Midterm 2, STATS 531/631 W26

In class on 4/15

Name: **UMID:**

Instructions. The test is closed book, and you are not allowed access to any notes. Any electronic devices in your possession must be turned off and remain in a bag on the floor.

For each question, circle one letter answer and provide some supporting reasoning.

Q1. Foundations of POMP models

Scientifically, our conclusions should not depend on the units of measurement we use, but we can make errors if we don't get the details right. Suppose our data are two years of weekly aggregated case reports of a disease and we have a continuous-time model solved numerically by an Euler timestep of size dt. Which of the following is a correct explanation of our options for properly implementing this in a pomp object called `po`?

(A). The measurement times, `time(po)`, should be in units of weeks, such as $1, 2, \dots, 104$. The latent process can be modeled using arbitrary time units, say days or weeks or years. The units of dt should match the time units of the **latent** process.
(B). The measurement times, `time(po)`, should be in units of weeks, such as $1, 2, \dots, 104$. The latent process can be modeled using arbitrary time units, say days or weeks or years. The units of dt should be in weeks (in practice, usually a fraction of a week) to match the units of the **measurement** times.
(C). The measurement times do not have to be in units of weeks. For example, we could use `time(po)`=1/52, 2/52, ..., 2. The latent process and dt should use the same units of time as the measurement times.
(D). The measurement times do not have to be in units of weeks. For example, we could use `time(po)`=1/52, 2/52, ..., 2. The latent process can also use arbitrary units of time, which do not necessarily match the units of the measurement times. The units of dt should match the units used for the **latent** process.
(E). The measurement times do not have to be in units of weeks. For example, we could use `time(po)`=1/52, 2/52, ..., 2. The latent process can also use arbitrary units of time, which do not necessarily match the units of the measurement times. The units of dt should match the units used for the **measurement** times.

Q2. Likelihood evaluation; the particle filter

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

Q3. Likelihood maximization; iterated filtering

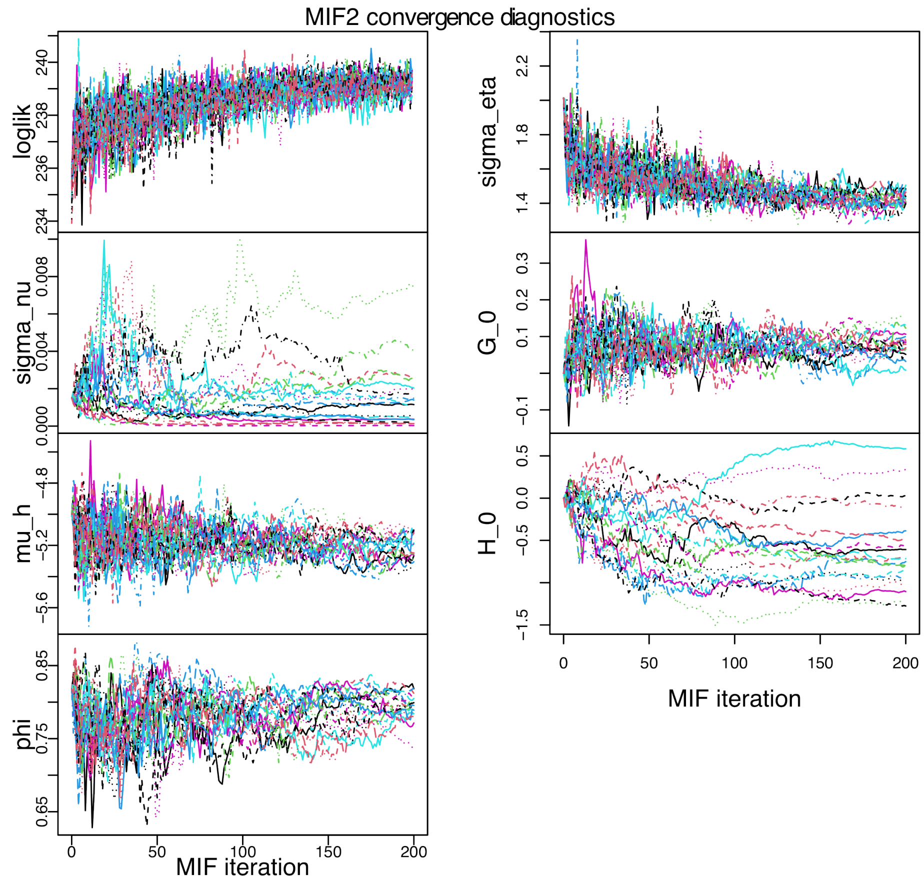

Figure 1: Iterated filtering diagnostic plot

The iterated filtering convergence diagnostics in figure 1 come from a student project investigating the market value of Gamestop. What is the best interpretation?

(A). Everything seems to be working fine. The likelihood is climbing. The replicated searches are giving consistent runs. The spread of convergence points for σ_ν and H_0 indicates weak identifability, which is a statistical fact worth noticing but not a weakness of the model.

(B). The consistently climbing likelihood is promising, but the failure of σ_ν and H_0 to converge needs attention. Additional searching is needed, experimenting with **larger** values of the random walk perturbation standard deviation for these parameters to make sure the parameter space is properly searched.

(C). The consistently climbing likelihood is promising, but the failure of σ_ν and H_0 to converge needs attention. Additional searching is needed, experimenting with **smaller** values of the random walk perturbation standard deviation for these parameters to make sure the parameter space is properly searched.

(D). The consistently climbing likelihood is promising, but the failure of σ_ν and H_0 to converge needs attention. This indicates weak identifiability which cannot be solved by improving the searching algorithm. Instead, we should change the model, or fix one or more parameters at scientifically plausible values, to resolve the identifiability issue before proceeding.

(E). Although the log likelihood seems to be climbing during the search, until the convergence problems with σ_ν and H_0 have been addressed we should not be confident about the successful optimization of the likelihood function or the other parameter estimates.

Q4. Data analysis: epidemiological models

A compartment model is first implemented as a system of ordinary differential equations (ODEs). This leads to qualitatively reasonable trajectories, but poor likelihood values. The researchers add stochasticity in an attempt to improve the fit of the model by interpreting the ODEs as rates of a Markov chain. The likelihood, maximized by iterated particle filtering, remains poor compared to ARMA benchmarks. In addition, the effective sample size for the particle filtering is low at many time points despite even using as many as 10^4 particles. Which of the following is the most promising next step?

(A). Increase to 10^5 particles, moving the computations to a cluster if necessary.
(B). Add noise to one or more rates to allow for overdispersion.
(C). Try adding extra features to the model to capture scientific details not present in the original model.
(D). Experiment with variations in the iterated filtering procedure; maybe more iterations, or a different cooling schedule.
(E). To address the possibility of reporting errors, see if the model fits better when the most problematic data points are removed.

Q5. Data analysis: financial models

The Heston model for volatility, V_n, is a stochastic volatility (SV) model with

$$V_n = (1-\phi)\theta + \phi V_{n-1} + \sqrt{V_{n-1}}\,\omega_n,$$

for $\omega_n \sim N[0, \sigma_\omega^2]$. The log return is $Y_n \sim N[0, V_n]$, conditional on V_n. A previous 531 project (W22, #14) fitted the Heston model to investment in Ethereum, a crypto currency. They obtained a log-likelihood of 34975.3, compared to 28587.4 for GARCH and 28977 for the SV model with leverage presented in class. Their iterated filtering convergence diagnostics are shown in figure 2. What is the best conclusion from this information?

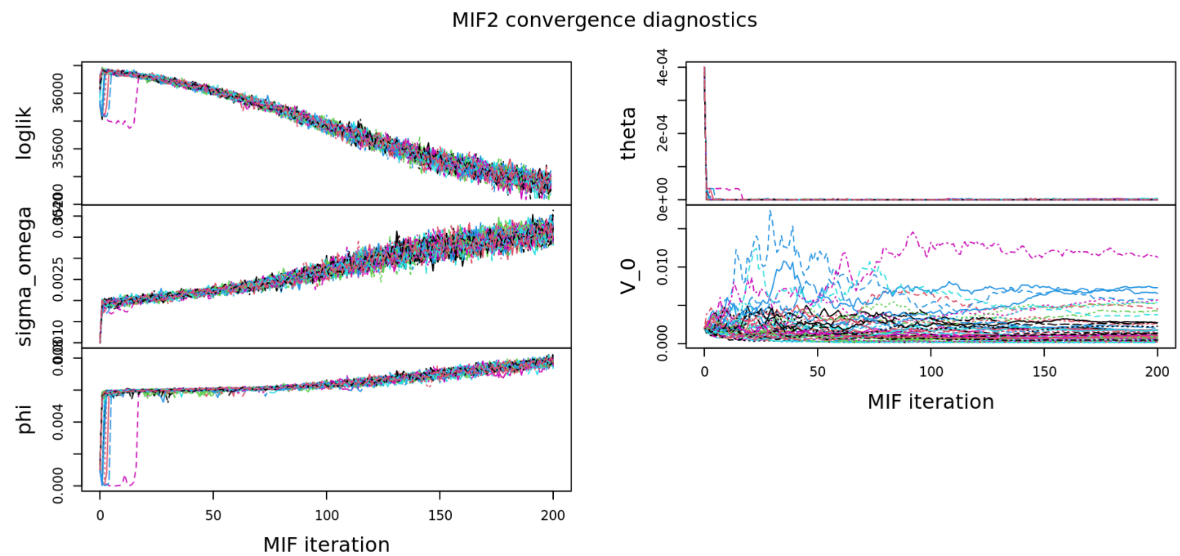

Figure 2: Diagnostic plot for fitting the Heston model

(A). The high likelihood shows this is a promising model despite the convergence problems identified in the figure. Attention to the diagnostics may lead to additional improvements.

(B). The most important diagnostic feature is the observation that the log-likelihood trace plot peaks and then declines. From the y-axis scale we see the decline is of order 1000 log units. This is evidence of substantial model misspecification which should be addressed.

(C). The most important dignostic feature is that the `theta` traces all drop quickly to zero. Since that is not a scientifically plausible value for the parameter, we can deduce that the model is unsuccessful despite its high likelihood.

(D). The most important diagnostic feature is that `phi` is close to zero and well identified. This shows that the volatility is close to constant, and is supported by the high likelihood.

(E). The decreasing likelihood and other convergence diagnostics problems show there is a problem with the model. Likely, there is a bug and the high likelihood obtained is simply an error.

Q6. Diagnostics and debugging for POMP models

A basic diagnostic plot is to simulate from the fitted model POMP model at the parameter value with the highest likelihood found so far in the search of the parameter space. If this simulation is qualitatively similar to the data, this suggests the data are consistent with the model. If the simulation is qualitatively different from the data, which of the following is the best next step?

(A). We should do additional global and local searches to make sure we have accurately maximized likelihood estimate. If the problem remains, we can try other approaches with more confidence.

(B). Visual comparisons of trajectories can be misleading. For example, there is no reason why a simulation from a model should match the timing of events in the dataset (e.g., recessions in economic data; outbreaks in epidemic data). Maybe our model is good despite the mismatching simulation.

(C). We should think hard about why the simulations don't match the data. There could easily be a missing scientific mechanism that needs to be included to explain the data. Unless we think about the science, we won't work this out.

(D). We should compare the log-likelihood against non-mechanistic benchmarks, such as ARMA models. If we can find a simple statistical model that has better AIC than our mechanistic model, that would strengthen the inference that there might be a problem with the data.

(E). We should inspect residuals (the difference between observed data and the one-step prediction mean) to help us identify which data points are explained poorly by the model.
