Supplementary Material to:

Partially Exchangeable Networks and Architectures for Learning Summary Statistics in Approximate Bayesian Computation

Contents

1 Approximate Bayesian computation rejection sampling

1.1 Settings for ABC rejection sampling "reference table" algorithm

In section 2 of the main paper we denote with x the ABC threshold. For g-and-k and α-stable models we consider for x the 0.1th percentile, and for AR(2) and MA(2) the 0.02th percentile of all distances. The number of proposals for g-and-k and α-stable models is $\tilde{N} = 100,000$, and for AR(2) and MA(2) $\tilde{N} = 500,000$.

1.2 The ABC distance function

In all our inference attempts we always used ABC rejection sampling and only needed to change the method used to compute the summary statistics. We employed the Mahalanobis distance

$$\Delta(s^*, s^{\text{obs}}) = \sqrt{(s^* - s^{\text{obs}})^\top A (s^* - s^{\text{obs}})},$$

where in our case A is the identity matrix, except when using *hand-picked* summary statistics for the g-and-k distribution, and in such case A is a diagonal matrix with diagonal elements $1/w^2$, with w a vector with entries $w = [0.22; 0.19; 0.53; 2.97; 1.90]$, as in Picchini and Anderson [2017].

2 Regularization

We use early-stopping for all networks. The early-stopping method used is to train the network over N epochs and then select the set of weights, out of the N sets, that generated the lowest evaluation error.

3 g-and-k distribution

- The full set of parameters for a g-and distribution is $[A, B, g, k, c]$, However, we follow the common practice of keeping c fixed to $c = 0.8$ and assume $B > 0$ and $k \geq 0$ Prangle [2017].
- Here is a procedure to simulate a single draw from the distribution: we first simulate a draw z from a standard Gaussian distribution, $z \sim N(0, 1)$, then we plug z into

$$Q = A + B \cdot (1 + c \cdot \tanh(g \cdot z/2)) \cdot z \cdot (1 + z^2)^k$$

  and obtain a realization Q from a g-and-k distribution.
- The network settings are presented in Table 1, 2, 3, and 4;
- The number of weights for the different networks are presented in Table 5;
- Values outside of the range $[-10, 50]$ are considered to be outliers and these values are replaced (at random) with values inside the data range. The data cleaning scheme is applied to both the observed and generated data;
- When computing the empirical distribution function we evaluate this function over 100 equally spaced points between 0 and 50;
- Number of training observations: $5 \cdot 10^5$, 10^5, 10^4, and 10^3. Evaluation data observations $5 \cdot 10^3$.

Table 1: g-and-k: Network settings for MLP small.

Layer	Dim. in	Dim. out	Activation
Input	1000	25	relu
Hidden 1	25	25	relu
Hidden 2	25	12	relu
Output	12	4	linear

Table 2: g-and-k: Network settings for MLP large.

Layer	Dim. in	Dim. out	Activation
Input	1000	100	relu
Hidden 1	100	100	relu
Hidden 2	100	50	relu
Output	50	4	linear

Table 3: g-and-k: Network settings MLP pre

Layer	Dim. in	Dim. out	Activation
Input	100	100	relu
Hidden 1	100	100	relu
Hidden 2	100	50	relu
Output	50	4	linear

Table 4: g-and-k: Network settings for PEN-0

ϕ network

Layer	Dim. in	Dim. out	Activation
Input	1	100	relu
Hidden 1	100	50	relu
Output	50	10	linear

ρ network

Layer	Dim. in	Dim. out	Activation
Input	10	100	relu
Hidden 1	100	100	relu
Hidden 2	100	50	relu
Output	50	4	linear

4 α-stable distribution

- The characteristic function $\varphi(x)$ for the α-stable distribution is given by Ong et al. [2018]

Table 5: g-and-k: Number of weights for the different networks

Network	# weights
MLP small	26039
MLP large	115454
MLP pre	25454
PEN-0	22214

$$\varphi(x) = \begin{cases} \exp\left(i\delta t - \gamma^{\alpha}|t|^{\alpha}\left(1 + i\beta \tan\frac{\pi\alpha}{2}\operatorname{sgn}(t)(|\gamma t|^{1-\alpha} - 1)\right)\right), & \alpha \neq 1, \\ \exp\left(i\delta t - \gamma|t|\left(1 + i\beta\frac{2}{\pi}\operatorname{sgn}(t)\log(\gamma|t|)\right)\right), & \alpha = 1, \end{cases}$$

where sgn is the sign function, i.e.

$$\operatorname{sgn}(t) = \begin{cases} -1 & \text{if } t < 0, \\ 0 & \text{if } t = 0, \\ 1 & \text{if } t > 0. \end{cases}$$

- The network settings are presented in Table 6, 7, 8, and 9;
- The number of weights for the different networks are presented in Table 10;
- Values outside of the range $[-10, 50]$ are considered to be outliers and these values are replaced (at random) with values inside the data range. The data cleaning scheme is applied to both the observed and generated data;
- All data sets are standardized using the "robust scalar" method, i.e. each data point y_i is standardized according to

$$\frac{y_i + Q_1(y)}{Q_3(y) - Q_1(y)}$$

  where Q_1 and Q_3 are the first and third quantiles respectively;
- When computing the empirical distribution function we evaluate this function over 100 equally spaced points between -10 and 100;
- The root-mean-squared error (RMSE) is computed as

$$\text{RMSE} = \sqrt{\frac{1}{R}\sum_{i=1}^{R}\{(\hat{\theta}_i^1 - \theta^1)^2 + (\hat{\theta}_i^2 - \theta^2)^2 + (\hat{\theta}_i^3 - \theta^3)^2 + (\hat{\theta}_i^4 - \theta^4)^2\}}$$

  where $\theta = [\theta^1, \theta^2, \theta^3, \theta^4]$ are ground-truth parameter values and $[\hat{\theta}_i^1, \hat{\theta}_i^2, \hat{\theta}_i^3, \hat{\theta}_i^4]_{1 \leq i \leq R}$ are ABC posterior means. R is the number of independent repetitions of the inference procedure;
- Number of training observations: $5 \cdot 10^5$, 10^5, 10^4, and 10^3. Evaluation data observations $5 \cdot 10^3$.

Table 6: α-stable: Network settings for MLP small.

Layer	Dim. in	Dim. out	Activation
Input	1002	25	relu
Hidden 1	25	25	relu
Hidden 2	25	12	relu
Output	12	4	linear

Table 7: α-stable: Network settings for MLP large.

Layer	Dim. in	Dim. out	Activation
Input	1002	100	relu
Hidden 1	100	100	relu
Hidden 2	100	50	relu
Output	50	4	linear

Table 8: α-stable: Network settings MLP pre.

Layer	Dim. in	Dim. out	Activation
Input	100	100	relu
Hidden 1	100	100	relu
Hidden 2	100	50	relu
Output	50	4	linear

Table 9: α-stable: Network settings for PEN-0.

ϕ network

Layer	Dim. in	Dim. out	Activation
Input	1	100	relu
Hidden 1	100	50	relu
Output	50	20	linear

ρ network

Layer	Dim. in	Dim. out	Activation
Input	22	100	relu
Hidden 1	100	100	relu
Hidden 2	100	50	relu
Output	50	4	linear

Table 10: α-stable: Number of weights for the different networks

Network	# weights
MLP small	26089
MLP large	115654
MLP pre	25454
PEN-0	23924

5 Autoregressive time series model

- The network settings are presented in Table 11, 12, 13, and 14;
- The number of weights for the different networks are presented in Table 15;
- Number of training observations: 10^6, 10^5, 10^4, and 10^3. Evaluation data observations 10^4.

6 Moving average time series with observational noise model

- The network settings are presented in Table 16, 17, 18, and 19;
- The number of weights for the different networks are presented in Table 20;
- Number of training observations: 10^6, 10^5, 10^4, and 10^3. Evaluation data observations $5 \cdot 10^5$.

Table 11: AR(2): Network settings for MLP small.

Layer	Dim. in	Dim. out	Activation
Input	100	55	relu
Hidden 1	55	55	relu
Hidden 2	55	25	relu
Output	25	2	linear

Table 12: AR(2): Network settings for MLP large.

Layer	Dim. in	Dim. out	Activation
Input	100	100	relu
Hidden 1	100	100	relu
Hidden 2	100	50	relu
Output	50	2	linear

Table 13: AR(2): Network settings for PEN-0.

ϕ network

Layer	Dim. in	Dim. out	Activation
Input	1	100	relu
Hidden 1	100	50	relu
Output	50	10	linear

ρ network

Layer	Dim. in	Dim. out	Activation
Input	10	50	relu
Hidden 1	50	50	relu
Hidden 2	50	20	relu
Output	20	2	linear

Table 14: AR(2): Network settings for PEN-2.

ϕ network

Layer	Dim. in	Dim. out	Activation
Input	3	100	relu
Hidden 1	100	50	relu
Output	50	10	linear

ρ network

Layer	Dim. in	Dim. out	Activation
Input	12	50	relu
Hidden 1	50	50	relu
Hidden 2	50	20	relu
Output	20	2	linear

Table 15: AR(2): Number of weights for the different networks

Network	# weights
MLP small	10087
MLP large	25352
PEN-0	9922
PEN-2	10222

Table 16: MA(2): Network settings for MLP small.

Layer	Dim. in	Dim. out	Activation
Input	100	60	relu
Hidden 1	60	60	relu
Hidden 2	60	25	relu
Output	25	2	linear

Table 17: MA(2): Network settings for MLP large.

Layer	Dim. in	Dim. out	Activation
Input	100	100	relu
Hidden 1	100	100	relu
Hidden 2	100	50	relu
Output	50	2	linear

Table 18: MA(2): Network settings for PEN-0.

ρ network

Layer	Dim. in	Dim. out	Activation
Input	1	100	relu
Hidden 1	100	50	relu
Hidden 2	50	10	relu

ϕ network

Layer	Dim. in	Dim. out	Activation
Input	10	50	relu
Hidden 1	50	50	relu
Hidden 2	50	20	relu
Output	20	2	linear

Table 19: MA(2): Network settings for PEN-10

ρ network

Layer	Dim. in	Dim. out	Activation
Input	11	100	relu
Hidden 1	100	50	relu
Hidden 2	50	10	relu

ϕ network

Layer	Dim. in	Dim. out	Activation
Input	20	50	relu
Hidden 1	50	50	relu
Hidden 2	50	20	relu
Output	20	2	linear

Table 20: MA(2): Number of weights for the different networks

Network	# weights
MLP small	11297
MLP large	25352
PEN-0	9922
PEN-10	11422

References

V. M. Ong, D. J. Nott, M.-N. Tran, S. A. Sisson, and C. C. Drovandi. Variational bayes with synthetic likelihood. *Statistics and Computing*, 28(4):971–988, 2018.

U. Picchini and R. Anderson. Approximate maximum likelihood estimation using data-cloning ABC. *Computational Statistics & Data Analysis*, 105:166–183, 2017.

D. Prangle. gk: An R package for the g-and-k and generalised g-and-h distributions. *arXiv:1706.06889*, 2017.