Bonus: How many particles is enough?

Katie Montovan

2025-07-26

The number of particles chosen should affect the variability of the estimate. Here is one way you could look at the `est` and `se` across a range of values for `Np`.

```
si_step <- Csnippet("
  double infections = rbinom(S,1-exp(-Beta*I*dt));
  double recoveries = rbinom(I,1-exp(-gamma*dt));
  S += -infections + recoveries;
  I += infections - recoveries;
  NewI += infections;
")

si_rinit <- Csnippet("
  S = nearbyint((1-phi)*N);
  I = nearbyint(phi*N);
  NewI = 0;
")

#Something about the rmeasure function is leading to differences
si_rmeas <- Csnippet("
  NewCases = rnbinom_mu(k,rho*NewI);
  ")

si_dmeas <- Csnippet("
    lik = dnbinom_mu(NewCases,k,rho*NewI, give_log);
  ")

SI_Model <- pomp(
  data = simuldata,
    times="day",
    t0=0,
    rinit=si_rinit,
    rprocess=euler(si_step,delta.t= 1/4),
    rmeasure=si_rmeas,
    dmeasure=si_dmeas,
    statenames=c("S","I","NewI"),
    accumvars="NewI",
    paramnames=c("Beta","gamma","rho","k","N","phi"),
  params= c(N=1000, #population size
            phi = 0.004, #fraction of population infected at time 0
            Beta=0.0003, # rate of spread of illness - per day
            gamma=.1, #recovery rate
            rho=.8, #fraction of new infections that are recorded
            k=5 )
```

1

```
  )
```

Check that simulations work.

```r
sims2 <- SI_Model |>
  simulate(
    params= c(N=1000, #population size
               phi = 0.004, #fraction of population infected at time 0
               Beta=0.0003, # rate of spread of illness - per day
               gamma=.1, #recovery rate
               rho=.8, #fraction of new infections that are recorded
               k=5 ), #changes the variability of reporting accuracy
    nsim=5,
    format="data.frame",
    include.data=FALSE
  )


sims2 |>
  ggplot(aes(x=day,y=NewCases,group=.id,color=.id=="data"))+
  geom_line()+
  guides(color="none")+
  theme_bw()
```
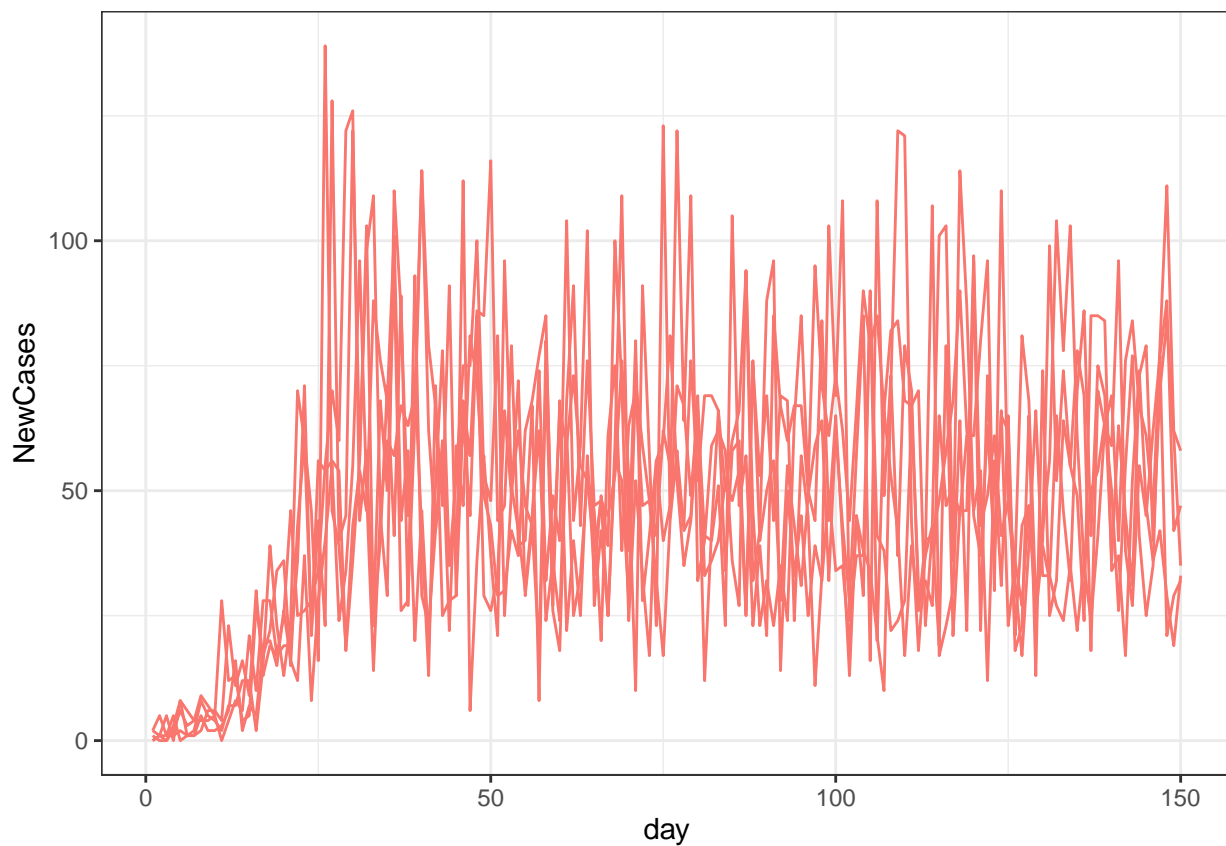

Run the particle filter for different values of Np.

```r
registerDoParallel()
registerDoRNG(294113554)
```

```
particles <- c(
  seq(1,100,by=5),
  seq(125,1000,by=25)
)
library(dplyr)
foreach (Npart=particles) %do% {
  foreach (i=1:20,.combine=c) %dopar% {
    library(pomp)
    SI_Model |>
      pfilter(Np=Npart)
  } -> pf
  pf |>
    logLik() |>
    logmeanexp(se=TRUE,ess=TRUE) -> L_pf
  c(Np=Npart,ll=L_pf)
} |>
  bind_rows() -> results
```

This plot shows the log-likelihood estimate and the bars show the 95% confidence interval for the log-likelihood based on the `se` estimate.

```
results |>
  ggplot(aes(x=Np,y=ll.est,ymin=ll.est-2*ll.se,ymax=ll.est+2*ll.se))+
  geom_point(color="purple")+
  geom_errorbar()+
  theme_bw()+
  labs(
    x="Number of particles in pfilter",
    y="Log-likelihood"
  )
```