Take Home Midterm

DUE: Wed. Mar 16th, 2020.

1 **Look at the detrended O2 isotope ratio time series that relates to the (topical) issues of longerterm cycles in climate and, perhaps, "ice-age" cycles in climatic indicators. As we have briefly discussed, this area has a long history related to the(so-called) Milankovitch cycles in the orbital dynamics of the Earth; e.g. https://en.wikipedia.org/wiki/Milankovitch cycles. See also, for example, West (1996), Bayesian time series: Models and computations for the analysi sof time series in the physical sciences, in Maximum Entropy and Bayesian Methods 15, Eds: K. Hanson & R. Silver, Kluwer, 1996, pp23–34.**

Reverse the time series so that left-to-right is now moving forward in time, with the end of the time series being "currently". Interest here will, in part, be on forecasting over the next many thousands of years. Some Matlab code to read in the data, reverse the time to be "forward", and plot etc., is linked here at the course Schedule page. Then, the course TVAR code for class examples can be used and slightly modified to address the following questions about analysis of this series.

(a) **Fit a TVAR(18) model to this forward-time oxygen isotope series. Use TVAR state discount factor** $\delta = 0.99$ **and volatility discount** $\beta = 0.975$. **Use the prior specifications:** $\mathrm{m0} = \mathrm{zeros(p, 1)}; \mathrm{m0(1)} = 1; \mathrm{n0} = 5; \mathrm{s0} = 0.02; \mathrm{C0} = \mathrm{eye(p)/2}$ **Look at the "plug-in" estimated decomposition of the time series and the corresponding "plug-in" estimates over time of the wavelengths of the 4 dominant quasi-periodic components in the series. Discuss and summarise what these visuals suggest.**

(b) **Use TVAR forecasting code tvarforecast.m to generate a reasonable number (several 000s) of synthetic futures over the next** $240,000$ **years. (That is, the next** $k = 80$ **time points, since our data are at 3 kyear intervals). In general terms, do you think this is a "good" model for the series? From visual displays of predictions via synthetic futures compared to the real data, what kinds of (mismatch?) features standout? Include relevant supporting figures.**

(c) **Peaks and troughs in the series relate to the earth orbital dynamics which link (through this noisy data) to quasi-periodicities in features of global climate. So predicting peaks and troughs is of interest. On the basis of your Monte Carlo predictions, make inferences on the following:**

i **The time (from now $=T$) to the next $240,000$ year maximum level in (detrended) O2, i.e., the time at which, over the coming $240,000$ years, the maximum level is predicted to be achieved- be careful to communicate full uncertainties with your forecast.**

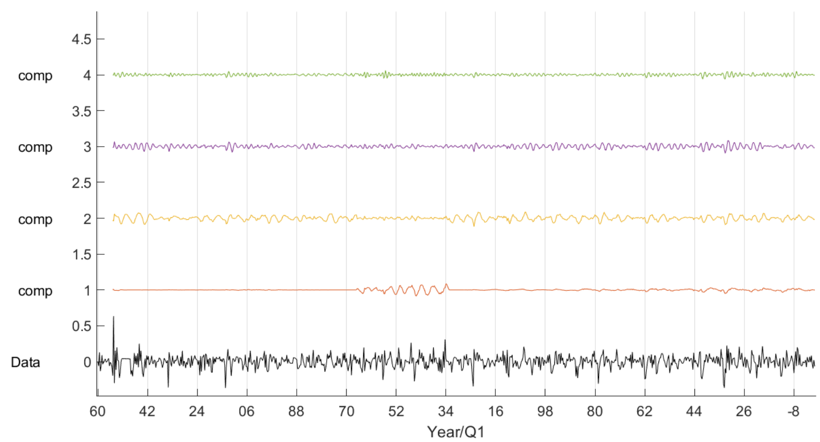
ii **What do you predict for the O2 value that will be achieved at that first maximum?**

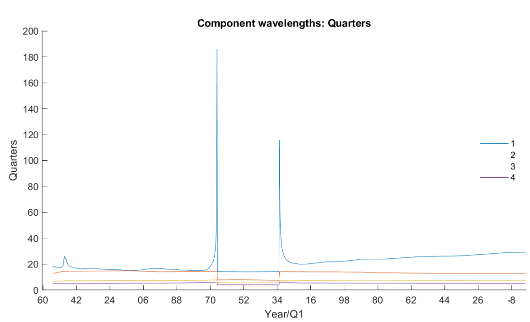
iii **As above, but now on the next 240,000year minimum.**

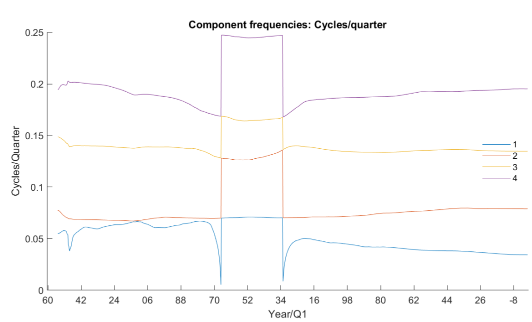
PROVE:

(a) Attached is the 4 dominant quasi cyclical processes. We find that mostly for the upper 2 processes has the same behaviour at all time, but the two lower processes seems to have compliment effects, as process 4 has more dominant behaviour at time 70-34 ($X = 313$), whereas the other have less. This suggest that the cyclical effects has been changed during these years.

Furthermore, looking at the wavelength, the change in the wavelength and frequency also happened at the same timestep. This suggest that at $X = 313$, a change of the process occurred

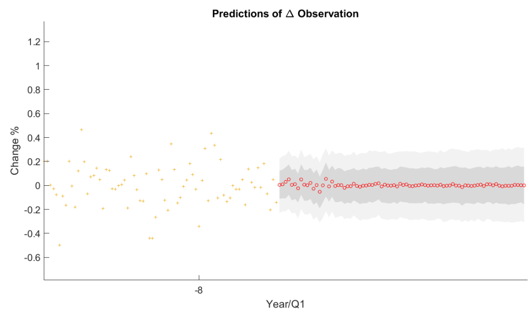


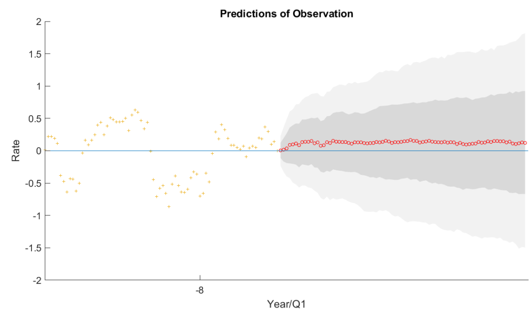




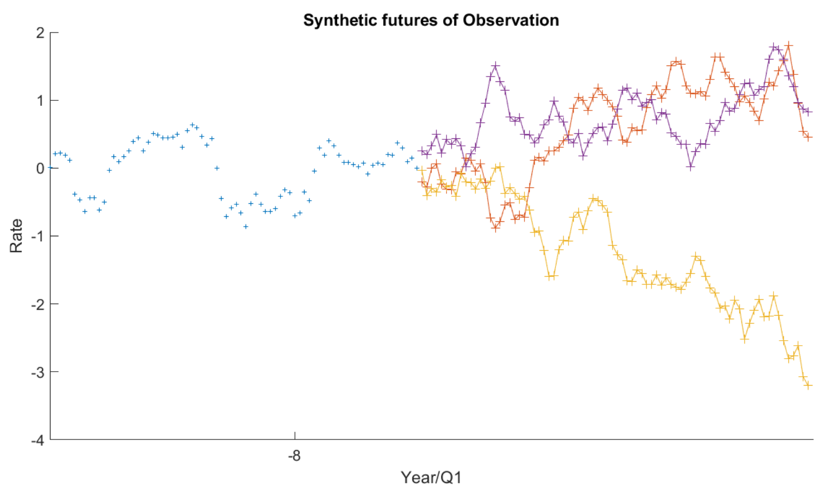
(b) I've over-predicted the later on process. But from the plot, we only need to look at a few more thousands of years.

First of all, in expectation of all MC samples, it is not hard to find that the prediction does not carry the cyclical effect in our original data. Maybe within a few iterations, the cyclical effect is still there, but it gradually vanished.





However, if we only look at a few single sampled process, we find that the model is able to capture the trend and pattern of previous data.

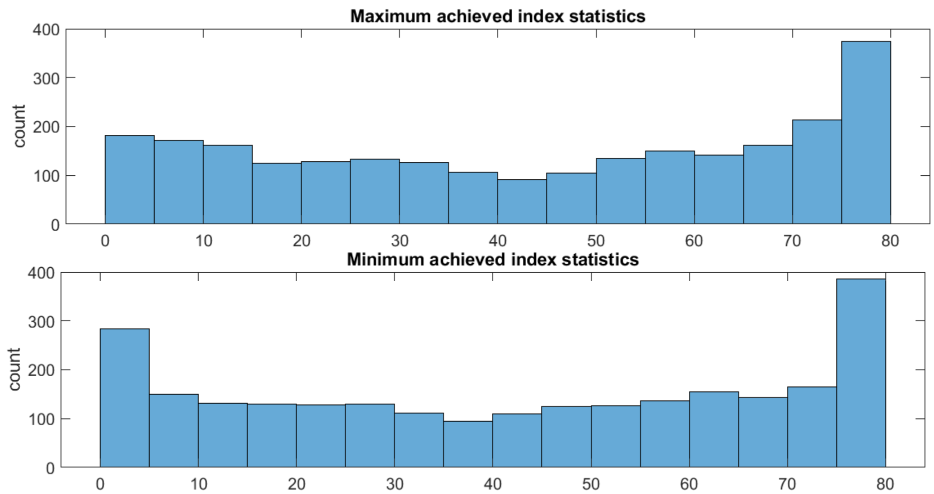


As a conclusion, the model is a good model in generating the prediction that captures the trend and cyclical process. However, there are too many different possibilities, and by averaging them will give us very stationary process that is not desired, and the prediction CI becomes very wide later on.

(c) For time step that achieved the maximum or minimum, we've calculated the following statistics using 2500 MC samples:

- Mode of maximum achieved time: 80 (240,000 years), Variance of maximum achieved time: 683.6928
- Mode of minimum achieved time: 80 (240,000 years), Variance of minimum achieved time: 715.5888

So, it seems like the model's prediction just follows an constant growth or decay autoregressive trend. Therefore, it leads most of the sampled paths to either gradually go higher and higher, or go lower and lower. And that magnitude of variance is unacceptable. Therefore, we're very uncertain about when the maximum or minimum will be achieved.



We've calculated the following statistics using 2500 MC samples:

- Mean of Max value: 1.3118, Variance of statistics: 0.9744
- Mean of Min value: -1.0844, Variance of statistics: 0.9029

The estimation still is not very certain, as we can see difference between max and min is about 2.5, which is just about 2 standard deviations. This means that the distribution of 2 statistics to overlap pretty much. So we should sample more paths to get a better estimate. Can use Chebychev inequality to approximate the needed sample size by controlling the variance.

2 **Full FFBS in the TVAR(18) analysis of the detrended Oxygen isotope series of Exercise 1. Run FFBS (use the tvarFFBS. m function) to generate a reasonably large MC sample from the full posterior over trajectories of TVAR states and volatilities.**

(a) **For each of these MC posterior samples, use the TVAR decomposition to compute the implied samples of the wavelengths and moduli of the 2 dominant (longest-wavelength) quasi-cyclical components in the series. Produce some relevant graphical summaries of some of the MC sampled trajectories and moduli over time. Comment on what is graphed in the context of retrospective analysis of this series and the use of TVAR DLMs, and why they are interesting in this applied study.**

(b) **Climatological and geological questions have been generated over potential changes in climate forcing mechanisms around 1million, i.e.,** $1,000$ **kyears, ago. These have been raised partly in response to the view that the 2 dominant earth-orbital dynamic mechanisms- of periods around 110kyears and 40kyears, respectively- seem to "look different" since that time that prior to that time.**

Use the MC samples of trajectories of moduli of the two dominant components to explore this via relevant posterior inferences. In particular:

i **Compute, plot and discuss some sampled trajectories of** $r_{t,1}/r_{t,2}$ **over time where** $r_{t,j}$ **is the modulus of component** j **at time** t

ii **Compute the MC estimate of** $\Pr(r_{t,1} > r_{t,2}|\mathcal{D}_T)$ **for each time** t **and plot over time.**

On the basis of these summaries of your analysis, comment on the question of change a million years ago.

There is code and examples on TVAR models in the course code base. Example code/scripts explore TVAR analyses of the quarterly changes in US inflation series and a subset of the EEG series from class discussion/slides.

PROVE:

(a) For wavelength, we observes that the 2 wavelengths are all relatively stable. However, we may observe that recently (at the end of the series), the wavelength posterior starts to have heavier tails. This tells us that recently, we're more uncertain about the wavelength. Things might be different.

For moduli, we find that the 2 dominant processes' modulis are all smaller than 1. Therefore, we conclude that the process is a stable process. Besides, the tail also get fatter CI, telling us the increasing uncertainty of the distribution.

(b) I've sampled 1000 MC samples from each time step. The vertical line represent the 1 million ago supposed to change time step. It seems like the ratio is staying

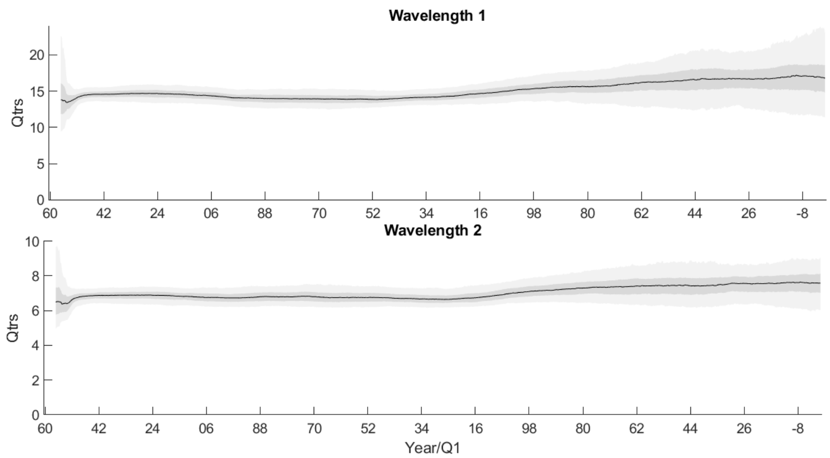


Figure 1: Wavelengths

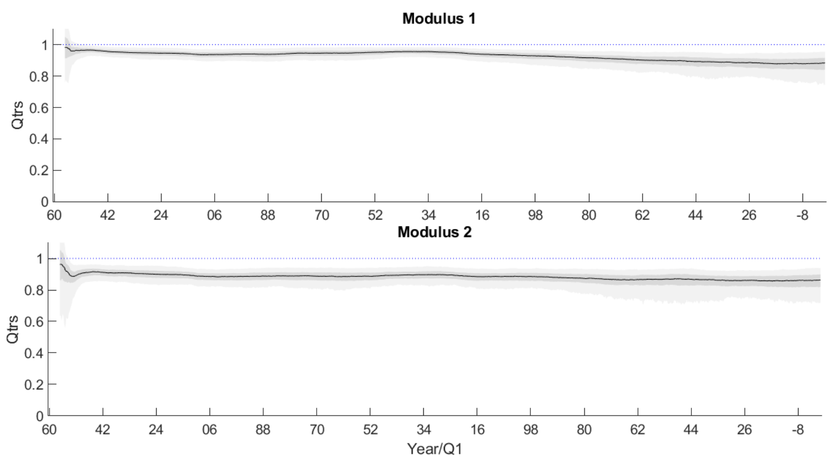


Figure 2: moduli

relatively stationary, whereas the probability of being bigger modulus is gradually increasing. I've also looked at the variance plot of process 1 and process 2, indicating that the second process (40k period) start to have larger variances. This could be the reason that the probability is going lower.

Therefore, I would conclude that after that 1 million years ago, 40k period process is gradually getting more dominant. We're getting into a climate cycle with higher frequency.

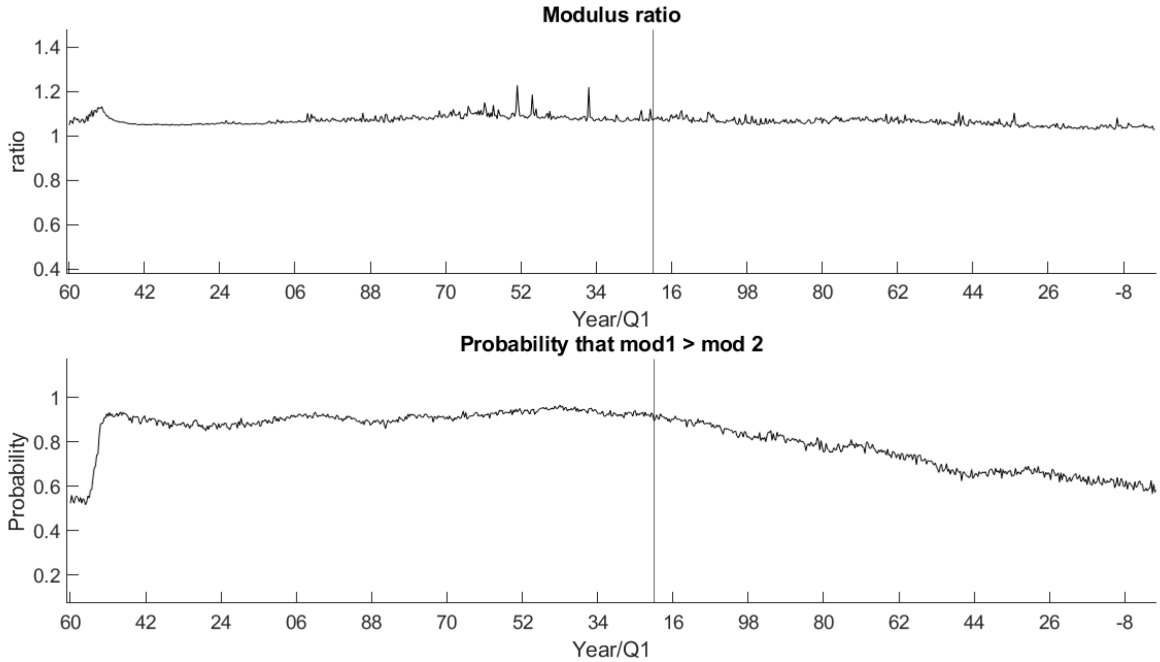


Figure 3: moduli

3 **The beta-gamma discount random walk model that we have used normal precisions (inverse variances) can be used in other contexts where a positive model parameter is "locally constant" but varies over time. One example is time series of counts at relatively low levels, including zero values, where Poisson/gamma models are very relevant**

Use/state theoretical results already defined in the normal DLM SV model- do not prove them from scratch. Refer to material in class slides and, particularly, HW5 on the beta/gamma model, some of which recapitulates Problem 4 of P&W Chapter 4**, Section** 4.6**, but with some differences in notation. That earlier context relates to the stochastic precision** $\phi_t = v_t^{-1}$ **in the normal DLM; the change here is that** ϕ_t **is a Poisson mean process. You have already worked through the relevant forward filtering, backwards smoothing and sampling results in HW#5 for the in the normal SV model. The application here to the Poisson time series model and essentials of analysis are the same with relevant notational changes.**

Suppose you have a time series problem with count data, $x_t = 0, 1, 2, \ldots,$ **over equally-spaced time** t**, and model the data as conditionally Poisson,** $x_t|\phi_t \sim Po\left(\phi_t\right)$

At a time $t-1$, assume that the current posterior on the mean (a.k.a. level) of the Poisson time series is $\phi_{t-1}|\mathcal{D}_{t-1} \sim Ga\left(a_{t-1}, a_{t-1}/m_{t-1}\right)$ for some shape parameter $a_{t-1} > 0$ and mean $E\left(\phi_{t-1}|\mathcal{D}_{t-1}\right) = m_{t-1} > 0$

(a) **The Poisson mean evolves over $(t-1,t)$ to $\phi_t = \phi_{t-1}\eta_t/\beta$ where η_t is independent of ϕ_{t-1} with $\eta_t \sim \mathrm{Be}\left(\beta a_{t-1}, (1-\beta)a_{t-1}\right)$ for some discount factor $\beta \in (0,1)$ What is the implied time t prior $p\left(\phi_t|\mathcal{D}_{t-1}\right)$?**

(b) **What is the 1 -step ahead forecast mean $E\left(x_t|\mathcal{D}_{t-1}\right)$?**

(c) **What is the 1 -step ahead predictive density $p\left(x_t|\mathcal{D}_{t-1}\right)$?**

(d) **Show that the posterior for $\phi_t|\mathcal{D}_t$ is gamma, $Ga\left(a_t, a_t/m_t\right)$ and give expressions for a_t, m_t. Show that the posterior mean $m_t \equiv E\left(\phi_t|\mathcal{D}_t\right)$ can be written as $m_t = r_t m_{t-1}$ where $r_t = \left(\beta a_{t-1} + x_t\right)/\left(\beta a_{t-1} + m_{t-1}\right)$. Comment on how m_t depends on x_t, and on the effect of large values of a_{t-1}**

(e) **Assuming a time series observed over times $1:T$, look back to time $t-1 < T$. Discuss how you can simulate the time $t-1$ retrospective distribution for $\phi_{t-1}|\phi_t, \mathcal{D}_T$ conditional on any value of ϕ_t. Comment on relevant theory and conditional independence structure.**

(f) **How do you simulate from the full posterior of the trajectory $\phi_{1:T} = \{\phi_1, \ldots, \phi_T\}$ given the full observed series $\mathcal{D}_T$?**

PROVE:

(a)

$$\phi_t|\mathcal{D}_{t-1} = \eta_t/\beta\phi_{t_1}|\mathcal{D}_{t-1}$$

As we've already derived in the previous homework, and we're not adding new data point. Therefore,

$$\phi_t|\mathcal{D}_{t-1} \sim \mathrm{Ga}(a_{t-1}\beta, \frac{a_{t-1}\beta}{m_{t-1}})$$

(b)

$$\begin{aligned}
\mathbb{E}[x_t|\mathcal{D}_{t-1}] &= \mathbb{E}_{\phi_t|\mathcal{D}_{t-1}}[\mathbb{E}_{x_t|\phi_t,\mathcal{D}_{t-1}}[x_t]] \\
&= \mathbb{E}_{\phi|\mathcal{D}_{t-1}}[\phi_t] \\
&= m_{t-1}
\end{aligned}$$

(c)

$$
\begin{aligned}
\mathbb{P}(x_t|\mathcal{D}_{t-1}) &= \int_{\mathbb{R}} \mathbb{P}(x_t|\phi_t, \mathcal{D}_{t-1})\mathbb{P}(\phi_t|\mathcal{D}_{t-1})d\phi_t \\
&= \frac{\left(\frac{a_{t-1}\beta}{m_{t-1}}\right)^{a_{t-1}\beta}}{\Gamma(a_{t-1}\beta)} \int_{\mathbb{R}} \phi_t^{a_{t-1}\beta-1} \exp\{-\frac{a_{t-1}\beta}{m_{t-1}}\phi_t\} \frac{\phi_t^{x_t}\exp\{-\phi_t\}}{x_t!} d\phi_t \\
&= \frac{\left(\frac{a_{t-1}\beta}{m_{t-1}}\right)^{a_{t-1}\beta}}{x_t!\Gamma(a_{t-1}\beta)} \int_{\mathbb{R}} \phi_t^{(a_{t-1}\beta+x_t)-1} \exp\{-(1+\frac{a_{t-1}\beta}{m_{t-1}})\phi_t\} d\phi_t \\
&= \frac{\Gamma(a_{t-1}\beta+x_t)}{\left(1+\frac{a_{t-1}\beta}{m_{t-1}}\right)^{(a_{t-1}\beta+x_t)}} \frac{\left(\frac{a_{t-1}\beta}{m_{t-1}}\right)^{a_{t-1}\beta}}{x_t!\Gamma(a_{t-1}\beta)} \\
&\quad \int_{\mathbb{R}} \frac{\left(1+\frac{a_{t-1}\beta}{m_{t-1}}\right)^{(a_{t-1}\beta+x_t)}}{\Gamma(a_{t-1}\beta+x_t)} \phi_t^{(a_{t-1}\beta+x_t)-1} \exp\{-(1+\frac{a_{t-1}\beta}{m_{t-1}})\phi_t\} d\phi_t \\
&= \frac{\Gamma(a_{t-1}\beta+x_t)}{\left(1+\frac{a_{t-1}\beta}{m_{t-1}}\right)^{(a_{t-1}\beta+x_t)}} \frac{\left(\frac{a_{t-1}\beta}{m_{t-1}}\right)^{a_{t-1}\beta}}{x_t!\Gamma(a_{t-1}\beta)} \\
&= \frac{\Gamma(a_{t-1}\beta+x_t)}{x_t!\Gamma(a_{t-1}\beta)} \frac{\left(\frac{a_{t-1}\beta}{m_{t-1}}\right)^{a_{t-1}\beta}}{\left(1+\frac{a_{t-1}\beta}{m_{t-1}}\right)^{a_{t-1}\beta}\left(1+\frac{a_{t-1}\beta}{m_{t-1}}\right)^{x_t}} \\
&= \frac{\Gamma(a_{t-1}\beta+x_t)}{x_t!\Gamma(a_{t-1}\beta)} \left(\frac{1}{1+\frac{a_{t-1}\beta}{m_{t-1}}}\right)^{x_t} \left(\frac{\frac{a_{t-1}\beta}{m_{t-1}}}{1+\frac{a_{t-1}\beta}{m_{t-1}}}\right)^{a_{t-1}\beta} \\
&= \binom{a_{t-1}\beta+x_t-1}{x_t} \left(\frac{1}{1+\frac{a_{t-1}\beta}{m_{t-1}}}\right)^{x_t} \left(1-\frac{1}{1+\frac{a_{t-1}\beta}{m_{t-1}}}\right)^{a_{t-1}\beta} \\
&\sim \mathrm{NB}(a_{t-1}\beta, \frac{1}{1+\frac{a_{t-1}\beta}{m_{t-1}}})
\end{aligned}
$$

(d)

$$
\begin{aligned}
\mathbb{P}(\phi_t|\mathcal{D}_t) &\propto \mathbb{P}(\phi_t|\mathcal{D}_{t-1})\mathbb{P}(x_t|\phi_{t-1},\mathcal{D}_{t-1}) \\
&\propto \phi_t^{(a_{t-1}\beta+x_t)-1}\exp\{-(1+\frac{a_{t-1}\beta}{m_{t-1}})\phi_t\} \\
&\sim \mathrm{Ga}(a_{t-1}\beta+x_t, \frac{m_{t-1}+a_{t-1}\beta}{m_{t-1}}) \\
&\sim \mathrm{Ga}(a_t, \frac{a_t}{m_t}) \\
a_t &= a_{t-1}\beta + x_t \\
m_t &= \frac{m_{t-1}(a_{t-1}\beta+x_t)}{a_{t-1}\beta+m_{t-1}} \\
&= m_{t-1}\left(\frac{a_{t-1}\beta+x_t}{a_{t-1}\beta+m_{t-1}}\right) \\
&= m_{t-1}\left(1+\frac{x_t-m_{t-1}}{m_{t-1}+a_{t-1}\beta}\right) = m_{t-1}r_t \\
r_t &= 1+\frac{x_t-m_{t-1}}{m_{t-1}+a_{t-1}\beta} = \frac{a_{t-1}\beta+x_t}{a_{t-1}\beta+m_{t-1}}
\end{aligned}
$$

We see that it seems like the bigger the x_t, the more "enlargement" effect that it will bring up to renormalize m_{t-1} to m_t. If it is smaller than m_{t-1}, it also shrinks m_t.

When we get a big a_{t-1}, we end up with a very hard to change renormalization of m_{t-1} to m_t, as big denominator will bring the multiplicative change very close to 1. Basically, it means a strong prior population, leaving the distribution very hard and rigid to adapt to upcoming data.

(e) This is the same as the question as in homework. We re-parametrize and find that

$$
\phi_{t-1} = \beta\phi_t + \gamma \qquad \gamma \sim \mathrm{Ga}((1-\beta)a_t, \frac{a_t}{m_t}) \quad \text{with } \gamma \perp\!\!\!\perp \phi_t
$$

We can just apply this result, and treat the reversed process as a re-parametrized original process.

(f) There are many ways. Take a prior for ϕ_0, then sample a path of ϕ_i till the end. Then reverse, use the backward sample to sample another path of ϕ_i until we reach ϕ_0. Then from the distribution of ϕ_0, we may re-sample back to ϕ_n ...

By doing this, we get many paths, and we can average them to get simulated processes.

4 **The time series intrusionevents.txt (in the course Data folder) represents a time series of monthly attempted intrusion events in a secure IT system over 52 biweekly periods. Assume that the Poisson/gamma discount model of Exercise 3 is adopted with the initial prior** $\phi_0|\mathcal{D}_0 \sim Ga(a_0, a_0/m_0)$ **with** $a_0 = 25, m_0 = 14.5$

(a) **Explore choice of discount factor** β. **Consider some values on the range 0.8:(0.01): 0.99 Choose a value you regard as relevant for this data set, and justify the choice.**

(b) **At the chosen value of** β, **implement FFBS to generate Monte Carlo samples of the full trajectory** $\phi_{1:T}$ **over time up to the end** $T = 52$. **Use this to generate a large MC sample size of trajectories. Summarize inference in relevant graphical display(s).**

(c) **The security protocols of the IT system were substantially modified in time period 44. Management is now interested in a formal statistical assessment of if- and, if so, how that has impacted the level of intrusion events. For example, did the underlying level of intrusion events change as a result of those modifications? And, if so, how? Discuss and provide responses to these general questions, including (i) some comments on the** x_t **data series itself, and then (ii) formal inference on** ϕ_{52}/ϕ_{44}

(d) **Management is also interested in reliable predictions of the total number of intrusion events in the next** k **time periods. Describe how you would formally address the question.**

PROVE:

(a) To explore β, I decide to find the MLE. As we've proved previously that

$$\mathbb{P}(x_t|\mathcal{D}_{t-1}) = \text{NB}(a_{t-1}\beta, \frac{1}{1 + \frac{a_{t-1}\beta}{m_{t-1}}})$$

Therefore, we know that

$$\begin{aligned}
\mathbb{P}(x_{1:T}) &= \mathbb{P}(x_{2:T}|x_1) \int_{\mathbb{R}} \mathbb{P}(x_1|\phi_1)\mathbb{P}(\phi_1)\phi_1 \\
&= \prod_{i=1}^{n-1} \mathbb{P}(x_{i+1}|\mathcal{D}_i) \int_{\mathbb{R}} \mathbb{P}(x_1|\phi_1)\mathbb{P}(\phi_1)\phi_1 \\
&= \prod_{i=0}^{n-1} \text{NB}(a_i\beta, \frac{1}{1 + \frac{a_i\beta}{m_i}})
\end{aligned}$$

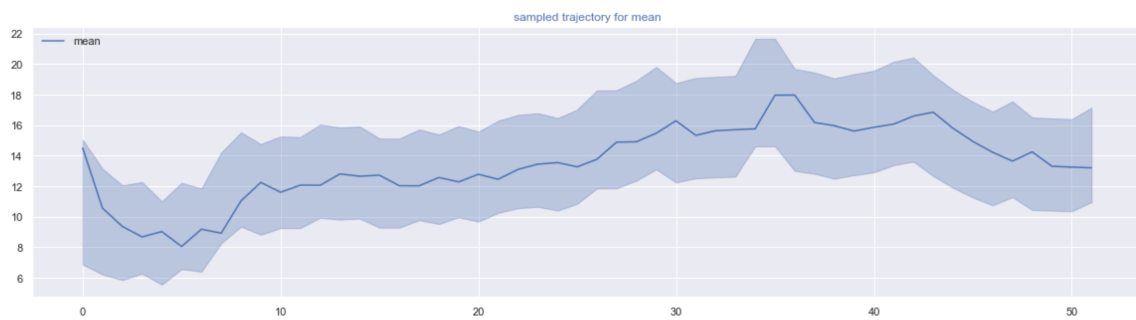
And, to easy calculation, we use log likelihood instead, we get

$$\ell(x_{1:T}) = \sum_{i=0}^{n} \log \binom{a_i\beta + x_{i+1} - 1}{x_{i+1}} - x_{i+1}\log(1 + \frac{a_i\beta}{m_i}) + a_i\beta\left(\log(\frac{a_i\beta}{m_i}) - \log(1 + \frac{a_i\beta}{m_i})\right)$$

$$= \sum_{i=0}^{n} \log \binom{a_i\beta + x_{i+1} - 1}{x_{i+1}} - x_{i+1}\log(m_i + a_i\beta) + (x_{i+1} + a_i\beta)\log(m_i) + a_i\beta\log(a_i\beta)$$

We may compare the likelihood value return by each β and choose the one that is the biggest. I calculated, and the answer is that $\beta = 0.82$ is the best choice.

(b) We find the sampled trajectory has the following distribution. The mean follows the following process, whereas the variance is relatively stable



(c) From the data itself, we don't see much of a difference. From the ϕ process, I also don't see any difference. For inference uppon ϕ_{52}/ϕ_{44}, we know it should follow a beta prime distribution , as from homework 5 we've shown ϕ_t has marginal distribution of $\text{Ga}(\beta^k a, \beta^k \frac{a}{m})$. The process has not changed, then we know that

$$\beta^{-8}\frac{\phi_{52}}{\phi_{44}} \sim \frac{\text{Ga}(\beta^8 a_{44}, \beta^8 \frac{a_{44}}{m_{44}})}{\beta^8 \text{Ga}(a_{44}, \frac{a_{44}}{m_{44}})} = \frac{\frac{a_{44}}{m_{44}}\beta^8 \text{Ga}(\beta^8 a_{44}, 1)}{\beta^8 \frac{a_{44}}{m_{44}} \text{Ga}(a_{44}, 1)} = \text{BetaPrime}(\beta^8 a_{44}, a_{44})$$

We form the following hypothesis

$$H_0 : \text{The trajectory has not changed for } \phi \text{ after } 44$$
$$H_a : \text{The trajectory has changed smaller for } \phi \text{ after } 44$$

By testing p-value for $\beta^{-8}\frac{\phi_{52}}{\phi_{44}}$, we find that p-value is 0.214. Not very significant. Therefore, we have no significant statistical evidence to reject the null hypothesis that the process has changed.

(d) Given the posterior distribution of all $\theta_{1:55}$, we may sample many paths to the next k steps. After that, by averaging them, we get a trend of mean and also variance. We can then predict x based on this predicted sequence.

```
from scipy.stats import betaprime

stats = (beta** 8) *    (phisPred[51][0] * phisPred[43][1])   /  (phisPred[51][1] * phisPred[43][0])
a = (beta** 8) * phisPred[43][0]
b = phisPred[43][0]
pvalue = betaprime.cdf(stats, a, b)
pvalue
```

```
0.21435352604525507
```

Figure 4: p-value

[Last revised: March 16, 2020]