D. Cicco, P. Arévalo, R. Carrasco-Davis, J. Abril, R. Kurtev, J. Borissova, J. Arredondo, E. Castillo-Navarrete, D. Rodriguez, D. Ruz-Mieres, A. Moya, L. Sabatini-Gacitúa, C. Sepúlveda-Cobo, and E. Camacho-Iñiguez. Alert classification for the alerce broker system: The light curve classifier, 2020.

[24] V. A. Villar, E. Berger, G. Miller, R. Chornock, A. Rest, D. O. Jones, M. R. Drout, R. J. Foley, R. Kirshner, R. Lunnan, E. Magnier, D. Milisavljevic, N. Sanders, and D. Scolnic. Supernova Photometric Classification Pipelines Trained on Spectroscopically Classified Supernovae from the Pan-STARRS1 Medium-deep Survey. *The Astrophysical Journal*, 884(1):83, Oct. 2019.

Paper Checklist

(a) Do the **claims** made in the abstract and introduction accurately reflect the paper's contributions and scope?
Yes.

(b) Have you read the ethics review guidelines and ensured that your paper conforms to them?
Yes.

(c) Did you discuss any potential **negative societal impacts** of your work?
No.

(d) Did you describe the **limitations** of your work?
Yes.

(e) Did you include the code, data, and instructions needed to **reproduce** the main experimental results (either in the supplemental material or as a URL)?
No.

(f) Did you specify all the **training details** (e.g., data splits, hyperparameters, how they were chosen)?
Yes, the Experimental Setup section details how MCMC was run as well as hyperparameter for the inference network training such as batch size, learning rate, number of epochs, etc.

(g) Did you report **error bars** (e.g., with respect to the random seed after running experiments multiple times)?
No, only mean and standard deviation is provided for the timing results on Table 2. The rest of the results correspond to a single experiment. However, results in Table 1 show different statistics for the average log-likelihood over the dataset.

(h) Did you include the amount of **compute** and the type of **resources** used (e.g., type of GPUs, internal cluster, or cloud provider)? Yes, a partial description of the hardware used is provided in the Experimental Setup section. Specifically, the number of CPU cores and the GPU model used for experiments.

(i) If your work uses existing assets, did you cite the creators?
Yes, we use data from the ALeRCE broker[8] as specified in Experimental Setup and provide a URL.

(j) Did you mention the license of the assets?
No.

(k) Did you include any new assets either in the supplemental material or as a URL?
No.

(l) Did you discuss whether and how consent was obtained from people whose data you're using/curating?
No, the data is publicly available.

(m) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content?
No.

Appendix

The architecture used for our inference network is shown in Figure 3, we use an LSTM to process the light curve data and then two fully connected networks of two layers each to estimate the parameters of the multivariate normal approximation.

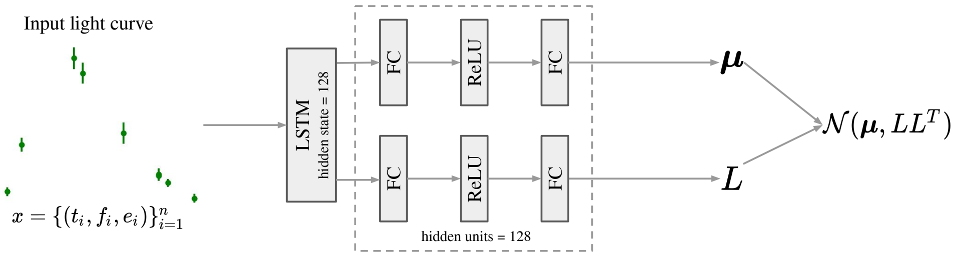

Figure 3: Inference network architecture.