Bootstrap filter

In order to test a Python implementation of the bootstrap filter(which can be found here: https://github.com/rlk-ama/dissertation/bootstrap-filter/filter.py) we used the following simple model:

$$X_t = \phi X_{t-1} + V_t$$

$$Y_t = X_t + W_t$$

where $V_t, W_t \overset{iid}{\sim} \mathcal{N}(0,1)$ and $\phi = 0.95$.
We compared the values of $\mathrm{E}[X_t|y_{0:t}]$ given by our implementation and the ones given by a Kalman filter (package KalmanFilter of library pykalman) for 100 time steps and 100 particles. We also calculated the ESS as shown in Figure 1.

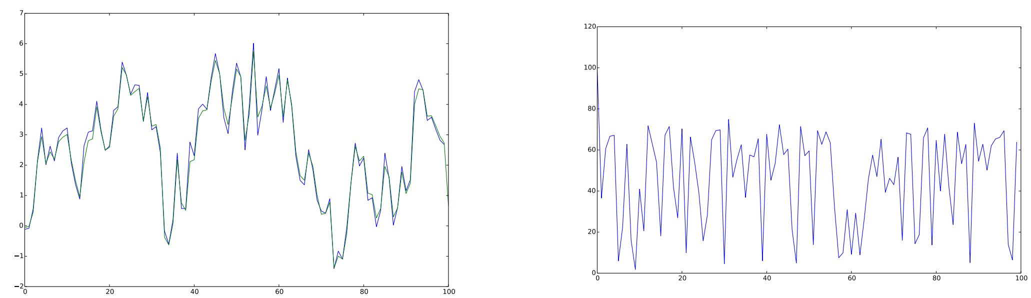


Figure 1: **(left)** Comparison of the expected value of the filtering distribution obtained using the Python implementation of the bootstrap filter **(blue)** and a Kalman filter **(green)**. **(right)** The ESS for the Python implementation of the bootstrap filter.

Moreover, we should check thatê

$$\frac{ESS_t}{N} \to K_t \text{ when } N \to \infty$$

where ESS_t is the effective sample size at step t, N the number of particles and K_t a constant. Figure 2 shows, for two different time steps, as N goes from 10 to 1000 that the above quantity indeed converges towards a constant. The ESS was calculated for simulation performed on simulated Ricker Map data using the gamma proposal described in the next section.

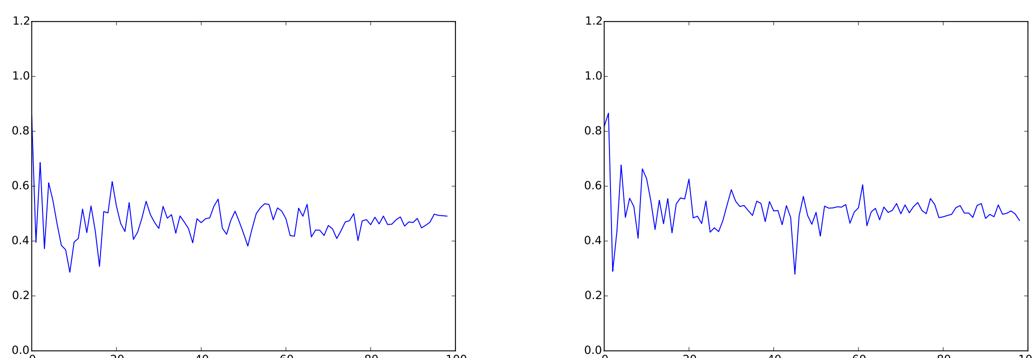


Figure 2: $\frac{ESS_t}{N}$ as N goes from 10 to 1000

Gamma approximation to Log-Normal distribution

The Ricker Map is the following model:

$$N_t = rN_{t-1}e^{-N_{t-1}}e^{Z_t}$$

$$Y_t \sim Poisson(\phi N_t)$$

where $Z_t \sim \mathcal{N}(0, \sigma^2)$.
Therefore $N_t \sim \log\mathcal{N}(\log(rN_{t-1}e^{-N_{t-1}}), \sigma^2)$. To approximate a Log-normal distribution we tried to minimize the Kullback-Leibler divergence from a Gamma to a Log-normal, ie we tried to minimize:

$$D_{KL}(P||Q)(\alpha, \theta) = \int_0^\infty p(z|\mu, \sigma^2)\log(\frac{p(z|\mu, \sigma^2)}{q(z|\alpha, \theta)})\mathrm{d}z$$

where p is the probability density function of a $\log\mathcal{N}(\mu, \sigma^2)$ and q of a Gamma with shape α and scale θ We have:

$$D_{KL}(P||Q)(\alpha, \theta) = C + (\alpha - 1)\log(\theta) + \log(\Gamma(\alpha)) - \alpha\mathrm{E_p}[\log(Z)] + \frac{1}{\theta}\mathrm{E_p}[Z]$$

Therefore:

$$\frac{\partial}{\partial\alpha}(D_{KL}(P||Q)) = \log(\theta) + \psi^{(0)}(\alpha) - \mathrm{E_p}[\log(Z)]$$

$$\frac{\partial}{\partial\theta}(D_{KL}(P||Q)) = \frac{\alpha}{\theta} - \frac{1}{\theta^2}\mathrm{E_p}[Z]$$

where $\psi^{(0)}$ is the digamma function.

Since $\mathrm{E_p}[\log(Z)] = \mu$ and $\mathrm{E_p}[Z] = e^{\mu + \frac{\sigma^2}{2}}$, we finally have that, setting the partial derivatives to zero:

$$\alpha = e^{\psi^{(0)}(\alpha) + \frac{\sigma^2}{2}}$$

$$\theta = \frac{1}{\alpha}e^{\mu + \frac{\sigma^2}{2}}$$

If we take $\psi^{(0)}(\alpha) \approx \log(\alpha) - \frac{1}{2\alpha}$ we finally have $\alpha = \frac{1}{\sigma^2}$ and $\theta = \frac{1}{\alpha}e^{\mu + \frac{\sigma^2}{2}}$. In our case we will thus approximate the distribution of N_t by

$$q(n_t|\alpha(n_{t-1}), \theta(n_{t-1})) = Gamma(\ \cdot\ ; \alpha(n_{t-1}), \theta(n_{t-1}))$$

where $\alpha(n_{t-1}) = \frac{1}{\sigma^2}$ and $\theta(n_{t-1}) = \sigma^2 e^{\log(rn_{t-1}e^{-n_{t-1}}) + \frac{\sigma^2}{2}}$, since $\mu = 0$.
Figure 3 shows the quality of the approximation which seems decent, although the tail of the log-normal seems heavier than the one of the gamma. This can be a problem when it comes to importance sampling.

Therefore our proposal for the bootstrap filter will be the following:

$$\begin{aligned} q_{t|t-1}(n_t|n_{t-1}, y_t) &\propto p(y_t|n_t)q(n_t|n_{t-1}) \\ &\propto e^{-\phi n_t}(\phi n_t)^{y_t} n_t^{\alpha(n_{t-1})-1} e^{-\frac{n_t}{\theta(n_{t-1})}} \end{aligned}$$

ie:

$$q_{t|t-1}(n_t|n_{t-1}, y_t) = Gamma(\ \cdot\ ; y_t + \alpha(n_{t-1}), \frac{\theta(n_{t-1})}{\theta(n_{t-1})\phi + 1})$$

Simulation carried out with $\log(r) = 3$, $\phi = 10$ and $\sigma = 0.3$. We also set $N_0 \sim Gamma(3, 1)$ in order to have N_0 around 3-4. A value of 3.8 as in Wood's paper for $\log(r)$ gave underflow and overflows in the Python code. Figure 4 shows results with prior proposal and Figure 5 with gamma proposal.

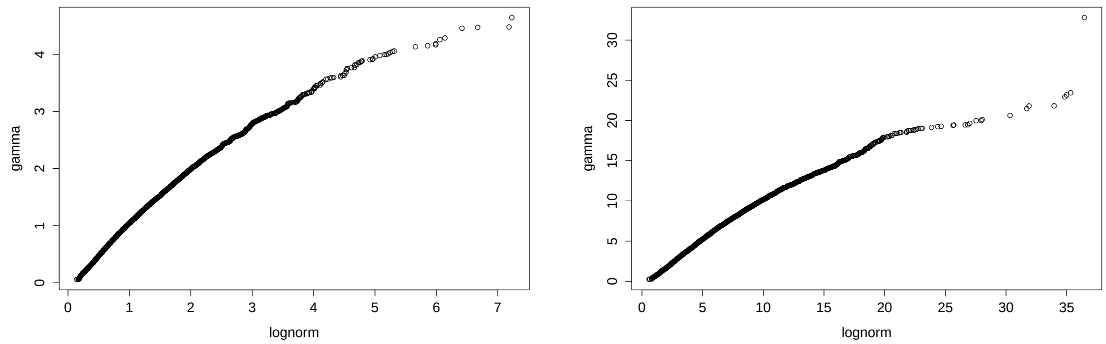


Figure 3: **(left)**QQ plot comparing a $\log\mathcal{N}(0, 0.3)$ and its gamma approximation. **(right)** QQ plot comparing a $\log\mathcal{N}(\log(5), 0.3)$ and its gamma approximation.

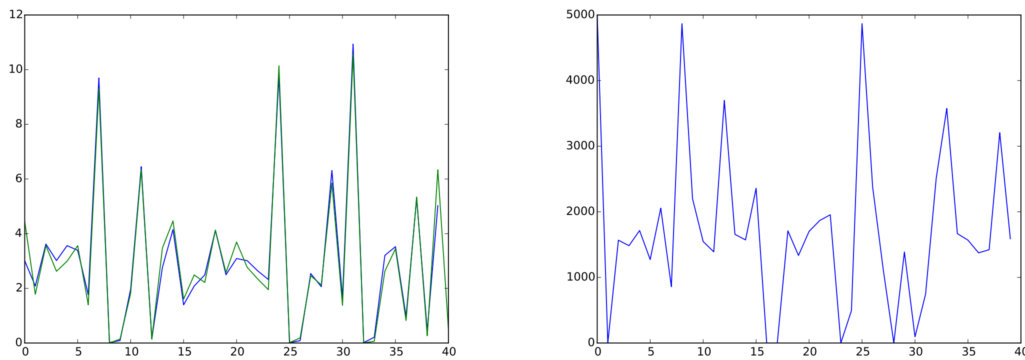


Figure 4: **(left)** Comparison of the expected value of the filtering distribution obtained using the prior proposal **(green)** simulated states **(blue)**. **(right)** The corresponding ESS.

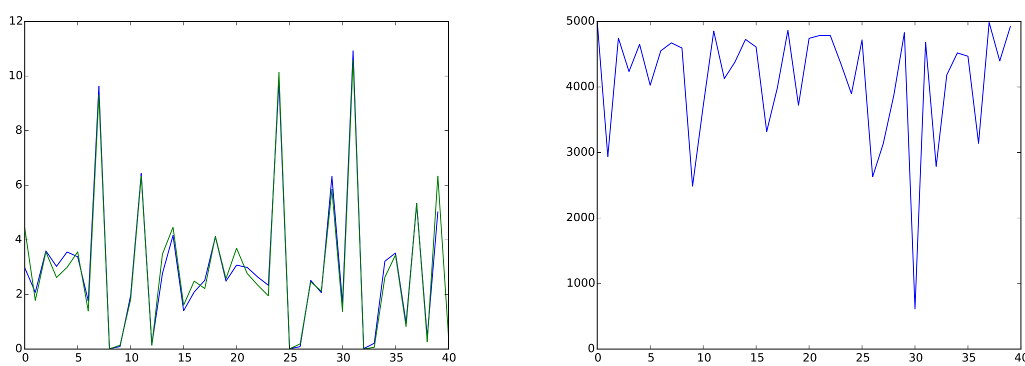


Figure 5: **(left)** Comparison of the expected value of the filtering distribution obtained using the gamma proposal **(green)** simulated states **(blue)**. **(right)** The corresponding ESS.

If on the contrary we minimized:

$$D_{KL}(P||Q)(\alpha,\theta) = \int_0^\infty q(z|\alpha,\theta)\log(\frac{q(z|\alpha,\theta)}{p(z|\mu,\sigma^2)})\mathrm{d}z$$

where p is the probability density function of a $\log\mathcal{N}(\mu,\sigma^2)$ and q of a Gamma with shape α and scale θ We have:

$$\begin{aligned} D_{KL}(P||Q)(\alpha,\theta) &= C - \frac{1}{\theta}\mathrm{E_q}[Z] + \alpha\mathrm{E_q}[\log(Z)] - \alpha\log(\theta) - \log(\Gamma(\alpha)) + \frac{1}{2\sigma^2}\mathrm{E_q}[(\log(Z)-\mu)^2] \\ &= C - \alpha + \alpha\psi^{(0)}(\alpha) - \log(\Gamma(\alpha)) + \frac{1}{2\sigma^2}(\psi^{(1)}(\alpha) + (\psi^{(0)}(\alpha) + \log(\theta) - \mu)^2) \end{aligned}$$

since $\mathrm{E_q}[Z] = \alpha\theta$, $\mathrm{E_q}[\log(Z)] = \psi^{(0)} + \log(\theta)$ and $\mathrm{Var_q}(\log(Z)) = \psi^{(1)}(\alpha)$
Therefore:

$$\frac{\partial}{\partial\alpha}(D_{KL}(P||Q)) = -1 + \frac{1}{2\sigma^2}\psi^{(2)}(\alpha) + \psi^{(1)}(\alpha)(\alpha + \frac{1}{\sigma^2}(\psi^{(0)}(\alpha) + \log(\theta) - \mu))$$

$$\frac{\partial}{\partial\theta}(D_{KL}(P||Q)) = \frac{1}{\theta\sigma^2}(\psi^{(0)}(\alpha) + \log(\theta) - \mu)$$

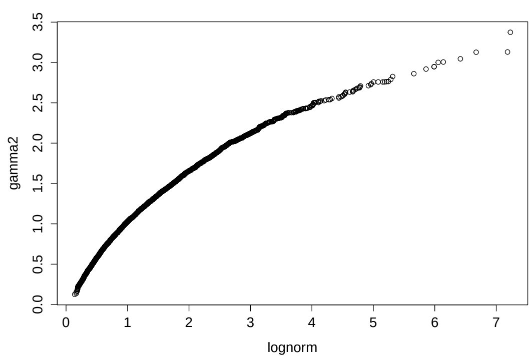
We finally have that, setting the partial derivatives to zero:

$$1 = \frac{1}{\sigma^2}\psi^{(2)}(\alpha) + \alpha\psi^{(1)}(\alpha)$$

$$\theta = e^{\mu - \psi^{(0)}(\alpha)}$$

If we take $\psi^{(0)}(\alpha) \approx \log(\alpha) - \frac{1}{2\alpha}$, $\psi^{(1)}(\alpha) \approx \frac{1}{\alpha} + \frac{1}{2\alpha^2} + \frac{1}{6\alpha^3}$ and $\psi^{(2)}(\alpha) \approx -\frac{1}{\alpha^2}$ we have $\alpha = \frac{6-\sigma^2}{3\sigma^2}$ and $\theta = \frac{1}{\alpha}e^{\mu + \frac{1}{2\alpha}}$.

Figure 6 shows the quality of the approximation. The tails of the approx are even lighter than those of the previous gamma approximation.



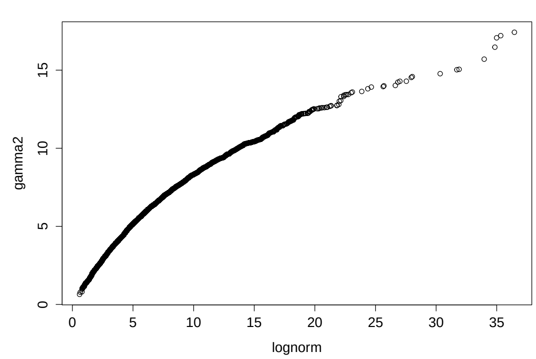


Figure 6: **(left)**QQ plot comparing a $\log\mathcal{N}(0,0.3)$ and its gamma approximation. **(right)** QQ plot comparing a $\log\mathcal{N}(\log(5),0.3)$ and its gamma approximation.

Figure 7 shows the result of the simulation with the gamma proposal based on our new approximation.

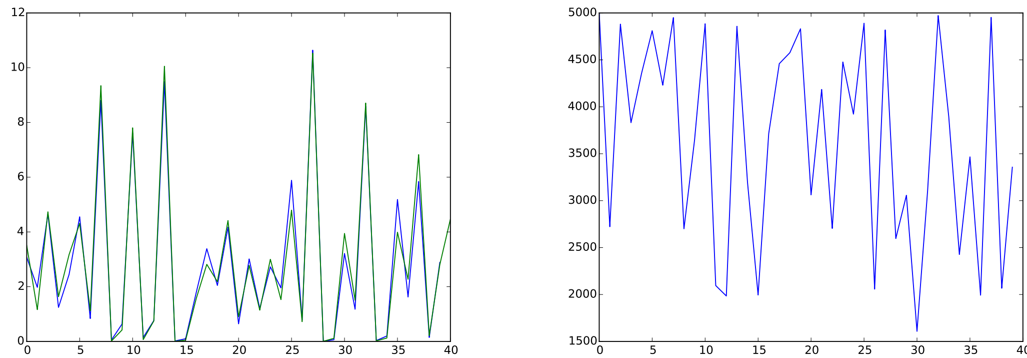


Figure 7: **(left)** Comparison of the expected value of the filtering distribution obtained using the prior proposal **(green)** simulated states **(blue)**. **(right)** The corresponding ESS.