



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Programa de Pós-Graduação em Estatística

Métodos Sequencias de Monte Carlo Bayesianos: Aspectos
Computacionais, Inferenciais e Aplicações

Felipe Carvalho Álvares da Silva

Minas Gerais – Brasil
Julho de 2016

Métodos Sequencias de Monte Carlo Bayesianos: Aspectos
Computacionais, Inferenciais e Aplicações

Felipe Carvalho Álvares da Silva

***Orientador:* Prof. Dr. Luiz Henrique Duczmal**

***Coorientadora:* Profa. Dra. Denise Burgarelli Duczmal**

Tese apresentada ao Instituto de Ciências Exatas – ICEx-
UFMG, como parte dos requisitos para obtenção do título
de Doutor . *VERSÃO REVISADA.*

UFMG – Belo Horizonte

Julho de 2016

Ficha catalográfica elaborada pelo Departamento Técnico do Sistema Integrado de Bibliotecas da UFMG, com os dados fornecidos pelo(a) autor(a)

S634m	<p>Silva, Felipe Carvalho Álvares da Métodos Sequencias de Monte Carlo Bayesianos: Aspectos Computacionais, Inferenciais e Aplicações / Felipe Carvalho Álvares da Silva; orientador Luiz Henrique Duczmal; coorientadora Denise Burgarelli Duczmal. - Belo Horizonte - MG, 2016. 196 p.</p> <p>Tese (Doutorado - Programa de Pós-Graduação em Estatística) - Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, 2016.</p> <p>1. Modelos Dinâmicos. 2. Métodos de Monte Carlo Sequenciais. 3. Filtros de Partículas. 4. Filtragem Bayesiana. 5. Aprendizagem de parâmetros. I. Duczmal, Luiz Henrique, orient. II. Duczmal, Denise Burgarelli, coorient. III. Título.</p>
-------	--

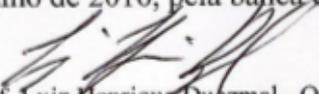
FOLHA DE APROVAÇÃO

"Métodos de Monte Carlo Sequenciais Bayesianos: Aspectos Computacionais, Inferenciais e Aplicações"

FELIPE CARVALHO ÁLVARES DA SILVA

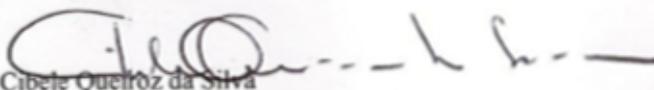
Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTATÍSTICA, como requisito para obtenção do grau de Doutor em ESTATÍSTICA, área de concentração ESTATÍSTICA E PROBABILIDADE.

Aprovada em 15 de julho de 2016, pela banca constituída pelos membros:


Prof. Luiz Henrique Duczmal - Orientador
DEST/UFMG


Prof. Denise Burgarelli Duczmal - Coorientadora
MAT/UFMG


Prof. Ricardo Hiroshi Caldeira Takahashi
MAT/UFMG


Prof. Cibele Queiroz da Silva
DEST/UnB


Prof. Saul de Castro Leite
DCC/UFJF


Prof. Sokol Ndreca
DEST/UFMG


Prof. Thiago Rezende dos Santos
DEST/UFMG

Belo Horizonte, 15 de julho de 2016.

*“Simplicity is a great virtue but it requires hard work to achieve it and education to appreciate it.
And to make matters worse: complexity sells better.”
(Edsger W. Dijkstra)*

RESUMO

Métodos Sequenciais de Monte Carlo, ou Filtros de Partículas, compreendem uma classe de algoritmos capazes de lidar com o problema da filtragem em modelos dinâmicos não-lineares e não-Gaussianos. Apesar de constituir um conjunto de procedimentos com uma vasta bagagem de propriedades já estabelecidas, tanto para a filtragem de estados latentes quanto para a estimação de parâmetros estáticos, muitas aplicações envolvendo tais algoritmos ainda apresentam resultados subótimos ainda despercebidos na literatura. É possível demonstrar que a degeneração de alguns métodos se deve a problemas na implementação ou na interpretação dos mesmos, e não a falhas nos algoritmos. Neste contexto, focando na estimação Bayesiana de parâmetros estáticos, buscamos evidenciar as verdadeiras fontes dos problemas de degeneração apontados na literatura e propomos o estabelecimento de diretrizes adequadas de implementação a partir da revisão e análise numérica dos principais algoritmos. Em especial, mostramos que, diferentemente dos resultados apresentados em diversas obras, mesmo problemas complexos podem ser tratados a partir de metodologias relativamente simples.

Palavras-chave: Modelos Dinâmicos, Métodos de Monte Carlo Sequenciais, Filtros de Partículas, Filtragem Bayesiana, Aprendizagem de parâmetros.

ABSTRACT

Sequential Monte Carlo methods, or Particle Filters, comprise a class of algorithms capable of dealing with the filtering problem in non-Gaussian and non-linear dynamic models. Although constituting a set of procedures with a large repertoire of properties already established, both for the filtering problem as for the estimation of static parameters, many applications involving such algorithms have sub-optimal results still unnoticed in the literature. It is possible to show that the degeneration of some procedures is overestimated due to interpretation or implementation issues, not to flaws in the algorithms. In this context, focusing on the Bayesian estimation of static parameters, we highlight the true sources of the well known degeneration phenomena pointed out by some works and we propose the establishment of appropriate implementation guidelines through the review and analysis of the main algorithms. In particular, we show that, unlike the results presented in several works, even complex problems can be treated by relatively simple methodologies.

Key-words: Dynamic Models, Sequential Monte Carlo methods, Particle Filters, Bayesian filtering, Parameter learning.

LISTA DE ILUSTRAÇÕES

Figura 1 – Representação gráfica da estrutura de dependência de uma cadeia Markoviana oculta	37
Figura 2 – Representação gráfica do esquema de reamostragem sistemática.	58
Figura 3 – Distribuições associadas a um único ponto considerando diferentes valores para o fator de desconto. Partindo do primeiro quadro e, percorrendo os demais em sentido horário, temos os comportamentos correspondentes às seguintes escolhas de δ : 0,99; 0,94; 0,89;0,84; 0,79; 0,74.	74
Figura 4 – Fatores de Fertilidade \times Tamanhos Efetivos Amostrais: comparação entre os mecanismos Multinomial, Residual, Sistemático e por Ramificação (escala percentual). O ensaio apresenta os fatores de fertilidade obtidos para 2500 medidas aleatórias independentes, sendo o suporte de cada uma composto por 10000 pontos.	86
Figura 5 – Distribuições <i>a posteriori</i> de θ obtidas a partir de 25 execuções do algoritmo PL considerando reamostragem multinomial e uma nuvem de 10000 partículas. 88	88
Figura 6 – Distribuições <i>a posteriori</i> de θ obtidas a partir de 25 execuções do algoritmo PL considerando reamostragem por ramificação e uma população de 5000 partículas.	89
Figura 7 – Resumo de 100 execuções, 5000 partículas cada, dos filtros LW convencional e LW com agenda dinâmica: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Colunas 1 e 2 ilustram, na devida ordem, os resultados obtidos pelos métodos LW e LW dinâmico. As curvas em preto correspondem ao resultado de referência: inferência obtida pelo algoritmo LW convencional a partir de 100000 pontos.	93
Figura 8 – Configuração espacial da população estática $\left\{ \beta^{(i)}, \tau^{2(i)} \right\}_{i=1}^N$ no Exemplo 2. A sequência de diagramas ilustra a evolução dos indivíduos após os 4 primeiros passos de reamostragem sem rejuvenescimento artificial. A nuvem de pontos é capaz de apresentar uma cobertura satisfatória do espaço mesmo na ausência de revitalização. O quinto passo de reamostragem é associado ao dispositivo de regularização (4.4) a fim de recuperar a diversidade.	95

Figura 9 – Resumo de 100 execuções, 5000 partículas cada, dos filtros PL convencional e PL com agenda dinâmica: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Colunas 1 e 2 ilustram, na devida ordem, os resultados obtidos pelos métodos PL e PL dinâmico. As curvas em preto correspondem ao resultado de referência: inferência obtida pelo algoritmo PL convencional a partir de 100000 pontos. Curvas em azul correspondem às quantidades reais utilizadas para a simulação do modelo.	97
Figura 10 – Resumo de 20 execuções, 2000 partículas cada, do filtro auxiliar dinâmico LW com e sem estratégia de reamostragem suavizada: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Colunas 1 e 2 ilustram, na devida ordem, os resultados obtidos sem suavização e com suavização. Curvas em azul correspondem aos valores reais utilizados para a simulação do modelo. . . .	101
Figura 11 – Distribuições <i>a priori</i> associadas aos parâmetros $(\alpha, \beta, \tau^2, \sigma^2)$	109
Figura 12 – Densidades estimadas via filtro de partículas – Modelo $AR(1) + \text{ruído}$ com 500 observações. As colunas da esquerda e da direita resumem, na devida ordem, os resultados obtidos de 25 execuções das estratégias PL e LW com 5000 partículas. As curvas em preto indicam a estimativa baseada em 100000 partículas de cada um dos dois métodos. Curvas em vermelho corresponde à estimativa construída via MCMC.	110
Figura 13 – Densidades estimadas via filtro de partículas – Modelo $AR(1) + \text{ruído}$ com 1000 observações. As colunas da esquerda e da direita resumem, na devida ordem, os resultados obtidos de 25 execuções das estratégias PL e LW com 5000 partículas. As curvas em preto indicam a estimativa baseada em 100000 partículas de cada um dos dois métodos. Curvas em vermelho corresponde à estimativa construída via MCMC.	111
Figura 14 – Distribuições <i>a priori</i> associadas aos parâmetros $(\beta, \tau^2, \sigma^2)$	112
Figura 15 – Tamanhos efetivos amostrais associados a 25 iterações do filtro LW otimizado considerado a perspectiva Auxiliar convencional.	113
Figura 16 – Tamanhos efetivos amostrais associados a 25 iterações do filtro LW otimizado considerado a perspectiva Auxiliar com distribuição proposta ótima.	113
Figura 17 – Densidades estimadas via filtro LW otimizado com número crescente de partículas. Do topo para a base estão representadas aproximações construídas com o auxílio de 2000, 15000 e 25000 partículas. Linhas verticas indicam os valores reais dos parâmetros utilizados para simulação.	114
Figura 18 – Distribuições <i>a priori</i> associadas aos parâmetros $(\beta, \tau^2, \sigma^2)$	115
Figura 19 – Exemplo de sequências de observações e estados latentes simulados de acordo com a especificação proposta por Storvik. A curva em azul expressa as observações (perturbadas por um ruído t) enquanto a trajetória em vermelho ilustra a sequência latente (ruídos Gaussianos).	116

Figura 20 – Resumo de 25 iterações do método LW otimizado para série de tamanho 150: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Implementações baseadas em enxames de 2000 partículas conforme proposto originalmente por Storvik. Trajetórias em azul representam quantidades reais adotadas na simulação da observações.	117
Figura 21 – Resumo de 25 iterações do método LW otimizado para série de tamanho 1000: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Implementações baseadas em enxames de 4000 partículas. Trajetórias em azul representam quantidades reais adotadas na simulação da observações.	117
Figura 22 – Sequências de observações e estados latentes, tamanho 1000, simuladas a partir dos parâmetros $(\beta, \tau^2, \sigma^2, \nu) = (0, 99; 1, 00; 1, 00; 5)$. A curva em azul expressa as observações (perturbadas por um ruído t) enquanto a trajetória em vermelho ilustra a sequência latente (ruídos Gaussianos).	118
Figura 23 – Resumo de 25 iterações do método LW otimizado para série de tamanho 500: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Implementações baseadas em enxames de 4000 partículas. Trajetórias em azul representam quantidades reais adotadas na simulação da observações.	118
Figura 24 – Filtragem da série de 150 observações com parâmetro ν desconhecido.	119
Figura 25 – Filtragem da série de 500 observações com parâmetro ν desconhecido.	119
Figura 26 – Filtragem da série de 1000 observações com parâmetro ν desconhecido.	120
Figura 27 – Distribuições <i>a priori</i> associadas aos parâmetros $(r_1, r_2, \tau^2, \sigma^2)$	124
Figura 28 – Exemplo de sequências de observações e estados latentes simulados de acordo com a especificação proposta por Prado e Lopes. A curva em azul expressa as observações enquanto a trajetória em vermelho ilustra o processo autoregressivo latente.	124
Figura 29 – Resumo de 15 iterações do método LW convencional com distribuição proposta ideal: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades reais adotadas na simulação da observações.	125
Figura 30 – Resumo de 15 iterações do método LW otimizado com distribuição proposta ideal: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades reais adotadas na simulação da observações.	125
Figura 31 – Resumo de 15 iterações do método LW otimizado com distribuição proposta ideal e distribuições <i>a priori</i> mais informativas: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades reais adotadas na simulação da observações.	126
Figura 32 – Distribuições <i>a priori</i> associadas aos parâmetros $(r_1, r_2, \lambda_2, \tau^2, \sigma^2)$	128

Figura 33 – Exemplo de sequências de observações e estados latentes simulados de acordo com a especificação proposta por Prado e Lopes. A curva em azul expressa as observações enquanto a trajetória em vermelho ilustra o processo autoregressivo latente.	128
Figura 34 – Resumo de 15 iterações do método LW otimizado: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades reais adotadas na simulação da observações.	129
Figura 35 – Distribuição das médias <i>a posteriori</i> construídas a partir de 150 execuções do filtro LW otimizado. Curvas em azul referem-se aos parâmetros reais tomados no processo de simulação.	129
Figura 36 – Resumo de 15 iterações do método LW otimizado considerando uma série de tamanho 1500: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades reais adotadas na simulação da observações.	130
Figura 37 – Distribuição das médias <i>a posteriori</i> construídas a partir de 150 execuções do filtro LW otimizado considerando uma série de 1500 observações e enxames de 8000 partículas. Curvas em azul referem-se aos parâmetros reais tomados no processo de simulação.	130
Figura 38 – Distribuições <i>a priori</i> correspondentes aos parâmetros μ , β e τ^2	135
Figura 39 – Séries de observações (azul) e volatilidades (vermelho) simuladas a partir do modelo (6.1).	136
Figura 40 – Resumo de desempenho do método de Liu e West otimizado (Σ_1) com 20000 partículas: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam os parâmetros reais tomados como referência para a geração das observações.	137
Figura 41 – Densidades <i>a posteriori</i> construídas pelo filtro de Liu e West otimizado com matriz de dependências Σ_1 (azul) e pelo método MCMC (vermelho).	137
Figura 42 – Resumo de desempenho do método de Liu e West otimizado (Σ_2) com 20000 partículas: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam os parâmetros reais tomados como referência para a geração das observações.	138
Figura 43 – Densidades <i>a posteriori</i> construídas pelo filtro de Liu e West otimizado com matriz de dependência Σ_2 (azul) e pelo método MCMC (vermelho).	138
Figura 44 – Taxas de câmbio Dólar Americano \times Euro (azul) e volatilidades (vermelho) estimadas via MCMC.	139

Figura 45 – Filtragem envolvendo taxas de câmbio Dólar Americano \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e matriz de covariância completa: médias a <i>posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam as quantidades estimadas via MCMC.	139
Figura 46 – Filtragem envolvendo taxas de câmbio Dólar Americano \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e matriz de covariância completa com decomposição em valores singulares: médias a <i>posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades estimadas via MCMC.	140
Figura 47 – Filtragem envolvendo taxas de câmbio Dólar Americano \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e blocos independentes de parâmetros: médias a <i>posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades estimadas via MCMC.	141
Figura 48 – Densidades estimadas via filtro LW otimizado estrutura de dependência bloco-independente. Cada curva corresponde a uma estimativa independente construída a partir de 50000 partículas.	141
Figura 49 – Comparação entre os paradigmas Auxiliar e <i>Bootstrap</i> . Os quadros representam o diagrama de dispersão envolvendo a estimativa média das quantidades $(\mu, \beta, \tau^2, X_t)$, considerando 50 execuções de cada método, ao longo dos 3139 instantes de iteração.	142
Figura 50 – Taxas de câmbio Rublo Russo \times Euro (azul) e volatilidades (vermelho) estimadas via MCMC.	143
Figura 51 – Filtragem envolvendo taxas de câmbio Rublo Russo \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e matriz de covariância completa: médias a <i>posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam as quantidades estimadas via MCMC.	143
Figura 52 – Filtragem envolvendo taxas de câmbio Rublo Russo \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e blocos independentes de parâmetros: médias a <i>posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades estimadas via MCMC.	144
Figura 53 – Densidades estimadas via filtro LW otimizado estrutura de dependência bloco-independente. Cada curva corresponde a uma estimativa independente construída a partir de 50000 partículas.	144

Figura 54 – Comparação entre os paradigmas Auxiliar e <i>Bootstrap</i> . Os quadros representam o diagrama de dispersão envolvendo a estimativa média das quantidades $(\mu, \beta, \tau^2, X_t)$, considerando 50 execuções de cada método, ao longo dos 3139 instantes de iteração.	145
Figura 55 – Discrepância observada entre as estimativas $\mathbb{E}[\beta y_{1:t}]$ construídas pelas perspectivas <i>Bootstrap</i> e Auxiliar.	146
Figura 56 – Taxas de câmbio Rupiah Indonésia \times Euro (azul) e volatilidades (vermelho) estimadas via MCMC.	146
Figura 57 – Filtragem envolvendo taxas de câmbio Rupiah Indonésia \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e matriz de covariância completa: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam as quantidades estimadas via MCMC.	147
Figura 58 – Filtragem envolvendo taxas de câmbio Rupiah Indonésia \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e blocos independentes de parâmetros: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades estimadas via MCMC.	147
Figura 59 – Densidades estimadas via filtro LW otimizado estrutura de dependência bloco-independente. Cada curva corresponde a uma estimativa independente construída a partir de 50000 partículas.	148
Figura 60 – Comparação entre os paradigmas Auxiliar e <i>Bootstrap</i> . Os quadros representam o diagrama de dispersão envolvendo a estimativa média das quantidades $(\mu, \beta, \tau^2, X_t)$, considerando 50 execuções de cada método, ao longo dos 3139 instantes de iteração.	148
Figura 61 – Taxas de câmbio Coroa Dinamarquesa \times Euro (azul) e volatilidades (vermelho) estimadas via MCMC.	149
Figura 62 – Filtragem envolvendo taxas de câmbio Coroa Dinamarquesa \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e matriz de covariância completa: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam as quantidades estimadas via MCMC.	150
Figura 63 – Filtragem envolvendo taxas de câmbio Coroa Dinamarquesa \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e blocos independentes de parâmetros: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades estimadas via MCMC.	150

Figura 64 – Erros cometidos pelo dispositivo de filtragem na estimação dos parâmetros estáticos (μ, β, τ^2) , ao longo de 50 execuções, considerando a utilização das estruturas de dependência Σ_1 e Σ_2	150
Figura 65 – Comparação entre os paradigmas Auxiliar e <i>Bootstrap</i> . Os quadros representam o diagrama de dispersão envolvendo a estimativa média das quantidades $(\mu, \beta, \tau^2, X_t)$, considerando 50 execuções de cada método, ao longo dos 3139 instantes de iteração.	151
Figura 66 – Discrepância observada entre as estimativas $\mathbb{E}[\beta y_{1:t}]$ e $\mathbb{E}[\tau^2 y_{1:t}]$ construídas pelas perspectivas <i>Bootstrap</i> e Auxiliar.	152
Figura 67 – Exemplo de evolução populacional obtida por simulação. As observações são dadas em escala real.	157
Figura 68 – Distribuições <i>a priori</i> correspondentes aos parâmetros $\mu_0, b_0, b_2, b_3, \tau^2$ e σ^2 na configuração artificial proposta por Chopin <i>et al.</i>	158
Figura 69 – Resumo de desempenho do método de Liu e West otimizado (perspectiva Auxiliar) com 50000 partículas: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam as quantidades reais utilizadas no processo de simulação.	158
Figura 70 – Distribuições <i>a posteriori</i> obtidas pelos métodos LW otimizado (azul) e SMC ² (vermelho).	159
Figura 71 – Distribuições <i>a posteriori</i> obtidas pelos métodos LW otimizado (azul) e SMC ² (vermelho).	159
Figura 72 – Distribuições <i>a posteriori</i> obtidas pelos métodos LW otimizado Auxiliar (azul) e SMC ² (vermelho).	160
Figura 73 – Diagramas de dispersão referentes aos parâmetros b_0, b_2 e b_3 obtidos a partir dos métodos SMC ² (topo), filtro LW otimizado (meio) e filtro LW sem agenda dinâmica (base).	161
Figura 74 – Resumo de desempenho do método de Liu e West otimizado (perspectiva <i>Bootstrap</i>) com 50000 partículas: médias <i>a posteriori</i> e bandas de 95% de credibilidade. Trajetórias em azul representam as quantidades reais utilizadas no processo de simulação.	162
Figura 75 – Distribuições <i>a posteriori</i> obtidas pelos métodos LW otimizado (azul) e SMC ² (vermelho).	162
Figura 76 – Distribuições <i>a posteriori</i> obtidas pelos métodos LW otimizado (azul) e SMC ² (vermelho).	163
Figura 77 – Comparação entre as estimativas construídas para cada parâmetro e os valores reais utilizados para simulação. Os diagramas ilustram a distribuição dos erros $\hat{\theta}_i - \theta$ considerando as perspectivas Auxiliar e <i>Bootstrap</i>	163
Figura 78 – Distribuições <i>a posteriori</i> obtidas pelos métodos LW otimizado <i>Bootstrap</i> (azul) e SMC ² (vermelho).	164

Figura 79 – Observações, em escala real, referentes à espécie <i>Accipter nisus</i> registrado sob o número 6575 no banco de dados global de populações.	165
Figura 80 – Distribuições <i>a priori</i> correspondentes aos parâmetros $\mu_0, b_0, b_2, b_3, \tau^2$ e σ^2 . 165	165
Figura 81 – Distribuições <i>a posteriori</i> obtidas pelos métodos LW otimizado Auxiliar (azul) e PMCMC (vermelho).	166
Figura 82 – Diagramas de dispersão amostral obtidos pelos métodos PMCMC (topo) e LW otimizado Auxiliar (base).	167
Figura 83 – Desempenho do método LW otimizado Auxiliar considerando 150000 partículas.	168
Figura 84 – Filtro LW sem a utilização de agenda dinâmica de rejuvenescimento. O diagrama no topo apresenta as densidades <i>a posteriori</i> estimadas pelo método em comparação com a estimativa construída via PMCMC. Abaixo temos os diagramas de dispersão obtidos após a filtragem da última observação. . . .	169
Figura 85 – Desempenho comparativo dos métodos LW otimizado, PMCMC e SMC ² . .	170
Figura 86 – Comparação entre TEAs e FFs obtidos através do método de seleção multinomial.	183
Figura 87 – Comparação entre TEAs e FFs. À esquerda segue a relação entre TEA e FF obtida através do método de seleção residual. À direita segue a comparação entre os FFs associados às abordagens residual e multinomial.	184
Figura 88 – Comparação entre TEAs e FFs obtidos através dos métodos de seleção sistemático e por ramificação.	184
Figura 89 – Desempenho comparativo entre métodos de reamostragem. Da esquerda para a direita: ramificação \times sistemática, ramificação \times residual e ramificação \times multinomial.	185
Figura 90 – Comparação corrigida entre os métodos LW, Storvik e PL. Percentis de $p(\theta y^f)$ (2,5%; 50%; 97,5%) baseados em 100 replicações de cada filtro com $N = 1000$ partículas (trajetórias cinzas). Curvas pretas são baseadas no filtro PL com $N = 100.000$ partículas. Filtro LW (coluna à esquerda), filtro de Storvik (coluna central) e PL (coluna à direita). As linhas (de cima para baixo) representam as componentes de $\theta = (\alpha, \beta, \tau^2, \sigma^2)$	191
Figura 91 – Comparação corrigida entre os filtros LW, Storvik e PL. Raíz do erro quadrático médio de $R = 100$ replicações de cada filtro. Todos os filtros são baseados em $N = 1000$ partículas e a raíz do erro quadrático médio é calculada com respeito a uma longa execução do filtro PL ($N = 100.000$).	192
Figura 92 – Comparação entre diferentes interpretações do filtro de Liu e West. A coluna da esquerda implementa a proposta sugerida por Lopes e Tsay enquanto a coluna da direita considera a sugestão deste trabalho. Em ambos os casos foram consideradas 20 execuções independentes com 5000 partículas cada. .	193

LISTA DE ALGORITMOS

Algoritmo 1 – Amostragem por Rejeição	47
Algoritmo 2 – Metropolis-Hastings	52
Algoritmo 3 – Amostrador de Gibbs	53
Algoritmo 4 – Amostragem Sequencial por Importância	55
Algoritmo 5 – Reamostragem Multinomial	57
Algoritmo 6 – Reamostragem Residual	58
Algoritmo 7 – Reamostragem Sistemática	59
Algoritmo 8 – Reamostragem por Ramificação	60
Algoritmo 9 – Amostragem Sequencial por Importância com Reamostragem	61
Algoritmo 10 – Amostra Índices	62
Algoritmo 11 – Filtro <i>Bootstrap</i>	64
Algoritmo 12 – Filtro <i>Bootstrap</i> com Reamostragem Adaptativa	65
Algoritmo 13 – Filtro Auxiliar	66
Algoritmo 14 – Filtro de Kitagawa	72
Algoritmo 15 – Filtro de Liu e West	73
Algoritmo 16 – Filtro de Storvik	75
Algoritmo 17 – Filtro PL	78
Algoritmo 18 – SMC2	80
Algoritmo 19 – PMCMC	81
Algoritmo 20 – Aprendizagem de Parâmetros <i>Bootstrap</i> com Agenda Dinâmica	91
Algoritmo 21 – Aprendizagem de Parâmetros Auxiliar com Agenda Dinâmica	96

LISTA DE SÍMBOLOS

\mathbb{R}^n — espaço Euclidiano n -dimensional

$\mathcal{B}(\Omega)$ — σ -álgebra dos Borelianos em Ω

$\sigma(X_s, 0 \leq s \leq t)$ — σ -álgebra gerada por X

$\mathcal{F}_b(\Omega)$ — funções mensuráveis limitadas definidas em Ω

$a_{k:l}$ — coleção de variáveis consecutivas da forma $(a_k, a_{k+1}, a_{k+2}, \dots, a_l)$.

$a_{k:k}$ — variável isolada a_k

(f, ν) — $(f, \nu) = \mathbb{E}_\nu[f(X)] = \int_\Omega f(x)\nu(dx)$

$\delta_x(\cdot)$ — função generalizada delta de Dirac

$\nu \ll \pi$ — ν é absolutamente contínua com respeito a π : $\pi(A) = 0 \Rightarrow \nu(A) = 0$

$\frac{d\nu}{d\pi}$ — derivada de Radon-Nikodym de ν com respeito a π .

SUMÁRIO

1	INTRODUÇÃO	25
1.1	Motivação	25
1.2	Revisão bibliográfica	27
1.3	Objetivos	30
1.4	Organização do Trabalho	31
2	CONCEITOS PRELIMINARES	33
2.1	Inferência Bayesiana	33
2.2	Cadeias de Markov	35
2.3	Cadeias Markovianas Ocultas	37
2.4	O Problema da Filtragem	39
2.5	Suavização	40
2.6	Parâmetros Estáticos	41
3	MÉTODOS DE MONTE CARLO	45
3.1	Amostragem por Rejeição	47
3.2	Amostragem por Importância	48
3.3	Monte Carlo via Cadeias de Markov	51
3.4	Amostragem Sequencial por Importância	53
3.5	Filtros de Partículas	62
3.5.1	<i>Filtragem Pura</i>	62
3.5.2	<i>Suavização</i>	68
3.5.3	<i>Aprendizagem de Parâmetros</i>	70
4	APRENDIZAGEM DE PARÂMETROS OTIMIZADA	83
4.1	Aprendizagem de Parâmetros Revisitada	83
4.2	Fator de Fertilidade e Reamostrador Ideal	85
4.3	Agenda Ótima de Rejuvenescimento	90
4.4	Reamostragem Suavizada	98
4.5	Janela Ótima \times Fator de Encolhimento	102
4.6	Estruturas de Dependência	104
5	ESTUDOS DE SIMULAÇÕES	107
5.1	AR(1) + Ruído	108

5.2	AR(1) + Ruído Não-Gaussiano	115
5.3	AR(p) + Ruído	121
6	APLICAÇÕES A DADOS REAIS	133
6.1	Volatilidade Estocástica	133
6.1.1	<i>Descrição do Modelo</i>	133
6.1.2	<i>Análise Experimental: Taxas de Câmbio</i>	134
6.2	Dinâmica Populacional	155
6.2.1	<i>Descrição do Modelo</i>	155
6.2.2	<i>Avaliação Experimental: Modelo θ-Estocástico</i>	156
7	CONSIDERAÇÕES FINAIS	171
7.1	Objetivos Alcançados	171
7.2	Conclusões	172
7.3	Trabalhos Futuros	173
	Referências	175
APÊNDICE A	<i>FATOR DE FERTILIDADE \times TAMANHO EFETIVO AMOSTRAL</i>	181
APÊNDICE B	<i>DETERMINAÇÃO DO COEFICIENTE DE SUAVIZAÇÃO λ</i>	187
APÊNDICE C	<i>COMENTÁRIOS SOBRE “PARTICLE FILTERS AND BAYESIAN INFERENCE IN FINANCIAL ECONOMETRICS” DE LOPES, H. E TSAY R.</i>	189
ANEXO A	<i>PACOTES COMPUTACIONAIS E PÁGINAS INTERESSANTES NA INTERNET</i>	195

INTRODUÇÃO

1.1 Motivação

A estimação dos estados (latentes) de um sistema que evolui ao longo do tempo gerando sequências ruidosas de medições, uma técnica formalmente denotada por *filtragem*, tem sido utilizada como alternativa de solução em um vasto conjunto de aplicações, dentre as quais podemos citar: reconhecimento de fala (Juang e Rabiner (1991), Gales e Young (2008)), processamento de imagens (Zhao e Li (2010), Han *et al.* (2011)), localização/rastreamento (Arulampalam *et al.* (2004), Nemeth, Fearnhead e Mihaylova (2014)), robótica (Thrun, Burgard e Fox (2005), Kwok e Rad (2006)), econometria (Genon-Catalot, Jeantheau e Laredo (2000), Lopes e Tsay (2011)), epidemiologia (Vidal-Rodeiro e Lawson (2006), Dukic, Lopes e Polson (2012)), etc. Problemas desta natureza são usualmente modelados de maneira eficaz por séries temporais provenientes de Modelos Markovianos Ocultos (*Hidden Markov Models* (HMMs)), uma classe de processos estocásticos fundamentada na observação indireta Y de uma sequência Markoviana latente X . A implementação de estruturas estocásticas como a de um HMM basicamente reduz o problema original à especificação de uma distribuição *a posteriori*¹ $P_{\theta}(X_{1:n}|Y_{1:n})$ referente à dinâmica de evolução dos estados ocultos condicionada às n medições disponíveis até o momento mais atual. Esta abordagem assegura a concepção de modelos de alta capacidade descritiva, capazes de se adequar a cenários bastante complexos.

Outro grande desafio relacionado ao problema da filtragem diz respeito à estimação dos parâmetros estáticos θ do modelo. A breve descrição acima assume que o modelo já é completamente especificado, focando seus esforços na estimação da sequência de estados latentes. A dificuldade encontrada na realização de inferência paramétrica neste cenário é fruto da impossibilidade de lidarmos com uma forma fechada para a verossimilhança $P_{\theta}(Y_{1:n})$ destes modelos na grande maioria dos casos.

¹ Vale ressaltar aqui que o termo "*a posteriori*" é utilizado neste ponto como um mero vício de linguagem, já que a especificação em si do problema não é atrelada à metodologia Bayesiana.

Durante vários anos a teoria da filtragem foi dominada por aplicações fundamentadas no filtro de Kalman (Kalman (1960), Kalman e Bucy (1961)) e suas variantes, as quais assumem que o problema pode ser descrito (ao menos aproximadamente) por processos Gaussianos com interações lineares tanto nas observações quanto na evolução dos estados latentes. A partir destes fundamentais é possível obter formas exatas tratáveis tanto para $P_\theta(X_{1:n}|Y_{1:n})$ quanto para $P_\theta(Y_{1:n})$, o que viabiliza o tratamento analítico da estimação de estados latentes e parâmetros estáticos. Apesar de construir soluções eficientes e sem a presença de erros aproximação, o filtro de Kalman trabalha com imposições muito fortes. A hipótese de um regime Gaussiano apresenta bons resultados em um número considerável de problemas, entretanto, na maioria dos casos, a hipótese de linearidade do modelo é muito grosseira devido à complexa interação entre variáveis não observáveis e medições. Na prática, não devemos então esperar uma formulação analiticamente tratável para a distribuição alvo do problema da filtragem.

Com o objetivo de simultaneamente usufruir da flexibilidade dos HMMs e superar a questão da intratabilidade, uma série de métodos computacionais baseados em simulações de Monte Carlo foram introduzidos ao longo das últimas décadas. Neste trabalho estaremos particularmente interessados em uma coleção específica de algoritmos denominados métodos Sequenciais de Monte Carlo (*Sequential Monte Carlo* (SMC)), também conhecidos como Filtros de Partículas (FPs), focando sua aplicação tanto na estimação dos estados latentes quanto na dos possíveis parâmetros estáticos do sistema. Algoritmos do tipo filtro de partículas têm sido utilizados ao longo das últimas décadas como um ferramental bastante robusto no tratamento do problema da filtragem envolvendo sistemas não-lineares e não Gaussianos. Sua grande contribuição em comparação às abordagens tradicionais apoiadas em métodos de Monte Carlo via Cadeias de Markov (*Markov Chain Monte Carlo* (MCMC)) é a eficiência computacional, todavia suas estimativas são mais imprecisas. Na espinha dorsal do método encontra-se uma versão sequencial do esquema de amostragem por importância, a qual propaga uma nuvem de variáveis aleatórias (partículas) recursivamente através da regra de Bayes. Diferentemente de outras técnicas mais tradicionais como o filtro de Kalman Estendido (Anderson e Moore (1979)), o “*Unscented Kalman filter*” (Julier e Uhlmann (1997)), ou métodos baseados em malhas de discretização (Bergman (1999)), a implementação dos filtros de partículas não faz uso de técnicas de linearização local, hipóteses rigorosas a respeito da representação funcional do modelo ou onerosos esquemas de discretização espacial, conferindo mais flexibilidade ao método. Além disso, a estrutura dos filtros de partículas pode ser explorada a fim de permitir a utilização de técnicas baseadas em MCMC ou maximização de verossimilhança para o problema de inferência envolvendo parâmetros estáticos, ou pode até mesmo ser modificada de modo a acomodar o problema de estimação paramétrica dentro do próprio esquema de amostragem por importância; em todos os casos os procedimentos são não triviais e apresentam desafios específicos.

1.2 Revisão bibliográfica

Variações do esquema de amostragem por importância, implementadas sob uma perspectiva sequencial, são conhecidas desde a década de 70 (Handschin (1970), Akashi e Kumamoto (1977)), porém não motivaram um número expressivo de aplicações por apresentar um comportamento instável. O problema da instabilidade, popularmente denotado por *degeneração amostral*, ocorre devido ao desequilíbrio gerado nos pesos de importância de cada estado ao longo de várias iterações; à medida que o algoritmo evolui, o número de partículas com peso insignificante é cada vez maior, aumentando naturalmente a variabilidade das estimativas. Visando contornar o problema da degeneração amostral, Gordon, Salmond e Smith (1993) introduziram um passo intermediário de reamostragem (*Filtro Bootstrap*), o qual seria responsável por eliminar partículas associadas a regiões de probabilidade desprezível. O novo método conseguiu de fato controlar melhor a variabilidade em torno das estimativas, fazendo com que os filtros de partículas se consolidassem definitivamente como uma ferramenta eficaz para o tratamento do problema da filtragem. A questão da reamostragem é um ponto chave na implementação dos filtros de partículas, exercendo influência significativa principalmente quando o modelo é influenciado pela presença de parâmetros estáticos desconhecidos. Desta forma, dedicaremos um pouco mais de atenção à revisão dos principais métodos de reamostragem no próximo parágrafo.

A primeira proposta de reamostragem baseou-se na ideia de que a nuvem de partículas pode ser considerada um vetor com distribuição multinomial cujas probabilidades são especificadas por seus respectivos pesos de importância. Apesar do relativo sucesso da metodologia, esta ideia de reamostragem trouxe uma nova dificuldade para a aplicação do filtro: a amostragem multinomial concentra a população de partículas em um conjunto restrito de pontos, diminuindo a diversidade e caracterizando um cenário de *empobrecimento amostral*. É fácil perceber que a excessiva propagação de réplicas restringe a exploração do espaço de estados do modelo a cada iteração, induzindo o método a gerar aproximações menos refinadas. Outras propostas tradicionais de reamostragem também fazem uso de uma determinada distribuição para o vetor de partículas e têm como premissa básica garantir que o número de réplicas de cada ponto seja proporcional a seu peso de importância. Nesta linha temos a *amostragem estratificada* proposta por Kitagawa (1996), a qual divide o intervalo $(0, 1]$ em N estratos (categorias), igualmente espaçados, visando a geração de N variáveis uniformes auxiliares; a amostra final é obtida a partir da comparação das variáveis auxiliares com uma função de soma cumulativa dos pesos. De maneira similar, o algoritmo de *amostragem sistemática* apresentado por Carpenter, Clifford e Fearnhead (1999) também explora a ideia de estratificação mas de uma maneira distinta. Nesta segunda formulação as categorias são definidas a partir de uma variável uniformemente distribuída em $(0, 1]$ gerada *a priori*, as $N - 1$ variáveis auxiliares restantes são definidas deterministicamente a partir da primeira e o algoritmo segue em um mecanismo similar à abordagem estratificada. Liu e Chen (1998) propuseram um procedimento de reamostragem em dois estágios: cada partícula é replicada deterministicamente de acordo com seu respectivo peso, em seguida a amostra é

completada a partir de um sorteio multinomial. Os pesos empregados no passo multinomial são expressos pela diferença (resíduo) entre o peso original de cada partícula e sua proporção de réplicas, com respeito à população total N , caracterizando assim um esquema de *amostragem residual*. [Fearnhead e Clifford \(2003\)](#) apresentaram um dispositivo análogo onde um conjunto inicial de partículas é preservado seguindo um critério de minimização dos erros de aproximação do algoritmo. Os demais pontos são obtidos da realização de um passo de amostragem sistemática. [Crisan, Del Moral e Lyons \(1999\)](#) construíram um procedimento inspirado em processos de ramificação, de modo que o número de replicações de cada partícula pudesse ser obtido de uma distribuição independente. Neste caso, o tamanho final da amostra pode ser variável ou não dependendo da escolha das distribuições. Outra linha de métodos de reamostragem presente na literatura parte da criação de novos valores (proles) na proximidade das melhores partículas ao invés da simples replicação das mesmas. A versão mais básica de métodos SMC, apresentada por [Gordon, Salmond e Smith \(1993\)](#), sugere o acréscimo de pequenas perturbações (“*roughening*”) às réplicas originais. [Li, Sattar e Sun \(2012\)](#) e [Fu e Jia \(2010\)](#) propõem técnicas determinísticas para o espalhamento dos pontos. Já [Higuchi \(1996\)](#), [Kwok, Gu e Zhou \(2005\)](#) e [Park et al. \(2009\)](#) argumentam que novas partículas podem ser geradas a partir da aplicação de operadores herdados dos algoritmos genéticos.

Alguns autores mostram que uma maior eficácia da metodologia poderia ser adquirida a partir da reestruturação de outros pontos além da reamostragem. Uma das propostas mais significativa na literatura, introduzida por [Pitt e Shephard \(1999\)](#), inverte a ordem dos passos de amostragem e reamostragem. O procedimento é realizado a partir da inserção de uma variável auxiliar, sugerindo a famosa conotação *filtro auxiliar*, de modo que a informação disponibilizada pela observação mais recente possa ser levada em consideração no processo de reamostragem. Outra sugestão bastante robusta é a utilização de estratégias do tipo *Rao-Blackwellization*. [Liu e Chen \(1998\)](#) mostram que em diversas situações o problema pode ser analisado por instâncias: uma parte menor do problema é tratado numericamente enquanto o restante problema, condicionado à primeira parte, é desenvolvido de maneira analítica. De uma maneira geral, a *Rao-Blackwellização* do algoritmo é capaz de reduzir a variabilidade final das estimativas, no entanto a estrutura necessária para seu desenvolvimento restringe o campo de aplicações. [Musso, Oudjane e Le Gland \(2001\)](#) argumentam que a degeneração amostral pode ser combatida através de um procedimento de regularização o qual consiste em reamostrar a nuvem de pontos a partir de uma distribuição aproximada por núcleos estimadores ao invés de realizar tal procedimento exclusivamente a partir dos pesos de importância. A estratégia de regularização apresenta resultados bastante satisfatórios e possui propriedades analíticas importantes já demonstradas, conforme podemos verificar nas obras de [Le Gland e Oudjane \(2004\)](#), [Künsch \(2005\)](#), [Crisan e Miguez \(2014\)](#). Quanto ao desenvolvimento de melhores distribuições propostas para o passo de amostragem, [Doucet, Godsill e Andrieu \(2000\)](#) sugerem o uso de distribuições aproximadas por um esquema de linearização do modelo, um dispositivo similar ao utilizado no filtro de Kalman estendido.

Aplicações envolvendo a estimação de parâmetros estáticos no contexto da filtragem eram restritas a problemas que possibilitavam a obtenção (ao menos aproximada) da função de verossimilhança das observações. Com o intuito de contornar tal restrição, Kitagawa (1998) sugeriu, em uma abordagem pioneira, inserir estes parâmetros na componente não observada do modelo, dando início ao conceito de *estados aumentados*. Vale ressaltar que os parâmetros desconhecidos do modelo não possui dinâmica natural, logo a sucessiva utilização de passos de reamostragem degenera esta nuvem de pontos ainda mais rapidamente. No mesmo trabalho, Kitagawa introduziu a ideia de evolução artificial inspirada em passeios aleatórios, a fim de acrescentar diversidade às partículas originalmente estáticas. Liu e West (2001) propuseram uma nova dinâmica de evolução artificial que preserva os dois primeiros momentos empíricos da população, uma propriedade alcançada através da aproximação de distribuições *a posteriori* por meio de misturas de funções Gaussianas. Apesar de figurar entre os métodos pioneiros, o filtro de Liu e West ainda é uma das abordagens de referência hoje devido à sua eficiência e generalidade nas hipóteses de construção do algoritmo. Uma segunda linha de algoritmos, trabalhada por Gilks e Berzuini (2001), Storvik (2002), Fearnhead (2002) e Carvalho *et al.* (2010), surgiu em seguida na tentativa de dar diversidade à população de parâmetros através da amostragem direta das distribuições *a posteriori* dos parâmetros estáticos. Nestes casos, as metodologias são condicionadas à existência de estatísticas suficientes capazes de ser atualizadas recursivamente. De uma maneira em geral estes trabalhos indicam que as abordagens dependentes da propagação de estatísticas suficientes apresentam resultados mais efetivos se comparados com o filtro de Liu e West, mas este apresenta uma estrutura mais flexível, podendo ser utilizado em uma maior variedade de contextos. Seguindo uma linha mais tradicional, uma proposta de grande impacto na literatura, introduzida por Andrieu, Doucet e Holenstein (2010), indica que os filtros de partículas podem ser utilizados como um passo intermediário de algoritmos MCMC para modelos de verossimilhança intratável. Em linhas gerais, filtros de partículas são utilizados para estabelecer estimativas não viciadas da função de verossimilhança, a qual pode ser utilizada para a realização de passos de amostragem por rejeição (e.g. Metropolis-Hastings). Na mesma linha, Singh (2015) estendem a metodologia de modo a considerar variações do tipo amostrador de Gibbs. Recentemente, Chopin, Jacob e Papaspiliopoulos (2013) apresentaram uma nova metodologia fundada no uso de passos MCMC associados aos filtros de partículas. Nesta aproximação, para cada indivíduo da população dos parâmetros estáticos, é lançado um filtro independente para a estimação dos estados latentes. O objetivo do método é utilizar as nuvens independentes para obter estimativas da função de verossimilhança para cada partícula $\theta^{(j)}$, eliminando assim a necessidade de propagação de estatísticas suficientes. A metodologia é tão abrangente quanto a proposta por Liu e West, e apresenta resultados ainda melhores. Entretanto, o método demanda um esforço computacional muito grande, podendo ser tão caro quanto versões mais convencionais do algoritmo MCMC aplicadas ao mesmo problema. No âmbito frequentista, Malik e Pitt (2011) e Poyiadjis, Doucet e Singh (2011) desenvolveram implementações capazes de explorar as partículas do sistema para construir uma aproximação consistente da função de

verossimilhança e realizar os procedimentos clássicos de inferência.

Por fim, destacamos que a flexibilidade oferecida pelos modelos de filtro de partículas habilita sua aplicação em cenários além dos HMMs. Este tipo de abordagem extrapola os objetivos deste trabalho, porém avaliamos que não podem deixar de ser citadas em qualquer trabalho de revisão acerca de métodos sequenciais de Monte Carlo. [Chopin \(2002\)](#) e [Fearnhead e Taylor \(2013\)](#) trabalham com um algoritmo que incorpora cada nova observação (proveniente de uma amostra de variáveis iid) sequencialmente e atualiza a incerteza a respeito dos parâmetros a partir das técnicas sequencias de Monte Carlo. A falta de dinâmica dos parâmetros é tratada neste caso a partir de um passo Metropolis-Hastings. [Lopes, Polson e Carvalho \(2012\)](#) desenvolvem um algoritmo similar porém explorando a ideia de propagação de estatísticas suficientes. Basicamente, o passo Metropolis no algoritmo de Chopin é substituído pela propagação de um conjunto auxiliar de estatísticas suficientes que se comporta como processo de estados latentes nos algoritmos de filtragem convencionais. Uma discussão mais abrangente a respeito de amostradores sequenciais em um contexto geral pode ser encontrada em [Del Moral, Doucet e Jasra \(2006\)](#).

1.3 Objetivos

Motivado pelo tradicional filtro de Liu e West ([LIU; WEST, 2001](#)), o objetivo da pesquisa era (a princípio) desenvolver novos métodos sequenciais eficientes, baseados em algoritmos estimadores de distribuição ([LARRANAGA; LOZANO, 2002](#)), capazes de incorporar ao problema da filtragem a incerteza a respeito de possíveis parâmetros estáticos sem a geração de novas exigências com relação às formas funcionais dos modelos em questão. A busca por um algoritmo inovador seria relevante uma vez que importantes registros na literatura ([STORVIK, 2002](#); [CARVALHO *et al.*, 2010](#); [LOPES; TSAY, 2011](#); [PRADO; LOPES, 2013](#); [KANTAS *et al.*, 2015](#)) asseguravam que apesar de o método de Liu e West ser capaz de gerar resultados satisfatórios em alguns casos, seu desempenho apresentaria um severo comportamento degenerativo em diversas situações, incluindo problemas relativamente simples. Todavia, uma análise mais criteriosa (verificação de algoritmos e códigos) dos trabalhos já consolidados indicaram que o fenômeno de degeneração do filtro de Liu e West era basicamente fruto de implementações subótimas e/ou até mesmo equivocadas. Resultados similares foram observados também com respeito a outros algoritmos, como os baseados na propagação de estatísticas suficientes, por exemplo. Desta forma, dedicamos a maior parte do trabalho aos aspectos computacionais do tópico, evidenciando o quanto os resultados apresentados por estes algoritmos podem ser eficazes desde que implementados sob certos cuidados.

Dadas as considerações elucidadas acima, o primeiro grande objetivo deste trabalho é evidenciar o efeito negativo que passos inadequados de reamostragem (especialmente o método multinomial) podem induzir nos algoritmos para a estimação de parâmetros estáticos. Tal efeito é ignorado de forma recorrente em diversos trabalhos, induzindo autores a atribuir problemas

de degeneração a outros fatores. Ainda nesta linha, buscamos desenvolver novas metodologias para amenizar o efeito de degeneração observado independentemente do reamostrador e da metodologia de amostragem de parâmetros estáticos. Desta forma, a primeira análise culminaria com um conjunto de diretrizes otimizadas de implementação capazes de corrigir problemas significativos de degeneração já identificados na literatura.

O segundo grande propósito é diretamente apontado para o método de Liu e West. Buscamos inicialmente estabelecer um critério prático para a calibragem dos parâmetros do algoritmo e mostrar como tal critério se relaciona com as escolhas tradicionais baseadas no conceito de fator de desconto. Em seguida, procuramos demonstrar experimentalmente que vários exemplos de degeneração apontados na literatura com respeito a este método são frutos de implementações subótimas e não de problemas referentes ao método. Por fim, discutiremos aplicações a dados reais de modo a ilustrar a verdadeira eficácia do algoritmo. A ideia em torno deste segundo objetivo não é identificar o método de Liu e West como a melhor opção possível e nem menosprezar implementações apresentadas por outros autores, mas sim indicar como uma simples ferramenta, tratada muitas vezes como um mero "quebra-galho", pode ser uma alternativa muito relevante a métodos computacionalmente mais onerosos.

1.4 Organização do Trabalho

A sequência da tese é composta por outros 6 capítulos. No capítulo 2 abordaremos brevemente o paradigma Bayesiano de inferência, apresentaremos conceitos cruciais para o bom entendimento do tópico tais como o de cadeia Markoviana Oculta e definiremos formalmente o problema da filtragem assim como seus principais desafios. No capítulo 3 discutiremos com mais detalhes o ferramental numérico englobado pela classe dos métodos sequenciais de Monte Carlo. Discutiremos os conceitos básicos a respeito de MCMC, amostragem por importância e a construção geral dos filtros de partículas. Abordaremos ainda de uma forma mais aprofundada as metodologias citadas anteriormente para o tratamento de parâmetros estáticos.

O capítulo 4 trata das principais contribuições deste trabalho: desenvolvimento de diretrizes de implementação e avaliação dos métodos sob as novas perspectivas. Para a avaliação numérica, será considerada uma extensa lista de experimentos simulados considerando os modelos mais recorrentes na literatura. O capítulo 5 fica a cargo da apresentação e da discussão destes resultados experimentais. O capítulo 6 é reservado a uma segunda avaliação experimental dos métodos, porém considerando dados reais. Este último conjunto de testes visa confirmar a eficácia das propostas em situações complexas que tiram os métodos da possível zona de conforto criada por avaliações a partir de dados simulados. O último capítulo trata das considerações finais da tese e projeta o futuro do trabalho.

CONCEITOS PRELIMINARES

2.1 Inferência Bayesiana

O campo da estatística Bayesiana (??) engloba um ferramental bastante robusto para a realização de inferências a respeito de fenômenos observados na natureza e em nosso cotidiano. Diferentemente do paradigma frequentista, a abordagem Bayesiana entende que a incerteza a respeito dos parâmetros de um modelo pode ser expressa a partir de distribuições de probabilidade, permitindo que estas entidades possam ser estudadas de maneira similar a uma variável aleatória qualquer. Esta incerteza é representada inicialmente por uma distribuição *a priori* a qual cumpre o papel de conhecimento prévio em torno dos alvos do processo inferencial. A informação *a priori* não precisa ser necessariamente precisa (informativa) ou restritiva, uma vez que seu real objetivo é oferecer um grau de liberdade adicional à modelagem do fenômeno em questão. Com o auxílio da regra de Bayes esta distribuição é combinada com a informação fornecida pelas observações através da função de verossimilhança, culminando assim na distribuição *a posteriori*

$$\underbrace{P(\theta|y_{1:n})}_{\text{posteriori}} \propto \underbrace{P(y_{1:n}|\theta)}_{\text{verossimilhança}} \cdot \underbrace{\pi(\theta)}_{\text{priori}}. \quad (2.1)$$

A relação (2.1) deve ainda ser complementada pela constante de normalização

$$P(y_{1:n}) = \mathbb{E}_{\pi}[P(y_{1:n}, \theta)], \quad (2.2)$$

a qual nem sempre pode ser tratada analiticamente devido à complexa forma funcional das densidades envolvidas. Este problema pode ser agravado caso os domínios de integração estejam ligados a espaços de dimensão elevada, o que inviabiliza a utilização de métodos clássicos de quadratura. Desta forma, há uma grande motivação para a busca de metodologias numéricas mais robustas de integração como as aproximações de Monte Carlo por exemplo.

A informação contida em uma distribuição *a posteriori* sintetiza de maneira elegante a incerteza acerca do modelo e a partir dela é possível extrair conhecimentos cruciais para a tomada de decisão a respeito das quantidades desconhecidas do estudo. Com o intuito de obter as melhores estimativas possíveis, especifica-se uma função de perda $L(\delta, \theta)$ a qual pode ser utilizada para quantificar o risco assumido em se utilizar δ como uma estimativa para a quantidade fixa θ :

$$R(\delta, \theta) := \mathbb{E}_{Y_{1:n}|\theta}[L(\delta(Y_{1:n}), \theta)]. \quad (2.3)$$

Ao associar a incerteza acerca de θ à função de risco esperado (2.3) obtemos o conceito de *Risco Bayesiano*:

$$\begin{aligned} \mathcal{R}(\delta, \theta) &= \mathbb{E}_{\theta}[R(\delta, \theta)] \\ &= \int_{\theta} \int_{\mathcal{Y}^n} L(\delta(y_{1:n}), \theta) p(dy_{1:n}|\theta) \pi(d\theta) \\ &= \int_{\mathcal{Y}^n} \left[\int_{\theta} L(\delta(y_{1:n}), \theta) p(d\theta|y_{1:n}) \right] p(dy_{1:n}) \\ &\geq \int_{\mathcal{Y}^n} \left[\min_{\delta} \int_{\theta} L(\delta(y_{1:n}), \theta) p(d\theta|y_{1:n}) \right] p(dy_{1:n}) \\ &= \mathbb{E}_{y_{1:n}} \left[\min_{\delta} \mathbb{E}_{\theta|y_{1:n}}[L(\delta(Y_{1:n}), \theta)] \right]. \end{aligned} \quad (2.4)$$

Desta forma, podemos adotar o seguinte estimador

$$\hat{\theta} := \arg \min_{\delta} \mathbb{E}_{\theta|y_{1:n}}[L(\delta, \theta)]. \quad (2.5)$$

Claramente, o emprego de funções de perda distintas acarretará na construção de estimativas finais distintas. Uma escolha bastante comum, a qual será adotada neste trabalho, é a função de perda quadrática

$$L(\delta, \theta) = (\delta - \theta)^t \mathbf{Q}(\delta - \theta), \quad (2.6)$$

onde \mathbf{Q} é uma matriz simétrica positiva definida. A expressão resultante em (2.6) é uma forma quadrática em δ , portanto seu mínimo pode ser obtido igualando o respectivo gradiente a zero. Neste caso, obtemos a estimativa final

$$\hat{\theta} = \mathbb{E}_{\theta|y_{1:n}}(\theta), \quad (2.7)$$

correspondente à media *a posteriori* de θ .

A inferência Bayesiana conta também com critérios objetivos para a seleção de um modelo paramétrico M_j dentre uma família de candidatos $\{M_j\}_{j=1}^N$. Este passo é indispensável já que o processo inferencial parte da especificação de um modelo probabilístico M para o problema de interesse. Caso o modelo definido não forneça uma descrição apropriada ao fenômeno em questão, o mecanismo de inferência pode culminar em resultados espúrios. Seguindo a discussão proposta em [Gamerman e Lopes \(2006\)](#), o modelo M_j , associado ao vetor de parâmetros θ_j , é

responsável pela distribuição *a posteriori*

$$p(\theta_j|y_{1:n}, M_j) = \frac{p(y_{1:n}|\theta, M_j)\pi(\theta_j|M_j)}{p(y_{1:n}|M_j)}. \quad (2.8)$$

A razão de chances do modelo M_j com respeito ao modelo M_k é dada por

$$\frac{P(M_j|y_{1:n})}{P(M_k|y_{1:n})} = \frac{p(y_{1:n}|M_j)}{\underbrace{p(y_{1:n}|M_k)}_{B_{jk}}} \cdot \frac{P(M_j)}{P(M_k)}, \quad (2.9)$$

sendo B_{jk} denominado o *fator de Bayes*. O fator de Bayes pode ser entendido como a razão de verossimilhanças de M_j para M_k . Desta forma, a relação $B_{jk} > 1$ evidencia que o modelo M_j é mais plausível que o modelo M_k para a análise dos dados $y_{1:n}$.

2.2 Cadeias de Markov

Um processo estocástico pode ser descrito sucintamente como uma coleção de variáveis aleatórias

$$X = \{X_t, t \geq 0\}$$

definida em um espaço mensurável (Ω, \mathcal{F}) . Para os propósitos deste trabalho, Ω se resumirá ao espaço Euclidiano n -dimensional \mathbb{R}^n equipado com a σ -álgebra de Borel $\mathcal{B}(\mathbb{R}^n)$, a menor σ -álgebra de \mathbb{R}^n contendo todos os abertos. O índice $t \in [0, +\infty)$ associado a cada variável aleatória (v.a.) X_t é convenientemente interpretado como o tempo.

Definição 1. Um processo estocástico X é dito mensurável se, para cada $A \in \mathcal{B}(\mathbb{R}^n)$, o conjunto $\{(t, \omega), X_t(\omega) \in A\}$ pertence à σ -álgebra produto $\mathcal{B}([0, +\infty)) \otimes \mathcal{F}$. Ou seja, se a aplicação

$$(t, \omega) \mapsto X_t(\omega) : ([0, +\infty) \times \Omega, \mathcal{B}([0, +\infty)) \otimes \mathcal{F}) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$$

é mensurável.

Ao associar σ -álgebras a um dado processo estocástico, temos como objetivo rastrear a informação disponível. Desta forma, considerando o fluxo temporal da dinâmica envolvida, podemos estabelecer noções de passado, presente e futuro e então avaliar o quanto um observador externo do processo pode inferir do estado presente comparado ao montante de informação conhecida do passado ou ao montante a ser revelado em pontos futuros.

Definição 2. Uma filtração associada ao espaço amostral (Ω, \mathcal{F}) é definida por uma família crescente $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$ de σ -álgebras de \mathcal{F} : $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ para quaisquer $0 \leq s < t$.

Dizemos que o processo estocástico X é adaptado à filtração $\{\mathcal{F}_t\}$ se, para cada $t \geq 0$, X_t é uma v.a. \mathcal{F}_t -mensurável. A escolha mais simples de filtração é aquela gerada pelo processo estocástico em questão: $\mathcal{F}_t^X := \sigma(X_s, 0 \leq s \leq t)$, a menor σ -álgebra com respeito a qual X_s é

mensurável para qualquer $s \in [0, t]$. Claramente, todo processo X é adaptado à filtração \mathcal{F}_t^X . A afirmação $A \in \mathcal{F}_t^X$ pode ser interpretada como: no instante de tempo t , um observador externo de X sabe dizer se o evento A já ocorreu. Um espaço mensurável (Ω, \mathcal{F}) associado a uma medida de probabilidade P e a uma filtração \mathbb{F} é denotado por espaço de probabilidade filtrado e é representado por $(\Omega, \mathcal{F}, \mathbb{F}, P)$.

Definição 3. Sejam $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ dois espaços mensuráveis. Um núcleo de transição de $(\Omega_1, \mathcal{F}_1)$ para $(\Omega_2, \mathcal{F}_2)$ é uma função $Q : \Omega_1 \times \Omega_2 \rightarrow [0, +\infty]$ satisfazendo

1. $\forall x \in \Omega_1, Q(x, \cdot)$ é uma medida positiva em $(\Omega_2, \mathcal{F}_2)$;
2. $\forall A \in \mathcal{F}_2$, a função $x \mapsto Q(x, A)$ é \mathcal{F}_1 -mensurável.

Se $Q(x, \Omega_2) = 1, \forall x \in \Omega_1$, dizemos que Q é um núcleo de transição, ou simplesmente um núcleo. Se $\Omega_1 = \Omega_2 = \Omega$ (similarmente $\mathcal{F}_1 = \mathcal{F}_2 = \mathcal{F}$) e $Q(x, \Omega) = 1, \forall x \in \Omega$, dizemos que Q define um núcleo de transição Markoviano em (Ω, \mathcal{F}) .

Dizemos que um núcleo de transição Q admite uma densidade com respeito à medida μ em Ω_2 se existir uma função não-negativa $q : \Omega_1 \times \Omega_2 \rightarrow [0, +\infty]$, mensurável com respeito à σ -álgebra produto $\mathcal{F}_1 \otimes \mathcal{F}_2$, tal que

$$Q(x, A) = \int_A q(x, y) \mu(dy), \quad A \in \mathcal{F}_2.$$

A função q é denominada a *função densidade de transição* ou *função densidade condicional*.

Definição 4 (Cadeia de Markov (homogênea)). Sejam $(\Omega, \mathcal{F}, \mathbb{F}, P)$ um espaço de probabilidade filtrado. Um processo estocástico $\{X_k\}_{k \geq 0}$ é dito uma *Cadeia de Markov Homogênea* sob P , com respeito à filtração \mathbb{F} e ao núcleo de transição Q , se ele for \mathbb{F} -adaptado e para quaisquer $k \geq 0$ e $A \in \Omega$,

$$P(X_{k+1} \in A | \mathcal{F}_k) = Q(X_k, A). \quad (2.10)$$

A distribuição de X_0 é denominada a *distribuição inicial*.

Uma cadeia de Markov define um processo estocástico cuja distribuição de probabilidade dos estados no instante k depende exclusivamente da configuração dos estados na posição imediatamente anterior $k - 1$. Uma propriedade fundamental destas cadeias é que suas distribuições finito-dimensionais podem ser completamente especificadas pela distribuição inicial em conjunto com o núcleo de transição da cadeia:

$$\mathbb{E}[f(X_0, \dots, X_k)] = \int_{\Omega^{k+1}} f(x_0, \dots, x_k) \nu(dx_0) \prod_{j=1}^k Q(x_{j-1}, dx_j),$$

onde ν expressa a distribuição inicial da cadeia, $k \geq 0$ e $f \in \mathcal{F}_b(\Omega^{k+1})$.

Em diversas situações é desejável que a cadeia de Markov sob estudo apresente um comportamento de estabilidade à medida que as observações evoluem no tempo, isto é, buscamos a existência de uma medida π tal que se $X_n \sim \pi$ então $X_{n+1} \sim \pi$, para n suficientemente grande. Formalmente, π deve satisfazer a relação de invariância

$$\pi(A) = \int Q(x,A)\pi(dx), \quad \forall A \in \mathcal{F}. \quad (2.11)$$

Nestes casos, a cadeia é dita estacionária com medida invariante π . Esta classe específica de cadeias será particularmente interessante para o desenvolvimento de métodos MCMC (Capítulo 3).

2.3 Cadeias Markovianas Ocultas

Modelos Markovianos Ocultos podem ser entendidos, de uma maneira informal, como cadeias de Markov observadas sob a influência de alguma fonte de ruídos. A metodologia lida basicamente com um processo estocástico $X \subset \mathbb{R}^{n_x}$, o qual representa a evolução dos estados latentes do sistema segundo uma estrutura Markoviana, e sua respectiva sequência de observações parciais $Y \subset \mathbb{R}^{n_y}$. O objetivo principal desses modelos é avaliar funções associadas à sequência de estados latentes $x_{1:n} = \{x_1, x_2, \dots, x_n\}$ condicionada aos valores observados parcialmente; um procedimento que passa fundamentalmente pela especificação da densidade condicional $p_\theta(x_{1:N}|y_{1:N})$, $\theta \in \mathbb{R}^{n_\theta}$. Uma descrição mais rigorosa do modelo é apresentada a seguir.

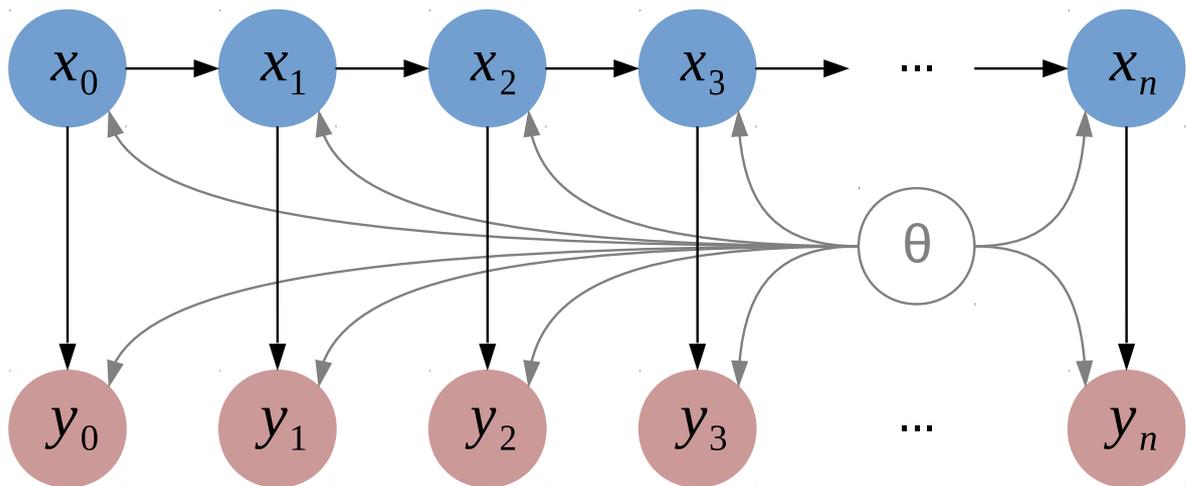


Figura 1 – Representação gráfica da estrutura de dependência de uma cadeia Markoviana oculta; $\{y_k\}_{k=0}^n$ ilustra a sequência de observações, $\{x_k\}_{k=0}^n$ a cadeia de Markov latente e θ o vetor de parâmetros estáticos.

Definição 5 (Modelo Markoviano Oculto). Sejam $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ espaços mensuráveis e suponha que Q e G representem, respectivamente, um núcleo de transição Markoviano em $(\Omega_1, \mathcal{F}_1)$ e um núcleo de transição de $(\Omega_1, \mathcal{F}_1)$ para $(\Omega_2, \mathcal{F}_2)$. Considere ainda o núcleo Markoviano definido no espaço produto $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ por

$$T((x, y), C) = \iint_C Q(x, dx') G(x', dy'), \quad (x, y) \in X \times Y, C \in \mathcal{F}_1 \otimes \mathcal{F}_2. \quad (2.12)$$

A cadeia de Markov $\{X_k, Y_k\}_{k \geq 0}$ com núcleo de transição T e distribuição inicial $V \otimes G$, onde V é uma medida de probabilidade em $(\Omega_1, \mathcal{F}_1)$, é denominada *modelo oculto de Markov*.

Neste trabalho estaremos interessados no caso particular de modelos cujas relações de transição possam ser representadas por meio de funções densidade de probabilidade em seu sentido mais usual (com respeito à medida de Lebesgue). Para tal, precisamos deixar claras as hipóteses de dominação acerca do modelo.

Definição 6. Sejam λ e μ medidas de probabilidade em $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$, respectivamente, tais que para todo $x \in \Omega_1$, $Q(x, \cdot)$ e $G(x, \cdot)$ sejam absolutamente contínuos com respeito a λ e μ , $G(x, \cdot) \ll \mu(\cdot)$ e $Q(x, \cdot) \ll \lambda(\cdot)$, com funções densidade de transição $q(x, \cdot)$ e $g(x, \cdot)$. Neste caso, para qualquer $A \in \mathcal{F}_1$, temos que

$$Q(x, A) = \int_A q(x, x') \lambda(dx')$$

e o modelo é dito *completamente dominado*. O núcleo de transição Markoviano conjunto T é então dominado pela medida produto $\lambda \otimes \mu$ e admite a função densidade de transição

$$t((x, y), (x', y')) := q(x, x') g(x', y'). \quad (2.13)$$

Ressaltamos que nos modelos completamente dominados, a notação v será utilizada para representar a *função densidade de transição* do estado inicial X_0 (com respeito a λ) ao invés da própria distribuição de probabilidade. Com esta notação, o problema pode ser formalmente descrito por:

$$\begin{cases} v(x_0) & \text{distribuição inicial;} \\ q(x_t | x_{t-1}), \quad t > 1, & \text{dinâmica de evolução dos estados latentes;} \\ g(y_t | x_t), \quad t \geq 0, & \text{dinâmica das observações parciais,} \end{cases}$$

onde v, q e g são funções densidade de probabilidade. Desta forma, é suficiente especificar v, q e g para obter uma descrição completa do problema.

A definição de cadeia oculta de Markov carrega nas entrelinhas um conceito crucial para o desenvolvimento de soluções para o problema da filtragem: independência condicional das observações. É possível demonstrar que (CAPPÉ; MOULINES; RYDEN, 2005), condicionadas ao processo latente $\{X_k\}_{k \geq 0}$, as observações $\{Y_k\}_{k \geq 0}$ são independentes e, para cada n , a distribuição condicional de Y_n depende apenas de X_n .

2.4 O Problema da Filtragem

O objetivo primário dos modelos ocultos Markovianos é a realização de inferências a respeito da sequência não-observada X_0, X_1, \dots, X_t , condicionada a um conjunto de observações $Y_0, Y_1, Y_2, \dots, Y_t$. Nesta seção formalizaremos esta questão além de apresentar alguns dos desafios impostos por ela. Seguindo as definições apresentadas anteriormente, a discussão será focada em modelos completamente dominados por medidas de Lebesgue.

Definição 7 (Verossimilhança). Seja $\{X_k, Y_k\}_{k \geq 0}$ um modelo oculto de Markov em $(\Omega_1 \times \Omega_2)$ com núcleo de transição $T = Q \times G$ completamente dominado, conforme (2.13). A verossimilhança das observações é a função densidade de probabilidade de Y_0, Y_1, \dots, Y_t , definida para qualquer sequência de observações $(y_0, y_1, \dots, y_t) \in \Omega_2^{t+1}$, por

$$L^{(t)}(y_{0:t}) = \int \int \cdots \int v(x_0) g(y_0|x_0) \left\{ \prod_{k=1}^t g(y_k|x_k) q(x_k|x_{k-1}) \right\} dx_0 dx_1 \dots dx_t. \quad (2.14)$$

A construção da função de verossimilhança das observações é um quesito básico em qualquer problema de inferência estatística. No entanto, nos modelos em questão, esta relação em separado não é sensível ao comportamento do objeto de estimação (sequência oculta) como nos problemas clássicos de inferência paramétrica. O foco da inferência deve ser então um elo de ligação entre os espaços das observações e dos estados latentes.

Definição 8 (Filtragem e Predição). Sejam $r^{(t)}$ e $p^{(t)}$ as funções densidade de probabilidade de $X_{0:t}$ e X_{t+1} condicionadas à sequência de observações $Y_{0:t}$:

$$X_{0:t}|Y_{0:t} \sim r^{(t)}, \quad X_{t+1}|Y_{0:t} \sim p^{(t)}.$$

O processo de inferência das densidades $r^{(t)}$ e $p^{(t)}$ é denominado filtragem e predição, respectivamente.

Em aplicações envolvendo aprendizagem em tempo real, estaremos interessados na inferência a respeito do estado mais atual, dada a sequência de observações. Nestes casos, o foco do problema da filtragem é a densidade marginal

$$m^{(t)}(x_t|y_{0:t}) = \int \cdots \int r^{(t)}(x_{0:t}|y_{0:t}) dx_0 \dots dx_{t-1}.$$

A densidade alvo do problema da filtragem é suficiente para descrever uma relação de transição entre os espaços Ω_1 e Ω_2 :

- dado um evento $A \in \mathcal{F}_1^{\otimes t+1}$, $y_{0:n} \mapsto \int \cdots \int_A r^{(t)}(x_{0:t}|y_{0:t}) dx_{0:t}$ é uma função Ω_2^{t+1} -mensurável;
- dada uma sequência $y_{0:t}$, $A \mapsto \int \cdots \int_A r^{(t)}(x_{0:t}|y_{0:t}) dx_{0:t}$ é uma medida de probabilidade em $(\Omega_1^{t+1}, \mathcal{F}_1^{\otimes t+1})$.

Neste cenário, a verossimilhança das observações aparece apenas como um instrumento complementar para a obtenção de $f^{(t)}$. Esta relação de complementação fica evidente ao analisarmos as seguintes recursões obtidas a partir do Teorema de Bayes:

$$r^{(t)}(x_{0:t}|y_{0:t}) = r^{(t-1)}(x_{0:t-1}|y_{0:t-1})L^{(t-1)}(y_{0:t-1})\frac{g(y_t|x_t)q(x_t|x_{t-1})}{L^{(t)}(y_{0:t})}, \quad (2.15)$$

$$p^{(t-1)}(x_t|y_{0:t-1}) = \int q(x_t|x_{t-1})m^{(t-1)}(x_{t-1}|y_{0:t-1})dx_{t-1}, \quad (2.16)$$

$$m^{(t)}(x_t|y_{0:t}) = L^{(t-1)}(y_{0:t})\frac{g(y_t|x_t)p^{(t-1)}(x_t|y_{0:t-1})}{L^{(t)}(y_{0:t})}. \quad (2.17)$$

As equações acima induziriam naturalmente um algoritmo de estimação para o problema da filtragem caso as expressões envolvidas pudessem ser tratadas analiticamente em casos gerais. Contudo, são raras as subclasses do modelo cujas estruturas admitem este tratamento.

Uma vez solucionado o problema da filtragem, estaremos interessados em avaliar quantidades da forma:

$$I(h) = \mathbb{E}[h(X_{0:t})|y_{0:t}] = \int \cdots \int h(x_{0:t})r^{(t)}(x_{0:t}|y_{0:t})dx_0 \cdots dx_t, \quad (2.18)$$

onde $h : \Omega_1^{t+1} \rightarrow \mathbb{R}$ é uma função integrável com respeito à densidade filtrada. Escolhas bastante usuais de integradores são

$$h(x_{0:t}) = x_{0:t} \quad \text{e} \quad h(x_{0:t}) = x_t x_t^T - \mathbb{E}[x_t|y_{0:t}]\mathbb{E}^T[x_t|y_{0:t}],$$

as quais resultam em estimativas para média e variância condicionais, respectivamente. Mesmo obtendo expressões fechadas para a densidade filtrada $f^{(t)}$, na prática ainda podemos encontrar dificuldades na conclusão da análise devido à relação integral acima. O custo de lidarmos com a ausência de restrições fortes na especificação funcional das densidades de transição é basicamente o surgimento de integrandos complicados o suficiente para impossibilitar o desenvolvimento de soluções puramente analíticas.

2.5 Suavização

Suavização corresponde a um procedimento intimamente relacionado à questão da filtragem. Enquanto os dispositivos de filtragem focam na inferência dos estados latentes $X_{0:t}$ mais atuais condicionados nas medições $Y_{0:t}$ disponíveis até o presente instante t , os métodos de suavização visam estimar a distribuição de um estado X_t dadas todas as medições até um instante posterior Y_{t+k} . Como um resultado imediato do acesso a informações extras, temos a construção de estimativas mais suaves se comparadas às obtidas no passo de filtragem.

Definição 9 (Suavização). Assuma que em um dado instante de tempo k temos acesso às medições $Y_{0:t}$, com $t > k$ e considere $\phi^{(k,t)}$ a distribuição de X_k condicionada a esta sequência de informações:

$$X_k|Y_{0:t} \sim \phi^{(k,t)}, \quad k < t.$$

O processo de inferência da densidade $\phi^{(k,t)}$ é denominado suavização.

Uma análise mais precisa do problema da suavização pode ser feita a partir dos resultados já estabelecidos para densidade de filtragem $r^{(t)}(x_{0:t}|y_{1:t})$, para tal basta utilizarmos um passo de integração marginal com respeito às variáveis latentes $(X_{0:k-1}, X_{k+1:t})$.

$$\begin{aligned} \phi^{(k,t)}(x) &= \int r^{(t)}(x_{0:t}|y_{0:t}) dx_{0:k-1} dx_{k+1:t} \\ &= \frac{1}{L^{(t)}(y_{0:t})} \int \cdots \int \left\{ \left[v(x_0)g(y_0|x_0) \prod_{j=1}^k q(x_j|x_{j-1})g(y_j|x_j) dx_j \right] \right. \\ &\quad \left. \times \left[q(x_{k+1}|x)g(y_{k+1}|x_{k+1}) \prod_{j=k+2}^t q(x_j|x_{j-1})g(y_j|x_j) dx_j \right] \right\}, \end{aligned}$$

onde $\prod_{j=1}^k q(x_j|x_{j-1})g(y_j|x_j) dx_j$ e $\prod_{j=k+2}^t q(x_j|x_{j-1})g(y_j|x_j) dx_j$ são vazios para $k=0$ e $t=1$, respectivamente. Para simplificar a notação, é comum definir as quantidades

$$\alpha^{(k)}(y_{0:k}) := \int \cdots \int v(x_0)g(y_0|x_0) \prod_{j=1}^k q(x_j|x_{j-1})g(y_j|x_j) dx_j \quad (2.19)$$

e

$$\begin{aligned} \beta^{(t,t)} &\equiv 1, \\ \beta^{(k,t)}(y_{k+1:t}, x) &:= \int \cdots \int q(x_{k+1}|x)g(y_{k+1}|x_{k+1}) \prod_{j=k+2}^t q(x_j|x_{j-1})g(y_j|x_j) dx_j, \end{aligned} \quad (2.20)$$

denominadas sequencias sucessivas e retroativas, respectivamente. Assim, chegamos à seguinte família de recursões sucessivas-retroativas (“*forward-backward recursions*”) para o problema de suavização:

$$\phi^{(k,t)}(x) = \frac{1}{L^{(t)}(y_{0:t})} \int \alpha^{(k)}(y_{0:k}) \beta^{(k,t)}(y_{k+1:t}, x), \quad k < t. \quad (2.21)$$

Fórmulas recursivas mais sofisticadas podem ser desenvolvidas para o tratamento do problema da estimação. Tais abordagens não são discutidas aqui pois nosso objetivo é apresentar apenas os pilares da teoria e não cobrir todo o tema; uma excelente revisão das recursões de suavização é apresentada no trabalho de [Cappé, Moulines e Ryden \(2005\)](#).

2.6 Parâmetros Estáticos

Nas seções anteriores, assumimos que as funções de densidade v , q e g eram completamente conhecidas, permitindo assim que a inferência fosse concentrada no problema dos estados latentes. Contudo, é comum conhecermos apenas as famílias correspondentes a cada uma das densidades, de modo que sua especificação completa passaria pela seleção adequada de um vetor

de parâmetros $\theta \in \mathbb{R}^{n_\theta}$. A formulação mais adequada do modelo seria então da forma:

$$\begin{cases} v_\theta(x_0) & \text{distribuição inicial;} \\ q_\theta(x_t|x_{t-1}), \quad t > 1, & \text{dinâmica de evolução dos estados latentes;} \\ g_\theta(y_t|x_t), \quad t \geq 0, & \text{dinâmica das observações parciais.} \end{cases}$$

Neste caso, a função de verossimilhança herda prontamente a dependência de θ :

$$L_\theta^{(t)}(y_{0:t}) = \int \int \cdots \int v_\theta(x_0) g_\theta(y_0|x_0) \left\{ \prod_{k=1}^t g_\theta(y_k|x_k) q_\theta(x_k|x_{k-1}) \right\} dx_0 dx_1 \cdots dx_t. \quad (2.22)$$

Diferentemente da sequência de estados latentes $X_{1:t}$, o vetor θ é uma quantidade fixa, isto é, não possui dinâmica de evolução natural. Por esta razão θ é comumente denominado o conjunto de *parâmetros estáticos* do modelo.

A inferência destes parâmetros pode seguir as metodologias tradicionais, frequentista ou Bayesiana, partindo da função de verossimilhança (2.14). No âmbito frequentista, métodos baseados na maximização da função de verossimilhança $L_\theta^{(t)}$ normalmente fazem uso da função *score* e da matriz de informação, as quais estão relacionadas às duas primeiras derivadas da função *log-verossimilhança* com respeito a θ . Em um primeiro momento, é possível explorar as identidades de Fisher-Louis (Cappé, Moulines e Ryden (2005)) de modo a expressar essas quantidades em termos das densidades de transição do modelo:

$$\nabla_\theta l^{(t)}(\theta) = \int \nabla_\theta \log p_\theta(x_{0:t}, y_{0:t}) r_\theta^{(t)}(x_{0:t}|y_{0:t}) dx_{0:t}, \quad (2.23)$$

$$-\nabla_\theta^2 l^{(t)}(\theta) = \{\nabla_\theta^2 l^{(t)}(\theta)\} \{\nabla_\theta^2 l^{(t)}(\theta)\}^T - \int [\nabla_\theta^2 p_\theta(x_{0:t}, y_{0:t}) + \{\nabla_\theta^2 \log p_\theta(x_{0:t}, y_{0:t})\} \{\nabla_\theta^2 \log p_\theta(x_{0:t}, y_{0:t})\}^T] r_\theta^{(t)}(x_{0:t}|y_{0:t}) dx_{0:t}, \quad (2.24)$$

onde

$$l^{(t)}(\theta) := \log L_\theta^{(t)}$$

e

$$p_\theta(x_{0:t}, y_{0:t}) = v_\theta(x_0) g_\theta(y_0|x_0) \prod_{k=1}^t g_\theta(y_k|x_k) q_\theta(x_k|x_{k-1}).$$

Entretanto, apesar da simplicidade usual das densidades de transição q e g , assim como da condição inicial v , as integrais acima impossibilitam o tratamento analítico das quantidades de interesse. No contexto Bayesiano, um problema análogo surge na especificação da distribuição *a posteriori*

$$p(\theta|y_{0:t}) = \frac{L_\theta^{(t)}(y_{0:t}) p(\theta)}{\int L_\theta^{(t)}(y_{0:t}) p(\theta) d\theta}. \quad (2.25)$$

Novamente, a complexidade do passo de integração impossibilita o desenvolvimento de uma tratativa analítica para o problema.

De uma forma geral vimos que os problemas de filtragem, suavização e aprendizagem de parâmetros estáticos são formalmente bem definidos, no entanto compartilham a dependência de integrais múltiplas, de dimensões elevadas, com respeito a funções não triviais. Em alguns casos é possível realizar ajustes finos no modelo, tais como transformações ou condicionamentos parciais, de modo a diminuir a complexidade das recursões envolvidas, porém estes casos são menos numerosos. Na prática tratamentos analíticos para estas integrais não são acessíveis, o que fornece uma grande motivação para o desenvolvimento de um ferramental numérico robusto.

MÉTODOS DE MONTE CARLO

Métodos de Monte Carlo englobam uma família de algoritmos fundamentados na simulação de eventos probabilísticos com o intuito de construir aproximações numéricas para problemas complexos. Neste trabalho estaremos particularmente interessados em procedimentos do tipo *Integração de Monte Carlo*, uma subclasse de procedimentos ligados à aproximação de integrais em cenários onde os tradicionais métodos de quadratura não apresentam desempenho satisfatório. De uma forma geral, o problema se resume ao cômputo de integrais da forma

$$(f, \pi) = \mathbb{E}_\pi[f(X)] = \int_{\Omega} f(x)\pi(dx), \quad (3.1)$$

onde Ω representa o espaço amostral de interesse, $f : \Omega \rightarrow \mathbb{R}$ é uma função mensurável qualquer e π é uma medida de probabilidade cuja interpretação varia de acordo com a aplicação.

Integradores de Monte Carlo tratam o problema (3.1) através de aproximações da medida contínua π por meio de medidas aleatórias da forma

$$\hat{\pi}(dx) = \sum_{i=1}^N \omega \cdot \delta_{x^{(i)}}(dx), \quad (3.2)$$

onde $\omega : \Omega \rightarrow \mathbb{R}$ é uma função auxiliar de ponderação, $\delta_x(\cdot)$ é o operador delta de Dirac e $\{x^{(i)}\}_{i=1}^N$ é um suporte aleatório obtido por meio da realização de algum modelo probabilístico de tratamento simples. Tal perspectiva é consistente por ser capaz de assegurar que a proximidade entre as distribuições $\hat{\pi}$ e π aumenta à medida que o tamanho amostral N cresce. Assim, podemos substituir o cálculo de integrais complexas pela estimativa da média ponderada:

$$\begin{aligned} (f, \pi) &\approx (f, \hat{\pi}) \\ &= \sum_{i=1}^N \int_{\Omega} (f \cdot \omega)(x) \delta_{x^{(i)}}(dx) = \sum_{i=1}^N (\omega \cdot f)(x^{(i)}), \end{aligned} \quad (3.3)$$

cujas consistência é assegurada pela Lei dos Grandes Números (LGN).

Em cenários ideais, podemos considerar que o mecanismo de amostragem do suporte é capaz de simular realizações diretamente do modelo π de modo a construir aproximações simplificadas

$$\hat{\pi}(dx) = \frac{1}{N} \sum_{i=1}^N \delta_{X^{(i)}}(dx). \quad (3.4)$$

Neste caso, temos a seguinte estimativa:

$$(f, \hat{\pi}) = \int_{\Omega} f(x) \hat{\pi}(dx) = \frac{1}{N} \sum_{i=1}^N f(x^{(i)}), \quad (3.5)$$

a qual também converge para a solução do problema (3.1) por um argumento análogo ao considerado na relação (3.3). Além disso, temos que o dispositivo oferece estimativas não viciadas cuja variância é dada por

$$\begin{aligned} \text{Var}[(f, \hat{\pi})] &= \frac{\text{Var}_{\pi}[f(X)]}{N} \\ &= \frac{1}{N} \int_{\Omega} (f(x) - \mathbb{E}_{\pi}[f(X)])^2 \pi(dx), \end{aligned} \quad (3.6)$$

indicando que a precisão da estimativa aumenta à medida que o tamanho amostral N cresce. Podemos assegurar ainda que, sob condições de regularidade, a distribuição das estimativas construídas por meio de integradores de Monte Carlo pode ser estudada com o auxílio do Teorema Central do Limite (TCL):

$$\sqrt{N} \left((f, \hat{\pi}) - (f, \pi) \right) \xrightarrow{\mathcal{D}} N \left(0, \text{Var}_{\pi}[f(X)] \right), \quad (3.7)$$

à medida que $N \rightarrow +\infty$. Desta forma, em linhas gerais, o problema da integração é substituído pela busca de uma sistemática eficiente para a simulação de valores provenientes da distribuição alvo π . Dentre as virtudes da integração de Monte Carlo podemos citar o fato de que $\left\{ x^{(i)} \right\}_{i=1}^N$ garante automaticamente uma cobertura das regiões mais relevantes do domínio de integração (espaço amostral) e sua taxa de convergência da ordem de $1/\sqrt{N}$ independe da dimensão do problema. Em contrapartida, nem sempre é possível amostrar diretamente do modelo π e mesmo metodologias de amostragem indireta podem ser computacionalmente complexas.

As próximas seções apresentam sucintamente alguns dos principais métodos de simulação de variáveis aleatórias: amostragem por rejeição, amostragem por importância e Monte Carlo via Cadeias de Markov (*Markov Chain Monte Carlo* (MCMC)). A abordagem visa elucidar apenas os conceitos básicos empregados no campo dos métodos SMC, os quais constituem o principal objeto de estudo deste trabalho. As seções seguintes apresentam com mais riqueza de detalhes as ideias principais acerca dos algoritmos de amostragem sequencial ou *filtros de partículas*. A discussão abordará tanto os conceitos de filtragem pura quanto a aprendizagem de parâmetros estáticos, a partir de uma revisão dos principais algoritmos encontrados na literatura.

3.1 Amostragem por Rejeição

Obter amostras oriundas do modelo π , passo indispensável para a implementação dos integradores de Monte Carlo, pode ser uma tarefa bastante complexa analítica e computacionalmente. Nos casos mais simples seria possível explorar, por exemplo, a utilização de métodos de transformação como os dispositivos de Box-Muller ou da transformação inversa. No entanto, tais métodos não podem ser aplicados de forma ampla e objetiva, dificultando a construção eficaz de algoritmos mais robustos. Uma abordagem tradicional que pode ser estendida a uma família mais vasta de problemas, denominada *amostragem por rejeição*, sugere que o procedimento pode ser realizado de maneira indireta através da simulação de pontos provenientes de uma distribuição proposta q de tratamento mais simples. As instâncias geradas indiretamente a partir do modelo q são aceitas ou rejeitadas segundo um critério quantitativo capaz de assegurar ao final do algoritmo a obtenção de uma amostra aleatória da distribuição alvo π . A premissa básica para a aplicação do algoritmo é a escolha de uma proposta absolutamente contínua com respeito ao alvo, isto é, devemos escolher q tal que

$$\pi(A) = 0 \implies q(A) = 0, \quad \forall A \in \mathcal{B}(\Omega). \quad (3.8)$$

Tal relação entre medidas é usualmente expressa por $q \ll \pi$ e assegura que a distribuição proposta seja incapaz de realizar observações de probabilidade nula com respeito ao modelo original. Além disso, é preciso garantir a existência de uma cota superior κ para a densidade de π com respeito a q :

$$\frac{d\pi}{dq}(x) \leq \kappa, \quad \forall x \in \Omega.$$

Algoritmo 1: Amostragem por Rejeição

Entrada: Tamanho amostral (N), medidas alvo (π) e proposta (q);

Saída: Amostra $\{X^{(i)}\}_{i=1}^N \sim \pi$;

```

1 para  $i \leftarrow 1$  até  $N$  faça
2   repita
3     amostre  $x \sim q$ ;
4     amostre  $U \sim u(0, 1)$ ;
5     até  $U \leq \frac{1}{\kappa} \frac{d\pi}{dq}(x)$ ;
6      $X^{(i)} \leftarrow x$ ;
7 fim
8 retorna  $\{X^{(i)}\}_{i=1}^N$ ;
```

O dispositivo prático de rejeição pode ser explorado também em cenários onde a forma funcional de π é conhecida a menos de uma constante de normalização:

$$\pi(dx) = \frac{1}{Z} \eta(dx), \quad Z = \mathbb{E}_\eta[1],$$

a qual pode ser inserida no algoritmo de forma direta.

O método da amostragem por rejeição pode ser justificado por um argumento geométrico simples e pode ser explorado nos mais variados contextos, entretanto, sua eficiência é fortemente dependente do parâmetro κ . A taxa de aceitação do método pode ser obtida da seguinte forma:

$$\begin{aligned}
 P\left(U \leq \frac{1}{\kappa} \frac{d\pi}{dq}(X)\right) &= \mathbb{E}_q \left[P\left(U \leq \frac{1}{\kappa} \frac{d\pi}{dq}(X) \mid X\right) \right] \\
 &= \mathbb{E}_q \left[\frac{1}{\kappa} \frac{d\pi}{dq}(X) \right] \\
 &= \frac{1}{\kappa} \int_{\Omega} \frac{d\pi}{dq}(x) q(dx) \\
 &= \frac{1}{\kappa} \int_{\Omega} \pi(dx) = \frac{1}{\kappa}.
 \end{aligned} \tag{3.9}$$

Na prática, quando κ é associado a valores elevados (muito superiores a 1), a vasta maioria dos pontos gerados pelo modelo proposto serão descartados. Situações desta natureza são observadas principalmente quando a diferença entre as medidas π e q é muito significativa, o que ocorre frequentemente quando a complexidade de π é grande. Nestes casos, a busca por estratégias tão abrangentes quanto este dispositivo, porém de menor custo computacional, é tratada como um problema de grande relevância no contexto dos métodos de Monte Carlo.

3.2 Amostragem por Importância

O dispositivo de *amostragem por importância* pode ser entendido como uma solução imediata para o problema de eficiência apresentado pela técnica de amostragem por rejeição. Este segundo conceito também faz uso de um modelo proposto q o qual segue as mesmas condições de regularidades do caso anterior. Entretanto, o passo de rejeição é substituído por uma espécie de passo de “correção” que possibilita o aproveitamento de todas as observações geradas a partir de q , conferindo mais eficiência ao método.

O integrador de Monte Carlo obtido por meio desta perspectiva é associado a um problema transformado, o que pode ser descrito pela relação fundamental:

$$(f, \pi) = \left(f \frac{d\pi}{dq}, q \right), \quad q \ll \pi. \tag{3.10}$$

É comum resumirmos a relação acima através da introdução da seguinte quantidade

$$\omega(x) := \frac{d\pi}{dq}(x), \tag{3.11}$$

a qual é popularmente conhecida como *peso de importância* de x , ou simplesmente, *importância* de x . Em notação integral, obtemos:

$$(f, \pi) = \int_{\Omega} (\omega \cdot f)(x) q(dx). \tag{3.12}$$

A medida proposta q é escolhida de modo a garantir a utilização de um modelo similar a π e do qual seja possível obter realizações $\{x^{(i)}\}_{i=1}^N$ independentes e identicamente distribuídas. Neste caso, lidaremos com aproximações da forma:

$$\hat{\pi} = \frac{1}{N} \sum_{i=1}^N \omega \cdot \delta_{x^{(i)}}(dx) \quad \text{e} \quad (f, \hat{\pi}) = \frac{1}{N} \sum_{i=1}^N (\omega \cdot f)(x^{(i)}), \quad (3.13)$$

as quais consistem de estimativas não viciadas e consistentes em analogia com os resultados já mencionados no início do capítulo. Podemos considerar também a versão normalizada de (3.13):

$$(f, \tilde{\pi}) = \frac{\frac{1}{N} \sum_{i=1}^N (\omega \cdot f)(x^{(i)})}{\frac{1}{N} \sum_{k=1}^N \omega(x^{(k)})} = \sum_{i=1}^N (\tilde{\omega} \cdot f)(x^{(i)}), \quad (3.14)$$

onde

$$\tilde{\omega}(\cdot) := \omega(\cdot) / \sum_{k=1}^N \omega(x^{(k)}).$$

A correção adotada acima é assintoticamente consistente já que

$$\frac{1}{N} \sum_{k=1}^N \omega(x^{(k)}) \xrightarrow{q.c.} \mathbb{E}_q \left[\frac{d\pi}{dq}(X) \right] = \mathbb{E}_\pi[1] = 1,$$

à medida que $N \rightarrow +\infty$. Apesar de acrescentar um pequeno vício às estimativas, a versão normalizada (3.14) pode ser explorada em situações onde π é conhecida a menos de uma constante e, além disso, oferece um menor erro quadrático médio de aproximação conforme apontado por Liu (2004).

É razoável assumir que a variabilidade dos estimadores obtidos com o auxílio da amostragem por importância dependa diretamente da distribuição proposta q . De fato,

$$\begin{aligned} \text{Var}[(f, \hat{\pi})] &= \frac{1}{N} \text{Var}_q[(\omega \cdot f)(X)] \\ &= \frac{1}{N} \left[\int_{\Omega} (\omega \cdot f)^2(x) q(dx) - \left(\int_{\Omega} (\omega \cdot f)(x) q(dx) \right)^2 \right] \\ &= \frac{1}{N} \left[\mathbb{E}_q[(\omega \cdot f)^2(X)] - \left(\int_{\Omega} f(x) \pi(dx) \right)^2 \right] \\ &= \frac{1}{N} (\mathbb{E}_q[(\omega \cdot f)^2(X)] - \mathbb{E}_\pi^2[f(X)]). \end{aligned} \quad (3.15)$$

Desta forma, podemos tentar obter a proposta capaz de minimizar a dispersão das estimativas. Considerando q uma proposta qualquer e tomando

$$q^* = |f| \cdot \pi / \mathbb{E}_\pi[|f(X)|], \quad (3.16)$$

obtemos:

$$\begin{aligned}
\text{Var}_{q^*}[(\omega^* \cdot f)(X)] &= \mathbb{E}_{q^*}[(\omega^* \cdot f)^2(X)] - \mathbb{E}_{\pi}^2[f(X)] \\
&= \int_{\Omega} \frac{(\pi \cdot f^2)(x) \pi(dx)}{q^*(x)} - \mathbb{E}_{\pi}^2[f(X)] \\
&= \mathbb{E}_{\pi}^2[|f(X)|] - \mathbb{E}_{\pi}^2[f(X)] \\
&= \mathbb{E}_q^2\left[\left(\frac{|f|\pi}{q}\right)(X)\right] - \mathbb{E}_{\pi}^2[f(X)] \\
&\leq \mathbb{E}_q[(\omega \cdot f)^2(X)] - \mathbb{E}_{\pi}^2[f(X)] \\
&= \text{Var}_q[(\omega \cdot f)(X)], \tag{3.17}
\end{aligned}$$

sendo a penúltima passagem uma aplicação direta da desigualdade de Jensen. Assim, a relação (3.16) corresponde à escolha ótima de distribuição proposta.

A eficiência dos integradores de Monte Carlo construídos a partir da amostragem por importância pode ser acessada através de um estudo comparativo entre estes estimadores e aqueles obtidos por meio de simulação perfeita – como no caso do método da rejeição por exemplo. Liu (2004) mostra que a variabilidade do procedimento baseado em amostragem por importância pode ser dada alternativamente por:

$$\text{Var}^{(imp)}[(f, \pi)] = \frac{1}{N} \text{Var}_q[(\omega \cdot f)(X)] \approx \frac{1}{N} \text{Var}_{\pi}[f(X)] \{1 + \text{Var}_q[\omega(X)]\}. \tag{3.18}$$

Utilizando a variabilidade já conhecida do estimador obtido por meio de M observações provenientes de dispositivos de simulação perfeita,

$$\text{Var}^{(per)}[(f, \pi)] = \frac{1}{M} \text{Var}_{\pi}[f(X)], \tag{3.19}$$

temos que a eficiência relativa entre as duas abordagens segue da razão:

$$\frac{\text{Var}^{(imp)}[(f, \pi)]}{\text{Var}^{(per)}[(f, \pi)]} \approx \frac{M}{N} \{1 + \text{Var}_q[\omega(X)]\}. \tag{3.20}$$

Finalmente, notemos que o tamanho amostral N_{ef} – tamanho amostral em um experimento baseado em amostragem perfeita para garantir a mesma efetividade do estimador por importância – pode ser extraído imediatamente da relação acima:

$$\frac{N_{ef}}{N} \{1 + \text{Var}_q[\omega(X)]\} = 1 \iff N_{ef} = \frac{N}{1 + \text{Var}_q[\omega(X)]} < N. \tag{3.21}$$

Por esta razão, a quantidade acima é denotada o *Tamanho Efetivo da Amostra* (TEA). Como a função ω é conhecida a menos de uma constante em muitas situações, a variância presente no denominador acima é estimada através do coeficiente de variação dos pesos ainda não normalizados:

$$\text{CV}_q[\omega(X)] \approx \sum_{i=1}^N \left(\omega^{(i)} - \bar{\omega}\right)^2 / N \cdot \bar{\omega}^2, \tag{3.22}$$

onde $\omega^{(i)} = \omega(x^{(i)})$ e $\bar{\omega} = \frac{1}{N} \sum_{i=1}^N \omega^{(i)}$. Assim, o TEA do experimento pode ser expresso por

$$\begin{aligned} N_{ef} &= \frac{N}{1 + \text{Var}_q[\omega(X)]} \\ &\approx \frac{N}{1 + \text{CV}_q[\omega(X)]} \\ &\approx \frac{(N \cdot \bar{\omega})^2}{\sum_{i=1}^N (\omega^{(i)})^2} = \frac{1}{\sum_{i=1}^N (\tilde{\omega}^{(i)})^2}. \end{aligned} \quad (3.23)$$

Na prática, o tamanho efetivo amostral informa que a inferência baseada em amostragem por importância carrega um montante de informação equivalente àquela oferecida por meio de uma amostra iid de tamanho inferior. Este resultado é bastante intuitivo uma vez que as aproximações por importância ganham mais eficiência computacional através da construção de soluções um pouco mais grosseiras.

3.3 Monte Carlo via Cadeias de Markov

Métodos de Monte Carlo via Cadeias de Markov são baseados na geração de cadeias de Markov cuja distribuição estacionária coincide com o modelo alvo π . Neste caso, assim como na amostragem por rejeição, teremos ao final do algoritmo uma amostra governada pela medida π . De forma sucinta, a ideia é gerar uma cadeia suficientemente longa e tomar como amostra resultante um subconjunto das últimas observações geradas pela cadeia. Neste caso, o desafio passa a ser determinar procedimentos eficientes para a simulação de cadeias ergódicas com a distribuição estacionária adequada. A metodologia padrão apoia-se na obtenção de núcleos de transição K os quais satisfazem o *Princípio do Balanceamento Detalhado*:

$$\phi(x)K(x,y) = \phi(y)K(y,x), \quad (3.24)$$

para alguma função $\phi : \Omega \rightarrow \mathbb{R}$. Caso a condição de balanceamento detalhado seja satisfeita para o núcleo de transição K e ϕ seja uma função densidade de probabilidade, então a cadeia é reversível, portanto estacionária, e possui ϕ como densidade associada à distribuição invariante.

A versão mais utilizada deste algoritmo explora a perspectiva de Metropolis-Hastings (MH), a qual define a transição da cadeia a partir de uma mistura finita

$$K(x_n, x_{n+1}) = \alpha_n q(x_{n+1}|x_n) + (1 - \alpha_n) \delta_{x_n}(x_{n+1}), \quad (3.25)$$

onde q é uma densidade proposta (usualmente simétrica) e $\alpha_n \in (0, 1)$ é um parâmetro de aceitação que evolui sequencialmente como função de x_n e x_{n+1} . Dada a natureza da transição adotada por este dispositivo, temos que a evolução da cadeia pode ser implementada de forma direta por meio de um passo de rejeição. Uma variante simplificada do método é descrita pelo Algoritmo 13 a seguir.

Algoritmo 2: Metropolis-Hastings**Entrada:** Tamanho amostral (N), densidades de transição (K), inicial (ν) e alvo (π);**Saída:** Amostra aleatória $\{X^{(i)}\}_{i=1}^N \sim \pi$;

```

1 amostre  $X_0 \sim \nu$ ;
2  $n \leftarrow 1$ ;
3 repita
4   amostre  $\tilde{X}_n \sim K(X_{n-1}, \cdot)$ ;
5   amostre  $U \sim u(0, 1)$ ;
6    $\alpha \leftarrow \min\left(1, \frac{\pi(\tilde{X}_n)K(\tilde{X}_n, X_{n-1})}{\pi(X_{n-1})K(X_{n-1}, \tilde{X}_n)}\right)$ ;
7   se  $U \leq \alpha$  então
8      $X_n \leftarrow \tilde{X}_n$ ;
9   senão
10     $X_n \leftarrow X_{n-1}$ ;
11    $n \leftarrow n + 1$ ;
12 até critério de parada;
13 retorna  $N$  últimas observações da cadeia;
```

Um segundo dispositivo bastante popular na literatura consiste do *Amostrador de Gibbs*, o qual explora a ideia de distribuições condicionais completas para a construção de um núcleo de transição adequado ao princípio do balanceamento detalhado. Se $X \in \mathbb{R}^{n_x}$ é distribuída segundo π , então o conjunto de distribuições condicionais completas $\{p^{(i)}\}_{i=1}^{n_x}$ é composto pela família de modelos probabilísticos associados a:

$$X|X_{-i}, \quad i = 1, 2, \dots, n_x, \quad (3.26)$$

onde $X_{-i} = (X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_{n_x})^t$. A estratégia consiste basicamente de gerar novas observações iterativamente a partir da família de condicionais completas. Assim como o Algoritmo MH, o amostrador de Gibbs satisfaz o princípio do balanceamento detalhado com respeito à medida π , o que garante uma amostragem consistente.

A metodologia MCMC conta ainda com vasto conjunto de tópicos importantes. Uma análise mais completa envolveria o estabelecimento de critérios de parada para os algoritmos assim como o desenvolvimento de um dispositivo de seleção mais adequado com respeito à amostra final. Poderíamos também focar em pontos mais específicos como na especificação de núcleos de transição mais eficazes no Algoritmo MH ou na determinação de amostragem por blocos no Amostrador de Gibbs. Entretanto, estas discussões fogem dos objetivos deste trabalho. Excelentes revisões do tópico podem ser obtidas nas obras de [Gilks, Richardson e Spiegelhalter \(1995\)](#) e [Gamerman e Lopes \(2006\)](#). Apesar de sua flexibilidade, os métodos MCMC podem ser desvantajosos no contexto da inferência sequencial. É possível obter versões robustas destes algoritmos neste cenário porém os mesmos serão frequentemente complexos em termos de tempo e/ou memória. Desta forma, é conveniente concentrar esforços na busca por metodologias alternativas também.

Algoritmo 3: Amostrador de Gibbs**Entrada:** Tamanho amostral (N), densidades inicial (ν) e condicionais completas

$$\{p^{(i)}\}_{i=1}^{n_x};$$

Saída: Amostra $\{X^{(i)}\}_{i=1}^N \sim \pi$;1 amostre $X_0 \sim \nu$;2 $n \leftarrow 1$;3 **repita**4 **para** $i \leftarrow 1$ até n_x **faça**5 amostre $X_{in} \sim p^{(i)}(x|X_{-in})$;6 $X_n \leftarrow (X_{1n}, X_{2n}, \dots, X_{n_x n})^t$;7 $n \leftarrow n + 1$;8 **até critério de parada**;9 **retorna** N últimas observações da cadeia;

3.4 Amostragem Sequencial por Importância

Aplicações envolvendo estruturas de dimensões elevadas podem ser muito complexas para serem tratadas por meio das abordagens de simulação direta. Uma forma alternativa de lidar com tais aplicações seria considerar o problema multidimensional como um processo sequencial cuja dimensão cresce gradativamente a cada instante de evolução. Consideremos por exemplo o problema de amostrar da distribuição π_n definida no espaço produto $(\Omega^n, \mathcal{B}(\Omega^n))$. Um dispositivo sequencial de amostragem poderia explorar a fatoração

$$\begin{aligned} \pi_n(dx_{1:n}) &= \pi_{n-1}(dx_{1:n-1})\pi_{n|n-1}(dx_n|x_{1:n-1}) \\ &= \pi_1(dx_1) \prod_{i=2}^n \pi_{i|i-1}(dx_i|x_{1:i-1}), \end{aligned} \quad (3.27)$$

com o intuito de operar sempre em dimensões menores: amostra-se $X_1^{(1:N)} \sim \pi_1$ no instante de partida e então, condicionado às realizações já obtidas, amostra-se $X_i^{(1:N)} \sim \pi_{i|i-1}$ nos instantes $i = 2, 3, \dots, n$. Além de facilitar o processo de amostragem, esta perspectiva oferece interpretações frutíferas por se adaptar perfeitamente a problemas reais naturalmente sequenciais, como aplicações envolvendo processos estocásticos por exemplo.

A metodologia sequencial mais comum consiste de um extensão do dispositivo de amostragem por importância onde as distribuições propostas assumem a forma fatorada

$$\begin{aligned} q_n(dx_{1:n}) &= q_{n-1}(dx_{1:n-1})q_{n|n-1}(dx_n|x_{1:n-1}) \\ &= q_1(dx_1) \prod_{i=2}^n q_{i|i-1}(dx_i|x_{1:i-1}), \end{aligned} \quad (3.28)$$

e como consequência imediata, temos também uma generalização da condição básica de regularidade

$$q_1 \ll \pi_1 \quad \text{e} \quad q_{i|i-1} \ll \pi_{i|i-1}, \quad i = 2, \dots, n. \quad (3.29)$$

A densidade associada aos pesos de importância $\omega_n : \Omega^n \rightarrow \mathbb{R}$ resulta da combinação das fatorações (3.27-3.28), um procedimento que pode ser realizado de forma recursiva dando origem a um dispositivo prático de atualização:

$$\begin{aligned}
\omega_n(x_{1:n}) &= \frac{d\pi_n}{dq_n}(x_{1:n}) \\
&= \frac{d(\pi_{n-1} \times \pi_{n|n-1})}{d(q_{n-1} \times q_{n|n-1})}(x_{1:n-1}, x_n) \\
&= \frac{d\pi_{n-1}}{dq_{n-1}}(x_{1:n-1}) \cdot \frac{d\pi_{n|n-1}}{dq_{n|n-1}}(x_n) \\
&= \omega_{n-1}(x_{1:n-1}) \cdot \frac{d\pi_{n|n-1}}{dq_{n|n-1}}(x_n). \tag{3.30}
\end{aligned}$$

Considerando a filtragem $\mathcal{F}_n^X := \sigma(X_k, k = 1, 2, \dots, n)$, podemos ainda explorar o esquema de atualização acima para deduzir

$$\begin{aligned}
\mathbb{E}_{q_{n+1}}[\omega_{n+1}(X_{1:n+1}) | \mathcal{F}_n^X] &= \mathbb{E}_{q_{n+1}} \left[\omega_n(X_{1:n}) \cdot \frac{d\pi_{n+1|n}}{dq_{n+1|n}}(X_{1:n}, X_{n+1}) \middle| \mathcal{F}_n^X \right] \\
&= \omega_n(X_{1:n}) \cdot \mathbb{E}_{q_{n+1}} \left[\frac{d\pi_{n+1|n}}{dq_{n+1|n}}(X_{1:n}, X_{n+1}) \middle| \mathcal{F}_n^X \right] \\
&= \omega_n(X_{1:n}) \cdot \mathbb{E}_{\pi_{n+1}}[1 | \mathcal{F}_n^X] \\
&= \omega_n(X_{1:n}). \tag{3.31}
\end{aligned}$$

Neste caso, se esta sequência de ponderações satisfizer a condição $\mathbb{E}_{q_n}[|\omega_n(X_{1:n})|] < +\infty$, $\forall n$, então tal sequência constituirá um Martingale. Mais ainda, se pudermos assegurar que $\sup_n \mathbb{E}_{q_n}[\omega_n(X_{1:n})] < +\infty$, então teremos a construção de uma aproximação minimamente consistente devido ao Teorema da Convergência de Martingales (SHIRYAEV, 1995). Desta forma, além de atender à condição básica de regularidade (3.29), a escolha das distribuições propostas deve ser feita de modo a evitar o possível crescimento indiscriminado das esperanças $\mathbb{E}_{q_n}[\omega(X_{1:n})]$.

Ao final do n -ésimo passo teremos uma sequência de realizações $\left\{ x_{1:n}^{(i)} \right\}_{i=1}^N$ multidimensionais as quais podem ser entendidas como uma amostra de N trajetórias realizadas ao longo do tempo. A cada uma destas trajetórias está associado um peso de importância $\omega_n^{(i)} = \omega_n(x_{1:n}^{(i)})$, culminando então em uma medida aleatória no espaço produto $(\Omega^n, \mathcal{B}(\Omega^n))$:

$$\hat{\pi}_n(dx_{1:n}) = \frac{1}{N} \sum_{i=1}^N \omega_n^{(i)} \cdot \delta_{x_{1:n}^{(i)}}(dx_{1:n}). \tag{3.32}$$

Assim como na versão convencional, podemos considerar um passo de normalização de pesos, sobretudo quando as densidades $\frac{d\pi_{k|k-1}}{dq_{k|k-1}}$ são conhecidas a menos de uma constante:

$$\tilde{\pi}_n(dx_{1:n}) = \sum_{i=1}^N \tilde{\omega}_n^{(i)} \cdot \delta_{x_{1:n}^{(i)}}(dx_{1:n}), \quad \tilde{\omega}_n^{(i)} = \omega_n^{(i)} \bigg/ \sum_{j=1}^N \omega_n^{(j)}, \quad i = 1, 2, \dots, N. \tag{3.33}$$

Algoritmo 4: Amostragem Sequencial por Importância**Entrada:** Tamanho amostral (N) e distribuições alvo π_n e proposta (q_n);**Saída:** Aproximação discreta $\hat{\pi}_n(dx_{1:n}) = \sum_{i=1}^N \tilde{\omega}_n^{(i)} \cdot \delta_{x_{1:n}}^{(i)}(dx_{1:n})$;

```

1 para  $i \leftarrow 1$  até  $N$  faça
2   Amostre  $X_1^{(i)} \sim q_1$ ;
3   Calcule  $\omega_1^{(i)} = \frac{d\pi_1}{dq_1}(X_1^{(i)})$ ;
4 Normalize  $\tilde{\omega}_1^{(i)} = \omega_1^{(i)} / \sum_{j=1}^N \omega_1^{(j)}$ ,  $i = 1, 2, \dots, N$ ;
5 para  $k \leftarrow 2$  até  $n$  faça
6   para  $i \leftarrow 1$  até  $N$  faça
7     Amostre  $X_k^{(i)} \sim q_{k|k-1}^{(i)}$ ;
8     Calcule  $\omega_k^{(i)} = \tilde{\omega}_{k-1}^{(i)} \cdot \frac{d\pi_{k|k-1}}{dq_{k|k-1}}(X_{1:k-1}^{(i)}, X_k^{(i)})$ ;
9     Normalize  $\tilde{\omega}_k^{(i)} = \omega_k^{(i)} / \sum_{j=1}^N \omega_k^{(j)}$ ,  $i = 1, 2, \dots, N$ ;
10 retorna  $\{X_{1:n}^{(i)}, \tilde{\omega}_n^{(i)}\}_{i=1}^N$ ;
```

A grande desvantagem do mecanismo sequencial de amostragem por importância é a perda de precisão à medida que a dimensão n aumenta. Este fenômeno pode ser verificado a partir das propriedades de Martingales herdadas pela sequência $\{\omega_n\}$:

$$\begin{aligned}
\text{Var}_{q_n}[\omega_n(X_{1:n})] &= \text{Var}_{q_n} \left[\mathbb{E}_{q_{n+1}} \left[\omega_{n+1}(X_{1:n+1}) \middle| \mathcal{F}_n^X \right] \right] \\
&= \text{Var}_{q_{n+1}}[\omega_{n+1}(X_{1:n+1})] - \mathbb{E}_{q_n} \left[\text{Var}_{q_{n+1}} \left[\omega_{n+1}(X_{1:n+1}) \middle| \mathcal{F}_n^X \right] \right] \\
&\leq \text{Var}_{q_{n+1}}[\omega_{n+1}(X_{1:n+1})].
\end{aligned} \tag{3.34}$$

Ao lidarmos com um conjunto de pesos com alto grau de dispersão, teremos necessariamente uma fração considerável de trajetórias associadas a massas probabilísticas muito inferiores às demais. Em termos práticos, estaremos explorando regiões de baixa importância do domínio, o que causa perdas importantes à estimativa final. Este problema pode ser contornado a partir da criação de um procedimento capaz de remover as trajetórias menos prováveis e substituí-las por cópias daquelas distribuídas em regiões mais favoráveis. Evidentemente, tal procedimento necessita garantir a preservação das propriedades matemáticas do algoritmo original. Um mecanismo desta natureza é denominado *reamostragem* e constitui uma parte indispensável dos algoritmos sequenciais de amostragem.

Reamostragem

O procedimento de reamostragem pode ser entendido como um passo de seleção responsável por multiplicar ou descartar cada uma das trajetórias $x_{1:n}^{(i)}$ de acordo com sua importância $\tilde{\omega}_n^{(i)}$. A execução deste passo garante que elementos ocupando regiões de baixa relevância sejam descartados da população em favor de aqueles mais adaptados. A ponderação $1/N$ é atribuída a cada um dos elementos selecionados de modo que a nova população possa ser interpretada como uma espécie de amostra aleatória da medida $\tilde{\pi}$. O mecanismo pode ser descrito sucintamente como uma transformação na medida aleatória normalizada $\tilde{\pi}_n$:

$$\tilde{\pi}_n = \sum_{i=1}^N \tilde{\omega}_n^{(i)} \delta_{x_{1:n}^{(i)}}(dx_{1:n}) \rightsquigarrow \tilde{\pi}_n = \frac{1}{N} \sum_{i=1}^N \xi_n^{(i)} \delta_{x_{1:n}^{(i)}}(dx_{1:n}), \quad (3.35)$$

onde

$$\xi_n^{(i)} \in \{0, 1, 2, \dots, N\},$$

corresponde ao número de replicações da trajetória $x_{1:n}^{(i)}$. Tal transformação deve, por questões de consistência, atender à seguinte restrição

$$\mathbb{E} \left[\xi_n^{(i)} \mid \mathcal{F}_n^X \right] = N \tilde{\omega}_n^{(i)}, \quad \forall i = 1, 2, \dots, N, \quad (3.36)$$

a qual é denominada *condição de não enviesamento*. Basicamente, ao garantir o cumprimento desta última propriedade, o dispositivo assegura a construção de estimativas não viciadas da medida $\tilde{\pi}$:

$$\begin{aligned} \mathbb{E} \left[(f, \tilde{\pi}_n) \mid \mathcal{F}_n^X \right] &= \frac{1}{N} \sum_{i=1}^N f \left(x_{1:n}^{(i)} \right) \mathbb{E}_{q_n} \left[\xi_n^{(i)} \mid \mathcal{F}_n^X \right] \\ &= \sum_{i=1}^N f \left(x_{1:n}^{(i)} \right) \tilde{\omega}_n^{(i)} \\ &= (f, \tilde{\pi}), \quad \forall f \in \mathcal{F}_b(\Omega^n). \end{aligned} \quad (3.37)$$

Mesmo com a restrição imposta pela condição de não enviesamento, ainda é possível construir vários dispositivos de amostragem capazes de gerar o conjunto $\left\{ \xi_n^{(i)} \right\}_{i=1}^N$. Dentre as possibilidades disponíveis na literatura, seria razoável assumir que as melhores estratégias são aquelas capazes de garantir a dispersão mínima dos elementos:

$$\xi_n^{(i)} \in \left[\left\lfloor N \tilde{\omega}_n^{(i)} \right\rfloor, \left\lceil N \tilde{\omega}_n^{(i)} \right\rceil \right]. \quad (3.38)$$

Neste caso teríamos um processo de reamostragem praticamente determinístico, garantindo o acréscimo de um montante mínimo de ruído às estimativas finais do algoritmo. No entanto, é bastante comum que a metodologia de reamostragem seja adotada de acordo com a facilidade de implementação. As principais estratégias encontradas na literatura são listadas a seguir; uma revisão mais ampla pode ser encontrada no trabalho de [Li, Bolic e Djuric \(2015\)](#).

Reamostragem Multinomial

A reamostragem multinomial é uma técnica amplamente explorada por ser bastante intuitiva e dispor de rotinas já implementadas em diversas linguagens. O procedimento consiste em amostrar um vetor $\xi_n = \left(\xi_n^{(1)}, \xi_n^{(2)}, \dots, \xi_n^{(N)} \right)$ com o auxílio da distribuição multinomial com parâmetros N e $W_n = \left(\tilde{\omega}_n^{(1)}, \tilde{\omega}_n^{(2)}, \dots, \tilde{\omega}_n^{(N)} \right)$. A condição de não enviesamento segue como uma consequência imediata das propriedades relativas à distribuição multinomial:

$$\xi_n | \mathcal{F}_n^X \sim \text{Mult}(N, W_n) \implies \mathbb{E} \left[\xi_n^{(i)} | \mathcal{F}_n^X \right] = N \tilde{\omega}^{(i)}. \quad (3.39)$$

Algoritmo 5: Reamostragem Multinomial

Entrada: Tamanho amostral (N) e probabilidades (W_n);

Saída: Números de replicações ξ_n ;

- 1 Amostre $\xi_n \sim \text{Mult}(N, W_n)$;
 - 2 **retorna** ξ_n ;
-

O número de replicações $\xi_n^{(i)}$ de cada elemento pode oscilar no intervalo $[0, N]$, o que indica a construção de estimativas com alto grau de variabilidade. Além disso, o dispositivo multinomial coloca muita pressão seletiva na população: indivíduos associados a pouca, mas não desprezível, massa probabilística tendem a ser descartados com facilidade. O fenômeno de enxugamento acelerado da população pode induzir efeitos muito negativos no processo inferencial, sobretudo no que diz respeito à estimativa das caudas da distribuição de interesse.

Reamostragem Residual

O procedimento de reamostragem residual escolhe deterministicamente um primeiro conjunto de réplicas $\hat{\xi}_n$ segundo o dispositivo prático:

$$\hat{\xi}_n^{(i)} = \lfloor N \cdot \tilde{\omega}_n^{(i)} \rfloor, \quad i = 1, 2, \dots, N. \quad (3.40)$$

O objetivo principal deste passo é impossibilitar a eliminação de indivíduos associados a ponderações superiores ou iguais a $1/N$, induzindo assim uma menor pressão seletiva na população. Uma vez que o passo determinístico seleciona um montante \hat{N} de réplicas inferior ao tamanho original N da população, faz-se necessário a implementação de um mecanismo de amostragem complementar

$$\bar{\xi}_n \sim \text{Mult}(\bar{N}, \bar{W}_n), \quad (3.41)$$

com parâmetros residuais

$$\bar{N} = N - \hat{N} \quad \text{e} \quad \bar{\omega}_n^{(i)} = \frac{N \cdot \tilde{\omega}_n^{(i)} - \lfloor N \tilde{\omega}_n^{(i)} \rfloor}{\bar{N}}, \quad i = 1, 2, \dots, N.$$

O vetor final de réplicas é obtido da soma dos elementos obtidos após os dois primeiros passos

$$\xi_n = \hat{\xi}_n + \bar{\xi}_n. \quad (3.42)$$

Algoritmo 6: Reamostragem Residual**Entrada:** Tamanho amostral (N) e probabilidades (W_n);**Saída:** Números de replicações ξ_n ;

- 1 $\hat{N} \leftarrow 0$;
- 2 **para** $i \leftarrow 1$ até N **faça**
- 3 $\hat{\xi}_n^{(i)} \leftarrow \lfloor N \cdot \tilde{\omega}_n^{(i)} \rfloor$;
- 4 $\hat{\omega}_n^{(i)} \leftarrow \tilde{\omega}_n^{(i)} - \hat{\xi}_n^{(i)} / N$;
- 5 $\hat{N} \leftarrow \hat{N} + \hat{\xi}_n^{(i)}$;
- 6 **para** $i \leftarrow 1$ até N **faça**
- 7 $\bar{\omega}_n^{(i)} \leftarrow \hat{\omega}_n^{(i)} \cdot N / (N - \hat{N})$;
- 8 $\bar{W}_n \leftarrow (\bar{\omega}_n^{(1)}, \bar{\omega}_n^{(2)}, \dots, \bar{\omega}_n^{(N)})$;
- 9 Amostre $\bar{\xi}_n \sim \text{Mult}(\bar{N}, \bar{W}_n)$;
- 10 $\xi_n \leftarrow \hat{\xi}_n + \bar{\xi}_n$;
- 11 **retorna** ξ_n ;

Este procedimento reduz consideravelmente a variabilidade do processo de seleção se comparado com a perspectiva multinomial, contudo o espalhamento na distribuição de réplicas ainda não é mínimo: $\xi_n^{(i)} \in \left[\lfloor N \cdot \tilde{\omega}_n^{(i)} \rfloor, \lfloor N \cdot \tilde{\omega}_n^{(i)} \rfloor + \bar{N} \right]$.

Reamostragem Sistemática

Amostragem sistemática refere-se a uma nomenclatura adotada no âmbito estatístico para o tradicional *Amostrador Estocástico Universal* (*Stochastic Universal Sampler* (SUS)) desenvolvido na literatura dos algoritmos genéticos (BAKER, 1987). O dispositivo parte da criação de duas instâncias: um conjunto de N estratos proporcionais às ponderações de cada indivíduo e um apontador u_0 uniformemente distribuído no intervalo $(0, 1/N]$. Os indivíduos são escolhidos então através da construção de N ponteiros auxiliares, partindo de u_0 e igualmente espaçados por $1/N$, cujas posições são comparadas com a amplitude de cada estrato.

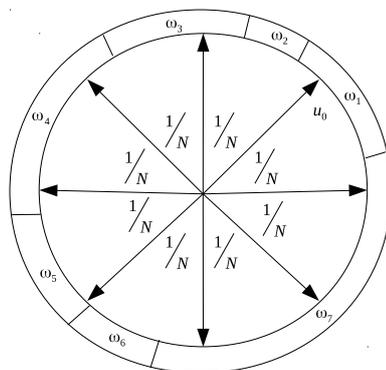


Figura 2 – Representação gráfica do esquema de reamostragem sistemática.

Algoritmo 7: Reamostragem Sistemática

Entrada: Tamanho amostral (N) e probabilidades (W_n);
Saída: Números de replicações ξ_n ;

- 1 Amostre $u \sim U(0, 1/N)$;
- 2 Soma $\leftarrow \tilde{\omega}_n^{(1)}$;
- 3 $\xi_n^{(j)} \leftarrow 0$;
- 4 Resíduo $\leftarrow N$;
- 5 $j \leftarrow 1$;
- 6 **repita**
- 7 **se** Soma $< u$ **então**
- 8 **repita**
- 9 $j \leftarrow j + 1$;
- 10 $\xi_n^{(j)} \leftarrow 0$;
- 11 Soma \leftarrow Soma $+ \tilde{\omega}_n^{(j)}$;
- 12 **até** Soma $\geq u$ ou $j = N$;
- 13 **senão**
- 14 $\xi_n^{(j)} \leftarrow \xi_n^{(j)} + 1$;
- 15 Resíduo \leftarrow Resíduo $- 1$;
- 16 $u \leftarrow u + 1/N$;
- 17 **até** $j = N$;
- 18 $\xi_n^{(N)} \leftarrow$ Resíduo;
- 19 **retorna** ξ_n ;

Como pontos a favor do dispositivo de amostragem sistemática destacam-se: o cumprimento da condição de não enviesamento, espalhamento mínimo do número de réplicas de cada elemento, complexidade de tempo linear com respeito ao tamanho amostral N e facilidade de implementação. Este mecanismo pode também ser explorado com o intuito de otimizar o desempenho da abordagem residual já que o passo complementar de seleção multinomial pode, sem restrições, ser substituído pela abordagem sistemática. Neste caso o laço principal continuaria oscilando entre 1 e N porém o espaçamento entre os apontadores seria alterado para $1/\bar{N}$. O algoritmo resultante é habitualmente denotado por *Amostragem Residual Sistemática* (LI; BOLIC; DJURIC, 2015). A dificuldade no trato matemático deste dispositivo é vista como o principal impedimento quanto à sua utilização (DOUC; CAPPÉ, 2005; BAIN; CRISAN, 2009); a ausência do formalismo quantitativo dificulta o estabelecimento de propriedades de convergência dos métodos sequenciais apoiados por esta abordagem.

Reamostragem por Ramificação

A ideia de amostragem por ramificação consiste em tratar o vetor ξ_n como um processo de ramificação N -dimensional. Basicamente, cada um dos elementos do vetor evolui segundo um modelo da forma:

$$\xi_n^{(i)} = \begin{cases} \lfloor N \cdot \tilde{\omega}_n^{(i)} \rfloor & \text{com probabilidade } 1 - \{N \cdot \tilde{\omega}_n^{(i)}\} \\ \lfloor N \cdot \tilde{\omega}_n^{(i)} \rfloor + 1 & \text{com probabilidade } \{N \cdot \tilde{\omega}_n^{(i)}\} \end{cases}. \quad (3.43)$$

onde

$$\{N \cdot \tilde{\omega}_n^{(i)}\} = N \cdot \tilde{\omega}_n^{(i)} - \lfloor N \cdot \tilde{\omega}_n^{(i)} \rfloor.$$

A dinâmica do processo é estabelecida de modo a garantir que o tamanho amostral mantenha-se fixo em N ; para tal basta estabelecer uma correlação negativa entre as N instâncias do processo.

Algoritmo 8: Reamostragem por Ramificação

Entrada: Tamanho amostral (N) e probabilidades (W_n);

Saída: Números de replicações ξ_n ;

```

1   $g \leftarrow N, \quad h \leftarrow N;$ 
2  para  $i \leftarrow 1$  até  $N - 1$  faça
3      Amostre  $u \sim U(0, 1);$ 
4      se  $\{N \cdot \tilde{\omega}_n^{(i)}\} + \{g - N \cdot \tilde{\omega}_n^{(i)}\} < 1$  então
5          se  $u < 1 - (\{N \cdot \tilde{\omega}_n^{(i)}\} / \{g\})$  então
6               $\xi_n^{(i)} \leftarrow \lfloor N \cdot \tilde{\omega}_n^{(i)} \rfloor;$ 
7          senão
8               $\xi_n^{(i)} \leftarrow \lfloor N \cdot \tilde{\omega}_n^{(i)} \rfloor + (h - \lfloor g \rfloor);$ 
9          senão
10             se  $u < 1 - (1 - \{N \cdot \tilde{\omega}_n^{(i)}\} / (1 - \{g\}))$  então
11                  $\xi_n^{(i)} \leftarrow \lfloor N \cdot \tilde{\omega}_n^{(i)} \rfloor + 1;$ 
12             senão
13                  $\xi_n^{(i)} \leftarrow \lfloor N \cdot \tilde{\omega}_n^{(i)} \rfloor + (h - \lfloor g \rfloor);$ 
14              $g \leftarrow g - N \cdot \tilde{\omega}_n^{(i)};$ 
15              $h \leftarrow h - \xi_n^{(i)};$ 
16  $\xi_n^{(N)} \leftarrow h;$ 
17 retorna  $\xi_n;$ 

```

As condições de não enviesamento e espalhamento mínimo seguem trivialmente da relação (3.43). Além disso, [Bain e Crisan \(2009\)](#) mostram que a forma quadrática associada à matriz de covariância do vetor ξ_n gerado pelo dispositivo de reamostragem por ramificação é limitada superiormente, um resultado importante para a determinação de taxas de convergência de amostradores sequenciais. Por esta razão, a amostragem por ramificação pode ser mais vantajosa que a sistemática.

Amostragem Sequencial por Importância com Reamostragem

O algoritmo de amostragem sequencial com reamostragem apresenta uma forma padrão independentemente da estratégia de reamostragem adotada. Além de acrescentar um novo passo correspondente à reamostragem, esta adaptação também revê a atualização dos pesos de importância já que o passo de seleção resulta em uma população de ponderação homogênea:

$$\tilde{\omega}_{n-1}^{(i)} = \frac{1}{N} \implies \omega_n^{(i)} \propto \frac{d\pi_{k|k-1}}{dq_{k|k-1}} \left(X_{1:k-1}^{(i)}, \tilde{X}_k^{(i)} \right), \quad \forall i = 1, 2, \dots, N.$$

Devido a questões computacionais, é interessante lidar com um algoritmo auxiliar que transforma o vetor ξ_n , obtido por qualquer método de reamostragem, em um vetor de índices cujos elementos referenciam os indivíduos selecionados. A rotina considerada aqui é identificada por `amostra_índices`; o dispositivo prático correspondente é apresentado na sequência.

Algoritmo 9: Amostragem Sequencial por Importância com Reamostragem

Entrada: Tamanho amostral (N) e distribuições alvo π_n e proposta (q_n);

Saída: Aproximação discreta $\hat{\pi}_n(dx_{1:n}) = \frac{1}{N} \sum_{i=1}^N \delta_{x_{1:n}}^{(i)}(dx_{1:n})$;

- 1 **para** $i \leftarrow 1$ até N **faça**
 - 2 Amostre $\tilde{X}_1^{(i)} \sim q_1$;
 - 3 Calcule $\omega_1^{(i)} = \frac{d\pi_1}{dq_1} \left(\tilde{X}_1^{(i)} \right)$;
 - 4 Normalize $\tilde{\omega}_1^{(i)} = \omega_1^{(i)} / \sum_{j=1}^N \omega_1^{(j)}$, $i = 1, 2, \dots, N$;
 - 5 $\{j_1, j_2, \dots, j_N\} \leftarrow \text{amostra_índices} \left(N, \{\tilde{\omega}_1^{(i)}\}_{i=1}^N \right)$;
 - 6 **para** $i \leftarrow 1$ até N **faça**
 - 7 $X_1^{(i)} \leftarrow \tilde{X}_1^{(j_i)}$, $i = 1, 2, \dots, N$;
 - 8 **para** $k \leftarrow 2$ até n **faça**
 - 9 **para** $i \leftarrow 1$ até N **faça**
 - 10 Amostre $\tilde{X}_k^{(i)} \sim q_{k|k-1}^{(i)}$;
 - 11 Calcule $\omega_k^{(i)} = \frac{d\pi_{k|k-1}}{dq_{k|k-1}} \left(X_{1:k-1}^{(i)}, \tilde{X}_k^{(i)} \right)$;
 - 12 Normalize $\tilde{\omega}_k^{(i)} = \omega_k^{(i)} / \sum_{j=1}^N \omega_k^{(j)}$, $i = 1, 2, \dots, N$;
 - 13 $\{j_1, j_2, \dots, j_N\} \leftarrow \text{amostra_índices} \left(N, \{\tilde{\omega}_k^{(i)}\}_{i=1}^N \right)$;
 - 14 **para** $i \leftarrow 1$ até N **faça**
 - 15 $X_{1:k}^{(i)} \leftarrow \left(X_{1:k-1}^{(j_i)}, \tilde{X}_k^{(j_i)} \right)$;
 - 16 **retorna** $\left\{ X_{1:n}^{(i)}, \frac{1}{N} \right\}_{i=1}^N$;
-

Algoritmo 10: Amostra Índices**Entrada:** Tamanho amostral (N) e probabilidades (W_n);**Saída:** Vetor de índices de reamostragem (J_n);

```

1  $j \leftarrow 0, \quad n \leftarrow 0;$ 
2  $\xi \leftarrow \text{Reamostra}(N, W_n);$  //Multinomial, Residual, Sistemática ou Ramificação
3 repita
4    $j \leftarrow j + 1;$ 
5   se  $\xi_n^{(j)} > 0$  então
6     repita
7        $n \leftarrow n + 1;$ 
8        $J_n^{(j)} \leftarrow j;$ 
9        $\xi_n^{(j)} \leftarrow \xi_n^{(j)} - 1;$ 
10    até  $\xi_n^{(j)} = 0;$ 
11 até  $n = N;$ 
12 retorna  $J_n;$ 

```

3.5 Filtros de Partículas

Filtros de Partículas são basicamente especializações dos algoritmos de amostragem sequencial para o tratamento do problema da filtragem. A vasta classe de problemas tratada sob esta perspectiva faz com que boa parte das aplicações da amostragem sequencial sejam absorvidas neste contexto. Desta forma, a não diferenciação das duas tratativas é um procedimento habitual na literatura.

3.5.1 Filtragem Pura

Em se tratando do problema da filtragem pura, conforme discutido na seção 2.4, estaremos interessados na construção sequencial de aproximações para a densidade

$$r^{(t)}(x_{0:t}|y_{0:t}) = r^{(t-1)}(x_{0:t-1}|y_{0:t-1})L^{(t-1)}(y_{0:t-1})\frac{g(y_t|x_t)q(x_t|x_{t-1})}{L^{(t)}(y_{0:t})}$$

$$\propto r^{(t-1)}(x_{0:t-1}|y_{0:t-1})g(y_t|x_t)q(x_t|x_{t-1}),$$

cuja dedução segue diretamente da estrutura Markoviana do problema. Assumindo que a quantidade $f^{(t-1)}(x_{0:t-1}|y_{0:t-1})$ já é conhecida, e denotando-a por ω_{t-1} , podemos prosseguir de forma análoga à metodologia discutida na seção anterior. Ou seja, basta definirmos uma distribuição proposta $q_{t|t-1}$, absolutamente contínua com respeito a $g(y_t|x_t)q(x_t|x_{t-1})$, e aplicar um passo de amostragem por importância. Diferentemente do caso mais geral, a densidade alvo a cada novo instante depende única e exclusivamente dos estados obtidos no passo imediatamente anterior, o que configura uma simplificação substancial do problema. O esquema de atualização poderá então ser sucintamente descrito por:

$$\omega_t \propto \omega_{t-1} \frac{g(y_t|x_t)q(x_t|x_{t-1})}{q_{t|t-1}}.$$

Como o dispositivo de filtragem faz uso de um enxame de partículas, correspondentes a uma amostra da distribuição proposta,

$$x_t^{(i)} \sim q_{t|t-1}^{(i)}, \quad i = 1, 2, \dots, N,$$

chega-se à conotação usual *Filtro de Partículas*. No caso específico da filtragem pura, a diferenciação entre os dois filtros de partículas quaisquer é dada basicamente pela escolha da distribuição proposta. Dentre as abordagens mais frequentemente encontradas na literatura destacam-se os filtros *Bootstrap* e *Auxiliar*, cujas especificações são apresentadas brevemente a seguir.

Filtro Bootstrap

É natural buscarmos distribuições propostas que garantam maior precisão às estimativas de filtragem. Seguindo a relação (3.16), podemos definir, ao menos formalmente, uma distribuição proposta ótima através da relação:

$$\begin{aligned} q_{t|t-1} &\propto g(y_t|x_t)q(x_t|x_{t-1}) \\ \therefore q_{t|t-1} &\propto p(x_t|x_{t-1}, y_t) = \frac{g(y_t|x_t)q(x_t|x_{t-1})}{p(y_t|x_{t-1})}. \end{aligned}$$

Desta forma, nestes casos ideais a propagação se resumirá a:

$$\omega_t \propto \omega_{t-1}p(y_t|x_{t-1}).$$

Em situações onde tal distribuição é desconhecida, ou analiticamente intratável, é comum eliminarmos a dependência da observação y_t de modo a construir um esquema de propagação cega:

$$q_{t|t-1} = q(x_t|x_{t-1}) \rightsquigarrow \omega_t \propto \omega_{t-1}g(y_t|x_t).$$

Tal escolha corresponde à proposta pioneira sugerida por [Gordon, Salmond e Smith \(1993\)](#) e o esquema de filtragem obtido a partir daí é popularmente denominado *Filtro Bootstrap*. O nome deve-se à associação do método à realização de sucessivos passos de reamostragem, configurando uma proposta similar à do tradicional método *Bootstrap*. A utilização deste passo garante a eliminação de indivíduos menos importantes do espaço amostral e assim garante mais foco à metodologia; um conceito rigorosamente análogo ao discutido para os amostradores sequenciais em um contexto geral. Uma peculiaridade da metodologia é que apenas os estados mais atuais precisam ser reamostrados já que estes constituíram a única informação relevante para o passo seguinte. Este detalhe assegura a implementação de um algoritmo ainda mais simples que o apresentado no contexto geral da amostragem sequencial por importância.

Algoritmo 11: Filtro *Bootstrap***Entrada:** Tamanho amostral (N), observações ($y_{1:n}$);**Saída:** Aproximação discreta $\hat{\pi}_n(dx_{1:n}) = \frac{1}{N} \sum_{i=1}^N \delta_{(x_{1:n}^{(i)})}(dx_{1:n})$;

```

1 para  $i \leftarrow 1$  até  $N$  faça
2   └ Amostre  $X_0^{(i)} \sim q_0$ ;
3 para  $j \leftarrow 1$  até  $n$  faça
4   └ para  $i \leftarrow 1$  até  $N$  faça
5     └ Amostre  $\tilde{X}_j^{(i)} \sim q(\cdot | X_{j-1}^{(i)})$ ;
6     └  $\omega_j^{(i)} \leftarrow g(y_t | \tilde{X}_j^{(i)})$ ;
7     └ Normalize  $\tilde{\omega}_j^{(i)} \leftarrow \omega_j^{(i)} / \sum_{k=1}^N \omega_j^{(k)} \quad i = 1, 2, \dots, N$ ;
8     └  $\{j_1, j_2, \dots, j_N\} \leftarrow \text{amostra\_índices}(N, \{\tilde{\omega}_j^{(i)}\}_{i=1}^N)$ ;
9     └ para  $i \leftarrow 1$  até  $N$  faça
10    └  $X_k^{(i)} \leftarrow \tilde{X}_k^{(j_i)}$ ;
11 retorna  $\{X_{1:n}^{(i)}, \frac{1}{N}\}$ ;
```

A realização excessiva dos passos de reamostragem pode induzir um enxugamento indesejado no conjunto de partículas e, conseqüentemente, inflar a variabilidade das aproximações. Neste caso, é possível definir convenientemente uma agenda de reamostragem que nos permita fazer uso deste passo apenas em instantes apropriados, os quais podem ser estabelecidos com o auxílio dos tamanhos efetivos amostrais:

$$TEA_j = \frac{\left(\sum_{i=1}^N \omega_j^{(i)}\right)^2}{\sum_{i=1}^N \left(\omega_j^{(i)}\right)^2} = \frac{1}{\sum_{i=1}^N \left(\tilde{\omega}_j^{(i)}\right)^2}, \quad j = 1, 2, \dots, n,$$

sendo $\{\tilde{\omega}_j^{(i)}\}$ a seqüência de pesos normalizados obtidos após a iteração j . Um passo de reamostragem é realizado no instante j apenas se (ARULAMPALAM *et al.*, 2002)

$$TEA_j < \frac{1}{2}N,$$

isto é, se a população ponderada condensar o mesmo montante de informação que uma amostra aleatória com metade do tamanho. Ao evitar a repetição desmoderada dos passos de seleção, não só obtemos um maior controle acerca da variabilidade das estimativas, mas também reduzimos a complexidade do algoritmo. Este tipo de vantagem é importante sobretudo em aplicações envolvendo séries longas de observações, situações nas quais o tempo de execução do filtro é naturalmente mais elevado.

Algoritmo 12: Filtro *Bootstrap* com Reamostragem Adaptativa**Entrada:** Tamanho amostral (N), observações ($y_{1:n}$);**Saída:** Aproximação discreta $\hat{\pi}_n(dx_{1:n}) = \sum_{i=1}^N \tilde{\omega}_n^{(i)} \delta_{(x_{1:n}^{(i)})}$;

```

1 para  $i \leftarrow 1$  até  $N$  faça
2   Amostre  $X_0^{(i)} \sim q_0$ ;
3    $\tilde{\omega}_0^{(i)} \leftarrow \frac{1}{N}$ ;
4 para  $j \leftarrow 1$  até  $n$  faça
5   para  $i \leftarrow 1$  até  $N$  faça
6     Amostre  $\tilde{X}_j^{(i)} \sim q(\cdot | X_{j-1}^{(i)})$ ;
7      $\omega_j^{(i)} \leftarrow \tilde{\omega}_{j-1}^{(i)} \cdot g(y_j | \tilde{X}_j^{(i)})$ ;
8     Normalize  $\tilde{\omega}_j^{(i)} \leftarrow \omega_j^{(i)} / \sum_{k=1}^N \omega_j^{(k)}$   $i = 1, 2, \dots, N$ ;
9      $TEA_j \leftarrow 1 / \sum_{i=1}^N (\tilde{\omega}_j^{(i)})^2$ ;
10    se  $TEA_j < N/2$  então
11       $\{j_1, j_2, \dots, j_N\} \leftarrow \text{amostra\_índices}(N, \{\tilde{\omega}_k^{(i)}\}_{i=1}^N)$ ;
12      para  $i \leftarrow 1$  até  $N$  faça
13         $X_k^{(i)} \leftarrow \tilde{X}_k^{(j_i)}$ ;
14         $\tilde{\omega}_j^{(i)} \leftarrow \frac{1}{N}$ ;
15 retorna  $\{X_{1:n}^{(i)}, \tilde{\omega}_n^{(i)}\}$ ;

```

Filtro Auxiliar

A distribuição proposta ótima permite que a informação mais atual proveniente das observações seja incorporada na dinâmica dos estados latentes favorecendo a previsão de características futuras da série. Este aspecto é importante pois nos permite lidar com uma aproximação melhor adaptada às observações, culminando em uma família mais equilibrada de pesos de importância. Neste caso, é importante tentar incorporar esta informação ao esquema sequencial de amostragem mesmo que a proposta ideal não seja acessível. Pitt e Shephard (1999) sugeriram uma aproximação da forma:

$$p(x_t | x_{t-1}, y_t) \approx \sum_{i=1}^N \underbrace{g(y_t | F(x_{t-1}^{(i)}))}_{\omega_{t,aux}^{(i)}} \cdot q(x_t | x_{t-1}),$$

onde F é uma função qualquer capaz de expressar alguma tendência futura da sequência latente, usualmente média ou mediana da dinâmica dos estados. Um novo indivíduo pode ser gerado desta estrutura selecionando inicialmente um índice auxiliar j (daí a nomenclatura) a partir das ponderações auxiliares

$$\omega_{t,aux}^{(i)} = g(y_t | F(x_{t-1}^{(i)})), \quad i = 1, 2, \dots, N,$$

e em seguida amostrando da distribuição $q(x_t|x_t^{(i)})$. Desta forma, teremos ao final da t -ésima iteração, uma amostra $\{x_t^{(i)}\}$ ponderada pelas quantidades:

$$\omega_t^{(i)} = \omega_t^{(i)} \frac{g(y_t|x_t^{(i)})q(x_t^{(i)}|x_{t-1}^{(i)})}{\omega_{t,aux}^{(i)}q(x_t^{(i)}|x_{t-1}^{(i)})} = \omega_t^{(i)} \frac{g(y_t|x_t^{(i)})}{\omega_{t,aux}^{(i)}}.$$

Uma população associada a ponderações homogêneas pode ser obtida realizando um novo passo de reamostragem com respeito a $\{\omega_t^{(i)}\}$. Este passo tem como objetivo identificar os indivíduos mais adaptados da amostra assim como no caso do Filtro *Bootstrap*. Desta forma, a abordagem seria associada a dois passos de reamostragem. Outra forma de chegar ao mesmo objetivo, a partir da realização de um passo de reamostragem apenas, é através da inclusão dos pesos mais atuais no processo de seleção dos índices auxiliares:

$$p(x_t|x_{t-1}, y_t) \approx \sum_{i=1}^N \omega_{t-1}^{(i)} \cdot \omega_{t,aux}^{(i)} \cdot q(x_t|x_{t-1}).$$

Assim, é possível selecionar uma família de índices auxiliares capaz de simultaneamente determinar os indivíduos mais adaptados e capturar tendências futuras das observações.

Algoritmo 13: Filtro Auxiliar

Entrada: Tamanho amostral (N), observações ($y_{1:n}$), função auxiliar de previsão (f);

Saída: Aproximação discreta $\hat{\pi}_n(dx_{1:n}) = \sum_{i=1}^N \tilde{\omega}_n^{(i)} \delta_{(x_{1:n}^{(i)})}(dx_{1:n})$;

1 **para** $i \leftarrow 1$ até N **faça**

2 Amostre $X_0^{(i)} \sim q_0$;

3 $\tilde{\omega}_0^{(i)} \leftarrow \frac{1}{N}$;

4 **para** $j \leftarrow 1$ até n **faça**

5 $\{k_1, k_2, \dots, k_N\} \leftarrow \text{amostra_índices} \left(N, \left\{ \tilde{\omega}_{j-1}^{(i)} \cdot g \left(y_j | f(X_{j-1}^{(i)}) \right) \right\}_{i=1}^N \right)$;

6 **para** $i \leftarrow 1$ até N **faça**

7 Amostre $\tilde{X}_j^{(i)} \sim q(\cdot | X_{j-1}^{(k_i)})$;

8 $\omega_j^{(i)} \leftarrow \frac{g(y_j | \tilde{X}_j^{(i)})}{g(y_j | F(x_{j-1}^{(k_i)}))}$;

9 Normalize $\tilde{\omega}_j^{(i)} \leftarrow \omega_j^{(i)} / \sum_{k=1}^N \omega_j^{(k)} \quad i = 1, 2, \dots, N$;

10 **retorna** $\{X_{1:n}^{(i)}, \tilde{\omega}_n^{(i)}\}$;

Verossimilhança

As medidas discretas construídas com a ajuda dos filtros de partículas podem ser reutilizadas de modo a gerar aproximações para a função de verossimilhança:

$$L^{(t)}(y_{0:t}) = \int \int \cdots \int v(x_0) g(y_0|x_0) \left\{ \prod_{k=1}^t g(y_k|x_k) q(x_k|x_{k-1}) \right\} dx_0 dx_1 \dots dx_t.$$

Esta aproximação nos habilita a lidar com aplicações envolvendo distribuições *a posteriori* no contexto inferencial Bayesiano. Assim como as demais construções realizadas até então, este passo pode ser executado sequencialmente considerando as estruturas parciais:

$$L^{(t)}(y_t|y_{0:t-1}).$$

Para tal, basta explorar a relação

$$\begin{aligned} L^{(t)}(y_t|y_{0:t-1}) &= \int L^{(t)}(y_t, x_t|y_{0:t-1}) dx_t \\ &= \int g(y_t, |x_t) p(x_t|y_{0:t-1}) dx_t, \end{aligned}$$

a qual nos assegura uma aproximação de Monte Carlo imediata para $L^{(t)}(y_t|y_{0:t-1})$, desde que tenhamos acesso a uma amostra da sequência preditiva $x_t|y_{0:t-1}$. Suponha que no instante t , tenhamos acesso à amostra ponderada

$$\left\{ x_{t-1}^{(i)}, \tilde{\omega}^{(i)} \right\}.$$

Neste caso, podemos considerar o passo de integração:

$$q(x_t|y_{0:t-1}) = \int q(x_t, x_{0:t-1}|y_{0:t-1}) dx_{0:t-1} \quad (3.44)$$

$$= \int q(x_t|x_{t-1}) f^{(t-1)}(x_{t-1}|y_{0:t-1}) dx_{0:t-1} \approx \sum_{i=1}^N q(x_t|x_{t-1}^{(i)}) \tilde{\omega}_t^{(i)}. \quad (3.45)$$

Dada uma amostra $\{\check{x}\}$ obtida da mistura (3.45), podemos obter a aproximação de Monte Carlo:

$$L^{(t)}(y_t|y_{0:t-1}) \approx \frac{1}{N} \sum_{i=1}^N g(y_t|\check{x}_t^{(i)}). \quad (3.46)$$

A aproximação é finalizada a partir da combinação das decomposições parciais:

$$L^{(t)}(y_{0:t}) = \frac{1}{N^{t+1}} \prod_{t=1}^N \left[\sum_{i=0}^N g(y_t|\check{x}_t^{(i)}) \right], \quad (3.47)$$

sendo a parcela referente ao instante $t = 0$ obtida diretamente da distribuição inicial do problema.

3.5.2 Suavização

Filtros de partículas também podem ser explorados para a construção de aproximações para o problema da suavização, isto é, para a estimação do modelo probabilístico de $x_k|y_{0:t}$, $k = 0, 1, 2, \dots, t$. Este tipo de aproximação demanda um esforço computacional muito intensivo se comparado ao problema da filtragem; tal esforço pode ser entendido rapidamente a partir de uma simples análise das relações (2.19-2.20). Contudo, esta metodologia é extremamente importante no contexto dos métodos sequenciais, sobretudo em se tratando da família de procedimentos do tipo MCMC.

A estruturação matemática dos modelos Markovianos ocultos nos permite lidar com a seguinte decomposição da densidade $r(x_{0:t}|y_{0:t})$ suavizada:

$$p(x_{0:t}|y_{0:t}) = r(x_t|y_{0:t}) \prod_{j=0}^{t-1} r(x_j|x_{j+1}, y_{0:j}).$$

Nota-se que $r(x_j|x_{j+1}, y_{0:j})$ define uma densidade de transição Markoviana, retroativa no tempo, a qual pode ser reescrita como

$$r(x_j|x_{j+1}, y_{0:j}) = \frac{q(x_{j+1}|x_j)r(x_j|y_{0:j})}{r(x_{j+1}|y_{0:j})}. \quad (3.48)$$

Combinando as duas expressões acima, podemos obter uma esquema de recursão retroativo para $\{r(x_j|y_{0:t})\}_{j=0}^t$:

$$\begin{aligned} r(x_j|y_{0:t}) &= \int r(x_j|x_{j+1}, y_{0:t})r(x_{j+1}|y_{0:t})dx_{j+1} \\ &= r(x_j|y_{0:t}) \int \frac{q(x_{j+1}|x_j)r(x_{j+1}|y_{0:t})}{r(x_{j+1}|y_{0:j})}dx_{j+1}. \end{aligned}$$

Tal relação explora uma sequência de densidades $\{r(x_j|y_{0:j})\}_{j=0}^t$, filtradas sucessivamente, na construção de uma recursão retroativa para $\{r(x_j|y_{0:t})\}_{j=0}^t$.

Em situações mais simples, como as dos modelos lineares Gaussianos, é possível tratar tais recursões analiticamente de modo a construir um método objetivo para o cálculo da distribuição de suavização. No caso geral, podemos contornar a intratabilidade da integral acima com o auxílio de aproximações provenientes de filtros de partículas. Basicamente, temos que dinâmica desenvolvida acima sugere a possibilidade de aproximar o modelo de suavização $f(x_{0:t}|y_{0:t})$ executando primeiramente um filtro de partículas de modo a obter uma sequência de aproximações

$$\hat{\pi}_j(dx_j|y_{0:j}) = \sum_{i=1}^N \tilde{\omega}_j^{(i)} \delta_{x_j^{(i)}}(dx_j), \quad j = 0, 1, \dots, t$$

para o problema da filtragem. Em seguida, amostramos $X_T \sim \hat{\pi}(dx_T|y_{0:T})$ e, para $j = T - 1, T - 2, \dots, 2, 1, 0$, amostramos $X_j \sim \hat{\pi}(dx_j|x_{j+1}, y_{0:n})$, sendo tal distribuição obtida de uma

aproximação de (3.48):

$$\begin{aligned} \hat{\pi}(dx_j|x_{j+1}, y_{0:j}) &\approx \frac{q(x_{j+1}|x_j)r(x_j|y_{0:j})}{f(x_{j+1}|y_{0:j})} \\ &\approx \frac{\sum_{i=1}^N \tilde{\omega}_j^{(i)} q(x_{j+1}|x_j^{(i)}) \delta_{x_j^{(i)}}(dx_j)}{\sum_{i=1}^N \tilde{\omega}_j^{(i)} q(x_{j+1}|x_j^{(i)})}. \end{aligned} \quad (3.49)$$

Desta forma, o método se resume a amostrar, a cada instante j , indivíduos da medida aleatória

$$\left\{ x_j^{(i)}, \check{\omega}_j^{(i)} \right\}_{i=1}^N,$$

onde

$$\check{\omega}_j^{(i)} = \frac{\tilde{\omega}_j^{(i)} q(x_{j+1}|x_j^{(i)})}{\sum_{i=1}^N \tilde{\omega}_j^{(i)} q(x_{j+1}|x_j^{(i)})},$$

$\{x_j^{(i)}\}$ é uma sequência obtida do passo de filtragem sucessiva e x_{j+1} é o ponto obtido no passo imediatamente anterior de amostragem retroativa. Uma discussão mais ampla a respeito de métodos de suavização baseados em filtros de partículas pode ser encontrada na obra de [Kitagawa \(1996\)](#).

3.5.3 Aprendizagem de Parâmetros

Estratégias sequenciais podem ser exploradas em duas frentes a fim de atacar o problema da estimação de parâmetros estáticos desconhecidos. A primeira delas consiste em acrescentar esta nova grandeza ao conjunto de estados latentes criando o conceito de vetor aumentado de estados

$$\mathbf{z}_n = \begin{bmatrix} x_n \\ \theta \end{bmatrix}. \quad (3.50)$$

Esta perspectiva incorpora diretamente os parâmetros desconhecidos na estrutura dos filtros de partículas de modo a oferecer um dispositivo de inferência em tempo real:

$$\text{Predição: } p(z_n | y_{1:n-1}) = p(x_n, \theta | y_{1:n-1}) = \int q(x_n | x_{n-1}, \theta) r(x_{n-1}, \theta | y_{1:n-1}) dx_{n-1}; \quad (3.51)$$

$$\text{Filtragem: } r(z_n | y_{1:n}) = r(x_n, \theta | y_{1:n}) \propto g(y_n | x_n, \theta) p(x_n, \theta | y_{1:n-1}); \quad (3.52)$$

$$\text{Evidência: } p(y_n | y_{1:n-1}) = \int g(y_n | z_n) p(z_n | y_{1:n-1}) dz_n. \quad (3.53)$$

Os principais representantes desta classe de métodos são os filtros de Kitagawa (KITAGAWA, 1998), Liu e West (LIU; WEST, 2001) e Storvik/*Parameter Learning* (STORVIK, 2002; CARVALHO *et al.*, 2010). A segunda abordagem utiliza a nuvem de pontos resultante dos algoritmos de filtragem para aproximar a função de verossimilhança das observações, relação (3.47), possibilitando o desenvolvimento de variações dos algoritmos MCMC. Os métodos PMCMC (ANDRIEU; DOUCET; HOLENSTEIN, 2010) e SMC² (CHOPIN; JACOB; PAPASPILIOPOULOS, 2013) podem ser apontados como os algoritmos mais relevantes segundo esta interpretação. Em casos mais simples, onde a densidade de suavização pode ser acessada de forma eficiente e o modelo permita a construção de um esquema de amostragem direta para os parâmetros estáticos, podemos ainda desenvolver variações do amostrador de Gibbs (SINGH, 2015). Tais estratégias são apresentadas em mais detalhes a seguir.

Filtro de Kitagawa

O filtro de Kitagawa, também denominado *Filtro de Partículas Auto-Organizativo*, tem muita relevância histórica por ser responsável pela introdução do conceito de vetor aumentado de estados. A estratégia consiste basicamente em atribuir uma dinâmica artificial, usualmente da forma

$$\theta_t = \theta_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \tau^2 \cdot \mathbf{I}_{n_\theta}),$$

que permita que a população estática obtida da distribuição *a priori* não se degenere ao longo das iterações do filtro. O ruído Gaussiano sugerido na versão mais usual do método pode gerar discrepâncias com respeito ao suporte do vetor de parâmetros. Em muitas situações, podemos estar interessados na estimativa de parâmetros cujo domínio de definição é dado por um intervalo da forma $[0, +\infty)$ ou $[a, b]$, $a, b \in \mathbb{R}$, $a < b$. Nestes casos, podemos realizar uma transformação prévia nos parâmetros de modo a expandir seu suporte para a reta e evitar problemas de inconsistência com relação ao ruído artificial. As transformações mais usuais são:

- $\theta \in [0, +\infty)$. O parâmetro θ pode ser expandido para a reta a partir de uma transformação logarítmica $\tilde{\theta} = \log(\theta)$. Após a realização do passo de rejuvenescimento artificial, os parâmetros devem ser retornados para a escala real através da transformação inversa $\theta = e^{\tilde{\theta}}$.
- $\theta \in (-\infty, 0]$. Tratamento análogo ao caso anterior: $\tilde{\theta} = -\log(\theta)$ e $\theta = e^{-\tilde{\theta}}$.
- $\theta \in [a, b]$. Caso o intervalo de definição seja limitado, é possível realizar uma transformação do tipo logito generalizado $\tilde{\theta} = \log(\theta - a) / \log(b - \theta)$. A transformação inversa é dada por $\theta = a + be^{\tilde{\theta}} / (1 + \tilde{\theta})$.

Ao indexar o vetor de parâmetros pelo parâmetro t , temos uma interpretação de evolução da distribuição *a posteriori* de θ ao longo do tempo, não necessariamente assumimos que tal estrutura fixa seja caracterizada por uma dinâmica de evolução. O vetor desconhecido θ pode, a partir desta construção, ser associado a um modelo finito de misturas:

$$p_{t-1}(\theta) := p(\theta | x_{0:t-1}, y_{0:t-1}) \approx \sum_{i=1}^N \tilde{\omega}_{t-1}^{(i)} N(\theta; \theta^{(i)}, \tau^2 \cdot \mathbf{I}_{n_\theta}),$$

onde $\{\theta^{(i)}, \tilde{\omega}_{t-1}^{(i)}\}$ correspondem aos indivíduos da população estática e suas respectivas ponderações de importância. Esta perspectiva nos permite estender as recursões de filtragem de modo a acomodar o vetor estático:

$$\begin{aligned} f(z_{0:t} | y_{0:t}) &= f(x_{0:t}, \theta | y_{0:t}) \\ &\propto f(y_t | x_t, \theta) p(x_{0:t}, \theta, y_{0:t-1}) \\ &\propto g(y_t | x_t, \theta) q(x_t | x_{0:t-1}, \theta) p(\theta | x_{0:t-1}, y_{0:t-1}) f(x_{0:t-1} | y_{0:t-1}). \end{aligned}$$

Assumindo que uma aproximação $\{\tilde{\omega}_{t-1}^{(i)}\}$ para a densidade $f(x_{0:t-1} | y_{0:t-1})$ foi construída no instante $t-1$, temos que

$$\begin{aligned} f(x_{0:t-1} | y_{0:t-1}) &= \int f(z_{0:t-1} | y_{0:t-1}) d\theta \\ &\approx \int \sum_{i=1}^N \tilde{\omega}_{t-1}^{(i)} \delta_{x_{0:t-1}, \theta^{(i)}}(x_{0:t-1}, \theta) d\theta \\ &= \sum_{i=1}^N \tilde{\omega}_{t-1}^{(i)} \delta_{x_{0:t-1}}^{(i)}(x_{0:t-1}). \end{aligned}$$

Ou seja, as mesmas ponderações podem ser utilizadas como uma aproximação para a densidade marginal $f(x_{0:t-1} | y_{0:t-1})$. Desta forma, chegamos ao seguinte dispositivo de atualização:

$$\begin{aligned} f(z_{0:t} | y_{0:t}) &= \omega_t \\ &= \tilde{\omega}_{t-1} \cdot \frac{g(y_t | x_t) q(x_t | x_{t-1}, \theta) p(\theta | x_{0:t-1}, y_{0:t-1})}{q_{t|t-1}(x_t | x_{t-1}, y_t, \theta) p_{t-1}(\theta)} \\ &\approx \tilde{\omega}_{t-1} \cdot \frac{g(y_t | x_t) q(x_t | x_{t-1}, \theta)}{q_{t|t-1}(x_t | x_{t-1}, y_t, \theta)}, \end{aligned} \tag{3.54}$$

assumindo que a mistura finita $p_{t-1}(\boldsymbol{\theta})$ é uma aproximação minimamente razoável para a densidade *a posteriori* $p(\boldsymbol{\theta}|x_{0:t-1}, y_{0:t-1})$.

Algoritmo 14: Filtro de Kitagawa

Entrada: Tamanho amostral (N), observações ($y_{1:n}$), distribuição *a priori* (p) e variância (σ^2);

Saída: Aproximação discreta $\hat{\pi}_n(dz_{1:n}) = \frac{1}{N} \sum_{i=1}^N \delta_{(z_{1:n}^{(i)})}(dz_{1:n})$;

```

1 para  $i \leftarrow 1$  até  $N$  faça
2   Amostre  $X_0^{(i)} \sim q_0$ ;
3   Amostre  $\boldsymbol{\theta}^{(i)} \sim p$ ;
4 para  $j \leftarrow 1$  até  $n$  faça
5   para  $i \leftarrow 1$  até  $N$  faça
6     Amostre  $\tilde{X}_j^{(i)} \sim q(\cdot | X_{j-1}^{(i)}, \boldsymbol{\theta}^{(i)})$ ;
7     Amostre  $\tilde{\boldsymbol{\theta}}^{(i)} \sim N(\boldsymbol{\theta}^{(i-1)}, \sigma^2)$ ;
8      $\boldsymbol{\omega}_j^{(i)} \leftarrow g(y_j | \tilde{X}_j^{(i)}, \tilde{\boldsymbol{\theta}}^{(i)})$ ;
9   Normalize  $\tilde{\boldsymbol{\omega}}_j^{(i)} \leftarrow \boldsymbol{\omega}_j^{(i)} / \sum_{k=1}^N \boldsymbol{\omega}_j^{(k)}$   $i = 1, 2, \dots, N$ ;
10   $\{j_1, j_2, \dots, j_N\} \leftarrow \text{amostra\_índices}(N, \{\tilde{\boldsymbol{\omega}}_j^{(i)}\}_{i=1}^N)$ ;
11  para  $i \leftarrow 1$  até  $N$  faça
12     $X_k^{(i)} \leftarrow \tilde{X}_k^{(j_i)}$ ;
13     $\boldsymbol{\theta}^{(i)} \leftarrow \tilde{\boldsymbol{\theta}}^{(j_i)}$ ;
14     $Z_{1:k}^{(i)} \leftarrow (X_{1:k-1}^{(i)} \cup X_k^{(i)}, \boldsymbol{\theta}^{(i)})$ ;
15 retorna  $\{Z_{1:n}^{(i)}, \frac{1}{N}\}$ ;
```

O método demonstrou ser eficaz em situações caracterizadas por séries curtas, entretanto, o mesmo apresenta uma instabilidade excessiva à medida que o número de observações aumenta. Esta situação ocorre pois o ruído artificial acrescido ao problema, mesmo se associado a valores pequenos para τ^2 , adiciona a cada instante de tempo um pouco mais de variabilidade às estimativas:

$$\begin{aligned} \text{Var}[\boldsymbol{\theta}] &= \mathbb{E}[\text{Var}[\boldsymbol{\theta}|I]] + \text{Var}[\mathbb{E}[\boldsymbol{\theta}|I]] \\ &= \mathbb{E}[\tau^2 \cdot \mathbf{I}_{n_\theta}] + \text{Var}[\boldsymbol{\theta}^{(I)}] = \tau^2 \cdot \mathbf{I}_{n_\theta} + \Sigma, \end{aligned}$$

onde I é uma variável aleatória latente responsável por rastrear os índices da mistura e

$$\bar{\boldsymbol{\theta}} = \sum_{i=1}^N \tilde{\boldsymbol{\omega}}_{t-1}^{(i)} \boldsymbol{\theta}^{(i)} \quad \text{e} \quad \Sigma = \sum_{i=1}^N \tilde{\boldsymbol{\omega}}_{t-1}^{(i)} (\boldsymbol{\theta}^{(i)} - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta}^{(i)} - \bar{\boldsymbol{\theta}})'$$

correspondem, na devida ordem, à média e a matriz de covariâncias de $\{\boldsymbol{\theta}^{(i)}\}$. Desta forma, ao final da filtragem de séries mais longas, teremos uma distribuição *a posteriori* excessivamente difusa, a qual não representa de forma adequada a incerteza a respeito dos parâmetros.

Filtro de Liu e West

O Filtro de Liu e West surge como uma estratégia capaz de dar uma dinâmica artificial à população estatística sem aumentar a variabilidade da mesma. Isto pode ser efeito assumindo uma dinâmica análoga à proposta por Kitagawa porém considerando uma correlação negativa entre o ruído adicional e o enxame de pontos. Este objetivo pode ser facilmente alcançado se comprimirmos a população $\{\theta^{(i)}\}$ em direção à média ponderada $\bar{\theta}$ através de combinações lineares convexas e, em seguida, introduzirmos o ruído responsável pelo rejuvenescimento artificial. O modelo assim definido é caracterizado pela nova mistura de Gaussianas:

$$p_{t-1}(\theta) = p(\theta|y_{0:t-1}) = \sum_{i=1}^N \tilde{\omega}_{t-1}^{(i)} N\left(\theta; a \cdot \theta^{(i)} + (1-a)\bar{\theta}, h^2 \Sigma\right), \quad (3.55)$$

onde $a \in (0, 1)$ e $h^2 = 1 - a^2$. Assim como no caso do filtro de Kitagawa, transformações convenientes devem ser realizadas nos parâmetros caso o suporte dos mesmos não seja a reta.

Algoritmo 15: Filtro de Liu e West

Entrada: Tamanho amostral (N), observações ($y_{1:n}$), fator de desconto (δ) e distribuição *a priori* (p);

Saída: Aproximação discreta $\hat{\pi}_n(dz_{1:n}) = \sum_{i=1}^N \tilde{\omega}_n^{(i)} \delta_{(z_{1:n})}^{(i)}(dz_{1:n})$;

- 1 $a \leftarrow (3\delta - 1)/2\delta$;
 - 2 $h^2 \leftarrow 1 - a^2$;
 - 3 **para** $i \leftarrow 1$ até N **faça**
 - 4 Amostre $\theta^{(i)} \sim p$;
 - 5 Amostre $X_0^{(i)} \sim q_0$;
 - 6 $\omega_0^{(i)} \leftarrow 1$;
 - 7 $\bar{\theta} \leftarrow \frac{1}{N} \sum_{i=1}^N \theta^{(i)}$;
 - 8 $V \leftarrow \frac{1}{N} \sum_{i=1}^N (\theta^{(i)} - \bar{\theta})(\theta^{(i)} - \bar{\theta})'$;
 - 9 **para** $j \leftarrow 1$ até n **faça**
 - 10 **para** $i \leftarrow 1$ até N **faça**
 - 11 $m^{(i)} \leftarrow a\theta^{(i)} + (1-a)\bar{\theta}$;
 - 12 $\omega_{aux}^{(i)} \leftarrow \tilde{\omega}_{j-1}^{(i)} \cdot p(y_j | G(X_{j-1}^{(i)}, m^{(i)}))$;
 - 13 $\{j_1, j_2, \dots, j_N\} \leftarrow \text{amostra_índices}(N, \{\omega_{aux}^{(i)}\}_{i=1}^N)$;
 - 14 **para** $i \leftarrow 1$ até N **faça**
 - 15 Amostre $\theta^{(i)} \sim N(m^{(j_i)}, V)$;
 - 16 Amostre $X_j^{(i)} \sim q(\cdot | X_{j-1}^{(j_i)}, \theta^{(i)})$;
 - 17 $\omega_j^{(i)} \leftarrow g(y_j | X_j^{(i)}, \theta^{(i)}) / \omega_{aux}^{(j_i)}$;
 - 18 $Z_{1:j}^{(i)} \leftarrow (X_{1:j-1}^{(i)} \cup X_j^{(i)}, \theta^{(i)})$;
 - 19 Normalize $\tilde{\omega}_j^{(i)} \leftarrow \omega_j^{(i)} / \sum_{k=1}^N \omega_j^{(k)}$, $i = 1, 2, \dots, N$;
 - 20 $\bar{\theta} \leftarrow \sum_{i=1}^N \tilde{\omega}_j^{(i)} \theta^{(i)}$;
 - 21 $V \leftarrow \sum_{i=1}^N \tilde{\omega}_j^{(i)} (\theta^{(i)} - \bar{\theta})(\theta^{(i)} - \bar{\theta})'$;
 - 22 **retorna** $\{Z_{1:n}^{(i)}, \tilde{\omega}_n^{(i)}\}$;
-

A nova aproximação preserva os dois primeiros momentos da distribuição amostral, assegurando que a informação não seja difundida ao longo do tempo:

$$\begin{aligned}\mathbb{E}[\theta] &= \mathbb{E}[\mathbb{E}[\theta|I]] \\ &= \mathbb{E}\left[a \cdot \theta^{(I)} + (1-a) \cdot \bar{\theta}\right] = a \cdot \bar{\theta} + (1-a) \cdot \bar{\theta} = \bar{\theta}; \\ \text{Var}[\theta] &= \mathbb{E}[\text{Var}[\theta|I]] + \text{Var}[\mathbb{E}[\theta|I]] \\ &= \mathbb{E}\left[h^2 \Sigma\right] + \text{Var}\left[a \cdot \theta^{(I)} + (1-a) \cdot \bar{\theta}\right] \\ &= h^2 \Sigma + a^2 \text{Var}\left[\theta^{(I)}\right] = \Sigma \cdot (h^2 + a^2) = \Sigma.\end{aligned}$$

O parâmetro de encurtamento $a \in (0, 1)$ é usualmente definido em função de um fator de desconto $\delta \in (0,95; 0,99)$:

$$a = \frac{3\delta - 1}{2\delta}.$$

Quanto maior o valor de δ , menor o será o desconto na informação oferecida pelo enxame de partículas. Neste caso, os indivíduos amostrados da aproximação por misturas serão muito próximos de seus ancestrais, mantendo assim uma população pouco rejuvenescida. Por outro lado, se utilizarmos fatores de desconto mais baixos, teremos a amostragem de pontos cujo modelo probabilístico tende ao de uma variável normalmente distribuída com média $\bar{\theta}$ e matriz de covariâncias Σ . A figura a seguir ilustra a distribuição associada a um dado ponto considerando diferentes fatores de desconto; o diagrama deixa claro que valores excessivamente pequenos de δ conduzem rapidamente o modelo probabilístico inicial da população a uma simples distribuição Gaussiana multivariada.

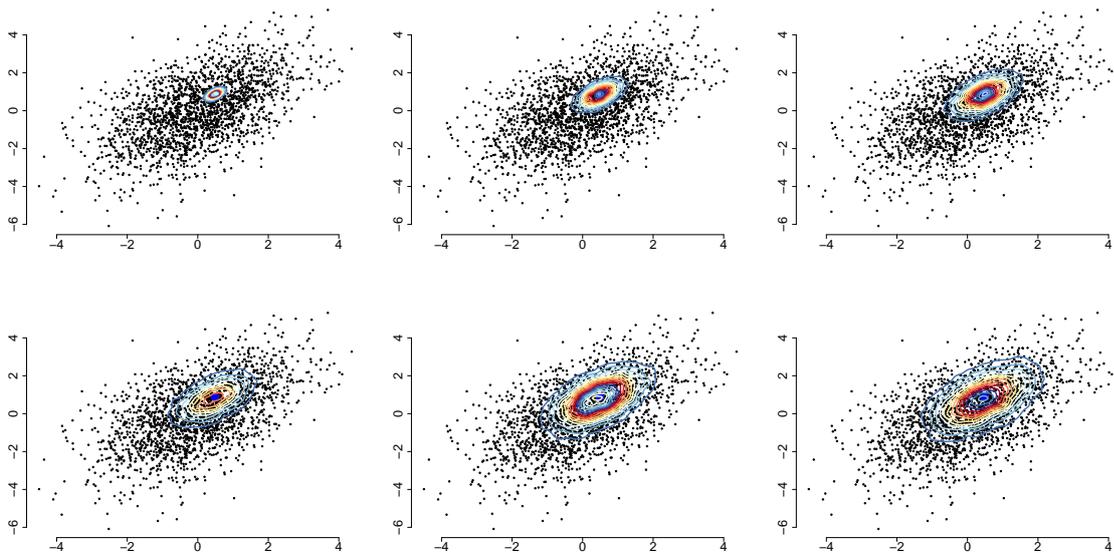


Figura 3 – Distribuições associadas a um único ponto considerando diferentes valores para o fator de desconto. Partindo do primeiro quadro e, percorrendo os demais em sentido horário, temos os comportamentos correspondentes às seguintes escolhas de δ : 0,99; 0,94; 0,89; 0,84; 0,79; 0,74.

A grande vantagem desta abordagem é que nenhuma hipótese a respeito da forma funcional do modelo é exigida. Desta forma, a abordagem de Liu e West constitui uma das metodologias mais gerais para o tratamento do problema da filtragem na presença de parâmetros estáticos desconhecidos. Destaca-se ainda que o método é muito eficiente em termos computacionais e é de implementação extremamente simples. O método é associado, em sua versão original, ao filtro Auxiliar. Basicamente, um novo ponto é gerado após a amostragem do índice auxiliar de modo que esta iteração possa ser interpretada agora como um processo de reamostragem suavizada dos parâmetros estáticos. Vários estudos podem ser encontrados na literatura a respeito desta metodologia, alguns indicando sua eficiência e outros apontando uma possível tendência degenerativa.

Filtro de Storvik

Metodologias mais específicas de rejuvenescimento podem ser adotadas além das propostas discutidas acima. Em particular, Storvik (2002) propõe amostrar diretamente de $p(\theta|x_{0:t}, y_{0:t})$, o que pode ser feito em situações onde tal distribuição pode ser tratada em função de um vetor de estatísticas suficientes $\mathcal{S}_t(x_{0:t}, y_{0:t})$ com respeito às trajetórias filtrada e observada. Ao fazer uso do conhecimento a respeito da forma funcional da distribuição *a posteriori*, esta abordagem garante a geração de aproximações mais precisas para o modelo probabilístico desconhecido. Assim, espera-se que tal procedimento seja capaz de gerar estimativas associadas a uma menor variabilidade quando comparadas às obtidas do filtro de Liu e West.

Algoritmo 16: Filtro de Storvik

Entrada: Tamanho amostral (N), observações ($y_{1:n}$) e distribuição *a priori* (p);

Saída: Aproximação discreta $\hat{\pi}_n(dz_{1:n}) = \frac{1}{N} \sum_{i=1}^N \delta_{(z_{1:n})}^{(i)}(dz_{1:n})$;

```

1 para  $i \leftarrow 1$  até  $N$  faça
2   Amostre  $X_0^{(i)} \sim q_0$ ;
3   Amostre  $\theta^{(i)} \sim p$ ;
4 para  $j \leftarrow 1$  até  $n$  faça
5   para  $i \leftarrow 1$  até  $N$  faça
6     Amostre  $\tilde{X}_j^{(i)} \sim q(\cdot | X_{j-1}^{(i)}, \theta^{(i)})$ ;
7     Amostre  $\tilde{\theta}^{(i)} \sim p(\theta^{(i-1)} | \mathcal{S}_{t-1}(x_{0:t-1}^{(i)}, y_{0:t-1}))$ ;
8      $\omega_j^{(i)} \leftarrow g(y_j | \tilde{X}_j^{(i)}, \tilde{\theta}^{(i)})$ ;
9   Normalize  $\tilde{\omega}_j^{(i)} \leftarrow \omega_j^{(i)} / \sum_{k=1}^N \omega_j^{(k)}$   $i = 1, 2, \dots, N$ ;
10   $\{j_1, j_2, \dots, j_N\} \leftarrow \text{amostra\_índices}(N, \{\tilde{\omega}_j^{(i)}\}_{i=1}^N)$ ;
11  para  $i \leftarrow 1$  até  $N$  faça
12     $X_k^{(i)} \leftarrow \tilde{X}_k^{(j_i)}$ ;
13     $\theta^{(i)} \leftarrow \tilde{\theta}^{(j_i)}$ ;  $Z_{1:k}^{(i)} \leftarrow (X_{1:k-1}^{(i)} \cup X_k^{(i)}, \theta^{(i)})$ ;
14 retorna  $\{Z_{1:n}^{(i)}, \frac{1}{N}\}$ ;
```

Em geral, estruturas favoráveis à utilização desta estratégia sempre podem ser obtidas quando a dinâmica do modelo Markoviano oculto é linear com respeito às quantidades estáticas. Considerando, por exemplo, processos unidimensionais

$$x_t = G(t, x_{t-1})\beta + \eta_t, \quad \eta_t \sim N(0, \tau^2),$$

onde G é uma aplicação de \mathbb{R}^2 em $\mathbb{R}^{n\beta}$, e assumindo uma distribuição *a priori* Normal-Gama Invertida para o par (β, τ^2) ,

$$\tau^2 \sim GI\left(\frac{1}{2}v_0, \frac{1}{2}v_0s_0^2\right) \quad \text{e} \quad \beta|\tau^2 \sim N(\mu_0, \sigma_2\Lambda_0^{-1}),$$

obtemos o seguinte esquema de atualização *a posteriori*:

$$\tau^2|x_{0:t} \sim GI(a_t, b_t) \quad \text{e} \quad \beta|\tau^2, x_{0:t} \sim N(\mu_t, \sigma^2\Lambda_t^{-1}),$$

onde

$$\begin{aligned} \mathbf{X}_t &= (x_1, x_2, \dots, x_t)', \\ \check{\mathbf{G}}_{t-1} &= (G_0(0, x_0), G_1(1, x_1), \dots, G_{t-1}(t-1, x_{t-1}))', \\ \mu_t &= (\check{\mathbf{G}}'_{t-1}\check{\mathbf{G}}_{t-1} + \Lambda_0)^{-1} (\Lambda_0\mu_0 + \check{\mathbf{G}}'_{t-1}\mathbf{X}_t), \\ \Lambda_t &= \check{\mathbf{G}}'_{t-1}\check{\mathbf{G}}_{t-1} + \Lambda_0, \\ a_t &= \frac{1}{2}(v_0 + t) \\ b_t &= \frac{1}{2}(v_0s_0^2 + \mathbf{X}'_t\mathbf{X}_t + \mu'_0\Lambda_0\mu_0 - \mu'_t\Lambda_t\mu_t). \end{aligned}$$

Neste caso, basta armazenarmos, a cada iteração, a família de estatísticas suficientes:

$$\mathcal{S}_t(x_{0:t}) = (\mathbf{X}'_t\mathbf{X}_t, \check{\mathbf{G}}'_{t-1}\check{\mathbf{G}}_{t-1}, \check{\mathbf{G}}'_{t-1}\mathbf{X}_t)'$$

Um procedimento análogo pode ser obtido se o processo de observações também apresentar um relacionamento linear com seus parâmetros estáticos. É importante ressaltar que a função G não precisa necessariamente definir uma relação linear com respeito aos argumentos t e x_t , o que confere mais generalidade ao dispositivo. Tal estrutura trata apenas do cenário mais comum de aplicações do método. Situações mais abrangentes também são passíveis de definição de um esquema de propagação de estatísticas suficientes, no entanto, tais situações normalmente não garantem a possibilidade de amostragem direta da distribuição *a posteriori* parcial. Nestes casos, a atualização pode ser feita por blocos de acordo com um dispositivo similar à de um amostrador de Gibbs. Esta última perspectiva assume que a amostra ponderada mais atual fornece uma boa aproximação para densidade alvo, logo, tal distribuição permanecerá invariante após a aplicação de um passo Gibbs. Devemos destacar que esta segunda forma de aproximação pode vir associada a um vetor de estatísticas suficientes de dimensão elevada, o que pode afetar significativamente a complexidade computacional do método.

Diferentemente do dispositivo de Liu e West, o filtro de Storvik foi associado inicialmente ao paradigma *Bootstrap*, caracterizando assim uma estratégia do tipo amostragem/reamostragem. Variações do mesmo algoritmo poderiam ser obtidas inserindo a dinâmica de rejuvenescimento artificial diretamente na estrutura do filtro Auxiliar. Outro ponto de destaque com respeito a este método é que a família composta pelo vetor de estatísticas suficientes $\left\{ \mathcal{S}_t \left(x_{0:t}^{(i)}, y_{0:t} \right) \right\}$ também é desprovida de dinâmica natural e, portanto, é passível de degeneração. Situações onde esta estrutura apresenta um enxugamento muito acelerado podem conduzir o dispositivo de filtragem a estimativas imprecisas.

Parameter Learning

O esquema *parameter learning*, proposto por [Carvalho et al. \(2010\)](#), trata basicamente de uma especialização do filtro de Storvik, a qual aplica o dispositivo de rejuvenescimento artificial baseado na propagação de estatísticas suficientes a uma versão do Filtro Auxiliar com distribuição proposta ótima. Neste caso, os pesos auxiliares serão dados por

$$\omega_{aux}^{(i)} = \tilde{\omega}_{t-1}^{(i)} \cdot g \left(y_t \mid x_{t-1}^{(i)}, \theta^{(i)} \right), \quad i = 1, 2, \dots, N$$

e, como consequência imediata, teremos um conjunto homogêneo de ponderações ao final de cada iteração:

$$\begin{aligned} \omega_t &\propto \tilde{\omega}_{t-1}^{(i)} \frac{g \left(y_t \mid x_t^{(i)}, \theta^{(i)} \right) q \left(x_t^{(i)} \mid x_{t-1}^{(i)}, \theta^{(i)} \right)}{\omega_{aux}^{(i)} \cdot q_{t|t-1} \left(x_t^{(i)} \mid y_t, x_{t-1}^{(i)}, \theta^{(i)} \right)} \\ &= \tilde{\omega}_{t-1}^{(i)} \cdot \frac{g \left(y_t \mid x_t^{(i)}, \theta^{(i)} \right) q \left(x_t^{(i)} \mid x_{t-1}^{(i)}, \theta^{(i)} \right)}{\tilde{\omega}_{t-1}^{(i)} \cdot g \left(y_t \mid x_{t-1}^{(i)}, \theta^{(i)} \right) \cdot \frac{g \left(y_t \mid x_t^{(i)} \right) q \left(x_t^{(i)} \mid x_{t-1}^{(i)}, \theta^{(i)} \right)}{g \left(y_t \mid x_{t-1}^{(i)}, \theta^{(i)} \right)}} = 1. \end{aligned}$$

Este esquema culmina em um enxame de pontos completamente diversificado e de ponderações homogêneas ao final de cada passo, o que constitui o maior diferencial a favor de sua aplicação. Os autores sugerem ainda que o passo de rejuvenescimento artificial seja feito após a propagação dos estados. Por esta razão, as densidades associadas aos parâmetros estáticos não foram explicitadas na relação acima. A interpretação neste caso é que o rejuvenescimento artificial constitui um passo Gibbs, o qual mantém invariante o modelo probabilístico associado ao vetor estático.

A abordagem é bastante frutífera pois, além de assegurar a construção de um algoritmo simples, garante também a obtenção de estimativas mais estáveis. Todavia, o dispositivo também faz uso de um vetor de estatísticas suficientes desprovido de dinâmica natural. Desta forma, é possível observar efeitos de degeneração nesta estratégia assim como nas demais. Como o dispositivo de rejuvenescimento artificial é rigorosamente o mesmo adotado na aproximação de Storvik, não faremos diferenciação destes métodos. A partir desta seção, referenciaremos ambas as perspectivas por *parameter learning* ou simplesmente PL.

Algoritmo 17: Filtro PL

Entrada: Tamanho amostral (N), observações ($y_{1:n}$) e distribuição *a priori* (p);

Saída: Aproximação discreta $\hat{\pi}_n(dz_{1:n}) = \frac{1}{N} \sum_{i=1}^N \delta_{(z_{1:n}^{(i)})}(dz_{1:n})$;

```

1 para  $i \leftarrow 1$  até  $N$  faça
2   Amostre  $\theta^{(i)} \sim p$ ;
3   Amostre  $X_0^{(i)} \sim q_0$ ;
4    $\omega_0^{(i)} \leftarrow 1$ ;
5 para  $j \leftarrow 1$  até  $n$  faça
6   para  $i \leftarrow 1$  até  $N$  faça
7      $\omega_{aux}^{(i)} \leftarrow p(y_j | G(X_{j-1}^{(i)}, m^{(i)}))$ ;
8      $\{j_1, j_2, \dots, j_N\} \leftarrow \text{amostra\_índices}(N, \{\omega_{aux}^{(i)}\}_{i=1}^N)$ ;
9     para  $i \leftarrow 1$  até  $N$  faça
10      Amostre  $X_j^{(i)} \sim q(\cdot | y_t, X_{j-1}^{(j_i)}, \theta^{(i)})$ ;
11      Propaga  $S_t^{(i)} \leftarrow \mathcal{S}(S_{t-1}^{(j_i)}, X_{1:t}^{(i)}, y_{1:t})$ ;
12      Amostre  $\theta^{(i)} \sim p(\cdot | S_t^{(i)})$ ;
13       $Z_{1:j}^{(i)} \leftarrow (X_{1:j-1}^{(i)} \cup X_j^{(i)}, \theta^{(i)})$ ;
14 retorna  $\{Z_{1:n}^{(i)}, \frac{1}{N}\}$ ;
```

SMC²

O próximo método constitui uma extensão do algoritmo *IBIS* (*Iterated Batch Importance Sampling*) desenvolvido por [Chopin \(2002\)](#) com o intuito de prover um dispositivo sequencial eficiente para a realização de inferências no contexto mais tradicional das amostras aleatórias. A metodologia *IBIS* parte primeiramente de uma amostra de indivíduos $\{\theta^{(i)}\}$ gerada da distribuição *a priori* $p(\theta)$. Como esses indivíduos tratam de uma amostra independente e identicamente distribuída, podemos associar a cada um deles a ponderação $1/N$. Nos instantes seguintes as ponderações são atualizadas a partir da verossimilhança parcial das observações:

$$\omega_t^{(i)} \propto \omega_{t-1}^{(i)} p(y_t | y_{1:t-1}, \theta^{(i)}), \quad i = 1, 2, \dots, N,$$

com a convenção $p(y_1 | y_{1:0}, \theta^{(i)}) = p(y_1 | \theta^{(i)})$. Caso algum critério de degeneração seja satisfeito no instante t , uma nova população é amostrada da mistura finita

$$\frac{1}{\sum_{i=1}^N \omega_t^{(i)}} \sum_{i=1}^N \omega_t^{(i)} K_t(\theta^{(i)}, \cdot),$$

onde K é um núcleo de transição invariante com respeito à distribuição *a posteriori* do modelo. Em geral, K pode ser tomado como uma transição do tipo Metropolis-Hastings baseada em passeios aleatórios, o que confere uma maior abrangência ao método por constituir uma escolha

isenta de hipóteses adicionais a respeito do modelo. Para evitar problemas de consistência caso os parâmetros não estejam definidos em toda a reta, podemos basear a transição em passeios aleatórios refletidos ou simplesmente considerar transformações análogas às descritas para os métodos de Kitagawa e de Liu e West. Ao final do rejuvenescimento a amostra ponderada dá lugar à nova população, a qual é caracterizada por pesos constantes. O dispositivo *IBIS* oferece um ferramental alternativo à amostragem MCMC e tem como principal vantagem uma complexidade computacional incomparavelmente menor. Desta forma, é natural tentar estender esta abordagem ao cenário caracterizado por cadeias Markovianas ocultas. Uma das grandes dificuldades encontradas neste caso é o desconhecimento da expressão analítica da função de verossimilhança. A ausência desta forma funcional inviabiliza em um primeiro momento o desenvolvimento da estratégia *IBIS*. No entanto, uma aproximação consistente pode ser obtida se substituirmos o funcional desconhecido por uma estimativa não viciada construída com o auxílio dos filtros de partículas. O método SMC^2 trata basicamente desta aproximação. A ideia é embutir um filtro de partículas no dispositivo sequencial *IBIS* de modo a fornecer as estimativas da função de verossimilhança associadas a cada indivíduo $\theta^{(i)}$. Os conjuntos de partículas $x_{1:n}^{(1:N_x, i)}$ construídos em cada um dos i filtros independentes, podem ser reaproveitados ao final do algoritmo para a realização de inferência a respeito da sequência latente.

O algoritmo SMC^2 demanda, ao contrário de sua perspectiva de origem, uma carga excessiva de recursos computacionais já que um filtro *Bootstrap* independente deve ser lançado para cada indivíduo no procedimento *IBIS* a fim de construir uma estimativa da função de verossimilhança. Em particular, se o método utiliza $N_\theta = 1000$ indivíduos para a aproximação da distribuição *a posteriori* e um número pequeno de partículas $N_x = 100$ na execução de cada filtro independente, teremos uma estimativa construída a partir de $N_x \times N_\theta = 100000$ pontos. Esta configuração simples ilustra o quanto o método pode ser computacionalmente complexo. O desempenho pode ainda ser mais insatisfatório uma vez que os autores sugerem que o número de partículas N_x , utilizado no passo de filtragem pura, seja dobrado a cada instante onde algum critério de degeneração é satisfeito. É comum, por exemplo, finalizar o método com um montante de partículas superior 10^8 indivíduos e com um tempo de execução superior a uma dezena de horas. Todavia, devemos destacar que a estrutura do método permite o desenvolvimento de uma implementação paralelizada. Esta alternativa constitui um desafio computacional considerável porém pode garantir a obtenção de tempos de execução mais atrativos. Apesar da pouca eficiência, temos que o método SMC^2 define um método consistente, o qual não carece de nenhuma hipótese específica a respeito da forma funcional do modelo de interesse. Este fato garante a aplicação do método a qualquer modelo Markoviano oculto.

Algoritmo 18: SMC2

Entrada: Tamanhos amostrais (N_θ, N_x) , observações $(y_{1:n})$, distribuição *a priori* (p) , densidade de transição (K) e algoritmo de filtragem (FP) ;

Saída: Aproximação discreta $\hat{\pi}_n(dz_{1:n}) = \sum_{i=1}^N \tilde{\omega}_n^{(i)} \delta_{(z_{1:n}^{(i)})}(dz_{1:n})$;

```

1 para  $i \leftarrow 1$  até  $N_\theta$  faça
2   Amostre  $\theta^{(i)} \sim p_\theta$ ;
3   Estime  $p(y_1 | \theta^{(i)})$  e  $x_1^{(1:N_x, 1)}$  a partir de  $FP$ ;
4    $\omega_1^{(i)} \leftarrow \hat{p}(y_1 | \theta^{(i)})$ ;
5 para  $j \leftarrow 1$  até  $n$  faça
6   para  $i \leftarrow 1$  até  $N_\theta$  faça
7     Estime  $p(y_t | y_{1:t-1}, \theta^{(i)})$  e  $x_{1:j}^{(1:N_x, t)}$  a partir de  $FP$ ;
8      $\omega_j^{(i)} \leftarrow \omega_{j-1}^{(i)} \hat{p}(y_t | y_{1:t-1}, \theta^{(i)})$ ;
9    $TEA_j \leftarrow \left( \sum_{i=1}^{N_\theta} \omega_j^{(i)} \right)^2 / \sum_{i=1}^{N_\theta} \left( \tilde{\omega}_j^{(i)} \right)^2$ ;
10  se  $TEA_j < N_\theta / 2$  então
11     $\{j_1, j_2, \dots, j_N\} \leftarrow \text{amostra\_índices} \left( N, \{ \omega_j^{(i)} / \sum_{i=1}^{N_\theta} \omega_j^{(i)} \}_{i=1}^{N_\theta} \right)$ ;
12    para  $i \leftarrow 1$  até  $N_\theta$  faça
13      Amostre  $\theta^{(i)} \sim K_j \left( \theta^{(j_i)}, x_{1:j}^{(1:N_x, j_i)}, \cdot \right)$ ;
14       $Z_{1:j}^{(i)} \leftarrow \{ x_{1:j}^{(1:N_x, j_i)}, \theta^{(i)} \}$ ;
15       $\omega_j^{(i)} \leftarrow 1$ ;
16 retorna  $\{ Z_{1:n}^{(i)}, \omega_n^{(i)} / \sum_{j=1}^{N_\theta} \omega_n^{(j)} \}$ ;
```

Um caso particular de destaque do método diz respeito aos modelos lineares Gaussianos ou apenas condicionalmente lineares Gaussianos. Modelos desta natureza permitem a construção da função de verossimilhança a partir das recursões do tradicional Filtro de Kalman, evitando assim a utilização intermediária de uma sequência de filtros de partículas. Outra situação favorável é obtida quando o modelo em questão habilita a amostragem direta de parâmetros estáticos com o auxílio de distribuições baseadas em estatísticas suficientes das trajetórias observada e filtrada. A amostragem direta evita a utilização de passos Métropolis-Hastings no desenvolvimento do núcleo de transição K . Nestes casos, não teremos problemas com respeito à taxa de aceitação de novos indivíduos e, conseqüentemente, o número de partículas N_x não precisará ser incrementado. A estratégia SMC² surge como um método muito atrativo para o tratamento do problema da filtragem considerando a existência de parâmetros estáticos desconhecidos nestes cenários específicos.

PMCMC

Métodos do tipo PMCMC (*Particle Markov Chain Monte Carlo*) englobam um conjunto de estratégias MCMC apoiadas por filtros de partículas. A abordagem mais geral encontrada nesta classe refere-se a uma aproximação da popular dinâmica Metropolis-Hastings com o intuito de amostrar da distribuição $p(x_{0:n}, \theta | y_{0:n})$. Seguindo a forma tradicional do método, poderíamos apenas gerar um novo ponto a partir de uma densidade de transição da forma

$$q((x_{0:t}, \theta)' | (x_{0:t}, \theta)) = q(\theta' | \theta) p_{\theta'}(x'_{0:t} | y_{0:t}),$$

e aceitar a nova geração com probabilidade

$$1 \wedge \frac{p_{\theta'}(y_{0:t}) p(\theta') q(\theta | \theta')}{p_{\theta}(y_{0:t}) p(\theta) q(\theta' | \theta)}.$$

No contexto dos modelos Markovianos ocultos, esta estratégia simplificada seria impraticável por dois motivos: impossibilidade de amostrar de $p_{\theta'}(x_{0:t} | y_{0:t})$ e desconhecimento da forma funcional associada à verossimilhança $p_{\theta}(y_{0:t})$. Entretanto, procedendo de forma análoga ao método SMC², poderíamos utilizar um filtro de partículas para aproximar as duas quantidades desconhecidas. O restante do dispositivo segue de forma idêntica à versão clássica do algoritmo Metropolis-Hastings, conforme sumarizado no diagrama a seguir.

Algoritmo 19: PMCMC

Entrada: Tamanho amostral (N), observações ($y_{1:n}$), distribuição *a priori* (v), densidade de transição (K) e algoritmo de filtragem (FP);

Saída: Amostra aleatória $\{\theta^{(i)}\}_{i=1}^N \sim \pi$;

- 1 $i \leftarrow 0$;
 - 2 Amostre $\theta^{(0)} \sim v$;
 - 3 Execute FP para obter $X_{1:t}^{(0)} \sim \hat{p}(\cdot | y_{1:t}, \theta^{(0)})$ e $\hat{p}(y_{1:n} | \theta^{(0)})$;
 - 4 **repita**
 - 5 $i \leftarrow i + 1$;
 - 6 Amostre $u \sim U(0, 1)$;
 - 7 Amostre $\tilde{\theta}^{(i)} \sim K$;
 - 8 Execute FP para obter $\tilde{X}_{1:t}^{(i)} \sim \hat{p}(\cdot | y_{1:t}, \tilde{\theta}^{(i)})$ e $\hat{p}(y_{1:n} | \tilde{\theta}^{(i)})$;
 - 9 **se** $u < \min\left(1, \frac{\hat{p}(y_{1:n} | \tilde{\theta}^{(i)}) v(\tilde{\theta}^{(i)})}{\hat{p}(y_{1:n} | \theta^{(i-1)}) v(\theta^{(i-1)})} \cdot \frac{K(\tilde{\theta}^{(i)}, \theta^{(i-1)})}{K(\theta^{(i-1)}, \tilde{\theta}^{(i)})}\right)$ **então**
 - 10 $\theta^{(i)} \leftarrow \tilde{\theta}^{(i)}$;
 - 11 $X_{1:t}^{(i)} \leftarrow \tilde{X}_{1:t}^{(i)}$;
 - 12 $\hat{p}(y_{1:n} | \theta^{(i)}) \leftarrow \hat{p}(y_{1:n} | \tilde{\theta}^{(i)})$;
 - 13 **senão**
 - 14 $\theta^{(i)} \leftarrow \theta^{(i-1)}$;
 - 15 $X_{1:t}^{(i)} \leftarrow X_{1:t}^{(i-1)}$;
 - 16 $\hat{p}(y_{1:n} | \theta^{(i)}) \leftarrow \hat{p}(y_{1:n} | \theta^{(i-1)})$;
 - 17 **até** Critério de parada;
 - 18 **retorna** N últimas observações da cadeia;
-

Andrieu, Doucet e Holenstein (2010) mostram que o método PMCMC tem $f(x_{0:t}, \theta | y_{0:t})$ como distribuição invariante para qualquer número N de partículas utilizado no dispositivo numérico de filtragem, o que caracteriza um resultado surpreendente. Todavia, o tamanho amostral N tem um impacto significativo no desempenho do algoritmo. Valores elevados de N são capazes de gerar estimativas associadas a um menor montante de variabilidade na comparação com aproximações baseados em valores pequenos do mesmo parâmetro, mas o custo computacional da construção de $\hat{p}_\theta(y_{0:t})$ cresce significativamente com N . De uma forma geral, os autores sugerem que o valor de N seja escolhido de forma a assegurar que a variância das aproximações para o logaritmo da verossimilhança figure no intervalo $(1, 2; 1, 3)$. Como a variância destas estimativas cresce linearmente com o número t de observações, segue que a complexidade computacional é da ordem de $O(t^2)$ operações por iteração. Desta forma, temos um algoritmo computacionalmente intensivo, o qual demanda um esforço computacional superior até mesmo ao demandado pela abordagem SMC².

Uma aproximação por meio de partículas do amostrador de Gibbs também pode ser obtida de forma similar. Esta estratégia, no entanto, é mais restritiva que a aproximação M-H pois necessita reamostrar diretamente das distribuições $p(\theta | x_{0:t}, y_{0:t})$ e $p(x_{0:t} | y_{0:t})$. Mesmo em situações onde é possível amostrar diretamente da distribuição condicional completa dos parâmetros estáticos, podemos ter problemas nas taxas de convergência do método caso seja possível amostrar de $p(x_{0:t} | y_{0:t})$ apenas de maneira aproximada. Desta forma, temos que o método define uma boa estratégia de amostragem basicamente em situações onde $p(\theta | x_{0:t}, y_{0:t})$ e $p(x_{0:t} | y_{0:t}, \theta)$ são acessíveis via algoritmos do tipo filtragem sucessiva-amostragem retroativa.

APRENDIZAGEM DE PARÂMETROS OTIMIZADA

Este capítulo introduz algumas diretrizes de implementação desenvolvidas com o intuito de otimizar o desempenho dos algoritmos de filtragem associados à aprendizagem de parâmetros estáticos, com destaque para os métodos de Liu e West e Storvik/*Parameter Learning*. A proposta desenvolvida aqui preserva a estruturação matemática dos métodos e concentra mais esforços em metodologias capazes de refinar o balanceamento entre diversidade e foco das nuvens de partículas.

4.1 Aprendizagem de Parâmetros Revisitada

O conceito de modelo auto-organizativo introduzido por [Kitagawa \(1998\)](#) estabeleceu uma associação promissora entre filtros de partículas e inferência paramétrica. O estado da arte acerca desta perspectiva é a concatenação dos parâmetros estáticos desconhecidos com o vetor de estados latentes, dando origem a uma estrutura expandida

$$Z_t = (X_t, \theta), \quad t = 0, 1, 2, \dots, n,$$

cuja estimação pode ser realizada sequencialmente com o auxílio dos métodos sequenciais de Monte Carlo. Originalmente, a componente θ foi associada a uma dinâmica artificial, mais precisamente um passeio aleatório, a fim de garantir diversidade a uma grandeza naturalmente desprovida de dinâmica. O método funcionou bem em estudos envolvendo séries pequenas de observações entretanto apresentou resultados muito desfavoráveis em cenários gerais devido à perda de informação associada à adição de um ruído externo ao problema. Variantes mais recentes deste conceito colheram melhores frutos a partir da introdução de dispositivos de rejuvenescimento artificial mais sofisticados. Estas abordagens podem ser interpretadas, em sua essência, como instâncias específicas de um procedimento de regularização o qual substitui a

medida aleatória inerte

$$\hat{\pi}_t(y_{1:t}, dz_{1:t}) = \sum_{i=1}^N \tilde{\omega}_t^{(i)} \delta_{z_{1:t}^{(i)}}(dz_{1:t}),$$

por uma mistura finita da forma

$$\tilde{\pi}_t(y_{1:t}, dz_{1:t}) = \sum_{i=1}^N \tilde{\omega}_t^{(i)} \delta_{x_{1:t}^{(i)}}(dx_{1:t}) \times K\left((y_{1:t}, z_{1:t}^{(i)}), d\theta\right), \quad (4.1)$$

onde K representa um núcleo de transição associado ao rejuvenescimento artificial de θ . Integrando a componente latente da cadeia, chegamos a uma aproximação genérica da incerteza *a posteriori* acerca dos parâmetros estáticos

$$\hat{p}(y_{1:t}, d\theta) = \sum_{i=1}^N \tilde{\omega}_t^{(i)} K\left((y_{1:t}, z_{1:t}^{(i)}), d\theta\right). \quad (4.2)$$

A regularização pode ser realizada tomando a transição K como um núcleo estimador Gaussiano (LIU; WEST, 2001), resultando na seguinte densidade aproximada

$$\hat{p}(\theta|y_{1:t}) = \sum_{i=1}^N \tilde{\omega}_t^{(i)} N\left(\theta; a \cdot \theta^{(i)} + (1-a) \cdot \bar{\theta}, h^2 \cdot \mathbf{V}\right), \quad (4.3)$$

onde

$$\bar{\theta} = \sum_{i=1}^N \tilde{\omega}_t^{(i)} \theta^{(i)}, \quad \mathbf{V} = \sum_{i=1}^N \tilde{\omega}_t^{(i)} (\theta^{(i)} - \bar{\theta})(\theta^{(i)} - \bar{\theta})^t \quad \text{e} \quad h^2 = 1 - a^2.$$

Esta estratégia culmina em um estimador não paramétrico da densidade $p(\theta|y_{1:t})$, o qual é isento de hipóteses adicionais a respeito da forma funcional do modelo e garante o controle da variabilidade das estimativas através do parâmetro de encolhimento $a \in (0, 1)$. Estimativas mais robustas podem ser obtidas em casos particulares onde a densidade $p(\theta|x_{1:t}, y_{1:t})$ é sumarizada precisamente a partir de um vetor de estatísticas suficientes $\mathcal{S}_t(x_{1:t}, y_{1:t})$ (GILKS; BERZUINI, 2001; STORVIK, 2002; FEARNHEAD, 2002; CARVALHO *et al.*, 2010). Neste caso, a incerteza *a posteriori* é definida a partir de uma aproximação de Monte Carlo

$$\begin{aligned} \hat{p}(\theta|y_{1:t}) &= \int_{\Omega^t} p(\theta, x_{1:t}|y_{1:t}) dx_{1:t} \\ &= \int_{\Omega^t} p(\theta|\mathcal{S}_t(x_{1:t}, y_{1:t})) \left(\int_{\Theta} f(x_{1:t}, \theta'|y_{1:t}) d\theta' \right) dx_{1:t} \\ &\approx \int_{\Omega^t} p(\theta|\mathcal{S}_t(x_{1:t}, y_{1:t})) \left(\int_{\Theta} \sum_{i=1}^N \tilde{\omega}_t^{(i)} \delta_{(x_{1:t}, \theta')^{(i)}}(x_{1:t}, \theta') d\theta' \right) dx_{1:t} \\ &= \int_{\Omega^t} p(\theta|\mathcal{S}_t(x_{1:t}, y_{1:t})) \left(\sum_{i=1}^N \tilde{\omega}_t^{(i)} \delta_{x_{1:t}^{(i)}}(x_{1:t}) \right) dx_{1:t} \\ &\approx \sum_{i=1}^N \tilde{\omega}_t^{(i)} p\left(\theta|\mathcal{S}_t(y_{1:t}, x_{1:t}^{(i)})\right) \end{aligned} \quad (4.4)$$

alimentada pela nuvem de partículas proveniente do dispositivo numérico de filtragem aumentada.

Apesar do grande potencial já evidenciado experimentalmente em diversos trabalhos, é possível demonstrar que os resultados alcançados por estes algoritmos ainda não são completamente satisfatórios por não desmenbrar características próprias das metodologias de filtragem pura daquelas associadas aos métodos de amostragem de parâmetros estáticos. Desta forma, temos até então apenas análises simplificadas as quais tendem a associar resultados negativos oriundos das metodologias de reamostragem, assim como de más escolhas de distribuições propostas, aos dispositivos de amostragem de parâmetros. Tais perspectivas geram uma atmosfera de bastante desconfiança em torno dos algoritmos, sobretudo com respeito ao filtro de Liu e West. Além disso, destacamos que nenhum argumento formal na literatura justifica a replicação perpétua do passo de rejuvenescimento da população estática. A utilização desmoderada dos dispositivos de amostragem de parâmetros pode acrescentar um montante de ruído desnecessário ao problema da filtragem além de elevar o custo computacional do algoritmo.

Buscando estabelecer uma análise mais frutífera de tais métodos, a perspectiva proposta neste trabalho identifica, independentemente do algoritmo de filtragem, os legítimos vícios e virtudes das metodologias básicas de amostragem de parâmetros no contexto sequencial. Dentre as principais contribuições extraídas da nova análise, destacamos: o desenvolvimento de uma agenda dinâmica de rejuvenescimento capaz de conferir mais eficiência aos métodos, o estabelecimento de uma estratégia de reamostragem suavizada que alivia possíveis efeitos de degeneração e a determinação de critérios objetivos para a calibragem de parâmetros do algoritmo de Liu e West. Uma série de experimentos computacionais é explorada a fim de evidenciar a efetividade das novas propostas.

4.2 Fator de Fertilidade e Reamostrador Ideal

O monitoramento da degeneração amostral constitui um fator de extrema relevância no âmbito dos métodos sequenciais de Monte Carlo. A abordagem mais frequente apoia-se no rastreamento dos tamanhos efetivos amostrais a cada iteração do processo de filtragem. Este procedimento é bastante efetivo e seu sucesso pode ser explicado por meio de dois argumentos:

matematicamente, os tamanhos efetivos amostrais são inversamente proporcionais à distância chi-quadrado entre as medidas alvo e proposta no passo de amostragem por importância. Tamanhos efetivos amostrais pequenos indicam que o modelo proposto difere muito do objetivo, favorecendo a geração de amostras ponderadas empobrecidas.

informalmente, esta medida pode ser entendida como o número de indivíduos necessários em uma população iid para garantir o mesmo montante de informação oferecido pela amostra ponderada. Tamanhos efetivos amostrais pequenos indicam neste caso que a população ponderada conta com muitos indivíduos de baixa relevância.

Uma forma alternativa de avaliar efeitos de degeneração pode ser estabelecida a partir do monitoramento da diversidade amostral. O acompanhamento deste atributo equivale à observação sequencial dos *fatores de fertilidade* (FF), uma medida introduzida por Baker (1987) no contexto dos algoritmos genéticos. O conceito consiste de contar o número de indivíduos selecionados para reprodução futura

$$FF = \frac{\# \text{ indivíduos distintos selecionados}}{\# \text{ indivíduos na população}}, \quad (4.5)$$

o que nos dá uma bom parâmetro a respeito do montante de informação disponível (capacidade de exploração espacial) para o desenvolvimento futuro da população. É evidente que as quantidades pontuais expressas por meio dos FFs são fortemente dependentes da metodologia de seleção, contudo podemos mostrar que, independentemente do dispositivo de seleção empregado, há uma forte correlação entre os fatores de fertilidade e os tamanhos efetivos amostrais (Apêndice A).

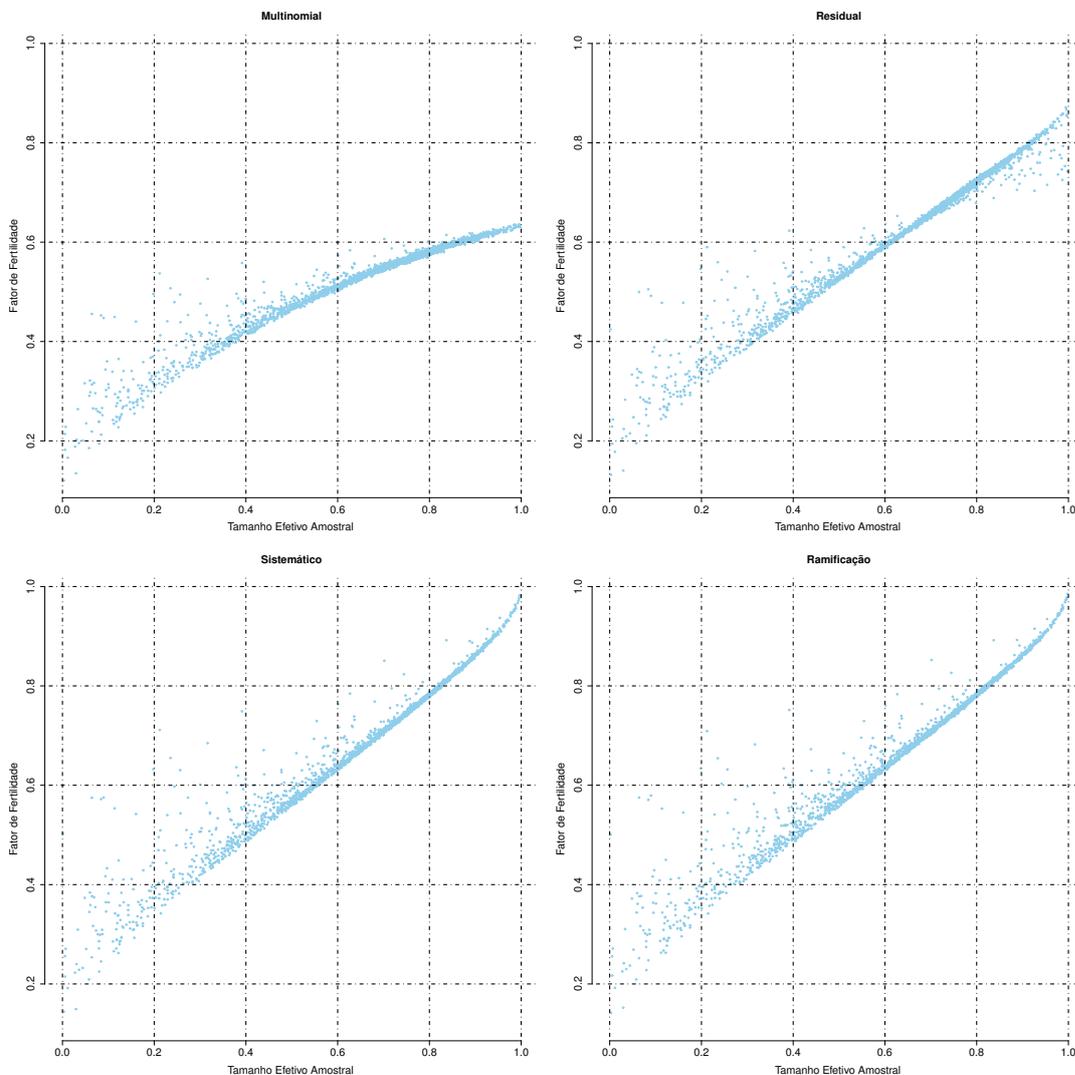


Figura 4 – Fatores de Fertilidade × Tamanhos Efetivos Amostrais: comparação entre os mecanismos Multinomial, Residual, Sistemático e por Ramificação (escala percentual). O ensaio apresenta os fatores de fertilidade obtidos para 2500 medidas aleatórias independentes, sendo o suporte de cada uma composto por 10000 pontos.

O ensaio ilustrado na Figura 4 evidencia o elo entre as duas medidas considerando alguns dos principais esquemas de seleção encontrados na literatura: multinomial, residual, sistemático e por ramificação. A amostra estudada neste experimento é baseada na construção de 5000 medidas aleatórias independentes, cada qual obtida por meio de um passo autônomo de amostragem por importância. Cada medida aleatória é associada a um único tamanho efetivo amostral e a quatro fatores de fertilidade correspondentes aos mecanismos de reamostragem em questão. Além de evidenciar a forte relação entre tamanho efetivo amostral e fator de fertilidade, o experimento resumido pelos diagramas de dispersão ilustra um relevante aspecto a respeito da escolha dos mecanismos de reamostragem: capacidade de preservação da diversidade populacional. Para uma mesma medida aleatória, independentemente de seu TEA associado, os mecanismos de escolha residual, sistemático e por ramificação impulsionam uma menor pressão seletiva na nuvem de partículas, garantindo um preenchimento mais rico do suporte da distribuição de interesse após o passo de reamostragem. Esta particularidade é usualmente ignorada em análises de degeneração tradicionais pois os TEAs concentram apenas informações com respeito às medidas aleatórias, deixando de lado o que ocorre após a reamostragem. Nestes casos, a escolha do dispositivo de reamostragem é deixada em segundo plano e o dispositivo de seleção multinomial acaba sendo adotado por mera conveniência.

A compreensão alternativa de rejuvenescimento artificial como um procedimento de regularização mais geral, torna evidente o possível déficit na informação disponível para a construção de aproximações da densidade *a posteriori* $p_t(\theta|y_{1:t})$ caso as populações $\{\theta^{(i)}\}$ e $\{\mathcal{S}_t(y_{1:t}, x_{i:t}^{(i)})\}$ passem por um processo acelerado de enxugamento. O método de Liu e West, em sua forma tradicional, é reconhecidamente sensível a este fenômeno, de modo que sua aplicação é usualmente recomendada com cautela. Em se tratando da metodologia PL, este problema é mascarado devido à possibilidade de rejuvenescimento da população de parâmetros por meio de dispositivos de simulação direta. Nestes casos, boa parte dos autores ignora o fato de que a população de estatísticas suficientes também é desprovida de dinâmica natural e exploram apenas a metodologia de reamostragem multinomial. O próximo exemplo ilustra o quanto esta perspectiva pode ser desastrosa.

Exemplo 1. *Kantas et al. (2015)* estudam uma instância do modelo linear dinâmico

$$\begin{cases} y_t = x_t + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma^2) \\ x_t = \beta x_{t-1} + \eta_t, & \eta_t \sim N\left(0, \frac{1}{10}\right), t = 1, 2, 3, \dots, 5000, \\ x_0 \sim N(0, 1) \end{cases}$$

cujo vetor de parâmetros desconhecidos é $\theta = (\beta, \sigma^2)$. O experimento consiste em analisar o comportamento do algoritmo PL considerando $\theta = \left(\frac{1}{2}, 1\right)$ e o par independente de distribuições *a priori*

$$\beta \sim U(-1, 1) \quad \text{e} \quad \sigma^2 \sim GI(1, 1),$$

a fim de demonstrar experimentalmente o comportamento degenerativo desta metodologia. O rejuvenescimento dos indivíduos estáticos é realizado sequencialmente a partir da amostragem

direta dos modelos *a posteriori* parciais

$$\beta|x_{1:t}, y_{1:t} \sim N_{(-1,1)} \left(\frac{\sum_{i=1}^t x_i x_{i-1}}{\sum_{i=1}^t x_{i-1}^2}, \frac{1}{10 \cdot \sum_{i=1}^t x_{i-1}^2} \right) \quad \text{e} \quad \sigma^2|x_{1:t}, y_{1:t} \sim GI \left(\frac{2+t}{2}, \frac{2 + \sum_{i=1}^t (y_i - x_i)^2}{2} \right),$$

o qual pode ser atualizado sequencialmente em conjunto com as demais quantidades do algoritmo de filtragem a partir do vetor de estatísticas suficientes

$$S_t = \left(\sum_{i=1}^t x_i x_{i-1}, \sum_{i=1}^t x_{i-1}^2, \sum_{i=1}^t (y_i - x_i)^2 \right).$$

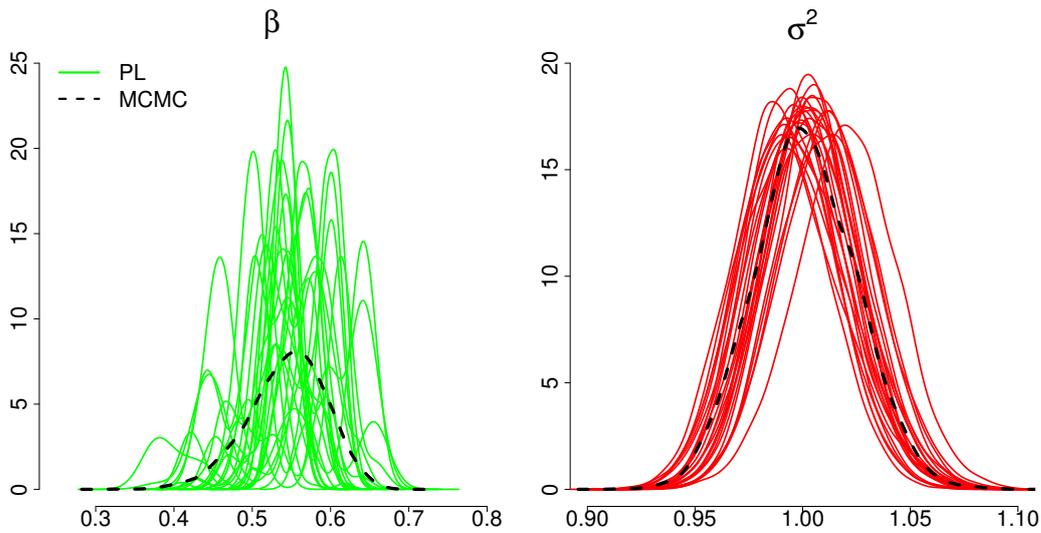


Figura 5 – Distribuições *a posteriori* de θ obtidas a partir de 25 execuções do algoritmo PL considerando reamostragem multinomial e uma nuvem de 10000 partículas.

De fato, se considerarmos a escolha mais trivial de reamostragem (multinomial), o enxugamento drástico da população de estatísticas suficientes acarretará um processo de degeneração acelerado do algoritmo conforme ilustrado na Figura 5. Este fenômeno é acorbetado em situações onde a região mais relevante do espaço paramétrico é pequena – $\beta \in (\frac{9}{10}, 1)$, por exemplo – caracterizando distribuições com alto poder discriminativo com respeito às quantidades desconhecidas. Nestes casos, mesmo após o esgotamento das estatísticas suficientes, ainda temos capacidade de identificar corretamente os parâmetros de interesse. Na configuração apresentada neste exemplo, mesmo a adoção de uma população relativamente grande, 10000 indivíduos, não é capaz de assegurar uma estimativa razoável a todos os parâmetros. Ressaltamos apenas que a sequência de distribuições *a posteriori* do parâmetro de volatilidade σ^2 depende diretamente da estatística suficiente

$$S_3 = \sum_{i=1}^t (y_i - x_i)^2,$$

a qual é função do conjunto de observações. Como a sequência de observações é isenta de degeneração, temos que este parâmetro sofre uma menor influência da perda de diversidade das demais quantidades.

A preservação dos fatores de fertilidade associados ao método de reamostragem por ramificação pode ser explorada com o intuito de contornar este problema. Ao manter a população de estatísticas suficientes mais diversificada, estamos implicitamente preservando um maior número de trajetórias filtradas distintas, o que assegura uma cobertura mais honesta da parte não observada do modelo. Uma sequência de estimativas do filtro PL apoiado pelo método da ramificação e por uma nuvem de apenas 5000 partículas é apresentada na Figura 6: o ganho é evidente.

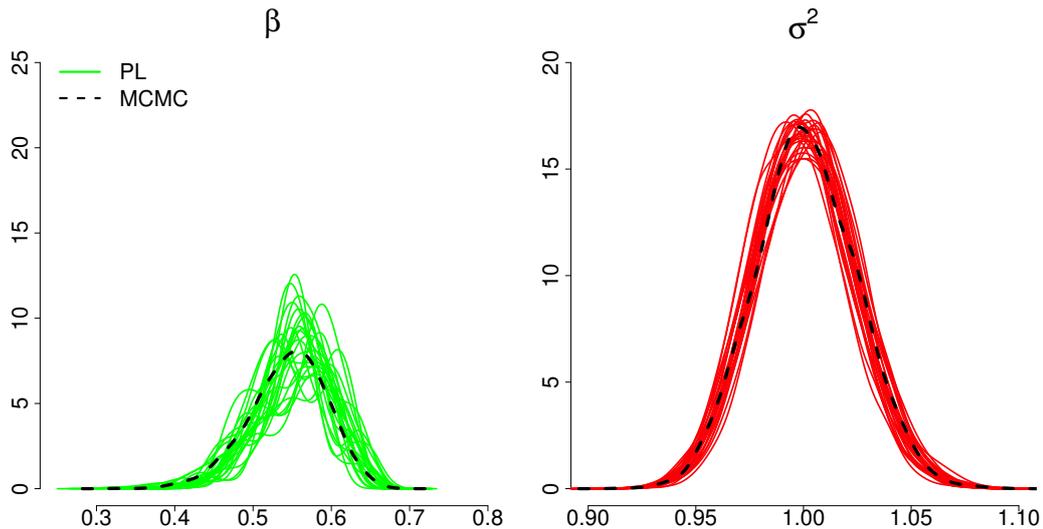


Figura 6 – Distribuições a *posteriori* de θ obtidas a partir de 25 execuções do algoritmo PL considerando reamostragem por ramificação e uma população de 5000 partículas.

Apesar de observarmos um ganho incontestável, verifica-se ainda um bom montante de irregularidade nas estimativas finais, o que é natural neste caso já que o número de indivíduos adotado foi relativamente pequeno para um problema envolvendo uma série tão longa de observações. É importante ressaltar que estamos lidando com aproximações por importância, o que garante a obtenção de resultados mais eficientes porém seguramente menos precisos que métodos baseados na estratégia MCMC.

De uma forma geral, temos que a introdução do fator de fertilidade como medida auxiliar no monitoramento de fenômenos de degeneração demonstrou-se bastante útil para o desenvolvimento de uma análise mais detalhada acerca dos métodos sequenciais implementados na presença de parâmetros estáticos desconhecidos. O entendimento desta grandeza indica que conclusões fundadas na utilização da reamostragem multinomial podem não ser confiáveis e, portanto, devem ser revistas. Destacamos, por exemplo, que [Chopin, Jacob e Papaspiliopoulos \(2013\)](#) tomam como referência o estudo deste mesmo problema, porém considerando a versão discutida no trabalho de [Doucet et al. \(2009\)](#), para acentuar a desconfiança acerca dos modelos auto-organizativos e assim apresentar mais uma justificativa para a construção da metodologia SMC². O estabelecimento de uma análise mais rica deixa claro que estratégias de tamanha

complexidade devem ser utilizadas apenas em cenários onde a degeneração é extrema; a simples associação dos algoritmos PL e LW a um passo de reamostragem mais eficaz já permite a aplicação de tais metodologias em um cenário bastante geral.

4.3 Agenda Ótima de Rejuvenescimento

Metodologias convencionais apoiadas por modelos auto-organizativos assumem que o passo de rejuvenescimento artificial deve ser integrado ao algoritmo de filtragem como uma iteração natural, exigindo assim a execução rigorosa deste procedimento a cada instante de evolução. Todavia, uma perspectiva de filtragem mais eficiente pode ser fundamentada por uma simples reflexão: *caso a população estática não passe por um processo de enxugamento, há necessidade de revitalizá-la?* Seguindo este ponto de vista alternativo, buscamos desenvolver uma agenda dinâmica de amostragem a qual permite que a fase de rejuvenescimento seja explorada apenas em situações caracterizadas por um processo de degeneração mais acentuado. O objetivo é conferir simultaneamente mais eficiência e precisão aos dispositivos numéricos de filtragem.

Filtro Bootstrap

O argumento acima sugere de imediato o desenvolvimento de uma estratégia alternativa de rejuvenescimento artificial. Basicamente, é possível vincular o esquema de revitalização artificial ao passo de reamostragem, o qual pode ser executado apenas em instantes com respeito aos quais o TEA torna-se menor que uma determinada cota inferior γ . Esta estratégia define naturalmente uma sequência de tempos aleatórios de amostragem, dando origem a uma versão complementar do filtro *Bootstrap* adaptativo (Algoritmo 12) capaz de acomodar a estrutura dos parâmetros desconhecidos. A nova perspectiva é condizente com a interpretação dos dispositivos de rejuvenescimento artificial como regularizações da forma (4.2) aplicadas a medidas aleatórias. Assim, a amostragem de parâmetros será executada apenas em situações onde o fenômeno de degeneração pode ser observado (após reamostragem), o que gera uma boa economia de recursos computacionais além de evitar a inserção desnecessária de ruídos ao problema.

Nossa abordagem é baseada na calibragem mais frequente do método adaptativo, isto é, $\gamma = \frac{1}{2}$ (ARULAMPALAM *et al.*, 2002; CHOPIN, 2002; LOPES; TSAY, 2011; CHOPIN; JACOB; PAPASPILIOPOULOS, 2013; LI *et al.*, 2014). Além disso, o núcleo de transição K adotado no pseudocódigo acima pode ser interpretado como um passo genérico de rejuvenescimento, podendo assumir as características das abordagens LW ou PL; a escolha dependerá exclusivamente da forma funcional do modelo. Reforçamos que nosso objetivo é dissociar as ações do método de amostragem de parâmetros daquelas ligadas ao algoritmo de filtragem pura, logo todos os procedimentos desenvolvidos daqui em diante farão uso do núcleo K para indicar genericamente a fase de amostragem de parâmetros estáticos.

Algoritmo 20: Aprendizagem de Parâmetros *Bootstrap* com Agenda Dinâmica**Entrada:** Tamanho amostral (N), distribuições *a priori* (p_0, q_0) e proposta (q);**Saída:** Aproximação discreta $\hat{\pi}_n(dz_{1:n}) = \frac{1}{N} \sum_{i=1}^N \delta_{(z_{1:n}^{(i)})}(dz_{1:n})$;

```

1 para  $i \leftarrow 1$  até  $N$  faça
2   Amostre  $X_0^{(i)} \sim q_0$ ;
3   Amostre  $\theta^{(i)} \sim p_0$ ;
4    $\tilde{\omega}_0^{(i)} \leftarrow \frac{1}{N}$ ;
5 para  $t \leftarrow 1$  até  $n$  faça
6   para  $i \leftarrow 1$  até  $N$  faça
7     Amostre  $\tilde{X}_t^{(i)} \sim q(\cdot | X_{t-1}^{(i)}, y_t, \theta^{(i)})$ ;
8      $\omega_t^{(i)} \leftarrow \tilde{\omega}_{t-1}^{(i)} \cdot g(y_t | \tilde{X}_t^{(i)}, \theta^{(i)}) f(\tilde{X}_t^{(i)} | x_{t-1}^{(i)}, \theta^{(i)}) / q(\tilde{X}_t^{(i)} | X_{t-1}^{(i)}, y_t, \theta^{(i)})$ ;
9   Normalize  $\tilde{\omega}_t^{(i)} \leftarrow \omega_t^{(i)} / \sum_{k=1}^N \omega_t^{(k)}$   $i = 1, 2, \dots, N$ ;
10  se  $TEA(\tilde{\omega}_t^{(1)}, \dots, \tilde{\omega}_t^{(N)}) < \frac{1}{2}$  então
11     $\{j_1, j_2, \dots, j_N\} \leftarrow \text{amostra\_índices}(N, \{\tilde{\omega}_t^{(i)}\}_{i=1}^N)$ ;
12    para  $i \leftarrow 1$  até  $N$  faça
13       $\tilde{\omega}_j^{(i)} \leftarrow \frac{1}{N}$ ;
14       $X_{1:t}^{(i)} \leftarrow (X_{1:t-1}^{(j_i)}, \tilde{X}_t^{(j_i)})$ ;
15      Amostre  $\theta^{(i)} \sim K(\cdot | \theta^{(j_i)}, X_{1:t}^{(i)}, y_{1:t})$ ;
16 para  $i \leftarrow 1$  até  $N$  faça
17    $Z_{1:n}^{(i)} \leftarrow (X_{1:n}^{(i)}, \theta^{(i)})$ ;
18 retorna  $\{Z_{1:n}^{(i)}, \frac{1}{N}\}$ ;
```

Exemplo 2. O modelo linear dinâmico

$$\begin{cases} y_t = x_t + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma^2) \\ x_t = \alpha + \beta x_{t-1} + \eta_t, & \eta_t \sim N(0, \tau^2), \quad t = 1, 2, 3, \dots, 1000, \\ x_0 = \frac{\alpha}{1-\beta} \end{cases}$$

onde $\theta = (\alpha, \beta, \tau^2, \sigma^2)$ é um vetor de parâmetros fixos e desconhecidos, foi estudado por [Lopes e Tsay \(2011\)](#) com o intuito de comparar a capacidade inferencial dos algoritmos LW, de Storvik e PL (Apêndice C). A análise considerou $\theta = (0, 0; 0, 9; 0, 5; 1, 0)$ em conjunto com as seguintes distribuições *a priori*

$$\begin{aligned} \sigma^2 &\sim GI\left(\frac{10}{2}, \frac{10 \cdot 1}{2}\right), & \tau^2 &\sim GI\left(\frac{10}{2}, \frac{10 \cdot 0,5}{2}\right), \\ \alpha, \beta | \tau^2 &\sim N\left(\begin{bmatrix} 0 \\ 0,9 \end{bmatrix}, \tau^2 \cdot \mathbf{I}\right), & x_0 &\sim N(0, 10). \end{aligned}$$

O esquema de conjugação adotado para a incerteza inicial permite que o algoritmo PL atualize sequencialmente o vetor de parâmetros segundo um modelo Bayesiano de regressão linear.

Para este exemplo consideramos o filtro de Liu e West convencional e sua versão obtida a partir da adoção da agenda dinâmica de rejuvenescimento. Para cada algoritmo, realizamos 100

execuções independentes com 5000 partículas. Além disso, executamos o filtro de Liu e West convencional com 1000000 de partículas a fim de obter uma referência para comparação. O resumo do desempenho apresentado pelos dois métodos (Figura 7) evidencia que a agenda dinâmica não causa nenhuma perda inferencial ao dispositivo; as trajetórias filtradas permanecem praticamente inalteradas. Entretanto, do ponto de vista computacional, é possível verificar um ganho bastante significativo já que, neste caso específico, o passo de reamostragem/rejuvenescimento foi realizado em média ao longo de 20% das iterações. Utilizando a função *microbenchmark* do pacote *microbenchmark* disponível para a linguagem R, foi possível determinar com precisão o tempo de execução de cada metodologia: o filtro implementado a partir da agenda dinâmica apresentou uma redução de 30% no tempo de execução (os tempos absolutos são irrelevantes pois dependem necessariamente do hardware utilizado).

O ganho computacional variará de acordo com a natureza do modelo, assim como oscilará em diferentes bancos de dados dentro de uma mesma modelagem, pois o gatilho de amostragem é dinâmico. Problemas mais suscetíveis a degeneração farão mais uso do dispositivo, dissipando uma maior quantidade de recursos computacionais. Ganhos similares são observados caso a estratégia LW seja substituída pelo esquema de amostragem PL. Neste caso, o ganho é ainda mais significativo pois além de não aplicar o passo de rejuvenescimento perpetuamente, também não precisamos reamostrar os índices correspondentes ao vetor de estatísticas suficientes (dimensão 8 neste caso particular). Uma avaliação experimental mais completa considerando a estratégia PL assim como variações no vetor de parâmetros será apresentada no próximo capítulo. A ideia deste primeiro exemplo é apenas ilustrar o quanto as estratégias usuais exploram de maneira ineficiente os dispositivos de filtragem.

Filtro Auxiliar

O esquema dinâmico desenvolvido acima é vinculado à possibilidade de não executarmos o passo de reamostragem em algumas iterações, uma perspectiva impraticável no paradigma do filtro auxiliar já que a ação do algoritmo parte de um passo de reamostragem na tentativa de capturar alguma tendência futura nas observações. Por outro lado, voltando à reflexão proposta no primeiro parágrafo, projetamos a revitalização da população de parâmetros desconhecidos exclusivamente em situações caracterizadas por degeneração e não necessariamente após reamostragem. A ênfase neste último ponto se deve à eventualidade de nos abstermos do fenômeno de degeneração mesmo após a realização do passo de reamostragem, o que pode ser determinado a partir de uma análise de diversidade da população. Uma vez que o objetivo é precisar o número de indivíduos distintos no enxame de partículas, temos que a análise pode ser efetuada por meio do rastreamento dos fatores de fertilidade. Ainda que o passo de reamostragem seja exercido, é possível obter fatores de fertilidade altíssimos conforme verificamos no ensaio da Figura 8.

A diversidade da população pode ser acompanhada mesmo com a execução continuada do passo de reamostragem. Suponhamos que no instante $t - 1$ o enxame de partículas propagado

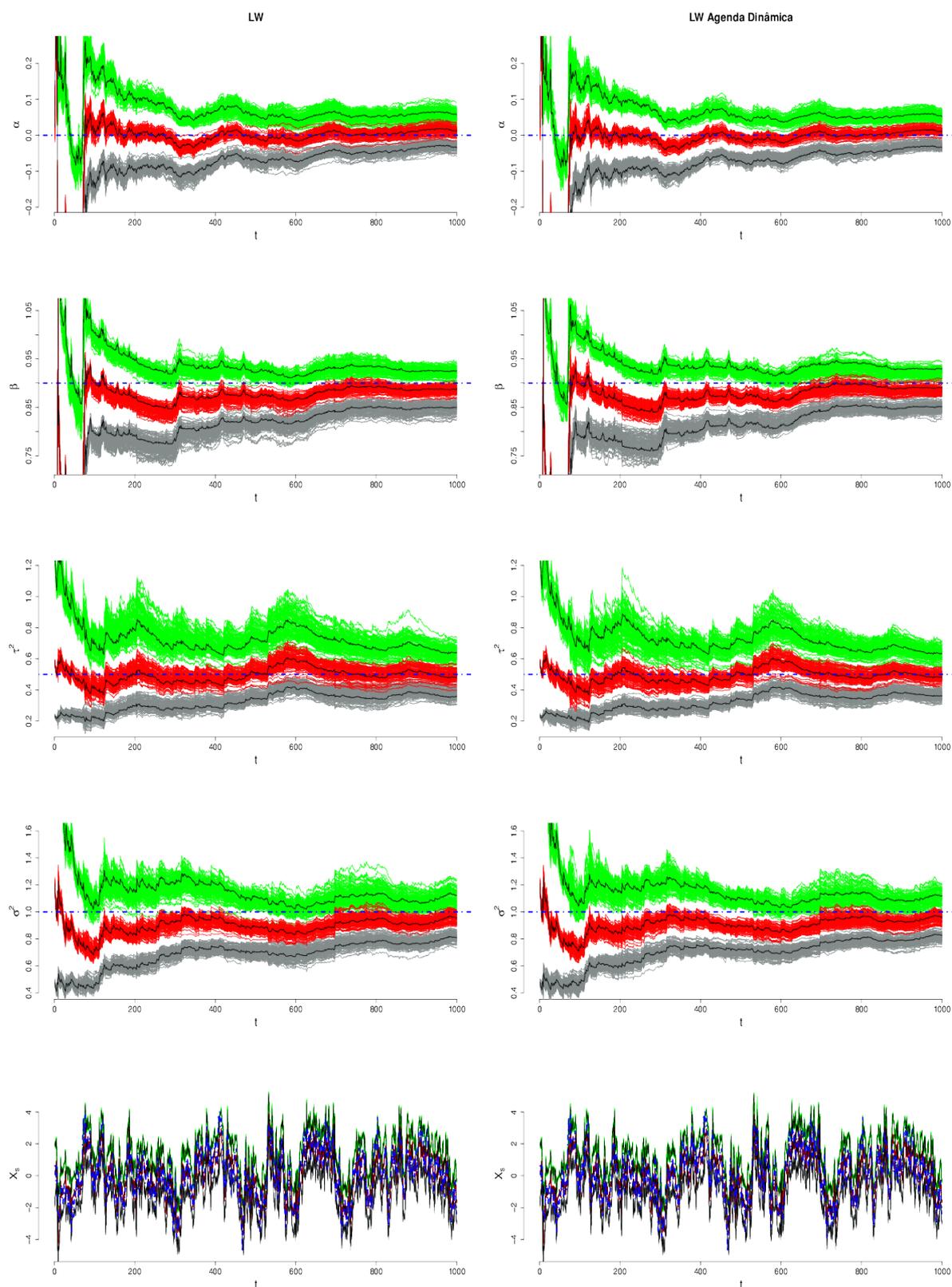


Figura 7 – Resumo de 100 execuções, 5000 partículas cada, dos filtros LW convencional e LW com agenda dinâmica: médias *a posteriori* e bandas de 95% de credibilidade. Colunas 1 e 2 ilustram, na devida ordem, os resultados obtidos pelos métodos LW e LW dinâmico. As curvas em preto correspondem ao resultado de referência: inferência obtida pelo algoritmo LW convencional a partir de 100000 pontos.

pelo algoritmo de filtragem tenha passado por um processo de rejuvenescimento artificial. A população no passo imediatamente seguinte será composta então por N indivíduos distintos, caracterizando um fator de diversidade de 100%. Se desconsiderarmos a fase de revitalização a partir do instante t , temos que a diversidade da população após k passos de seleção pode ser estimada pelo coeficiente de diversificação

$$D := \prod_{j=i}^k FF_{t+j}, \quad (4.6)$$

onde FF_i corresponde ao fator de fertilidade amostral associado à fase de seleção no instante i . O conceito de diversidade associado à quantidade D busca aproximar, mesmo que grosseiramente, o número de indivíduos distintos ainda presentes na nuvem de pontos após k passos de seleção. Notemos que, a rigor, é possível determinar com exatidão a diversidade da amostras contando explicitamente o número de indivíduos distintos, contudo este procedimento afetaria significativamente o custo computacional do algoritmo. A partir do rastreamento deste coeficiente, podemos desenvolver uma agenda dinâmica de rejuvenescimento artificial adequada à metodologia do filtro auxiliar: basta exercermos a amostragem de parâmetros quando D violar uma cota inferior γ . Assim como a agenda dinâmica de rejuvenescimento adaptada ao filtro *bootstrap* anteriormente, consideramos a cota $\gamma = \frac{1}{2}$ de modo que o novo algoritmo também faça uso de mecanismos de amostragem paramétrica exclusivamente em situações onde o conjunto de partículas demonstra um comportamento verdadeiramente degenerativo.

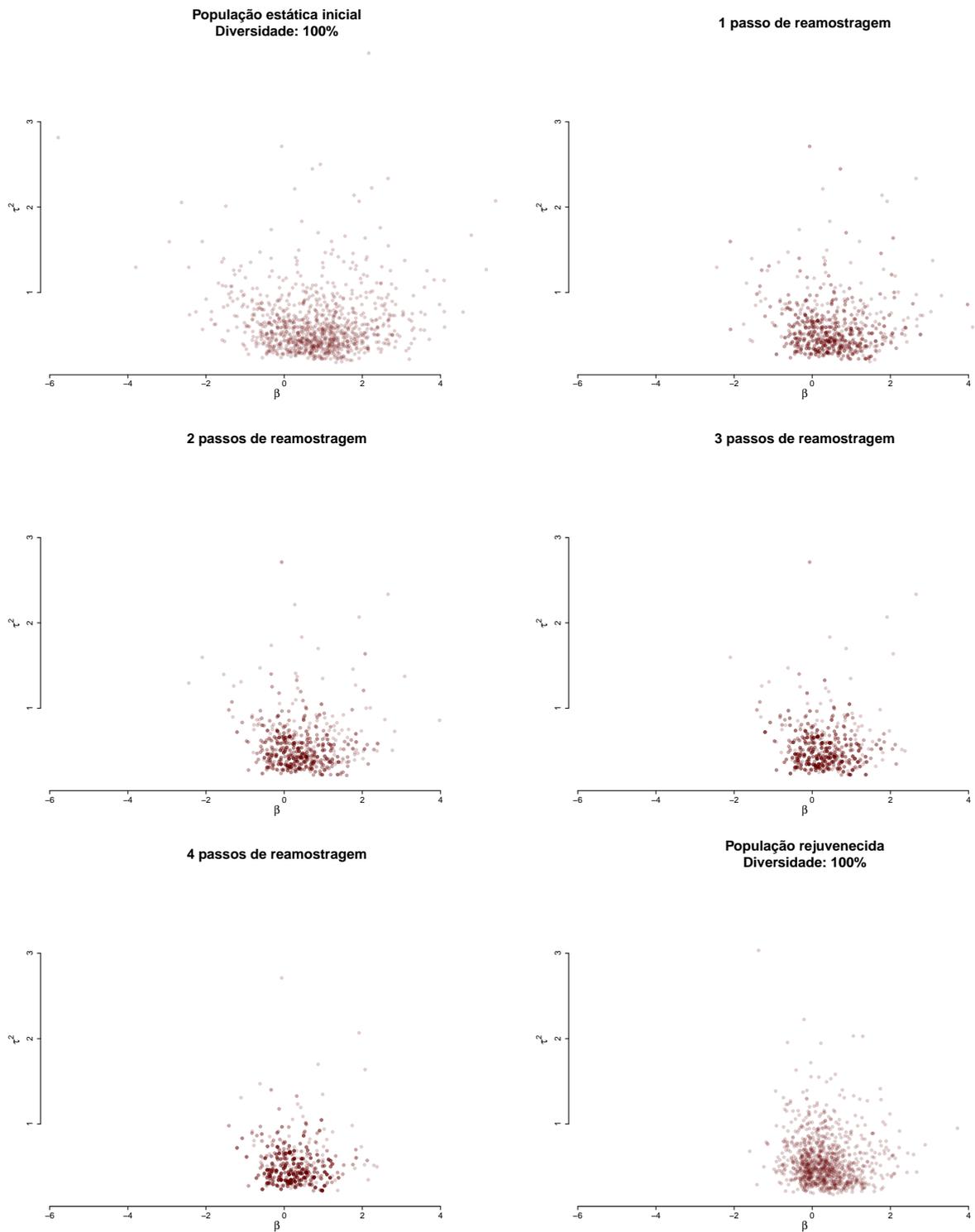


Figura 8 – Configuração espacial da população estática $\{\beta^{(i)}, \tau^{2(i)}\}_{i=1}^N$ no Exemplo 2. A sequência de diagramas ilustra a evolução dos indivíduos após os 4 primeiros passos de reamostragem sem rejuvenescimento artificial. A nuvem de pontos é capaz de apresentar uma cobertura satisfatória do espaço mesmo na ausência de revitalização. O quinto passo de reamostragem é associado ao dispositivo de regularização (4.4) a fim de recuperar a diversidade.

Algoritmo 21: Aprendizagem de Parâmetros Auxiliar com Agenda Dinâmica**Entrada:** Tamanho amostral (N), distribuições *a priori* (p_0, q_0) e proposta (q);**Saída:** Aproximação discreta $\hat{\pi}_n(dz_{1:n}) = \sum_{i=1}^N \tilde{\omega}_n^{(i)} \delta_{(z_{1:n}^{(i)})}$;

```

1 para  $i \leftarrow 1$  até  $N$  faça
2   Amostre  $X_0^{(i)} \sim q_0$ ;
3   Amostre  $\theta^{(i)} \sim p_0$ ;
4    $\tilde{\omega}_0^{(i)} \leftarrow \frac{1}{N}$ ;
5 para  $t \leftarrow 1$  até  $n$  faça
6    $\{j_1, j_2, \dots, j_N\} \leftarrow \text{amostra\_índices} \left( N, \left\{ \tilde{\omega}_{t-1}^{(i)} \cdot g \left( y_t | h(X_{t-1}^{(i)}) \right) \right\}_{i=1}^N \right)$ ;
7   se  $FF(j_1, j_2, \dots, j_N) < \frac{1}{2}$  então
8     para  $i \leftarrow 1$  até  $N$  faça
9       Amostre  $\tilde{\theta}^{(i)} \sim K(\cdot | \theta^{(j_i)}, X_{1:t}^{(i)}, y_{1:t})$ ;
10    senão
11      para  $i \leftarrow 1$  até  $N$  faça
12         $\tilde{\theta}^{(i)} \leftarrow \theta^{(j_i)}$ ;
13    para  $i \leftarrow 1$  até  $N$  faça
14       $\theta^{(i)} \leftarrow \tilde{\theta}^{(i)}$ ;
15      Amostre  $\tilde{X}_t^{(i)} \sim q(\cdot | X_{t-1}^{(j_i)}, y_t, \theta^{(i)})$ ;
16       $\omega_t^{(i)} \leftarrow g(y_t | \tilde{X}_t^{(i)}, \theta^{(i)}) f(\tilde{X}_t^{(i)} | X_{t-1}^{(j_i)}, \theta^{(i)}) / g(y_t | h(X_{t-1}^{(j_i)})) \cdot q(\tilde{X}_t^{(i)} | X_{t-1}^{(j_i)}, y_t, \theta^{(i)})$ ;
17       $X_{1:t} \leftarrow (X_{1:t-1}^{(j_i)}, \tilde{X}_t^{(i)})$ ;
18    Normalize  $\tilde{\omega}_t^{(i)} \leftarrow \omega_t^{(i)} / \sum_{k=1}^N \omega_t^{(k)}$   $i = 1, 2, \dots, N$ ;
19 para  $i \leftarrow 1$  até  $N$  faça
20    $Z_{1:n}^{(i)} \leftarrow (X_{1:n}^{(i)}, \theta^{(i)})$ ;
21 retorna  $\{Z_{1:n}^{(i)}, \tilde{\omega}_n^{(i)}\}$ ;
```

Exemplo 2 (Continuação). O esquema de aprendizagem de parâmetros auxiliar com agenda dinâmica também foi implementado para o tratamento do modelo linear dinâmico discutido anteriormente. A abordagem PL foi utilizada neste segundo experimento buscando reforçar que nosso conceito independe do método de amostragem; a ideia é assegurar que as abordagens são consistentes com um mesmo resultado apesar de estarem associadas a montantes de variabilidade distintos. De maneira análoga ao primeiro ensaio, consideramos o filtro PL convencional e sua versão obtida a partir da agenda dinâmica sugerida no Algoritmo 21. O resultado ótimo de referência utilizado agora corresponde a uma execução do filtro PL convencional com 1000000 de partículas. A Figura 9 a seguir resume o desempenho apresentado pelos dois métodos.

Assim como no caso anterior, o gatilho dinâmico de rejuvenescimento não trouxe nenhuma influência negativa para a inferência final oferecida pelo dispositivo de filtragem; as trajetórias foram mais uma vez bastante similares na comparação entre métodos. O passo de rejuvenescimento foi executado (em média) em 22,8% das iterações. Já o ganho no tempo

de execução foi de 21,5%, um resultado naturalmente inferior ao primeiro caso (mas ainda bastante significativo) já que agora o passo de reamostragem é executado perpetuamente devido às características do filtro auxiliar.

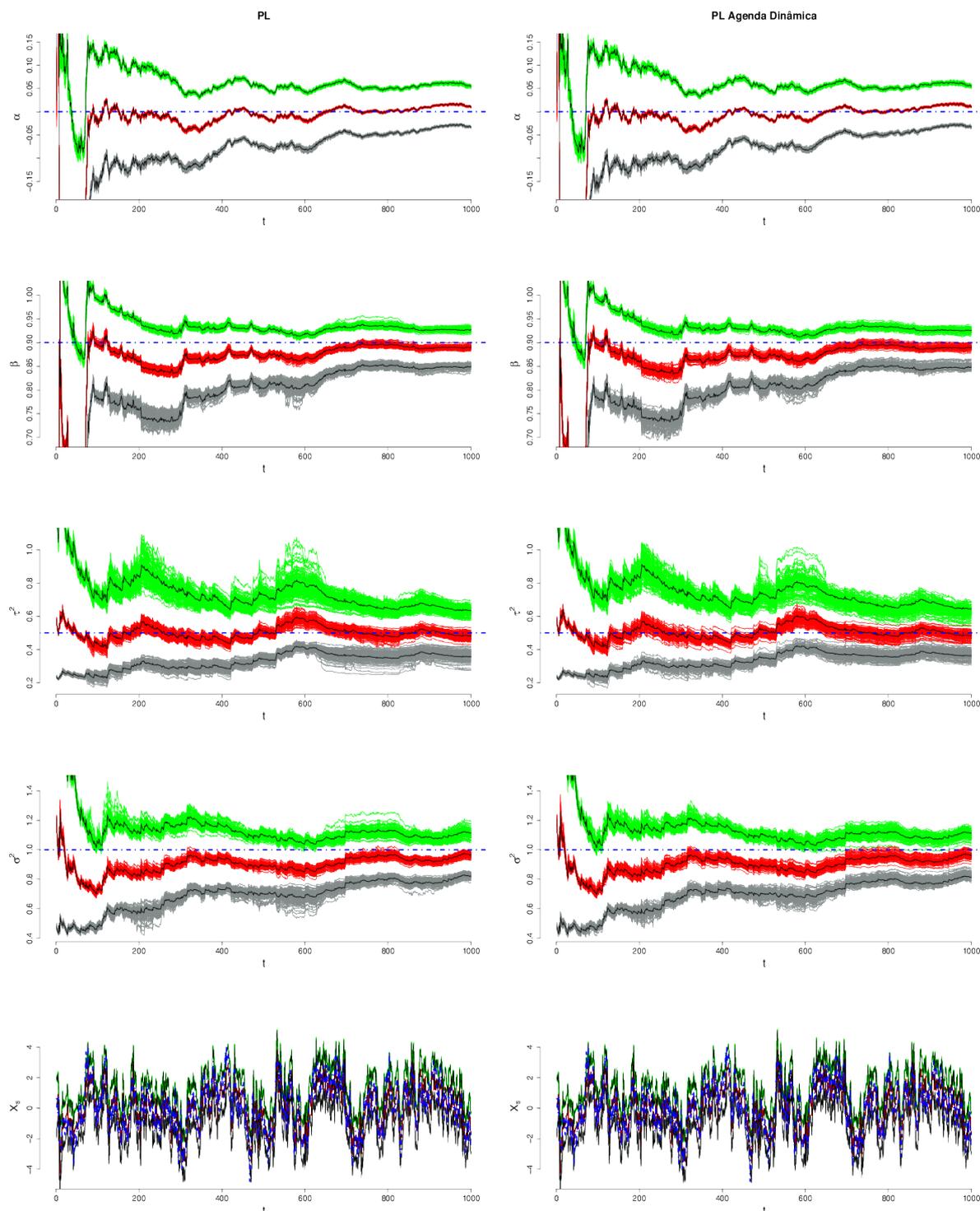


Figura 9 – Resumo de 100 execuções, 5000 partículas cada, dos filtros PL convencional e PL com agenda dinâmica: médias a *posteriori* e bandas de 95% de credibilidade. Colunas 1 e 2 ilustram, na devida ordem, os resultados obtidos pelos métodos PL e PL dinâmico. As curvas em preto correspondem ao resultado de referência: inferência obtida pelo algoritmo PL convencional a partir de 100000 pontos. Curvas em azul correspondem às quantidades reais utilizadas para a simulação do modelo.

4.4 Reamostragem Suavizada

A introdução do conceito de reamostragem no contexto das simulações sequenciais (GORDON; SALMOND; SMITH, 1993) foi responsável pela grande popularização dos filtros de partículas. Este dispositivo tem como finalidade a remoção de partículas de baixa relevância em prol da fração mais adaptada dos indivíduos, evitando assim o problema do empobrecimento amostral. Em linhas gerais, o dispositivo de reamostragem seleciona cópias dos melhores indivíduos para propagação futura e atribui pesos homogêneos a cada uma delas. Seguindo a perspectiva das medidas aleatórias, o método se resume à substituição:

$$\hat{\pi}_t(dz_{1:t}) = \sum_{i=1}^N \tilde{\omega}_t^{(i)} \delta_{z_{1:t}^{(i)}}(dz_{1:t}) \rightsquigarrow \check{\pi}_t(dz_{1:t}) = \frac{1}{N} \sum_{i=1}^N \xi_t^{(i)} \delta_{z_{1:t}^{(i)}}(dz_{1:t}), \quad (4.7)$$

onde $\xi_t^{(i)}$ representa o número de réplicas do indivíduo $z_{1:t}^{(i)}$. Apesar de remediar o problema do empobrecimento amostral, a introdução da reamostragem determinou o surgimento de uma segunda adversidade: enxugamento acentuado da nuvem de partículas ou, simplesmente, degeneração amostral. A fim de combater este revés secundário, uma série de métodos alternativos de reamostragem foram propostos, conforme discutido no Capítulo 3 (Li, Bolic e Djuric (2015) para uma revisão mais ampla), na tentativa de impor uma menor pressão seletiva na população; resultados bastante satisfatórios foram obtidos em função dos novos métodos, entretanto o problema não foi completamente resolvido. Resultados ainda muito longe do ideal são observados sobretudo em aplicações envolvendo algoritmos na presença de parâmetros estáticos desconhecidos, os quais constituem um cenário onde o processo de reamostragem é deixado em segundo plano.

Novas estratégias de combate à degeneração amostral podem ser desenvolvidas a partir de uma interpretação mais ampla que trata os passos de reamostragem (4.7) como instâncias específicas de uma metodologia fundamentada na construção de aproximações da medida discreta $\hat{\pi}_t$ por meio de um passo adicional de amostragem por importância. Suponha que o suporte finito da medida aleatória $\hat{\pi}_t$ seja dado pela família $\chi = \{z_{1:t}^{(i)}\}_{i=1}^N$ e considere um conjunto alternativo de pesos $\gamma = \{\alpha^{(i)}\}_{i=1}^N$. Uma aproximação por importância de $\hat{\pi}_t$ (RUBIN, 1987; LIU; CHEN; LOGVINENKO, 2001) pode ser especificada a partir da amostragem de N elementos $\{z_{1:t}^{(j_1)}, z_{1:t}^{(j_2)}, \dots, z_{1:t}^{(j_N)}\}$ de χ segundo os pesos γ e da subsequente atribuição de pesos

$$\check{\omega}_t^{(j_i)} = \frac{\tilde{\omega}_t^{(j_i)} / \alpha^{(j_i)}}{\sum_{i=1}^N \tilde{\omega}_t^{(j_i)} / \alpha^{(j_i)}}, \quad i = 1, 2, \dots, N, \quad \text{e} \quad \check{\omega}_t^{(i)} \equiv 0, \quad i \notin \{j_1, j_2, \dots, j_N\},$$

a cada indivíduo selecionado, o que resulta na nova medida aleatória:

$$\check{\pi}_t(dz_{1:t}) = \sum_{i=1}^N \xi_t^{(i)} \check{\omega}_t^{(i)} \delta_{z_{1:t}^{(i)}}(dz_{1:t}). \quad (4.8)$$

É evidente que se considerarmos a escolha $\alpha^{(i)} \equiv \tilde{\omega}_t^{(i)}$ no esquema descrito acima, então teremos $\check{\omega}_t^{(i)} \equiv 1/N$ e a aproximação resultante coincidirá com a perspectiva de reamostragem empregada na extensa maioria dos métodos SMC. Versões mais generalizadas podem ser obtidas a partir de

escolhas adequadas para as ponderações secundárias $\alpha^{(i)}$. Pitt e Shephard (1999), por exemplo, sugerem uma especificação capaz de refletir alguma tendência futura com respeito às observações. Já Liu, Chen e Logvinenko (2001) propõem a definição da família γ a partir de funções crescentes dos pesos originais $\tilde{\omega}_t^{(i)}$ a fim de conferir mais equilíbrio às ponderações iniciais sem deturpar a informação original sumarizada pela primeira medida aleatória.

O objetivo deste trabalho é explorar a liberdade de escolha das ponderações $\alpha^{(i)}$ buscando assegurar à nuvem de partículas um balanceamento mais fino entre diversidade e foco. Esta tentativa não é inteiramente nova na literatura já que alguns autores (LIU; CHEN; LOGVINENKO, 2001) sugerem a escolha

$$\alpha^{(i)} = \sqrt{\tilde{\omega}^{(j)}}, \quad j = 1, 2, \dots, N, \quad (4.9)$$

a qual confere um certo grau de normalização aos pesos. Contudo, esta simples abordagem pode ser potencializada de modo a favorecer dois aspectos cruciais do algoritmo:

foco: a aplicação desmoderada do dispositivo acima pode dificultar a eliminação de partículas pouco adaptadas, afetando diretamente o foco do algoritmo. Assim como no caso da revitalização artificial, seria interessante, na tentativa de obter um algoritmo mais efetivo, associar o método a uma coleção de tempos aleatórios ao invés de replicá-lo necessariamente a cada passo de seleção. Neste caso, estabelecemos uma sequência de tempos aleatórios caracterizada por

$$\tau = \left\{ t \in \{1, 2, \dots, n\} \mid \text{TEA} \left(\tilde{\omega}_t^{(1)}, \tilde{\omega}_t^{(2)}, \dots, \tilde{\omega}_t^{(N)} \right) < \frac{1}{10} \right\}, \quad (4.10)$$

de modo a aplicar o passo de reamostragem suavizada apenas em situações indiscutivelmente desfavoráveis à nuvem de partículas. A escolha da cota superior 1/10 na relação acima foi dada experimentalmente e visa especificar uma situação limite onde a aproximação por importância oferece um resultado inferencial minimamente aceitável.

diversidade: a função raiz pode apresentar um desempenho pouco satisfatório com respeito à suavização dos pesos caso a variabilidade dos mesmos seja muito alta já que apenas uma porção pequena da população apresenta pesos significativos neste cenário. Portanto, para conseguir um resultado minimamente favorável, seria necessária a especificação de uma transformação da forma

$$\alpha^{(i)} = \left(\tilde{\omega}^{(i)} \right)^\lambda, \quad \lambda < \frac{1}{2}. \quad (4.11)$$

Nosso objetivo é escolher um expoente λ que assegure a obtenção de medidas aleatórias suavizadas a fim de preservar um fator de diversidade mínimo na população. Mantendo a coerência com a definição da família de tempos aleatórios τ , definimos o expoente de suavização (Apêndice B) como a raiz da equação:

$$\text{TEA} \left(\left(\tilde{\omega}_t^{(1)} \right)^\lambda, \left(\tilde{\omega}_t^{(2)} \right)^\lambda, \dots, \left(\tilde{\omega}_t^{(N)} \right)^\lambda \right) - \frac{1}{10} = 0. \quad (4.12)$$

Neste caso, λ pode ser obtido facilmente através de algum algoritmo de busca de raízes de equações em domínios unidimensionais.

Tanto os tempos aleatórios de suavização quanto os coeficientes λ foram determinados de maneira objetiva através dos tamanhos efetivos amostrais da população. Notemos que o desenvolvimento deste dispositivo é atrelado exclusivamente às características das medidas aleatórias mais atuais na tentativa de prever um comportamento excessivamente degenerativo. Como esta previsão independe de qualquer mecanismo de reamostragem, temos que a exploração dos TEAs em detrimento aos fatores de fertilidade é de fato mais sensata neste contexto. Por fim, destacamos que a abordagem pode ser implementada de maneira direta tanto na perspectiva *Bootstrap* quanto na perspectiva Auxiliar de filtragem, basta substituir o procedimento de reamostragem adequadamente.

Exemplo 3. O modelo AR(1)+“ruído”

$$\begin{cases} y_t = x_t + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma^2) \\ x_t = \beta x_{t-1} + \eta_t, & \eta_t \sim N(0, \tau^2), \quad t = 1, 2, 3, \dots, 300, \\ x_0 \sim N(0, 1) \end{cases}$$

onde $\theta = (\beta, \tau^2, \sigma^2)$ é um vetor de parâmetros fixos e desconhecidos, foi estudado por [Prado e Lopes \(2013\)](#) com o intuito de introduzir uma nova metodologia de filtragem, baseada na propagação de estatísticas suficientes, para modelos da forma AR(p)+“ruído”. A fim de reforçar a importância do novo método, os autores consideraram a versão convencional do método de Liu e West e verificaram um comportamento excessivamente degenerativo por parte do mesmo. Os autores destacaram também que a estratégia LW apresentou baixíssimos tamanhos efetivos amostrais, o que certamente contribuiu de forma determinante para a degeneração observada.

Neste trabalho realizamos um estudo do mesmo modelo porém considerando a proposta de reamostragem suavizada associada ao Algoritmo 21 e à estratégia LW. As especificações sugeridas no trabalho de [Prado e Lopes \(2013\)](#) foram seguidas rigorosamente; número de partículas $N = 2000$, $\theta = (0, 95; 0, 1; 0, 02)$ e distribuições *a priori*

$$x_0 \sim N(0, 1), \quad \beta \sim U(0, 1), \quad \tau^2 \sim GI(0, 01; 0, 01) \quad \text{e} \quad \sigma^2 \sim GI(0, 01; 0, 01).$$

Foram executadas 20 realizações do algoritmo de filtragem com e sem o mecanismo de reamostragem suavizada. Os resultados apresentados na Figura 10 demonstram que a introdução do conceito de reamostragem suavizada assegurou mais estabilidade ao esquema de filtragem. O ganho apresentado é surpreendentemente grande já que o passo de suavização foi realizado apenas 1% das vezes (3 iterações) em todas as 20 realizações. A maior estabilidade oferecida pela suavização também teve um impacto positivo na agenda dinâmica de amostragem: as médias correspondentes às frações de rejuvenescimento utilizadas em ambas as implementações foram respectivamente 12% e 20% para as implementações com e sem a estratégia de suavização. Temos assim um excelente indicativo de que o método de Liu e West pode de fato oferecer boas inferências desde que implementado de maneira adequada.

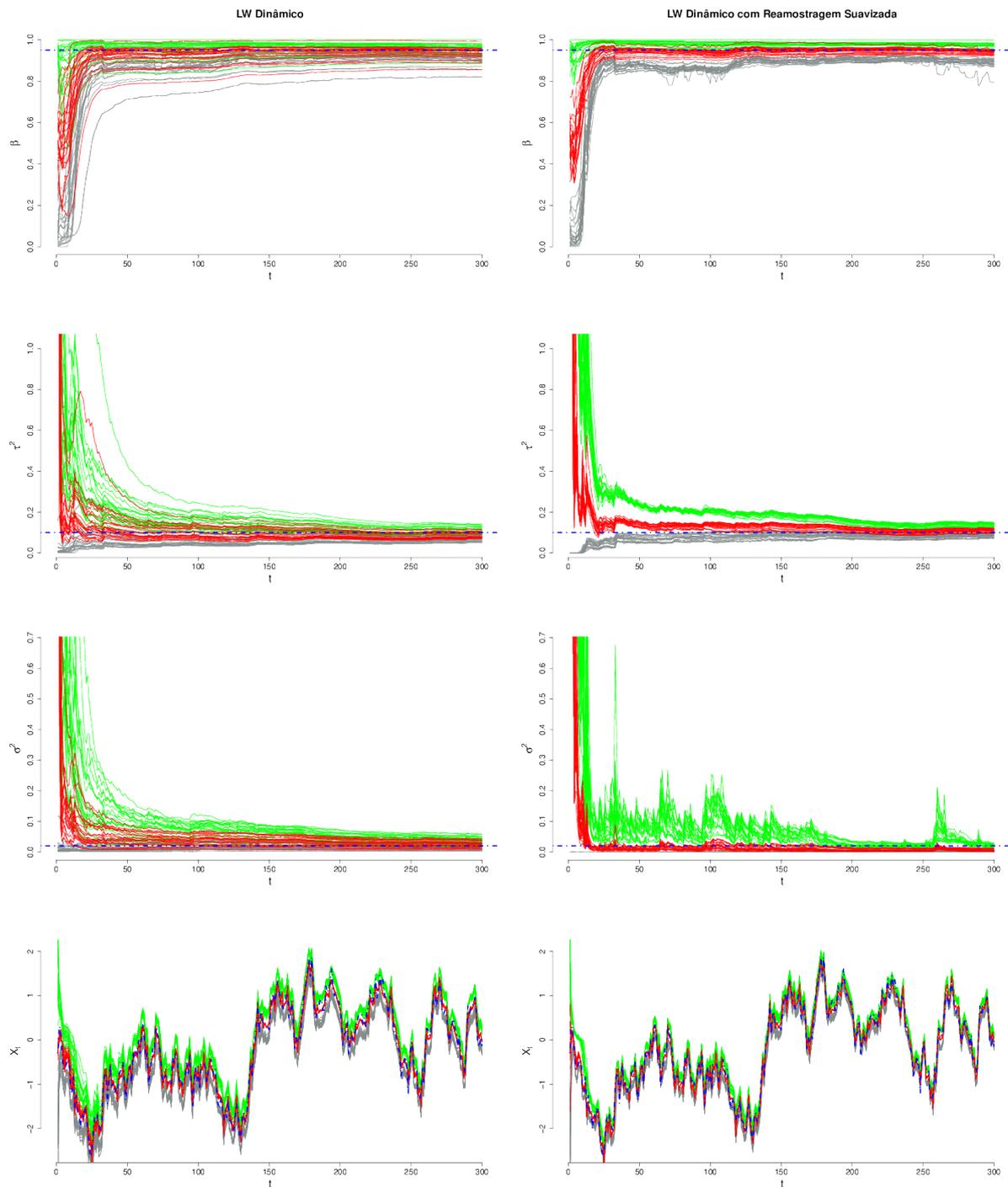


Figura 10 – Resumo de 20 execuções, 2000 partículas cada, do filtro auxiliar dinâmico LW com e sem estratégia de reamostragem suavizada: médias *a posteriori* e bandas de 95% de credibilidade. Colunas 1 e 2 ilustram, na devida ordem, os resultados obtidos sem suavização e com suavização. Curvas em azul correspondem aos valores reais utilizados para a simulação do modelo.

4.5 Janela Ótima × Fator de Encolhimento

A escolha da janela h no método de Liu e West é dada usualmente de maneira indireta através do parâmetro de encurtamento a que, por sua vez, é definido em função de um fator de desconto δ (WEST; HARRISON, 1997; LIU; WEST, 2001). Apesar dos bons resultados já verificados na literatura, esta simples estratégia pode ser considerada subótima por dois fatores:

- a escolha arbitrária de δ torna o algoritmo muito dependente da calibragem de cada usuário, o que inclusive constitui uma das maiores críticas associadas ao método;
- a mera escolha $\delta \in [0.95, 0.99]$ pode não ser uma perspectiva tão satisfatória na associação com estimadores não paramétricos. O tamanho ideal da janela h deve satisfazer, segundo a literatura dos métodos baseados em núcleos estimadores, algumas premissas básicas como manter uma relação inversa de crescimento com respeito ao número de pontos envolvidos, por exemplo. Ao negligenciar algumas destas premissas, a escolha aleatória de δ pode induzir um comportamento degenerativo o qual é normalmente atribuído ao algoritmo como um todo, gerando uma atmosfera de desconfiança com respeito ao método.

Visando eliminar estes dois problemas primários, buscamos construir um critério objetivo capaz de explorar tanto os resultados formais já conhecidos da estatística não-paramétrica quanto a versatilidade oferecida pelo conceito de fator de desconto. Como resultado teremos um método mais eficaz cujos parâmetros podem ser determinados automaticamente de forma simples e objetiva.

O primeiro passo na construção de novas diretrizes é tentar medir o erro e a variabilidade atrelados à aproximação de uma função densidade de probabilidade através de núcleos estimadores. É possível mostrar (SILVERMAN, 1986) que o vício e a variância associados à estimativa não paramétrica de uma densidade f por meio de um núcleo estimador K podem ser aproximados pelas expressões

$$\frac{1}{4}h^4 \left(\int t^2 K(t) dt \right)^2 \int f''(x)^2 dx \quad \text{e} \quad \frac{1}{Nh} f(x) \int K(t) dt, \quad (4.13)$$

respectivamente. Desta forma, temos de imediato duas premissas básicas para a construção de uma estimativa consistente:

$$h \rightarrow 0 \quad \text{e} \quad Nh \rightarrow +\infty \text{ à medida que } N \rightarrow +\infty. \quad (4.14)$$

Em outras palavras, é de fato importante considerar valores pequenos para h a fim de construir boas aproximações em termos de núcleos estimadores, entretanto este valor deve encolher segundo uma taxa superior a N^{-1} . Devemos esperar estimativas de grande variabilidade caso a taxa de decrescimento da janela seja inferior ao esperado, um problema frequentemente observado na forma de degeneração em aplicações do método de Liu e West onde o fator de

encurtamento é muito alto ($a \approx 1$). Infelizmente não há um critério objetivo para balancear a escolha de h em função de N em um contexto genérico, todavia temos algumas sugestões clássicas na literatura as quais apresentam excelentes resultados. Silverman (1986) desenvolveu uma proposta eficaz, a qual é tomada como referência no caso geral:

$$h := \left(\frac{4}{N(2+d)} \right)^{\frac{1}{4+d}}, \quad (4.15)$$

onde d é a dimensão do suporte da função de interesse. Caso a densidade em estudo apresente sinais evidentes de multimodalidade, Silverman sugere ainda a utilização de um janela reduzida pela metade $h' = \frac{1}{2}h$.

Ao definir um novo critério para a escolha direta da janela h , devemos percorrer o caminho inverso para a obtenção do parâmetro de encolhimento a . É preciso ter um mínimo de cautela neste ponto já que uma perda excessiva de informação será verificada caso a especificação de h resulte em fatores de encurtamento inadequados. Em linhas gerais, se escolha inicial de h culminar em valores muito pequenos para a , então teremos um alto grau de desconto e o modelo de rejuvenescimento se aproximará de uma única distribuição Gaussiana multivariada; informações importantes como multimodalidade ou possíveis assimetrias seriam perdidas neste caso. Contudo, dada a relação direta entre os dois parâmetros

$$a = \sqrt{1 - h^2},$$

é possível estabelecer uma cota inferior κ que nos permita amarrar o conceito de fator de desconto aos resultados matemáticos herdados da inferência não-paramétrica. A partir da imposição $a \geq \kappa$,

$$a \geq \kappa \iff \sqrt{1 - \left(\frac{4}{N^*(2+d)} \right)^{\frac{2}{4+d}}} \geq \kappa \iff N^* \geq \frac{4}{(2+d)(1-\kappa^2)^{\frac{4+d}{2}}}, \quad (4.16)$$

asseguramos um controle adequado do fator de desconto além de obtermos um cota inferior N^* para o tamanho amostral que permita a escolha de h segundo a relação (4.15).

A cota mínima $\kappa = 0,9$ foi adotada neste trabalho de modo a obter um equilíbrio razoável entre taxas de encolhimento não muito pequenas (preservando a informação amostral) e tamanhos mínimos amostrais não muito grandes (preservando a eficiência computacional do método). Assim, a expressão (4.16) se reduziu a

$$N^* \geq 4 / \left((2+d) \cdot 0,19^{\frac{4+d}{2}} \right). \quad (4.17)$$

A tabela a seguir apresenta alguns valores de tamanho amostral mínimo N^* em função da dimensão d do problema.

Dimensionalidade (d)	Tamanho amostral mínimo ($\lceil N^* \rceil$)
1	84
2	146
3	268
4	512
5	1006
6	2020
7	4118
8	8503
9	17733
10	37291

Tabela 1 – Tamanhos amostrais mínimos: obtenção de um balanceamento adequado entre fator de encolhimento e janela ótima.

É importante ressaltar que os valores de N^* obtidos da relação (4.17) asseguram apenas que os parâmetros do algoritmo podem ser calibrados de modo a favorecer tanto a perspectiva dos fatores de desconto quanto os resultados estabelecidos no ramo da estatística não paramétrica. Esta determinação não garante resultados de convergência das aproximações e certamente, em aplicações práticas, será interessante utilizar tamanhos amostrais superiores a N^* . Os tamanhos amostrais apresentados acima indicam que este algoritmo pode apresentar problemas em dimensões muito elevadas. Todavia, podemos esperar que aplicações associadas a espaços paramétricos de dimensão inferior a nove apresentem boas aproximações a um custo computacional acessível.

4.6 Estruturas de Dependência

A estrutura de dependência dos parâmetros estáticos presentes em um modelo de filtragem é representada no método de Liu e West através da matriz de covariância amostral da população. Este procedimento é bastante intuitivo e normalmente não é questionado na literatura dos filtros de partículas, sendo considerado o dispositivo ótimo para a construção da estimativa do modelo probabilístico dos parâmetros estatísticos. Esta escolha, no entanto, não é tão imediata no contexto da inferência não-paramétrica. Assim como no caso da definição de uma janela ótima em cenários gerais, não há um critério indiscutível a respeito da estrutura de dependência a ser implementada de uma forma geral nos núcleos estimadores Gaussianos.

O maior atrativo oferecido pela matriz de covariância amostral é a preservação das variâncias marginais dos parâmetros. Esta propriedade é indispensável para o dispositivo de Liu e West pois evita que a informação a respeito da incerteza *a posteriori* seja difundida ao longo das iterações do dispositivo de filtragem. Outra proposta capaz de desempenhar o mesmo objetivo

seria a utilização de uma matriz diagonal formada pelas variâncias marginais dos parâmetros,

$$\begin{pmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix},$$

sugerindo que a evolução dos mesmos pode ser dada de forma independente. Tal proposta é frequentemente explorada na construção de estimativas não-paramétricas e foi implementada por [Flury e Shephard \(2009\)](#) na tentativa de aprimorar o método de Liu e West. A estratégia demonstrou ser capaz de oferecer resultados satisfatórios em alguns casos mas não necessariamente em uma situação geral. [Larranaga e Lozano \(2002\)](#) discutem uma terceira perspectiva, adotada no campo dos algoritmos estimadores de distribuição, que consiste em avaliar a melhor estrutura de dependência dos indivíduos através de medidas de seleção de modelos. Esta perspectiva trata separadamente subgrupos de dimensões menores (blocos construtivos) os quais podem ser atualizados de forma independente ou não, o que sugere a utilização de uma estrutura intermediária situada entre a matriz de covariâncias completa e a matriz diagonal simplificada:

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & 0 & 0 \\ \sigma_{21} & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & \sigma_{34} \\ 0 & 0 & \sigma_{43} & \sigma_4^2 \end{pmatrix}, \quad \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & \sigma_{23} & 0 \\ 0 & \sigma_{32} & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{pmatrix}, \quad \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ 0 & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ 0 & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{pmatrix}, \quad \text{etc.}$$

A estrutura de dependência entre subgrupos que melhor se ajustar aos dados é considerada a ideal para a construção de um modelo probabilístico. Esta última abordagem é bastante promissora porém é inviável no campo dos métodos SMC devido à sua complexidade computacional.

Na ausência de uma estratégia objetiva para a especificação da melhor estrutura de dependências em um contexto geral, sugerimos apenas uma análise qualitativa do modelo de interesse caso a implementação convencional do método de Liu e West não apresente resultados satisfatórios. Esta reflexão é relevante pois a forte correlação entre um subgrupo de parâmetros pode causar problemas de mau condicionamento na matriz de covariâncias, conduzindo o algoritmo a soluções muito instáveis. Nestes casos, podemos testar no mínimo a utilização de uma matriz diagonal e executar novamente o método. A fim de demonstrar a importância desta análise, apresentamos em mais detalhes no [Capítulo 6](#) uma aplicação onde o simples entendimento desta questão nos habilita a transformar um método completamente espúrio em um dispositivo apto a gerar inferências compatíveis com algoritmos MCMC.

ESTUDOS DE SIMULAÇÕES

Os modelos discutidos neste capítulo foram escolhidos de modo a manter a coerência com os principais trabalhos encontrados na literatura e visam justificar experimentalmente, de maneira ampla e objetiva, as proposições tratadas neste trabalho. Em particular, selecionamos aplicações encontradas em algumas das revistas mais importantes da área de modo a reforçar a importância de se realizar uma análise numérica mais detalhada do tópico. Dedicaremos uma maior atenção a abordagem LW pois avaliamos que a otimização de uma metodologia mais simples e mais abrangente constitui uma contribuição mais relevante. Os resultados evidenciam de forma clara que, apesar de simples, nossas estratégias conferem melhorias significativas tanto na eficiência computacional quanto na acurácia dos resultados.

5.1 AR(1) + Ruído

O modelo linear dinâmico

$$\begin{cases} y_t = x_t + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma) \\ x_t = \alpha + \beta x_{t-1} + \eta_t, & \eta_t \sim N(0, \tau^2), \quad t = 1, 2, 3, \dots, n, \\ x_0 \sim N(0, 1) \end{cases}$$

habitualmente denominado *modelo AR(1) + “ruído”*, já foi exaustivamente estudado (LIU; WEST, 2001; POLSON; STROUD; MÜLLER, 2008; PETRIS; PETRONE; CAMPAGNOLI, 2009; PRADO; WEST, 2010; LOPES; TSAY, 2011; PRADO; LOPES, 2013) devido à sua estrutura simplificada, a qual oferece uma ótima possibilidade de estudo comparativo com respeito a diferentes estratégias de filtragem. Nosso objetivo ao analisar este modelo é desmistificar alguns resultados apresentados na literatura, os quais foram desenhados, equivocadamente, de modo a evidenciar a possível degeneração da proposta de Liu e West.

Cenário I: Lopes e Tsay (2011)

Neste primeiro cenário analisamos os resultados apresentados por Lopes e Tsay, os quais são extremamente desfavoráveis à estratégia LW. Esta análise em particular é de grande relevância pois os autores concluem que o filtro de Liu e West apresenta um comportamento degenerativo mesmo se implementado a partir da distribuição proposta ótima $p(x_t|y_t, x_{t-1}, \theta)$. Desta forma, deveríamos esperar um comportamento irregular da mistura de Gaussianas mesmo na situação mais favorável possível. Contudo, verifica-se que as conclusões traçadas nesta obra são baseadas em erros de implementação conforme demonstrado no Apêndice C. Ao corrigir o código disponibilizado pelos autores, percebemos que a degeneração observada na construção de estimativas pontuais dos parâmetros desaparece completamente. É possível verificar ainda que a consistência da estratégia LW vai além do problema das estimativas pontuais. Ao comparar as abordagens PL e LW, a partir de uma implementação adequada, verificamos que ambas são capazes de gerar boas estimativas para a forma funcional das distribuições *a posteriori* acerca dos parâmetros estáticos. A fim de manter a coerência com o ensaio proposto originalmente por Lopes e Tsay, baseamos nossa análise na seguinte configuração:

- parâmetros para simulação: $\theta = (\alpha, \beta, \tau^2, \sigma^2) = (0, 0; 0, 9; 0, 5; 1, 0)$;

- distribuições *a priori*:

$$\sigma^2 \sim GI(n_0/2; n_0\sigma_0^2/2), \quad \tau^2 \sim GI(v_0/2; v_0\tau_0^2/2) \quad \text{e} \quad \alpha, \beta | \tau^2 \sim N(b_0, \tau^2 B_0),$$

onde

$$n_0 = v_0 = 10, \quad \tau_0^2 = 0, 5, \quad \sigma_0^2 = 1, 0, \quad b_0 = (0, 0; 0, 9)' \quad \text{e} \quad B_0 = \mathbf{I}_2;$$

- distribuições propostas:

$$x_0 \sim N(0, 1) \quad \text{e} \quad x_t | y_t, x_{t-1}, \theta \sim N\left(\frac{y_t \sigma^{-2} + (\alpha + \beta x_{t-1}) \tau^{-2}}{\sigma^{-2} + \tau^{-2}}, \frac{1}{\sigma^{-2} + \tau^{-2}}\right), \quad t = 1, 2, \dots, n.$$

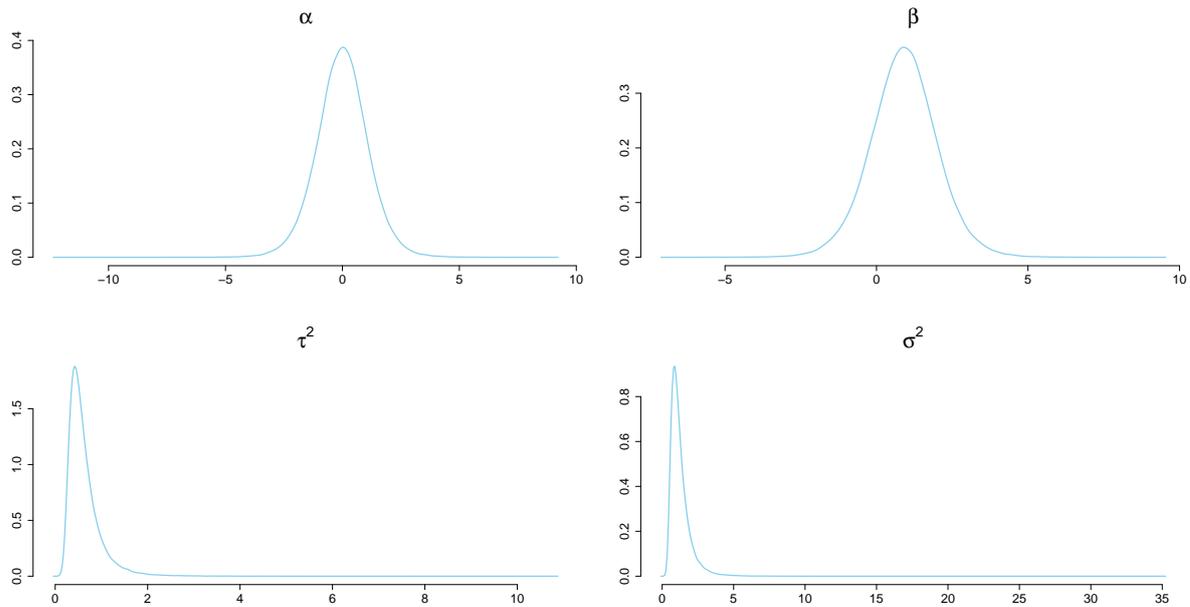


Figura 11 – Distribuições *a priori* associadas aos parâmetros $(\alpha, \beta, \tau^2, \sigma^2)$.

Ao invés de avaliar o desempenho dos métodos a partir de comparações baseadas em uma única execução do filtro PL baseado em 100000 partículas, conforme sugerido originalmente, consideramos a implementação do algoritmo particle Gibbs com 20000 iterações e *burn-in* de 10000. O modelo favorece a implementação deste tipo de algoritmo uma vez que a estrutura linear Gaussiana habilita a construção direta de trajetórias suavizadas via FFBS. Desta forma, esperamos obter uma análise de desempenho mais justa com respeito a ambas as estratégias. Simulamos inicialmente uma série de 500 observações e consideramos 25 execuções das perspectivas LW e PL todas com 5000 partículas. Consideramos também uma execução de cada algoritmo com 100000 pontos a fim de analisar o comportamento assintótico dos dispositivos. Os resultados foram satisfatórios em ambos os casos, reforçando que a degeneração atribuída à metodologia LW deve-se de fato a implementações subótimas. A concordância entre as duas tratativas foi tão significativa que até mesmo as taxas médias de rejuvenescimento artificial e suavização de reamostragem foram as mesmas: 21,4% e 0,2%, respectivamente. Destacamos apenas que os resultados baseados na aproximação por núcleos estimadores apresenta um pouco mais de variabilidade, mas de forma alguma, caracterizam uma solução degenerada. A figura a seguir apresenta as densidades construídas ao longo das 25 execuções de cada método; a semelhança entre as estimativas construídas via filtros de partículas e via MCMC é evidente.

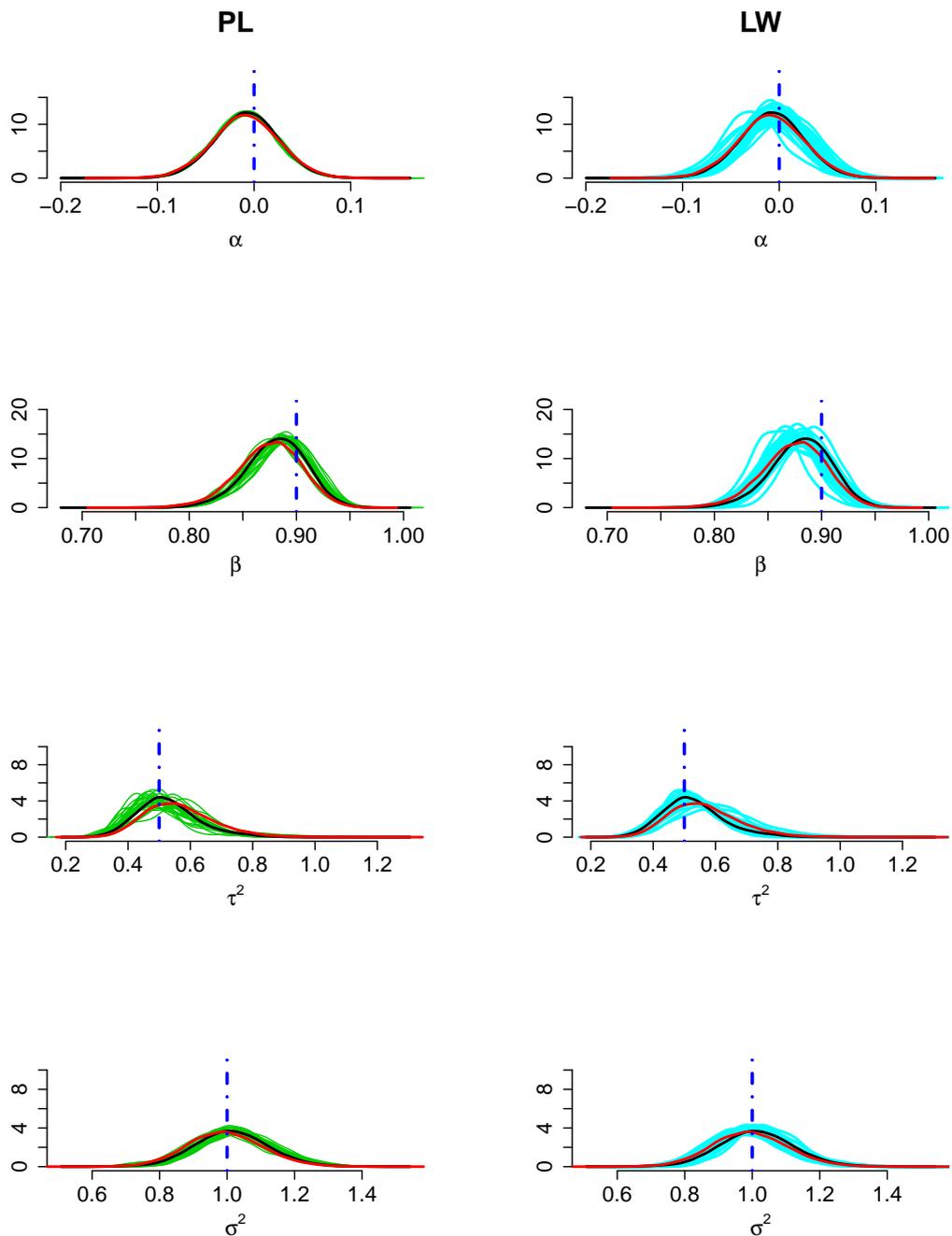


Figura 12 – Densidades estimadas via filtro de partículas – Modelo $AR(1) + \text{ruído}$ com 500 observações. As colunas da esquerda e da direita resumem, na devida ordem, os resultados obtidos de 25 execuções das estratégias PL e LW com 5000 partículas. As curvas em preto indicam a estimativa baseada em 100000 partículas de cada um dos dois métodos. Curvas em vermelho corresponde à estimativa construída via MCMC.

Repetindo o ensaio para um problema baseado em 1000 observações e, considerando novamente implementações baseadas em 5000 partículas, observa-se que o filtro LW otimizado continua estável. Além disso, nota-se que a abordagem PL começa a demonstrar pequenos problemas na estimação do parâmetro τ^2 associado à volatilidade dos estados latentes. Este último resultado é um reflexo da falta de dinâmica natural associada ao vetor de estatísticas

suficientes. Em resumo, temos que os dois métodos são capazes de gerar resultados muito similares, diferentemente da conclusão anterior de Lopes e Tsay. O algoritmo LW apresentou uma variabilidade maior devido à natureza do método, mas ainda assim retornou ótimos resultados.

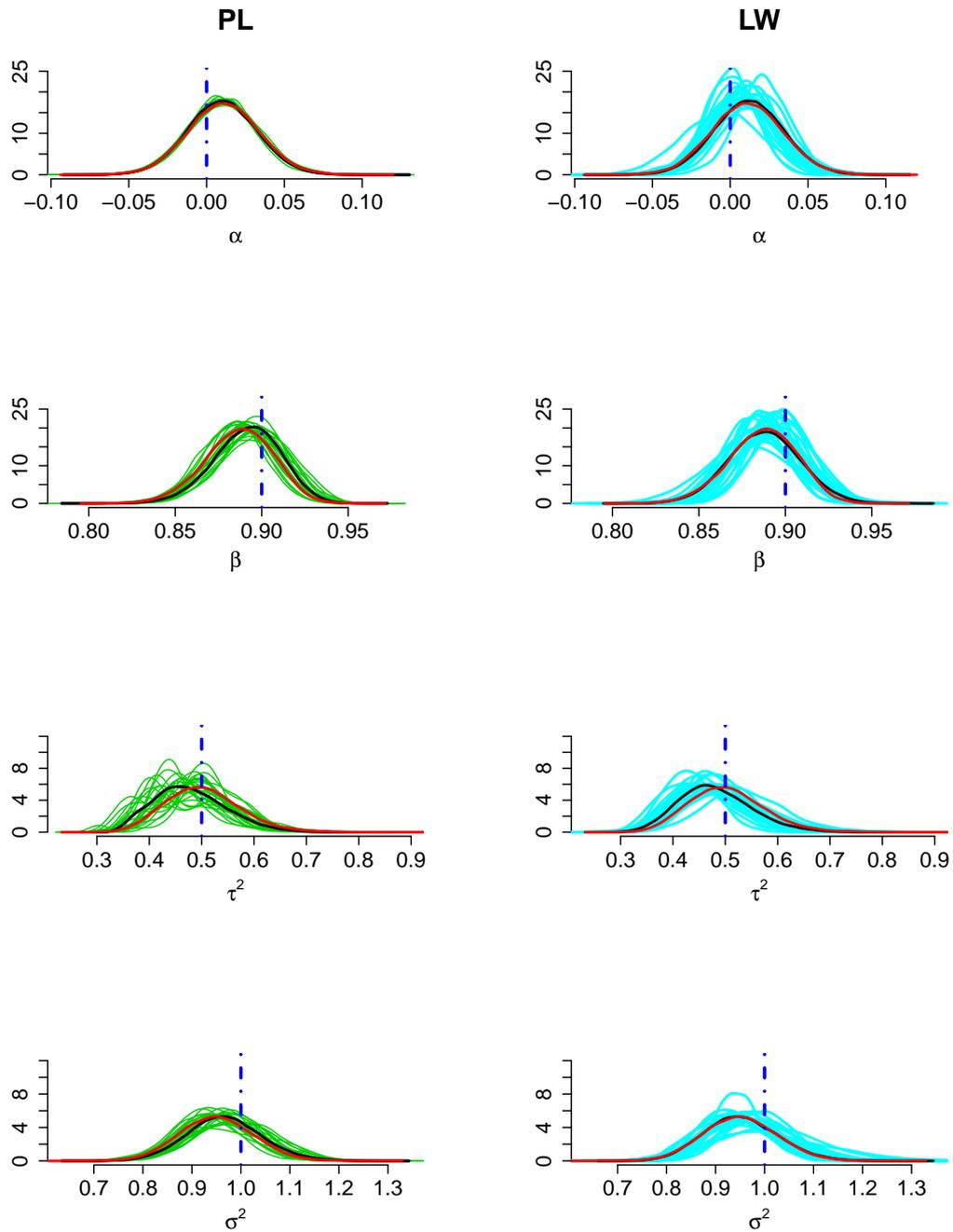


Figura 13 – Densidades estimadas via filtro de partículas – Modelo $AR(1) + ruído$ com 1000 observações. As colunas da esquerda e da direita resumem, na devida ordem, os resultados obtidos de 25 execuções das estratégias PL e LW com 5000 partículas. As curvas em preto indicam a estimativa baseada em 100000 partículas de cada um dos dois métodos. Curvas em vermelho corresponde à estimativa construída via MCMC.

Cenário II: Prado e Lopes (2013)

Analisamos neste segundo cenário alguns dos resultados apresentados por Prado e Lopes. Na obra, os autores apresentam experimentos que evidenciam uma maior instabilidade associada à estratégia LW. Entretanto, acreditamos que os autores fazem uso de um confronto inadequado ao comparar uma estratégia baseada na utilização da distribuição proposta ótima a uma estratégia baseada na proposta Auxiliar convencional. Neste caso, ao revisitar o ensaio proposto por Prado e Lopes, temos uma ótima oportunidade de dissociar as estratégias de filtragem pura das metodologias de rejuvenescimento artificial e assim identificar as verdadeiras fontes de instabilidade no problema da filtragem na presença de parâmetros estáticos desconhecidos.

Seguindo as configurações do primeiro experimento proposto por Prado e Lopes, temos:

- o parâmetro α é considerado conhecido e igual a zero, logo, tal quantidade não é explicitada no problema;
- parâmetros para simulação: $\theta = (\beta, \tau^2, \sigma^2) = (0,95; 0,1; 0,02)$;
- distribuições *a priori*:

$$\beta \sim U(0,1), \quad \tau^2 \sim GI(0,01;0,01) \quad \text{e} \quad \sigma^2 \sim GI(0,01;0,01);$$

- distribuições propostas:

$$x_0 \sim N(0,1) \quad \text{e} \quad x_t | x_{t-1}, y_t, \theta \sim \sum_{i=1}^N \omega_{aux,t}^{(i)} \cdot p\left(x_t | x_{t-1}, \beta^{(i)}, \tau^{2(i)}\right),$$

onde

$$\omega_{aux,t} \propto \frac{1}{\sigma^{(i)}} e^{-\frac{1}{2\sigma^{2(i)}} (y_t - \beta^{(i)} x_{t-1})^2};$$

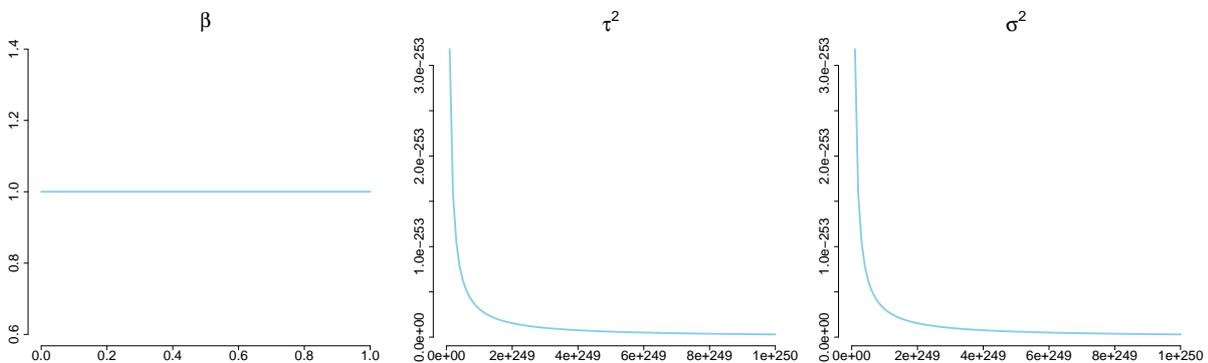


Figura 14 – Distribuições *a priori* associadas aos parâmetros $(\beta, \tau^2, \sigma^2)$.

O experimento baseou-se na geração de uma série de tamanho 300 e os algoritmos de filtragem foram implementados a partir de 2000 partículas apenas. Mesmo considerando algumas das propostas sugeridas neste trabalho, verificamos um comportamento inadequado do filtro LW

otimizado. Abaixo são ilustradas as sequência de tamanhos efetivos amostrais obtidas de 25 execuções deste algoritmo considerando a distribuição proposta cega utilizada nas versões convencionais do filtro Auxiliar. Os resultados coincidiram fielmente com os apresentados pelo autores no ensaio original e de fato indicam que esta implementação específica apresenta sérios problemas de instabilidade.

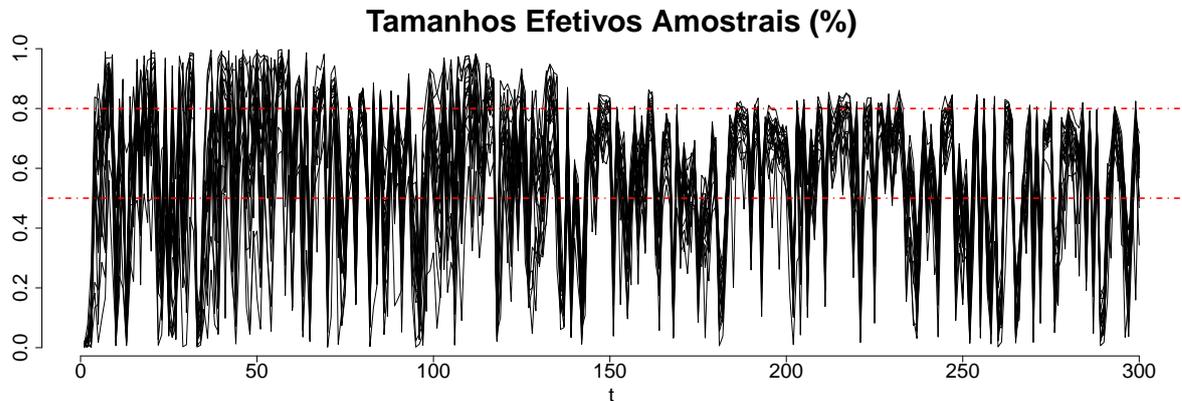


Figura 15 – Tamanhos efetivos amostrais associados a 25 iterações do filtro LW otimizado considerado a perspectiva Auxiliar convencional.

Considerando agora a implementação do filtro LW otimizado baseada na proposta ótima

$$x_t | y_t, x_{t-1}, \theta \sim N \left(\frac{y_t \sigma^{-2} + \beta x_{t-1} \tau^{-2}}{\sigma^{-2} + \tau^{-2}}, \frac{1}{\sigma^{-2} + \tau^{-2}} \right),$$

temos um comportamento completamente distinto. Nas 25 execuções da implementação alternativa, verificamos que os tamanhos efetivos amostrais relativos permaneceram, em média, superiores a 80% ao longo de 85% das iterações conforme ilustrado abaixo. Reforçamos que a

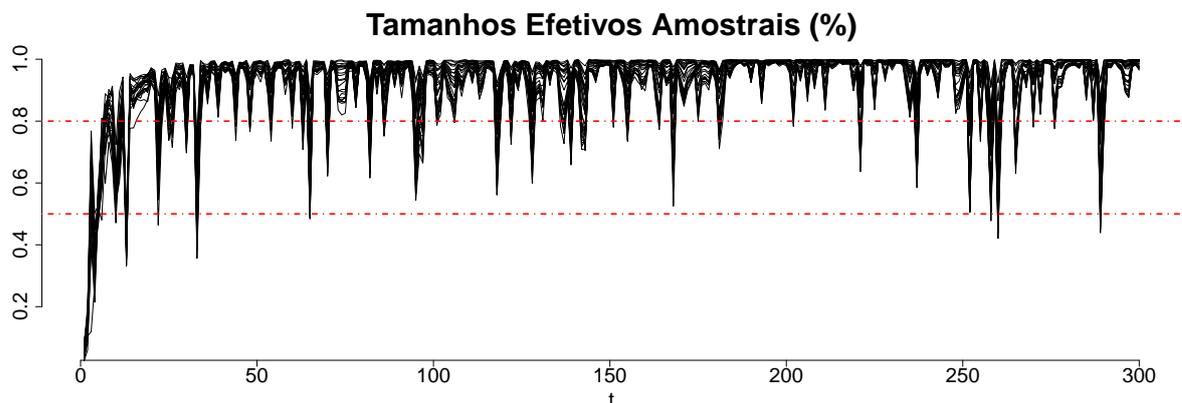


Figura 16 – Tamanhos efetivos amostrais associados a 25 iterações do filtro LW otimizado considerado a perspectiva Auxiliar com distribuição proposta ótima.

única modificação com respeito ao ensaio anterior deu-se na distribuição proposta na passo de amostragem por importância. Desta forma, concluímos que o problema de degeneração apontado por Prado e Lopes não deveria ser associado ao dispositivo de rejuvenescimento baseado nas

mistura de distribuições Gaussianas, mas sim à escolha inadequada para a dinâmica de amostragem da sequência latente. O resumo completo das soluções baseadas na distribuição ótima correspondem à segunda coluna da Figura 10 apresentada no capítulo anterior.

É possível verificar ainda que o método é consistente com respeito à estimação das formas funcionais da distribuições *a posteriori* mesmo considerando esta família vaga de informações iniciais. As estimativas são naturalmente carregadas de variabilidade se considerarmos apenas as 2000 partículas sugeridas no experimento original. Entretanto, a variabilidade é reduzida gradativamente à medida que o enxame de partículas cresce, conforme verificado nos diagramas a seguir. Nota-se que as distribuições estimadas pelo filtro fornecem informações além dos dois primeiros momentos; os resultados indicam algumas densidades mais assimétricas e leptocúrticas na comparação com o modelo Gaussiano.

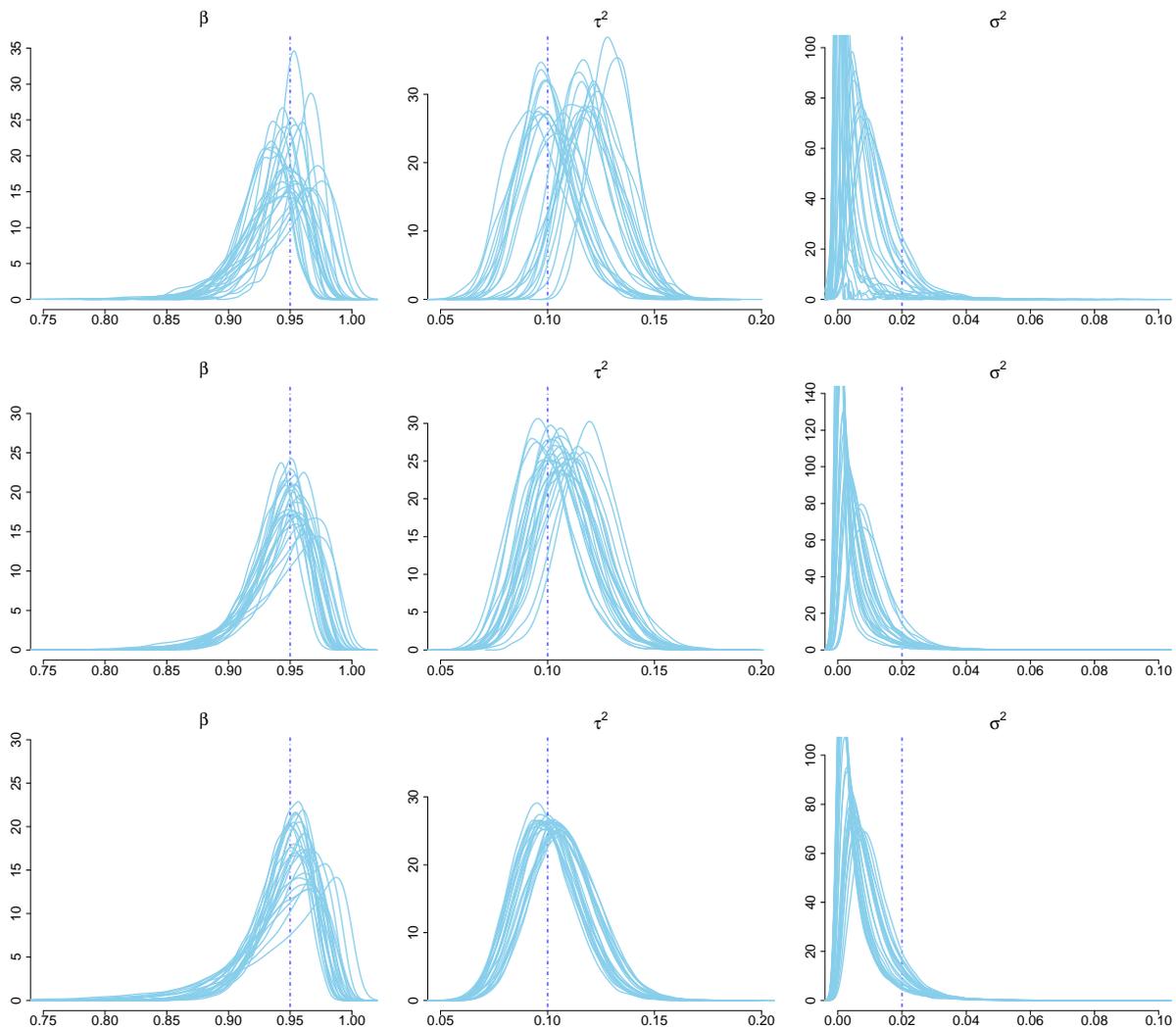


Figura 17 – Densidades estimadas via filtro LW otimizado com número crescente de partículas. Do topo para a base estão representadas aproximações construídas com o auxílio de 2000, 15000 e 25000 partículas. Linhas verticas indicam os valores reais dos parâmetros utilizados para simulação.

5.2 AR(1) + Ruído Não-Gaussiano

A estrutura apresentada na seção anterior pode ser empregada em situações mais amplas a partir de modificações na natureza dos ruídos. Em particular, [Storvik \(2002\)](#) sugere a utilização de um ruído não-Gaussiano nas observações de modo a considerar inovações de cauda pesada:

$$\begin{cases} y_t = x_t + \sigma \varepsilon_t, & \varepsilon_t \sim t_\nu \\ x_t = \alpha + \beta x_{t-1} + \eta_t, & \eta_t \sim N(0, \tau^2) \\ x_0 \sim N(0, 1) \end{cases}, \quad t = 1, 2, 3, \dots, n.$$

A análise deste modelo é bastante relevante pois Storvik o toma como exemplo para apontar o comportamento degenerativo de abordagens mais simples como os filtros de Kitagawa e de Liu e West, justificando assim a utilização de dispositivos de filtragem baseados na propagação de estatísticas suficientes. Nosso objetivo é verificar que a versão otimizada do filtro LW sugerida aqui é capaz de lidar com este modelo de maneira simples e extremamente satisfatória. Em um primeiro momento, consideramos rigorosamente a mesma configuração paramétrica sugerida por Storvik. Em seguida, aumentaremos a complexidade do modelo acrescentando a quantidade ν ao vetor de parâmetros desconhecidos e, por fim, mostraremos que o método pode ser consistente até mesmo com respeito ao problema de seleção de modelos.

Cenário I: [Storvik \(2002\)](#)

Consideramos inicialmente a configuração sugerida por Storvik:

- parâmetros para simulação: $\theta = (\alpha, \beta, \tau^2, \sigma^2, \nu) = (0, 0; 0, 9; 1, 0; 1, 0; 5)$ sendo os parâmetros α e ν considerados conhecidos e, portanto, excluídos do processo inferencial;
- distribuições *a priori*

$$\beta \sim N(0, \sqrt{10}), \quad \tau^2 \sim GI(0, 5; 0, 5) \quad \text{e} \quad \sigma^2 \sim GI(0, 5; 0, 5);$$

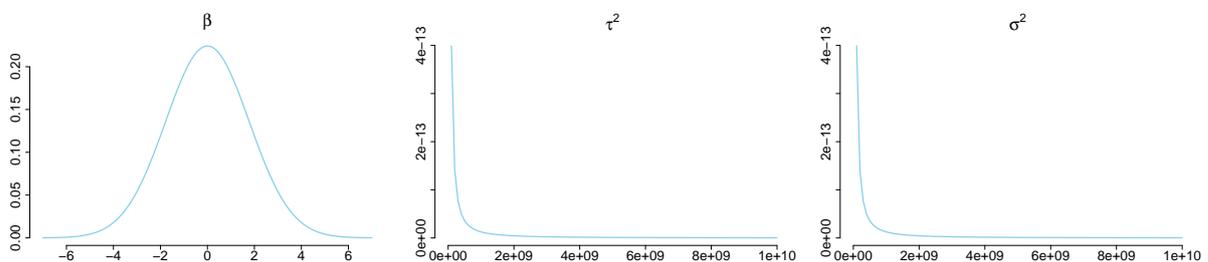


Figura 18 – Distribuições *a priori* associadas aos parâmetros $(\beta, \tau^2, \sigma^2)$.

A distribuição proposta ótima é acessível a este modelo devido à sua estrutura condicionalmente linear Gaussiana. Para facilitar a implementação de tal distribuição, basta introduzirmos uma sequência de variáveis auxiliares

$$\delta_t \sim GI(\nu/2, \nu/2)$$

e reparametrizar o processo de observações da seguinte forma:

$$\hat{y}_t = x_t + \sigma \sqrt{\delta_t} \hat{\varepsilon}_t, \quad \hat{\varepsilon}_t \sim N(0, 1).$$

Neste caso,

$$\hat{y}_t | \delta_t, x_{t-1}, \theta \sim N(\beta x_{t-1}, \tau^2 + \delta_t \sigma^2)$$

e

$$x_t | y_t, x_{t-1}, \theta \sim N\left(\frac{y_t (\delta_t \sigma^2)^{-1} + \beta x_{t-1} \tau^{-2}}{(\delta_t \sigma^2)^{-1} + \tau^{-2}}, \frac{1}{(\delta_t \sigma^2)^{-1} + \tau^{-2}}\right), \quad t = 1, 2, \dots, n.$$

A sequência $\{\delta_t\}$ deve ser incorporada ao modelo como parte da estrutura latente, a qual pode ser atualizada a cada instante de tempo de forma independente, segundo seu modelo de definição, ou através da estrutura condicional:

$$\begin{aligned} p(\delta_t | y_t, x_t, \theta) &\propto p(y_t | \delta_t, x_t, \theta) p(\delta_t) \\ &\propto \delta_t^{-\frac{1}{2}} e^{-\frac{(y_t - x_t)^2}{2\delta_t \sigma^2}} \cdot \delta_t^{-\frac{\nu}{2} - 1} e^{-\frac{1}{2\delta_t}} \\ &= \delta_t^{-\frac{\nu+1}{2} - 1} e^{-\frac{1}{2\delta_t} \left(\frac{(y_t - x_t)^2}{\sigma^2} + \nu \right)} \sim GI\left(\frac{\nu+1}{2}, \frac{1}{2} \left(\nu + \frac{(y_t - x_t)^2}{\sigma^2} \right)\right). \end{aligned}$$

Storvik implementou um dispositivo de filtragem com 2000 partículas, o qual constituiu uma das versões pioneiras dos algoritmos de amostragem sequencial baseados na propagação de estatísticas suficientes. Segundo as conclusões apresentadas pelo autor, as demais estratégias de filtragem não seriam capazes de obter o mesmo êxito e apresentariam um comportamento extramamente degenerativo mesmo considerando distribuições *a priori* mais informativas para as variâncias ($GI(10; 10)$) e um montante superior de partículas ($N = 5000$). Contudo, de maneira

Observações x Estados Latentes

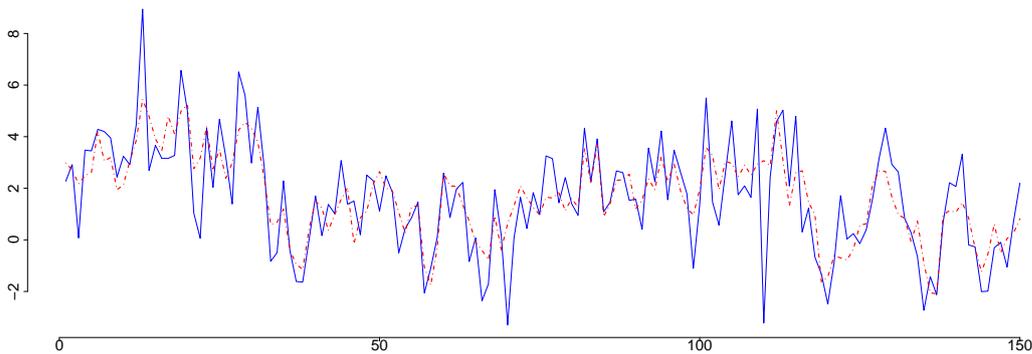


Figura 19 – Exemplo de sequências de observações e estados latentes simulados de acordo com a especificação proposta por Storvik. A curva em azul expressa as observações (perturbadas por um ruído t) enquanto a trajetória em vermelho ilustra a sequência latente (ruídos Gaussianos).

análoga ao estudo de caso apresentado na seção anterior, podemos verificar que as conclusões de Storvik também são baseadas em implementações subótimas. Considerando rigorosamente as especificações utilizadas por Storvik no desenvolvimento de seu algoritmo de filtragem, foi

possível obter resultados muito satisfatórios a partir do filtro LW otimizado. As figuras a seguir ilustram o bom desempenho do algoritmo considerando um experimento similar ao proposto por Storvik. O método não apenas rastreou bem os parâmetros estáticos desconhecidos como também estimou de forma estável a sequência de estados latentes. Diferentemente do ensaio analisado na seção anterior, sequer conseguimos reproduzir os efeitos degenerativos extremos apontados pelo autor original do estudo, mesmo considerando a tratativa Auxiliar convencional.

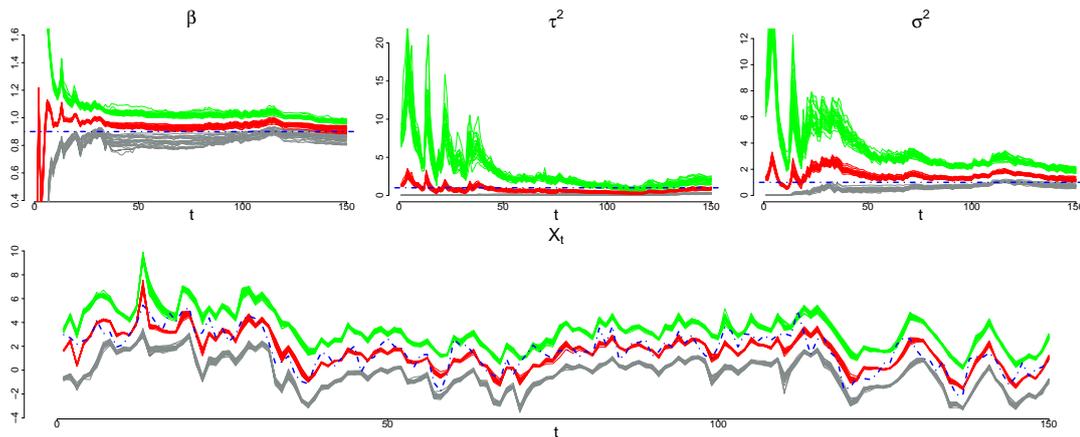


Figura 20 – Resumo de 25 iterações do método LW otimizado para série de tamanho 150: médias *a posteriori* e bandas de 95% de credibilidade. Implementações baseadas em enxames de 2000 partículas conforme proposto originalmente por Storvik. Trajetórias em azul representam quantidades reais adotadas na simulação da observações.

Em seu estudo, Storvik indica a construção de resultados inadequados a partir da estratégia LW mesmo considerando implementações baseadas em 5000 partículas. Desta forma, visando contradizer tal conclusão e evidenciar o potencial das aproximações baseadas em misturas Gaussianas, consideramos a implementação de um cenário similar utilizando 4000 partículas para o processamento de uma série de tamanho 500 (ao menos 3 vezes mais longa que a série do experimento original).

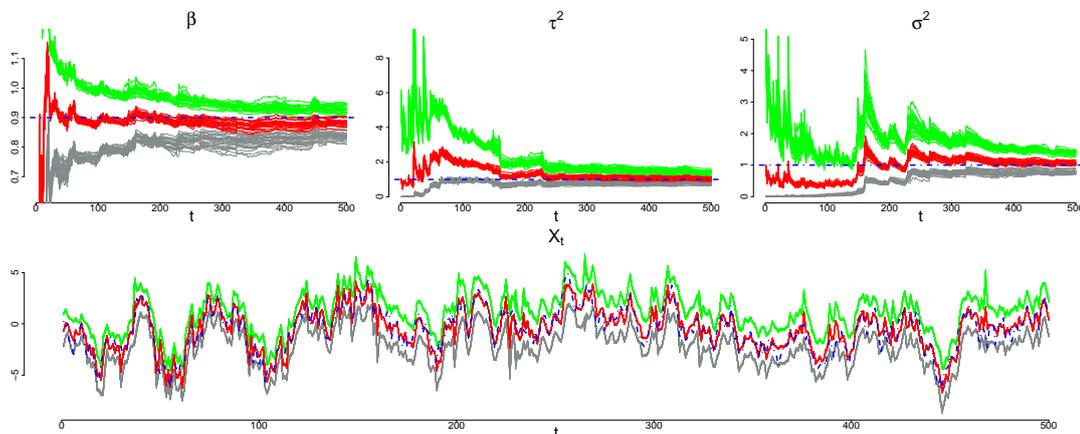


Figura 21 – Resumo de 25 iterações do método LW otimizado para série de tamanho 1000: médias *a posteriori* e bandas de 95% de credibilidade. Implementações baseadas em enxames de 4000 partículas. Trajetórias em azul representam quantidades reais adotadas na simulação da observações.

O método LW otimizado ainda foi capaz de gerar resultados satisfatórios em configurações mais adversas. Avaliamos o desempenho do método na filtragem de uma série de tamanho 1000 cujo processo latente tem coeficiente autoregressivo $\beta = 0,99$. Buscamos neste caso criar

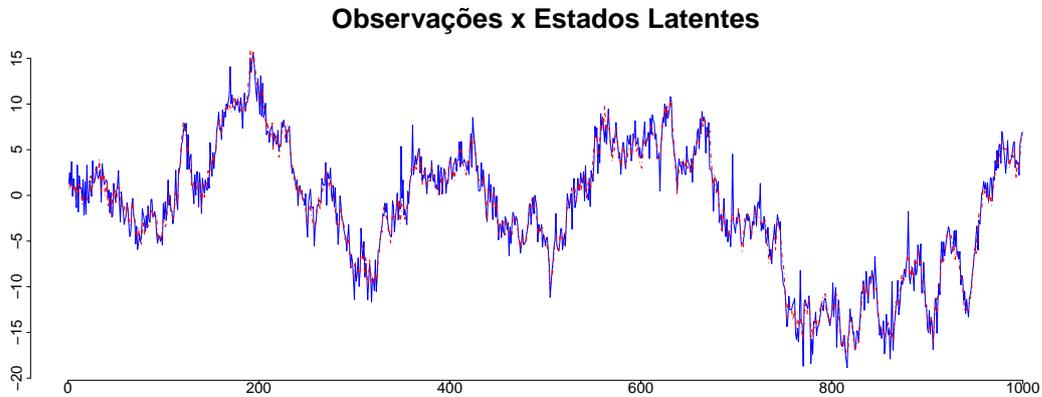


Figura 22 – Sequências de observações e estados latentes, tamanho 1000, simuladas a partir dos parâmetros $(\beta, \tau^2, \sigma^2, \nu) = (0,99; 1,00; 1,00; 5)$. A curva em azul expressa as observações (perturbadas por um ruído t) enquanto a trajetória em vermelho ilustra a sequência latente (ruídos Gaussianos).

um cenário mais propenso à degeneração: séries mais longas e próximas da não estacionariedade. O número de partículas foi mantido em 4000 e o desempenho do algoritmo foi novamente satisfatório. Concluimos que a estratégia LW foi capaz de lidar com o problema dos parâmetros estáticos desconhecidos de forma satisfatória. Esperamos assim reforçar a afirmação de que a simples aproximação baseada na mistura de distribuições Gaussianas é tão consistente quanto as estratégias baseadas na propagação de estatísticas suficientes.

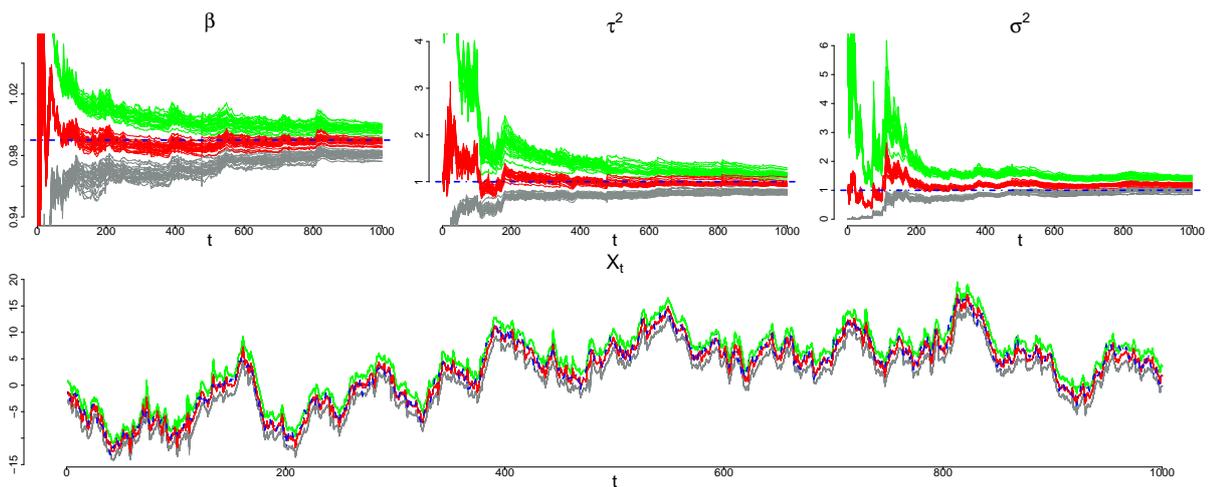


Figura 23 – Resumo de 25 iterações do método LW otimizado para série de tamanho 500: médias *a posteriori* e bandas de 95% de credibilidade. Implementações baseadas em enxames de 4000 partículas. Trajetórias em azul representam quantidades reais adotadas na simulação da observações.

Cenário II: v desconhecido

Neste segundo ensaio, incorporamos o parâmetro v , correspondente aos graus de liberdade das inovações t , ao vetor de quantidades desconhecidas. Por simplicidade, adotamos uma distribuição uniforme no intervalo $(2,30)$. Além disso, elevamos o número de partículas em 50% em cada um dos três ensaios discutidos anteriormente devido à variabilidade adicional observada após a inserção de mais uma quantidade desconhecida. O algoritmo não foi capaz de identificar com precisão o parâmetro v na série de tamanho 500, entretanto, as demais estimativas sofreram pouca ou nenhuma alteração. É possível que a combinação de uma distribuição *a priori* não informativa com uma sequência pequena de observações tenha favorecido esta situação. Se estivessemos em um regime de degeneração deveríamos observar um comportamento inadequado com respeito às demais quantidades, o que não foi o caso.

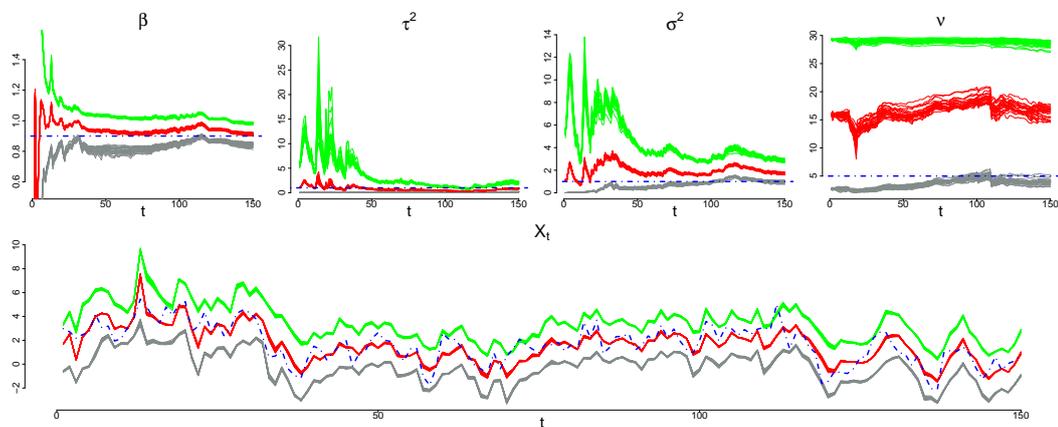


Figura 24 – Filtragem da série de 150 observações com parâmetro v desconhecido.

Estimativas mais razoáveis foram obtidas para as séries mais longas, no entanto, ainda verifica-se uma certa imprecisão sobretudo nas caudas da distribuição associada ao parâmetro v . As estimativas associadas às demais quantidades permanecerem praticamente inalteradas, reforçando a hipótese de que o filtro não tenha passado por um processo de degeneração.

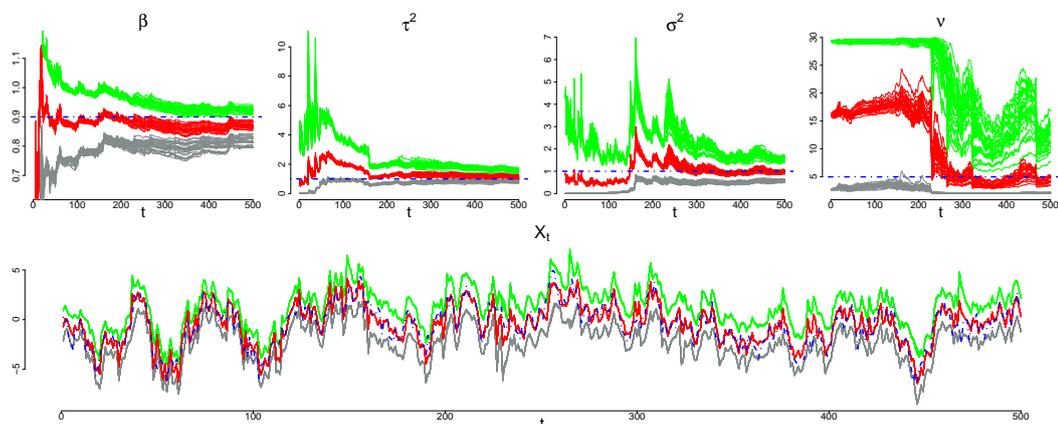


Figura 25 – Filtragem da série de 500 observações com parâmetro v desconhecido.

Resultados mais precisos certamente poderiam ser obtidos a partir da adoção de distribuições *a priori* distintas. Lopes e Polson (2016), por exemplo, sugerem o emprego da distribuição de Jeffreys e demonstram experimentalmente melhoras significativas com respeito às estimativas de v . Os autores desenvolvem ainda um esquema de filtragem baseado na propagação de estatísticas suficientes para o tratamento de problemas similares a este. Não consideramos outras

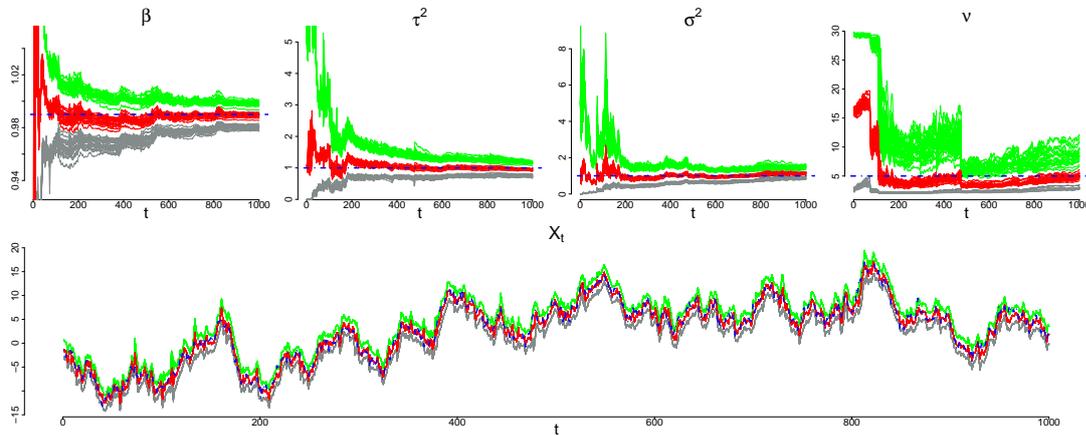


Figura 26 – Filtragem da série de 1000 observações com parâmetro v desconhecido.

possibilidades de implementação pois nosso objetivo era basicamente avaliar o comportamento do filtro LW otimizado em cenários possivelmente favoráveis à sua degeneração. Ao constatar um desempenho estável mesmo em situações baseadas em informações iniciais vagas, podemos esperar um comportamento bastante satisfatório na associação com distribuições *a priori* menos problemáticas.

Cenário III: seleção de modelos

O último cenário de testes visa ilustrar a consistência do método no quesito seleção de modelos. Para tal avaliamos a capacidade do filtro LW otimizado em discriminar o modelo baseado em inovações t de um modelo autoregressivo com ruído convencional, considerando os três experimentos discutidos anteriormente com v desconhecido. Uma maior dificuldade de

Experimento	150 obs.	500 obs.	1000 obs.
Fator de Bayes	1,7976	27976,7100	1011,9250

Tabela 2 – Médias dos fatores de Bayes (inovações $t \times$ inovações Gaussianas) obtidas das 25 execuções dos experimentos abordados no Cenário II.

identificação foi verificada apenas no experimento envolvendo 150 observações. Já havíamos obtido neste caso uma estimativa pontual em torno de 18, para os graus de liberdade v , além de uma distribuição *a posteriori* próxima da informação uniforme utilizada *a priori*. É razoável imaginar uma maior dificuldade de diferenciação dos modelos nestas circunstâncias. Os demais experimentos identificaram muito bem o modelo original, mantendo a coerência com as estimativas já observadas no cenário anterior.

5.3 AR(p) + Ruído

Prado e Lopes (2013) consideraram uma extensão do modelo AR(1)+“ruído” capaz de acomodar processos autoregressivos mais complexos:

$$\begin{cases} y_t = x_t + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma^2) \\ x_t = \sum_{i=1}^p \phi_i x_{t-i} + \eta_t, & \eta_t \sim N(0, \tau^2), \quad t = 1, 2, 3, \dots, n. \\ x_0 \sim N(0, 1) \end{cases}$$

Os autores ressaltam que esta abordagem equivale à construção de um modelo ARMA(p,p), conforme argumentado anteriormente por Granger e Morris (1976), o que reforça a relevância do exemplar. A dinâmica do processo pode ser reescrita convenientemente de modo a acomodar a estrutura de um modelo Markoviano oculto:

$$\begin{cases} y_t = \mathbf{F}' \mathbf{z}_t + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma^2) \\ \mathbf{z}_t = \mathbf{G}(\phi) \mathbf{z}_{t-1} + \eta_t, & \eta_t \sim N(\mathbf{0}, \mathbf{W}) \end{cases},$$

onde $\mathbf{F}' = (1, 0, \dots, 0) \in \mathbb{R}^p$, $\mathbf{W} = \tau^2 \mathbf{F} \mathbf{F}'$, $\mathbf{z}_t = (x_t, x_{t-1}, \dots, x_{t-p+1})' \in \mathbb{R}^p$ é o vetor latente e

$$\mathbf{G}(\phi) = \begin{pmatrix} \phi_{1:(p-1)} & \phi_p \\ \mathbf{I}_{p-1} & \mathbf{0}_{p-1} \end{pmatrix}$$

a matriz de evolução dos estados, sendo $\mathbf{I}_{p-1} \in \mathbb{R}^{(p-1) \times (p-1)}$ a matriz identidade e $\mathbf{0}_{p-1}$ um vetor de zeros. Desta forma, temos a possibilidade de trabalhar com modelos ARMA(p,p) a partir de uma alternativa eficiente baseada em métodos sequenciais de Monte Carlo.

A fim de determinar a estacionariedade da sequência latente podemos associar a estrutura autoregressiva acima a um polinômio característico da forma

$$\begin{aligned} \Phi(u) &= 1 - \phi_1 u - \phi_2 u^2 - \dots - \phi_p u^p \\ &= \prod_{j=1}^C (1 - r_j e^{-2\pi i / \lambda_j} u) (1 - r_j e^{2\pi i / \lambda_j} u) \times \prod_{j=C+1}^{C+R} (1 - r_j u), \end{aligned}$$

onde $\{r_j e^{\pm 2\pi i / \lambda_j}\}_{j=1}^C$ e $\{r_j\}_{j=C+1}^{C+R}$ são os recíprocos das raízes do polinômio. O processo será estacionário se o módulo dos recíprocos destas raízes for inferior a 1. Buscando garantir a estacionariedade do processo latente, Lopes e Prado sugeriram a adoção de distribuições *a priori* com respeito aos $2R + C$ parâmetros

- r_j , $j = 1, 2, \dots, R$;
- (r_j, λ_j) , $j = R + 1, R + 2, \dots, R + C$.

Uma vez obtidos os parâmetros descritos acima, podemos recuperar o vetor $\phi_{1:p}$ associado aos parâmetros iniciais do modelo.

Os experimentos originais focam bastante na degeneração da estratégia LW, sugerindo que esta não constitui uma abordagem válida para o tratamento do problema. Neste caso, a melhor alternativa seria a utilização de variações de algoritmos apoiados pela propagação de estatísticas suficientes. Nosso objetivo é apontar a verdadeira fonte de degeneração dos filtros e mostrar porque as sugestões propostas aqui garantem o bom funcionamento da perspectiva LW. Este ganho é muito relevante pois, apesar de apresentar resultados favoráveis, a metodologia sugerida por Prado e Lopes necessita de uma calibragem imprecisa dos parâmetros do algoritmo. Em particular, a proposta de distribuições *a priori* adotada pelos autores não garante a construção sequencial de distribuições *a posteriori* passíveis de atualização via um esquema de propagação de estatísticas suficientes. Neste caso, é preciso obter uma distribuição Gaussiana que aproxima bem tais distribuições e permita a construção de tal esquema de conjugação; a calibragem dos parâmetros desta aproximação não é especificada pelos autores. Além disso, mesmo com a construção aproximada do esquema de estatísticas suficientes, não é possível amostrar diretamente da distribuição *a posteriori*

$$r_1, \dots, r_R, r_{R+1}, \lambda_{R+1}, \dots, r_{R+C}, \lambda_{R+C}, \sigma^2, \tau^2 | y_{1:t}, x_{1:t}.$$

Neste caso, a atualização é feita separadamente por blocos de maneira análoga à execução de um passo Gibbs:

$$\begin{aligned} & r_1^* | r_2, r_3, \dots, r_R, r_{R+1}, \lambda_{R+1}, \dots, r_{R+C}, \lambda_{R+C}, \sigma^2, \tau^2 | y_{1:t}, x_{1:t} \\ & r_2^* | r_1^*, r_3^*, \dots, r_R, r_{R+1}, \lambda_{R+1}, \dots, r_{R+C}, \lambda_{R+C}, \sigma^2, \tau^2 | y_{1:t}, x_{1:t} \\ & \quad \vdots \\ & r_{R+1}^*, \lambda_{R+1}^* | r_1^*, r_2^*, \dots, r_R^*, r_{R+2}, \lambda_{R+2}, \dots, r_{R+C}, \lambda_{R+C}, \sigma^2, \tau^2 | y_{1:t}, x_{1:t} \\ & r_{R+2}^*, \lambda_{R+2}^* | r_1^*, r_2^*, \dots, r_R^*, r_{R+1}^*, \lambda_{R+1}^*, \dots, r_{R+C}, \lambda_{R+C}, \sigma^2, \tau^2 | y_{1:t}, x_{1:t} \\ & \quad \vdots \\ & r_{R+C}^*, \lambda_{R+C}^* | r_1^*, r_2^*, \dots, r_R^*, r_{R+1}^*, \lambda_{R+1}^*, \dots, r_{R+C-1}^*, \lambda_{R+C-1}^*, \sigma^2, \tau^2 | y_{1:t}, x_{1:t} \\ & \tau^{2*} | r_1^*, r_2^*, \dots, r_R^*, r_{R+1}^*, \lambda_{R+1}^*, \dots, r_{R+C}^*, \lambda_{R+C}^*, \sigma^2 | y_{1:t}, x_{1:t} \\ & \sigma^{2*} | r_1^*, r_2^*, \dots, r_R^*, r_{R+1}^*, \lambda_{R+1}^*, \dots, r_{R+C}^*, \lambda_{R+C}^*, \tau^{2*} | y_{1:t}, x_{1:t} \end{aligned}$$

Esta segunda aproximação constitui mais uma fonte de ruído na estimação da distribuição *a posteriori*. Por fim, os autores indicam também que em uma aplicação geral envolvendo p coeficientes autorregressivos, seria necessária a construção de um vetor de estatísticas suficientes de dimensão $5C + 2R + 2$, onde C e R são os números de raízes complexas e reais, respectivamente. Este montante de informação gera um custo extra tanto em termos de tempo quanto em termos de memória, caracterizando um ponto desfavorável à utilização da estratégia. Todos estes pormenores são completamente erradicados pelo filtro LW otimizado, o que garante ao método muito mais simplicidade e eficiência.

Da discussão traçada nas seções anteriores a respeito do modelo $AR(1) + \text{ruído}$, temos que a escolha da distribuição proposta para a propagação dos estados constitui um passo crucial no desenvolvimento de algoritmos mais estáveis. Dada a estrutura do problema, é possível acessar a proposta ótima independentemente da ordem p :

$$x_t | y_t, x_{t-1}, \theta \sim N \left(\frac{y_t \sigma^{-2} + \tau^{-2} \sum_{i=1}^p \phi_i x_{t-i}}{\sigma^{-2} + \tau^{-2}}, \frac{1}{\sigma^{-2} + \tau^{-2}} \right). \quad (5.1)$$

A perspectiva adotada, *Bootstrap* ou Auxiliar, é pouco relevante em situações como esta onde a distribuição ideal é conhecida e analiticamente tratável. Nestes casos, seria interessante considerar o paradigma *Bootstrap* já que este faz uso de menos passos de reamostragem conferindo assim, no mínimo, mais eficiência aos dispositivos de filtragem. Em se tratando deste estudo especificamente, os experimentos foram baseadas na perspectiva Auxiliar de modo a manter a coerência com os estudos desenhados por Prado e Lopes. Além disso, ao adotar a abordagem Auxiliar, podemos evidenciar que é possível manter o fenômeno de degeneração sob controle monitorando os fatores de fertilidade mesmo reamostrando a cada instante de iteração.

AR(2) + Ruído

Considerando inicialmente o modelo autoregressivo de ordem 2, observado com ruído, temos a seguinte representação:

$$\begin{cases} y_t &= x_t + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma^2) \\ x_t &= \phi_1 x_{t-1} + \phi_2 x_{t-2} + \eta_t, & \eta_t \sim N(0, \tau^2), \quad t = 1, 2, 3, \dots, n. \\ x_{-1} &= 0, x_0 \sim N(0, 1) \end{cases}$$

Prado e Lopes propuseram um estudo deste problema a partir da seguinte configuração, a qual conta apenas com coeficientes reais:

- parâmetros para simulação:

$$(\phi_1, \phi_2, \tau^2, \sigma^2) = (0, 1000; 0, 8075; 0, 10000; 0, 0200);$$

- os recíprocos r_1 e r_2 podem ser obtidos da configuração:

$$(1 - r_1 u)(1 - r_2 u) = 1 - \phi_1 u - \phi_2 u^2 \implies \begin{cases} r_1 + r_2 = \phi_1 \\ -r_1 r_2 = \phi_2 \end{cases} \implies r_1 = 0,95 \text{ e } r_2 = -0,85;$$

- distribuições *a priori*:

$$r_1 \sim U(0, 1), \quad r_2 \sim U(-1, 0), \quad \tau^2 \sim GI(0, 01; 0, 01) \quad \text{e} \quad \sigma^2 \sim GI(0, 01; 0, 01);$$

- tamanho do enxame de partículas: $N = 6000$;

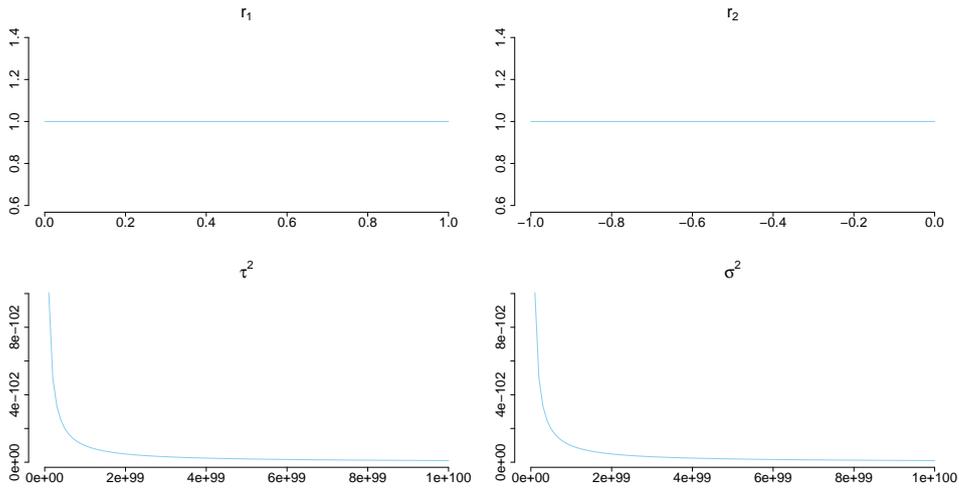


Figura 27 – Distribuições *a priori* associadas aos parâmetros $(r_1, r_2, \tau^2, \sigma^2)$.

O ensaio foi baseado em uma série de tamanho 400, cuja evolução é representada abaixo. Este cenário caracteriza de fato uma situação desfavorável à estratégia LW a família de distri-

Observações x Estados Latentes

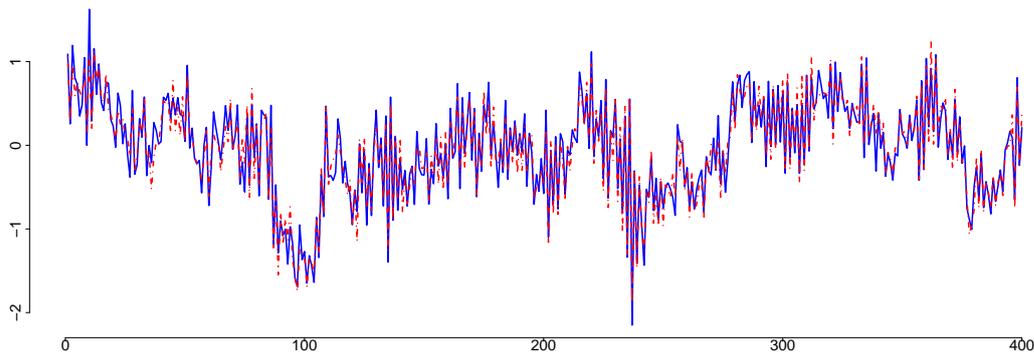


Figura 28 – Exemplo de sequências de observações e estados latentes simulados de acordo com a especificação proposta por Prado e Lopes. A curva em azul expressa as observações enquanto a trajetória em vermelho ilustra o processo autoregressivo latente.

buições *a priori* é muito vaga, o que certamente pode dificultar a busca do algoritmo por foco. A fim de tentar reproduzir os resultados apresentados por Prado e Lopes, implementamos primeiramente a estratégia LW com distribuição proposta ideal porém sem considerar as estratégias de otimização sugeridas em nosso trabalho. Os resultados apontaram estimativas carregadas de variabilidade porém não necessariamente degeneradas.

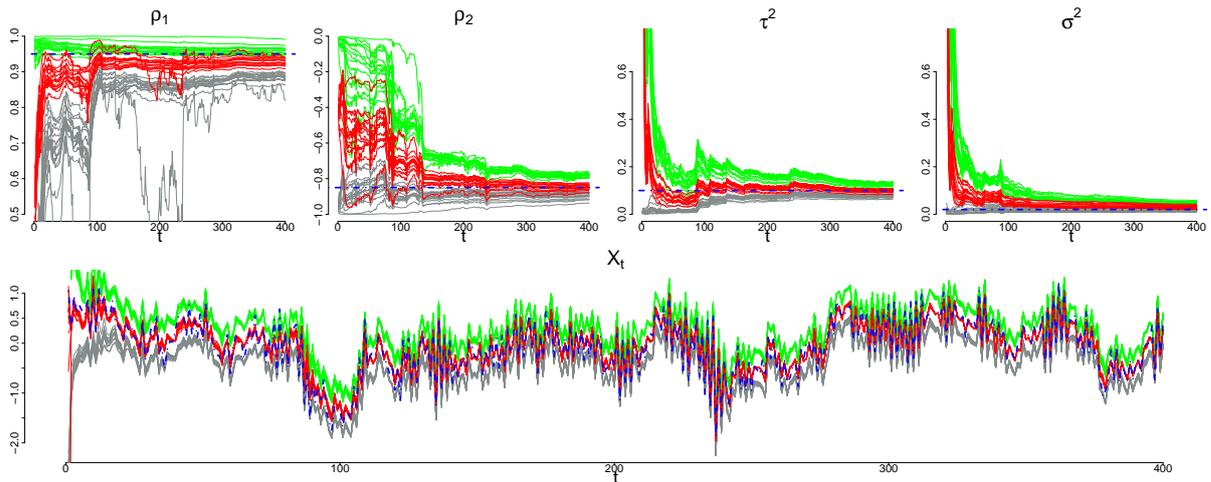


Figura 29 – Resumo de 15 iterações do método LW convencional com distribuição proposta ideal: médias *a posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades reais adotadas na simulação da observações.

A variabilidade observada no diagrama acima pode ser explicada pelo conjunto vago de informações tomado *a priori*. Monitorando a diversidade da população a cada instante, observamos um fator de fertilidade médio de 2,5% após o primeiro passo de reamostragem. O grande enxugamento deve-se de fato ao espalhamento desordenado da população inicial, fruto da falta de precisão dos modelos probabilísticos considerados no ponto de partida. Neste caso, a estratégia LW já parte com um montante de informação extremamente reduzido e, conseqüentemente, demonstrará uma maior instabilidade nos instantes futuros. O conceito de reamostragem suavizada surge como uma ferramenta de auxílio exatamente em cenários como este. A fim de evitar a degeneração imediata, podemos reamostrar da população a partir de um conjunto de pesos não tão heterogêneo. Ao final do passo de seleção teremos então indivíduos marcados por um maior balanceamento entre diversidade e foco. O próximo diagrama ilustra o efeito da reamostragem suavizada neste problema. Ao longo das 15 novas iterações, verificamos

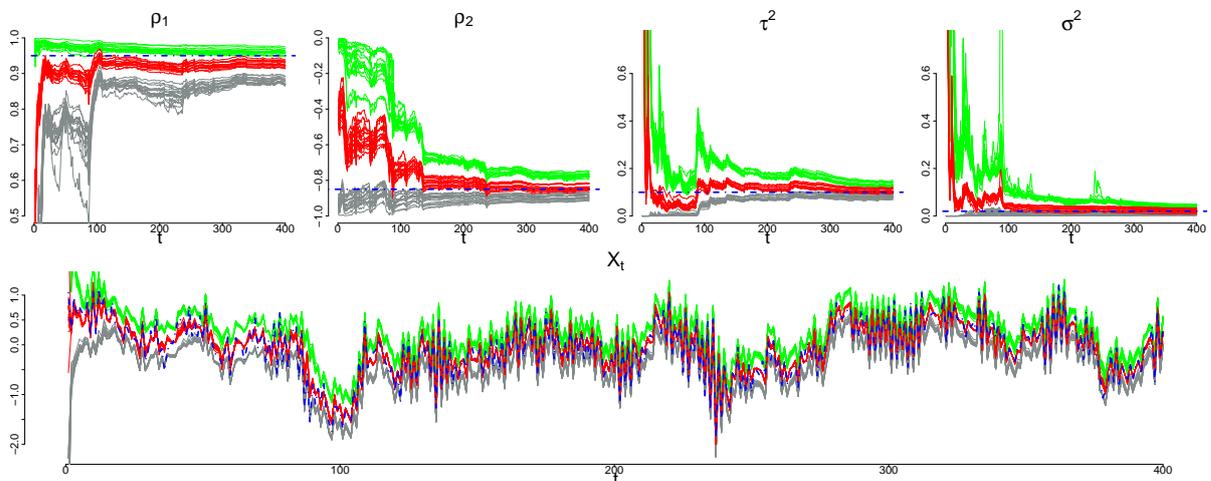


Figura 30 – Resumo de 15 iterações do método LW otimizado com distribuição proposta ideal: médias *a posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades reais adotadas na simulação da observações.

que o passo de reamostragem suavizada foi realizado em média ao longo de 0,5% das iterações. Não coincidentemente, tais iterações coincidem exatamente com os dois instantes iniciais de filtragem, justificando que o maior problema era de fato o espalhamento desmoderado imposto pela distribuição inicial dos parâmetros.

Um detalhe curioso apontado por Prado e Lopes é que os resultados do algoritmo de Liu e West não seriam sensíveis à distribuição *a priori* adotada. Os autores comentam que mesmo considerando um par de distribuições iniciais mais informativa para os parâmetros de volatilidade

$$\tau^2 \sim GI(3; 2 \times 0, 1) \quad \text{e} \quad \sigma^2 \sim GI(3; 2 \times 0, 02),$$

ainda verificaríamos um comportamento inadequado da mistura de Gaussianas. Este resultado de fato caracterizaria um argumento forte contra a utilização da estratégia LW já que a mesma apresentaria estimativas inadequadas mesmo em uma versão mais adestrada do modelo. Todavia, adotando tais distribuições na implementação do filtro LW otimizado, verificamos uma melhora substancial nas estimativas. A figura a seguir resume a inferência resultante da utilização do novo par de distribuições para os parâmetros τ^2 e σ^2 ; todos os demais parâmetros da implementação foram mantidos inalterados.

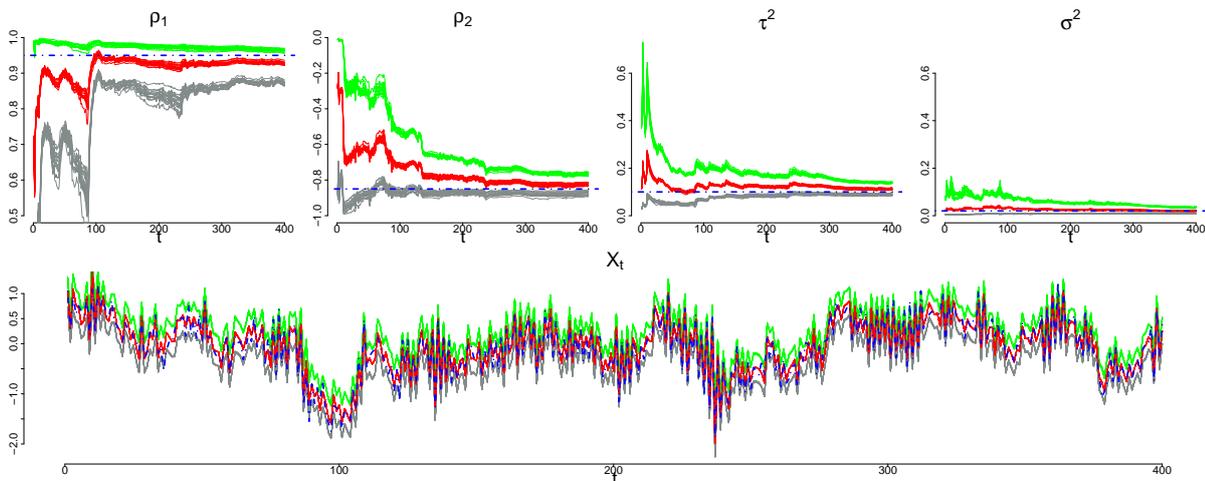


Figura 31 – Resumo de 15 iterações do método LW otimizado com distribuição proposta ideal e distribuições *a priori* mais informativas: médias *a posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades reais adotadas na simulação da observações.

Este último resultado mostra definitivamente que a principal causa do fenômeno de degeneração apontado por Prado e Lopes era de fato a distribuição proposta do passo de amostragem por importância e não a aproximação utilizada no rejuvenescimento artificial. O conceito de reamostragem suavizada entrou neste contexto como um passo de refinamento responsável por dar mais equilíbrio à população em situações muito adversas. No conjunto da obra temos um resultado muito relevante já que, neste contexto, o filtro LW otimizado é incomparavelmente mais simples que a abordagem PL proposta por Prado e Lopes além de mais eficiente tanto em termos de custo de tempo e de memória.

AR(3) + Ruído

O próximo ensaio trata de processos autoregressivos latentes de ordem 3, os quais também foram abordados na obra original de Prado e Lopes:

$$\begin{cases} y_t &= x_t + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma^2) \\ x_t &= \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + \eta_t, & \eta_t \sim N(0, \tau^2), \quad t = 1, 2, 3, \dots, n. \\ x_{-2} &= x_{-1} = 0, & x_0 \sim N(0, 1) \end{cases}$$

Os autores escolheram adequadamente os parâmetros de modo a induzir o surgimento de raízes complexas com respeito ao polinômio característico Φ . Mais precisamente, a configuração adotada foi:

- parâmetros para simulação:

$$(r_1, r_2, \lambda_2, \tau^2, \sigma^2) = (-0,950; 0,950; 16,00; 1,000, 0,250);$$

- os coeficientes autorregressivos podem ser recuperados da relação:

$$(1 - r_1 u)(1 - r_2 e^{-2\pi i/\lambda_2} u)(1 - r_2 e^{2\pi i/\lambda_2} u) = 1 - \phi_1 u - \phi_2 u^2 - \phi_3 u^3$$

$$\implies \begin{cases} \phi_1 = r_1 + 2r_2 \cos(2\pi/\lambda_2) \\ \phi_2 = -2r_1 r_2 \cos(2\pi/\lambda_2) - r_2^2; \\ \phi_3 = r_1 r_2^2 \end{cases}$$

- distribuições *a priori*:

$$r_1 \sim NT_{(-1,0;0,0)}(-0,5; 1,0), \quad r_2 \sim NT_{(0,5;1,0)}(0,8; 1,0), \quad \lambda_2 \sim NT_{(12,20)}(16,0; 2,0),$$

$$\tau^2 \sim GI(2,00; 0,25) \quad \text{e} \quad \sigma^2 \sim GI(2,0; 1,0),;$$

onde $NT_{(a,b)}(\mu, \sigma^2)$ corresponde à distribuição Normal com parâmetros μ e σ^2 truncada no intervalo (a, b) ;

- tamanho do enxame de partículas: $N = 2000$;

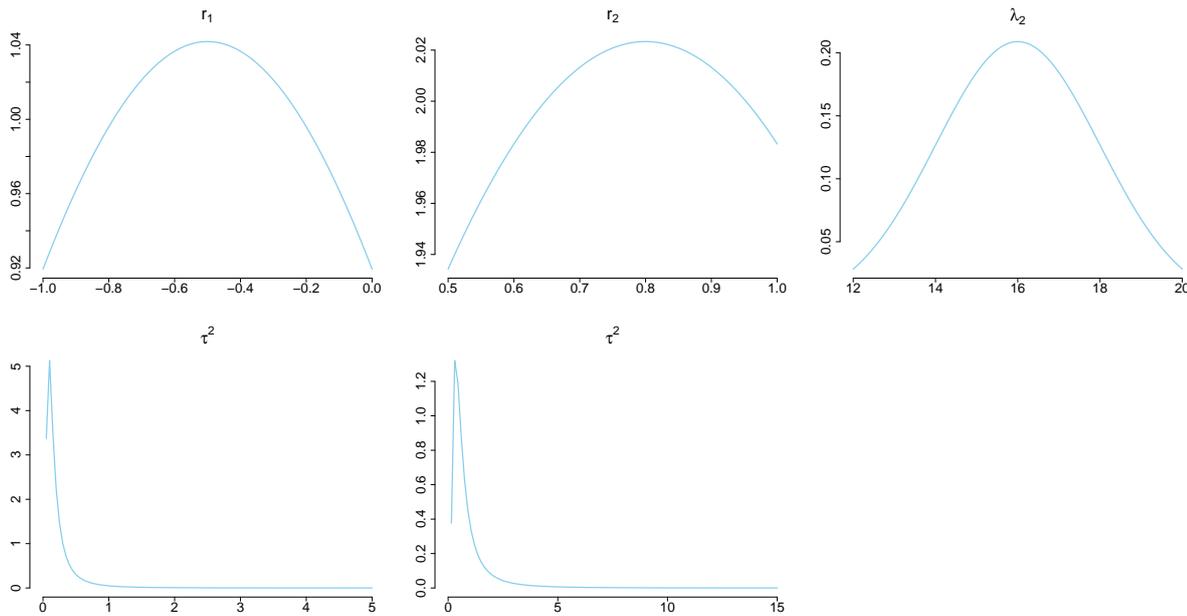


Figura 32 – Distribuições *a priori* associadas aos parâmetros $(r_1, r_2, \lambda_2, \tau^2, \sigma^2)$.

O experimento foi baseado na simulação de uma série de tamanho 250 e os autores consideraram apenas a implementação da estratégia PL. A utilização do filtro de Liu e West foi descartada no experimento original devido ao comportamento inadequado observado nos cenários mais simples. Nosso objetivo é verificar que a versão otimizada deste filtro também é capaz de gerar bons resultados mesmo em se tratando de uma aplicação, a princípio, mais complicada.

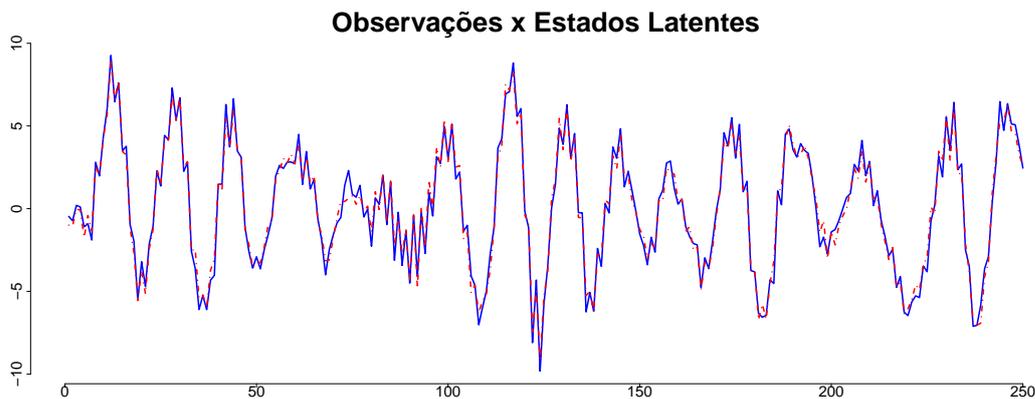


Figura 33 – Exemplo de sequências de observações e estados latentes simulados de acordo com a especificação proposta por Prado e Lopes. A curva em azul expressa as observações enquanto a trajetória em vermelho ilustra o processo autoregressivo latente.

O filtro LW otimizado foi implementado seguindo rigorosamente as especificações propostas por Prado e Lopes para o bom funcionamento de sua estratégia baseada na propagação de estatísticas suficientes. Novamente, o filtro LW demonstrou-se muito estável e capaz de boas estimativas para os 5 parâmetros estáticos assim como para a trajetória latente. Como as distribuições *a priori* não foram tão difusas como as propostas no cenário anterior, o dispositivo

de reamostragem suavizada praticamente não fez diferença ao problema. Já a agenda dinâmica

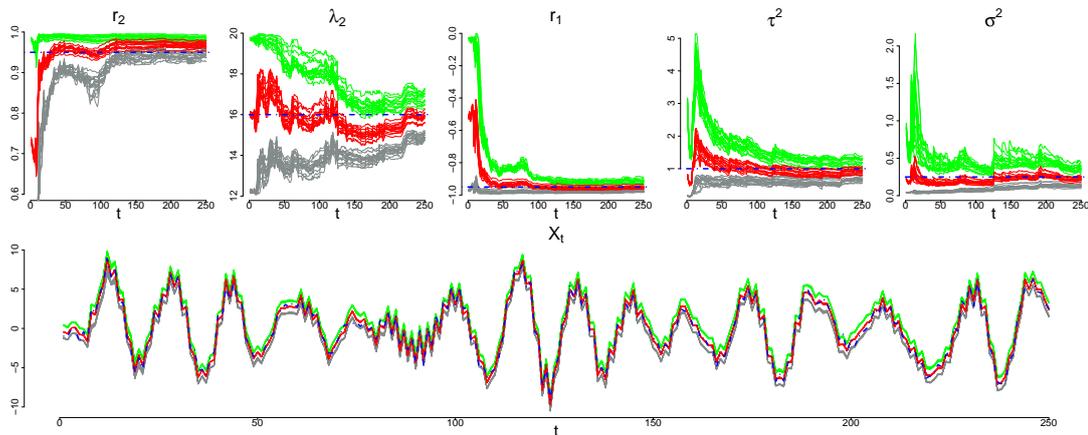


Figura 34 – Resumo de 15 iterações do método LW otimizado: médias *a posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades reais adotadas na simulação da observações.

de rejuvenescimento demonstrou grande utilidade no quesito eficiência computacional. Ao longo das 15 estimativas, o processo de rejuvenescimento foi realizado em média em 22% das iterações. Como o espaço paramétrico tem dimensão 5, a amostragem de parâmetros toma boa parte dos recursos computacionais utilizados no processo de filtragem, neste caso, verificamos uma economia de aproximadamente 40% de tempo execução na comparação com o dispositivo baseado no rejuvenescimento perpétuo.

A fim de avaliar o desempenho do método com um pouco mais de precisão, consideramos 150 execuções do mesmo mantendo fixas as especificações adotadas anteriormente. Os resultados mostraram-se mais uma vez favoráveis ao método, conforme podemos observar nos *boxplots* abaixo. O parâmetro r_2 tendeu a superestimar levemente o parâmetro real, entretanto este

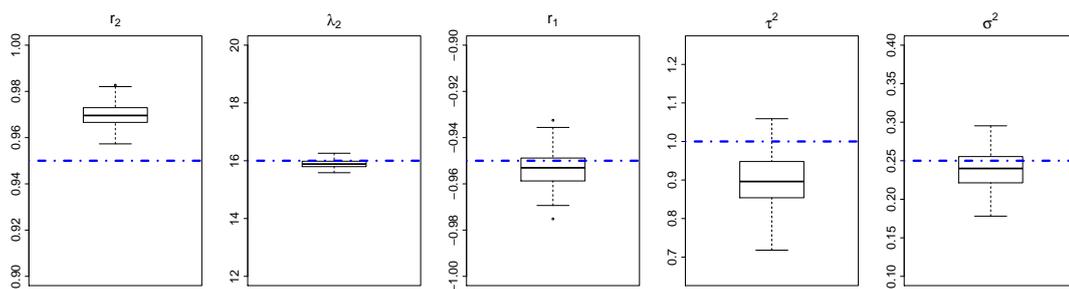


Figura 35 – Distribuição das médias *a posteriori* construídas a partir de 150 execuções do filtro LW otimizado. Curvas em azul referem-se aos parâmetros reais tomados no processo de simulação.

resultado é coerente com a escolha da distribuição *a priori*. A distribuição de partida do parâmetro (Figura 32) dá um pouco mais de prioridade a valores próximos a 1 e, provavelmente, o pequeno tamanho da série não foi suficiente para compensar tal ponderação inicial. De uma forma geral, o método demonstrou ser uma ferramenta bastante útil para o tratamento deste problema.

Por fim, realizamos um último experimento considerando uma série de tamanho 1500 gerada a partir do vetor de parâmetros trabalhado nos experimentos anteriores. Este estudo é

particularmente interessante uma vez que Prado e Lopes sugerem que dispositivos sequenciais, mesmo considerando a abordagem baseada na propagação de estatísticas suficientes, não devem ser utilizadas em aplicações envolvendo séries longas. Todavia, os resultados obtidos neste último ensaio indicam que o filtro LW otimizado é capaz de gerar bons resultados mesmo em se tratando de longas séries de observações. O próximo diagrama ilustra o comportamento do

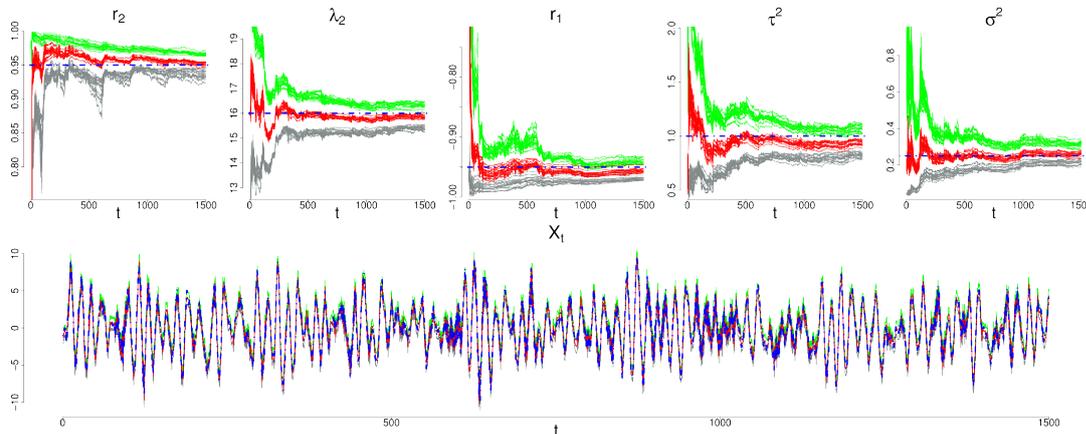


Figura 36 – Resumo de 15 iterações do método LW otimizado considerando uma série de tamanho 1500: médias *a posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades reais adotadas na simulação da observações.

método ao longo de 15 execuções para a filtragem de uma série de tamanho 1500. Como estamos lidando com uma série de tamanho 6 vezes superior ao adotado anteriormente, implementamos o filtro de partículas a partir de um enxame de 8000 pontos. Não só os parâmetros foram bem identificados como o processo latente também foi filtrado de maneira bastante eficaz em todas as execuções.

Assim como no caso anterior, repetimos o experimento 150 vezes de modo a avaliar as distribuições das estimativas finais (médias *a posteriori*) acerca dos parâmetros estáticos. Os resumos a seguir ilustram que as estimativas foram muito satisfatórias com respeito a todos os parâmetros. Além de consistentes, destacamos que as execuções foram muito eficientes. Cada

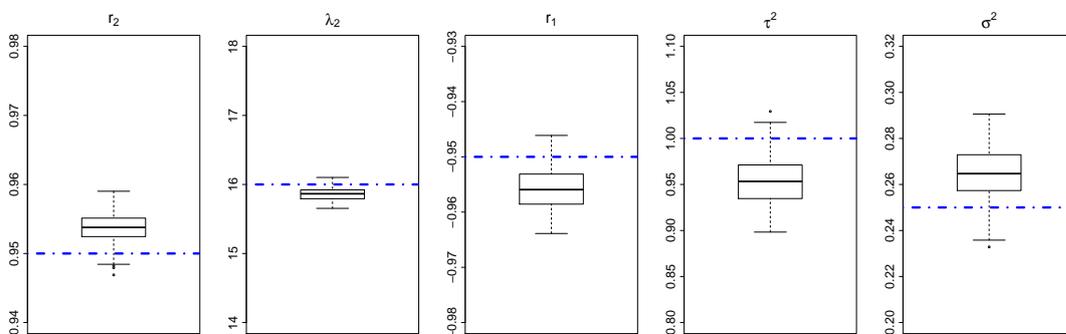


Figura 37 – Distribuição das médias *a posteriori* construídas a partir de 150 execuções do filtro LW otimizado considerando uma série de 1500 observações e enxames de 8000 partículas. Curvas em azul referem-se aos parâmetros reais tomados no processo de simulação.

rodada não levou mais que 15 segundos de execução sequencial em linguagem R considerando

um processador Intel Core i7-2670QM CPU 2.20GHz \times 8. Além disso, destacamos que a complexidade de memória do algoritmo é bastante satisfatória. A sugestão inicial de Lopes e Prado, por exemplo, necessitaria de 11 vetores para o armazenamento das estatísticas suficientes além de outros 6 vetores para o armazenamento das quantidades filtradas, totalizando 17 vetores de dimensão N (tamanho do enxame de partículas). Já o filtro LW otimizado faz uso de apenas 6 vetores, configurando assim uma redução de quase $2/3$ no consumo total de memória. Devemos destacar também a simplicidade na implementação da estratégia LW. A construção do método é dada de forma direta, diferentemente da proposta de Prado e Lopes que conta com uma especificação complexa e pouco clara. Desta forma, acreditamos que este estudo possa, além de esclarecer algumas de degeneração, estimular ainda mais a utilização de dispositivos sequencias no tratamento deste elegante modelo tratado originalmente por Prado e Lopes.

APLICAÇÕES A DADOS REAIS

6.1 Volatilidade Estocástica

O modelo de volatilidade estocástica abordado aqui é relativamente simples e já foi exaustivamente estudado na literatura. Em particular, algoritmos MCMC bastante eficientes para o tratamento deste e de outros modelos mais complexos de volatilidade podem ser encontrados na literatura (KIM; SHEPHARD; CHIB, 1998; CHIB; NARDARI; SHEPHARD, 2002; OMORI *et al.*, 2007; NAKAJIMA; OMORI, 2009; KASTNER; Frühwirth-Schnatter, 2014). Nosso objetivo aqui é apenas ilustrar a importância da análise qualitativa do problema para a implementação do algoritmo de Liu e West conforme discutido na Seção 4.6. Para tal, selecionamos um conjunto de séries cujos resultados inferenciais obtidos a partir das implementações convencionais do método são indiscutivelmente insatisfatórios. Entretanto, ao explorar algumas das sugestões propostas neste trabalho podemos chegar a soluções eficientes e muito próximas das oferecidas por algoritmos MCMC.

6.1.1 Descrição do Modelo

Este estudo é baseado na parametrização mais clássica do modelo de volatilidade estocástica, o qual é dado em função da média estacionária do processo latente (JACQUIER; POLSON; ROSSI, 1994; KIM; SHEPHARD; CHIB, 1998):

$$\begin{cases} y_t = e^{h_t/2} \varepsilon_t, & \varepsilon_t \sim N(0, 1) \\ h_t = \mu + \beta(h_{t-1} - \mu) + \eta_t, & \eta_t \sim N(0, \tau^2) \\ h_0 | \mu, \beta, \tau \sim N(\mu; \tau^2 / (1 - \beta^2)) \end{cases}, \quad t = 1, 2, \dots, n. \quad (6.1)$$

onde ε_t e η_s são ruídos independentes para quaisquer $t, s \in \{1, 2, \dots, n\}$.

Esta parametrização pode, em alguns casos, ser bastante problemática pois o método de Liu e West apresenta uma grande dificuldade com respeito à estimação da média estacionária μ . Tal parâmetro é muito sensível à variação na dinâmica dos dados, o que induz uma maior instabilidade na evolução de sua incerteza. Além disso, verifica-se que o parâmetro μ apresenta uma correlação muito acentuada com respeito à persistência β . Por conseguinte, a utilização da matriz de covariâncias convencional no método de Liu e West pode fazer com que a grande instabilidade de μ seja refletida diretamente neste segundo parâmetro, amplificando muito a variabilidade das estimativas. Buscando evidenciar este fenômeno, fizemos um estudo comparativo envolvendo a utilização da estrutura de covariâncias convencional assim como uma segunda matriz de dependências que trata o espaço paramétrico de forma independente:

$$\Sigma_1 \begin{pmatrix} \mu \\ \beta \\ \tau^2 \end{pmatrix} = \begin{pmatrix} \hat{\sigma}_\mu^2 & \hat{\sigma}_{\mu\beta} & \hat{\sigma}_{\mu\tau^2} \\ \hat{\sigma}_{\beta\mu} & \hat{\sigma}_\beta^2 & \hat{\sigma}_{\beta\tau} \\ \hat{\sigma}_{\tau^2\mu} & \hat{\sigma}_{\tau\beta} & \hat{\sigma}_\tau^2 \end{pmatrix} \quad \text{e} \quad \Sigma_2 \begin{pmatrix} \mu \\ \beta \\ \tau^2 \end{pmatrix} = \begin{pmatrix} \hat{\sigma}_\mu^2 & 0 & 0 \\ 0 & \hat{\sigma}_\beta^2 & 0 \\ 0 & 0 & \hat{\sigma}_\tau^2 \end{pmatrix} \quad (6.2)$$

A substituição da estrutura convencional Σ_1 pela proposta alternativa Σ_2 não implica na eliminação da dependência entre os parâmetros; a troca significa apenas que tal dependência é desconsiderada ao longo do processo de rejuvenescimento dos parâmetros.

Os experimentos revelam que o algoritmo LW otimizado pode degenerar completamente se ignorarmos o efeito induzido pela estrutura de dependências, mesmo considerando as demais propostas de otimização sugeridas nesta obra. Até onde sabemos, nenhum outro trabalho chamou a atenção para a análise deste ponto durante a implementação do filtro de Liu e West. Desta forma, esperamos contribuir com uma análise mais completa indicando todo o potencial do algoritmo assim como verdadeiras causas associadas aos cenários de degeneração.

6.1.2 Análise Experimental: Taxas de Câmbio

Exploramos neste experimento o banco de dados considerado por [Kastner e Frühwirth-Schnatter \(2014\)](#), o qual consiste de séries diárias referentes a taxas de câmbio de 23 moedas com respeito ao Euro no período de 3 de janeiro de 2000 a 4 de abril de 2012. A título de curiosidade, acrescentamos aos dados a série correspondente à taxa de câmbio do Real com respeito ao Euro no mesmo período. O desempenho do método de Liu e West otimizado foi comparado ao do algoritmo MCMC apresentado por Kastner e Frühwirth-Schnatter e disponível na biblioteca *stochvol* para a linguagem de programação R.

As realizações de todos os 24 ensaios foram baseadas em um mesmo conjunto de distribuições *a priori*:

$$\mu \sim N(-10; 100), \quad \frac{\beta + 1}{2} \sim \text{Beta}(20; 1, 5),$$

$$\tau^2 \sim \chi^2(1) \quad \text{e} \quad h_0 \sim N(\mu; \tau^2/(1 - \beta^2)).$$

A incerteza inicial a respeito do parâmetro de persistência segue a sugestão usual (KIM; SHEPHARD; CHIB, 1998) adotada na literatura. Tal escolha atribui uma massa probabilística não desprezível para valores no intervalo (0,7;1,0) de modo a evitar a especificação de um modelo muito restritivo. Os demais parâmetros são associados a distribuições não informativas de modo a investigar a capacidade inferencial dos métodos em diferentes cenários desconsiderando qualquer conhecimento prévio a respeito do comportamento dos dados. A adoção desta estrutura de distribuições *a priori* é bastante relevante para análise proposta neste estudo uma vez que vários autores sugerem que o método de Liu e West não deve ser implementado a partir de distribuições muito difusas. Com respeito à dinâmica do processo latente, consideramos

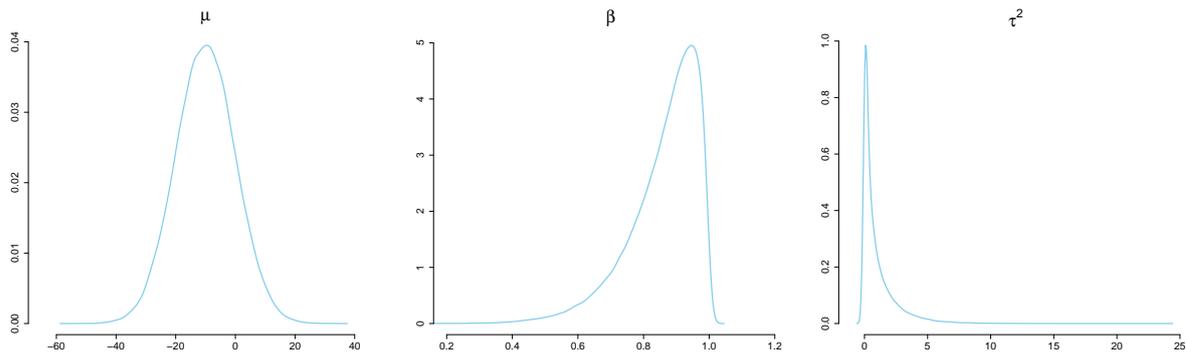


Figura 38 – Distribuições *a priori* correspondentes aos parâmetros μ , β e τ^2 .

distribuições propostas cegas análogas às utilizadas no estudo de modelos de volatilidade com dados simulados apresentado no capítulo anterior:

$$q(h_t | h_{t-1}, y_t, \theta) = \frac{1}{\sqrt{2\pi\tau^2}} \exp \left\{ -\frac{1}{2\tau^2} \left(h_t - \mu - \beta(h_{t-1} - \mu) \right)^2 \right\}.$$

Destacamos ainda que serão apresentados em mais detalhes apenas alguns dos experimentos, em particular aqueles que apresentaram oportunidades ricas para o desenvolvimento de uma análise mais completa dos métodos. O detalhamento completo das 24 séries traria apenas um montante extra de repetição ao texto. Uma tabela resumo contendo um resultado comparativo amplo de ambas as metodologias é apresentada ao final da seção.

Dados Artificiais

Apresentamos a princípio um breve experimento envolvendo uma série simulada diretamente da parametrização (6.1) na tentativa de obter bases de comparação para as possíveis implementações das metodologias tratadas aqui. Neste primeiro estudo avaliaremos o comportamento do algoritmo considerando as estruturas de dependência Σ_1 e Σ_2 em uma situação ótima onde a série de observações segue rigidamente o modelo proposto.

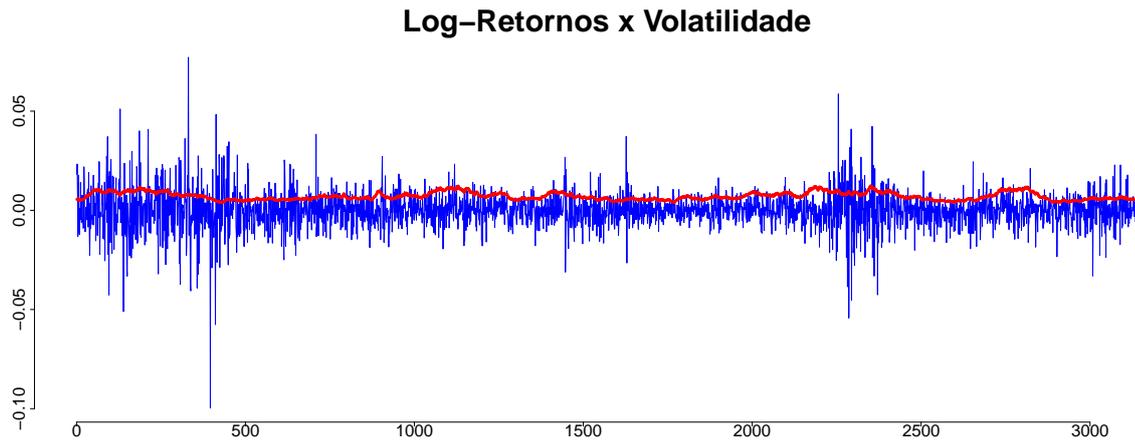


Figura 39 – Séries de observações (azul) e volatilidades (vermelho) simuladas a partir do modelo (6.1).

Os dados foram simulados a partir dos parâmetros estimados pelo método MCMC para a série de taxas referentes à comparação Dólar Americano×Euro:

$$(\mu, \beta, \tau^2) = (-10, 100; 0, 993; 0, 070^2).$$

Temos então um problema relativamente complicado já que o processo latente está próximo da não-estacionariedade. Consideramos inicialmente a implementação do Filtro Auxiliar com abordagem LW e matriz de covariâncias completas e, neste caso, o método demonstrou-se estável e capaz de gerar resultados satisfatórios.

A Figura 42 mostra que as trajetórias filtradas não sofrem alterações bruscas devido a oscilações provenientes das observações, um resultado bastante natural sabendo que a série foi construída diretamente do modelo. Destacamos que a discrepância observada entre a trajetória $x_{1:t}$ estimada pelo filtro de partículas e a obtida pelo método MCMC apresentam diferenças consideráveis, sobretudo nos primeiros instantes de tempo, pois uma trata do processo de filtragem simples enquanto a outra trata do caso suavizado.

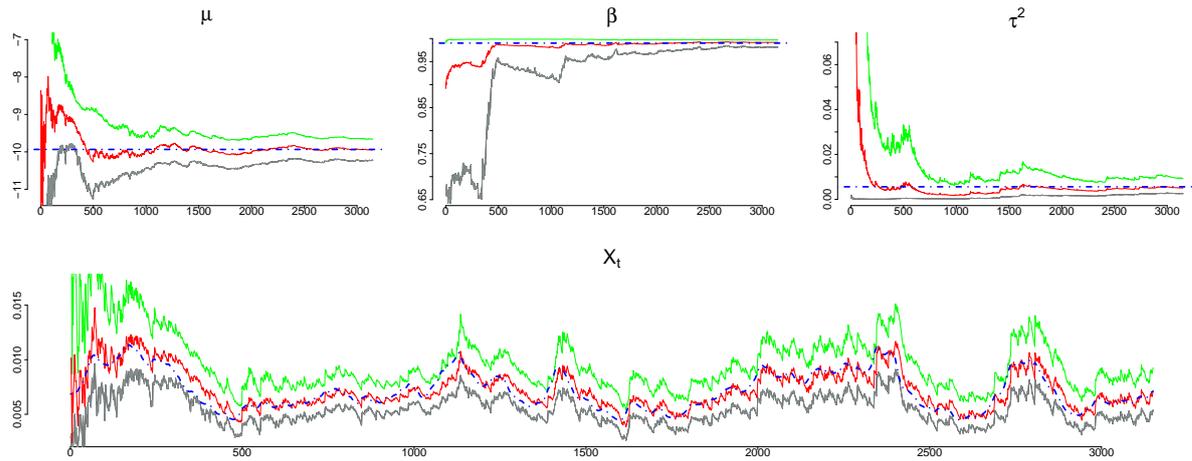


Figura 40 – Resumo de desempenho do método de Liu e West otimizado (Σ_1) com 20000 partículas: médias *a posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam os parâmetros reais tomados como referência para a geração das observações.

A próxima ilustração trata das distribuições *a posteriori* obtidas após a filtragem da última observação. Esta comparação é importante por evidenciar a capacidade inferencial do método de Liu e West otimizado mesmo em situações caracterizadas por séries longas de observações. Além de não observarmos degenerações, temos que a semelhança entre as inferências obtidas pelos métodos LW otimizado e MCMC é muito evidente.

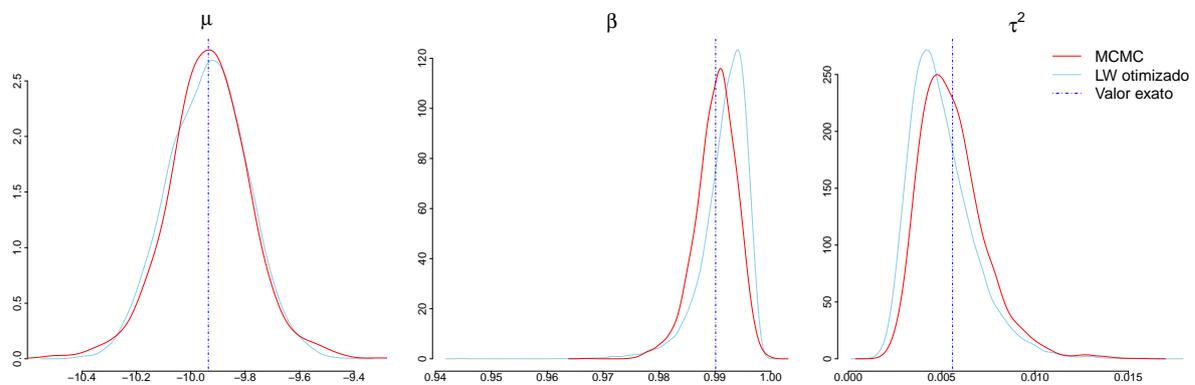


Figura 41 – Densidades *a posteriori* construídas pelo filtro de Liu e West otimizado com matriz de dependências Σ_1 (azul) e pelo método MCMC (vermelho).

Implementamos em seguida o mesmo filtro de partículas entrando considerando a matriz Σ_2 a qual atualiza o vetor de parâmetros estáticos de forma independente. O resultado foi mais uma vez satisfatório e não demonstrou nenhuma variação significativa com relação à conclusão obtida no primeiro ensaio. As distribuições *a posteriori* oferecidas por esta segunda implementação também foram comparadas às obtidas por meio do método MCMC e novamente o resultado obtido pelas aproximações por importância foi excelente se comparado às distribuições consideradas ótimas. Nota-se que nas duas situações o algoritmo foi capaz de identificar até mesmo a leve assimetria associada aos parâmetros β e σ^2 , demonstrando que a metodologia é capaz de capturar informações além dos dois primeiros momentos desde que implementada

adequadamente. Estes aspectos seriam facilmente perdidos caso um limite não fosse imposto ao fator de desconto δ na perspectiva original.

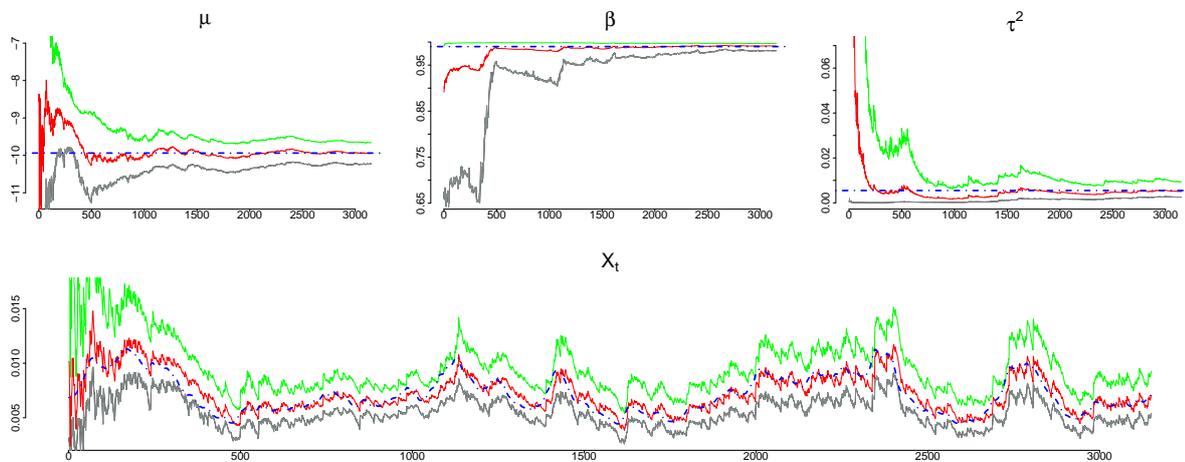


Figura 42 – Resumo de desempenho do método de Liu e West otimizado (Σ_2) com 20000 partículas: médias *a posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam os parâmetros reais tomados como referência para a geração das observações.

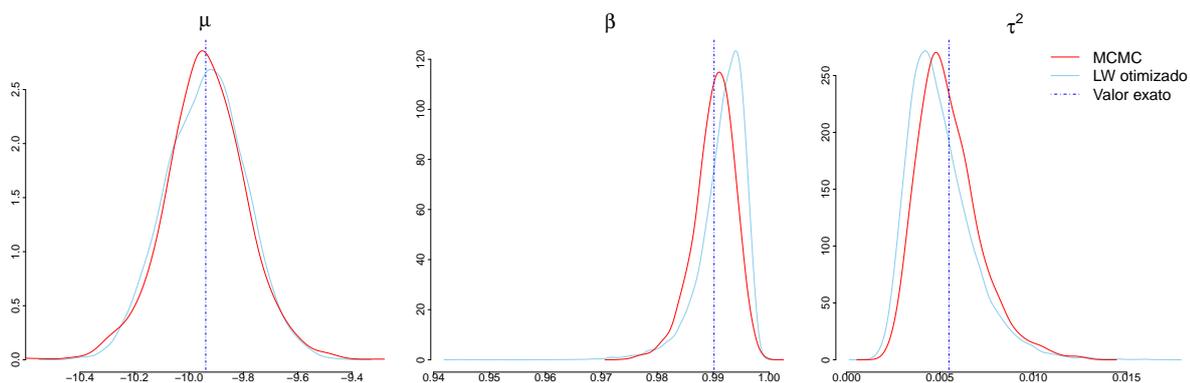


Figura 43 – Densidades *a posteriori* construídas pelo filtro de Liu e West otimizado com matriz de dependência Σ_2 (azul) e pelo método MCMC (vermelho).

Concluimos neste primeiro estudo que as duas tratativas são capazes de gerar inferências satisfatórias em cenários simulados, o que nos dá um bom indício de que de fato podemos utilizar estruturas de dependência mais genéricas no passo de rejuvenescimento artificial. Dando um passo adiante, investigaremos a seguir a relevância desta nova estrutura em aplicações envolvendo dados reais de modo a caracterizar cenários mais complexos constituídos por séries igualmente próximas da não estacionariedade porém sujeitas a interferências externas possivelmente não mensuradas pelo modelo.

Dólar Americano

Ilustramos na sequência o comportamento do mesmo dispositivo de filtragem para uma série real de taxas de câmbio relativas à comparação Dólar Americano \times Euro.

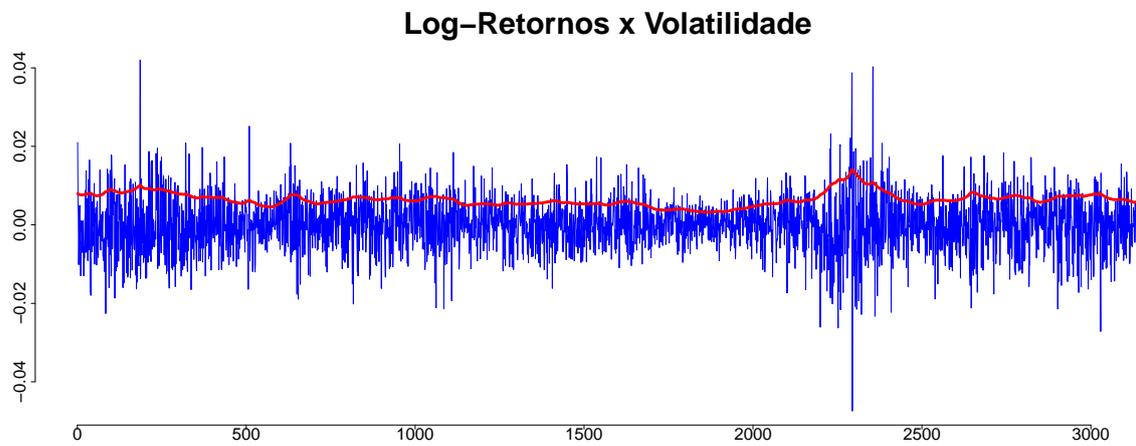


Figura 44 – Taxas de câmbio Dólar Americano \times Euro (azul) e volatilidades (vermelho) estimadas via MCMC.

Uma instabilidade significativa já pode ser verificada ao longo da evolução da incerteza *a posteriori* com respeito aos parâmetros μ e β . A fragilidade das estimativas é observada especificamente nesta dupla de parâmetros devido à forte correlação entre as duas quantidades. Durante o período de maior instabilidade verifica-se um coeficiente de correlação linear superior a 0,99. Podemos observar claramente que o início do período de conturbação na filtragem de β coincide com o início da fase de instabilidade na filtragem μ . A qualidade da inferência com respeito aos parâmetros estáticos é severamente comprometida a partir deste instante.

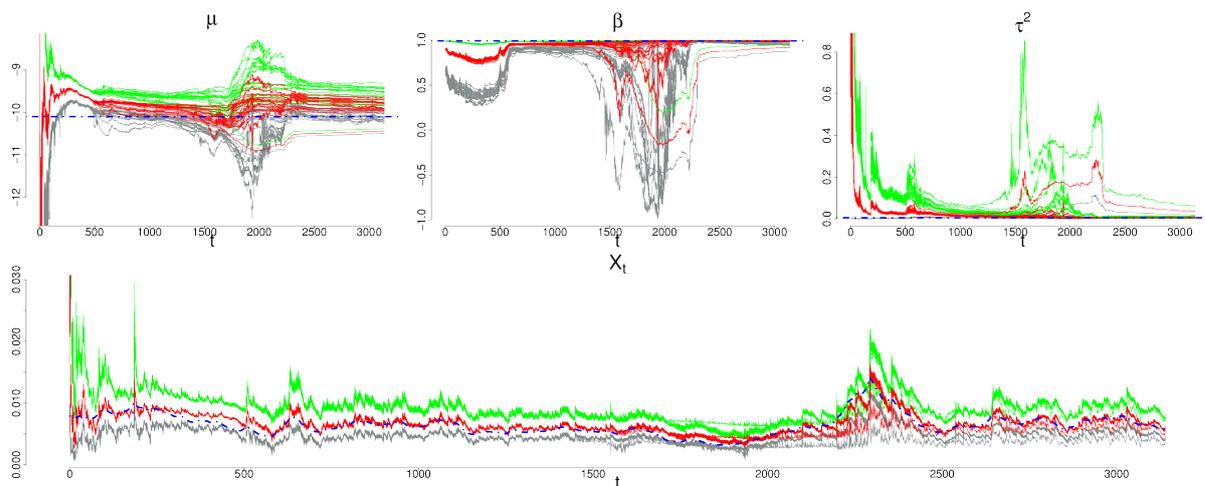


Figura 45 – Filtragem envolvendo taxas de câmbio Dólar Americano \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e matriz de covariância completa: médias *a posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam as quantidades estimadas via MCMC.

Caso implementado cegamente, este experimento constituiria um exemplar perfeito para a invalidação da estratégia LW mesmo considerando algumas das sugestões de otimização propostas aqui. Entretanto, tal comportamento pode ser entendido a partir de uma análise qualitativa do modelo. É natural que parâmetros associados a propriedades estacionárias apresentem uma maior dificuldade de estimação em tempo real uma vez que séries verdadeiras de dados não necessariamente seguem um modelo matemático de forma rígida. Neste caso em particular, a forte correlação entre μ e os demais parâmetros favorece a potencialização do problema, motivando a obtenção de resultados indevidos conforme ilustrado acima. A combinação da instabilidade de μ com a forte correlação entre parâmetros induz o surgimento de problemas de mau condicionamento na matriz de covariância, o que fatalmente reflete de forma negativa no rejuvenescimento da população via distribuição Gaussiana multivariada. A sugestão imediata para a correção deste tipo de problema seria a utilização de algoritmos mais estáveis de decomposição matricial no processo de geração das variáveis Gaussianas multivariadas responsáveis pelo rejuvenescimento artificial. Poderíamos, por exemplo, basear a geração de novos indivíduos na decomposição em valores singulares (DVS) da matriz Σ_1 na tentativa de contornar os problemas de mal condicionamento da mesma. Todavia, mesmo com a utilização da abordagem mais estável temos novamente um cenário degenerativo conforme ilustrado abaixo.

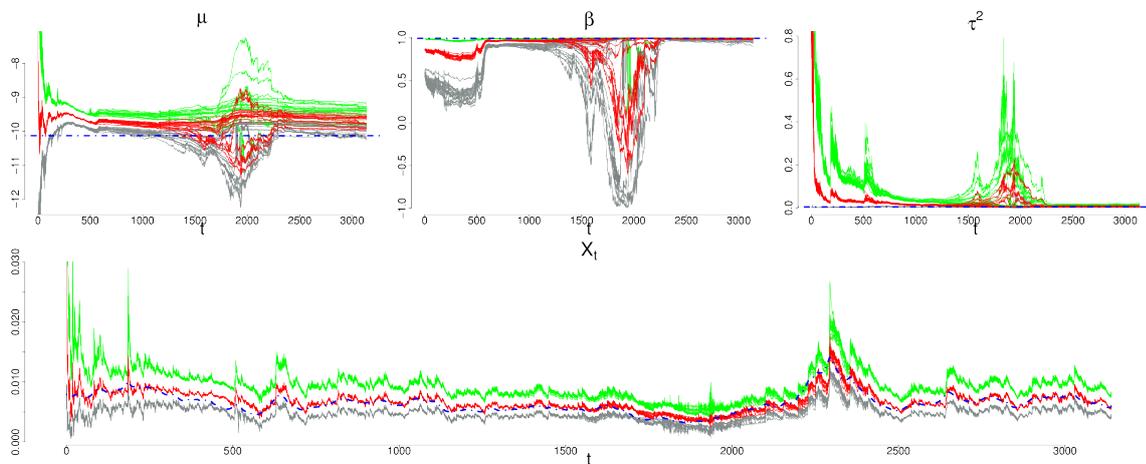


Figura 46 – Filtragem envolvendo taxas de câmbio Dólar Americano \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e matriz de covariância completa com decomposição em valores singulares: médias a *posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades estimadas via MCMC.

Ao entender esta peculiaridade do modelo, podemos implementar o filtro LW otimizado considerando uma atualização independente de parâmetros a partir da matriz Σ_2 descrita na relação (6.2). Os resultados oferecidos pela nova estrutura de dependência erradicam a influência da fragilidade de μ sobre os demais parâmetros, garantindo assim a construção de inferências mais robustas. O próximo diagrama sintetiza os resultados obtidos pelo filtro LW otimizado considerando blocos independentes de parâmetros. Nota-se que a evolução da incerteza com respeito a μ continua apresentado uma instabilidade proporcional à escala do parâmetro, contudo tal fragilidade não é propagada para as demais quantidades. Desta forma, a própria matriz de

dependências Σ_2 permanece mais estável ao longo da execução do algoritmo, o que garante que o processo de rejuvenescimento Gaussiano seja realizado de forma adequada.

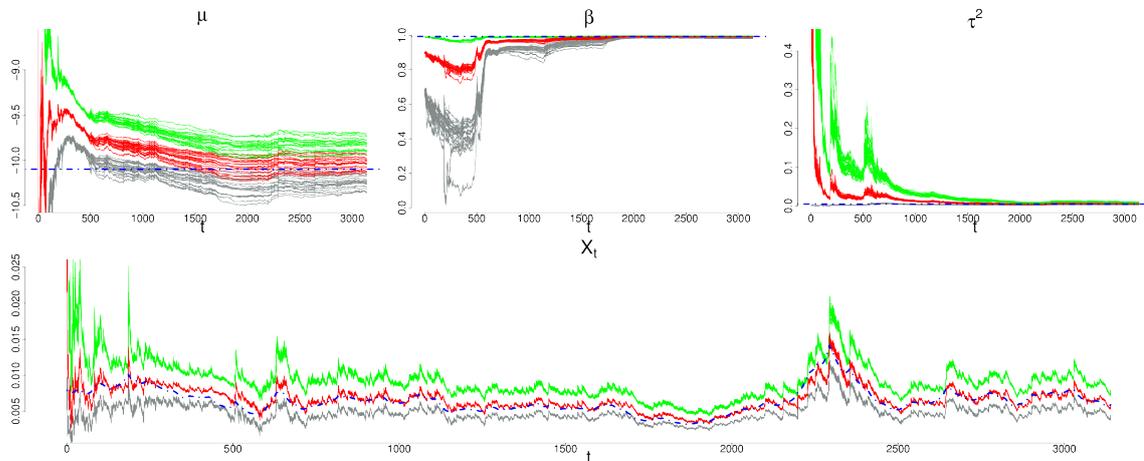


Figura 47 – Filtragem envolvendo taxas de câmbio Dólar Americano \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e blocos independentes de parâmetros: médias a *posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades estimadas via MCMC.

Ilustramos abaixo as densidades estimadas segundo esta perspectiva, considerando enxames de 50000 partículas. O número de pontos foi incrementado consideravelmente com relação aos experimentos anteriores pois a estimação da função de densidade constitui um problema naturalmente mais complexo. Ainda assim, o algoritmo continuou com um desempenho muito satisfatório em termos de tempo execução.

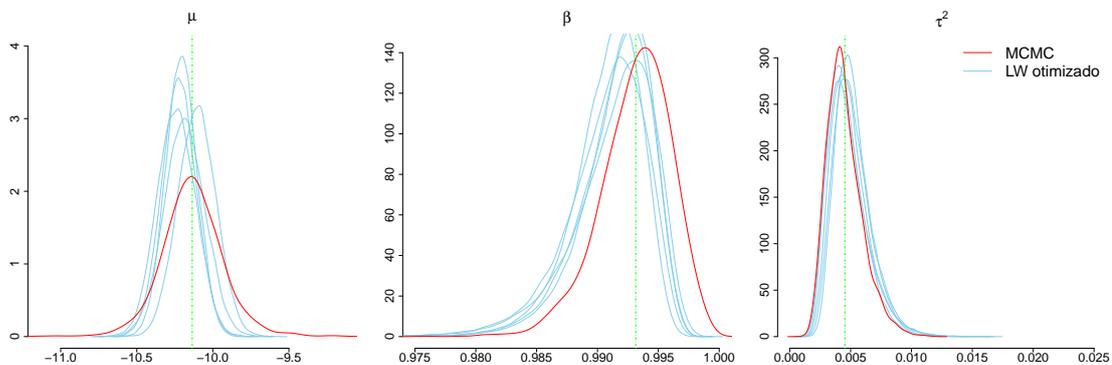


Figura 48 – Densidades estimadas via filtro LW otimizado estrutura de dependência bloco-independente. Cada curva corresponde a uma estimativa independente construída a partir de 50000 partículas.

Comportamentos idênticos foram obtidos considerando o Filtro *Bootstrap* associado à abordagem LW; os resultados não serão apresentados em maiores detalhes para evitar repetições. Apresentamos apenas um diagrama de dispersão a seguir comparando as estimativas médias, com respeito a 50 realizações, de cada um dos algoritmos ao longo dos 3139 instantes de tempo. Nota-se uma grande consistência entre as duas abordagens. As estimativas construídas a partir das duas perspectivas de filtragem, com respeito às 24 séries, serão sumarizados ao final da seção para efeitos de comparação.

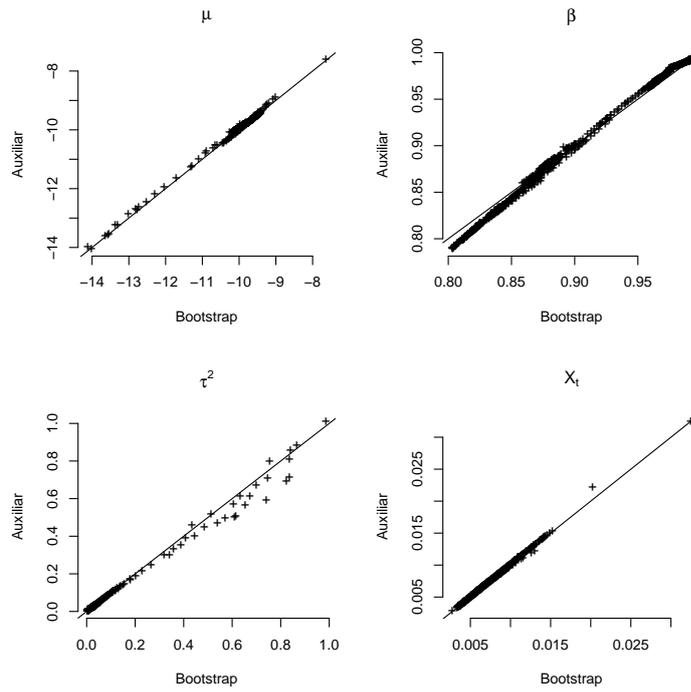


Figura 49 – Comparação entre os paradigmas Auxiliar e *Bootstrap*. Os quadros representam o diagrama de dispersão envolvendo a estimativa média das quantidades $(\mu, \beta, \tau^2, X_t)$, considerando 50 execuções de cada método, ao longo dos 3139 instantes de iteração.

Rublo Russo

A análise apresentada na seção anterior foi repetida para a série de taxas de câmbio relativas à comparação Rublo Russo×Euro.

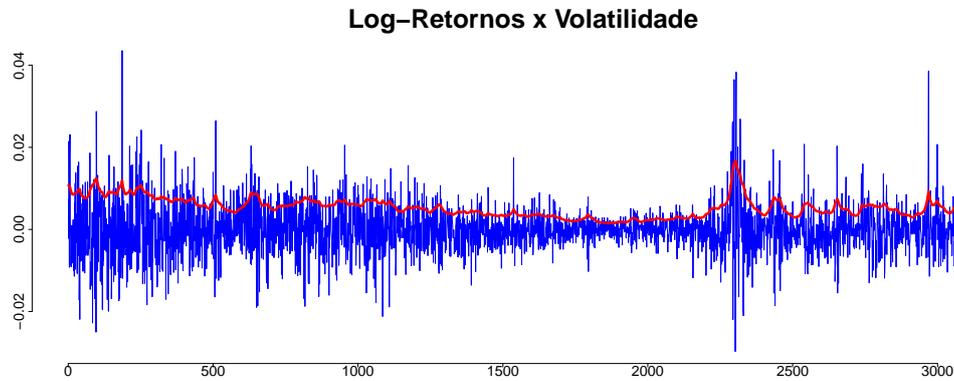


Figura 50 – Taxas de câmbio Rublo Russo × Euro (azul) e volatilidades (vermelho) estimadas via MCMC.

Os parâmetros estimados via MCMC neste caso foram

$$(\mu, \beta, \tau^2) = (-10,628; 0,988; 0,150^2),$$

os quais revelam uma persistência igualmente importante porém demonstram um montante de variabilidade significativamente superior na comparação com a série referente ao Dólar Americano. A utilização da estrutura de dependências tradicional apresentou resultados ainda mais catastróficos neste experimento. A maior fragilidade do algoritmo já era esperada nesta situação uma vez que a série de observações apresenta variações importantes em seu comportamento a partir do instante $t = 1750$. Todavia, o problema foi solucionado mais uma vez pela simples substituição da estrutura de dependência implementada no passo de rejuvenescimento.

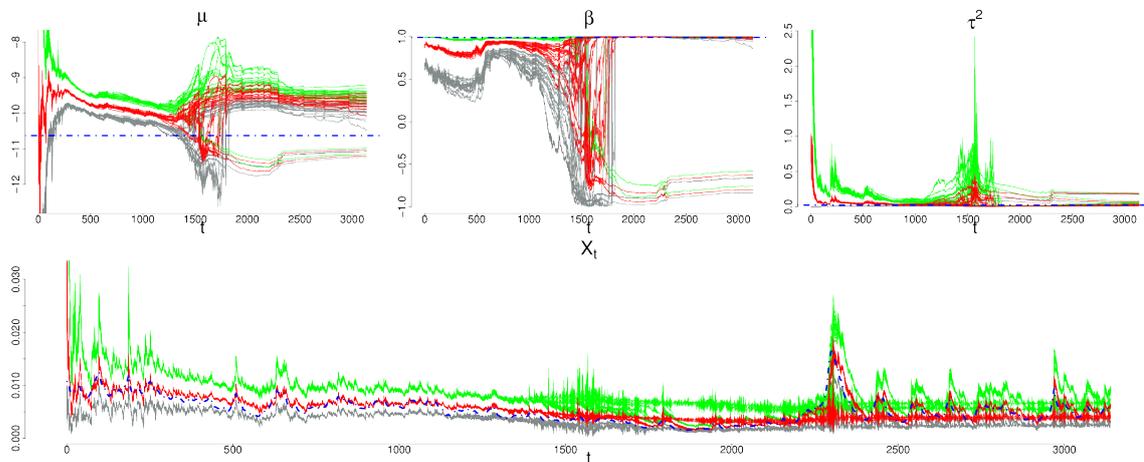


Figura 51 – Filtragem envolvendo taxas de câmbio Rublo Russo × Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e matriz de covariância completa: médias a *posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam as quantidades estimadas via MCMC.

Em resumo, o comportamento degenerativo se repetiu ao considerarmos a matriz de covariâncias completa no dispositivo de filtragem, assim como uma nova melhora foi verificada ao implementarmos o rejuvenescimento por blocos independentes. Este resultado é bastante relevante pois a verificação de desempenhos satisfatórios em situações problemáticas provenientes de séries distintas, pode ser entendida como um bom indicador da consistência do método. Observa-se apenas um vício discreto (na comparação com a escala do parâmetro) na estimativa do nível estacionário, o que certamente é fruto de um leve enxugamento da amostra induzido pelo aglomerado de volatilidade iniciado a partir do instante $t = 2250$. Nota-se também que este pequeno desvio não agrega nenhuma influência negativa à estimação do processo de volatilidade X_t .

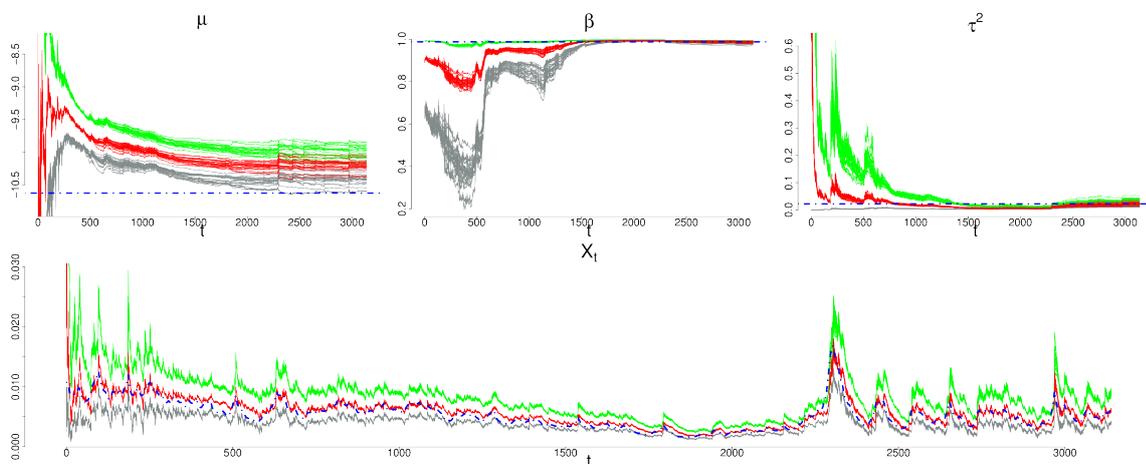


Figura 52 – Filtragem envolvendo taxas de câmbio Rublo Russo \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e blocos independentes de parâmetros: médias *a posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades estimadas via MCMC.

Analisamos também as estimativas funcionais das distribuições *a posteriori* construídas com o auxílio do dispositivo de filtragem. Nota-se que μ foi o único parâmetro a indicar um vício

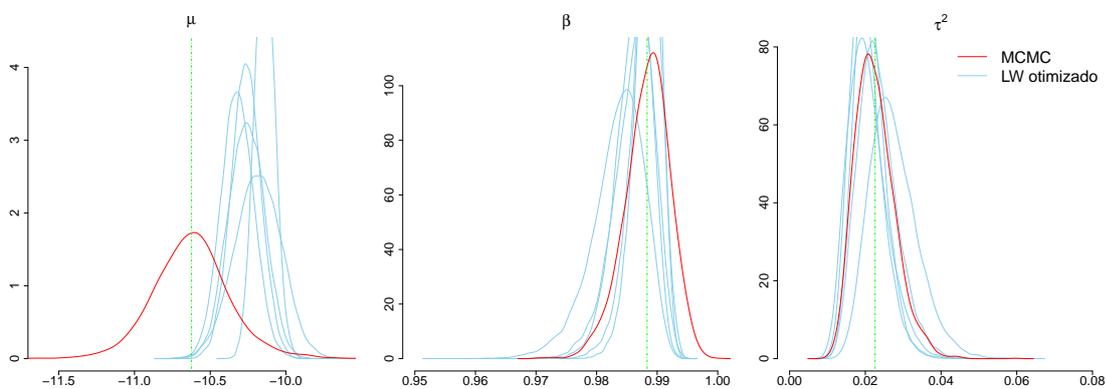


Figura 53 – Densidades estimadas via filtro LW otimizado estrutura de dependência bloco-independente. Cada curva corresponde a uma estimativa independente construída a partir de 50000 partículas.

mais importante, em conformidade com os resultados já observados na análise das trajetórias filtradas. Entretanto, mesmo caracterizando um caso explícito de degeneração, temos ainda

assim uma estimativa bastante razoável considerando a escala do parâmetro. Destacamos que a aproximação foi baseada na estratégia Auxiliar convencional; seria possível ainda considerar a distribuição proposta ótima através da construção de uma aproximação baseada em misturas de normais para os erros das observações.

Fazendo uma comparação entre os resultados construídos pelos paradigmas *Bootstrap* e Auxiliar, verificamos um comportamento muito consistente nas estimativas. Observa-se alguma diferença apenas nas estimativas do parâmetro β : a inferência construída pela perspectiva *Bootstrap* tende a subestimar discretamente os resultados obtidos por meio da abordagem Auxiliar. A maior diferença é verificada nas primeiras iterações onde a influência da distribuição *a priori* pouco informativa ainda é muito importante. A tentativa, por parte do dispositivo auxiliar, de prever tendências futuras das observações contribui para uma maior precisão do método nestas situações. O desvio entre as duas estimativas é praticamente erradicado na segunda metade da execução do experimento, conforme ilustrado na Figura 55.

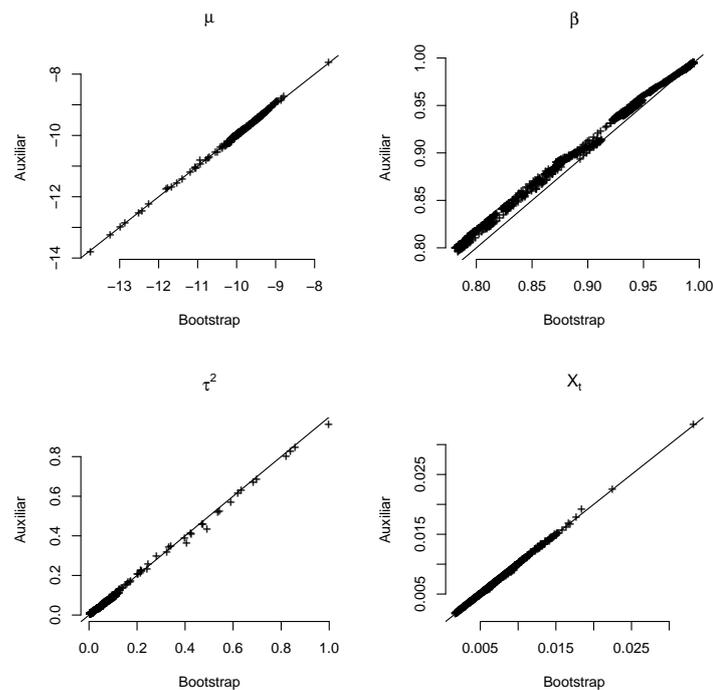


Figura 54 – Comparação entre os paradigmas Auxiliar e *Bootstrap*. Os quadros representam o diagrama de dispersão envolvendo a estimativa média das quantidades $(\mu, \beta, \tau^2, X_t)$, considerando 50 execuções de cada método, ao longo dos 3139 instantes de iteração.

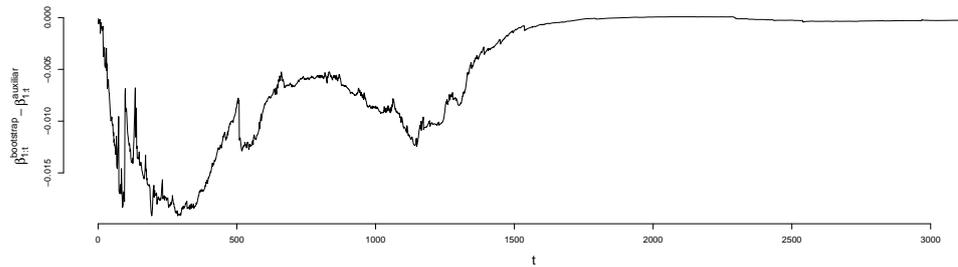


Figura 55 – Discrepância observada entre as estimativas $\mathbb{E}[\beta|y_{1:t}]$ construídas pelas perspectivas *Bootstrap* e *Auxiliar*.

Rúpia Indonésia

O próximo conjunto de observações refere-se às taxas de câmbio obtidas da comparação Rupiah Indonésia \times Euro. Este exemplo é interessante pois revela uma situação onde o próprio algoritmo MCMC apresentou uma maior dificuldade durante a construção das estimativas *a posteriori*.

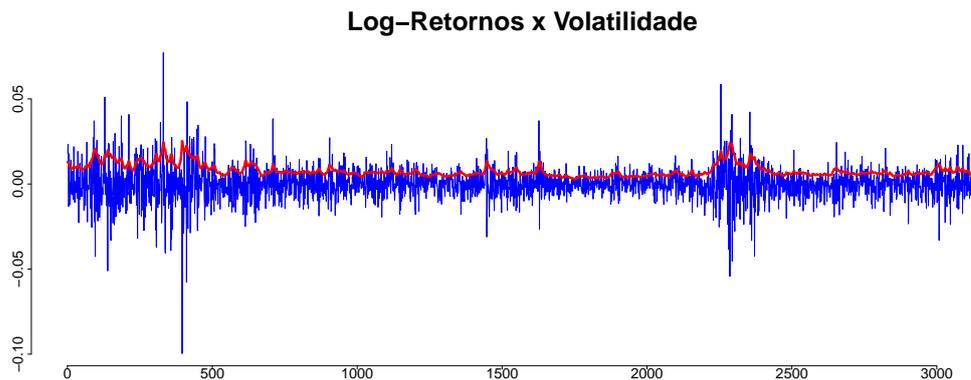


Figura 56 – Taxas de câmbio Rupiah Indonésia \times Euro (azul) e volatilidades (vermelho) estimadas via MCMC.

[Kastner e Frühwirth-Schnatter \(2014\)](#) mostram que o algoritmo MCMC gera observações altamente correlacionadas especificamente neste contexto, caracterizando assim um significativo fator de ineficiência. As estimativas

$$(\mu, \beta, \tau^2) = (-9,867; 0,966; 0,232^2).$$

foram construídas tomando um atraso (lagging) de 115 iterações entre cada geração do algoritmo na tentativa de assegurar uma amostra aproximadamente independente. Neste caso, a amostra final de tamanho 3000 foi extraída de uma sequência de 345000 indivíduos, o que ajuda a ilustrar o tamanho do esforço computacional empregado.

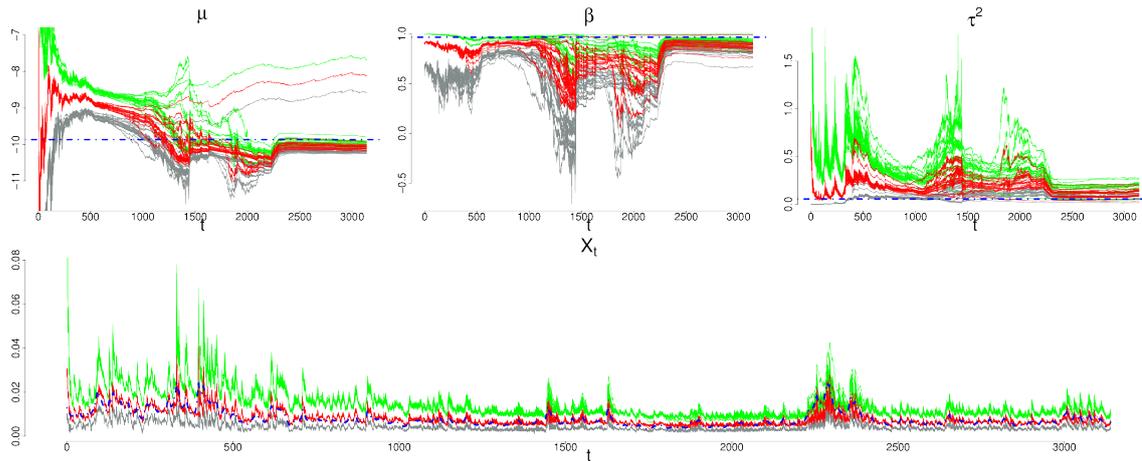


Figura 57 – Filtragem envolvendo taxas de câmbio Rupiah Indonésia \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e matriz de covariância completa: médias *a posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam as quantidades estimadas via MCMC.

Uma nova sequência de degenerações do método LW foi observada ao implementarmos a estrutura de dependências convencional apesar de não estarmos lidando agora com um parâmetro de persistência β muito próximo de 1. Ainda assim, a alta correlação evidenciada pelo método MCMC é um bom indicativo de que a saída para o problema pode estar diretamente associada à estrutura de dependências dos parâmetros. Novamente, a utilização da estrutura Σ_2 foi suficiente para corrigir o problema.

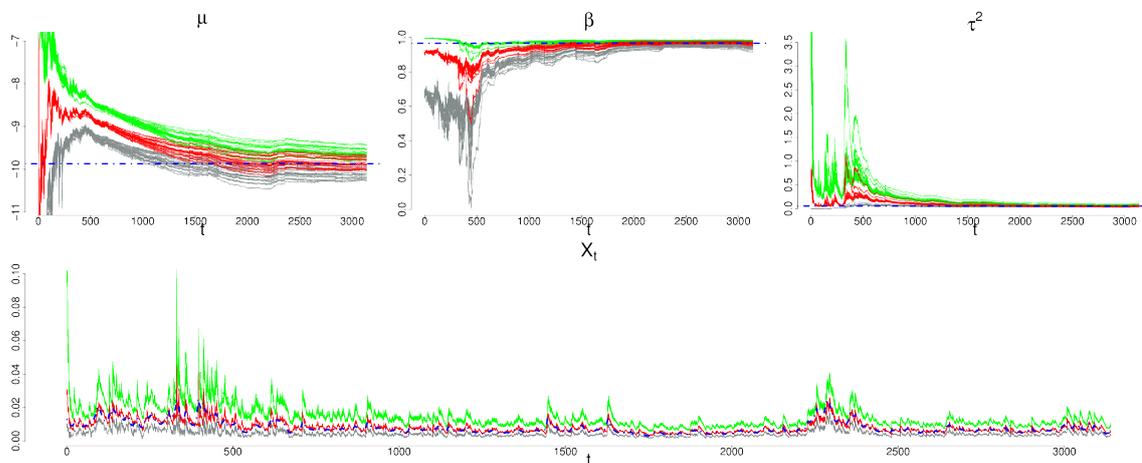


Figura 58 – Filtragem envolvendo taxas de câmbio Rupiah Indonésia \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e blocos independentes de parâmetros: médias *a posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades estimadas via MCMC.

Este resultado evidencia que o método de Liu e West otimizado é capaz de gerar inferências consistentes mesmo em situações onde os algoritmos MCMC apresentam uma certa dificuldade. Até mesmo a forma funcional das distribuições *a posteriori* apresentaram um comportamento satisfatório na comparação com as estimativas provenientes da estratégia MCMC.

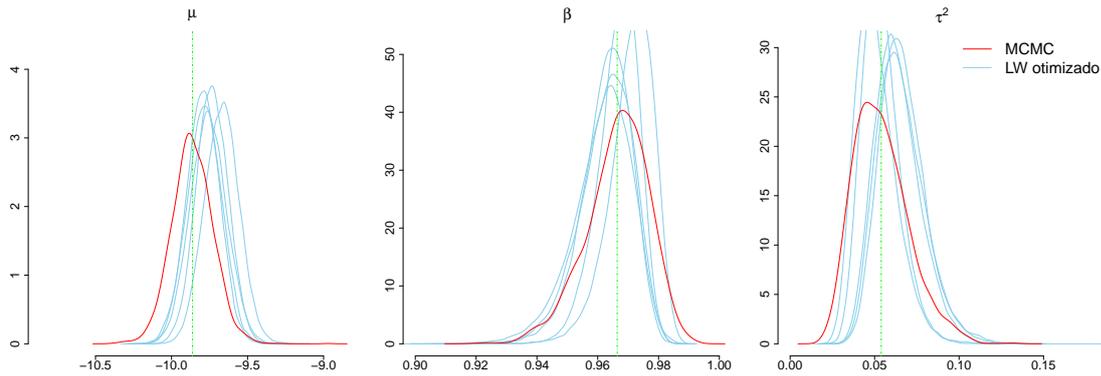


Figura 59 – Densidades estimadas via filtro LW otimizado estrutura de dependência bloco-independente. Cada curva corresponde a uma estimativa independente construída a partir de 50000 partículas.

O ganho é relevante pois além de tratar de uma implementação mais simples, a metodologia sequencial oferece uma estimativa mais rica ao precisar em tempo real a influência de pequenos blocos de observações na incerteza *a posteriori*. Além disso, a utilização de um atraso de 115 iterações no método MCMC induziu um tempo computacional mais de 10 vezes superior ao obtido pelo filtro de partículas mesmo considerando a concorrência desleal entre as linguagens de programação (MCMC/C++ \times filtro de partículas/R).

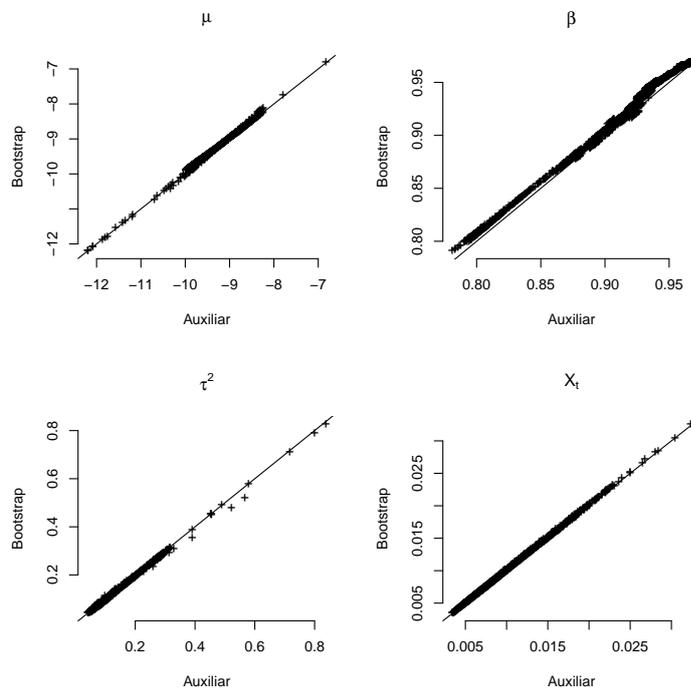


Figura 60 – Comparação entre os paradigmas Auxiliar e *Bootstrap*. Os quadros representam o diagrama de dispersão envolvendo a estimativa média das quantidades $(\mu, \beta, \tau^2, X_t)$, considerando 50 execuções de cada método, ao longo dos 3139 instantes de iteração.

Coroa Dinamarquesa

Por fim, detalhamos um conjunto de dados cuja análise via filtros de partículas é satisfatória mesmo considerando matrizes de covariâncias completas. Ao ilustrar um cenário mais favorável, buscamos evidenciar, assim como no caso simulado, que a utilização de uma estrutura alternativa de dependências também gera bons resultados mesmo em situações onde a perspectiva convencional já é suficiente para a construção de estimativas aceitáveis.

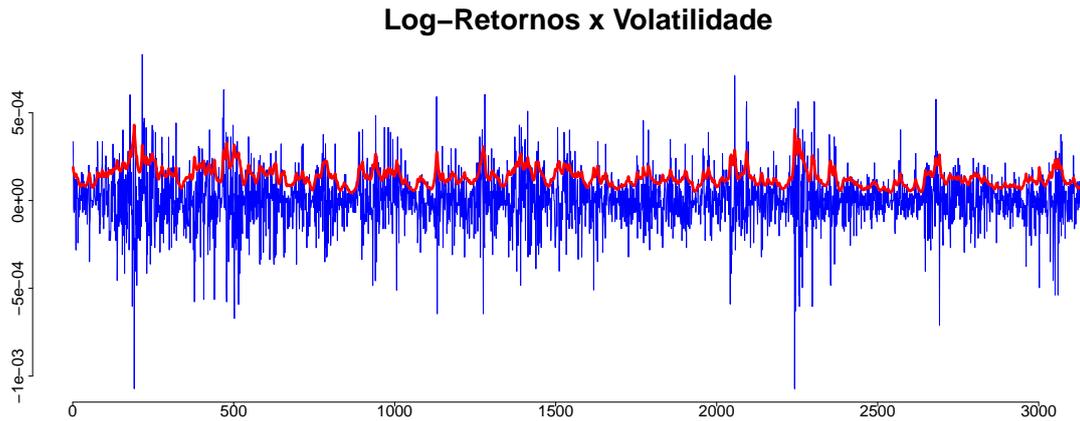


Figura 61 – Taxas de câmbio Coroa Dinamarquesa \times Euro (azul) e volatilidades (vermelho) estimadas via MCMC.

Os dados correspondem à sequência de taxas de câmbio referentes à comparação Coroa Dinamarquesa \times Euro a qual pode ser associada à seguinte tripla de parâmetros estáticos, estimados via MCMC,

$$(\mu, \beta, \sigma^2) = (-18, 0; 0, 916; 0, 38^2).$$

A série de volatilidade conta com um parâmetro de persistência bastante distante da fronteira da não-estacionariedade, logo devemos de fato esperar estimativas satisfatórias sem maiores esforços. Além disso, nota-se que a série apresenta um nível estacionário muito pequeno na comparação com os estudos anteriores. Esta peculiaridade não gera maiores complicações já que uma distribuição *a priori* vaga foi atribuída ao parâmetro μ com o intuito de varrer adequadamente o espaço paramétrico. O mecanismo de reamostragem suavizada evita que a população seja enxugada de forma acelerada garantindo a construção de estimativas estáveis. O resumo a seguir mostra que o método é capaz de filtrar os três parâmetros desconhecidos sem maiores complicações.

A utilização da estrutura de dependências Σ_2 também é capaz de gerar boas estimativas, conforme identificado no resumo a seguir. É possível até observar uma certa redução na variabilidade das estimativas, contudo o objetivo é simplesmente verificar a consistência do método com respeito às situações onde a metodologia tradicional comporta-se de forma satisfatória.

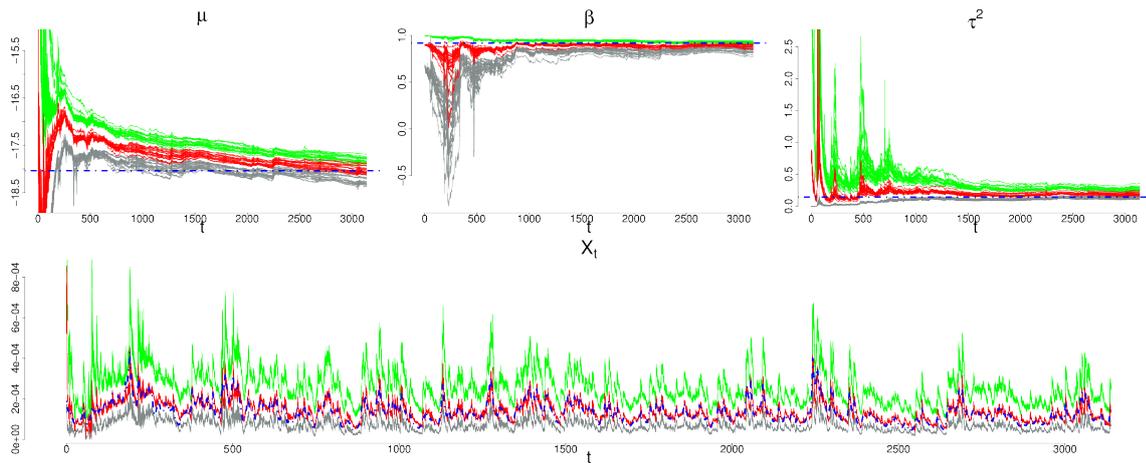


Figura 62 – Filtragem envolvendo taxas de câmbio Coroa Dinamarquesa \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e matriz de covariância completa: médias *a posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam as quantidades estimadas via MCMC.

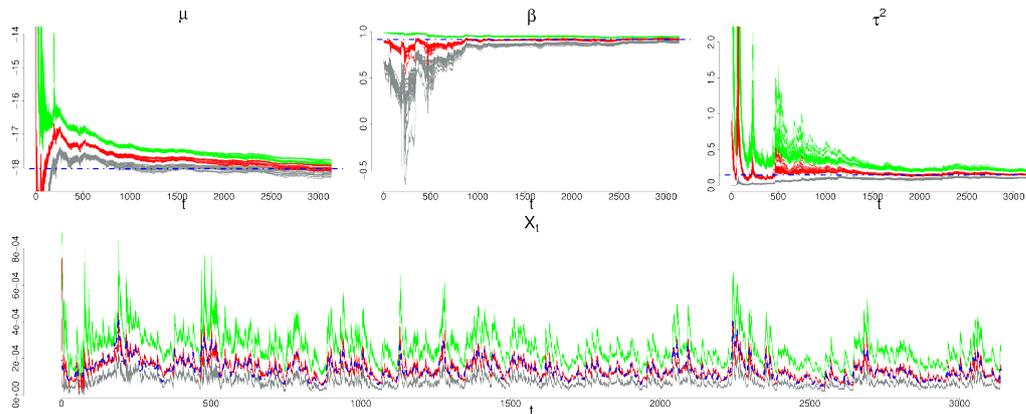


Figura 63 – Filtragem envolvendo taxas de câmbio Coroa Dinamarquesa \times Euro. Resumo de desempenho do método de Liu e West otimizado com 20000 partículas e blocos independentes de parâmetros: médias *a posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam quantidades estimadas via MCMC.

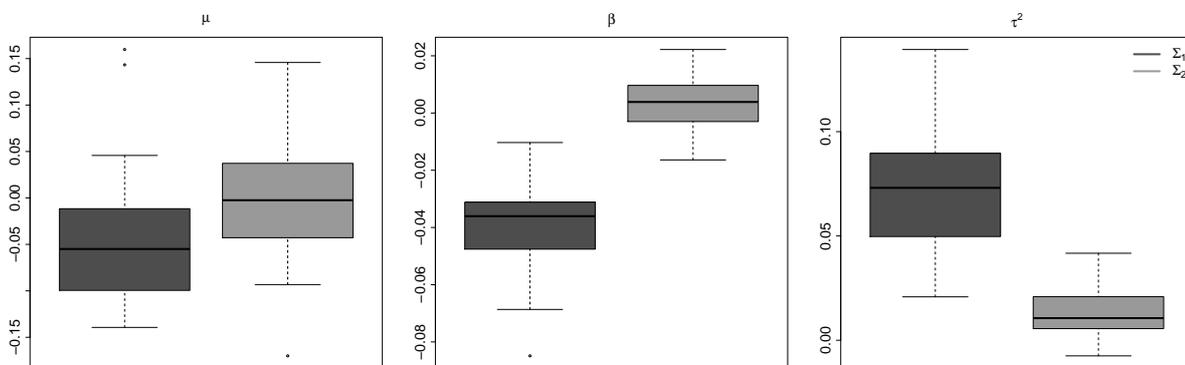


Figura 64 – Erros cometidos pelo dispositivo de filtragem na estimação dos parâmetros estáticos (μ, β, τ^2) , ao longo de 50 execuções, considerando a utilização das estruturas de dependência Σ_1 e Σ_2 .

A comparação entre as abordagens *Bootstrap* e Auxiliar aplicada a este cenário nos

permite tirar algumas conclusões importantes. A filtragem dos parâmetros apresenta uma certa dificuldade inicial já que a incerteza *a priori* com respeito ao nível estacionário não está centrada em torno do valor real como nas situações anteriores. Além disso, a variabilidade associada ao processo latente é bastante elevada considerado o contexto dos modelos de volatilidade estocástica. Neste caso, a previsão realizada na perspectiva Auxiliar é capaz de conferir um maior foco ao dispositivo de filtragem, o que garante uma eliminação mais eficiente dos indivíduos menos adaptados da amostra. Temos um fenômeno similar ao observado na série referente ao Rublo Russo, entretanto observado agora em uma escala maior. Podemos observar na Figura 66 que novamente a diferença entre as duas perspectivas praticamente desaparece à medida que a filtragem avança.

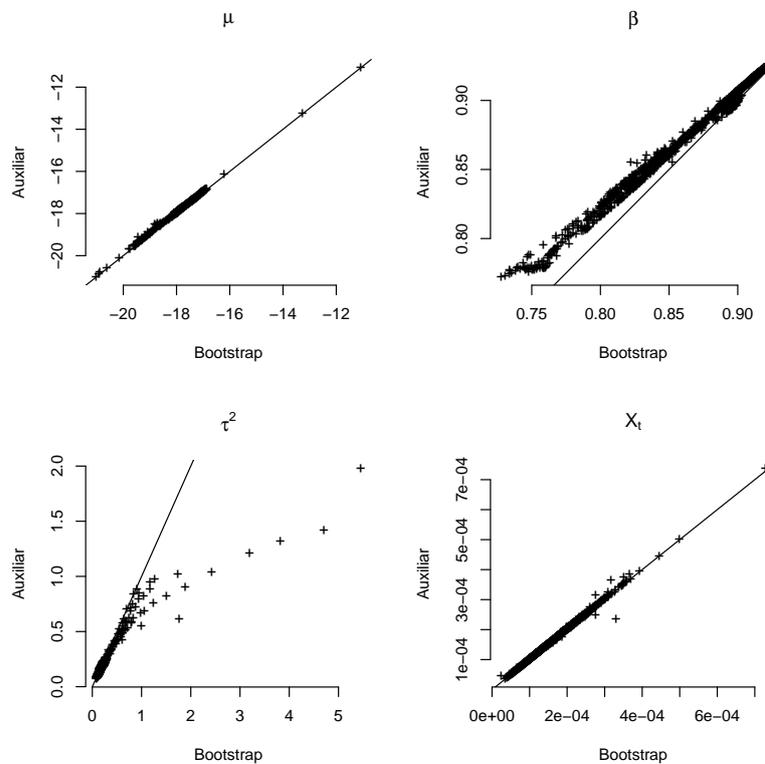


Figura 65 – Comparação entre os paradigmas Auxiliar e *Bootstrap*. Os quadros representam o diagrama de dispersão envolvendo a estimativa média das quantidades $(\mu, \beta, \tau^2, X_t)$, considerando 50 execuções de cada método, ao longo dos 3139 instantes de iteração.

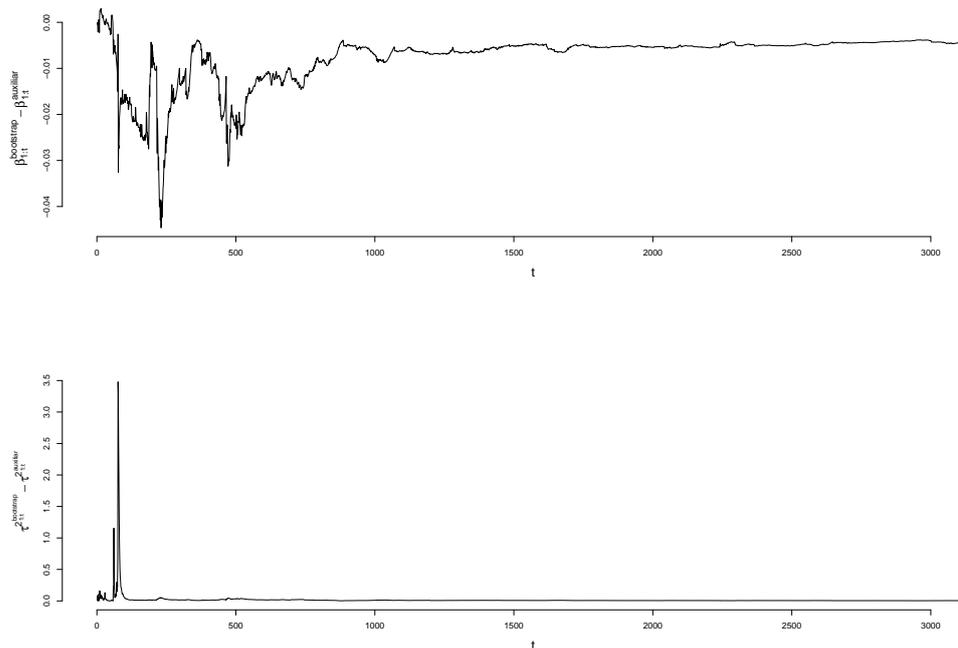


Figura 66 – Discrepância observada entre as estimativas $\mathbb{E}[\beta|y_{1:t}]$ e $\mathbb{E}[\tau^2|y_{1:t}]$ construídas pelas perspectivas *Bootstrap* e *Auxiliar*.

Resultado Geral

O conjunto de dados tratado neste estudo foi escolhido de modo a ilustrar o comportamento do filtro LW otimizado em situações responsáveis por introduzir dificuldades de naturezas distintas. A partir de comparações com o caso ótimo, método MCMC, foi possível comprovar experimentalmente a capacidade inferencial do filtro de Liu e West otimizado de forma honesta. Dentre as sugestões propostas neste trabalho destacamos principalmente a relevância de uma análise qualitativa do modelo na escolha da estrutura de dependências utilizada no dispositivo de rejuvenescimento. Outro ponto importante é a escolha dos parâmetros a e h^2 do algoritmo; o mesmo mecanismo de calibragem automática foi utilizado em todos os 24 cenários e demonstrou ser muito eficaz. Até onde sabemos, nenhuma outra proposta na literatura havia trabalhado com uma escolha genérica em um conjunto tão amplo de séries. De uma forma geral, a metodologia apresentou resultados satisfatórios considerando tanto a perspectiva *Bootstrap* quanto a abordagem *Auxiliar*. Observamos apenas pequenas discrepâncias na filtragem das primeiras observações de cada série, o que constitui um resultado natural já que, na ausência da distribuição ótima, espera-se que a perspectiva *Auxiliar* confira mais foco às estimativas. A tabela a seguir resume os resultados obtidos por cada uma das metodologias. A primeira coluna de parâmetros indica as estimativas construídas através do método MCMC com 3000 pontos, as quais podemos considerar como inferência ótima no desenvolvimento de um estudo comparativo. As duas colunas restantes apresentam uma média de 50 estimativas construídas com o auxílio do filtro LW otimizado com 15000 partículas considerando, respectivamente, as implementações

Bootstrap e *Auxiliar*. Os valores destacados entre parênteses correspondem à raiz quadrada do erro quadrático médio cometido ao longo das 50 execuções na comparação com as estimativas oriundas do algoritmo MCMC. Nota-se, em todas as situações, que as estimativas do filtro LW são condizentes com o resultado ótimo esperado. Observamos ainda uma redução média de 35% (*Bootstrap*) e 20% (*Auxiliar*) no tempo computacional demandado pelas estratégias otimizadas em comparação com filtros implementados sem a utilização de agendas dinâmicas de rejuvenescimento. Os métodos foram implementados na linguagem R e o tamanho do enxame de partículas foi escolhido visando assegurar uma menor variabilidade das estimativas finais além de um tempo de execução significativamente superior ao apresentado pelo método MCMC. O tempo de uma execução do filtro de partículas foi inferior à metade do tempo demandado pela metodologia ótima. Destacamos que, apesar de tratar de uma biblioteca desenvolvida para a linguagem R, o dispositivo baseado em cadeias Markovianas desenvolvido por [Kastner e Frühwirth-Schnatter \(2014\)](#) foi implementado na linguagem C++. Temos assim um bom indicativo da eficiência computacional dos filtros de partículas: estimativas baseadas em 15000 pontos construídas em uma plataforma de linguagem interpretada apresenta um desempenho de tempo ao menos duas vezes superior em comparação com uma abordagem apoiada por 3000 pontos implementada em linguagem compilada. Desta forma, podemos esperar um desempenho de alta eficiência a partir da implementação do filtro LW otimizado em linguagens mais eficientes.

Moeda	MCMC			LW <i>Bootstrap</i>			LW <i>Auxiliar</i>		
	μ	β	τ	μ	β	τ	μ	β	τ
Dólar Australiano	-10,3	0,976	0,17	-10,207 (0,088)	0,972 (0,006)	0,174 (0,018)	-10,206 (0,083)	0,976 (0,004)	0,156 (0,017)
Dólar Canadense	-10,1	0,987	0,09	-10,130 (0,069)	0,981 (0,007)	0,100 (0,016)	-10,083 (0,095)	0,007 (0,006)	0,016 (0,013)
Coroa Tcheca	-11,5	0,953	0,28	-11,660 (0,157)	0,957 (0,008)	0,263 (0,029)	-11,993 (0,159)	0,962 (0,012)	0,246 (0,044)
Coroa Dinamarquesa	-18,0	0,916	0,38	-18,051 (0,062)	0,914 (0,008)	0,400 (0,025)	-18,014 (0,072)	0,918 (0,008)	0,393 (0,023)
Dólar de Hong Kong	-10,2	0,993	0,07	-10,103 (0,083)	0,991 (0,002)	0,072 (0,006)	-10,038 (0,145)	0,991 (0,002)	0,072 (0,006)
Rupiah Indonésia	-9,9	0,966	0,23	-9,926 (0,130)	0,960 (0,011)	0,235 (0,018)	-9,970 (0,120)	0,964 (0,007)	0,228 (0,017)
Yen Japonês	-10,0	0,989	0,12	-9,757 (0,280)	0,988 (0,002)	0,122 (0,005)	-9,660 (0,375)	0,990 (0,002)	0,117 (0,006)
Won Sul-Coreano	-10,0	0,987	0,14	-10,134 (0,162)	0,986 (0,003)	0,136 (0,012)	-10,131 (0,184)	0,988 (0,003)	0,130 (0,015)
Ringgit Malaio	-10,3	0,990	0,08	-10,178 (0,149)	0,988 (0,003)	0,085 (0,006)	-10,100 (0,224)	0,990 (0,002)	0,083 (0,005)
Peso Mexicano	-9,8	0,977	0,16	-9,520 (0,296)	0,983 (0,008)	0,138 (0,027)	-9,496 (0,315)	0,984 (0,008)	0,142 (0,026)
Dólar Neo-Zelandês	-10,0	0,963	0,17	-10,087 (0,094)	0,961 (0,006)	0,178 (0,016)	-10,090 (0,095)	0,961 (0,005)	0,178 (0,013)
Coroa Norueguesa	-11,1	0,970	0,18	-11,324 (0,228)	0,967 (0,006)	0,173 (0,019)	-11,311 (0,216)	0,970 (0,005)	0,162 (0,025)
Peso Filipino	-10,1	0,981	0,12	-10,203 (0,126)	0,979 (0,007)	0,109 (0,0160)	-10,062 (0,146)	0,987 (0,009)	0,095 (0,027)
Zloty Polonês	-10,4	0,975	0,19	-10,510 (0,121)	0,974 (0,002)	0,183 (0,011)	-10,486 (0,096)	0,975 (0,002)	0,180 (0,013)
Leu Romeno	-11,1	0,970	0,31	-10,283 (0,830)	0,978 (0,008)	0,306 (0,025)	-10,171 (0,941)	0,979 (0,009)	0,312 (0,023)
Rublo Russo	-10,6	0,988	0,15	-10,206 (0,407)	0,988 (0,003)	0,149 (0,017)	-10,247 (0,363)	0,988 (0,003)	0,142 (0,024)
Dólar de Singapura	-10,6	0,995	0,07	-10,335 (0,353)	0,993 (0,003)	0,074 (0,005)	-10,082 (0,546)	0,995 (0,001)	0,074 (0,006)
Coroa Sueca	-11,3	0,991	0,11	-11,232 (0,083)	0,988 (0,003)	0,115 (0,006)	-11,220 (0,094)	0,988 (0,003)	0,115 (0,006)
Franco Suíço	-12,0	0,985	0,21	-12,240 (0,248)	0,980 (0,006)	0,201 (0,027)	-12,253 (0,260)	0,981 (0,005)	0,199 (0,021)
Baht Tailandês	-10,2	0,98	0,13	-10,025 (0,243)	0,983 (0,006)	0,122 (0,013)	-9,867 (0,371)	0,987 (0,008)	0,117 (0,016)
Lira Turca	-9,8	0,966	0,27	-10,164 (0,380)	0,966 (0,007)	0,283 (0,018)	-10,190 (0,403)	0,968 (0,007)	0,274 (0,017)
Libra Esterlina	-10,8	0,992	0,10	-10,922 (0,177)	0,992 (0,001)	0,091 (0,009)	-10,767 (0,125)	0,993 (0,001)	0,090 (0,010)
Dólar Americano	-10,1	0,993	0,07	-10,090 (0,067)	0,991 (0,002)	0,072 (0,004)	-10,056 (0,081)	0,991 (0,002)	0,073 (0,005)
Real Brasileiro	-9,4	0,960	0,26	-9,352 (0,124)	0,969 (0,010)	0,231 (0,030)	-9,322 (0,150)	0,971 (0,012)	0,224 (0,036)

Tabela 3 – Comparação MCMC \times Filtros de partículas. A primeira coluna de valores resume as estimativas obtidas a partir do método MCMC com 3000 pontos. As duas seguintes apresentam as médias dos parâmetros estimados ao longo de 50 iterações de cada filtro de partículas com 15000 pontos; valores entre paranteses indicam a raiz do erro quadrático médio, na comparação com o método MCMC, considerando as 50 iterações.

6.2 Dinâmica Populacional

Nesta seção lidaremos com um modelo populacional o qual foi estudado por [Peters, Hosack e Hayes \(2010\)](#) a partir de uma variação de algoritmo PMCMC e por [Chopin, Jacob e Papaspiliopoulos \(2013\)](#) (material suplementar) através do método SMC². Esta aplicação é interessante pois a estruturação matemática do modelo dificulta a utilização de algoritmos baseados em estatísticas suficientes e, neste caso, a melhor alternativa até então seria de fato a utilização de algoritmos como PMCMC ou SMC². Contudo, verificaremos que o método de Liu e West otimizado pode também oferecer soluções muito satisfatórias a partir de um custo computacional incomparavelmente menor.

6.2.1 Descrição do Modelo

A dinâmica populacional considerada neste conjunto de experimentos pode ser sucintamente modelada pelo par de equações

$$\begin{cases} y_t = \log(N_t) + \sigma \varepsilon_t, & \varepsilon_t \sim N(0, 1) \\ \log(N_t) = \log(N_t) + b_0 + b_2(N_t)^{b_3} + \tau \eta_t, & \eta_t \sim N(0, 1) \\ \log(N_0) = \mu_0 \end{cases}, \quad t = 1, 2, \dots, n \quad (6.3)$$

onde N_t representa o tamanho populacional real a ser estimado, y_t corresponde ao logaritmo dos tamanhos populacionais observados e ε_t, η_t , com $t \in \{1, 2, \dots, n\}$, são coleções de variáveis independentes e identicamente distribuídas segundo o modelo Gaussiano padrão. O vetor

$$\theta = (\mu_0, b_0, b_2, b_3, \tau^2, \sigma^2)$$

define os parâmetros estáticos desconhecidos do problema. Neste caso, a população inicial μ_0 referente à sequência de estados latentes também é tratada como um parâmetro desconhecido. O modelo é completado pelas especificações

parâmetro	dist. <i>a priori</i>
μ_0	$N(0, 4)$
b_0	$N(0, 1)$
b_2	$N(0, 1)$
b_3	$N(0, 1)$
σ^2	$GI\left(\frac{n}{2}, \frac{n-2}{10}\right)$
τ^2	$GI\left(\frac{n}{2}, \frac{n-2}{10}\right)$

as quais seguem rigorosamente a sugestão proposta no experimento original de [Peters, Hosack e Hayes \(2010\)](#).

6.2.2 Avaliação Experimental: Modelo θ -Estocástico

O modelo de observações (6.3) traduz o fato de que as populações de interesse são monitoradas a partir de mecanismos imperfeitos de observação, podendo então estar sujeitos a qualquer fonte de variabilidade introduzida pelo dispositivo de geração/coleta dos dados. Seguindo a discussão proposta originalmente, este modelo conta com duas peculiaridades com respeito à distribuição *a posteriori* de θ : serrilhamento e multimodalidade. Nosso objetivo é demonstrar experimentalmente que o filtro LW otimizado é capaz de reproduzir as duas propriedades, mesmo em se tratando de uma abordagem mais simples, de modo a ilustrar que o método vai além de uma aproximação razoável capaz de preservar apenas os dois primeiros momentos das distribuições *a posteriori*. A avaliação será baseada em duas etapas: um conjunto simulado e um conjunto real. Consideraremos as duas perspectivas de filtragem (*Auxiliar* e *Bootstrap*) e buscaremos evidenciar a importância das estratégias de otimização propostas neste trabalho. A fim de estabelecer uma comparação justa, apresentaremos também os resultados obtidos pelos métodos PMCMC e SMC² considerando um número similar de partículas.

Apesar de dificultar a implementação de métodos baseados na propagação de estatísticas suficientes, a estrutura deste problema é favorável ao desenvolvimento de filtros de partículas por permitir a construção da distribuição proposta ideal para os passos de amostragem por importância. A estrutura linear das observações associada ao fato de que as sequências de ruídos são Gaussianas culmina no seguinte modelo:

$$\log(N_t) | \log(N_{1:t-1}), \theta, y_t \sim N \left(\frac{y_t \cdot \sigma^{-2} + \left(\log(N_{t-1}) + b_0 + b_2 \cdot N_{t-1}^{b_3} \right) \tau^{-2}}{\sigma^{-2} + \tau^{-2}}, \frac{1}{\sigma^{-2} + \tau^{-2}} \right). \quad (6.4)$$

Neste caso, as ponderações utilizadas tanto pela perspectiva *Auxiliar* quanto pela *Bootstrap* são governadas pelo modelo probabilístico associado a $Y_t | N_{t-1}$:

$$W_t(\log(N_t), \theta | y_t) \sim N \left(\log(N_t) + b_0 + b_2 \cdot N_{t-1}^{b_3}, \sigma^2 + \tau^2 \right). \quad (6.5)$$

Esta propriedade é crucial para a construção de um dispositivo de filtragem ótima neste cenário já que as distribuições envolvidas apresentam um comportamento naturalmente complicado. É de se esperar que a utilização de distribuições propostas cegas tenham um desempenho bastante inadequado nesta aplicação específica.

Um último detalhe de implementação diz respeito ao tamanho h da janela utilizada na aproximação por núcleos estimadores do dispositivo LW. Como nosso objetivo é evidenciar uma possível característica de multimodalidade do problema, adotamos a opção

$$h = \frac{1}{2} \left(\frac{4}{N \cdot (2 + d_\theta)} \right)^{\frac{1}{4 + d_\theta}},$$

em conformidade com as sugestões propostas por [Silverman \(1986\)](#) e [Liu, Chen e Logvinenko \(2001\)](#).

Dados Artificiais

Consideramos inicialmente um cenário artificial visando analisar o comportamento dos algoritmos em uma situação "bem comportada". O experimento foi baseado na escolha de parâmetros proposta por [Chopin, Jacob e Papaspiliopoulos \(2013\)](#):

$$(\mu_0, b_0, b_2, b_3, \tau^2, \sigma^2) = (\log(1,270); 0,150; -0,125; 0,100; 0,470^2; 0,390^2).$$

Os autores argumentam que esta configuração estimula o surgimento de picos de crescimento populacional. Situações como esta, caracterizadas por mudanças relativamente bruscas, favore-

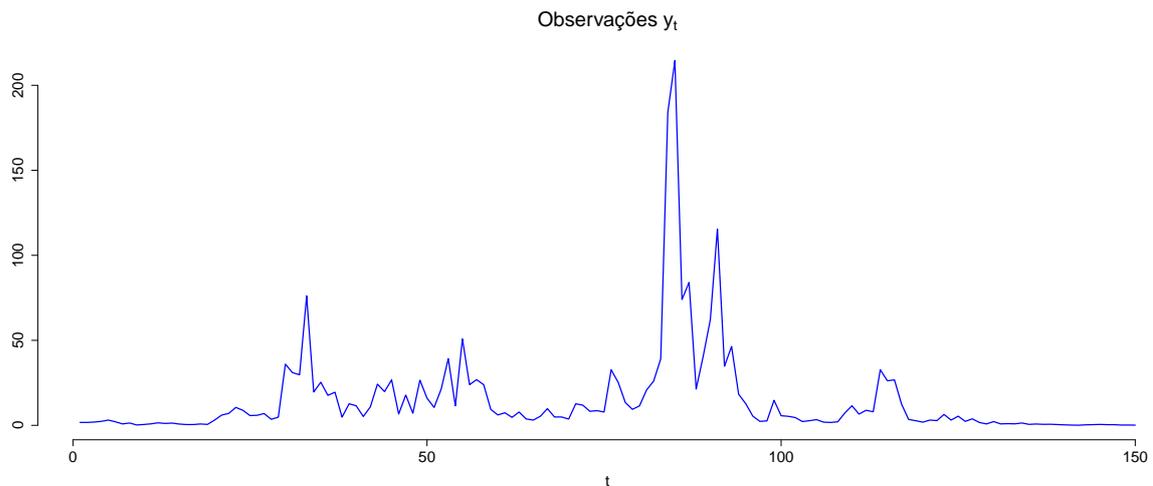


Figura 67 – Exemplo de evolução populacional obtida por simulação. As observações são dadas em escala real.

cem a degeneração de algoritmos do tipo filtro de partículas, logo, este cenário seria bastante frutífero para uma análise a respeito do poder de rejuvenescimento assegurado pelo método SMC². Os autores sequer consideraram a implementação do método de Liu e West por confiar que a estratégia apresentaria um comportamento degenerado análogo ao apontado pelos mesmos em experimentos anteriores.

Diferentemente do ensaio original sugerido por Peters *et al.*, um par de distribuições *a priori* não-informativas foi atribuído aos parâmetros de volatilidade na implementação trabalhada por Chopin *et al.*:

$$\tau^2 \sim GI(2, 1) \quad \text{e} \quad \sigma^2 \sim GI(2, 1).$$

Além disso, o método SMC² foi implementado a partir de uma nuvem inicial de 10^6 partículas:

$$N_x = 1000 \quad \text{e} \quad N_\theta = 1000.$$

Neste primeiro estudo, o filtro LW otimizado teve sua implementação baseada no conjunto de distribuições *a priori* sugerido por Chopin *et al.* a fim de realizar uma comparação justa entre as estratégias propostas nos dois trabalhos. A implementação baseou-se em 50000 partículas e, portanto, demandou no máximo 5% dos recursos computacionais (considerando tanto a complexidade de tempo quanto a de memória) utilizados pelo dispositivo SMC².

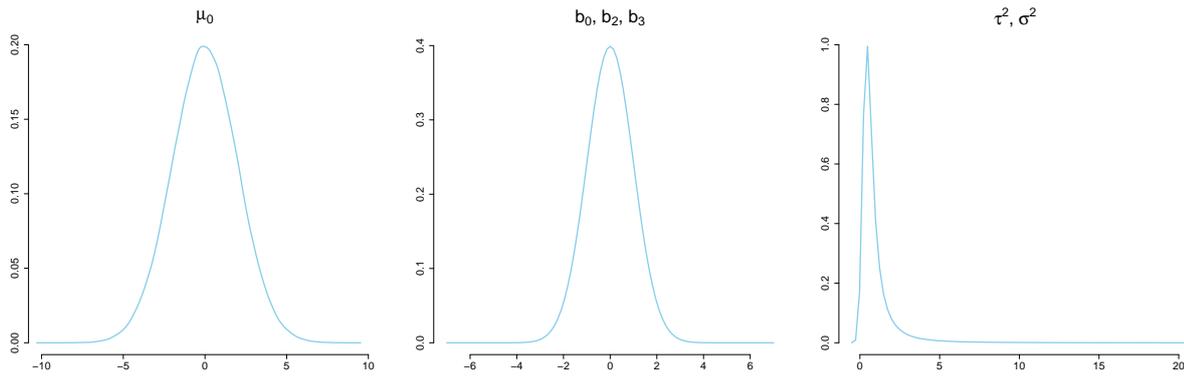


Figura 68 – Distribuições *a priori* correspondentes aos parâmetros μ_0 , b_0 , b_2 , b_3 , τ^2 e σ^2 na configuração artificial proposta por Chopin *et al.*

LW otimizado (Auxiliar) \times SMC²

Em um primeiro momento buscamos analisar o comportamento do filtro LW otimizado, implementado sob a perspectiva Auxiliar, com respeito ao problema da degeneração. Para tal acompanhamos a evolução do processo de filtragem a cada instante de iteração e comparamos com os parâmetros reais de simulação. O resumo a seguir, baseado em 25 execuções, mostra que o algoritmo é capaz de gerar estimativas muito estáveis para o problema, contrariando assim as expectativas de Chopin *et al.*

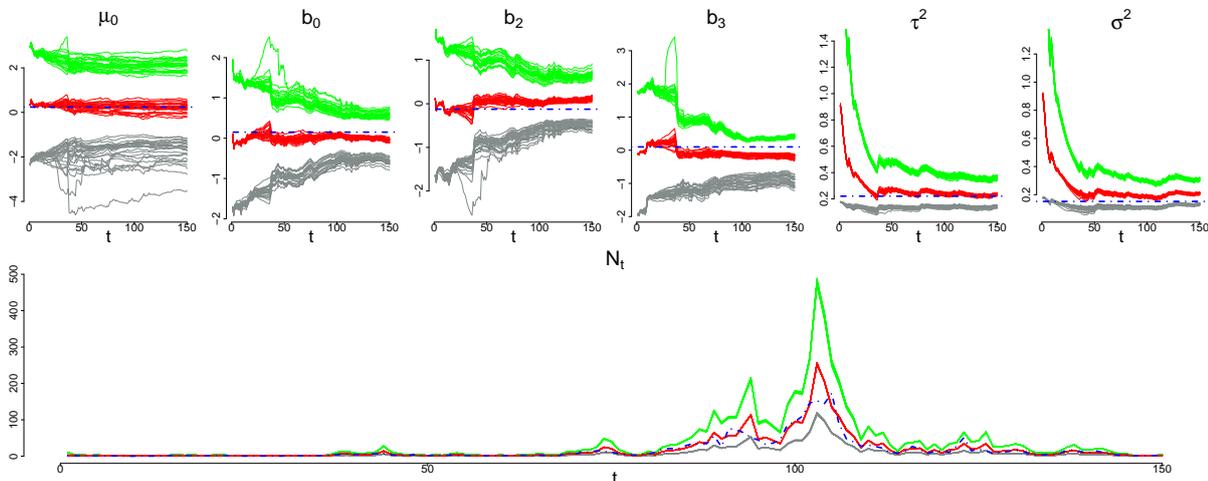


Figura 69 – Resumo de desempenho do método de Liu e West otimizado (perspectiva Auxiliar) com 50000 partículas: médias *a posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam as quantidades reais utilizadas no processo de simulação.

O segundo passo consiste em analisar as possíveis características de serrilhamento e multimodalidade associadas à incerteza final acerca dos parâmetros estáticos. Ilustramos a seguir as distribuições obtidas pelo filtro LW otimizado (azul) em comparação com as estimativas oferecidas pelo esquema SMC² (vermelho). Apesar da grande coerência entre as duas metodologias, nossa abordagem aparenta evidenciar com mais propriedade as características esperadas para tais distribuições. Isso certamente deve-se ao fato de que o método SMC² aproxima um modelo

probabilístico em 6 dimensões através de um enxame de apenas 1000 partículas. É possível construir aproximações mais ricas através de amostras de tamanho superior, contudo devemos ressaltar que cada nova partícula utilizada no tratamento do problema dos parâmetros estáticos, nesta implementação específica, será associada a ao menos 1000 outras partículas exploradas a favor da aproximação da função de verossimilhança. Desta forma, mesmo que o método seja implementado em paralelo, com o auxílio das unidades de processamento gráfico, teríamos ainda assim um custo de memória bastante elevado.

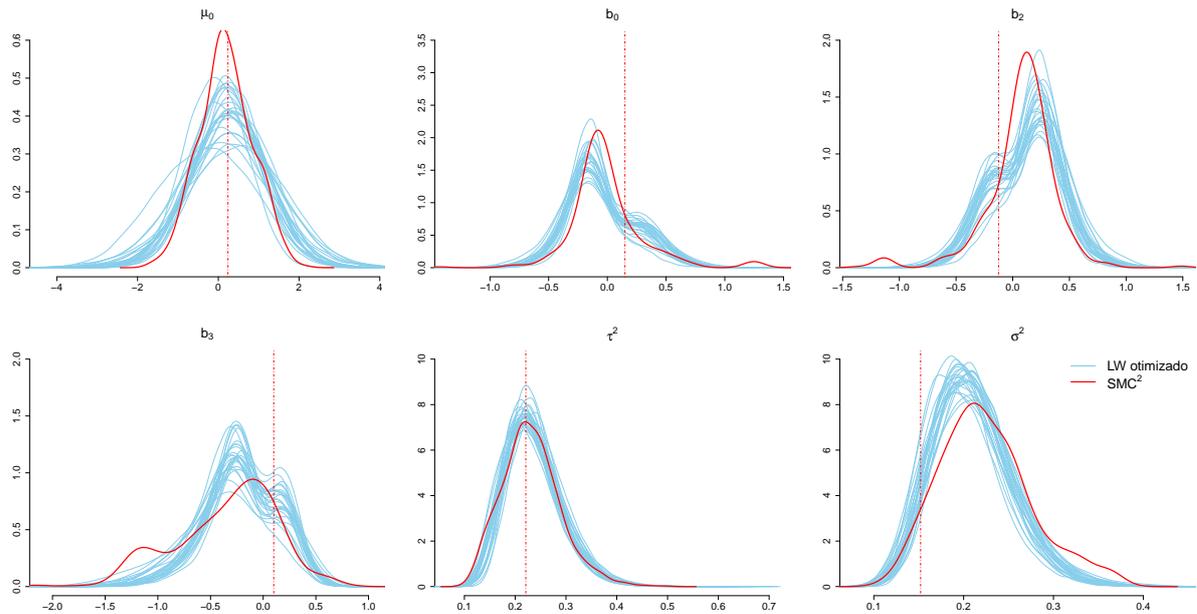


Figura 70 – Distribuições *a posteriori* obtidas pelos métodos LW otimizado (azul) e SMC² (vermelho).

As duas construções também concordam com respeito à filtragem da sequência latente N_t . Ambas foram capazes de inferir adequadamente a sequência mesmo considerando os picos populacionais observados em torno do instante $t = 100$.

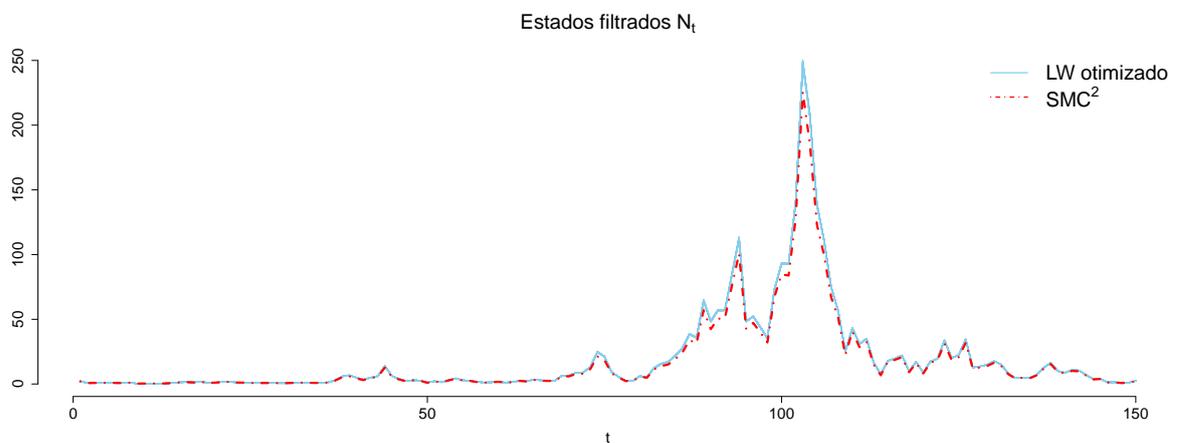


Figura 71 – Distribuições *a posteriori* obtidas pelos métodos LW otimizado (azul) e SMC² (vermelho).

Este experimento configura uma ótima oportunidade para evidenciar com mais propriedade os benefícios obtidos a partir da adoção de uma agenda dinâmica de rejuvenescimento. Em primeiro lugar, temos um ganho computacional importante: o método baseado na amostragem dinâmica apresentou uma redução de 37% no tempo de execução (resultado obtido novamente com o auxílio do pacote *microbenchmark*). A economia de recursos é naturalmente mais satisfatória do que a apresentada no problema da volatilidade estocástica já que agora estamos trabalhando com um espaço paramétrico de dimensão 6. Neste caso, a geração de parâmetros ocupa uma parte computacionalmente mais significativa do problema. O segundo ganho é observado na esfera da inferência. A próxima figura ilustra o desempenho do algoritmo baseado na amostragem perpétua de parâmetros ao longo das 150 iterações; é possível observar que o montante extra de ruído inserido ao longo dos passos de rejuvenescimento adicionais induzem uma maior suavização das distribuições *a posteriori*, mascarando os efeitos de serrilhamento e multimodalidade. A diferença mais significativa é observada com respeito aos parâmetros b_0 , b_2 e b_3 , os quais

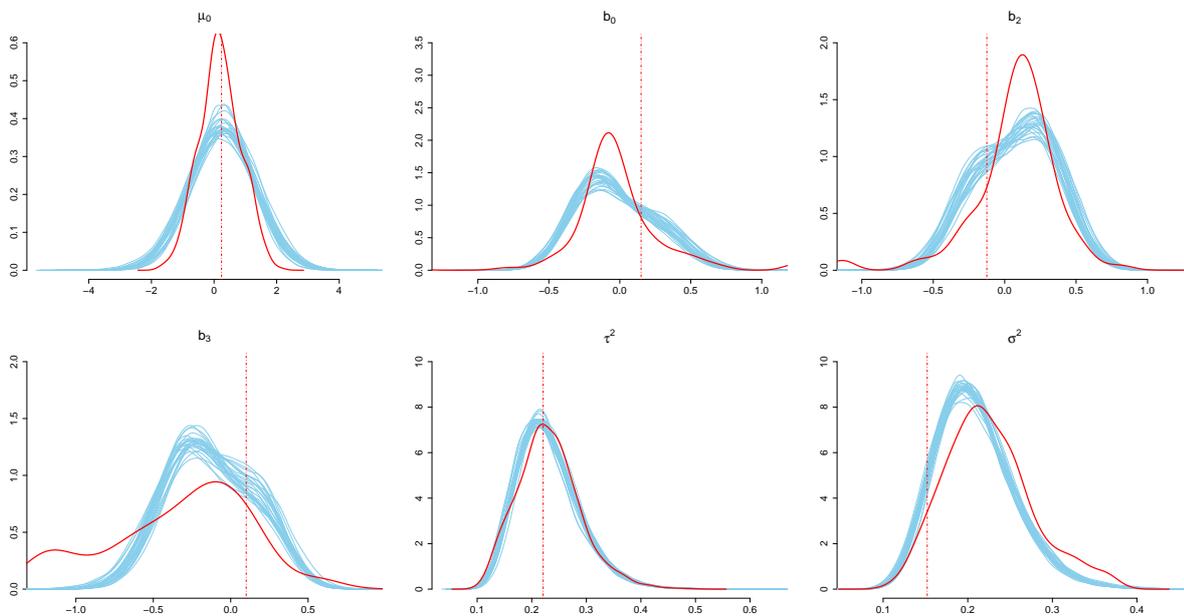


Figura 72 – Distribuições *a posteriori* obtidas pelos métodos LW otimizado Auxiliar (azul) e SMC² (vermelho).

caracterizam de maneira mais contundente os aspectos peculiares do problema. A fim de reforçar a maior coerência da perspectiva baseada em uma agenda dinâmica, apresentamos também os diagramas de dispersão, correspondentes a estes três parâmetros, obtidos a partir do método SMC² em comparação com os construídos pelo método LW com e sem a perspectiva dinâmica. Os resultados ilustram de maneira clara que a perpetuação do passo de rejuvenescimento de fato adiciona um montante indesejado de ruído à inferência final algoritmo.

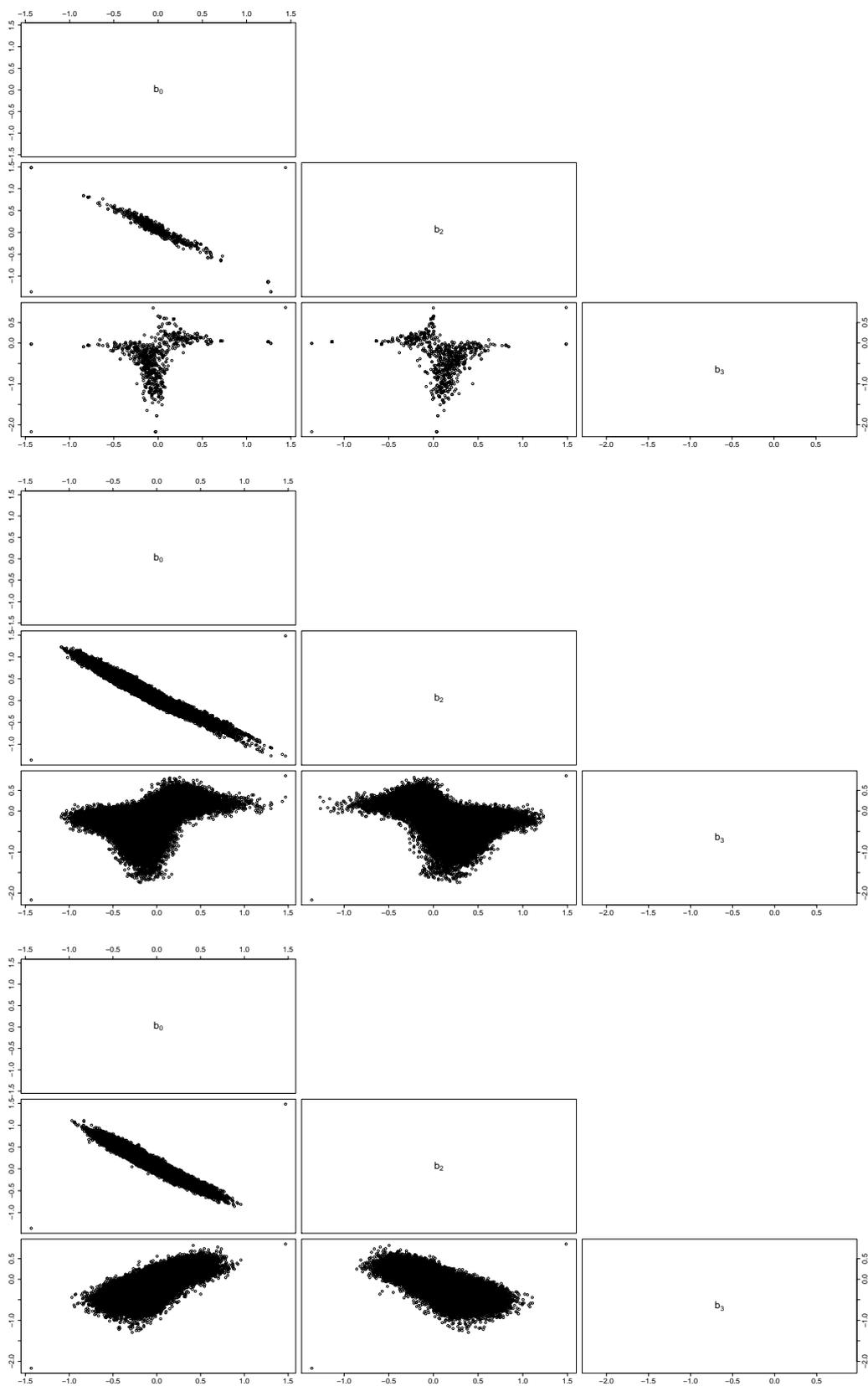


Figura 73 – Diagramas de dispersão referentes aos parâmetros b_0 , b_2 e b_3 obtidos a partir dos métodos SMC² (topo), filtro LW otimizado (meio) e filtro LW sem agenda dinâmica (base).

LW otimizado (*Bootstrap*) \times SMC²

Uma análise similar à apresentada no parágrafo anterior foi realizada considerando o paradigma *Bootstrap*. O objetivo neste segundo conjunto de testes é evidenciar que as estratégias de otimização geram resultados ainda mais satisfatórios quando aplicadas à perspectiva *Bootstrap* associada à distribuição ótima. Resumos análogos aos apresentados para a tratativa Auxiliar são apresentados a seguir para efeito de comparação.

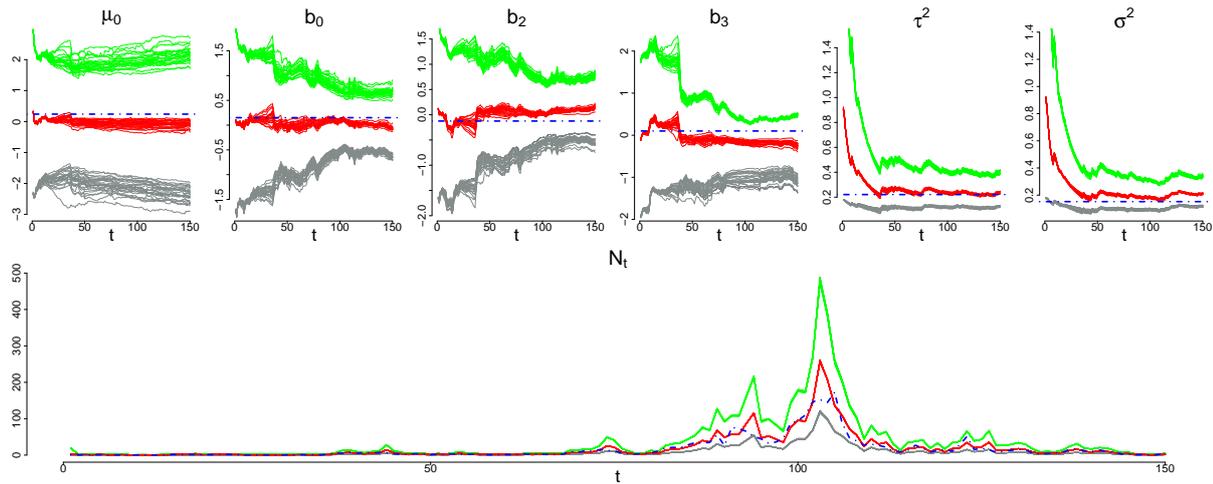


Figura 74 – Resumo de desempenho do método de Liu e West otimizado (perspectiva *Bootstrap*) com 50000 partículas: médias *a posteriori* e bandas de 95% de credibilidade. Trajetórias em azul representam as quantidades reais utilizadas no processo de simulação.

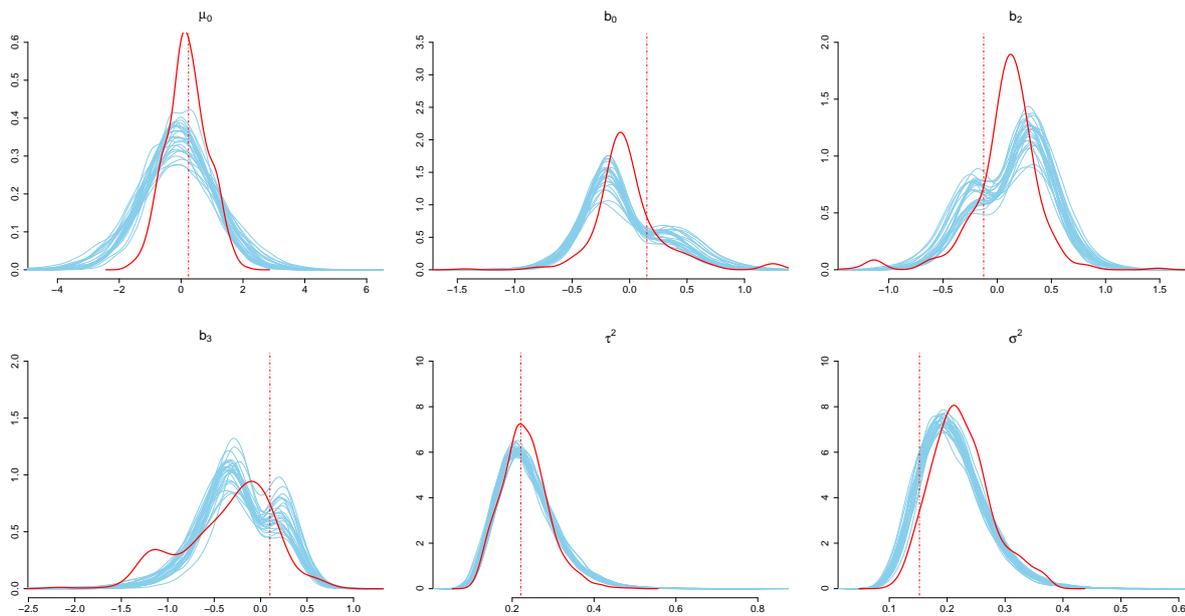


Figura 75 – Distribuições *a posteriori* obtidas pelos métodos LW otimizado (azul) e SMC² (vermelho).

As estimativas provenientes da abordagem *Bootstrap* aparentam identificar a característica multimodal do problema de forma um pouco mais precisa. Uma melhor noção a respeito

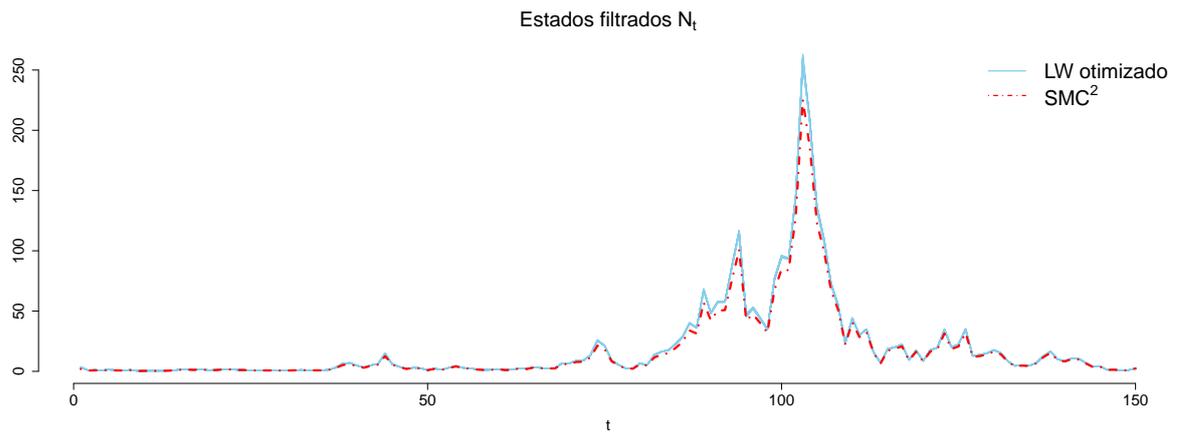


Figura 76 – Distribuições *a posteriori* obtidas pelos métodos LW otimizado (azul) e SMC^2 (vermelho).

da discrepância observada no desempenho das duas perspectivas pode ser obtida a partir da comparação entre as estimativas geradas por cada uma delas e os valores reais dos parâmetros. O próximo resultado confirma que o método *Bootstrap* de fato gerou melhores estimativas neste contexto.

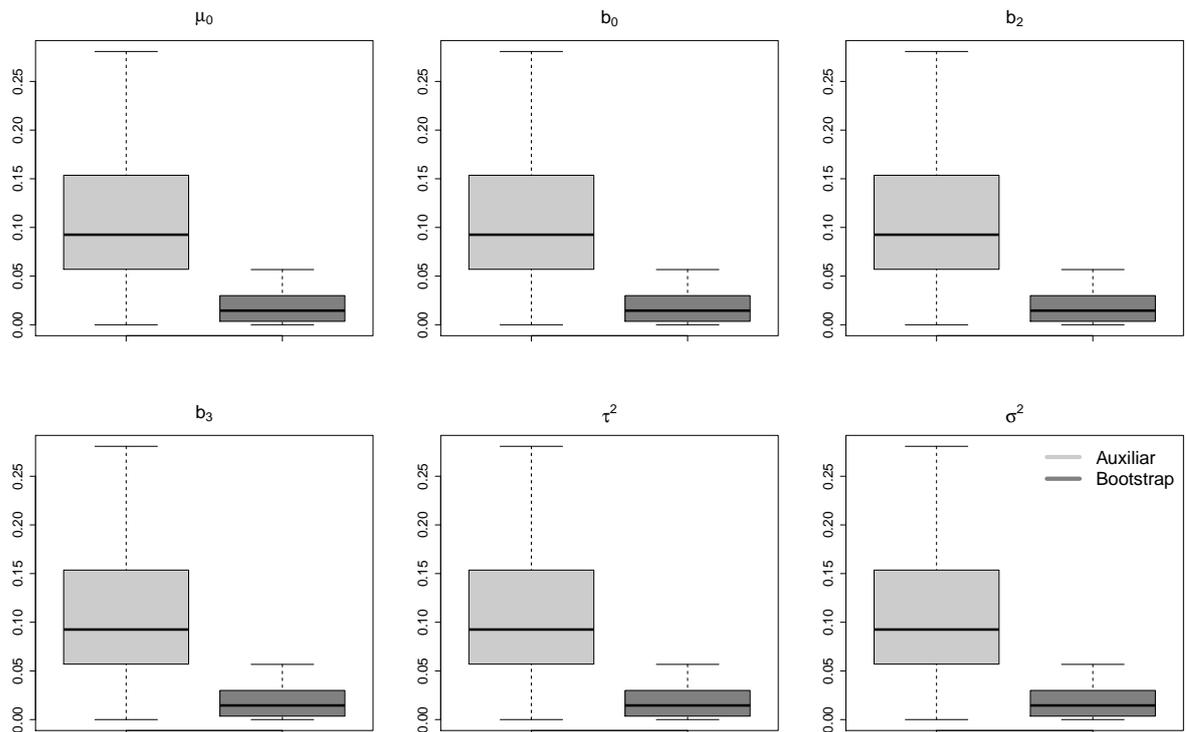


Figura 77 – Comparação entre as estimativas construídas para cada parâmetro e os valores reais utilizados para simulação. Os diagramas ilustram a distribuição dos erros $\hat{\theta}_i - \theta$ considerando as perspectivas Auxiliar e *Bootstrap*.

Os ganhos oferecidos pela adoção de uma agenda dinâmica de rejuvenescimento também podem ser evidenciados claramente na perspectiva *Bootstrap*. Com respeito ao ganho de eficiência computacional, temos que o resultado foi ainda mais interessante: obtemos uma economia de 47% no tempo médio de execução do método. Este resultado já era esperado uma vez que a perspectiva *Bootstrap* economiza também nos passos de reamostragem, diferentemente do que ocorre na abordagem Auxiliar. Tratando agora da questão inferencial, observamos ganhos similares aos já verificados na análise anterior. O montante adicional de ruído adicionado pela execução perpétua do dispositivo de rejuvenescimento atropela as peculiaridades associadas a este conjunto de distribuições *a posteriori* gerando estimativas mais suavizadas que o desejado.

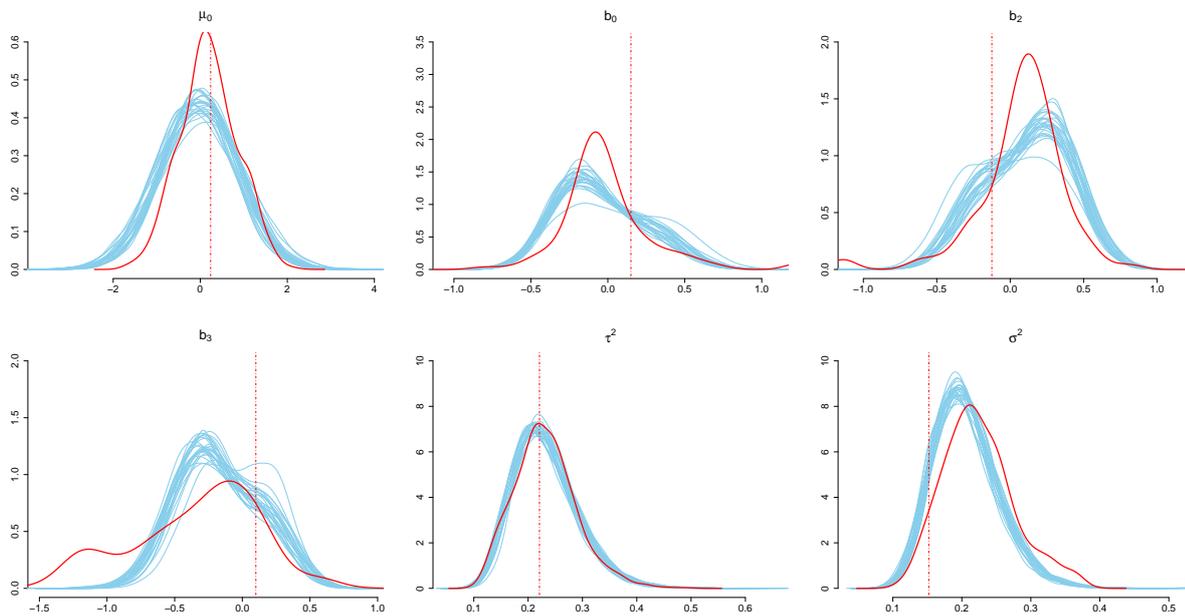


Figura 78 – Distribuições *a posteriori* obtidas pelos métodos LW otimizado Bootstrap (azul) e SMC^2 (vermelho).

A discrepância entre as duas tratativas poderia ser reforçada a partir da comparação entre os diagramas de dispersão obtidos por ambas, todavia os resultados foram omitidos para evitar repetições desnecessárias.

Dados reais

Consideramos nesta segunda etapa de testes observações reais referentes à evolução populacional de uma espécie particular de gaviões (*Acciption nisus*) no sul da Escócia. Os dados estão registrados sob o número 6575 no banco de dados global de populações (NERC Centre for Population Biology, Imperial College, 1999) e são de acesso livre. Peters *et al.* adotaram

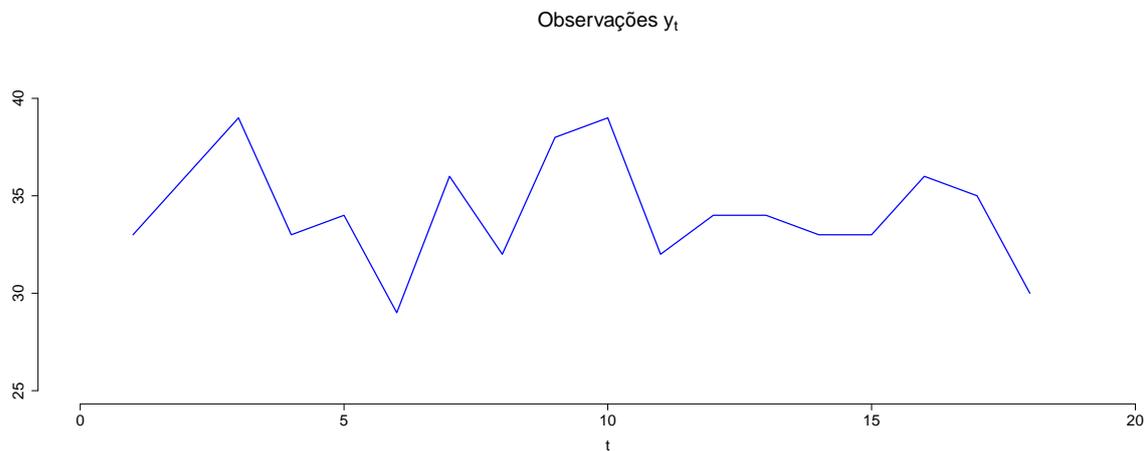


Figura 79 – Observações, em escala real, referentes à espécie *Acciption nisus* registrado sob o número 6575 no banco de dados global de populações.

este conjunto de observações devido à forte característica de serrilhamento e multimodalidade já evidenciada em trabalhos anteriores. A ideia original dos autores era explorar algoritmos do tipo PMCMC a fim de construir uma alternativa válida e eficiente para o tratamento de dados associados a distribuições mais irregulares como as associadas a este contexto. Por se tratar de um conjunto com $n = 18$ observações, foram atribuídas as seguintes distribuições *a priori*

$$\tau^2 \sim GI\left(\frac{18}{2}, \frac{18-2}{10}\right) \quad \text{e} \quad \sigma^2 \sim GI\left(\frac{18}{2}, \frac{18-2}{10}\right)$$

para os parâmetros de volatilidade, caracterizando uma proposta mais informativa.

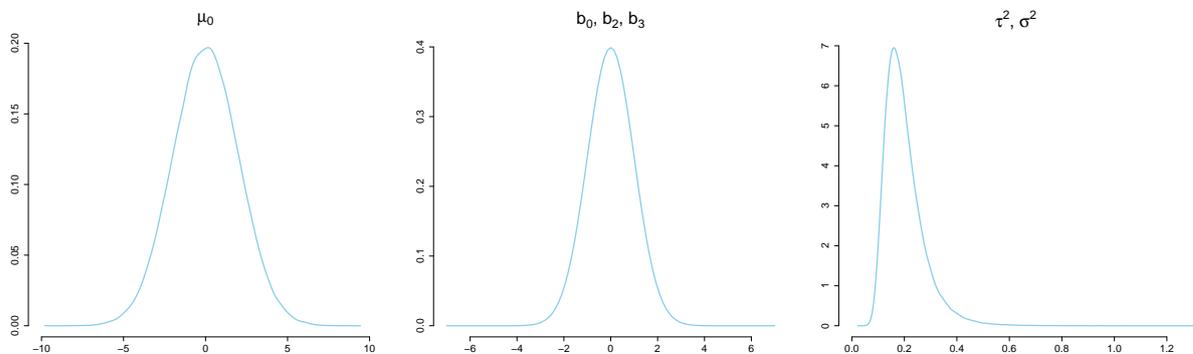


Figura 80 – Distribuições *a priori* correspondentes aos parâmetros μ_0 , b_0 , b_2 , b_3 , τ^2 e σ^2 .

Uma sequência de testes foi conduzida na tentativa de comparar a capacidade inferencial do filtro LW otimizado com respeito às metodologias PMCMC e SMC². A avaliação resumiu-se

neste caso às distribuições *a posteriori* dos parâmetros estáticos. O algoritmo PMCMC foi implementado seguindo um roteiro análogo ao sugerido por Peters *et al.*: 500 partículas no filtro de partículas utilizado para aproximar a função de verossimilhança, *burn-in* de 150000 e a construção da inferência deu-se com respeito às 50000 últimas partículas geradas pela cadeia. Já o método SMC² foi implementado de forma ligeiramente distinta em comparação com a sugestão de seus criadores. Neste último caso adotamos, com respeito aos parâmetros de dispersão, as distribuições *a priori* propostas no experimento original de Peters *et al.* em detrimento à escolha não informativa $GI(2, 1)$. As demais informações *a priori* seguiram a escolha já discutida no estudo anterior e os tamanhos populacionais foram: $N_x = 1000$ e $N_\theta = 5000$. Além disso, cada passo de rejuvenescimento fez uso de 5 iterações do dispositivo de rejeição Metropolis-Hastings. Por fim, os filtros LW otimizados seguiram rigorosamente a configuração proposta por Peters *et al.* e suas implementações foram baseadas em nuvens de 50000 partículas.

PMCMC

A primeira análise conduzida neste segundo conjunto de testes refere-se à dupla LW otimizado e PMCMC. Ambas as abordagens foram capazes de reproduzir as características de serrilhamento e multimodalidade das distribuições *a posteriori*. Mais do que isso, houve uma grande coerência entre as aproximações obtidas pelas duas abordagens, conforme ilustrado a seguir. Nota-se que o filtro de partículas não destacou as duas modas referentes ao parâmetro μ_0 com a mesma nitidez alcançada pelo dispositivo PMCMC, entretanto a aproximação apresentada é bastante razoável considerando que estamos substituindo um amostrador ótimo por uma aproximação por importância. Estas relações são reforçadas pelos diagramas de dispersão na Figura 82.

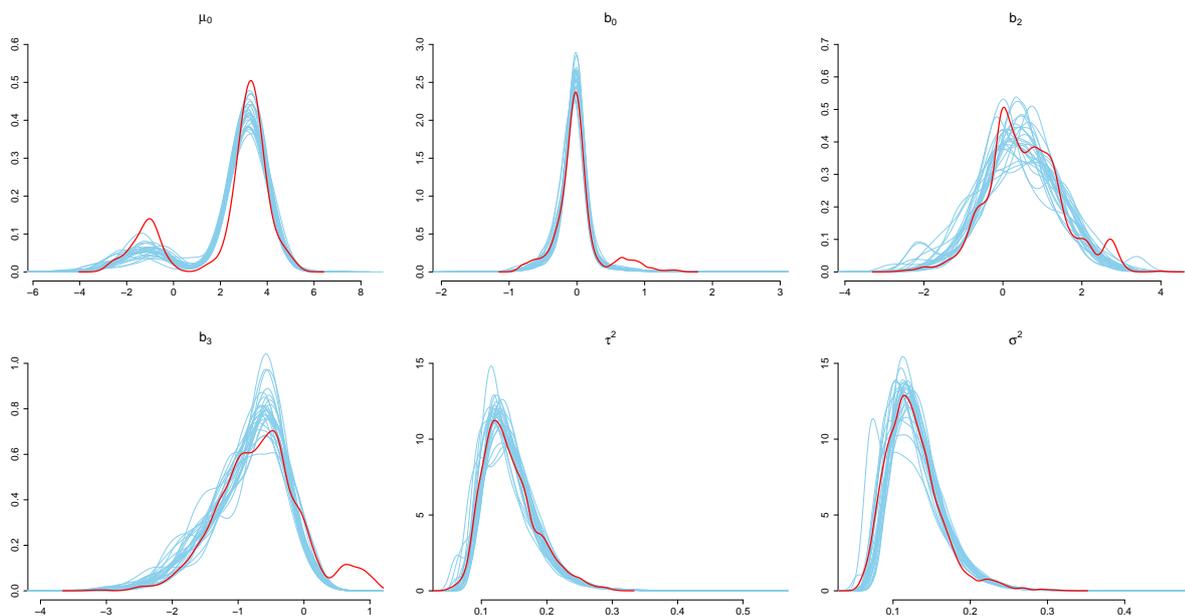


Figura 81 – Distribuições *a posteriori* obtidas pelos métodos LW otimizado Auxiliar (azul) e PMCMC (vermelho).

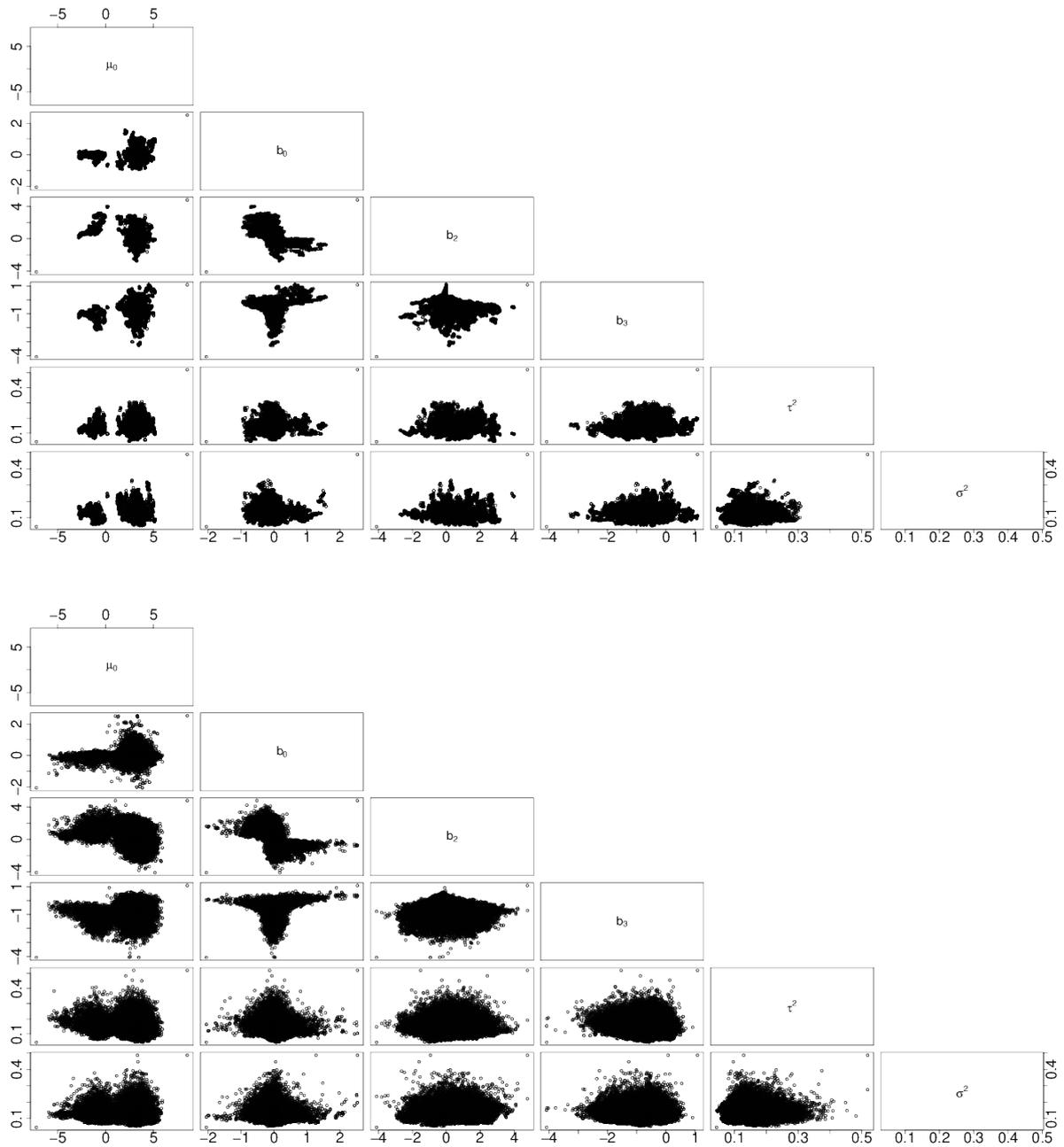


Figura 82 – Diagramas de dispersão amostral obtidos pelos métodos PMCMC (topo) e LW otimizado Auxiliar (base).

Podemos observar também um ganho interessante relativo à calibragem automática dos parâmetros a e h^2 associados à metodologia de Liu e West. Mesmo aumentando consideravelmente o tamanho da nuvem de partículas, não observa-se excesso de suavização nas estimativas *a posteriori* oferecidas pelo algoritmo. Tanto a bimodalidade observada em μ_0 quanto o comportamento serrilhado destacado em b_2 e b_3 continuam perceptíveis já que a janela h^2 (responsável pela suavização) é reduzida naturalmente à medida que a população cresce.

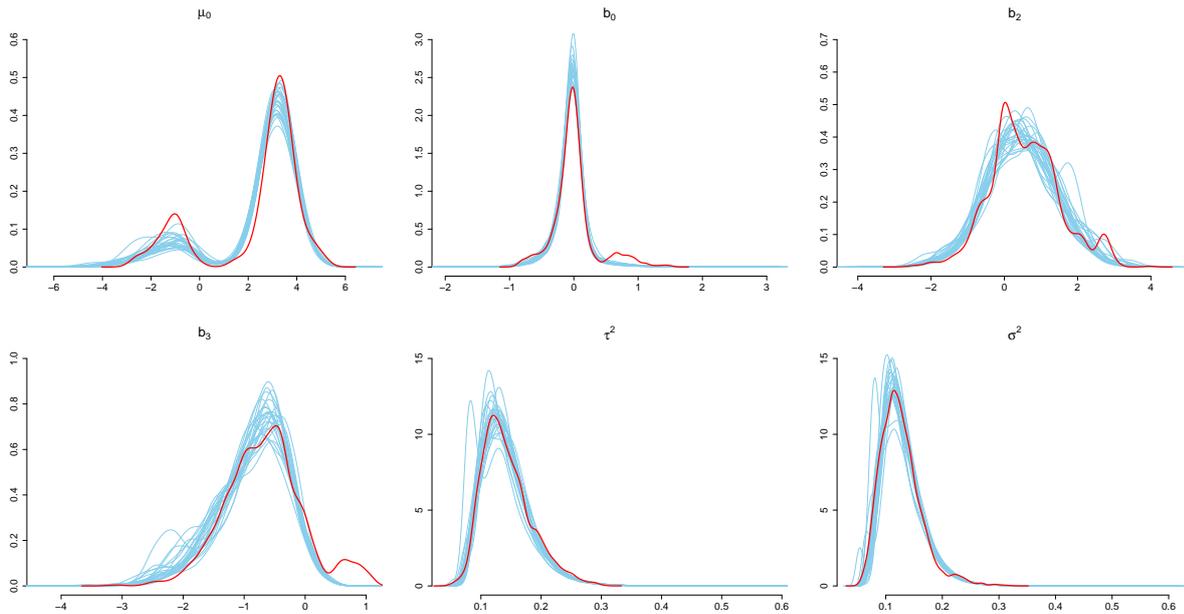


Figura 83 – Desempenho do método LW otimizado Auxiliar considerando 150000 partículas.

Assim como no cenário simulado, a agenda dinâmica demonstrou ser um passo fundamental na aproximação das distribuições. Os diagramas a seguir evidenciam que a utilização perpétua do mecanismo de rejuvenescimento inseriu novamente um montante extra de ruído nas estimativas, culminando com um nível de suavização indesejado. Neste caso, os aspectos peculiares das distribuições *a posteriori* passam a ser mascarados.

Destacamos graficamente apenas o desempenho da abordagem Auxiliar para reforçar que o dispositivo pode, em geral, manter a diversidade da amostra mesmo realizando o passo de reamostragem a cada iteração. O desenvolvimento da agenda dinâmica neste cenário só foi possível a partir da introdução do conceito de fator de fertilidade, o qual permitiu acessar a diversidade da população de maneira simples e eficiente. Esta medida permite que a população seja renovada apenas em situações indispensáveis, minimizando a introdução de interferências no processo inferencial; um ponto bastante relevante já que a aproximação de distribuições *a posteriori* por meio de misturas, como o próprio nome já sugere, trata apenas de uma aproximação para o modelo probabilístico real e portanto acrescentará necessariamente algum erro às estimativas. Apesar de observamos resultados idênticos na perspectiva *Bootstrap* associada a uma agenda dinâmica de rejuvenescimento, tais resultados não são surpreendentes uma vez que o método reduz significativamente o uso dos passos de reamostragem.

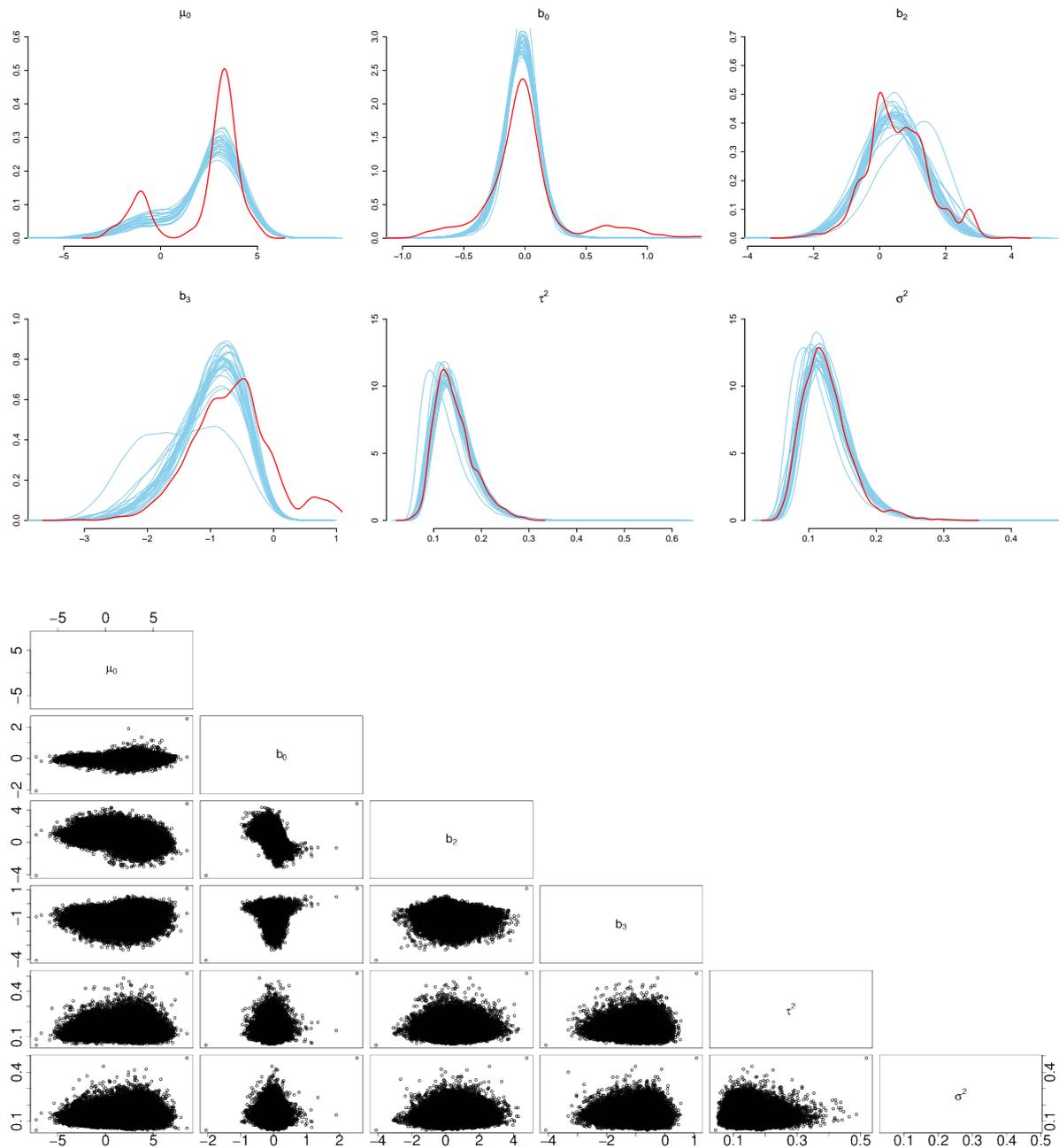


Figura 84 – Filtro LW sem a utilização de agenda dinâmica de rejuvenescimento. O diagrama no topo apresenta as densidades *a posteriori* estimadas pelo método em comparação com a estimativa construída via PMCMC. Abaixo temos os diagramas de dispersão obtidos após a filtragem da última observação.

SMC²

O método SMC² constitui, sem sombra de dúvidas, uma das estratégias mais consistentes para a realização de inferência em modelos Markovianos ocultos devido a seus resultados de convergência formalmente demonstrados (CHOPIN; JACOB; PAPASPILIOPOULOS, 2013). Todavia, alguns potenciais problemas de degeneração ainda passam despercebidos na análise puramente teórica do método. Em particular, os autores do algoritmo não justificam a utilização de distribuições *a priori* não informativas com respeito aos parâmetros σ^2 e τ^2 na aplicação envolvendo os dados da espécie *Accipter nisus*. Na tentativa de buscar um entendimento mais amplo do método, consideramos então as distribuições mais informativas propostas por Peters *et al.* e avaliamos o desempenho do algoritmo. Reforçamos que estas foram as distribuições adotadas na implementação do método LW otimizado, o qual obteve estimativas consistentes com as oferecidas pelo algoritmo PMCMC. O resultado foi, curiosamente, desfavorável ao algoritmo SMC² conforme ilustrado abaixo.

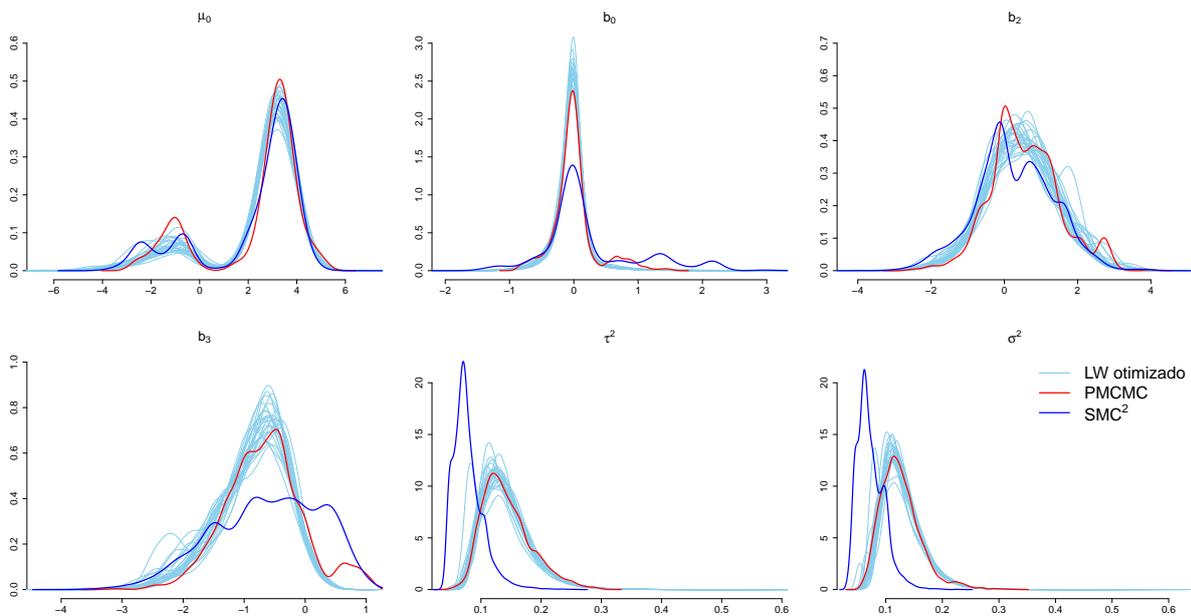


Figura 85 – Desempenho comparativo dos métodos LW otimizado, PMCMC e SMC².

O desempenho se repetiu considerando até mesmo uma implementação baseada em $N_x = 5000$ pontos na aproximação da verossimilhança, $N_\theta = 20000$ na construção das distribuições *a posteriori* e 10 passos Metropolis-Hatings após cada reamostragem. Os parâmetros não foram elevados ainda mais devido ao custo computacional do método. Esta última configuração, por exemplo, faz uso de um mínimo de 10^8 variáveis de ponto flutuante. Desta forma, acreditamos que o problema do método não está atrelado a questões assintóticas. Esta discussão favorece a abertura de uma análise mais precisa deste promissor algoritmo. Possivelmente, resultados mais robustos podem ser obtidos se associarmos o dispositivo SMC² a algumas das propostas de otimização sugeridas neste trabalho.

CONSIDERAÇÕES FINAIS

7.1 Objetivos Alcançados

Este trabalho teve como objetivo rever os principais algoritmos de filtragem considerando problemas com parâmetros estáticos desconhecidos, identificar problemas de implementação ainda despercebidos na literatura e propor melhorias simples e eficientes. De uma forma sucinta, podemos destacar:

- apresentamos um estudo alternativo que dissociou a metodologia de filtragem pura da sistemática de amostragem de parâmetros estáticos, permitindo assim uma análise mais precisa a respeito dos vícios e virtudes dos principais métodos abordados na literatura;
- revisitamos os principais trabalhos da área e demonstramos que os resultados negativos apresentados pelo método de Liu e West eram basicamente fruto de implementações equivocadas e/ou subótimas;
- verificamos experimentalmente que uma escolha de parâmetros objetiva (e bastante tradicional no contexto da aproximação de curvas por meio de núcleos estimadores) é capaz de levar o método de Liu e West a resultados muito consistentes, eliminando assim a ideia de que o algoritmo é desfavorável por necessitar de uma calibragem pouco precisa de seus parâmetros;
- desenvolvemos propostas otimizadas de implementação as quais foram capazes de atacar de forma eficiente o problema de degeneração amostral a partir da utilização da informação oferecida pelos tamanhos efetivos amostrais e pelos fatores de fertilidade;
- apresentamos um extenso estudo de simulações, assim como estudos com base em dados reais, a fim de justificar experimentalmente os resultados discutidos no texto;

7.2 Conclusões

Os resultados apresentados neste trabalho indicaram primeiramente que a escolha do método de reamostragem exerce grande influência nos algoritmos de filtragem envolvendo parâmetros estáticos desconhecidos. É bastante comum encontrar na literatura expressões da forma “poderíamos utilizar outros métodos mas, por simplicidade, utilizaremos a reamostragem multinomial”, o que deixa claro como a seleção adequada do reamostrador é deixada em segundo plano. Mesmo as estratégias baseadas na propagação de estatísticas suficientes, muitas vezes julgadas isentas do fenômeno de degeneração, mostraram grande sensibilidade com respeito ao método de reamostragem. A partir dos experimentos e discussões propostas aqui podemos concluir que os amostradores de variância mínima – amostragem sistemática ou por ramificação – nunca devem ser substituídos pela tradicional reamostragem multinomial em problemas envolvendo parâmetros estáticos.

Outro ponto a ser destacado neste estudo diz respeito ao ganho de eficiência e acurácia alcançado a partir das propostas tratadas nesta obra. Ao reforçar a interpretação proposta por Kitagawa para modelos com parâmetros desconhecidos, entendemos que o rejuvenescimento da população a cada instante era desnecessário. Concluímos então, a partir dos experimentos propostos, que o rejuvenescimento baseado em tempos de parada – obtidos a partir do tamanho efetivo amostral em variações do filtro *Bootstrap* ou dos fatores de fertilidade em versões do filtro auxiliar – é uma alternativa capaz de simultaneamente contribuir com a eficiência computacional e com a precisão dos métodos. É interessante notar que mesmo utilizando o passo de reamostragem a cada instante do filtro Auxiliar, ainda foi possível manter a diversidade da população e garantir assim a construção de uma inferência satisfatória. Tal resultado não seria obtido sem a introdução do conceito de fertilidade.

Tratando mais especificamente do método de Liu e West, ficou evidente que os problemas de degeneração associados ao algoritmo eram basicamente fruto de implementações equivocadas e/ou subótimas. O simples fato de considerarmos o dispositivo de reamostragem por ramificação em detrimento do multinomial já foi capaz de trazer melhoras de desempenho importantes para o algoritmo. Destacamos também que as disparidades de desempenho deste procedimento com respeito às variações da metodologia apoiada na propagação de estatística suficientes estavam associadas às discrepâncias nos pesos de importância utilizados em cada implementação. Na prática observamos que ambas as tratativas chegam a um resultado similar caso consideremos o mesmo esquema de desenvolvimento para os pesos de importância. Além disso, verificamos que a calibragem dos parâmetros do algoritmo não deve preservar apenas a ideia original do fator de desconto ignorando as sugestões provenientes do contexto da inferência não-paramétrica. Segundo nossa abordagem é possível privilegiar concomitantemente os dois aspectos a fim de obter um algoritmo mais robusto. Os resultados foram validados vários experimentos envolvendo tanto dados simulados quanto observações reais. Destacamos ainda que a estratégia de reamostragem suavizada demonstrou ser uma ferramenta importante em situações caracterizadas

por distribuições *a priori* não informativas. A utilização deste dispositivo evita que a população seja enxugada muito rapidamente devido a um espalhamento inadequado da população nos instantes iniciais permitindo assim a utilização de informações iniciais mais vagas e contrariando a hipótese de que o filtro LW é desvantajoso por ser refém da escolha de distribuições iniciais mais precisas. Por fim, mostramos também que o método de Liu e West otimizado é de fato uma alternativa a ser considerada em problemas complexos tratados até então a partir de algoritmos pouco eficientes como PMCMC ou SMC². Reforçamos que a ideia não é tentar substituir as demais abordagens pelo método detalhado aqui, mas sim acrescentar à literatura mais uma ferramenta válida, eficiente e de implementação simples.

7.3 Trabalhos Futuros

A sequência deste estudo passa pelo desenvolvimento de resultados formais de convergência para filtros de partículas implementados na presença de parâmetros desconhecidos. Seria interessante tentar entender o quanto a presença de parâmetros desconhecidos pode desacelerar a taxa de convergência dos filtros. Apesar de não configurar necessariamente uma certeza, é razoável imaginar que de fato há uma desaceleração na convergência uma vez que os parâmetros desconhecidos trazem um grau de complexidade maior para o problema. Além de nos deixar mais confortáveis com respeito aos cenários aos quais estas metodologias podem ser executadas, este tipo de estudo poderia auxiliar na determinação de tamanhos mais adequados para as nuvens de partículas envolvidas.

As discussões traçadas aqui também deixam espaço para o desenvolvimento de uma estratégia objetiva para a escolha da estrutura de dependência utilizada pela mistura de distribuições gaussianas do algoritmo de Liu e West. Na prática temos concretamente apenas as duas sugestões padrão: matriz diagonal de variâncias ou matriz completa de covariâncias. Estratégias como a sugerida na aplicação do modelo de volatilidade estocástica são obtidas a partir de uma análise qualitativa do modelo, o que inviabiliza uma generalização. O estabelecimento de uma metodologia quantitativa eficiente certamente conferiria mais robustez ao algoritmo.

Além das contribuições evidenciadas neste trabalho, a especificação de uma implementação mais consistente para o método de Liu e West nos permite ainda imaginar aplicações desta abordagem em outros cenários. Devemos lembrar que a ideia original de aproximação de distribuições *a posteriori* por meio de misturas apresentada por West (1993) visava aplicações associadas à inferência acerca de observações iid e não sequenciais. Poderíamos, por exemplo, tentar associar esta aproximação ao algoritmo IBIS (CHOPIN, 2002): situações onde o passo Metropolis-Hastings apresenta baixas taxas de aceitação poderiam ser substituídas pela aproximação utilizada por Liu e West. Consequentemente, o mesmo raciocínio poderia ser automaticamente estendido para o algoritmo SMC², o qual sofre ainda mais com o problema de baixas taxas de aceitação. Seria interessante também avaliar o impacto de outras estratégias

propostas aqui com respeito à implementação do método SMC². No experimento da dinâmica populacional, por exemplo, verificamos tamanhos efetivos amostrais muito baixos após a filtragem da primeira observação. Este fato pode ter contribuído para a degeneração da informação a respeito das variâncias, o que consequentemente afetou as estimativas finais. É possível que a inserção do procedimento de reamostragem suavizada com respeito à filtragem dos parâmetros estáticos ofereça alguma contribuição com respeito a problemas desta natureza.

Extrapolando mais os tópicos abordados neste trabalho, poderíamos tentar adaptar o conceito de estimação de parâmetros via filtros de partículas para problemas de filtragem em tempo contínuo. Tais aplicações envolvem o tratamento de equações diferenciais estocásticas (ordinárias e parciais) e possuem um grau de complexidade ainda maior. Algumas abordagens interessantes de filtragem pura já podem ser encontradas na literatura; uma discussão mais detalhada do tópico pode ser obtida em [Bain e Crisan \(2009\)](#). Entretanto, a estimativa paramétrica neste contexto ainda é muito dependente de métodos MCMC, desta forma o desenvolvimento de alternativas eficientes seria uma contribuição bastante relevante.

REFERÊNCIAS

- AKASHI, H.; KUMAMOTO, H. Random sampling approach to state estimation in switching environments. v. 13, n. 4, p. 429–434, 1977. Citado na página 27.
- ANDERSON, B.; MOORE, J. **Optimal Filtering**. Englewood Cliffs, NJ: Prentice Hall, 1979. Citado na página 26.
- ANDRIEU, C.; DOUCET, A.; HOLENSTEIN, R. Particle markov chain monte carlo methods. **Journal of the Royal Statistical Society: Series B**, v. 72, n. 3, 2010. Citado 3 vezes nas páginas 29, 70 e 82.
- ARULAMPALAM, S.; MASKELL, S.; GORDON, N.; CLAPP, T. A tutorial on particle filters for online nonlinear/nongaussian bayesian tracking. **IEEE Transactions on Signal Processing**, v. 50, n. 2, p. 174–188, 2002. Citado 2 vezes nas páginas 64 e 90.
- ARULAMPALAM, S.; RISTIC, B.; GORDON, N.; MANSELL, T. Bearings-only tracking of manoeuvring targets using particle filters. **EURASIP Journal on Applied Signal Processing**, v. 15, p. 2351–2365, 2004. Citado na página 25.
- BAIN, A.; CRISAN, D. **Fundamentals of Stochastic Filtering**. New York: Springer-Verlag, 2009. Citado 4 vezes nas páginas 59, 60, 174 e 185.
- BAKER, J. E. Reducing bias and inefficiency in the selection algorithm. In: INTERNATIONAL CONFERENCE ON GENETIC ALGORITHMS AND THEIR APPLICATION, 2. [S.l.], 1987. p. 14–21. Citado 2 vezes nas páginas 58 e 86.
- BERGMAN, N. **Recursive Bayesian Estimation Navigation and Tracking Applications**. Tese (Doutorado) — Linkoping University, 1999. Citado na página 26.
- CAPPÉ, O.; MOULINES, E.; RYDEN, T. **Inference in Hidden Markov Models**. New York: Springer-Verlag, 2005. Citado 3 vezes nas páginas 38, 41 e 42.
- CARPENTER, J.; CLIFFORD, P.; FEARNHEAD, P. An improved particle filter for non-linear problems. In: IEEE INTERNATIONAL CONFERENCE ON RADAR, SONAR AND NAVIGATION, 146. [S.l.], 1999. p. 2–7. Citado 2 vezes nas páginas 27 e 192.
- CARVALHO, C.; JOHANNES, M.; LOPES, H.; POLSON, N. Particle learning and smoothing. **Statistical Science**, v. 25, n. 1, p. 88–106, 2010. Citado 6 vezes nas páginas 29, 30, 70, 77, 84 e 189.
- CHIB, S.; NARDARI, F.; SHEPHARD, N. Markov chain monte carlo methods for stochastic volatility models. **Journal of Econometrics**, v. 108, p. 281–316, 2002. Citado na página 133.
- CHOPIN, N. A sequential particle filter method for static models. **Biometrika**, v. 89, n. 3, p. 539–551, 2002. Citado 4 vezes nas páginas 30, 78, 90 e 173.

- CHOPIN, N.; JACOB, P.; PAPASPILIOPOULOS, O. SMC² : an efficient algorithm for sequential analysis of state space models. **Journal of the Royal Statistical Society: Series B**, v. 75, n. 3, p. 397–426, 2013. Citado 7 vezes nas páginas [29](#), [70](#), [89](#), [90](#), [155](#), [157](#) e [170](#).
- CRISAN, D.; Del Moral, P.; LYONS, T. Discrete filtering using branching and interacting particle systems. **Markov Processes Related Fields**, v. 5, n. 3, p. 293–319, 1999. Citado 2 vezes nas páginas [28](#) e [185](#).
- CRISAN, D.; MIGUEZ, J. Particle-kernel estimation of the filter density in state-space models. **Bernoulli**, v. 20, n. 4, p. 1879–1929, 2014. Citado na página [28](#).
- Del Moral, P.; DOUCET, A.; JASRA, A. Sequential monte carlo samplers. **Journal of the Royal Statistical Society: Series B**, v. 68, n. 3, p. 411–436, 2006. Citado na página [30](#).
- DOUC, R.; CAPPÉ, O. Comparison of resampling schemes for particle filtering. In: **IEEE. Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis (ISPA), 2005**. [S.l.], 2005. p. 64–69. Citado na página [59](#).
- DOUCET, A.; GODSILL, S. J.; ANDRIEU, C. On sequential monte carlo sampling methods for bayesian filtering. **Statistics and Computing**, v. 10, p. 197–208, 2000. Citado na página [28](#).
- DOUCET, A.; KANTAS, N.; SINGH, S.; MACIEJOWSKI, J. An overview of sequential monte carlo methods for parameter estimation in general state-space models. In: **Proceedings of the International Federation of Automatic Control Meeting on System Identification**. [S.l.: s.n.], 2009. Citado na página [89](#).
- DUKIC, V.; LOPES, H.; POLSON, N. Tracking epidemics with google flu trends data and a state-space seir model. **Journal of the American Statistical Association**, v. 107, n. 500, p. 1410–1426, 2012. Citado na página [25](#).
- FEARNHEAD, P. Markov chain monte carlo, sufficient statistics, and particle filters. **Journal of Computational and Graphical Statistics**, v. 11, n. 4, p. 848–862, 2002. Citado 2 vezes nas páginas [29](#) e [84](#).
- FEARNHEAD, P.; CLIFFORD, P. On-line inference for hidden markov models via particle filters. **Journal of the Royal Statistical Society: Series B**, v. 65, n. 4, p. 887–899, 2003. Citado na página [28](#).
- FEARNHEAD, P.; TAYLOR, B. An adaptive sequential monte carlo sampler. **Bayesian Analysis**, v. 8, n. 2, p. 411–438, 2013. Citado na página [30](#).
- FLURY, T.; SHEPHARD, N. **Learning and filtering via simulation: smoothly jittered particle filters**. [S.l.]: Department of Economics, University of Oxford, 2009. Citado na página [105](#).
- FU, X.; JIA, Y. An improvement on resampling algorithm of particle filters. **IEEE Transactions on Signal Processing**, v. 58, n. 10, 2010. Citado na página [28](#).
- GALES, M.; YOUNG, S. The application of hidden markov models in speech recognition. **Foundations and Trends in Signal Processing**, v. 1, n. 3, p. 195–304, 2008. Citado na página [25](#).

- GAMERMAN, D.; LOPES, H. **Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference**. 2nd. ed. [S.l.]: Taylor & Francis, 2006. (Chapman & Hall/CRC Texts in Statistical Science). ISBN 9781584885870. Citado 2 vezes nas páginas 34 e 52.
- GENON-CATALOT, V.; JEANTHEAU, T.; LAREDO, C. Stochastic volatility as hidden markov models and statistical applications. **Bernoulli**, v. 6, n. 6, p. 1051–1079, 2000. Citado na página 25.
- GILKS, W.; RICHARDSON, S.; SPIEGELHALTER, D. **Markov Chain Monte Carlo in Practice**. [S.l.]: Taylor & Francis, 1995. (Chapman & Hall/CRC Interdisciplinary Statistics). ISBN 9780412055515. Citado na página 52.
- GILKS, W. R.; BERZUINI, C. Following a moving target - monte carlo inference for dynamic bayesian models. **Journal of the Royal Statistical Society: Series B**, v. 63, p. 1–20, 2001. Citado 2 vezes nas páginas 29 e 84.
- GORDON, N.; SALMOND, D.; SMITH, A. F. M. Novel approach to non-linear/non-gaussian bayesian state estimation. **Proc. Inst. Elect. Eng. F**, v. 140, n. 2, p. 107–113, 1993. Citado 4 vezes nas páginas 27, 28, 63 e 98.
- GRANGER, C. W.; MORRIS, M. J. Time series modelling and interpretation. **Journal of the Royal Statistical Society. Series A (General)**, JSTOR, p. 246–257, 1976. Citado na página 121.
- HAN, H.; DING, Y.; HAO, K.; LIANG, X. An evolutionary particle filter with the immune genetic algorithm for intelligent video target tracking. **Computers and Mathematics with Applications**, v. 62, n. 7, p. 2685–2695, 2011. Citado na página 25.
- HANDSCHIN, J. E. Monte carlo techniques for prediction and filtering of non-linear stochastic processes. **Automatica**, v. 6, n. 4, p. 555–563, 1970. Citado na página 27.
- HIGUCHI, T. Genetic algorithm and monte carlo filter. **Proc. Institute Statistical Mathematics**, v. 44, n. 1, p. 19–30, 1996. Citado na página 28.
- JACQUIER, E.; POLSON, N.; ROSSI, P. Bayesian analysis of stochastic volatility models. **Journal of Business & Economic Statistics**, v. 12, p. 371–417, 1994. Citado na página 133.
- JUANG, B. H.; RABINER, L. R. Hidden markov models for speech recognition. **Technometrics**, v. 33, n. 3, p. 251–272, 1991. Citado na página 25.
- JULIER, S.; UHLMANN, J. New extension of the kalman filter to nonlinear systems. In: SPIE INTERNATIONAL SYMPOSIUM ON AEROSPACE/DEFENSE SENSING, SIMULATION AND CONTROLS. [S.l.], 1997. Citado na página 26.
- KALMAN, R. A new approach to linear filtering and prediction problems. **Journal of Basic Engineering**, v. 82, n. 1, p. 35–45, 1960. Citado na página 26.
- KALMAN, R.; BUCY, R. New results in linear filtering and prediction theory. **Journal of Basic Engineering**, v. 83, n. 2, p. 95–107, 1961. Citado na página 26.
- KANTAS, N.; DOUCET, A.; SINGH, S.; MACIEJOWSKI, J.; CHOPIN, N. On particle methods for parameter estimation in state-space models. **Statistical Science**, v. 30, n. 3, 2015. Citado 2 vezes nas páginas 30 e 87.

KASTNER, G.; Frühwirth-Schnatter, S. Ancillarity-sufficiency interweaving strategy (asis) for boosting {MCMC} estimation of stochastic volatility models. **Computational Statistics & Data Analysis**, v. 76, p. 408–423, 2014. Citado 4 vezes nas páginas [133](#), [134](#), [146](#) e [153](#).

KIM, S.; SHEPHARD, N.; CHIB, S. Stochastic volatility: likelihood inference and comparison with arch models. **Review of Economic Studies**, v. 65, p. 361–393, 1998. Citado 2 vezes nas páginas [133](#) e [135](#).

KITAGAWA, G. Monte carlo filter and smoother for non-gaussian nonlinear state space models. **Journal of Computational and Graphical Statistics**, v. 5, n. 1, p. 1–25, 1996. Citado 2 vezes nas páginas [27](#) e [69](#).

_____. A self-organized state-space model. **Journal of the American Statistical Association**, v. 93, n. 443, p. 1203–1215, 1998. Citado 3 vezes nas páginas [29](#), [70](#) e [83](#).

KÜNSCH, H. R. Recursive monte carlo filters: Algorithms and theoretical analysis. **Annals of Statistics**, v. 33, p. 1983–2021, 2005. Citado na página [28](#).

KWOK, N. M.; GU, F.; ZHOU, W. Evolutionary particle filter: re-sampling from the genetic algorithm perspective. In: IEEE/RSJ INTERNATIONAL CONFERENCE ON INTELLIGENT ROBOTS AND SYSTEMS (IROS). Edmonton, Alberta, Canada, 2005. p. 2935–2940. Citado na página [28](#).

KWOK, N. M.; RAD, A. B. A modified particle filter for simultaneous localization and mapping. **Journal of Intelligent and Robotic Systems**, v. 46, n. 4, p. 365–382, 2006. Citado na página [25](#).

LARRANAGA, P.; LOZANO, J. **Estimation of distribution algorithms: A new tool for evolutionary computation**. [S.l.]: Springer Science & Business Media, 2002. Citado 2 vezes nas páginas [30](#) e [105](#).

Le Gland, F.; OUDJANE, N. Stability and uniform approximation of nonlinear filters using the hilbert metric and application to particle filters. **Annals of Applied Probability**, v. 14, p. 1032–1044, 2004. Citado na página [28](#).

LI, T.; BOLIC, M.; DJURIC, P. Resampling methods for particle filtering: Classification, implementation, and strategies. **IEEE Signal Processing Magazine**, v. 32, n. 3, p. 70–86, 2015. Citado 3 vezes nas páginas [56](#), [59](#) e [98](#).

LI, T.; SATTAR, T. P.; SUN, S. Deterministic resampling: Unbiased sampling to avoid sample impoverishment in particle filters. **Signal Processing**, v. 92, n. 7, p. 1637–1645, 2012. Citado na página [28](#).

LI, T.; SUN, S.; SATTAR, T. P.; CORCHADO, J. M. Fight sample degeneracy and impoverishment in particle filters: A review of intelligent approaches. **Expert Systems with applications**, Elsevier, v. 41, n. 8, p. 3944–3954, 2014. Citado na página [90](#).

LIU, J. **Monte Carlo strategies in scientific computing**. New York, NY, USA: Springer-Verlag, 2004. Citado 2 vezes nas páginas [49](#) e [50](#).

LIU, J.; CHEN, R. Sequential monte-carlo methods for dynamic systems. **Journal of the American Statistical Association**, v. 39, n. 443, p. 1032–1044, 1998. Citado 2 vezes nas páginas [27](#) e [28](#).

- LIU, J.; CHEN, R.; LOGVINENKO, T. A theoretical framework for sequential importance sampling with resampling. In: _____. **Sequential Monte Carlo in Practice**. New York: Springer-Verlag, 2001. p. 225–246. Citado 3 vezes nas páginas 98, 99 e 156.
- LIU, J.; WEST, M. Combined parameter and state estimation in simulation-based filtering. In: _____. **Sequential Monte Carlo in Practice**. New York: Springer-Verlag, 2001. p. 197–223. Citado 7 vezes nas páginas 29, 30, 70, 84, 102, 108 e 189.
- LOPES, H.; POLSON, N.; CARVALHO, C. Bayesian statistics with a smile: A resampling-sampling perspective. **Brazilian Journal of Probability and Statistics**, v. 26, n. 4, p. 358–371, 2012. Citado na página 30.
- LOPES, H.; TSAY. Particle filters and bayesian inference in financial econometrics. **Journal of Forecasting**, v. 30, p. 168–209, 2011. Citado 7 vezes nas páginas 25, 30, 90, 91, 108, 189 e 192.
- LOPES, H. F.; POLSON, N. G. Particle learning for fat-tailed distributions. **Econometric Reviews**, Taylor & Francis, p. 1–26, 2016. Citado na página 120.
- MALIK, S.; PITT, M. Particle filters for continuous likelihood evaluation and maximisation. **Journal of Econometrics**, v. 165, n. 2, p. 190–209, 2011. Citado na página 29.
- MUSSO, C.; OUDJANE, N.; Le Gland, F. Improving regularised particle filters. In: _____. **Sequential Monte Carlo in Practice**. New York: Springer-Verlag, 2001. p. 247–271. Citado na página 28.
- NAKAJIMA, J.; OMORI, Y. Leverage, heavy-tails and correlated jumps in stochastic volatility models. **Computational Statistics & Data Analysis**, v. 53, p. 2335–2353, 2009. Citado na página 133.
- NEMETH, C.; FEARNHEAD, P.; MIHAYLOVA, L. Sequential monte carlo methods for state and parameter estimation in abruptly changing environments. **IEEE Transactions on Signal Processing**, v. 62, n. 5, p. 1245–1255, 2014. Citado na página 25.
- NERC Centre for Population Biology, Imperial College. The global population dynamics database. 1999. Disponível em: <<http://www.sw.ic.ac.uk/cpb/cpb/gpdd.html>>. Citado na página 165.
- OMORI, Y.; CHIB, S.; SHPHARD, N.; NAKAJIMA, J. Stochastic volatility with leverage: Fst and efficient likelihood inference. **Journal of Econometrics**, v. 140, p. 425–449, 2007. Citado na página 133.
- PARK, S.; HWANG, J. P.; KIM, E.; KANG, H. J. A new evolutionary particle filter for the prevention of sample impoverishment. **IEEE Transactions on Evolutionary Computation**, v. 13, n. 4, p. 801–809, 2009. Citado na página 28.
- PETERS, G.; HOSACK, G.; HAYES, K. Ecological non-linear state space model selection via adaptive particle markov chain monte carlo. **Arxiv preprint arXiv:1005.2238**, 2010. Citado na página 155.
- PETRIS, G.; PETRONE, S.; CAMPAGNOLI, P. **Dynamic linear models**. [S.l.]: Springer, 2009. Citado na página 108.

- PITT, M.; SHEPHARD, N. Filtering via simulation: auxiliary particle filters. **Journal of the American Statistical Association**, v. 94, n. 446, p. 590–599, 1999. Citado 3 vezes nas páginas 28, 65 e 99.
- POLSON, N.; STROUD, J.; MÜLLER, P. Practical filtering with sequential parameter learning. **Journal of the Royal Statistical Society: Series B**, v. 70, n. 2, p. 413–428, 2008. Citado na página 108.
- POYIADJIS, G.; DOUCET, A.; SINGH, S. Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. **Biometrika**, v. 98, n. 1, p. 65–80, 2011. Citado na página 29.
- PRADO, R.; LOPES, H. Sequential parameter learning and filtering in structured autoregressive state-space models. **Statistics and Computing**, v. 23, n. 1, p. 43–57, 2013. Citado 5 vezes nas páginas 30, 100, 108, 112 e 121.
- PRADO, R.; WEST, M. **Time Series: Modeling, Computation, and Inference**. 1st. ed. [S.l.]: Chapman & Hall/CRC, 2010. ISBN 978-1-4200-9336-0. Citado na página 108.
- RUBIN, D. B. A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The sir algorithm. **Journal of the American Statistical Association**, v. 82, n. 398, p. 543–546, 1987. Citado na página 98.
- SHIRYAEV, A. N. **Probability**. 2nd. ed. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 1995. ISBN 0-387-94549-0. Citado na página 54.
- SILVERMAN, B. W. **Density Estimation for Statistics and Data Analysis**. London: Chapman & Hall, 1986. ISBN 0-412-24620-1. Citado 3 vezes nas páginas 102, 103 e 156.
- SINGH, N. C. S. On particle gibbs sampling. **Bernoulli**, v. 21, n. 3, p. 1855–1883, 2015. Citado 2 vezes nas páginas 29 e 70.
- STORVIK, G. Particle filters in state space models with the presence of unknown static parameters. **IEEE Transactions on Signal Processing**, v. 50, n. 2, p. 281–289, 2002. Citado 7 vezes nas páginas 29, 30, 70, 75, 84, 115 e 189.
- THRUN, S.; BURGARD, W.; FOX, D. **Probabilistic Robotics**. Cambridge, MA: MIT Press, 2005. Citado na página 25.
- VIDAL-RODEIRO, C.; LAWSON, A. Online updating of space-time disease surveillance models via particle filters. **Statistical Methods in Medical Research**, v. 15, n. 5, p. 423–444, 2006. Citado na página 25.
- WEST, M. Approximating posterior distributions by mixture. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 409–422, 1993. Citado na página 173.
- WEST, M.; HARRISON, J. **Bayesian Forecasting and Dynamic Models (2nd ed.)**. New York, NY, USA: Springer-Verlag, 1997. Citado na página 102.
- ZHAO, J.; LI, Z. Particle filter based on particle warm optimization resampling for vision tracking. **Expert Systems with Applications**, v. 37, n. 12, p. 8910–8914, 2010. Citado na página 25.

FATOR DE FERTILIDADE × TAMANHO EFETIVO AMOSTRAL

Este experimento analisa a relação entre fator de fertilidade e tamanho efetivo amostral considerando quatro dos principais dispositivos de reamostragem encontrados na literatura: multinomial, residual, sistemático e por ramificação. O principal objetivo da análise é estabelecer uma novo critério de avaliação de degeneração no contexto geral dos métodos sequenciais de Monte Carlo. Como primeira aplicação, discutiremos os algoritmos de reamostragem sob uma perspectiva alternativa, a qual nos permite selecionar o método mais adequado de forma objetiva e didática.

Metodologia

O ensaio é baseado no estudo de uma coleção independente de medidas aleatórias construídas como estimativas da distribuição normal padrão a partir do método de amostragem por importância. As aproximações foram produzidas tomando uma família de distribuições propostos Gaussianas $N(\mu_i, \sigma_i^2)$ cujos parâmetros foram escolhidos aleatoriamente segundo o modelo

$$\mu_i, \frac{\sigma_i^2}{3} \sim N(0, 1) \times \text{Beta}(5, 2), \quad i = 1, 2, \dots, R.$$

O procedimento é resumido pela iteração dos seguintes passos:

- 1:** amostre $\mu_i, \frac{\sigma_i^2}{3} \sim N(0, 1) \times \text{Beta}(5, 2)$;
- 2:** amostre $x^{(j)} \sim N(\mu_i, \sigma_i^2)$, $j = 1, 2, \dots, M$;
- 3:** calcule as ponderações $\omega^{(i)} = \exp \left\{ -\frac{1}{2} \left(x^{(j)} \right)^2 + \frac{1}{2\sigma_i^2} \left(x^{(j)} - \mu_i \right)^2 \right\}$, $j = 1, 2, \dots, M$;
- 4:** armazene a aproximação discreta $\left\{ x^{(j)}, \omega^{(j)} / \sum_{k=1}^M \omega^{(k)} \right\}_{j=1}^M$.

Considerando $R = 5000$ e $M = 2R$, teremos ao final do conjunto de iterações acima uma coleção de 5000 medidas aleatórias distintas,

$$\mathcal{M} = \left\{ \left(x_1^{(i)}, \tilde{\omega}_1^{(i)} \right)_{i=1}^M, \left(x_2^{(i)}, \tilde{\omega}_2^{(i)} \right)_{i=1}^M, \dots, \left(x_R^{(i)}, \tilde{\omega}_R^{(i)} \right)_{i=1}^M \right\}$$

onde

$$\tilde{\omega}_k^{(i)} = \frac{e^{-\frac{1}{2}(x_k^{(i)})^2 + \frac{1}{2\sigma_k^2}(x_k^{(i)} - \mu_k)^2}}{\sum_{i=1}^M e^{-\frac{1}{2}(x_k^{(i)})^2 + \frac{1}{2\sigma_k^2}(x_k^{(i)} - \mu_k)^2}}, \quad x_k^{(i)} \sim N(\mu_k, \sigma_k^2), \quad i = 1, 2, \dots, M, k = 1, 2, \dots, R,$$

cada uma associada a um único tamanho efetivo amostral

$$\text{TEA}_k = \frac{1}{M \cdot \sum_{i=1}^M \tilde{\omega}_i^2}, \quad k = 1, 2, \dots, R.$$

Associamos ainda cada elemento de \mathcal{M} a quatro fatores de fertilidade

$$\begin{aligned} FF_k^{(i)} &= FF_k \left(\left(x_1^{(i)}, \tilde{\omega}_1^{(i)} \right)_{i=1}^M \right) \\ &= \frac{\text{\#número de indivíduos distintos selecionados a partir do método } k}{M}, \end{aligned}$$

onde $k =$ Multinomial, Residual, Sistemático, Ramificação.

A escolha das distribuições para a dupla (μ, σ^2) foi dada de modo a garantir um conjunto bastante diversificado de tamanhos efetivos amostrais. O experimento concentra esforços na análise de propriedades associadas a medidas aleatórias em um contexto geral, independentemente da complexidade da medida aproximada assim como da dimensão de seu respectivo suporte. Como estes atributos específicos são irrelevantes para o ensaio, famílias de distribuições Gaussianas unidimensionais foram utilizadas por mera conveniência.

Discussão

Tamanhos efetivos amostrais muito elevados são associados a medidas aleatórias de alta entropia. Nestes casos, as ponderações associadas a cada elemento do suporte aleatório são bastante homogêneas $\left(\tilde{\omega}_k^{(i)} \approx \frac{1}{M} \right)$ sugerindo que a amostra construída por importância resume um montante de informação similar ao de uma amostra aleatória. Ao aplicar um dispositivo de reamostragem nesta circunstância devemos esperar que uma fração ínfima da população seja eliminada caracterizando um alto fator de fertilidade. Em contrapartida, verifica-se que medidas aleatórias conectadas a conjuntos excessivamente desequilibrados de pesos resultam em TEAs muito pequenos. A desarmonia entre as importâncias associadas a cada indivíduo neste cenário induz uma discrepância significativa na população já que uma pequena parcela desta será ligada a massas de probabilidade desproporcionalmente grandes. Como resultado teremos uma enorme pressão seletiva exercida sobre a amostra e, conseqüentemente, fatores de fertilidade muito baixos serão observados após o processo de reamostragem. Desta forma, considerando as duas situações, é bastante razoável esperar uma correlação positiva entre as variáveis FF e TEA.

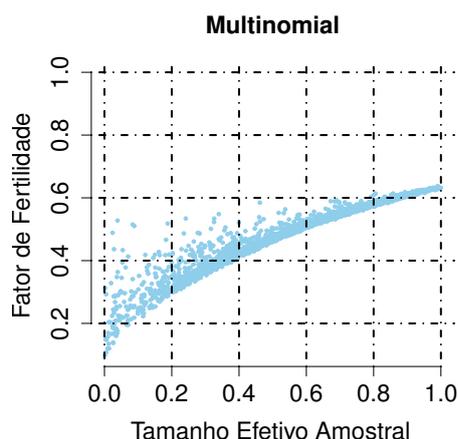


Figura 86 – Comparação entre TEAs e FFs obtidos através do método de seleção multinomial.

O resultado ilustrado acima demonstra de forma clara a relação entre os TEAs e os FFs considerando o esquema multinomial. Mais precisamente, obtemos o coeficiente de correlação

$$\rho_{\text{Mult}} = 0,9701,$$

caracterizando uma dependência fortemente positiva entre as variáveis. Este efeito nos permite tomar o FF como uma medida de degeneração a ser explorada em análises posteriores ao passo de seleção mantendo uma relação de coerência com análises prévias construídas a partir dos TEAs. Analisar o comportamento do enxame de partículas posteriormente nos habilita a observar o fenômeno de degeneração amostral a partir de uma perspectiva mais concreta, favorecendo a proposição de estratégias mais robustas.

Efeitos similares são observados se desenvolvermos o mesmo ensaio com respeito a outras metodologias de reamostragem, entretanto conclusões distintas, e bastante elucidativas, podem ser extraídas a partir da comparação entre os métodos. Adotando o dispositivo de seleção residual, por exemplo, obtemos novamente um coeficiente de correlação bastante elevado:

$$\rho_{\text{Res}} = 0,9855.$$

Todavia, se compararmos o desempenho dos fatores de fertilidade, é possível identificar um ganho significativo da abordagem residual. A saída ilustrada na Figura 89 indica que o desempenho dos métodos é equivalente caso estejamos lidando com medidas aleatórias associadas a tamanhos efetivos pequenos ($\text{TEA} < 0,4$). Porém, percebe-se um ganho de diversidade à medida que os TEAs aumentam. Este resultado é expressivo pois indica que o dispositivo de seleção multinomial descarta boa parte da população mesmo em situações onde a amostra ponderada se assemelha a uma amostra aleatória, o que pode ser problemático principalmente em situações onde o enxame de partículas considerado é isento de dinâmica de evolução natural. A comparação realizada em um conjunto relativamente grande de medidas independentes mostra que o resultado não trata de uma mera coincidência, justificando a necessidade de uma maior atenção à escolha do procedimento de seleção adotado durante a implementação de um filtro de partículas.

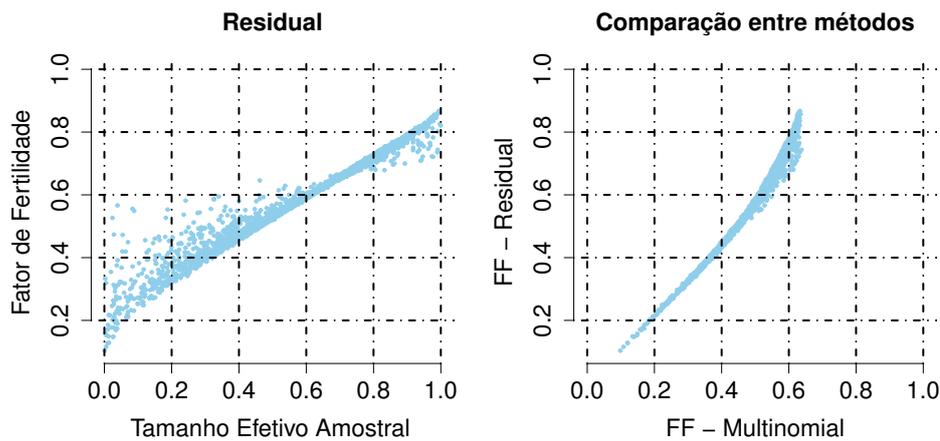


Figura 87 – Comparação entre TEAs e FFs. À esquerda segue a relação entre TEA e FF obtida através do método de seleção residual. À direita segue a comparação entre os FFs associados às abordagens residual e multinomial.

Repetindo a análise para os métodos de seleção sistemático e por ramificação, obtivemos mais uma vez a confirmação de uma dependência fortemente positiva entre FF e TEA,

$$\rho_{\text{Sist}} = 0.9782 \quad \rho_{\text{Ram}} = 0.9782581,$$

além de observar um novo ganho de diversidade. Os resultados apresentados pela Figura 88

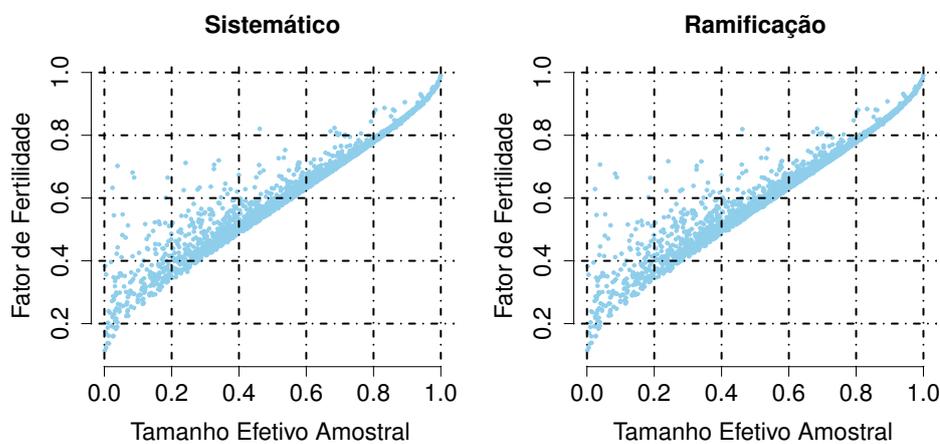


Figura 88 – Comparação entre TEAs e FFs obtidos através dos métodos de seleção sistemático e por ramificação.

justificam de forma alternativa a otimalidade destas duas tratativas: a diversidade da população é preservada ao máximo. Notemos que ambos os dispositivos preservam quase totalmente a diversidade da população ao lidar com amostras caracterizadas por altos tamanhos efetivos ($\text{TEA} \approx 1$). Em outras palavras, tais métodos não enxugam a população ponderada quando esta pode ser tomada como uma ótima aproximação de medida aleatória. A mesma propriedade não é verificada nos dois procedimentos anteriores conforme apresentado nas Figuras 86 e 89.

A última análise traçada neste ensaio visa comparar a capacidade de preservação de diversidade entre os quatro métodos. Esta análise é relevante para a escolha do método mais

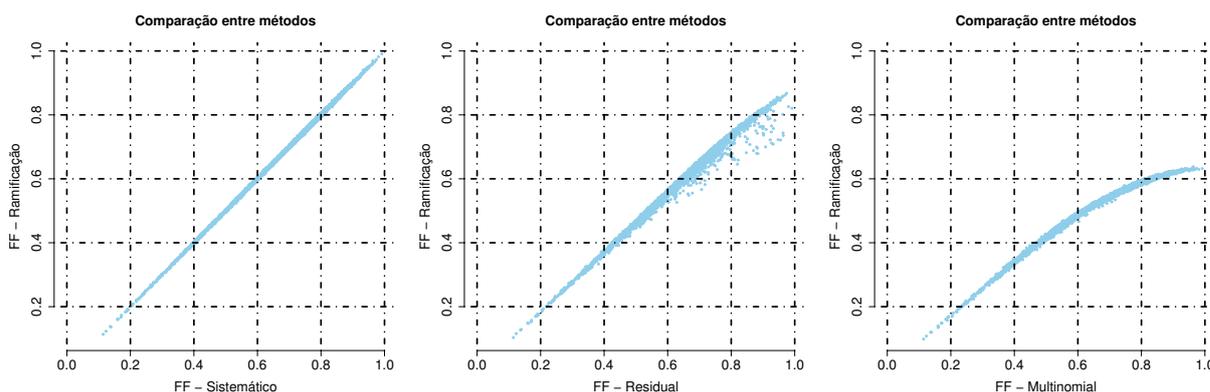


Figura 89 – Desempenho comparativo entre métodos de reamostragem. Da esquerda para a direita: ramificação \times sistemática, ramificação \times residual e ramificação \times multinomial.

adequado já que a complexidade computacional de todos eles é equivalente – $O(N)$, onde N é o tamanho amostral. A figura acima deixa claro que o método multinomial retorna resultados inferiores, os quais podem ser descritos como um grande desperdício de informação. Este desperdício só pode ser compensado aumentando o tamanho do enxame de pontos, o que fatalmente tornaria o procedimento menos atrativo computacionalmente. A abordagem residual apresenta um bom ganho de diversidade com respeito ao procedimento multinomial porém é superada facilmente pelas duas últimas tratativas. Em resumo, sua utilização não deve comprometer severamente o desempenho do filtro porém não oferece um resultado otimizado. As metodologias sistemática e por ramificação demonstram um comportamento rigorosamente idêntico do ponto de vista prático; a única diferença significativa entre elas está associada ao tratamento analítico das mesmas. O método de seleção por ramificação pode ser tratado analiticamente de maneira mais simples (BAIN; CRISAN, 2009; CRISAN; Del Moral; LYONS, 1999), o que favorece a demonstração de propriedades de convergência e estabilidade dos filtros de partículas como um todo.

DETERMINAÇÃO DO COEFICIENTE DE SUAVIZAÇÃO λ

O coeficiente λ utilizado na construção do procedimento de reamostragem suavizada foi definido inicialmente pela relação

$$\text{TEA} \left(\left[\tilde{\omega}_t^{(1)} \right]^\lambda, \left[\tilde{\omega}_t^{(2)} \right]^\lambda, \dots, \left[\tilde{\omega}_t^{(N)} \right]^\lambda \right) - \frac{1}{10} = 0$$

Em outras palavras, λ é definido como um possível zero da função

$$f(L) = \frac{\left[\sum_{i=1}^N \left(\tilde{\omega}_t^{(i)} \right)^L \right]^2}{N \cdot \sum_{i=1}^N \left(\tilde{\omega}_t^{(i)} \right)^{2L}} - \frac{1}{10}.$$

Reforçamos que esta definição é explorada exclusivamente em situações onde o tamanho efetivo amostral observado foi inferior à cota inferior $\gamma = 1/10$, isto é, buscamos o coeficiente apenas em instantes $t \in \tau$ (4.10).

Para garantir a existência do expoente ideal λ , notemos inicialmente que f é formada por uma composição aritmética de funções exponenciais, o que assegura sua continuidade. Desta forma, basta mostrarmos que f muda de sinal no intervalo $(0, 1]$. Por construção, temos que

$$f(1) < 0, \quad \forall t \in \tau.$$

Além disso, cada ponderação $\tilde{\omega}_t^{(i)}$ é positiva e inferior a 1. A primeira afirmação segue da definição do procedimento de amostragem por importância a qual parte da escolha de distribuições propostas absolutamente contínuas com respeito à medida alvo. Já o limite superior é obtido do processo de normalização dos pesos. Assim,

$$\left(\tilde{\omega}_t^{(i)} \right)^L \xrightarrow{L \downarrow 0} 1, \quad \forall i = 1, 2, \dots, N \implies f(L) \xrightarrow{L \downarrow 0} \frac{9}{10} > 0.$$

Concluimos então que a função f possui de fato uma raiz no intervalo $(0, 1]$. Podemos ainda verificar a unicidade de λ mostrando que f é uma função monótona no intervalo $(0, 1]$. Derivando f , obtemos:

$$\begin{aligned} f'(L) &= \frac{d}{dL} \left\{ \frac{\left[\sum_{i=1}^N (\tilde{\omega}_t^{(i)})^L \right]^2}{N \cdot \sum_{i=1}^N (\tilde{\omega}_t^{(i)})^{2L}} \right\} \\ &= \frac{2N \cdot \left\{ \left[\sum_{i=1}^N (\tilde{\omega}_t^{(i)})^{2L} \right] \left[\sum_{i=1}^N (\tilde{\omega}_t^{(i)})^L \cdot \ln \tilde{\omega}_t^{(i)} \right] \left[\sum_{i=1}^N (\tilde{\omega}_t^{(i)})^L \right] - \left[\sum_{i=1}^N (\tilde{\omega}_t^{(i)})^L \right]^2 \left[\sum_{i=1}^N (\tilde{\omega}_t^{(i)})^{2L} \cdot \ln \tilde{\omega}_t^{(i)} \right] \right\}}{N^2 \cdot \left[\sum_{i=1}^N (\tilde{\omega}_t^{(i)})^{2L} \right]^2} \\ &= \frac{1}{\left[\sum_{i=1}^N (\tilde{\omega}_t^{(i)})^L \right]} \cdot \text{TEA}_t \cdot \sum_{i=1}^N (\tilde{\omega}_t^{(i)})^L \cdot \ln \tilde{\omega}_t^{(i)} - \frac{1}{\left[\sum_{i=1}^N (\tilde{\omega}_t^{(i)})^{2L} \right]} \cdot \text{TEA}_t \cdot \sum_{i=1}^N (\tilde{\omega}_t^{(i)})^{2L} \cdot \ln \tilde{\omega}_t^{(i)}, \end{aligned}$$

onde

$$\text{TEA}_t = \text{TEA} \left(\left[\tilde{\omega}_t^{(1)} \right]^L, \left[\tilde{\omega}_t^{(2)} \right]^L, \dots, \left[\tilde{\omega}_t^{(N)} \right]^L \right) = \frac{\left[\sum_{i=1}^N (\tilde{\omega}_t^{(i)})^L \right]^2}{N \cdot \sum_{i=1}^N (\tilde{\omega}_t^{(i)})^{2L}}.$$

Como $\left(\tilde{\omega}_t^{(i)} \right)^L > \left(\tilde{\omega}_t^{(i)} \right)^{2L}$, $i = 1, 2, \dots, N$, segue que

$$\begin{aligned} f'(L) &\leq \frac{1}{\left[\sum_{i=1}^N (\tilde{\omega}_t^{(i)})^{2L} \right]} \cdot \text{TEA}_t \cdot \sum_{i=1}^N (\tilde{\omega}_t^{(i)})^L \cdot \ln \tilde{\omega}_t^{(i)} - \frac{1}{\left[\sum_{i=1}^N (\tilde{\omega}_t^{(i)})^{2L} \right]} \cdot \text{TEA}_t \cdot \sum_{i=1}^N (\tilde{\omega}_t^{(i)})^{2L} \cdot \ln \tilde{\omega}_t^{(i)} \\ &= \underbrace{\left[\frac{1}{\left[\sum_{i=1}^N (\tilde{\omega}_t^{(i)})^{2L} \right]} \cdot \text{TEA}_t \right]}_{\geq 0} \times \left[\sum_{i=1}^N \underbrace{\left(\ln \tilde{\omega}_t^{(i)} \right)}_{\leq 0} \underbrace{\left(\left(\tilde{\omega}_t^{(i)} \right)^L - \left(\tilde{\omega}_t^{(i)} \right)^{2L} \right)}_{\geq 0} \right] \leq 0, \end{aligned}$$

conforme queríamos verificar.

Os valores de λ podem ser determinados de maneira eficiente através de procedimentos de busca unidimensionais como o método da bissecção uma vez que o domínio de interesse é muito pequeno. Por mera simplicidade, adotamos o procedimento *uniroot* do pacote *rootSolve* disponível para linguagem R. A função implementa uma variante do algoritmo de Newton-Raphson em Fortran, possibilitando o cálculo de raízes de funções em tempo bastante ágil.

COMENTÁRIOS SOBRE “PARTICLE FILTERS AND BAYESIAN INFERENCE IN FINANCIAL ECONOMETRICS” DE LOPES, H. E TSAY R.

Em sua revisão acerca dos métodos sequenciais de Monte Carlo aplicados à econometria, [Lopes e Tsay \(2011\)](#) traçam uma discussão bastante frutífera a respeito dos algoritmos do tipo filtro de partículas, sobretudo em cenários influenciados pela presença de parâmetros estáticos desconhecidos. O estudo demonstra experimentalmente a funcionalidade dos métodos de Liu e West ([LIU; WEST, 2001](#)), de Storvik ([STORVIK, 2002](#)) e *Parameter Learning* ([CARVALHO et al., 2010](#)), construindo assim um compêndio indispensável a todos os profissionais interessados em iniciar seus estudos na área.

O presente estudo revisa um dos experimentos apresentados no artigo além de abordar alguns poucos detalhes acerca da implementação geral dos métodos na tentativa de enriquecer ainda mais a discussão proposta originalmente. O trabalho foca em dois pontos principais:

- 1) verificação de que o caso particular de degeneração apresentado pelo filtro de Liu e West é fruto de um pequeno equívoco de código;
- 2) demonstração de que simples modificações no conceito de reamostragem podem potencializar notadamente a inferência realizada por meio do filtro de Liu e West.

Esperamos que os comentários apresentamos aqui possam contribuir de maneira significativa para um melhor entendimento do esquema de filtragem de Liu e West, reforçando a ideia de que dispositivos relativamente simples podem ser capazes de ofertar análises consistentes mesmo em cenários complexos.

Comparação entre LW, Storvik e PL revisitada

Revisamos inicialmente o experimento apresentado no Exemplo 4 do artigo original. Este experimento é particularmente importante por tratar de uma das poucas aplicações encontradas na literatura onde o filtro LW é implementado em associação ao filtro auxiliar completamente adaptado. Neste caso, deveríamos esperar um desempenho superior desta abordagem com relação à sua implementação usual por meio do filtro auxiliar, contudo os resultados sumarizados nas Figuras 3 e 4 da obra indicam que o algoritmo pode apresentar um comportamento espúrio mesmo em se tratando de uma situação ótima. A obtenção de resultados adversos durante a análise de um problema bastante simples fez deste ensaio uma importante referência na literatura para justificar o possível comportamento degenerativo do método. Todavia, uma análise minuciosa do código disponibilizado pelos autores (<http://hedibert.org/r-code-to-our-journal-of-forecasting-review-paper/>) nos permitiu concluir que a instabilidade observada ao longo das execuções do método é oriunda de um sutil deslize na implementação. A rigor, verifica-se que o algoritmo LW é codificado de maneira precisa, no entanto as implementações associadas aos dispositivos de Storvik e PL não tratam corretamente o esquema de propagação de estatísticas suficientes induzindo ambos os procedimentos à construção de estimativas inadequadas. Uma vez que o dispositivo PL baseado em um número expressivo de partículas é tomado como referência de solução ótima (exata) para o experimento e a implementação do mesmo apresenta falhas, temos que o método LW deve necessariamente apresentar um desempenho aquém do esperado. A fim de facilitar o entendimento do problema apontado aqui, apresentamos o trecho específico correspondente ao descuido de implementação; basicamente, o parâmetro σ^2 é atualizado segundo a oitava entrada da matriz de estatísticas suficientes quando na verdade deveria explorar a oitava coluna da mesma (linhas 130 e 197 do código).

Lopes e Tsay

```
# Sampling parameters
```

```
sig2s = 1/rgamma(N,s[8]/2,s[,9]/2)
```

Correção

```
# Sampling parameters
```

```
sig2s = 1/rgamma(N,s[,8]/2,s[,9]/2)
```

Os resultados obtidos a partir da correção proposta são apresentados nas Figuras 92 e 91 a seguir; as quais preservam, por uma questão de coerência, os parâmetros gráficos utilizados no experimento original. Reforçamos que neste primeiro momento foi utilizada a implementação do filtro LW disponibilizada pelos autores sem nenhuma alteração. As correções são aplicadas apenas aos métodos de Storvik e PL de acordo com o trecho de código sugerido acima. Podemos concluir seguramente que o filtro de Liu e West ostenta de fato uma maior aglomeração de variabilidade em suas estimativas, contudo o mesmo não demonstra um comportamento verdadeiramente degenerativo conforme especulado em um primeiro momento.

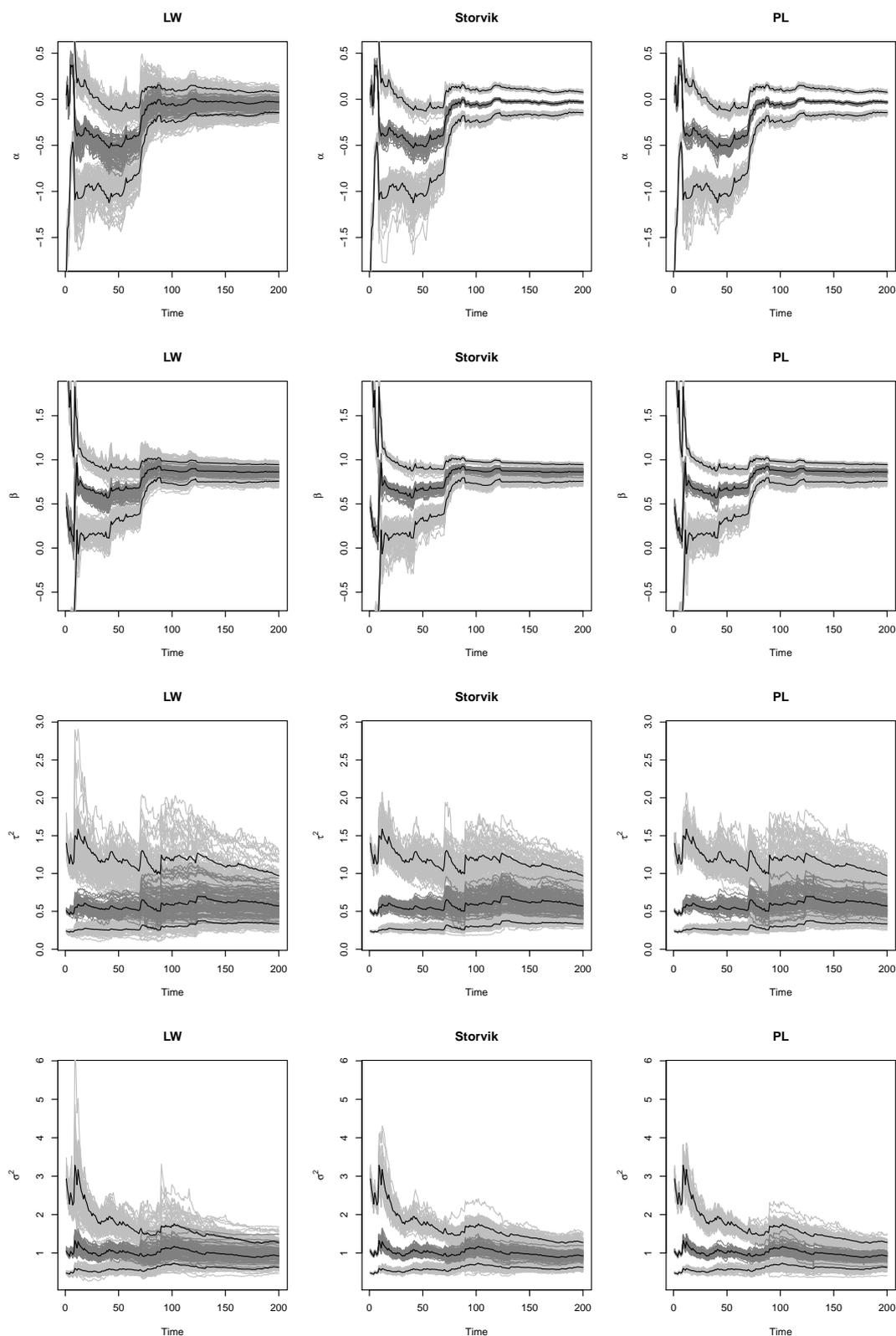


Figura 90 – Comparação corrigida entre os métodos LW, Storvik e PL. Percentis de $p(\theta|y^t)$ (2,5%;50%;97,5%) baseados em 100 replicações de cada filtro com $N = 1000$ partículas (trajetórias cinzas). Curvas pretas são baseadas no filtro PL com $N = 100.000$ partículas. Filtro LW (coluna à esquerda), filtro de Storvik (coluna central) e PL (coluna à direita). As linhas (de cima para baixo) representam as componentes de $\theta = (\alpha, \beta, \tau^2, \sigma^2)$.

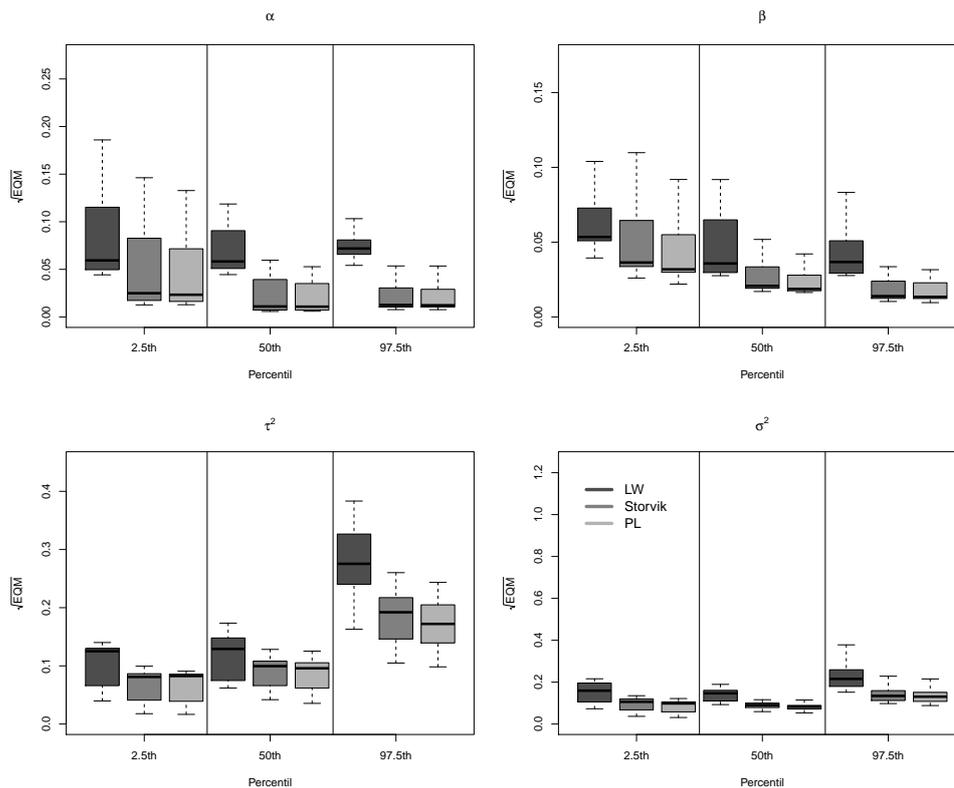


Figura 91 – Comparação corrigida entre os filtros LW, Storvik e PL. Raíz do erro quadrático médio de $R = 100$ replicações de cada filtro. Todos os filtros são baseados em $N = 1000$ partículas e a raíz do erro quadrático médio é calculada com respeito a uma longa execução do filtro PL ($N = 100.000$).

Sugestões sobre reamostragem

A utilização indevida de dispositivos de reamostragem pode conduzir os filtros de partículas a resultados mais imprecisos em um contexto geral devido ao fenômeno de degeneração amostral. Quando implementados na presença de parâmetros estáticos desconhecidos, estes algoritmos sofrem ainda mais com o problema por basearem sua inferência na informação finita disponibilizada por populações desprovidas de dinâmica natural. Desta forma, estratégias como a empregada no desenvolvimento do filtro de Liu e West deveriam realizar um único passo de reamostragem buscando evitar o enxugamento desnecessário da nuvem de partículas estáticas. Em outras palavras, a sugestão proposta por Lopes e Tsay (2011) deveria ser substituída pela proposta original do filtro LW a qual realiza apenas o primeiro passo de reamostragem considerando os pesos

$$\omega_t^{(i)} \propto \omega_t^{(i-1)} \cdot p(y_t | g(x_{t-1})^{(i)}, m^{(i)}).$$

Além disso, é importante investir na utilização de conceitos mais robustos de reamostragem como o empregado pelo dispositivo sistemático de Carpenter, Clifford e Fearnhead (1999), por exemplo. Estas perspectivas preservam com mais efetividade a diversificação do enxame amostral, garantido assim um montante de informação mais razoável para o desenvolvimento do método de Liu e West.

O experimento abaixo ilustra o ganho obtido a partir da simples sugestão acima. Neste exemplo consideramos o modelo de volatilidade estocástica com ruído t – *student* aplicado à série de log-retornos mensais estudada no artigo. Para uma mesma quantidade de pontos, a versão proposta por Lopes e Tsay apresenta um acúmulo de variabilidade muito superior. A melhora obtida nos resultados da coluna à direita indicam que boa parte desta variabilidade se deve à interpretação específica do algoritmo de Liu e West apresentada pelos autores, e não à estratégia de uma forma global.

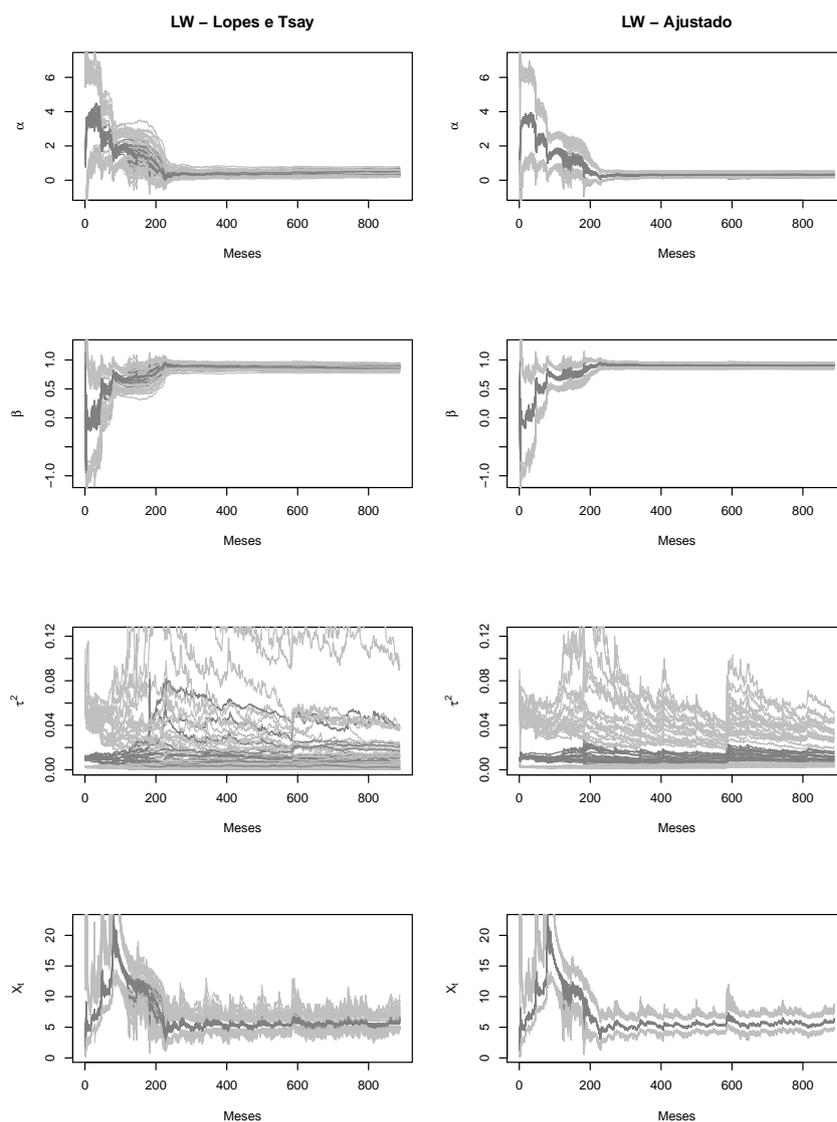


Figura 92 – Comparação entre diferentes interpretações do filtro de Liu e West. A coluna da esquerda implementa a proposta sugerida por Lopes e Tsay enquanto a coluna da direita considera a sugestão deste trabalho. Em ambos os casos foram consideradas 20 execuções independentes com 5000 partículas cada.

PACOTES COMPUTACIONAIS E PÁGINAS INTERESSANTES NA INTERNET

<<https://cran.r-project.org/web/packages/rootSolve/>> Pacote *rootSolve* utilizado como apoio para o cálculo de raízes de funções unidimensionais;

<<https://cran.r-project.org/web/packages/microbenchmark/>> Pacote *microbenchmark* utilizado como apoio para a medição de tempos de execução de procedimentos implementados na linguagem R.

<<https://cran.r-project.org/web/packages/mvnfast/>> Pacote *mvnfast* utilizado como apoio para a geração mais eficiente de variáveis aleatórias Gaussianas multivariadas a partir de interfaces para implementação em C e MPI.

<<https://cran.r-project.org/web/packages/stochvol/>> Pacote *stochvol* utilizado como apoio para o estudo de modelos de volatilidade estocástica a partir de métodos MCMC.

<<https://github.com/nchopin/py-smc2>> Código fonte do algoritmo SMC² implementado em python por Pierre Jacob. A implementação facilita a utilização do algoritmo uma vez que o procedimento principal é implementado a partir de classes mais gerais as quais recebem como parâmetro um modelo qualquer definido pelo usuário.

<<http://www.keithschwarz.com/darts-dice-coins/>> Descrição do método de amostragem por tabelas de pseudônimos, um algoritmo alternativo para a geração de amostras multinomias em complexidade linear;

<http://www.stats.ox.ac.uk/~doucet/smc_resources.html> Excelente compêndio da literatura dos métodos SMC disponibilizado pelo professor Arnaud Doucet da universidade de Oxford;

[<https://sdw.ecb.europa.eu/>](https://sdw.ecb.europa.eu/) Banco de dados estatísticos do Banco Central Europeu; fonte de dados interessante para o estudo de modelos econométricos.