Poisson approximate likelihood compared to the particle filter

Yize Hao, Aaron A. Abkemeier and Edward L. Ionides

Draft compiled on September 12, 2024

Abstract

Filtering algorithms are fundamental for inference on partially observed stochastic dynamic systems, since they provide access to the likelihood function and hence enable likelihood-based or Bayesian inference. A novel Poisson approximate likelihood (PAL) filter was introduced by Whitehouse et al. (2023). PAL employs a Poisson approximation to conditional densities, offering a fast approximation to the likelihood function for a certain subset of partially observed Markov process models. A central piece of evidence for PAL is the comparison in Table 1 of Whitehouse et al. (2023), which claims a large improvement for PAL over a standard particle filter algorithm. This evidence, based on a model and data from a previous scientific study by Stocks et al. (2018), might suggests that researchers confronted with similar models should use PAL rather than previous particle filter methods. Alternatively, the improvement in likelihood found by Whitehouse et al. (2023) compared to Stocks et al. (2018) might indicate a flaw in the model or numerical methods for that previous work. We show that neither of these conclusions is valid, and the comparison of log-likelihood values made by Whitehouse et al. (2023) is flawed because the PAL calculations were carried out using a dataset scaled differently from the previous study. If PAL and the particle filter are used on this new scale, the superficial advantage of PAL largely disappears. On simulations where the model is correctly specified, the particle filter outperforms PAL.

1 Introduction

This article results from an investigation of the results presented by Whitehouse et al. (2023) (henceforth, WWR) in their Table 1. WWR were given the opportunity to submit a correction, after we shared the results of our investigation with them, but they declined. The theory developed by WWR shows that PAL has some potentially useful scaling properties, but the numerical results in their Table 1 appear to show much stronger performance than a standard particle filter on an example of scientific interest but moderate size. We present a correction of this error so that researchers considering whether to implement Poisson approximate likelihood (PAL) are appropriately informed about its benefits.

Table 1 of WWR reanalyzes the model and data of Stocks et al. (2018) (henceforth, SBH) for which the likelihood was calculated using a particle filter. SBH found strong evidence for the importance of overdispersion in a stochastic dynamic model for their epidemiological data. This is significant because most earlier research on population dynamics avoided consideration of overdispersion, perhaps due to the lack of available statistical methodology for fitting overdispersed nonlinear stochastic dynamic models. The conclusions of SBH hinge on a comparison of likelihoods, and so the results of WWR discredit those conclusions by indicating that SBH based their reasoning on inaccurately computed likelihoods. An important consequence of correcting Table 1 of WWR is that the results of SBH stand undiminished.

SBH and WWR each fitted three different rotavirus models. The first has equidispersion (i.e., no overdispersion) in the measurement model and the dynamic model, and is called EqEq by WWR. The second, EqOv, includes overdispersion in only the measurement model. The third, OvOv, includes overdispersion in both these model components. We focus on OvOv, which WWR found to be the best fitting model.

We show that most of the apparent advantage for PAL on the OvOv model, compared to a particle filter, arises because WWR used a different scaling of the data from SBH. Two models for the same data can properly be compared by their likelihood, even if the models have entirely different structures. One can make allowance for the number of estimated parameters using a quantity such as Akaike's information criterion (Akaike, 1974). However, if data are rescaled, a correction is required to make likelihoods comparable. For example, if one model describes a dataset in grams and another describes it in kilograms, then the latter model will earn an increased log-likelihood of $\log(10^3)$ for each data point simply because of the change in scale. Presenting a direct comparison of a likelihood for the data in grams with a likelihood for the data in kilograms would evidently be inappropriate.

Table 1: AIC for the OvOv rotavirus model, computed using two filtering methods. PAL is the Poisson approximate likelihood, implemented using the code of WWR. PF is the particle filter, implemented using the pomp R package King et al. (2016). The first two lines are from WWR, Table 1, and lines 3–5 are our own computations. We used 50000 particle for both PF and PAL. PF was repeated 36 times to reduce the Monte Carlo variance, but this step was not necessary for PAL. PF was maximized using iterated filtering, and PAL was maximized using coordinate gradient descent.

	Method	Result	AIC
1.	PF	Table 1 of WWR, originally from SBH	20134
2.	PAL	Table 1 of WWR	13778
3.	PAL	Data, model and parameters from WWR	13959
4.	PF	Data, model and parameters matching 3.	60607
5.	PF	Changing initial values	942239
6.	PAL	Data, model and parameters matching 5.	18104
7.	ARMA	Sec. 7.3.2 of WWR	23043
8.	ARMA	Data from WWR	12751

SBH fitted their model to a dataset derived by dividing the original reported count data by an estimated reporting rate, to put their data on the scale of the actual number of cases in the population, whereas WWR fitted directly to the report data. The reporting rate used by SBH varied over time, but was generally around 7%. On approximately 1200 data points, this corresponds to a discrepancy of around $-1200\log(0.07) \approx 3200$ log-likelihood units, largely explaining the difference reported in Table 1 and interpreted by WWR as evidence supporting PAL. The comparison can be corrected either by applying the method of SBH to the data of WWR or vice versa. Since the method of SBH is applicable to a more general class of models, and supported by published software, it was convenient to apply the SBH method to the model and data of WWR. The large discrepancy in log-likelihood disappears at this point by recomputing the likelihood of the model using PAL and particle filter separately (see Table 1). This re-analysis does, however, show a discrepancy between two methods. We continued our investigation to establish the cause of this.

Inspection of log-likelihood anomalies (Wheeler et al., 2024; Li et al., 2024; Hao, 2024) showed that the initial conditions for the latent process in January 2001 were fixed at values which were incompatible with the trajectory of the data early in the time series. By contrast, SBH fixed the initial conditions 6 years before the first measurement, giving time for the system to reach its equilibrum distribution. Line 5 of Table 1 shows that this improvement enables PF to reach close to the values attained by PAL. Further, at the MLE obtained by maximizing PF, we found that PF beats PAL (line 6).

WWR compare their fitted models with a log-ARMA(2,1) benchmark AIC value of 23043 (line 7), inferring that all the mechanistic models possessing overdispersion have better statistical fit than a simple loglinear time series model. However, this value corresponds to a log-ARMA(2,1) model fitted not to WWR's data but to SBH's data. Thus, line 7 can properly compared to line 1 but not to any other line of Table 1. We independently fitted a log-ARMA(2,1) model to the SBH data and obtained a similar AIC value (23085). Carrying out the same computation for the data fitted by WWR gives a log-ARMA benchmark AIC value of 12751 (line 8). Thus, the statistical fit of the mechanistic model considered by WWR is inferior to a simple log-ARMA model. This holds for all the variants in lines 2–6 of Table 1, regardless of whether PF or PAL is used. The goal of mechanistic modeling is not necessarily to beat a simple statistical benchmark, but this can be an indication that additional model development could be worthwhile (Wheeler et al., 2024). A correct interpretation of Table 1 is therefore very different to the conclusions drawn by WWR from comparing line 2 inappropriately to lines 1 and 7.

In the presence of model misspecification, it becomes harder to compare likelihood evaluation methods. A likelihood approximation, such as PAL, may potentially obtain a higher value than the exact likelihood if it compensates for the misspecification. Log-likelihood is a proper scoring rule for forecasts Gneiting and Raftery (2007), and both the particle filter and PAL construct their log-likelihood estimates via a sequence of one-step forecasts. Therefore, if the model is correctly specified, the approximation error in PAL can only decrease the expected log-likelihood. We tested this on simulated data for which the model of WWR is correctly specified. For this simulation study, the particle filter out-performs PAL (Figure 1). On average, the particle filter likelihood estimate is -5887.1 log units higher than the PAL. We know from the benchmark AIC value in

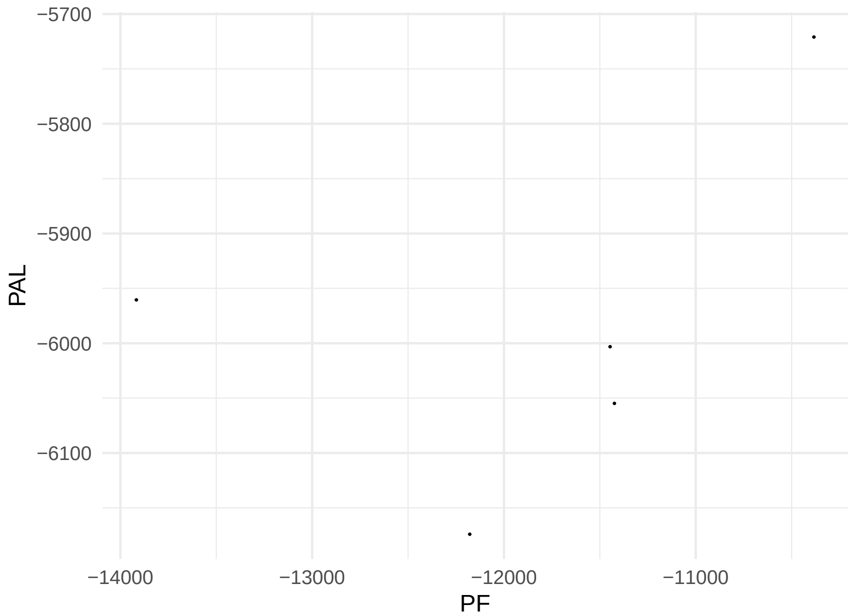

Figure 1: Log-likelihoods computed using two filtering methods for 100 randomly simulated datasets at the MLE of the OvOv model. Simulations with one or more zero counts were disqualified since they resulted in errors for the PAL implementation. We used 50000 particle for both PF and PAL. PF was repeated 36 times to reduce the Monte Carlo variance, but this step was not necessary for PAL. The red line corresponds to equality of the two estimates.

Table 1. line 8, that there is substantial model misspecification. Therefore, this real-data example may not be well suited to comparing the filtering skill of PAL and PF.

We conclude that PAL is a potentially useful algorithm, with some favorable theoretical properties. However, the corrected evidence does not indicate an advantage for using PAL in situations where the particle filter is effective.

An extended description of our reanalysis of WWR is provided by Hao (2024). That thesis also contains some additional results reinforcing the conclusions presented in this article. The source code for this article is available at `https://github.com:ionides/pal-vs-pf` and archived at [EI: ZENODO - TODO].

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Hao, Y. (2024). PAL versus SMC: Two approaches in compartmental modeling. a case study with rotavirus in Germany during 2001-2008. University of Michigan Honors Thesis.

King, A. A., Nguyen, D., and Ionides, E. L. (2016). Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, 69(12):1–43.

Li, J., Ionides, E. L., King, A. A., Pascual, M., and Ning, N. (2024). Inference on spatiotemporal dynamics for networks of biological populations. *Journal of the Royal Society Interface*, 21(216):20240217.

Stocks, T., Britton, T., and Höhle, M. (2018). Model selection and parameter estimation for dynamic epidemic models via iterated filtering: application to rotavirus in Germany. *Biostatistics*, 21(3):400–416.

Wheeler, J., Rosengart, A., Zhuxun, J., Tan, K. H. E., Treutle, N., and Ionides, E. L. (2024). Informing policy via dynamic models: Cholera in Haiti. *PLOS Computational Biology*, 20:e1012032.

Whitehouse, M., Whiteley, N., and Rimella, L. (2023). Consistent and fast inference in compartmental models of epidemics using Poisson Approximate Likelihoods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(4):1173–1203.