Lesson 3: Likelihood-based inference for POMP models

Spencer J. Fox   Qianying (Ruby) Lin   Jesse Wheeler

Objectives

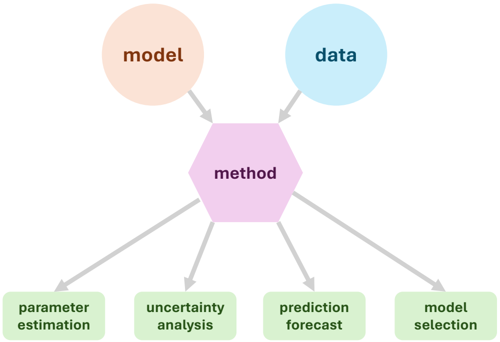
Students completing this lesson will:

1. Gain an understanding of the nature of the problem of likelihood computation for POMP models.
2. Be able to explain the simplest particle filter algorithm.
3. Gain experience in the visualization and exploration of likelihood surfaces.
4. Be able to explain the tools of likelihood-based statistical inference that become available given numerical accessibility of the likelihood function.

Overview I

A general framework of epidemiological inference includes three layers:

- The input: a model of interest and the given data
- A method for inference
- Inferences include estimation, uncertainty, prediction and forecast, and model selection.



Overview II

Methods for inference can be categorized into three groups:

- Optimization-based: minimize a cost function (e.g., SSE, MSE, MAE) that measures the difference between observed data and model predictions
- Likelihood-based: maximize a likelihood function, which represents the probability of observing the given data given the parameters
- Summary Statistics-based: use a set of features of the data instead of the full set of data

In this lesson, we focus on the likelihood-based method because

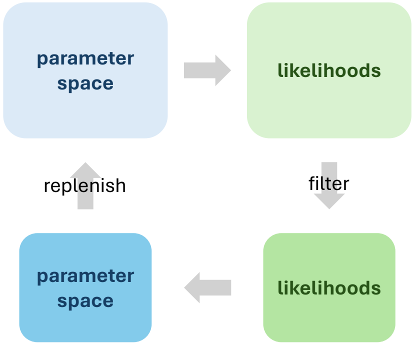
- it fits for stochastic models and
- it incorporates all data (i.e., full-information).

The likelihood I

- The basis for modern frequentist, Bayesian, and information-theoretic inference.
- Method of maximum likelihood introduced by Fisher (1922).
- The likelihood function itself is a representation of the what the data have to say about the parameters.
- A good general reference on likelihood is by Pawitan (2001).

The likelihood II

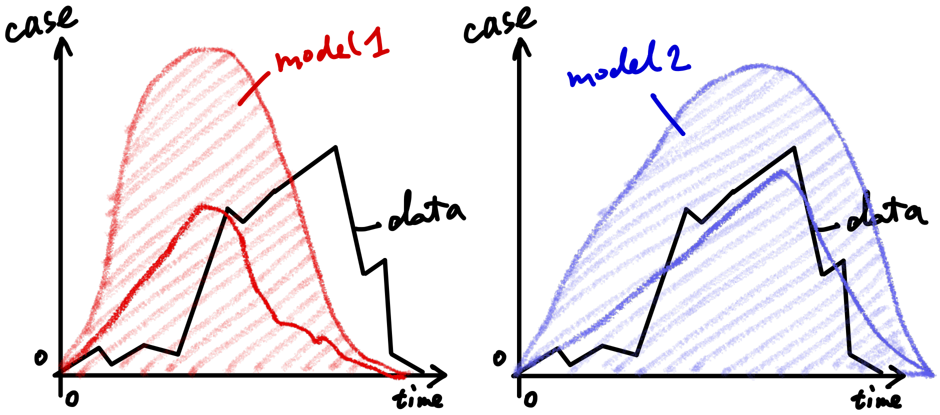
- Goal: fit the model to the data and conduct statistical inferences, such as parameter estimation.
- The likelihood, thus, can be considered as a metric to assess the *goodness* of the proposed parameters.
- By exploring the space of parameters, we can eventually obtain the maximum likelihood estimator (MLE).



Thus, the objective of this lesson is to discuss how we compute the likelihood given a model of interest with a proposed set of parameters in both theory and in `pomp`.

Definition of the likelihood function I

- How likely the data are drawn from a distribution or sampled from a model?



Definition of the likelihood function II

- Notations:
  - $y^*_{1:N}$: the data, a sequence of N observations
  - $f_{Y_{1:N}}(y_{1:N};\theta)$: the statistical model, a probability distribution for each value of a parameter vector θ
  - $Y_{1:N} \sim f_{Y_{1:N}}(y_{1:N};\theta)$: a random variable drawn from distribution $f_{Y_{1:N}}(y_{1:N};\theta)$
- The likelihood function is used to measure the "how likely'':

$$\mathcal{L}(\theta) = f_{Y_{1:N}}(y^*_{1:N};\theta).$$

- The log-likelihood function:

$$\ell(\theta) = \log \mathcal{L}(\theta) = \log f_{Y_{1:N}}(y^*_{1:N};\theta).$$

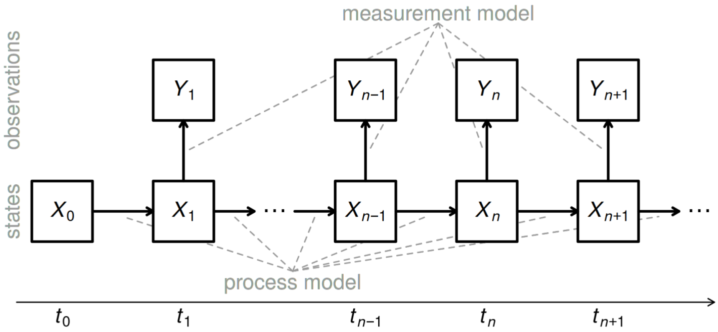
Simulation is easy for complex models

- $f_{Y_{1:N}}(y_{1:N};\theta)$ is simple and with an explicit expression:
  - the simulation of $Y_{1:N}$ is direct, e.g., $Y_k \sim N(0,1)$ for $k = 1, \ldots, N$
  - the likelihood function is explicit
- $f_{Y_{1:N}}(y_{1:N};\theta)$ is complex or even without an explicit expression:
  - the simulation of $Y_{1:N}$, given the underlying dynamical model, is a bit more complex but convenient
  - the likelihood function exists with a complicated expression or even without an explicit expression

Thus, we can develop numerical methods to compute the complex or implicit likelihood functions!

The likelihood for a POMP model I

Recall the following schematic diagram, showing dependence among variables in a POMP model.



The likelihood for a POMP model II

Recall the following definitions and properties:

- **Measurements**: Y_n, at time t_n depend on the latent process, X_n, at that time.
- **The Markov property**: latent process variables depend on their value at the previous timestep.
  1. The distribution of the state X_{n+1}, conditional on X_n, is independent of the values of X_k, $k < n$ and Y_k, $k \leq n$.
  2. The distribution of the measurement Y_n, conditional on X_n, is independent of all other variables.
- The **latent process**: $X(t)$, may be defined at all times, but we are particularly interested in its value at observation times. Therefore, we write

$$X_n = X(t_n).$$

  - We write collections of random variables using the notation $X_{0:N} = (X_0, \dots, X_N)$.

The likelihood for a POMP model III

- The **one-step transition density**, $f_{X_n|X_{n-1}}(x_n|x_{n-1};\theta)$, together with the **measurement density**, $f_{Y_n|X_n}(y_n|x_n;\theta)$ and the **initial density**, $f_{X_0}(x_0;\theta)$, specify the entire joint density via

$$
\begin{aligned}
& f_{X_{0:N},Y_{1:N}}(x_{0:N},y_{1:N};\theta) \\
& \qquad = f_{X_0}(x_0;\theta)\,\prod_{n=1}^{N} f_{X_n|X_{n-1}}(x_n|x_{n-1};\theta)\, f_{Y_n|X_n}(y_n|x_n;\theta).
\end{aligned}
$$

- **The marginal density for sequence of measurements**: $Y_{1:N}$, evaluated at the data, $y^*_{1:N}$, is

$$
\mathcal{L}(\theta) = f_{Y_{1:N}}(y^*_{1:N};\theta) = \int f_{X_{0:N},Y_{1:N}}(x_{0:N},y^*_{1:N};\theta)\, dx_{0:N}.
$$

Special case: deterministic latent process

- When the latent process is non-random, the log-likelihood for a POMP model closely resembles a nonlinear regression model.
- In this case, we can write $X_n = x_n(\theta)$, and the log-likelihood is

$$\ell(\theta) = \sum_{n=1}^{N} \log f_{Y_n|X_n}\big(y_n^*|x_n(\theta);\theta\big).$$

- If we have a Gaussian measurement model, where Y_n given $X_n = x_n(\theta)$ is conditionally normal with mean $\hat{y}_n\big(x_n(\theta)\big)$ and constant variance σ^2, then the log-likelihood contains a sum of squares which is exactly the criterion that nonlinear least squares regression seeks to minimize.

General case: stochastic unobserved state process

- For a POMP model, the likelihood takes the form of an integral:

$$
\begin{aligned}
\mathcal{L}(\theta) &= f_{Y_{1:N}}(y^*_{1:N};\theta) \\
&= \int f_{X_0}(x_0;\theta) \prod_{n=1}^{N} f_{Y_n|X_n}(y^*_n|x_n;\theta)\, f_{X_n|X_{n-1}}(x_n|x_{n-1};\theta)\, dx_{0:N}.
\end{aligned} \tag{1}
$$

- This integral is high dimensional and, except for the simplest cases, can not be reduced analytically.

Monte Carlo likelihood: direct simulation I

Spoiler Alert: This section serves to introduce the concept of the **particle filter** and the approach of Monte Carlo integration by first proposing an intuitive and a simpler method. This simple method usually **does NOT work** on anything but **very short** time series.

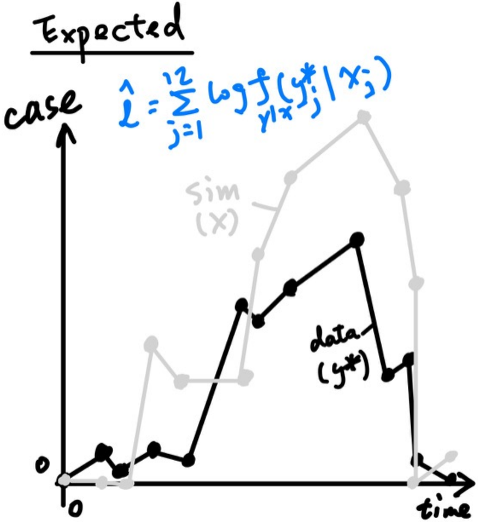
1. Let's rewrite the likelihood integral using an equivalent factorization. As an exercise, you could check how the equivalence of Equation 1 and Equation 2 follows algebraically from the Markov property and the definition of conditional density.

$$
\begin{aligned}
\mathcal{L}(\theta) &= f_{Y_{1:N}}(y^*_{1:N};\theta) \\
&= \int \left\{ \prod_{n=1}^{N} f_{Y_n|X_n}(y^*_n|x_n;\theta) \right\} f_{X_{0:N}}(x_{0:N};\theta)\, dx_{0:N}.
\end{aligned} \tag{2}
$$

Monte Carlo likelihood: direct simulation II

2. Notice, using the representation in Equation 2, that the likelihood can be written as an expectation,
$$\mathcal{L}(\theta) = \mathbb{E}\left[\prod_{n=1}^{N} f_{Y_n|X_n}(y_n^*|X_n;\theta)\right],$$
where the expectation is taken with $X_{0:N} \sim f_{X_{0:N}}(x_{0:N};\theta)$.
3. Now, using a law of large numbers, we can approximate an expectation by the average of a Monte Carlo sample. Thus,
$$\mathcal{L}(\theta) \approx \frac{1}{J}\sum_{j=1}^{J}\prod_{n=1}^{N} f_{Y_n|X_n}(y_n^*|X_n^j;\theta),$$
where $\{X_{0:N}^j, j = 1, \dots, J\}$ is a Monte Carlo sample of size J drawn from $f_{X_{0:N}}(x_{0:N};\theta)$.

Monte Carlo likelihood: direct simulation III



Sequential Monte Carlo: The particle filter I

Fortunately, we can compute the likelihood for a POMP model by a much more efficient algorithm than direct Monte Carlo integration:

1. We proceed by factorizing the likelihood in a different way:

$$\begin{aligned}\mathcal{L}(\theta) = f_{Y_{1:N}}(y^*_{1:N};\theta) &= \prod_{n=1}^{N} f_{Y_n|Y_{1:n-1}}(y^*_n|y^*_{1:n-1};\theta)\\ &= \prod_{n=1}^{N} \int f_{Y_n|X_n}(y^*_n|x_n;\theta)\, f_{X_n|Y_{1:n-1}}(x_n|y^*_{1:n-1};\theta)\, dx_n,\end{aligned}$$

with the understanding that $f_{X_1|Y_{1:0}} = f_{X_1}$.

Sequential Monte Carlo: The particle filter II

2. The Markov property leads to the **prediction formula:**

$$
\begin{aligned}
&f_{X_n|Y_{1:n-1}}(x_n|y^*_{1:n-1};\theta) \\
&\quad = \int f_{X_n|X_{n-1}}(x_n|x_{n-1};\theta)\, f_{X_{n-1}|Y_{1:n-1}}(x_{n-1}|y^*_{1:n-1};\theta)\, dx_{n-1}.
\end{aligned}
$$

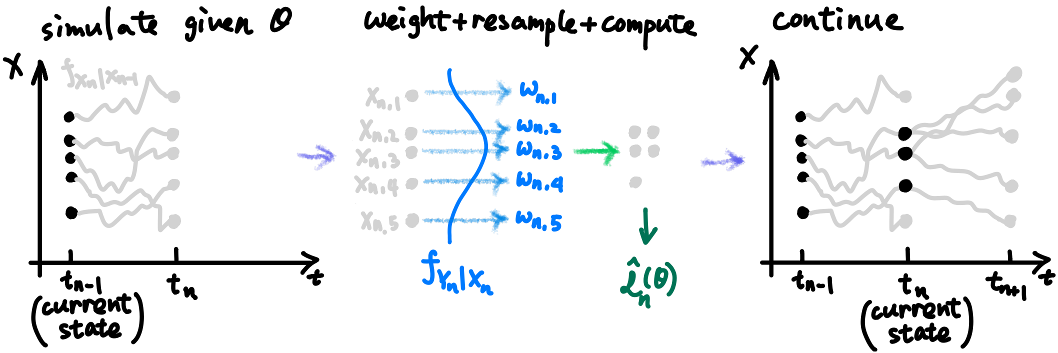
3. Bayes' theorem gives the **filtering formula:**

$$
\begin{aligned}
&f_{X_n|Y_{1:n}}(x_n|y^*_{1:n};\theta) \\
&\quad = f_{X_n|Y_n,Y_{1:n-1}}(x_n|y^*_n,y^*_{1:n-1};\theta) \\
&\quad = \frac{f_{Y_n|X_n}(y^*_n|x_n;\theta)\, f_{X_n|Y_{1:n-1}}(x_n|y^*_{1:n-1};\theta)}{\int f_{Y_n|X_n}(y^*_n|u_n;\theta)\, f_{X_n|Y_{1:n-1}}(u_n|y^*_{1:n-1};\theta)\, du_n}.
\end{aligned}
$$

Sequential Monte Carlo: The particle filter III

- This suggests that we keep track of two key distributions at each time t_n,
  - The **prediction distribution** is $f_{X_n|Y_{1:n-1}}(x_n|y^*_{1:n-1})$.
  - The **filtering distribution** is $f_{X_n|Y_{1:n}}(x_n|y^*_{1:n})$.
- The prediction and filtering formulas give us a two-step recursion:
  - The prediction formula gives the prediction distribution at time t_n using the filtering distribution at time t_{n-1}.
  - The filtering formula gives the filtering distribution at time t_n using the prediction distribution at time t_n.
- The **particle filter** use Monte Carlo techniques to sequentially estimate the integrals in the prediction and filtering recursions. Hence, the alternative name of **sequential Monte Carlo (SMC)**.

Sequential Monte Carlo: The particle filter IV

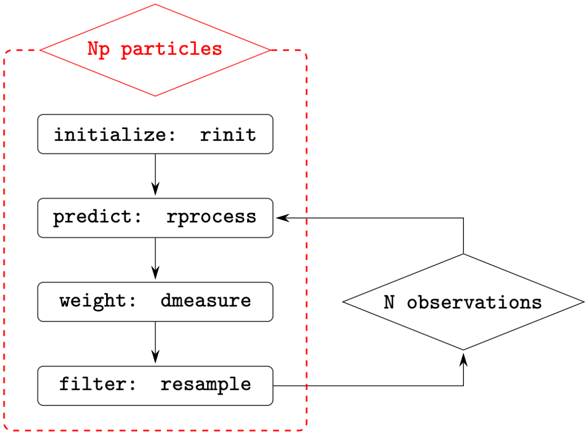


The full log-likelihood then has approximation

$$\ell(\theta) = \log \mathcal{L}(\theta) = \sum_n \log \mathcal{L}_n(\theta) \approx \sum_n \log \hat{\mathcal{L}}_n(\theta).$$

Sequential Monte Carlo: conclusion

- It can be shown that the particle filter provides an unbiased estimate of the likelihood (Kitagawa (1987), Arulampalam et al. (2002), Doucet, Freitas, and Gordon (2001), King, Nguyen, and Ionides (2016)).
- This implies a consistent but biased estimate of the log-likelihood.



Parallel computing: general

It will be helpful to parallelize most of the computations. Most machines nowadays have multiple cores and using this computational capacity is as simple as:

1. letting `R` know you plan to use multiple processors;
2. using the parallel for loop provided by the `foreach` package; and
3. paying proper attention to the use of parallel random number generators (RNG).

For example:

```
library(foreach)                # load foreach
library(doFuture)               # load doFuture/future
plan(multisession)              # using multiple R sessions
                                # handling global environment
```

The second line tells `foreach` that we will use the `doFuture` backend, along with `%dofuture%` when we set up the `foreach` loop. The third line explicitly exports the global environment to each of the session on Windows machines; macOS/Linux does that automatically.

Parallel computing: Exercise

The following codes (or the `/scripts/exercise_parallel.R` script) is an example of setting up a parallel computing scheme.

```
library(foreach)
library(doFuture)
plan(multisession, workers=8)
source("model_measSIR.R")

foreach(
  i=1:20, .combine="c", .options.future = list(seed=TRUE),
  .errorhandling = "remove"
) %dofuture% {
  measSIR |> pfilter(Np=5000)
} -> pfs

pfs |> logLik() |> logmeanexp(se=TRUE)
```

References

Arulampalam, M. S., S. Maskell, N. Gordon, and T. Clapp. 2002. "A Tutorial on Particle Filters for Online Nonlinear, Non-Gaussian Bayesian Tracking." *IEEE Trans Signal Process* 50: 174–88. https://doi.org/10.1109/78.978374.

Doucet, Arnaud, Nando de Freitas, and Neil Gordon, eds. 2001. *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag.

Fisher, R. A. 1922. "On the Mathematical Foundations of Theoretical Statistics." *Philos Trans R Soc London A* 222: 309–68. https://doi.org/10.1098/rsta.1922.0009.

King, Aaron A., Dao Nguyen, and Edward L. Ionides. 2016. "Statistical Inference for Partially Observed Markov Processes via the R Package Pomp." *J Stat Softw* 69 (12): 1–43. https://doi.org/10.18637/jss.v069.i12.

Kitagawa, Genshiro. 1987. "Non-Gaussian State-Space Modeling of Nonstationary Time Series." *J Am Stat Assoc* 82 (400): 1032–41. https://doi.org/10.1080/01621459.1987.10478534.

Pawitan, Yudi. 2001. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford: Clarendon Press.

License, acknowledgments, and links

- This lesson is prepared for the Simulation-based Inference for Epidemiological Dynamics module at the Summer Institute in Statistics and Modeling in Infectious Diseases, SISMID.
- The materials build on previous versions of this course and related courses.
-
- Produced with R version 4.4.2 and pomp version 6.1.
- Compiled on 2024-07-24.

Back to Lesson
R code for this lesson