Poisson approximate likelihood compared to the particle filter

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

Draft compiled on August 23, 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. If a poorly fitting specification of initial value parameters is amended, the particle filter also outperforms PAL on the actual data.

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 Table 1 appear to show much stronger performance than a standard particle filter on an example of scientific interest but moderate size. Our task here is to correct the error in Table 1 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 earlier research on population dynamics largely avoided consideration of overdispersion, perhaps because of the lack of avaialable statistical methodology to fit such 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. 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
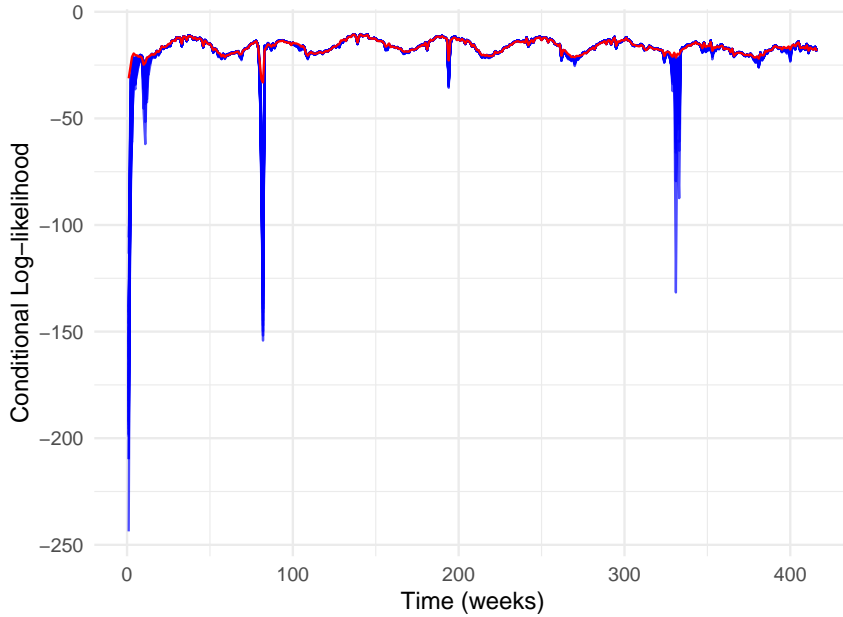
Figure 1: Conditional log-likelihoods computed using PAL (red) and PF (blue) for the rotavirus model and MLE from WWR. We used 50000 particle for both PF and PAL. PF averaged over 36 replications to reduce the Monte Carlo variance, but this step was not necessary for PAL. The main sources of likelihood shortfall of PF are at time points $t = 1, 2, 3, 11, 81, 194$, and $325 - 333$.

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 code of 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. [EI: SHOULD WE HAVE A LINE FOR PAL AT THE IF2 MLE FROM LINE 5? PERHAPS WE CAN'T READILY MAXIMIZE PAL FOR THIS MODEL (OR CAN WE?) BUT SOME SORT OF COMPARISON WOULD BE HELPFUL]

	Method	Result	AIC
1.	PAL	Table 1 of WWR	13778.08
2.	PF	Table 1 of WWR, originally from SBH	20134.38
3.	PAL	Model, data and MLE from WWR	13798.60
4.	PF	Model, data and MLE from WWR	14507.15
5.	PF	Data from WWR, estimated initial values	13630.26

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. We discovered that the initial conditions for the latent process in January 2001 (their time 0) were estimated by SBH. However, in WWR, they were assumed fixed and made the model fitted poorly.

Figure 1 indicates that the worst conditional log-likelihood values, estimated using the particle filter, arise at the start of the time series. This could be a problem with data collection, or some

Table 2: *Maximum likelihood estimation of parameters for the OvOv model of WWR by PAL (taken from WWR), and by iterated particle filtering implemented using* `mif2` *in* `pomp` *(King et al., 2016). The initial distribution parameter* $\lambda_0 = (S_{10}, I_{10}, R_{10}, S_{20}, I_{20}, R_{20}, S_{30}, I_{30}, R_{30})$ *are assumed to be fixed in Whitehouse et al. (2023). The iterated particle filtering estimates were obtained by using 3 rounds of refinement, with 100 filtering iterations, 50,000 particles, and 36 replicates in each round with the top 12 best fits in terms of likelihood chosen to be the starting value for the next round.*

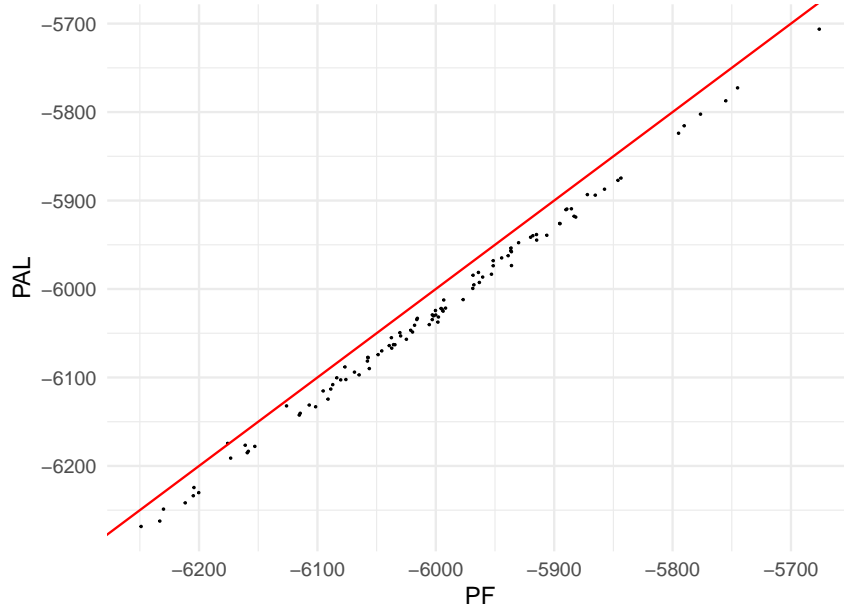| Parameter	PAL	iterated filtering
β_1	11.48	9.84
β_2	0.25	0.24
β_3	0.35	0.74
ϕ	0.14	0.17
ρ	0.16	0.18
σ_q^2	0.021	0.019
σ_ξ	66.89	13.96
S_{10}	3876549	6145778
I_{10}	30351	2181
R_{10}	1315221	468260
S_{20}	57139612	37469849
I_{20}	871	1942
R_{20}	302852	31572281
S_{30}	19573727	4540574
I_{30}	2550	774
R_{30}	131092	2171187
AIC	13778.08	13630.26

Figure 2: *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.*

subtle issue with the science of the epidemiological system, but a simple possibility is that the initial values of the latent state process might be poorly specified. For this model, the initial values were fixed based on scientific reasoning, rather than being estimated. If we instead estimate the initial values, we obtain the results in Table 2.

[EI: SINCE THE MEAN DURATION OF IMMUNITY IS 1 YR, HAVING 31×10^6 in $R_{2,0}$ IMPLIES THAT ABOUT HALF OF ALL SUSCEPTIBLE INDIVIDUALS GET ROTAVIRUS EVERY YEAR ACCORDING TO THE MODEL. IS THAT THE CASE FOR THE MODEL? FOR THE DATA, THE PER-CAPITA RATE IN THIS AGE GROUP IS RATER LOW. THE SIMPLEST WAY TO AVOID THIS PROBLEM WOULD BE TO DO WHAT SBH DID (DESCRIBED IN THEIR ONLINE APPENDIX C) - EFFECTIVELY, DRAWING X_0 FROM THE STATIONARY DISTRIBUTION GIVEN THE MODEL. THIS HAS THE ADVANTAGE OF INCLUDING NO NEW PARAMETERS AND BEING CONSISTENT WITH SBH. WE DON'T KNOW IN ADVANCE EXACTLY HOW SUCCESSFUL IT WOULD BE...]

Discussion

PAL provides an approximate filter, whereas the particle filter provides a Monte Carlo estimate of an exact filter. Possible explanations for PAL obtaining a higher log-likelihood than a particle filter include:

1. The particle filter could have high Monte Carlo variance. This results in negative bias in its log-likilihood estimate due to Jensen's inequality, since the particle filter provides an unbiased likelihood estimate.

2. If the model is misspecified, the approximation error in PAL could lead to a higher log-likelihood estimate than that of the actual misspecified model.

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 Whitehouse et al. (2023) is correctly specified. For this simulation study, the particle filter out-performs PAL (Figure 2) which supports hypothesis 2, above. On average, the particle filter likelihood estimate is 25.3 log units higher than the PAL. Consequently, we looked for model misspecification by comparing the conditional log-likelihood estimates at each time point, as described in the Methods section.

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.

Methods

We first replicated the model of WWR in the R package `pomp` (King et al., 2016). Comparing summary statistics of simulations confirms that both model versions are essentially identical, as described by Hao (2024). We therefore implement PAL using the code from Whitehouse et al. (2023), comparing directly with the particle filter implemented using `pomp`.

This article has focused on likelihood evaluation, but the inferences presented also require the likelihood estimate to be maximized. WWR used direct numerical maximization, whereas SBH used iterated filtering to maximize the particle filter likelihood estimate implemented using the `mif2` algorithm in `pomp`. We follow the approaches adopted by these two papers. The optimization method is not directly pertinent to the comparison of the filtering methods, so we do not discuss optimization in further detail here. 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.

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.

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.