Quiz on Lesson 3

Edward L. Ionides, Aaron A. King, Qianying Lin

July 19, 2022

Question 1. Suppose that 10 replications of a particle filter, each using 10^3 particles, runs in 15 minutes with no parallelization. To look for a more precise likelihood evaulation, you consider running 20 replicates, each with 10^4 particles. How many minutes will this take, if you distribute the calculation across 4 cores?

(A) 50

(B) 60

(C) 75

(D) 120

(E) 300

Solution. C. Using the linear dependence, also called proportionality, of the computing effort on various algorithmic parameters, we calculate

$$5 \times (10000/1000) \times (20/10) \times (1/4) = 75.$$

Question 2. A particle filter is repeated 5 times to evaluate the likelihood at a proposed maximum likelihood estimate, each time with 10^4 particles. Suppose the log likelihood estimates are -2446.0, -2444.0, -2443.0, -2442.0, -2440.0. Which of the following is an appropriate estimate for the log likelihood at this parameter value and its standard error.

(A) Estimate $= -2443.0$, with standard error 1.0

(B) Estimate $= -2443.0$, with standard error 2.2

(C) Estimate $= -2443.0$, with standard error 5.0

(D) Estimate $= -2441.4$, with standard error 2.2

(E) Estimate $= -2441.4$, with standard error 1.4

Solution. E. Answers A, B and C estimate using a mean on the log scale. However, the particle filter provides an unbiased likelihood estimate on a natural scale but not on a log scale. Note that the particle filter also has some bias for most quantities on a natural scale, which reduces to zero as the number of particles tends to infinity, but it happens to be unbiased for the likelihood. The standard error for the log of the mean of the likelihoods can be computed by the delta method or a jack-knife, for example using the logmeanexp function in pomp.

```
ll <- c(-2446,-2444,-2443,-2442,-2440)
mean(ll)
[1] -2443
sd(ll)
[1] 2.236068
sd(ll)/sqrt(length(ll))
[1] 1
library(pomp)
logmeanexp(ll,se=TRUE)
                         se
-2441.422198      1.380688
```

Question 3. What is the log likelihood (to the nearest unit) of the Dacca cholera data for the POMP model constructed in pomp via

```
d <- dacca(deltaI=0.08)
```

with cholera mortality rate 8% and other parameters fixed at the default values.

(A) -3764

(B) -3765

(C) -3766

(D) -3767

(E) -3768

Solution. A, calculated as follows:

```
d <- dacca(deltaI=0.08)
library(doParallel)
my_cores <- detectCores()
registerDoParallel(my_cores)
bake(file="Q9.rds",{
  foreach(i=1:32,.combine=c) %dopar% {
    library(pomp)
    logLik(pfilter(d,Np=10000))
  }
}) -> cholera_loglik
logmeanexp(cholera_loglik,se=TRUE)
                         se
-3764.1894659      0.1320433
```

Question 4. Effective sample size (ESS) is one of the main tools for diagnosing the success of a particle filter. If you plot an object of class `pfilterd_pomp` (created by applying `pfilter` to a pomp object), the ESS is displayed. Suppose one or more time points have low ESS (say, less than 10) even when using a fairly large number of particles (say, 10^4). What is the proper interpretation?

(A) There is a problem with data, perhaps an error recording an observation.

(B) There is a problem with the model which means that it cannot explain something in the data.

(C) The model and data have no major problems, but the model happens to be problematic for the particle filter algorithm.

(D) At least one of A, B and C.

(E) Either A or B or both, but not C. If the model fits the data well, the particle filter is guaranteed to work well.

Solution. D. An example of a situation where the model fits the data well, but filtering is hard, arises when the measurement error is small relative to the process noise. In this case, the particles are scattered by the process noise and very few of them are compatible with the data due to the precise measurement. Thus, almost all the particles must be discarded as unfeasible given the data, corresponding to a low ESS.