Estatística Computacional - Lista 4

Gabriel Lima Novais

November 21, 2019

Exercise 1 (Kalman filter)

(Q1) Dadas as distribuições normais das variáveis W_t e V_t, junto com a regra da cadeia de Markov das variáveis X_t e Y_t, temos que:

- $f(x_t|x_{t-1}) = N(\phi x_{t-1}, \sigma_V^2)$
- $g(y_t|x_t) = N(x_t, \sigma_W^2)$

Veja que, (com uma forma de conferir o resultado acima) temos o seguinte resultado, quando buscampos o valore esperado e a variância:

$$E(x_t) = \phi E(x_{t-1}) + E(V_t) \implies E(x_t) = \phi x_{t-1} + 0 = \phi x_{t-1}$$

e ainda,

$$Var(x_t) = \phi^2 Var(x_{t-1}) + Var(V_t) \implies 0 + \sigma_V^2 = \sigma_V^2$$

(Q2) De acordo com o processo de atualização:

$$p(x_{t+1}|y) = \int f(x_{t+1}|x_t)p(x_t|y)dx_t =$$

$$= \int \frac{1}{\sqrt{2\pi\sigma_V^2}} exp(-\frac{(x_{t+1} - \phi x_t)^2}{2\sigma_V^2}) \frac{1}{\sqrt{2\pi\sigma_t^2}} exp(-\frac{(x_t - m_t)^2}{2\sigma_t^2})dx_t =$$

$$= \int \frac{1}{2\pi\sigma_V\sigma_t} exp(-\frac{(x_{t+1} - \phi x_t)^2}{2\sigma_V^2} - \frac{(x_t - m_t)^2}{2\sigma_t^2})dx_t =$$

$$= \int \frac{1}{\sqrt{4\pi^2\sigma_V^2\sigma_t^2}} exp(\frac{-(x_{t+1}^2\sigma_t^2 + m_t^2\sigma_V^2)}{4\sigma_V^2\sigma_t^2} - \frac{x_t^2(\phi^2\sigma_t^2 + \sigma_V^2) - 2(\sigma_t^2\phi x_{t+1} + m_t^2\sigma_V^2)x_t}{4\sigma_V^2\sigma_t^2})dx_t =$$

$$= \frac{1}{\sqrt{4\pi^2\sigma_V^2\sigma_t^2}} exp(\frac{-(x_{t+1}^2\sigma_t^2 + m_t^2\sigma_V^2)}{4\sigma_V^2\sigma_t^2}) \int exp(-\frac{x_t^2(\phi^2\sigma_t^2 + \sigma_V^2) - 2(\sigma_t^2\phi x_{t+1} + m_t^2\sigma_V^2)x_t}{4\sigma_V^2\sigma_t^2})dx_t =$$

Calculando a integral Gaussiana:

$$= \frac{1}{\sqrt{4\pi^2\sigma_V^2\sigma_t^2}} exp(\frac{-(x_{t+1}^2\sigma_t^2 + m_t^2\sigma_V^2)}{4\sigma_V^2\sigma_t^2}) \sqrt{\frac{2\pi\sigma_V^2\sigma_t^2}{\phi^2\sigma_t^2 + \sigma_V^2}} exp(\frac{4(\sigma_t^2\phi x_{t+1} + m_t^2\sigma_V^2)^2}{4\sigma_V^2\sigma_t^2(\phi^2\sigma_t^2 + \sigma_V^2)}) =$$

Simplificando tudo e depois rearranjando teremos que:

$$= \frac{1}{\sqrt{2\pi(\phi^2\sigma_t^2 + \sigma_V^2)}} exp(\frac{-(x_{t+1} - \phi m_t)^2}{2(\phi^2\sigma_t^2 + \sigma_V^2)}) \sim N(\phi m_t, \phi^2\sigma_t^2 + \sigma_V^2)$$

Pronto, obtemos a distribuição procurada.

(Q3) Sabemos que, por Bayes:

$$p(x_{t+1}|y_{t+1}) = \frac{g(y_{t+1}|x_{t+1})p(x_{t+1}|y_t)}{\int g(y_{t+1}|x_{t+1})p(x_{t+1}|y_t)dx_{t+1}}$$

Agora aplicando o as questões 1 e 2 anteriores e fazendo a proporcionalidade para achar o Kernel central da distribuição temos que:

$$p(x_{t+1}|y_{t+1}) = \frac{g(y_{t+1}|x_{t+1})p(x_{t+1}|y_t)}{\int g(y_{t+1}|x_{t+1})p(x_{t+1}|y_t)dx_{t+1}} \propto exp(-\frac{(y_{t+1} - x_{t+1})^2}{2\sigma_W^2} - \frac{(x_{t+1} - m_{t+1|t})^2}{2\sigma_{t+1|t}^2})$$

$$\propto exp(-\frac{(y_{t+1}^2 - 2y_{t+1}x_{t+1} + x_{t+1}^2)}{2\sigma_W^2} - \frac{(x_{t+1}^2 - 2x_{t+1}m_{t+1|t} + m_{t+1|t}^2)}{2\sigma_{t+1|t}^2}) \propto$$

$$\propto exp(-\frac{(-2y_{t+1}x_{t+1} + x_{t+1}^2)}{2\sigma_W^2} - \frac{(x_{t+1}^2 - 2x_{t+1}m_{t+1|t})}{2\sigma_{t+1|t}^2}) \propto$$

$$\propto exp(x_{t+1}^2(\frac{1}{\sigma_{t+1|t}^2} + \frac{1}{\sigma_W^2}) - 2x_{t+1}(\frac{m_{t+1|t}}{\sigma_{t+1|t}^2} + \frac{y_{t+1}}{\sigma_W^2})) \propto$$

$$\propto exp((\frac{1}{\sigma_{t+1|t}^2} + \frac{1}{\sigma_W^2})(x_{t+1}^2 - 2x_{t+1}\frac{(\frac{m_{t+1|t}}{\sigma_{t+1|t}^2} + \frac{y_{t+1}}{\sigma_W^2})}{(\frac{1}{\sigma_{t+1|t}^2} + \frac{1}{\sigma_W^2})})) \propto$$

$$\propto exp((\frac{1}{\sigma_{t+1|t}^2} + \frac{1}{\sigma_W^2})(x_{t+1}^2 - 2x_{t+1}\frac{(\frac{m_{t+1|t}}{\sigma_{t+1|t}^2} + \frac{y_{t+1}}{\sigma_W^2})}{(\frac{1}{\sigma_{t+1|t}^2} + \frac{1}{\sigma_W^2})} + (\frac{(\frac{m_{t+1|t}}{\sigma_{t+1|t}^2} + \frac{y_{t+1}}{\sigma_W^2})}{(\frac{1}{\sigma_{t+1|t}^2} + \frac{1}{\sigma_W^2})})^2)) \propto$$

$$\propto exp((\frac{1}{\sigma_{t+1|t}^2} + \frac{1}{\sigma_W^2})(x_{t+1} - \frac{(\frac{m_{t+1|t}}{\sigma_{t+1|t}^2} + \frac{y_{t+1}}{\sigma_W^2})}{(\frac{1}{\sigma_{t+1|t}^2} + \frac{1}{\sigma_W^2})})^2) \propto$$

$$\propto exp(\frac{(x_{t+1} - \frac{(\frac{m_{t+1|t}}{\sigma_{t+1|t}^2} + \frac{y_{t+1}}{\sigma_W^2})}{(\frac{1}{\sigma_{t+1|t}^2} + \frac{1}{\sigma_W^2})})^2}{(\frac{1}{\sigma_{t+1|t}^2} + \frac{1}{\sigma_W^2})^{-1}}) \sim N(\frac{(\frac{m_{t+1|t}}{\sigma_{t+1|t}^2} + \frac{y_{t+1}}{\sigma_W^2})}{(\frac{1}{\sigma_{t+1|t}^2} + \frac{1}{\sigma_W^2})}, (\frac{1}{\sigma_{t+1|t}^2} + \frac{1}{\sigma_W^2})^{-1})$$

E essa é a distribuição pedida, de forma que a média e variância encontram-se em função dos parâmetros anteriores.

(Q4) Este item é semelhante ao item 2 que foi feito anteriormente. Veja que:

$$p(y_{t+1}|y_t) = \int g(y_{t+1}|x_{t+1})p(x_{t+1}|y_t)dx_{t+1}$$

Mas sabemos que:

- $p(x_{t+1}|y_t) = N(\phi m_t, \phi^2\sigma_t^2 + \sigma_V^2)$
- $g(y_{t+1}|x_{t+1}) = N(x_t, \sigma_W^2)$

Então tal como foi feito no item 2 podemos fazer as contas todas e concluir qual seria a distribuição, ou então podemos fazer por analogia, uma vez que os cálculos são os mesmos mudando apenas algumas variáveis. Ou seja, sabemos que:

$$\int N(\phi x_t, \sigma_V^2)N(m_t, \sigma_t^2)dx_t = N(\phi m_t, \phi^2\sigma_t^2 + \sigma_V^2)$$

Para este problema devemos calcular:

$$\int N(x_t, \sigma_W^2)N(_t, \phi^2\sigma_t^2 + \sigma_V^2)dx_t = N(A, B)$$

Então simplesmente comparando a integral de cima com a de baixo podemos concluir que:

$$A = \phi m_t$$
$$B = \phi^2\sigma_t^2 + \sigma_V^2 + \sigma_W^2$$

E então está calculada a distribuição pedida.

Exercise 2 (SIS filter)

(Q1) A variável v_t é uma aproximação da distribuição de x dado y, isto é, de $p(x_t, y_{0:t})$. O algoritmo é chamado deste jeito, pois filtra essa distribuição.

(Q2) Através das estruturas de independência condicional do Hidden Markov Model podemos escrever que:

$$p_{Y_0}(y_0) = \int p_{Y_0,X_0}(x_0, y_0)dx_0 = \int p_{Y_0|X_0}(y_0|x_0)p_{X_0}(x_0)dx_0 = \int g(y_0|x_0)\mu(x_0)dx_0$$

E de forma similar ao que foi feito no desenvolvimento acima podemos então obter que:

$$p_{Y_0,Y_1}(y_0, y_1) = \int p_{Y_0,X_0,Y_1,X_1}(x_0, y_0, x_1, y_1)dx_0 =$$

$$= \int p_{Y_1|X_0,Y_0,X_1}(y_1, x_0, x_1, y_0)p_{X_1|X_0,Y_0}(x_1|x_0, y_0)p_{Y_0|X_0}(y_0|x_0)p_{X_0}(x_0)dx_{0:1} =$$

$$= \int g(y_1|x_1)f(x_1|x_0)g(y_0|x_0)\mu(x_0)dx_{0:1}$$

(Q3) Dado o que foi feito no item anterior é natural pensar que os estimadores sejam provenientes das integrais, e desta maneira se tornem médias das funções principais de acordo com o tamanho da amostra. Isto é:

$$\hat{p}_{Y_0}(y_0) = \frac{1}{n}\sum_{i=0}^{n} g(y_0|X_0^i)$$

$$\hat{p}_{Y_0,Y_1}(y_0, y_1) = \frac{1}{n}\sum_{i=0}^{n} g(y_1|X_1^i)g(y_0|X_0^i)$$

(Q4) Como a equação a ser demonstrada possui tanto índices em t como em $(t-1)$, será interessante colocar as estruturas do valor esperado como função do período atual e anterior. Assim:

$$E((\prod_{k=0}^{t} g(y_k|x_k))^2) = \int (\prod_{k=0}^{t} g(y_k|x_k))^2 \prod_{k=0}^{t} \mu(x_k)dx_{0:t} =$$

$$= \int (\prod_{k=0}^{t_1} g(y_k|x_k))^2 (\prod_{k=0}^{t-1} \mu(x_k))(g(y_t|x_t))^2 \mu(x_t)dx_{0:t} =$$

$$= \int (\prod_{k=0}^{t_1} g(y_k|x_k))^2 (\prod_{k=0}^{t-1} \mu(x_k))dx_{0:t-1} \int (g(y_t|x_t))^2 \mu(x_t)dt =$$

$$= E((\prod_{k=0}^{t-1} g(y_k|x_k))^2)E((g(y_t|x_t))^2)$$

Por outro lado para o termo sem estar elevado ao quadrado teremos que:

$$E(\prod_{k=0}^{t} g(y_k|x_k)) = \int \prod_{k=0}^{t} g(y_k|x_k) \prod_{k=0}^{t} \mu(x_k)dx_{0:t} = \int \prod_{k=0}^{t} g(y_k|x_k)\mu(x_k)dx_{0:t}$$

Utilizando o item 2 podemos obter que:

$$E(\prod_{k=0}^{t} g(y_k|x_k)) = p_{Y_{0:t}}(y_{0:t})$$

Agora para facilitar as contas, verifique que na diferença de duas variâncias uma com ínidice em t e outra em t-1 a seguinte igualdade é verdadeira:

$$V(a_t) - V(a_{t-1}) = E((a_t)^2) - E^2(a_t) - (E((a_{t-1})^2) - E^2(a_{t-1}))$$

Mas se $E(a_t) = E(a_{t-1}) = c$, então podemos simplificar a equação acima de forma que encontramos a seguinte expressão:

$$V(a_t) - V(a_{t-1}) = E((a_t)^2) - E((a_{t-1})^2$$

Portanto, como este é o caso das variâncias pedidas na equação a ser demonstrada uma vez que:

$$E\left(\left(\frac{\prod_{k=0}^{t} g(y_k|x_k)}{p_{Y_{0:t}}(y_{0:t})}\right)\right)^2 = 1 = E\left(\left(\frac{\prod_{k=0}^{t-1} g(y_k|x_k)}{p_{Y_{0:t-1}}(y_{0:t-1})}\right)\right)^2$$

Então podemos escrever que:

$$V\left(\left(\frac{\prod_{k=0}^{t} g(y_k|x_k)}{p_{Y_{0:t}}(y_{0:t})}\right)^2\right) - V\left(\left(\frac{\prod_{k=0}^{t-1} g(y_k|x_k)}{p_{Y_{0:t-1}}(y_{0:t-1})}\right)^2\right) =$$

$$= \frac{E((\prod_{k=0}^{t} g(y_k|x_k))^2)}{p^2_{Y_{0:t}}(y_{0:t})} - \frac{E((\prod_{k=0}^{t-1} g(y_k|x_k))^2)}{p^2_{Y_{0:t-1}}(y_{0:t-1})} =$$

$$= \frac{E((\prod_{k=0}^{t-1} g(y_k|x_k))^2 E((g(y_t|x_t))^2}{p^2_{Y_{0:t-1}}(y_{0:t})p^2_{Y_t}(y_t)} - \frac{E((\prod_{k=0}^{t-1} g(y_k|x_k))^2)}{p^2_{Y_{0:t-1}}(y_{0:t-1})} =$$

$$= \frac{E((\prod_{k=0}^{t-1} g(y_k|x_k))^2}{p^2_{Y_{0:t-1}}(y_{0:t})}(\frac{E((g(y_t|x_t))^2)}{p^2_{Y_t}(y_t)} - 1)$$

E assim obtemos a identidade da questão. Para obter a desigualdade positiva, basta verificar que se o valor esperado de uma função convexa é maior do que a função do valor esperado, então sabemos que o que está entre parênteses será maior que 1, e como o que está fora do parênteses é um termo ao quadrado então teremos que o valor será sempre positivo (Verificar a desigualdade de Jensein).

(Q5) Dúvida!

(Q6) Tomando como verdade a questão 5, podemos verificar que se a variância possui uma evolução com t de forma exponencial, pode ser inviável fazer o algoritmo para valores muito grandes da amostra, isto é , com t tendendo para o infinito.

Simulation question (Reversible jump MCMC)

Inicialmente foi necessário a criação de funções que pudessem implementar um Metropolis-Hastings para cada um dos modelos. Então criou-se o algoritmo de Reversible Jump para "alternar" os Kernels do modelo, e para isso, conforme vimos em sala de aula, podemos implementar a variavel auxiliar seguindo uma distribuição de Cauchy. Essa variável auxiliar é importante, pois com ela calculamos a taxa de aceitação do Reversible Jump. Idealmente a taxa de visita deveria ser proporcional a razão das taxas de normalização dos modelos, o que nos forneceria (inclusive tal como foi implementado em sala de aula) um valor em torno de $\sqrt{2\pi} \approx 2.5067$. A probabilidade de "jump" do modelo foi colocada em 0.3. Desta maneira o que se observou foi que a proporção de visitas foi de 2.4 para o primeiro modelo, e de 2.85 para o segundo modelo, que de fato são números próximos ao que deveria ser. Para os parâmetros podemos verificar que os gráficos abaixo demosntram as densidades de cada um dos dois modelos. E por fim o outro gráfico, que se assemelha a um código de barras, representa os pulos entre os Kernels.

Gráficos mencionados acima:

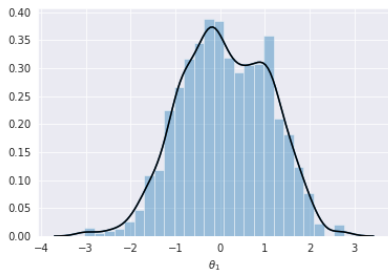


Figure 1: Primeiro Modelo

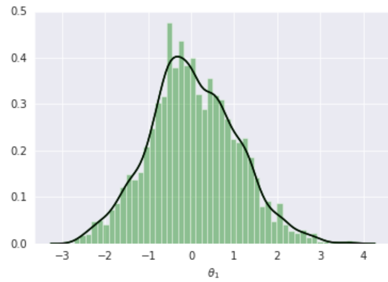


Figure 2: Segundo Modelo θ_1

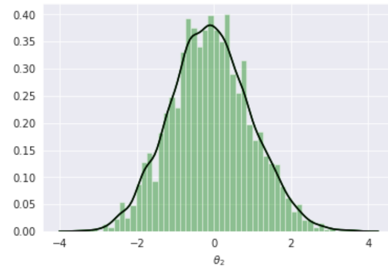
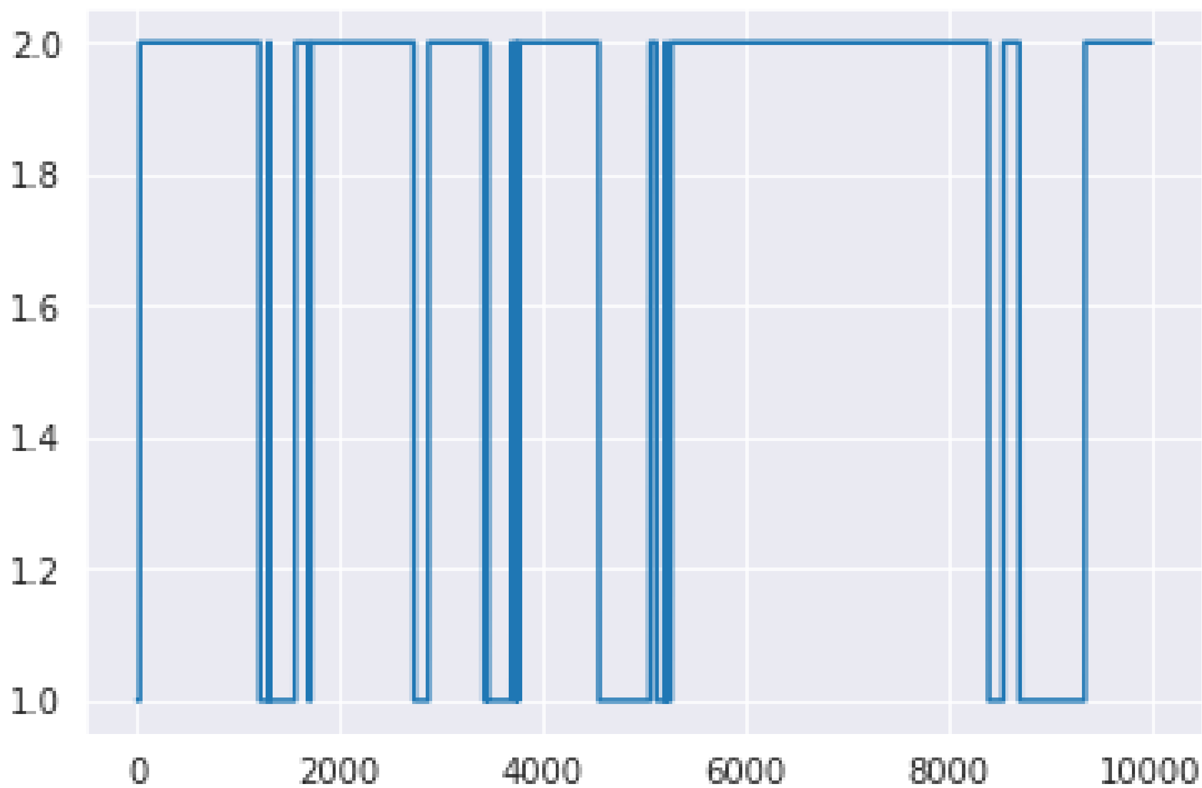


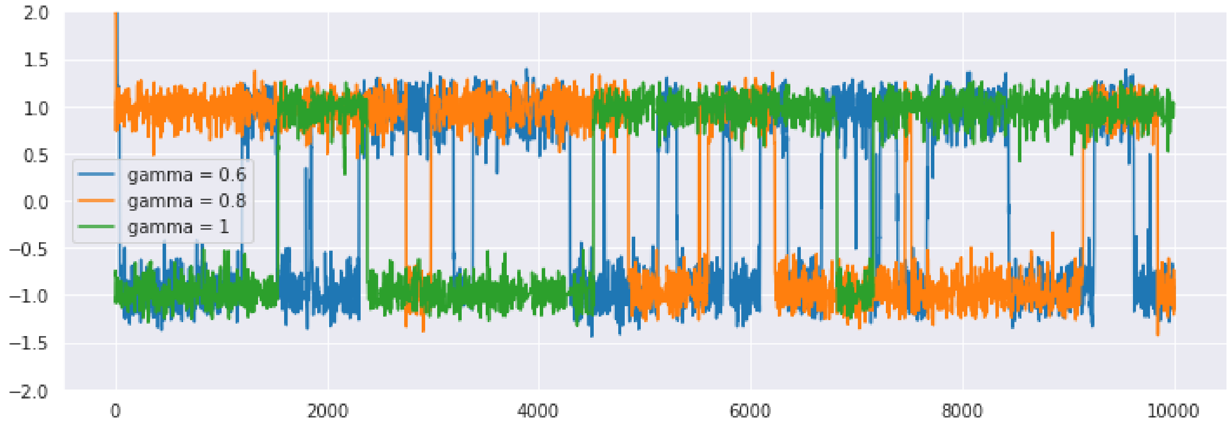
Figure 3: Segundo Modelo θ_2

Visitas feitas em cada modelo:

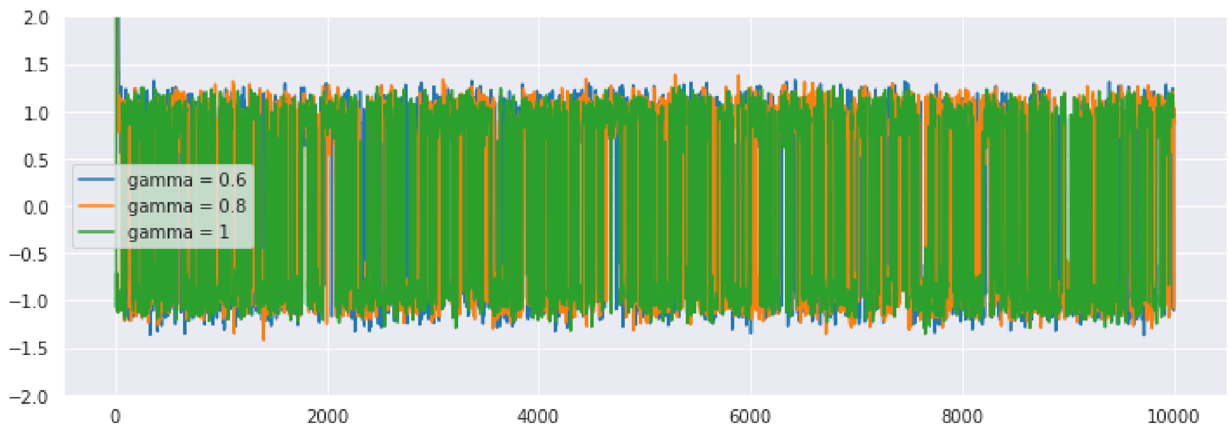


Simulation question (Parallel tempering)

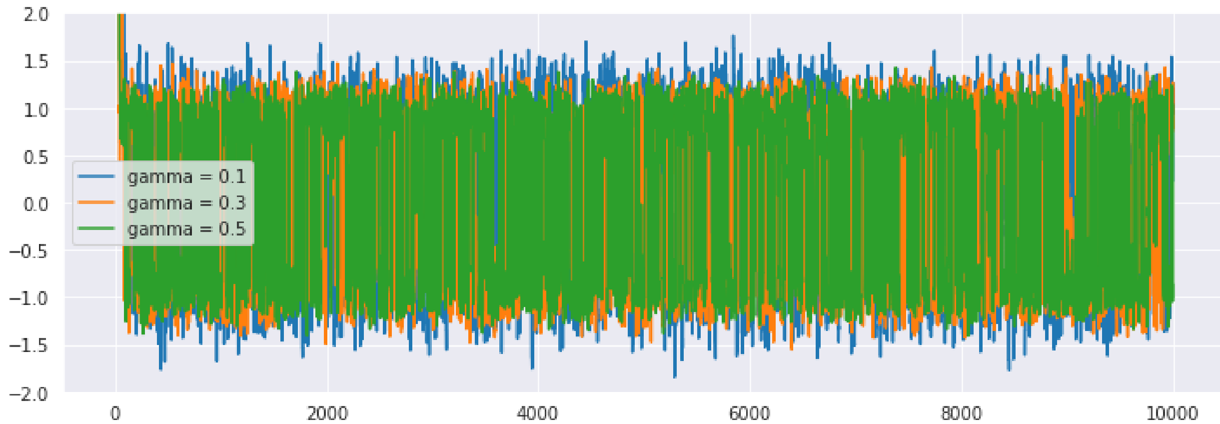
Após construir as funções descritas no enunciado, foi estabelecido uma quantidade de N=10 gammas que evoluem de 0 até 1 de maneira igual com incremento de 0.1. Logo cada distribuição π_k foi calculada com base nesses valores de gamma. O gráfico das cadeias para valores altos de gamma, quando são elaboradas simulações de forma independente, é o que segue:



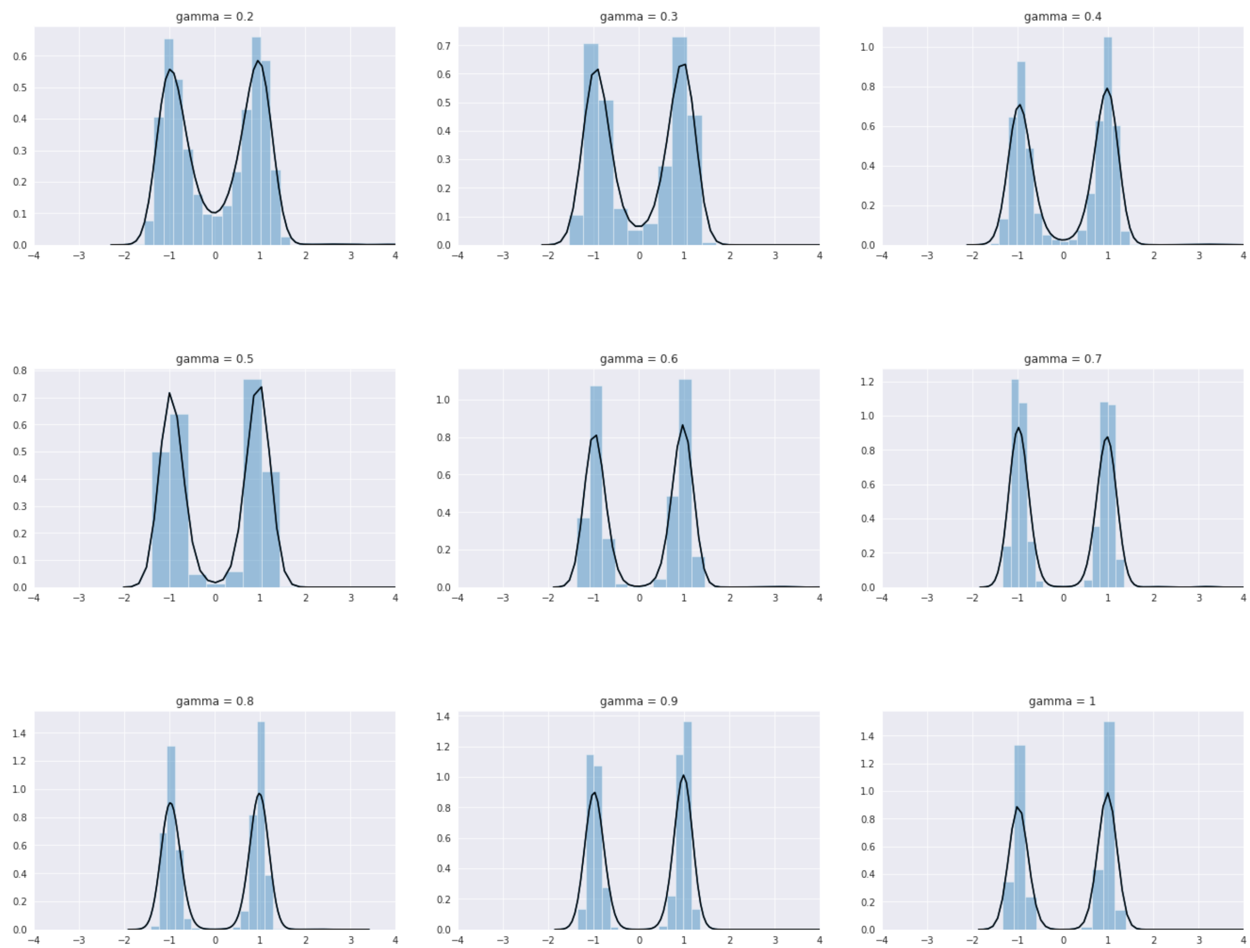
Agora fazendo o Parellel Tempering, as cadeias obtidas para os mesmos gammas do gráfico anterior podem ser vistos na figura abaixo:



Uma interessante conclusão obtida dessa comparação entre os dois gráficos é de que fazendo o Parallel Tempering as cadeias percorrem mais os diferentes "cores" do que no caso em que são utilizadas simulações MCMC independentes. Veja que para valores baixos de gamma, podemos obter o seguinte gráfico:



E para as distribuições veja que conforme o gamma vai aumentando progressivamente até chegar no seu máximo valor (no nosso caso 1) as modas vão se separando e ficando portanto "menos misturadas":



Simulation question (linear Gaussian model – SIS and SIR)

Como o modelo em questão é Gaussiano e linear, podemos então aplicar o filtro de Kalman de modo que poderemos a partir desse ponto estimar o valor esperado de x_t dados os valores de $y_1, y_2, ..., y_t$. O primeiro passo tal como foi descrito no item 1, foi justamente simular 100 observações deste modelo, o que foi feito por meio de uma distribuição normal. Depois, aplicamos o filtro de Kalman para obtermos um resultado que possa ser comparado com os métodos de SIS e SIR. Então foi escolhida a prior proposal e feita a normalização dos pesos, com isso o SIS foi utilizado apenas com o intuito de filtrar as médias, assim como o SIR. Então os resultado obtidos e comparados ao FIltro de Kalman para quando o t é fixado em 100 e o número N se situa em 100, 500 e 1000 é dado por:

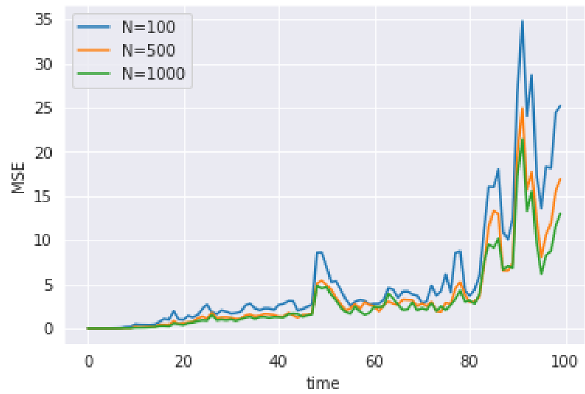


Figure 4: SIS

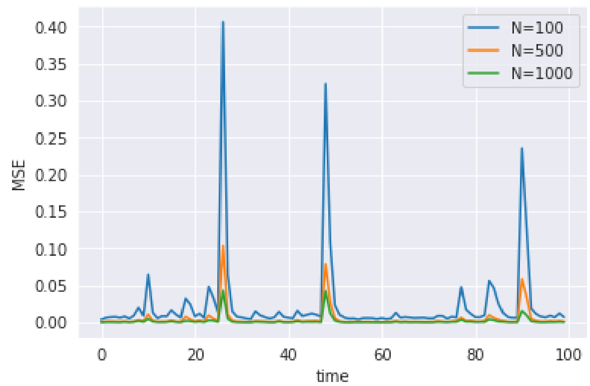


Figure 5: SIR

E as médias dos erros podem ser comparados nos gráficos abaixo, que mostram as médias obtidas pelo SIS e pelo SIR, de forma que fica claro como a introdução do SIR deixou a curva de médias mais fina, ou seja, com um erro absoluto em relação a média verdadeira menor:

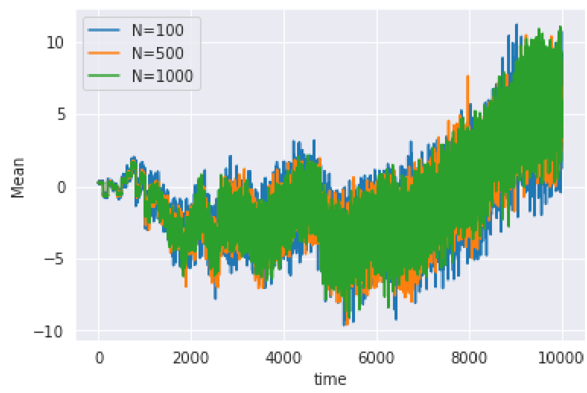


Figure 6: SIS

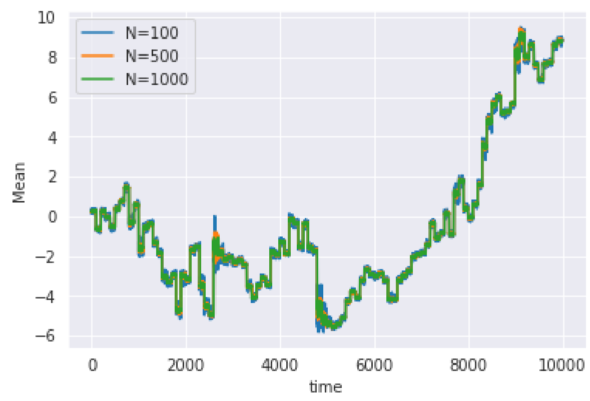


Figure 7: SIR