

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

**Modelos Dinâmicos com estimação via Filtro de
Partículas**

IGOR FERREIRA DO NASCIMENTO

12/0001781

Brasília

2014

IGOR FERREIRA DO NASCIMENTO

12/0001781

Modelos Dinâmicos com estimação via Filtro de Partículas

Relatório apresentado à disciplina Estágio Supervisionado II do curso de mestrado em Estatística, Departamento de Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para o grau de Mestre em Estatística.

Orientadora: Professora Doutora Cibele Queiroz da Silva

Brasília

2014

Dedicatória

A mim e a todos que acreditam em mim.

*”Inteligência é capacidade que se tem de aceitar o que está
ao redor.”*

William Faulkner

Agradecimentos

Agradeço, majoritariamente, a Deus.

À dedicada, atenciosa e paciente professora Cibele, que orientou esta dissertação e me ajudou a concluí-la em meio as inúmeras dificuldades ao longo do trabalho.

Aos demais professores do departamento de Pós-graduação em Estatística.

Aos meus chefes, pela compreensão.

Resumo

Esta dissertação apresenta, de forma detalhada, as metodologias discutidas nos principais artigos seminais na área de Filtro de Partículas (FP). O FP é uma técnica alvo de recentes estudos que apresenta soluções numéricas para problemas de difícil solução analítica. Tal técnica será abordada nos modelos dinâmicos e, para isso, necessitará de recursos como Amostragem de Importância e Monte Carlo Sequencial, também discutidos neste trabalho. Os resultados encontrados nos FP seminais que foram estudados nesta dissertação, Gordon et al. (1993) e Pitt e Shephard (1999), tiveram estimativas próximas às soluções analítica encontradas no Filtro de Kalman. Além disso, foram comparadas duas técnicas de estimação dos parâmetros estáticos dos artigos de Liu e West (2001) e de Storvik (2002). Foram obtidos bons resultados, tanto para estimação dos estados latentes, quanto para os parâmetros estáticos. Dessa forma, a técnica de FP mostrou-se ser uma excelente opção de estimação por meio de aproximações numéricas *online*, no âmbito dos Modelos Dinâmicos.

Palavras chaves: Filtro de partículas, Modelos dinâmicos, Monte Carlo Sequencial, Amostragem de Importância, Estimação *online*.

Abstract

The technique FP has been the subject of recent studies and provides numerical solutions to difficult problems analytical solution. This technique will be discussed in Dynamic Models and, therefore, require resources such as Importance Sampling and Sequential Monte Carlo, also discussed in this paper. The results found in seminal FP that were studied in this dissertation, Gordon et al. (1993) and Pitt e Shephard (1999), provides close estimates to analytical solutions found in the Kalman filter estimates. In addition, two techniques for the estimation of static parameters of articles Liu e West (2001) and Storvik (2002) were compared. Good results were obtained both for estimates of latent states, as for the static parameters. Thus, the FP technique proved to be an excellent choice of estimates by means of numerical approximations online, under the Dynamic Models.

Keywords: Particle Filter, Dynamic Models, Monte Carlo sequential, Importance sampling, online estimation.

Lista de Tabelas

5.1	Estimativas para os parâmetros estáticos do modelo dinâmico binomial utilizando os métodos de Storvik (2002) e Liu e West (2001) e com filtragem utilizando os métodos de Gordon et al. (1993) e Pitt e Shephard (1999).	80
6.1	Resumo para estimação via Filtro de Partículas.	92

Lista de Figuras

2.1	Modelo de Espaço de Estados.	5
2.2	Simulação para um Modelo Dinâmico Linear Gaussiano	13
2.3	Estimativas dos estados latentes pelo Filtro de Kalman	14
2.4	Estimativas suavizadas dos estados latentes pelo Filtro de Kalman	15
2.5	Comparação entre as estimativas suavizadas e filtradas	16
4.1	Representação da localização de um robô ,em azul, numa sala com obstruções. . .	29
4.2	Disposição das partículas (localizações hipotéticas), em rosa, na sala.	29
4.3	Sensor identificando as partículas mais próximas à atual posição do robô.	31
4.4	Selecionando partículas mais prováveis como representantes da posição do robô. .	31
4.5	Gerando novos palpites em torno dos palpites mais prováveis.	32
4.6	Filtro de partículas Bootstrap - esquema para gerar partículas segundo o ciclo de atualização: $p(\theta_{t-1} D_{t-1}) \rightarrow p(\theta_t D_{t-1}) \rightarrow p(\theta_t D_t)$	44
4.7	Equação de transição não-linear	47
4.8	FP Bootstrap em um modelo não-linear	48
4.9	Densidades do processo de estimação BF	49
4.10	FP Auxiliar: Estimativas dos estados latentes	58
4.11	Densidades do processo de estimação FP auxiliar	59

4.12	Processo ARCH	60
4.13	Estimação para o modelo ARCH	61
4.14	Processo dinâmico: Gaussiano linear de ordem 1	63
4.15	FP simples: Gaussiano linear de ordem 1	66
4.16	FP ótimo: Gaussiano linear de ordem 1	68
4.17	Tabela e Boxplot dos desvios quadráticos, com relação ao valor verdadeiro dos estados latentes $\theta_{1:T}$, das metodologias Filtro de Kalman, FP básico e FP ótimo.	68
5.1	Modelo dinâmico binomial simulado (vide equação 5.8).	79
5.2	Estimativas para os percentis do parâmetro α no modelo Binomial utilizando o processo de estimação dos parâmetros via Storvik (2002) e o método de filtragem de Gordon et al. (1993).	81
5.3	Estimativas para os percentis do parâmetro β no modelo Binomial utilizando o processo de estimação dos parâmetros via Storvik (2002) e o método de filtragem de Gordon et al. (1993).	82
5.4	Estimativas para os percentis do parâmetro σ^2 no modelo Binomial utilizando o processo de estimação dos parâmetros via Storvik (2002) e o método de filtragem de Gordon et al. (1993). A transformação escolhida para esse parâmetro foi a logaritmica.	83
5.5	Estimativas para os percentis dos estados latentes $x_{1:100}$ no modelo Binomial utilizando o processo de estimação dos parâmetros via Storvik (2002) e o método de filtragem de Gordon et al. (1993).	84

5.6	Estimativas para os percentis do parâmetro α no modelo Binomial utilizando o processo de estimação dos parâmetros via Liu e West (2001) e o método de filtragem de Gordon et al. (1993).	84
5.7	Estimativas para os percentis do parâmetro β no modelo Binomial utilizando o processo de estimação dos parâmetros via Liu e West (2001) e o método de filtragem de Gordon et al. (1993).	85
5.8	Estimativas para os percentis do parâmetro σ^2 no modelo Binomial utilizando o processo de estimação dos parâmetros via Liu e West (2001) e o método de filtragem de Gordon et al. (1993). A transformação escolhida para esse parâmetro foi a logaritmica.	85
5.9	Estimativas para os percentis dos estados latentes $x_{1:100}$ no modelo Binomial utilizando o processo de estimação dos parâmetros via Liu e West (2001) e o método de filtragem de Gordon et al. (1993).	86
5.10	Estimativas para os percentis do parâmetro α no modelo Binomial utilizando o processo de estimação dos parâmetros via Storvik (2002) e o método de filtragem de Pitt e Shephard (1999).	86
5.11	Estimativas para os percentis do parâmetro β no modelo Binomial utilizando o processo de estimação dos parâmetros via Storvik (2002) e o método de filtragem de Pitt e Shephard (1999).	87

5.12	Estimativas para os percentis do parâmetro σ^2 no modelo Binomial utilizando o processo de estimação dos parâmetros via Storvik (2002) e o método de filtragem de Pitt e Shephard (1999). A transformação escolhida para esse parâmetro foi a logaritmica.	87
5.13	Estimativas para os percentis dos estados latentes $x_{1:100}$ no modelo Binomial utilizando o processo de estimação dos parâmetros via Storvik (2002) e o método de filtragem de Pitt e Shephard (1999).	88
5.14	Estimativas para os percentis do parâmetro α no modelo Binomial utilizando o processo de estimação dos parâmetros via Liu e West (2001) e o método de filtragem de Pitt e Shephard (1999).	88
5.15	Estimativas para os percentis do parâmetro β no modelo Binomial utilizando o processo de estimação dos parâmetros via Liu e West (2001) e o método de filtragem de Pitt e Shephard (1999).	89
5.16	Estimativas para os percentis do parâmetro σ^2 no modelo Binomial utilizando o processo de estimação dos parâmetros via Liu e West (2001) e o método de filtragem de Pitt e Shephard (1999). A transformação escolhida para esse parâmetro foi a logaritmica.	89
5.17	Estimativas para os percentis dos estados latentes $x_{1:100}$ no modelo Binomial utilizando o processo de estimação dos parâmetros via Liu e West (2001) e o método de filtragem de Pitt e Shephard (1999).	90

Sumário

Resumo	iv
Abstract	v
1 Introdução	1
2 Inferência bayesiana em modelos Markovianos	3
2.1 Introdução	3
2.2 Modelos de Espaço de Estados	3
2.3 Modelos Dinâmicos Lineares	6
2.4 Filtragem	8
2.5 Predição	9
2.6 Recursões do tipo FFBS	9
2.7 Suavização	11
2.8 Modelos Dinâmicos Lineares Gaussianos - MDLG	12
2.8.1 Filtragem	13
2.8.2 Suavização	14
3 O Método de Monte Carlo	17
3.1 Amostragem de Importância	20

3.2	Amostragem de Importância Sequencial	24
3.3	Reamostragem (SIR)	26
4	Filtro de Partículas	28
4.1	Filtro de partículas em geral	32
4.1.1	Reamostragem	37
4.2	Filtro de Partículas Bootstrap (Gordon et al., 1993)	38
4.2.1	Introdução	39
4.2.2	Estimador Bayesiano Recursivo	39
4.2.3	FP Bootstrap Bayesiano Gordon et al. (1993)	43
4.2.4	Descrição Matemática	43
4.2.5	Algoritmo - FB Bootstrap	46
4.2.6	Exemplo de aplicação do FB Bootstrap	47
4.3	FP Auxiliar (Pitt e Shephard, 1999)	48
4.3.1	Descrição matemática	49
4.3.2	SIR	51
4.3.3	SIR Adaptativo	53
4.3.4	SIR Auxiliar	54
4.3.5	Algoritmo relativo ao FP auxiliar	57
4.3.6	Exemplos de aplicação do FP auxiliar	57
4.4	Filtro de Partículas Básico	60
4.4.1	FP Simples	64
4.4.2	FP ótimo	65

4.5	Suavização	69
5	Estimação de parâmetros	71
5.1	Liu e West (2001)	71
5.2	Storvik (2002)	74
5.3	Exemplo de aplicação do método de Storvik (2002)	77
5.4	Comparando os métodos	79
6	Considerações finais	91
6.1	Sumário Filtro de Partículas	91
6.2	Conclusão	92
	Referências	94
	Apêndice	96
A	Código em R para estimação via Filtro de Partículas Gordon et al. (1993) e Liu e West (2001)	96
B	Código em R para estimação via Filtro de Partículas Pitt e Shephard (1999) e Liu e West (2001)	98
C	Código em R para estimação via Filtro de Partículas Gordon et al. (1993) e Storvik (2002)	100
D	Código em R para estimação via Filtro de Partículas Pitt e Shephard (1999) e Storvik (2002)	102

Capítulo 1

Introdução

Os trabalhos em séries temporais nos fornecem inúmeros modelos. No entanto, ainda há espaço para aperfeiçoamentos e desenvolvimentos. Neste trabalho os esforços serão concentrados nos modelos de Espaço de Estados em uma abordagem via modelos dinâmicos com o uso de Filtro de Partículas (FP).

Os modelos que envolvem análise de dados temporais compõem um vasto campo de pesquisas, devido a necessidade de que se possa compreender e prever o comportamento dos dados temporalmente. O estudo da série temporal é feito com propostas de modelos parcimoniosos com boa acurácia e ajuste.

Os FP estão fortemente associados aos modelos dinâmicos e apresentam soluções para os entraves computacionais presentes nesses modelos quando as suposições de **normalidade** e **linearidade** são relaxadas.

As contribuições da técnica de FP são datadas da década de 90, com o aprimoramento e desenvolvimento de máquinas com alta performance computacional. A partir de Gordon et al. (1993), ainda sem a nomenclatura de FP, inúmeros autores apresentaram desenvolvimentos consideráveis para tal técnica como apresentado em Pitt e Shephard (1999).

Devido à complexidade do tema e importância crescente na área dos modelos dinâmicos, esta dissertação tem como objetivo fazer um estudo dos filtros de partículas descrevendo, com detalhes, algumas das principais ideias propostas nos artigos seminais e que, até o momento, fazem parte da literatura mais citada na área, sendo alvo de tentativas de aprimoramento nas pesquisas atualmente.

Com o intuito de compreender a técnica FP aplicado ao estudo de séries temporais não-gaussianas, objetiva-se comparar as inferências obtidas por métodos via modelos dinâmicos lineares generalizados com as obtidas por meio de algumas técnicas de FP.

Esta dissertação está organizada da seguinte forma: no capítulo 2, introduz-se os modelos de Espaço de Estados e os modelos dinâmicos lineares gaussianos. No capítulo 3, descreve-se o método Monte Carlo sequencial. No capítulo 4, descreve-se alguns tipos de FP. No capítulo 5, são apresentadas formas de estimação do vetor de parâmetros desconhecidos e são comparados os resultados dos filtros estudados. No capítulo 6 são apresentadas as considerações finais desta dissertação.

Capítulo 2

Inferência bayesiana em modelos Markovianos

2.1 Introdução

Uma série temporal é um conjunto de observações avaliadas ordenadamente em um período, podendo ser igualmente espaçadas e ter correlação serial. Os modelos de séries temporais buscam, entre outras coisas, entender essa correlação temporal e quantificá-la.

O uso de modelos de séries temporais convencionais para tal finalidade pode representar inúmeras inconsistências devido a problemas estruturais e fragilidade da técnica. Os modelos de espaço de estado e, em especial, os modelos dinâmicos (West e Harrison, 1997) são boas alternativas.

2.2 Modelos de Espaço de Estados

Um **Processo Estocástico** é definido por um conjunto de valores observados ao longo de instantes no tempo, podendo ser representados por $(\theta_t)_{t>0}$, sendo $t = 1, \dots, N$ o indexador que representa o instante de mensuração da quantidade (θ_t) .

Considere, ainda, o vetor de observações entre os tempos $t = 1, \dots, N$ representados por $\theta_{1:N}$.

O **Espaço de Estados** é o conjunto de valores que tal processo estocástico pode assumir. Tais espaços são classificados por:

- Discreto (cadeia): θ_t assume valores em um conjunto enumerável ou finito.
- Contínuo (sequência): θ_t assume valores em um conjunto não-enumerável.

Com relação ao **Tempo**, tem-se as seguintes classificações:

- Discreto: t assume valores em um conjunto enumerável/discreto.
- Contínuo: t assume valores em um conjunto não-enumerável.

Em geral, a sequência temporal dos dados de um Processo Estocástico viola os principais pressupostos das técnicas de análise estatística, que são a **independência** e **permutabilidade**.

Uma das estruturas de dependência imposta aos Processos Estocásticos é a de um **Processo Markoviano**. Essa estrutura considera que as informações atuais dependem das informações passadas exclusivamente pela informação imediatamente anterior. Nesse caso, tem-se um processo envolvendo “perda de memória” devido ao “esquecimento” das informações do passado que não sejam as do período imediatamente anterior.

Um Processo Markoviano é uma **Cadeia de Markov** quando θ_t é definida em um Espaço de Estados discreto, e é uma **Cadeia de Markov em Tempo discreto** quando t também é discreto.

Considere uma sequência, **não observável**, θ_t que possua a estrutura de dependência de um Processo Markoviano. Tal sequência servirá para auxiliar a descrição e o estudo de y_t , uma quantidade **observável** do sistema temporal. Dessa forma, o modelo de Espaços de Estados é composto pela sequência não observável $\{\theta_t, t = 1, \dots, n\}$ e pela parte observável $\{y_t, t = 1, \dots, n\}$. Considere $y_{1:t}$ a representação do vetor da sequência $\{y_1, y_2, \dots, y_t\}$ e os seguintes pressupostos para os modelos a serem apresentados:

A.1: $\theta_t, t = 1, \dots, n$ a representação de variáveis aleatórias de um Processo Markoviano;

A.2: Cada uma das observações do vetor $y_{1:t}$ são condicionalmente independentes dado $\theta_t, t = 1, \dots, n$.

Essas propriedades permitem descrever a **distribuição conjunta** das observações e dos estados como o produto das seguintes **distribuições condicionais**:

$$\pi(y_{1:n}, \theta_{0:n}) = \pi(\theta_0) \prod_{k=1}^n \pi(\theta_k | \theta_{k-1}) \pi(y_k | \theta_k) \quad (2.1)$$

A figura 2.1 representa o esquema de dependência do sistema de estados e observações definidos.

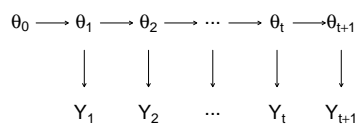


Figura 2.1: Modelo de Espaço de Estados.

O modelo de espaços de estados é composto pela **Equação de Observação** e a **Equação do Sistema** e, de acordo com as propriedades A.1 e A.2 definidas acima, tem-se, respectivamente:

$$y_t \sim g(\cdot|\theta_t) \quad (2.2)$$

$$\theta_t \sim f(\cdot|\theta_{t-1}) \quad (2.3)$$

Dessa forma, a equação (2.2) modela a série temporal y_t , que representa a parte manifesta do processo. Essa série de dados está sujeita a perturbações de um Processo Markoviano θ_t , por meio da equação (2.3). Essas funções também podem ser afetadas por **perturbações aleatórias**, que representam a parte não determinística do modelo.

As funções $g(\cdot)$ e $f(\cdot)$ descrevem os modelos por meio de formas muito gerais. Isso inclui formas não lineares, além de erros (perturbações) não aditivos e presença de covariáveis regressoras. Essas características também permitem considerar tal processo temporal como um **Modelo Dinâmico**.

2.3 Modelos Dinâmicos Lineares

Uma importante classe de modelos de Espaço de Estados são os **Modelos Dinâmicos Lineares** (MDL), que são caracterizados pela linearidade das funções $f(\cdot)$ e $g(\cdot)$ e por erros aditivos.

A função $g(\cdot)$, que modela a parte observável, é uma transformação linear dos estados latentes θ_t e uma perturbação aleatória v_t , que não possui restrição quanto

a distribuição de probabilidade a ele associada. Dessa forma, tem-se a seguinte

Equação das Observações:

$$y_t = F_t \theta_t + v_t. \quad (2.4)$$

O modelo (2.4) pode ser interpretado como uma regressão linear para a “resposta” y_t que é descrita pela matriz de planejamento F_t e um vetor de parâmetros θ_t . No entanto, θ_t é um vetor não estático. No caso de $\theta_t = \theta, \forall t$, tem-se um modelo de regressão usual.

Considere, ainda, a **Equação dos Estados** ser representada por:

$$\theta_t = G_t \theta_{t-1} + w_t. \quad (2.5)$$

A matriz G_t é conhecida como matriz de transição entre os estado θ_{t-1} e θ_t . Além disso, w_t representa a parte aleatória, não determinística, dessa transição temporal, e também não possui restrição quanto à distribuição de probabilidade a ele associada.

O interesse nos MDL é obter estimativas para a distribuição a *posteriori* dos estados, $\pi(\theta_s | D_t = y_1, \dots, y_t)$. A determinação do índice temporal s com relação a t determina o tipo de estimativa feita e podem ser determinadas em três situações, a saber:

- $s = t$, **processo de filtragem**: a cada nova informação, atualiza-se a distribuição de interesse;
- $s > t$, **processo de predição**: as informações atuais são utilizadas para projetar a distribuição de interesse;

- $s < t$, **processo de suavização**: é o processo de reestimação dos parâmetros obtidos na filtragem, utilizando todas as informações já observadas.

2.4 Filtragem

O processo de filtragem para os MDL consiste em obter estimativas para a distribuição a *posteriori* do processo latente iterativamente, à medida que se tem acesso a uma nova observação. Tal processo é conhecido como filtro *online*. A estrutura Markoviana presente nos modelos de Espaço de Estados auxilia o processo iterativo de estimação da distribuição de interesse $\pi(\theta_t|\cdot)$.

Considerando uma distribuição inicial $\theta_0 \sim \pi(\theta_0)$ e as propriedades A.1 e A.2 e o resultado da equação (2.1), apresentadas na Seção 2.2, o processo de filtragem pode ser resumido nas seguintes etapas:

- (i) **Predição do estado** θ_t condicionalmente às informações D_{t-1} até o tempo $t - 1$, isto é:

$$\pi(\theta_t|D_{t-1}) = \int \pi(\theta_t|\theta_{t-1})\pi(\theta_{t-1}|D_{t-1})d\theta_{t-1}. \quad (2.6)$$

- (ii) **Predição a um passo a frente** de y_t dado a informação até o tempo $t - 1$:

$$\pi(y_t|D_{t-1}) = \int \pi(y_t|\theta_t)\pi(\theta_t|D_{t-1})d\theta_t. \quad (2.7)$$

- (iii) **Atualização** da distribuição a *posteriori* $\pi(\theta_t|D_t)$:

$$\pi(\theta_t|D_t) = \frac{\pi(y_t|\theta_t)\pi(\theta_t|D_{t-1})}{\pi(y_t|D_t)}. \quad (2.8)$$

Dessa forma, tem-se:

$$\pi(\theta_t|D_t) \propto \pi(y_t|\theta_t)\pi(\theta_t|D_{t-1}). \quad (2.9)$$

2.5 Predição

Há ainda o interesse em fazer predições k -passos a frente condicionado às informações das observações e do estado latente até o tempo t . Esse processo de estimação é feito também em dois passos. Primeiramente, com a evolução da equação do estado latente, tem-se:

$$\pi(\theta_{t+k}|D_t) = \int \pi(\theta_{t+k}|\theta_{t+k-1})\pi(\theta_{t+k-1}|D_t)d\theta_{t+k-1} \quad (2.10)$$

Após essa etapa, pode-se obter estimativas para a observação futura y_{t+k} por meio da equação de observação:

$$\pi(y_{t+k}|D_t) = \int \pi(y_{t+k}|\theta_{t+k})\pi(\theta_{t+k}|D_t)d\theta_{t+k}. \quad (2.11)$$

2.6 Recursões do tipo FFBS

Considere a sigla FFBS para designar o processo de cálculo da distribuição a *posteriori* via método de *Forward Filtering and Backward Sampling*. Objetiva-se estimar tal distribuição a *posteriori* conjunta dos estados latentes por meio de métodos de simulação. Para tanto, considere a decomposição recursiva:

$$\pi(\theta_1, \dots, \theta_T | D_T) = \pi(\theta_1, \dots, \theta_T | y_1, \dots, y_T) \quad (2.12)$$

$$= \pi(\theta_T | D_T) \prod_{k=1}^{T-1} \pi(\theta_k | \theta_{k+1}, D_T). \quad (2.13)$$

Utilizando as propriedades A.1 e A.2, tem-se que:

$$\pi(\theta_1, \dots, \theta_T | D_T) = \pi(\theta_T | D_T) \prod_{k=1}^{T-1} \pi(\theta_k | \theta_{k+1}, D_k), \quad (2.14)$$

sendo,

$$\pi(\theta_k | \theta_{k+1}, D_k) = \frac{\pi(\theta_{k+1} | \pi(\theta_k) \pi(\theta_k | D_k))}{\pi(\theta_{k+1} | D_k)}. \quad (2.15)$$

Por meio do cálculo de $\pi(\theta_1, \dots, \theta_T | D_T)$ é possível obter as distribuições marginais $\{\pi(\theta_k | D_T)\}$, uma vez que:

$$\begin{aligned} \pi(\theta_k | D_T) &= \int \pi(\theta_k, \theta_{k+1} | D_T) d\theta_{k+1} \\ &= \int \pi(\theta_k | \theta_{k+1}, D_T) \pi(\theta_{k+1} | D_T) d\theta_{k+1} \\ &= \int \pi(\theta_k | \theta_{k+1}, D_k) \pi(\theta_{k+1} | D_T) d\theta_{k+1} \\ &= \pi(\theta_k | D_k) \int \frac{\pi(\theta_{k+1} | \theta_k)}{\pi(\theta_{k+1} | D_k)} \pi(\theta_{k+1} | D_T) d\theta_{k+1} \end{aligned}$$

A equação (2.14) sugere o seguinte algoritmo para obter amostras utilizando $\pi(\theta_1, \dots, \theta_T | D_T)$, segundo o FFBS:

- (1) Obtenha as distribuições filtradas $\pi(\theta_k | D_k)$
- (2) Amostre $\theta_T \sim \pi(\theta_T | D_T)$

(3) Para $k = T - 1, T - 2, \dots, 1$, amostra θ_k a partir de $\pi(\theta_k|\theta_{k+1}, D_k)$, isto é:

$$\theta_k \sim \pi(\theta_k|\theta_{k+1}, D_k)$$

onde $\pi(\theta_k|\theta_{k+1}, D_k)$ é dada pela equação (2.15).

2.7 Suavização

O processo de suavização reestima o processo latente (θ) utilizando as estimativas fornecidas pelo processo de filtragem. Então θ_t é reestimado utilizando média a *posteriori* no tempo t . Considere $T \geq t$, tem-se que:

$$\begin{aligned} E(\theta_t|D_T) &= \int \left[\int \theta_t \pi(\theta_t, \theta_{t+1}|D_T) d\theta_{t+1} \right] d\theta_t \\ &= \int \theta_t \left[\int \pi(\theta_{t+1}|D_T) \pi(\theta_t|\theta_{t+1}, D_T) d\theta_{t+1} \right] d\theta_t \\ &= \int \pi(\theta_{t+1}|D_T) \left[\int \theta_t \pi(\theta_t|\theta_{t+1}, D_T) d\theta_t \right] d\theta_{t+1} \\ &= \int \pi(\theta_{t+1}|D_T) E(\theta_t|\theta_{t+1}, D_T) d\theta_{t+1} \\ &= \int E(\theta_t|\theta_{t+1}, D_T) \pi(\theta_{t+1}|D_T) d\theta_{t+1} \\ &= E[E(\theta_t|\theta_{t+1}, D_T)|D_T]. \end{aligned}$$

O valor esperado da parte interna de (2.16) pode ser solucionado utilizando as propriedades A.1 e A.2 de Markov (vide seção 2.2), isto é, $\{y_{t+1}, \dots, y_T\}$ são condicionalmente independentes a θ_t , dado θ_{t+1} .

$$\begin{aligned} \pi(\theta_t, \theta_{t+1}|D_T) &= \pi(\theta_t|\theta_{t+1}, D_T) \pi(\theta_{t+1}|D_T) \\ &= \pi(\theta_t|\theta_{t+1}, D_t) \pi(\theta_{t+1}|D_T). \end{aligned}$$

Portanto,

$$\pi(\theta_t, \theta_{t+1}|D_T) = \pi(\theta_t, \theta_{t+1}|D_t).$$

Dessa forma, temos o seguinte resultado para média suavizada:

$$E(\theta_t|D_T) = E [E(\theta_t|\theta_{t+1}, D_t)|D_T]$$

Utilizando as mesmas propriedades Markovianas e propriedades de variância condicional (Ross, 2007), tem-se que a equação da variância suavizada é dada por:

$$V(\theta_t|D_T) = V [E(\theta_t|\theta_{t+1}, D_t)|D_T] + E [V(\theta_t|\theta_{t+1}, D_t)|D_T]$$

2.8 Modelos Dinâmicos Lineares Gaussianos - MDLG

Caracteriza-se como MDLG's os MDL's que possuem as variáveis de perturbação com distribuição Gaussiana. Em tais modelos, as integrais apresentadas na Seção 2.3 têm forma fechada. Isso se deve às propriedades da distribuição Normal Multivariada (Petris et al., 2009), seção 1.5.

Para tal modelo, considere que a função $f(\cdot)$, que modela a parte observável y_t , é composta por uma relação linear entre os estados latentes θ_t (unidimensional) e uma perturbação aleatória v_t com distribuição normal com variância igual a $V = 1$, isto é, $v_t \sim N(0, 1)$. Considere ainda a matriz de delineamento $F_t = 1$. Assim, tem-se que a equação das observações pode ser descrita por:

$$y_t = \theta_t + v_t \Rightarrow y_t \sim N(\theta_t, 1)$$

Para a equação dos estados, considere a matriz de delineamento que evolui dos estados latentes θ_{t-1} para o θ_t sendo $G_t = 1$. Considere ainda que, assim como na equação das observações, a perturbação aleatória w_t compõe a equação dos estados

de forma aditiva, sendo $w_t \sim N(0, 1)$.

$$\theta_t = \theta_{t-1} + w_t \Rightarrow \theta_t \sim N(\theta_{t-1}, 1)$$

A figura 2.2 representa uma simulação do modelo descrito.

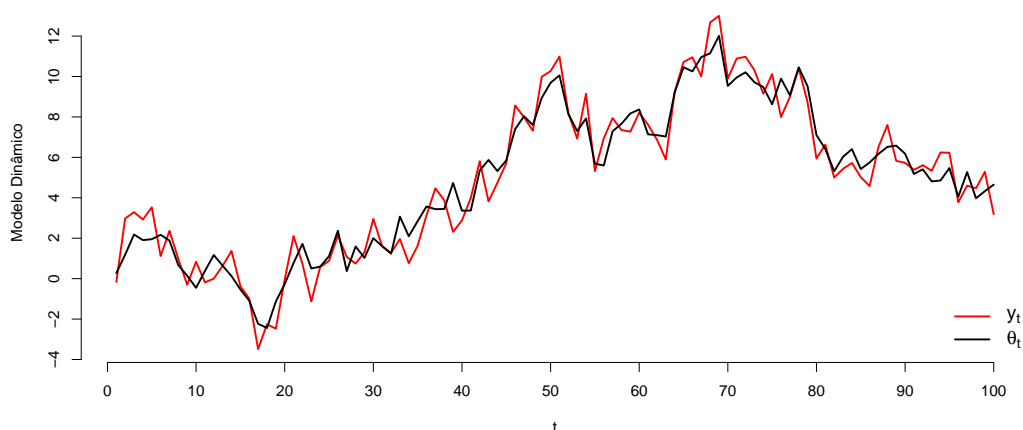


Figura 2.2: Simulação para um Modelo Dinâmico Linear Gaussiano

2.8.1 Filtragem

Sejam v_t e w_t perturbações com distribuição Gaussiana independentes com média zero e variâncias V e W , respectivamente. West e Harrison (1997) apresentam um esquema sequencial para o processo de filtragem, conhecido como **Filtro de Kalman**.

Considere ainda a distribuição inicial da cadeia com *priori* $(\theta_0|D_0) \sim N[m_0, C_0]$.

Dessa forma, o processo de filtragem é obtido sequencialmente por:

- (a) **Posteriori de θ em $t - 1$** : $(\theta_{t-1}|D_{t-1}) \sim N[m_{t-1}, C_{t-1}]$;
- (b) **Atualização da *priori* em $t - 1$** : $(\theta_t|D_{t-1}) \sim N[m_{t-1}, R_t]$;

(c) **Predição para y em t :** $(y_t|D_{t-1}) \sim N[f_t, Q_t]$;

(d) **Posteriori de θ em t :** $(\theta_t|D_t) \sim N[m_t, C_t]$;

Sendo: $R_t = C_{t-1} + W_t$, $f_t = m_{t-1}$, $Q_t = R_t + V_t$, $\theta_t = m_{t-1} + A_t e_t$, $C_t = A_t V_t$,
 $A_t = R_t/Q_t$ e $e_t = Y_t - f_t$.

Essa rotina de atualização e predição é repetida a cada nova observação disponível. A figura 2.3 apresenta o resultado do processo de filtragem descrito aplicado ao modelo proposto na seção 2.8, isto é $V_t = 1$ e $W_t = 1$.

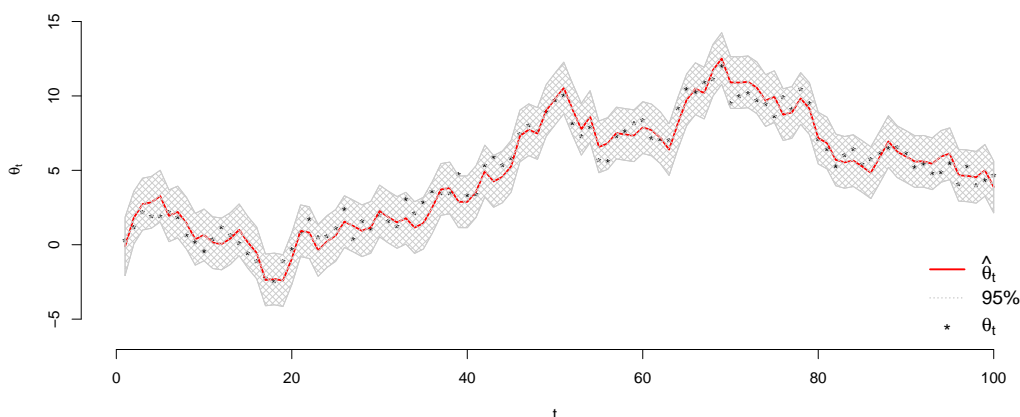


Figura 2.3: Estimativas dos estados latentes pelo Filtro de Kalman

2.8.2 Suavização

Considere o processo de suavização para encontrar a distribuição de $\pi(\theta_t|y_{1:T})$. Utilizando as mesmas representações descritas no processo de filtragem, o processo de suavização aplicado ao Modelo Dinâmico Linear Gaussiano descrito tem as estimativas suavizadas dos estados latentes, representados com m_t^s , obtidas por:

$$m_t^s = m_t + \frac{C_t}{R_t}(m_{t+1}^s - a_{t+1}),$$

sendo m_t^s a estimativa suavizada do estado latente obtida no tempo $t + 1$.

As estimativas suavizadas para a variância da distribuição *a posteriori* dos estados latentes, representadas por C_t^s , são calculada por meio da seguinte equação:

$$C_t^s = C_t + \left(\frac{C_t}{R_t} \right)^2 (R_t - C_{t+1}^s) \quad (2.16)$$

sendo C_t^s a estimativa suavizada da variância do estado latente obtida no tempo $t + 1$.

A figura 2.4 apresenta o resultado do processo de filtragem descrito aplicado ao modelo proposto na seção 2.8, isto é $V_t = 1$ e $W_t = 1$.

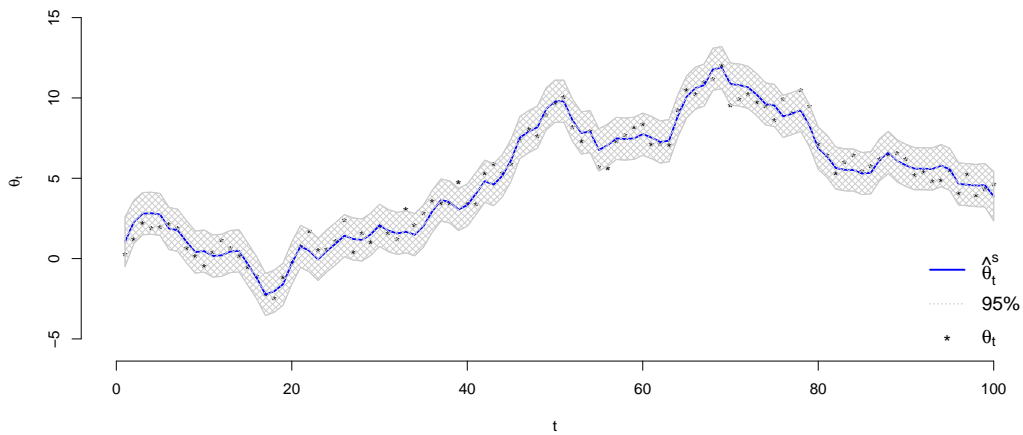


Figura 2.4: Estimativas suavizadas dos estados latentes pelo Filtro de Kalman

As estimativas suavizadas possuem uma maior precisão para às estimativas dos estados latentes. Tal precisão é aumentada por essa metodologia utilizar o conhecimento de todas as amostras, $y_{1:T}$. A figura 2.5 compara as estimativas pontuais e intervalares dos estados latentes. A região em cinza representa o intervalo das estimativas intervalares no processo *on-line* de filtragem. Comparada à região em

azul claro, associada as estimativas do processo de suavização, pode se perceber uma menor precisão do processo de filtragem.

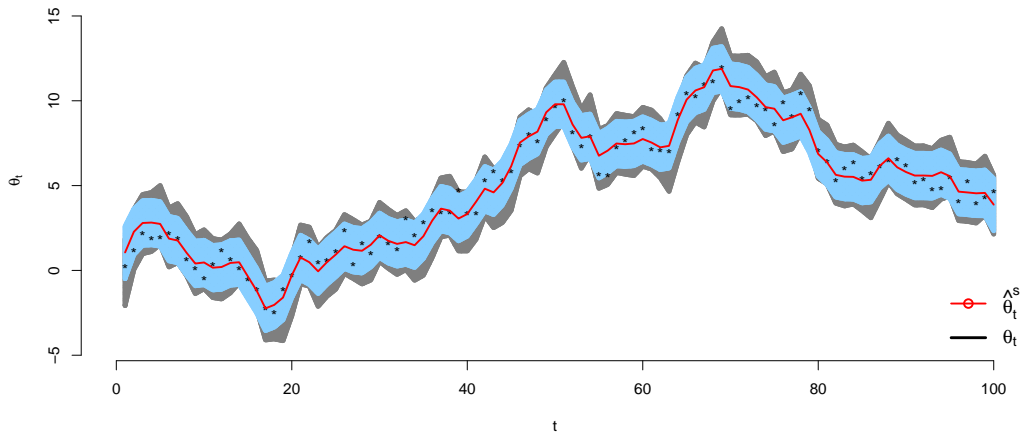


Figura 2.5: Comparação entre as estimativas suavizadas e filtradas

Mais detalhes sobre tal técnica de suavização e filtragem podem ser encontrados em West e Harrison (1997).

Capítulo 3

O Método de Monte Carlo

O método de Monte Carlo (MC) representa uma solução numérica para situações em que não se pode obter de forma analítica os valores desejados. Neste capítulo serão descritas algumas das técnicas associadas à metodologia de **MC**. Tais técnicas serão úteis para o entendimento do FP a ser estudado no capítulo 4.

Por meio do método de MC é possível estimar uma integral de forma numérica, considerando-se propriedades de primeira ordem de convergência em probabilidade. Considere o cálculo da seguinte integral:

$$\pi = \int_a^b f(\theta) d\theta. \quad (3.1)$$

Seja θ uma variável aleatória com distribuição **Uniforme**(a, b), isto é, $g_u(\theta) = \frac{1}{b-a}$. É possível reescrever a equação (3.1) como:

$$\pi = \int_a^b (b-a)f(\theta) \frac{1}{(b-a)} d\theta = (b-a) \int_a^b f(\theta) g_u(\theta) d\theta.$$

Dessa forma, a integral de interesse pode ser vista como o valor **esperado** da variável aleatória θ na função $f(\cdot)$:

$$\pi = (b-a)E_U[f(\theta)].$$

Para concluir o método, pode-se gerar N amostras de $\theta^i \sim g_u(a, b)$, $i = 1, \dots, N$, e avaliá-las em $f(\cdot)$. Dessa forma, a estimativa para π de forma numérica é dada por:

$$\hat{\pi} = (b - a) \sum_{i=1}^N \frac{f(\theta^i)}{N}.$$

É fácil notar que essa aproximação é não viesada pois:

$$E(\hat{\pi}) = \frac{(b - a)}{N} \sum_i^N E[f(\theta^i)].$$

Dessa forma, o valor esperado é:

$$E(\hat{\pi}) = \frac{(b - a)}{N} N \int_a^b \frac{f(\theta)}{(b - a)} d\theta = \int_a^b f(\theta) d\theta = \pi.$$

Com isso, tem-se que:

$$E(\hat{\pi}) = \int f(\theta) d\theta.$$

Por meio da **Lei Forte do Grandes Números**, é possível dizer que $\hat{\pi}$ converge quase certamente para π . A variância do estimador é dada por:

$$V_{\hat{\pi}} = \frac{1}{N^2} \sum_i^N [f(\theta^i) - \hat{\pi}]^2.$$

Utilizando o Teorema do Limite Central, para um N suficientemente grande, tem-se:

$$\frac{\hat{\pi} - E[f(\theta)]}{\sqrt{V_{\hat{\pi}}}} \sim N(0, 1).$$

Esse resultado mostra que o estimador π é **consistente**, pois $V_{\hat{\pi}}$ converge para zero, quando N cresce.

O uso da técnica de MC na solução de integrações é feita por meio do cálculo de uma soma ponderada de funções avaliadas em uma amostra simulada a partir

da distribuição de interesse. Considere, inicialmente, o problema da aproximação de uma densidade de probabilidade genérica $\pi_n(\theta_1, \dots, \theta_n)$ para n fixo, o qual será tratado como a **distribuição alvo de interesse**.

Na obtenção da distribuição aproximada de $\pi_n(\theta_1, \dots, \theta_n)$, gera-se N amostras independentes do vetor aleatório $\Theta_{1:n}^i \sim \pi_n(\theta_1, \dots, \theta_n)$, $i = 1, \dots, N$. Então, a aproximação de MC de $\pi_n(\theta_1, \dots, \theta_n)$ é dada por:

$$\hat{\pi}_n(\theta_1, \dots, \theta_n) = \frac{1}{N} \sum_i^N \delta_{\Theta_{1:n}^i}(\theta_{1:n}), \quad (3.2)$$

sendo $\delta_{\Theta_{1:n}^i}(\theta_{1:n})$ uma função indicadora de massa no ponto $\theta_{1:n}$. Baseado na equação (3.2), é possível obter aproximações para qualquer distribuição marginal:

$$\hat{\pi}_n(\theta_k) = \frac{1}{N} \sum_i^N \delta_{\Theta_k^i}(\theta_k). \quad (3.3)$$

De forma análoga, o valor esperado de uma função ϕ_n de $\Theta = (\theta_1, \dots, \theta_n)$, dada por:

$$E(\phi_n) = I_n(\phi_n) = \int \phi_n(\theta_1, \dots, \theta_n) \pi_n(\theta_1, \dots, \theta_n) d\theta_{1:n},$$

é aproximada por,

$$\hat{E}(\phi_n) = I_n^{MC}(\phi_n) = \int \phi_n(\theta_1, \dots, \theta_n) \hat{\pi}_n(\theta_1, \dots, \theta_n) d\theta_{1:n} = \frac{1}{N} \sum_{i=1}^N \phi_n(\Theta_{1:n}^i).$$

Verifica-se que $I_n^{MC}(\phi_n)$ é estimador não-viciado de $I_n(\phi_n)$:

$$E[I_n^{MC}(\phi_n)] = E \left[\frac{1}{N} \sum_{i=1}^N \phi_n(\Theta_{1:n}^i) \right] = E[\phi_n(\Theta_{1:n})] = I_n(\phi_n)$$

Além disso, a variância de $I_n^{MC}(\phi_n)$ é dada por:

$$V[I_n^{MC}(\phi_n)] = \frac{\sigma^2}{n},$$

sendo,

$$\sigma^2 = \left[\int \phi^2(\theta_1, \dots, \theta_n) \pi_n(\theta_1, \dots, \theta_n) d(\theta_1, \dots, \theta_n) \right] - [I_n(\phi_n)]^2.$$

A principal vantagem dos métodos de MC em relação aos outros é que a variância do erro nas aproximações decresce a uma taxa de $O(1/n)$ independentemente da dimensão do espaço \mathbb{R}^n . No entanto, há pelo menos dois grandes problemas com o método de MC básico:

- Problema 1: Se $\pi_n(\theta_1, \dots, \theta_n)$ é uma distribuição de probabilidade complexa e não é possível amostrar diretamente desta distribuição.
- Problema 2: Mesmo que soubéssemos exatamente como amostrar de $\pi_n(\theta_1, \dots, \theta_n)$, a complexidade computacional de tal esquema amostral é, tipicamente, pelo menos linear no número de variáveis, n . Dessa forma, um algoritmo com o qual amostre-se exatamente a partir de $\pi_n(\theta_1, \dots, \theta_n)$, de forma sequencial para cada vetor de n , teria complexidade computacional que cresce, pelo menos, linearmente com n .

3.1 Amostragem de Importância

De modo a abordar o problema 1, utiliza-se a **Amostragem de Importância**. Essa técnica é fundamental para todos os algoritmos a serem discutidos nesta dissertação.

A técnica de Amostragem de Importância baseia-se na utilização de uma **distribuição de importância**, $q(\cdot)$, de fácil amostragem e que seja definida no mesmo suporte da distribuição $\pi(\cdot)$. Desse modo,

$$E_{\pi}(\phi) = \int \phi(\theta)\pi(\theta)d\theta = \int \frac{\phi(\theta)\pi(\theta)}{q(\theta)}q(\theta)d\theta.$$

Nesse caso, as amostras são geradas a partir da distribuição $q(\cdot)$, e a quantidade $\frac{\pi(\theta)}{q(\theta)}$ fornece um ajuste da discrepância entre a distribuição geradora e a de interesse, $\pi(\cdot)$.

Considere que a distribuição-alvo em estudo é $\pi_n(\theta_{1:n})$, uma distribuição multi-variada que pode ser representada por:

$$\pi_n(\theta_{1:n}) = \frac{\gamma_n(\theta_{1:n})}{Z_n}, \quad (3.4)$$

sendo o denominador representado pela seguinte integral:

$$Z_n = \int \gamma_n(\theta_{1:n})d\theta_{1:n}. \quad (3.5)$$

Considere uma densidade de importância $q_n(\theta_{1:n})$ para auxiliar na estimação da distribuição de interesse, $\pi_n(\theta_{1:n})$. Considere ainda que $\pi_n(\theta_{1:n}) > 0$ implique em $q_n(\theta_{1:n}) > 0$. Das equações (3.4) e (3.5), tem-se as seguintes identidades, úteis na amostragem de importância:

$$\pi_n(\theta_{1:n}) = \frac{\gamma_n(\theta_{1:n})}{Z_n} = \frac{\gamma_n(\theta_{1:n})}{q_n(\theta_{1:n})} \frac{q_n(\theta_{1:n})}{Z_n}.$$

Desse modo,

$$\pi_n(\theta_{1:n}) = \frac{w_n(\theta_{1:n})q_n(\theta_{1:n})}{Z_n}, \quad (3.6)$$

em que

$$w_n(\theta_{1:n}) = \frac{\gamma_n(\theta_{1:n})}{q_n(\theta_{1:n})}, \quad (3.7)$$

e

$$Z_n = \int w_n(\theta_{1:n}) q_n(\theta_{1:n}) d\theta_{1:n}. \quad (3.8)$$

Em particular, seleciona-se uma densidade de importância $q_n(\theta_{1:n})$ que seja de fácil amostragem, por exemplo, uma densidade normal multivariada.

Considere N amostras independentes $(\theta_1, \dots, \theta_n)$ geradas a partir de $q_n(\theta_{1:n})$, isto é,

$$\Theta_{1:n}^i \sim q_n(\theta_{1:n}),$$

sendo $i = 1, \dots, N$.

Considerando a aproximação (3.3) para a densidade de importância $q_n(\theta_{1:n})$, tem-se:

$$\hat{q}_n(\theta_{1:n}) = \frac{1}{N} \sum_i^N \delta_{\Theta^i}(\theta_{1:n}),$$

que é a medida empírica das amostras $\Theta_{1:n}^i, i = 1, \dots, N$. Ao inserir esta aproximação de MC de $q_n(\theta_{1:n})$ nas equações (3.6) e (3.8), tem-se:

$$\hat{\pi}_n(\theta_{1:n}) = \frac{w_n(\theta_{1:n}) \hat{q}_n(\theta_{1:n})}{\int w_n(\theta_{1:n}) \hat{q}_n(\theta_{1:n}) d\theta_{1:n}}.$$

Utilizando as partículas/amostras para gerar a distribuição numérica para $\hat{q}_n(\theta_{1:n})$, tem-se:

$$\hat{\pi}_n(\theta_{1:n}) = \frac{\frac{1}{N} \sum_{i=1}^N w_n(\Theta_{1:n}^i) \delta_{\Theta_n^i}(\theta_{1:n})}{\int w_n(\Theta_{1:n}^i) \left[\frac{1}{N} \sum_{i=1}^N \delta_{\Theta_n^i}(\theta_{1:n}) \right] d\theta_{1:n}}.$$

Neste ponto, realiza-se a aproximação para $w_n(\Theta_{1:n}^i)$ utilizando as mesmas partículas/amostras geradas. Com isso, pode-se inverter a ordem do somatório e da integração para que esta contenha apenas os valores que dependam ou sejam funções da variável aleatória $\theta_{1:n}$. Desta forma, tem-se:

$$\hat{\pi}_n(\theta_{1:n}) = \frac{\frac{1}{N} \sum_{i=1}^N w_n(\Theta_{1:n}^i) \delta_{\Theta_n^i}(\theta_{1:n})}{\frac{1}{N} \sum_{i=1}^N w_n(\Theta_{1:n}^i) \int \delta_{\Theta_n^i}(\theta_{1:n}) d\theta_{1:n}}.$$

A quantidade $\int \delta_{\Theta_n^i}(\theta_{1:n}) d\theta_{1:n}$ denota a função δ de Dirac para avaliação da partícula gerada de Θ_n^i em todo espaço paramétrico de $\theta_{1:n}$. Com isso, tem-se que:

$$\int \delta_{\Theta_n^i}(\theta_{1:n}) d\theta_{1:n} = 1.$$

dessa forma, tem-se que:

$$\hat{Z}_n = \frac{1}{N} \sum_{i=1}^N w_n(\Theta_{1:n}^i).$$

Utilizando este resultado, pode-se reescrever a aproximação como:

$$\hat{\pi}_n(\theta_{1:n}) = \frac{\sum_{i=1}^N w_n(\Theta_{1:n}^i) \delta_{\Theta_{1:n}^i}(\theta_{1:n})}{\sum_{i=1}^N w_n(\Theta_{1:n}^i)}. \quad (3.9)$$

Considere W_n^i como o peso normalizado, isto é, $\sum_{i=1}^N W_n^i = 1$, calculado por:

$$W_n^i = \frac{w_n(\Theta_{1:n}^i)}{\sum_{j=1}^N w_n(\Theta_{1:n}^j)}.$$

A distribuição-alvo é aproximada por:

$$\hat{\pi}_n(\theta_{1:n}) = \sum_{i=1}^N W_n^i(\Theta_{1:n}^i) \delta_{\Theta_{1:n}^i}. \quad (3.10)$$

Esses pesos corrigem a discrepância entre a distribuição geradora da amostras e a densidade original na qual deseja-se avaliar a função $\gamma(\theta_{1:n})$. No contexto dos modelos de espaço de estados, a distribuição de interesse é $\pi(\cdot) = \pi(\theta_{0:t}|y_{1:t})$, o

valor da *posteriori* às observações do estado latente no tempo t . Além disso, há o interesse do cálculo dessa *posteriori* para os tempos $t = 1, \dots, N$ do sistema, a cada nova observação.

No contexto de filtragem, apresentado na seção 2.4, a distribuição-alvo é a distribuição a *posteriori* dos estados latentes, tal que:

$$\pi_n(\theta_{1:n}) = p(\theta_{1:n} | D_n = \{y_1, \dots, y_n\}). \quad (3.11)$$

Dessa forma, o numerador e o denominador da representação da *posteriori*, $p(\theta_{1:n} | D_n)$, apresentado pela equação (3.4), são, respectivamente:

$$\gamma_n(\theta_{1:n}) = \pi_n(\theta_{1:n}) = p(\theta_{1:n}, D_n), \quad (3.12)$$

e

$$Z_n = \int p(\theta_{1:n}, D_n) d\theta_{1:n} = p(D_n = \{y_1, \dots, y_n\}). \quad (3.13)$$

3.2 Amostragem de Importância Sequencial

A Amostragem de Importância se destaca pela facilidade de implementação e pela capacidade de solução de problemas de difícil tratamento analítico. Apesar da simplicidade desta técnica, em muitas situações, não é trivial obter amostras da distribuição $\pi(\cdot)$ devido à sua dimensão. Esse é o problema 2 presente na técnica de MC apresentado na seção 3.

Para contornar tal problema, utiliza-se a **Amostragem de Importância Sequencial**. Essa técnica propõe-se a obter as amostras desejadas de forma sequencial, com amostras unidimensionais das distribuições marginais de $\theta_{1:t}$ a cada tempo t .

Essa técnica faz parte do método de Monte Carlo Sequencial (MCS), que representa uma classe geral dos métodos de MC. Nele, amostra-se sequencialmente $\pi_n(\theta_{1:n})$, a distribuição de interesse, em $t = 1, \dots, n$. Seja essa distribuição definida no espaço produto \mathbb{R}^n , que possui dimensão que cresce com n .

Considere a mesma representação para a distribuição de interesse apresentada anteriormente:

$$\pi_n(\theta_{1:n}) = \frac{\gamma_n(\theta_{1:n})}{Z_n} \quad (3.14)$$

O método MCS aproxima a distribuição de interesse aplicando a técnica de amostragem de importância para cada uma das suas dimensões, que são avaliadas em $t = 1, \dots, n$. Considere $\pi_1(\theta_1)$ a representação da parte da distribuição de interesse no tempo $t = 1$ e $\pi_2(\theta_{1:2})$ em $t = 2$, e assim sucessivamente.

Analisando a equação dos pesos $w_{1:n}$ no tempo $t = n$, obtidas pelas relações algébricas da amostragem de importância para aproximar a distribuição-alvo, apresentadas nas equações (3.7) e (3.6), tem-se:

$$w(\theta_{1:n}) = \frac{\gamma(\theta_{1:n})}{q(\theta_{1:n-1})}.$$

Considere, ainda, que a distribuição de importância $q(\cdot)$ possa ser expressa da seguinte forma:

$$q(\theta_{1:n}) = q(\theta_{1:n-1})q(\theta_n|\theta_{1:n}). \quad (3.15)$$

Isto é,

$$q(\theta_{1:n}) = q(\theta_1) \prod_{k=2}^n q(\theta_k|\theta_{1:k-1}). \quad (3.16)$$

Dessa forma, tem-se que os pesos podem ser calculados de forma **sequencial** ao longo dos tempos $t = 1, \dots, n$. Considere como exemplo analisar os pesos $w_{1:2}$ no

tempo $t = 2$. Tais pesos podem ser escritos como:

$$w(\theta_{1:2}) = \frac{\gamma(\theta_{1:2})}{q(\theta_{1:2})} = \frac{\gamma(\theta_{1:2})}{q(\theta_1)q(\theta_2|\theta_1)}.$$

Isto é,

$$\begin{aligned} w(\theta_{1:2}) &= \frac{\gamma(\theta_{1:2})}{q(\theta_{1:2})} = \frac{\gamma(\theta_{1:2})}{q(\theta_1)q(\theta_2|\theta_1)} \\ &= \frac{\gamma(\theta_1)}{q(\theta_1)} \frac{\gamma(\theta_{1:2})}{\gamma(\theta_1)q(\theta_2|\theta_1)} \\ &= w_1(\theta_1) \frac{\gamma(\theta_{1:2})}{\gamma(\theta_1)q(\theta_2|\theta_1)}. \end{aligned}$$

Por indução, é possível chegar ao resultado para qualquer $t > 1$. Dessa forma, verifica-se que os pesos são atualizados iterativamente a cada instante, por meio da expressão:

$$w_n(\theta_{1:n}) = w_{n-1}(\theta_{1:n-1}) \frac{\gamma(\theta_{1:n})}{\gamma(\theta_{1:n-1})q(\theta_n|\theta_{n-1})}.$$

3.3 Reamostragem (SIR)

A técnica de **SIR**, do Inglês *Sampling Importance Resampling*, foi desenvolvida por Rubin (1992) e é um desdobramento da amostragem de importância. Deseja-se aproximar uma função $g(\cdot)$ da *posteriori* $\pi(\theta|y)$ utilizando n amostras $\theta_{1:n}$ de uma distribuição auxiliar $q(\cdot)$. No passo seguinte, tais valores são reamostrados proporcionalmente à verossimilhança:

$$w_i = \frac{p(\theta_i|y)/q(\theta_i)}{\sum_j p(\theta_j|y)/q(\theta_j)}, i = 1, \dots, n$$

No entanto, a distribuição de $p(\theta_i|y)$ nem sempre é conhecida para efetuar os cálculos. Com isso, utiliza-se a seguinte aproximação:

$$w_i = \frac{p(y|\theta_i)p(\theta_i)/Cq(\theta_i)}{\sum_j p(y|\theta_j)p(\theta_j)/Cq(\theta_j)}$$

sendo C , a constante normalizadora. Nesse caso, quando a distribuição de amostragem auxiliar é definida igual à *priori* $p(\theta_i)$ e anulando a constante normalizadora, tem-se:

$$w_i = \frac{p(y|\theta_i)}{\sum_j p(y|\theta_j)}.$$

A relação entre $p(\theta_i)$ e $q(\theta_i)$ definirá os tipos de filtros de partículas selecionados para o processo de estimação de θ_i .

Capítulo 4

Filtro de Partículas

Neste capítulo serão discutidas algumas metodologias propostas em artigos seminais na área de Filtro de Partículas. A notação utilizada nos mesmos apresenta discrepâncias. No entanto, pode-se considerar que estas limitações não interferem na compreensão do problema.

Para iniciar este capítulo é interessante discorrer sobre a definição do termo **partícula**. Tal definição auxiliará a compreensão da técnica abordada nessa dissertação.

Na física, define-se partícula subatômica a designação genérica de algo cuja dimensão é muito menor do que a de um átomo (o todo). No contexto a ser abordado, além de existir o todo (átomo) e uma parte dele (partícula), há outra parte do todo, desconhecida e que se deseja estimar.

No contexto estatístico, a partícula pode ser considerada como uma realização de um experimento, na tentativa de estimar características desconhecidas (parâmetros), e cada realização do experimento designa um valor hipotético (possível) para a característica a ser estimada. Em resumo, **partículas são realizações de um experimento cujos valores possíveis estão definidos no espaço paramétrico**

do valor desconhecido.

No contexto de espaços de estados, o processo de filtragem (vide seção 2.2) utilizando conceitos de partículas é denominado **Filtro de Partículas**.

Considere um exemplo de aplicação do Filtro de Partículas na área da robótica, (<http://www.aiqus.com/questions/25739/particle-filter>). O problema é encontrar a posição de um robô, representado pelo ponto azul, que se movimenta por uma sala limitada por paredes, representadas pela parte cinza (vide figura 4.1) . As demais representações na sala, nas cores verde e marrom, são outros obstáculos para o robô. Dessa forma, o robô pode se movimentar por qualquer parte na cor branca da sala.

Objetiva-se estimar a posição do robô utilizando-se **partículas** para representar localizações hipotéticas do robô na parte branca do desenho da sala. A figura 4.2 ilustra disposição das partículas geradas (pontos em rosa) ao longo das possíveis localizações. Milhares de valores (partículas) podem ser gerados de modo a “varrer” o espaço paramétrico.

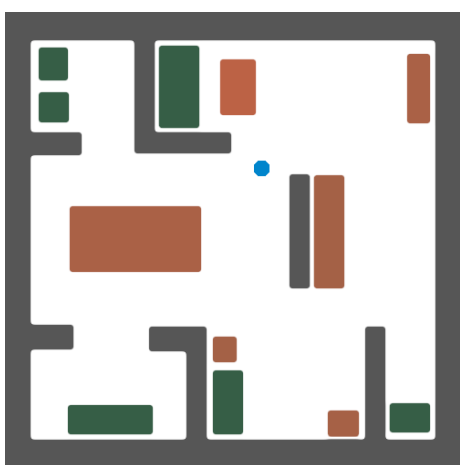


Figura 4.1: Representação da localização de um robô, em azul, numa sala com obstruções.

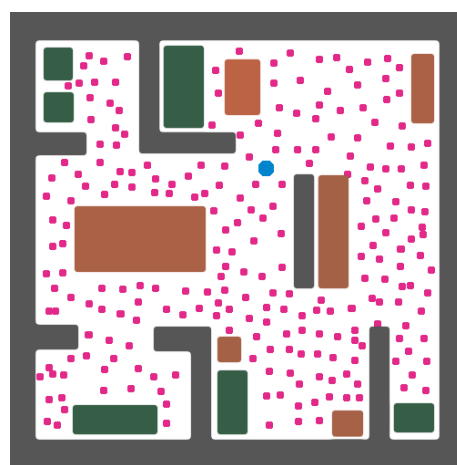


Figura 4.2: Disposição das partículas (localizações hipotéticas), em rosa, na sala.

Cada ponto apresentado na figura 4.2 é composto de 3 valores (x, y, z) , em que

(x, y) designam a posição no plano e z o ângulo informando a localização do robô. O próximo passo consiste na obtenção de dados (mensurações) feitas a partir de sensores, localizados no robô, de modo que seja possível calcular a distância das paredes a um círculo de raio r , por exemplo, 2 metros. Tais mensurações descrevem a **parte observável** do problema, isto é, o vetor Y_t dos modelos apresentados na seção 2.3.

Com tal vetor Y_t de mensurações, define-se quais são as partículas localizadas nas proximidades da posição do robô. Um obstáculo é considerado próximo ao robô caso esteja no círculo com raio de 2 metros. A figura 4.3 mostra tal raio de proximidade do robô e as partículas que foram consideradas próximas, considerando-se algum critério de proximidade. Tal critério dá maior peso às partículas com maior probabilidade de representarem a posição real do robô. A figura 4.4 mostra tal classificação das partículas com representações opacas (alta probabilidade) e semi-transparentes (baixa probabilidade).

Algumas das partículas, ainda que distantes, foram classificadas como próximas ao robô. Isso se deve à geometria da sala, próxima a tais pontos, que apresenta uma configuração parecida àquela em que os pontos realmente próximos estão inseridos. Como mostrado na figura 4.4, pontos com cores mais fortes representam posições mais prováveis de estarem próximas à posição real do robô, quando comparadas aos pontos com cores mais suaves. Na realidade, pode-se representar gradações de cores em uma escala contínua o intervalo $[0, 1]$. Nos exemplos, devido ao aspecto didático, apenas os extremos, $\{0, 1\}$ estão sendo representados.

Com isso, todas as partículas consideradas próximas tem a mesma **representatividade** (peso) para estimação da localização real do robô. No entanto, como mencionado, é possível definir funções com atribuições de peso mais gradativas (contínuas), para representar a localização. Isso seria feito, por exemplo, caso a função fosse uma medida de distância como a Euclidiana.

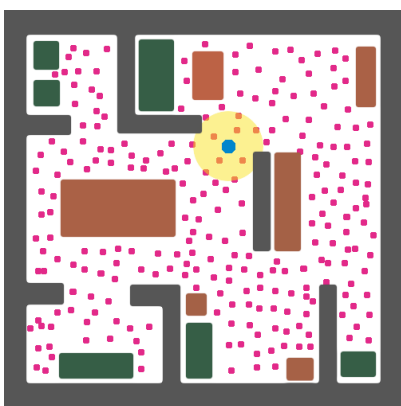


Figura 4.3: Sensor identificando as partículas mais próximas à atual posição do robô.

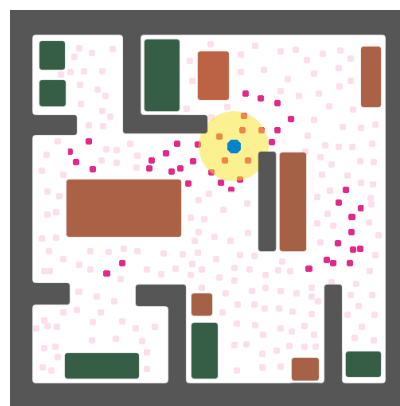


Figura 4.4: Selecionando partículas mais prováveis como representantes da posição do robô.

Como dito, o robô se movimenta nos espaços da sala e, por isso, esse processo de estimação da localização deve ser **dinâmico**. Para isso, considere que cada partícula **sobrevivente** do passo anterior possa ser selecionada, de acordo com o peso (distância), para reproduzir e gerar novas partículas com uma variação ao redor de sua atual localização. A justificativa para isso é que as partículas consideradas próximas ao robô (mais “aptas”) devem gerar outras partículas que possam auxiliar a encontrar a nova posição do robô, assim como feito no início do processo. Nesse procedimento, é possível selecionar uma partícula mais de uma vez, e é esperado que as partículas mais próximas sejam selecionadas mais do que as mais distantes. Esse é o processo de **reamostragem**, importante para eliminar partículas distantes que

foram selecionadas, embora possuam baixos pesos.

O processo de **reprodução** das partículas é conhecido como **propagação**, quando a *posteriori* (partículas sobreviventes) gera *novas partículas* para a próxima *posteriori*, a nova posição do robô.

Novamente, faz-se uma nova leitura (varredura) pelo sensor localizado no robô. Posteriormente, as partículas sobreviventes são avaliadas quanto às suas distâncias em relação às leituras e, novamente, são classificadas como próximas ou não próximas ao robô. Esse processo é conhecido como **atualização**. Nesse passo são calculados os novos pesos das partículas, e é feita a nova estimativa para a nova posição do robô.

A figura 4.5 ilustra uma situação para os passos de **propagação** e **atualização**.

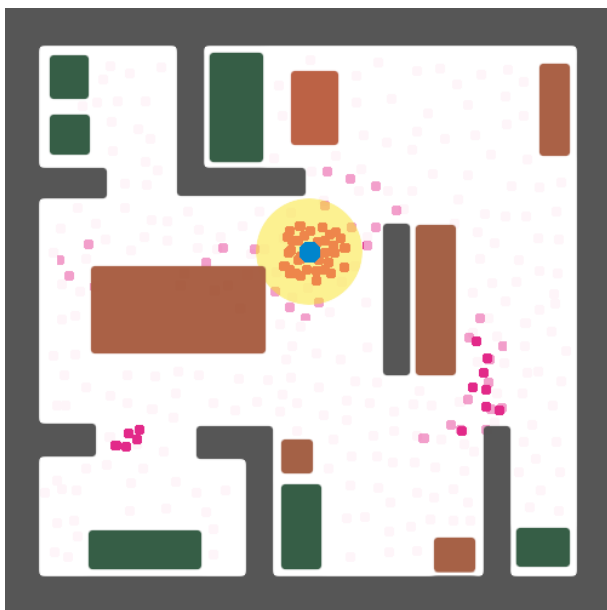


Figura 4.5: Gerando novos palpites em torno dos palpites mais prováveis.

4.1 Filtro de partículas em geral

Um modelo geral de espaços de estados pode ser descrito por:

Equação das observações:

$$z_k \sim p(z_k | x_k, \theta_1) \quad (4.1)$$

Equação do sistema:

$$x_k \sim p(x_k | x_{k-1}, \theta_2) \quad (4.2)$$

Em que z_k representa a observação no tempo t_k e $\{x_k\}$ é um processo estocástico não observável. Um problema importante na análise dos Modelos de Espaços de Estados é a estimação do processo não observável por meio de medições do processo observável. Para tanto, considere, em geral, as estimações como *offline* e *online*.

Estimação *offline*: Condicionado aos dados $z_{1:N} = (z_1, \dots, z_N)$, a estimação de $x_{1:N} = (x_1, \dots, x_N)$ é usualmente referenciada como estimação *offline*.

Estimação *online*: Quando o processo de estimação é sequencial, isto é, a estimação de x_k é baseada em $z_{1:k}$, para $k = 1, 2, \dots$, diz-se que o processo de estimação é *online*. Nesse caso, novas observações chegam em (horas/minutos/segundos) e estimação em tempo real é necessária.

Considere, inicialmente, que θ seja fixo ou conhecido. Os filtros de partículas possibilitam trabalhar com o processo de estimação *online*, daí o grande interesse no estudo de tal metodologia.

A ideia principal dos filtros de partículas é representar a distribuição a *posteriori* dos estados latentes até tempo k , isto é,

$$p(x_{1:k} | z_{1:k}) \quad (4.3)$$

por meio de um conjunto finito de amostras ou partículas que podem ser usadas para estimar qualquer propriedade de $p(x_{1:k} | z_{1:k})$, utilizando-se uma abordagem de

estimação de Monte Carlo, numa forma sequencial. Quando uma nova observação z_{k+1} é disponibilizada, as partículas são atualizadas de modo a representar a nova *posteriori*:

$$p(x_{1:k+1}|z_{1:k+1}). \quad (4.4)$$

Atualmente, há muitas versões de filtros de partículas. Duas abordagens são tipicamente utilizadas na construção do filtro. São elas: amostragem de importância sequencial e aproximação por histograma (amostragem de importância simples).

Abordagem I: A amostragem de importância sequencial (AIS).

No contexto de filtro, necessita-se calcular uma aproximação numérica para a distribuição $\{p(x_{1:n}|z_{1:n})\}_{n \geq 1}$, sequencialmente no tempo.

Nesse caso, de acordo com a notação da equação (3.4), a distribuição-alvo $\pi_n(x_{1:n})$, que é dada por

$$\pi_n(x_{1:n}) = \frac{\gamma_n(x_{1:n})}{Z_n}, \quad (4.5)$$

é tal que $\gamma_n(x_{1:n}) = p(x_{1:n}, z_{1:n})$, $Z_n = p(z_{1:n})$ e $\pi_n(x_{1:n}) = p(x_{1:n}|z_{1:n})$. Isto é,

$$\begin{aligned} p(x_{1:n}|z_{1:n}) &= \frac{p(x_{1:n}, z_{1:n})}{\int p(x_{1:n}, z_{1:n}) dx_{1:n}} = \frac{\gamma_n(x_{1:n})}{Z_n} \\ &= \frac{p(x_{n-1}|z_{1:n-1})p(x_n|x_{n-1})p(z_n|x_n)}{\int p(x_{n-1}|z_{1:n-1})p(x_n|x_{n-1})p(z_n|x_n) dx_{1:n}} \end{aligned} \quad (4.6)$$

Considerando a equação (3.6), e uma densidade de importância $q_n(\theta_{1:n})$, tem-se:

$$\begin{aligned} p(x_{1:n}|z_{1:n}) &= \frac{\gamma_n(x_{1:n})}{q_n(x_{1:n})} \frac{q_n(x_{1:n})}{Z_n} \\ &= w_n(x_{1:n}) \frac{q_n(x_{1:n})}{Z_n} \end{aligned} \quad (4.7)$$

Agora, considere a amostragem sequencial discutida na seção 3.2, o peso $w_n(x_{1:n})$

pode ser escrito, de forma sequencial, como:

$$\begin{aligned}
w_n(x_{1:n}) &= \frac{\gamma_n(x_{1:n})}{q_n(x_{1:n})} \\
&= \frac{\gamma_{n-1}(x_{1:n-1})}{q_{n-1}(x_{1:n-1})} \times \frac{\gamma_n(x_{1:n})}{\gamma_{n-1}(x_{1:n-1})q_n(x_n|x_{1:n-1})} \\
&= w_{n-1}(x_{1:n}) \times \alpha_n(x_{1:n}),
\end{aligned} \tag{4.8}$$

sendo que α_n é denominado por peso de importância incremental.

De acordo com Doucet e Johansen (2008), $q_n(x_n|x_{1:n-1})$ deve ser escolhido de modo a minimizar a variância dos pesos de importância $w_n(\cdot)$ a cada tempo n . A escolha ótima de distribuição de importância é dada por:

$$\begin{aligned}
q_n^{otm}(x_n|x_{1:n-1}) &= p(x_n|z_n, x_{n-1}) \\
&= \frac{p(z_n|x_n)p(x_n|x_{n-1})}{p(z_n|x_{n-1})}.
\end{aligned} \tag{4.9}$$

No entanto, em muitas situações não é possível amostrar a partir de (4.9), mas deve-se tentar obter aproximações da mesma. De qualquer forma, a equação (4.9) mostra que a distribuição de importância deve ser da forma:

$$q_n(x_n|x_{n-1}) = q(x_n|z_n, x_{n-1}). \tag{4.10}$$

Além disso, não há nada a ser ganho ao construir distribuições de importância que dependam de $(z_{1:n-1}, x_{1:n-2})$.

De (4.6), (4.8) e (4.10), observe que o peso de importância incremental $\alpha_n(x_{1:n})$ é escrito como:

$$\begin{aligned}
\alpha_n(x_{1:n}) &= \frac{\gamma_n(x_{1:n})}{\gamma_{n-1}(x_{1:n-1})q_n(x_n|z_n, x_{n-1})} \\
&= \frac{p(x_{1:n-1}|z_{1:n-1})p(x_n|x_{1:n-1})p(z_n|x_n)}{p(x_{1:n-1}|z_{1:n-1})q(x_n|z_n, x_{n-1})} \\
&= \frac{p(x_n|x_{n-1})p(z_n|x_n)}{q(x_n|z_n, x_{n-1})} = \alpha_n(x_{n-1}, x_n)
\end{aligned}$$

Sumarizando

Das equações (4.5),(4.7),(4.8) e a expressão (3.9) na seção 3.2, tem-se:

$$\hat{p}(x_{1:n}|z_{1:n}) = \frac{\sum_{i=1}^N w_n(x_{1:n}^i) \delta(x_{1:n} = x_{1:n}^i)}{\sum_{i=1}^N w_n(x_{1:n}^i)}, \quad (4.11)$$

em que

$$w_n(x_{1:n}) = w_{n-1}(x_{1:n-1}) \times \alpha_n(x_{1:n}),$$

com

$$\alpha_n(x_{1:n}) = \frac{p(x_n|x_{n-1})p(z_n|x_n)}{q(x_n|z_n, x_{n-1})}.$$

A fórmula (4.11) pode ser reescrita como:

$$\hat{p}(x_{1:n}|z_{1:n}) = \sum_{i=1}^N W_n(x_{1:n}^i) \delta(x_{1:n} = x_{1:n}^i), \quad (4.12)$$

em que

$$W_n(x_{1:n}^i) = \frac{w_n(x_{1:n}^i)}{\sum_{i=1}^N w_n(x_{1:n}^i)}$$

e $\delta(a = a_0)$ representa a função indicadora de massa no ponto a_0 . As amostras $\{X_n^i, i = 1, \dots, N\}$ são geradas a partir de $q(x_n|z_n, x_{1:n-1}^i)$. Com tal procedimento, obtém-se um conjunto de partículas ponderado $S_k = \{(X_{1:k}^i, w_{1:k}^i), i = 1, \dots, N\}$.

Algoritmo para o Filtro de Partículas via amostragem de importância sequencial

No tempo $n = 1$:

- Amostre $X_1^i \sim q_1(x_1)$.
- Calcule os pesos $w_1(X_1^i)$ e $W_1^i \propto w_1(X_1^i)$

No tempo $n \geq 2$:

- Amostre $X_n^i \sim q_n(x_n|z_n, X_{1:n-1}^i)$

- Calcule os pesos:

$$w_n(X_{1:n}^i) = w_{n-1}(X_{1:n-1}^i) \times \alpha_n(X_{1:n}^i),$$

$$W_n^i \propto w_n(X_{1:n}^i).$$

Obtenha o conjunto de partículas ponderando

$$S_k = \{(X_{1:k}^i, w_{1:k}^i), i = 1, \dots, N\}, k = 1, \dots, T$$

4.1.1 Reamostragem

Tanto a amostragem de importância (AI) quanto a amostragem de importância sequencial (AIS) conduzem as estimativas cujas variâncias aumentam, tipicamente exponencialmente, com n . Técnicas de reamostragem são ingredientes-chave nos métodos de Monte Carlo sequencial na solução (parcial) desse problema, para alguns casos específicos.

O processo de reamostragem está presente no processo de estimação para reduzir a discrepância, o que em casos extremos causa a chamada **degeneração dos pesos de importância**.

Há mais de um método de reamostragem e os mais discutidos são:

- Reamostragem Multinomial;
- Reamostragem Residual;
- Reamostragem Estratificada;
- Reamostragem Sistemática.

O método mais utilizado é o multinomial, em que o processo de reamostragem é realizado tratando a amostra como uma distribuição multinomial. O processo de reamostragem é um experimento n vezes, sendo n o tamanho da amostra, e em cada repetição, as amostras têm probabilidade de serem escolhidas igual ao peso das mesmas.

O processo de reamostragem mencionado nessa dissertação utiliza o método multinomial. Mais detalhes sobre os métodos de reamostragem são encontrados em Douc e Cappe (2006). Nas próximas seções discutiremos alguns artigos seminais na área de Filtro de Partículas.

4.2 Filtro de Partículas Bootstrap (Gordon et al., 1993)

Em Gordon et al. (1993), foi sugerida uma técnica que até o início da década de 90 não era conhecida como Filtro de Partículas. No entanto, a técnica utiliza os conceitos de amostragem sequencial de importância para as estimativas, isto é, à medida que novas observações são incorporadas ao estudo. Nesse artigo, os autores utilizam a amostragem de importância reamostrando os valores gerados, ponderados pela intensidade da verossimilhança. Atualmente, essa técnica é conhecida como Filtro de Partículas Bootstrap (FP Bootstrap), e também é associado à técnica SIR, do inglês, *Sampling Importance with Resampling*. Essa técnica tem como característica substituir o processo de atualização dos pesos, visto no **SIS**, pelo processo de reamostragem, selecionando partículas com grande massa de probabilidade.

4.2.1 Introdução

Os modelos de espaços de estados abordados até 1990 possuíam solução analítica e podiam ser estimadas no processo *on-line* sem grandes problemas, em especial os modelos Gaussianos. No entanto, quando o modelo possui parâmetros desconhecidos, a solução pelo filtro de Kalman pode não ser possível. Nesse contexto, utilizando-se uma modelagem Bayesiana, soluciona-se o problema por meio do MCMC. Utilizando-se tal técnica, a cada nova observação recalcula-se toda a distribuição a *posteriori* $\pi(\theta_{1:t}|y_{1:t})$, isto é, estima-se toda a cadeia de estados (Petris et al., 2009). Do ponto de vista computacional, as sucessivas reestimações da cadeia tornam o processamento muito lento. O Filtro de Partículas (FP) é uma ferramenta útil na tentativa de superar as limitações computacionais dos métodos numéricos de estimação por aproximação.

4.2.2 Estimador Bayesiano Recursivo

Nesta seção discutiremos a proposta de Gordon et al. (1993) para a construção de um algoritmo (Filtro Bootstrap) utilizado na implementação Bayesiana recursiva de um filtro. Considere o problema de estimação em tempo real discreto e $\theta_t \in \mathbb{R}^n$ um vetor de estados. Considere ainda $f : \mathbb{R}^n \times \mathbb{R}^m \Rightarrow \mathbb{R}^n$ uma função com propriedades recursivas em $t = 1, \dots, T$, que representa a evolução do estado θ_t , equação do sistema, tal que:

$$\theta_t = f(\theta_{t-1}, w_t), \quad (4.13)$$

sendo $w_t \in \mathbb{R}^m$, uma variável aleatória que segue uma distribuição $p_1(\cdot)$, com forma definida e conhecida. Em cada instante t , avalia-se $y_t \in \mathbb{R}^p$ por meio de uma função

$g : \mathbb{R}^n \times \mathbb{R}^r \Rightarrow \mathbb{R}^p$, tal que a equação das observações é dada por,

$$y_t = g(\theta_t, v_t), \quad (4.14)$$

sendo $v_t \in \mathbb{R}^r$ uma variável aleatória que segue uma distribuição $p_2(\cdot)$, com forma definida e conhecida. Considere w_t e v_t independentes, a distribuição inicial para $p(\theta_1|D_0) = p(\theta_1)$, sendo que $D_t = \{y_t, D_{t-1}\}$ representa a informação do sistema até instante t .

O objetivo é construir a função de densidade de probabilidade (f.d.p) para o estado θ_t , dadas as informações até o tempo t , isto é, a distribuição a *posteriori* $p(\theta_t|D_t)$. Essa densidade pode ser obtida seguindo dois passos importantes: **predição** e **atualização**. Esse processo de estimação por aproximação pode ser obtido utilizando os conceitos bayesianos nas relações descritas por meio das funções $f(\cdot)$ e $g(\cdot)$.

Supondo conhecida a distribuição a *posteriori* do estado latente no tempo $t - 1$, $p(\theta_{t-1}|D_{t-1})$, é possível obter a densidade a *priori* do estado latente θ_t por meio de

$$p(\theta_t|D_{t-1}) = \int p(\theta_t|\theta_{t-1})p(\theta_{t-1}|D_{t-1})d\theta_{t-1}. \quad (4.15)$$

Esta relação mostra que, para obter a distribuição de $p(\theta_t|D_{t-1})$, é necessário percorrer todo o espaço paramétrico de θ_{t-1} .

Considere que a distribuição dos erros de evolução, w_t , independe da recursividade do estado latente θ_t , $\forall t$, isto é, $p(w_{t-1}|\theta_{t-1}) = p(w_{t-1})$. Logo, a distribuição condicional do estado θ_t dado o estado θ_{t-1} é dada por:

$$p(\theta_t|\theta_{t-1}) = \int p(\theta_t|\theta_{t-1}, w_{t-1})p(w_{t-1}|\theta_{t-1})dw_{t-1}. \quad (4.16)$$

Nesta representação é possível perceber que obtém-se a distribuição $p(\theta_t|\theta_{t-1})$ avaliando todo o espaço paramétrico de w_{t-1} . Pode-se perceber que θ_t é obtido de forma determinística por meio de $f(\cdot|\theta_{t-1}, w_{t-1})$ caso θ_{t-1} e w_{t-1} sejam conhecidos, (vide equação 4.13). Com isso, a equação (4.16) é escrita como:

$$p(\theta_t|\theta_{t-1}) = \int \delta(\theta_t - f(\theta_{t-1}, w_{t-1}))p(w_{t-1})dw_{t-1}, \quad (4.17)$$

sendo $\delta(\cdot)$ a função Delta de Dirac, que assume o valor de 1 quando a diferença de $\{\theta_t - g(\theta_{t-1}, w_{t-1})\}$ é zero, e assume valor zero, caso contrário. Isto é, tal função representa uma variável aleatória degenerada tal que $Pr(\theta_t = g(\theta_{t-1}, w_{t-1})) = 1$. Sendo assim, define-se a função δ de Dirac por:

$$\delta(\theta_t - g(\theta_{t-1}, w_{t-1})) = \begin{cases} 1, & \text{se } \theta_t - g(\theta_{t-1}, w_{t-1}) = 0 \\ 0, & \text{c.c} \end{cases}$$

Dessa forma, a equação (4.15) pode ser reescrita como uma integral dupla, com relação aos suportes de θ_{t-1} e w_{t-1} , definido a **Equação de Propagação**:

$$p(\theta_t|D_{t-1}) = \int \left[\int p(\theta_t|\theta_{t-1}, w_{t-1})p(w_{t-1}|\theta_{t-1})dw_{t-1} \right] p(\theta_{t-1}|D_{t-1})d\theta_{t-1}.$$

Usando a independência entre θ_{t-1} e w_{t-1} , tem-se:

$$p(\theta_t|D_{t-1}) = \int \int f(\theta_t|\theta_{t-1}, w_{t-1})p(w_{t-1})p(\theta_{t-1}|D_{t-1})dw_{t-1}d\theta_{t-1}. \quad (4.18)$$

Com isso, pode-se obter amostras de $p(\theta_t|D_{t-1})$, por meio dos seguintes passos:

- 1- **amostrar θ_{t-1} a partir de $p(\theta_{t-1}|D_{t-1})$.** Esse é o ponto em que se obtém amostras passando por todo o suporte da distribuição de θ_{t-1} (integral com relação a θ_{t-1}).

- 2- **amostrar** w_{t-1} **a partir de** $p(w_{t-1})$. Nesse momento do processo, obtém-se amostras passando por todo o suporte da distribuição de w_{t-1} , associado à integral com relação a w_{t-1} .
- 3- Obter, de forma determinística, as **amostras da distribuição (4.18) por meio de cada par** (θ_{t-1}, w_{t-1}) aplicados na função $f(\cdot)$, isto é, gera-se:

$$\theta_t = f(\theta_{t-1}, w_t) \sim p(\theta_t|D_{t-1}).$$

Com o esquema apresentado, a equação (4.18) pode ser representada da seguinte maneira:

$$p(\theta_t|D_{t-1}) = \int \int \overbrace{f(\theta_t|\theta_{t-1}, w_{t-1})}^3 \overbrace{p(w_{t-1})}^2 \overbrace{p(\theta_{t-1}|D_{t-1})}^1 dw_{t-1}d\theta_{t-1}.$$

Utilizando o teorema de Bayes, a distribuição a *posteriori* do estado θ_t , $p(\theta_t|D_t)$, é expressa por:

$$\begin{aligned} p(\theta_t|D_t) &= \frac{g(\theta_t, D_t)}{p(D_t)} = \frac{p(\theta_t, D_{t-1}, y_t)}{p(D_{t-1}, y_t)} \\ &= \frac{p(y_t|\theta_t, D_{t-1})p(\theta_t|D_{t-1})p(D_{t-1})}{p(y_t|D_{t-1})p(D_{t-1})} \\ &= \frac{p(y_t|\theta_t, D_{t-1})p(\theta_t|D_{t-1})}{p(y_t|D_{t-1})} \\ &= \frac{g(y_t|\theta_t)p(\theta_t|D_{t-1})}{p(y_t|D_{t-1})}, \end{aligned}$$

sendo $p(y_t|D_{t-1})$ a constante normalizadora, obtida por:

$$p(y_t|D_{t-1}) \int g(y_t|\theta_t)p(\theta_t|D_{t-1})d\theta_t.$$

Dessa forma, representa-se a distribuição a *posteriori* do estado latente no tempo t por:

$$p(\theta_t|D_t) = \frac{g(y_t|\theta_t)p(\theta_t|D_{t-1})}{\int g(y_t|\theta_t)p(\theta_t|D_{t-1})d\theta_t}. \quad (4.19)$$

No denominador da equação (4.19), a densidade $h(y_t|\theta_t)$ é obtida por meio de:

$$g(y_t|\theta_t) = \int \delta(y_t - g(\theta_t, v_t))p(v_t)dv_t,$$

isto é, a distribuição das observações é obtida avaliando-se o estado latente θ_t em todo espaço paramétrico de v_t .

Soluções analíticas para esses processos podem ser obtidas pelo Filtro de Kalman, mas tais soluções são restritas a sistemas com funções $f(\cdot)$ e $g(\cdot)$ lineares e erros w_t e v_t aditivos e com distribuição gaussiana. Tais restrições são comumente violadas em sistemas mais complexos.

4.2.3 FP Bootstrap Bayesiano Gordon et al. (1993)

O FP Bootstrap proposto por Gordon et al. (1993) sugere uma metodologia numérica para lidar com algumas das limitações do Filtro de Kalman. O Filtro Bootstrap aproxima as distribuições representadas pelas equações (4.15) e (4.17) por meio de amostras sequenciais.

Tal metodologia utiliza os passos de **propagação**, gerando valores a partir da *priori* $p(\theta_t|D_{t-1})$ e **atualização** em que as amostras obtidas no passo anterior são reponderadas de modo a se obter uma amostra a partir da posteriori $p(\theta_t|D_t)$.

4.2.4 Descrição Matemática

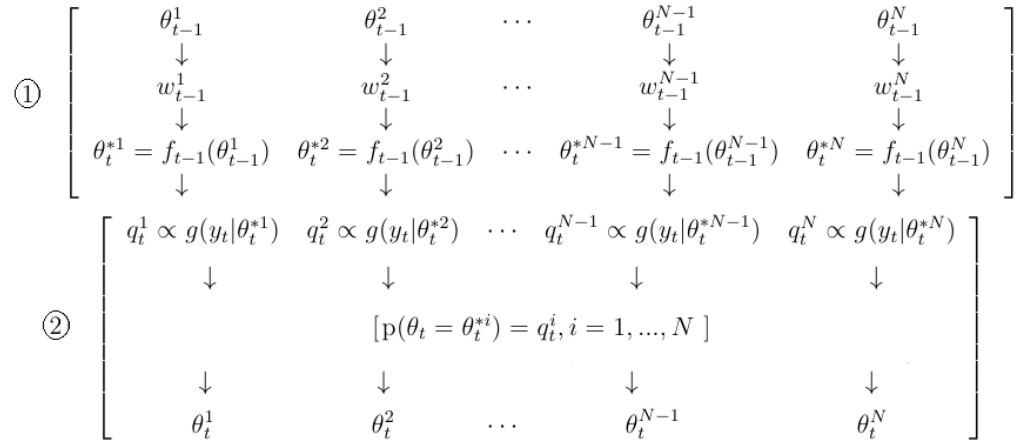
Seja a amostra $\{\theta_{t-1}^i : i = 1, \dots, N\}$ obtida a partir de $p(\theta_{t-1}|D_{t-1})$. O Filtro Bootstrap é um algoritmo que permite a obtenção de amostras a partir de $p(\theta_t|D_t)$ de acordo com os seguintes passos:

- 1 **Propagação** - Condicionamente à $\{\theta_{t-1}^i : i = 1, \dots, N\}$, tirar amostras a partir

da *priori* $p(\theta_t|D_{t-1})$.

2 **Atualização** - Utilizar as amostras obtidas no passo 1 para obter amostras a partir da *posteriori* $p(\theta_t|D_t)$.

No passo 1 (**propagação**), condicionalmente ao valor de θ_{t-1}^i e ao erro w_{t-1}^i , gerado a partir de $p(w_t)$, isto é, utilizando o par $(\theta_{t-1}^i, w_{t-1}^i)$, obtendo-se um valor candidato, $\theta_t^{*i} = f_{t-1}(\theta_{t-1}^i, w_{t-1}^i), i = 1, \dots, N$, gerado a partir de $p(\theta_t|D_{t-1})$. Observe a parte 1 da figura 4.6 e as relações descritas na expressão (4.18).



① Propagação, ② Atualização

$$q_t^i = \frac{g(y_t|\theta_t^{*1})}{\sum_{j=1}^N g(y_t|\theta_t^{*j})}, i = 1, \dots, N.$$

Figura 4.6: Filtro de partículas Bootstrap - esquema para gerar partículas segundo o ciclo de atualização: $p(\theta_{t-1}|D_{t-1}) \rightarrow p(\theta_t|D_{t-1}) \rightarrow p(\theta_t|D_t)$.

No passo 2 (**atualização**), a informação referente à observação atual, y_t , é incorporada ao processo para que, por meio da expressão (4.19), seja possível obter amostras a partir da *posteriori* $p(\theta_t|D_t)$ (distribuição filtrada). A partir da equação

(4.19) tem-se:

$$\overbrace{p(\theta_t|D_t)}^{\text{Posteriori}} = \overbrace{\left[\frac{g(y_t|\theta_t)}{\int g(y_t|\theta_t)p(\theta_t|D_{t-1})d\theta_t} \right]}^{\text{pesos}} \overbrace{p(\theta_t|D_{t-1})}^{\text{priori}}. \quad (4.20)$$

Os pesos,

$$\frac{g(y_t|\theta_t)}{\int g(y_t|\theta_t)p(\theta_t|D_t)d\theta_t},$$

podem ser aproximadas por meio de:

$$q_t^i = \frac{g(y_t|\theta_t^{i*})}{\sum_j^N g(y_t|\theta_t^{j*})}$$

com $\theta_t^{i*} \sim p(\theta_t|D_{t-1})$.

Segundo Smith e Gelfand (1992), um esquema do tipo *bootstrap* ponderado é útil para obter amostras a partir de distribuições como na expressão (4.20). Para tanto, deve-se obter $\{\theta_t^{i*}, i = 1, \dots, N\}$ a partir de $p(\theta_t^i|D_{t-1})$ e criar uma distribuição discreta tal que $p(\theta_t^i = \theta_t^{i*}) = q_t^i$. Posteriormente, obtém-se uma amostra, com reposição, de tamanho N , a partir dessa última distribuição para obter $\{\theta_t^i, i = 1, \dots, N\}$. A parte 2 da figura 4.6 ilustra o processo de atualização.

A justificativa matemática para o passo de **propagação**, no qual é gerada uma amostra a partir da *priori*, é trivial pois o θ_t^{i*} é obtido de forma determinística de $f(\theta_{t-1}^i, w_t^i)$, quando se conhece os valores de θ_{t-1}^i e w_t^i .

Para a justificativa do passo da **atualização**, o Teorema de Bayes pode ser implementado por meio de um *bootstrap* ponderado (Smith e Gelfand, 1992). Para isso, seja $\{\theta_i^*, i = 1, \dots, N\}$ uma amostra de uma f.d.p $G(\theta)$, mas deseja-se obter amostras de uma distribuição $H(\theta)$, tal que:

$$H(\theta) \propto L(\theta)G(\theta),$$

sendo $L(\theta)$ uma função conhecida.

Para gerar amostras $\{\theta_i, i = 1, \dots, N\}$ de $H(\theta)$, pode se reamostrar as partículas $\{\theta_i^*, i = 1, \dots, N\}$, geradas a partir $G(\theta)$, com pesos proporcionais a $L(\theta_i^*)$. Dessa forma, sendo

$$\pi_i = \frac{L(\theta_i^*)}{\sum_j^N L(\theta_j^*)},$$

tem-se que $P(\theta_j = \theta_i^*) = \pi_i$.

A distribuição empírica obtida por meio desse processo converge, em distribuição, para $H(\theta)$ (Smith e Gelfand, 1992). Para o contexto do filtro apresentado, temos que:

$$p(\theta_t|D_t) \propto p(y_t|\theta_t)p(\theta_t|D_{t-1}),$$

sendo $G(\theta) = p(\theta_t|D_{t-1})$, a *priori* da qual temos uma amostra $\{\theta_t^{i*}, i = 1, \dots, N\}$, $L(\theta) = p(y_t|\theta_t)$, a função de verossimilhança, e $H(\theta) = p(\theta_t|D_t)$, a distribuição a *posteriori* da qual deseja-se ter amostras $\{\theta_t^{i*}, i = 1, \dots, N\}$.

4.2.5 Algoritmo - FB Bootstrap

- 1.1 Para $t = 1$: gerar $\{\theta_0^i, i = 1, \dots, N\} \sim p(\theta_0)$;
- 1.2 Para $t > 1$: utilizar $\{\theta_{t-1}^i, i = 1, \dots, N\}$;
- 2 Gerar amostra para $w_t^i \sim p_1(w)$;
- 3 Obter valores de θ_{t-1}^{i*} , de forma determinística, $\theta_{t-1}^i = f(\theta_{t-1}^i, w_t^i)$;
- 4 Sendo v_t um valor gerado a partir de $p(v_t)$, atualiza-se o peso θ_{t-1}^{i*} usando

$$q_t^i = \frac{p(y_t|\theta_t^{i*}, v_t)}{\sum_j^N p(y_t|\theta_t^{j*}, v_t)};$$

- 5 Reamostrar $\{\theta_t^{i*}, i = 1, \dots, N\}$ N vezes com probabilidade igual a q_t^i .

4.2.6 Exemplo de aplicação do FB Bootstrap

O exemplo apresentado a seguir está presente em vários artigos de autores na área, entre estes no trabalho de Gordon et al. (1993). Trata-se de um Modelo Dinâmico não linear com erros aditivos.

Seja a equação do sistema descrita como:

$$\theta_t = f_t(\theta_{t-1}, w_t) = 0.5\theta_t + \frac{25\theta_{t-1}}{(1 + \theta_{t-1}^2)} + 8 \cos(1.2(t - 1)) + w_t.$$

A figura 4.7 mostra a função com a transição não-linear de θ_{t-1} para θ_t .

$$\theta_t = \frac{\theta_{t-1}}{2} + 25 \frac{\theta_{t-1}}{(1 + \theta_{t-1}^2)} + 8 \cos(1.2(t-1))$$

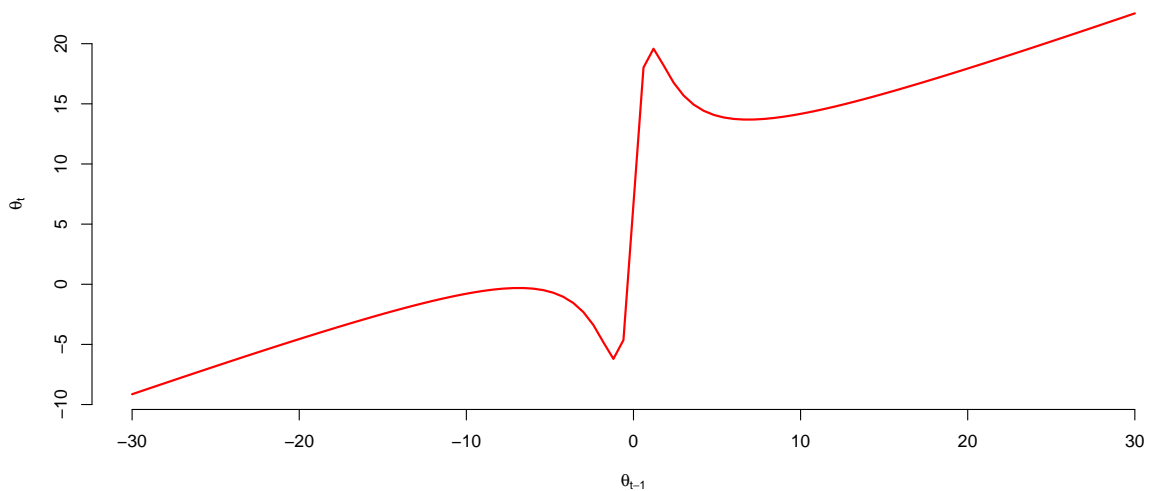


Figura 4.7: Equação de transição não-linear

Considere $w_t \sim N(0, W)$ e a parte do modelo observável representada por:

$$y_t = g(\theta_t, v_t) = \frac{\theta_t^2}{20} + v_t,$$

com $v_t \sim N(0, V)$.

Dessa forma, implementando os passos apresentados no algoritmo referente ao FP Bootstrap (vide seção 4.2.5), a figura 4.8 apresenta o resultado da estimação dos

estados latentes do modelo proposto considerando-se $N = 500$ partículas. Observe-se que a metodologia descrita nesta seção proporciona um bom ajuste.

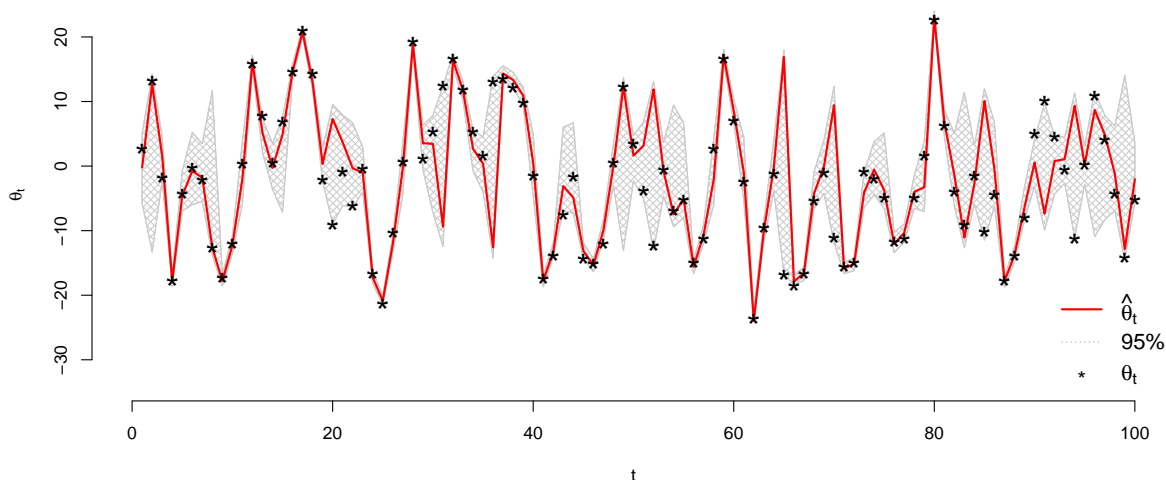


Figura 4.8: FP Bootstrap em um modelo não-linear

O Gráfico 4.9 mostra a distribuição das partículas ao longo do processo de estimação para algumas iterações. É apresentada a distribuição das partículas por meio de histogramas, além disso, é comparado o valor real (representado pela linha pontilhada na cor vermelha) e a mediana da distribuição das partículas (representado pela linha pontilhada na cor azul).

4.3 FP Auxiliar (Pitt e Shephard, 1999)

O FP Auxiliar, proposto por Pitt e Shephard (1999), tem como principal característica a geração de partículas com grandes massas de probabilidade. Isso se deve à utilização das informações da parte observável do sistema para gerar propostas mais eficientes para os estados latentes.

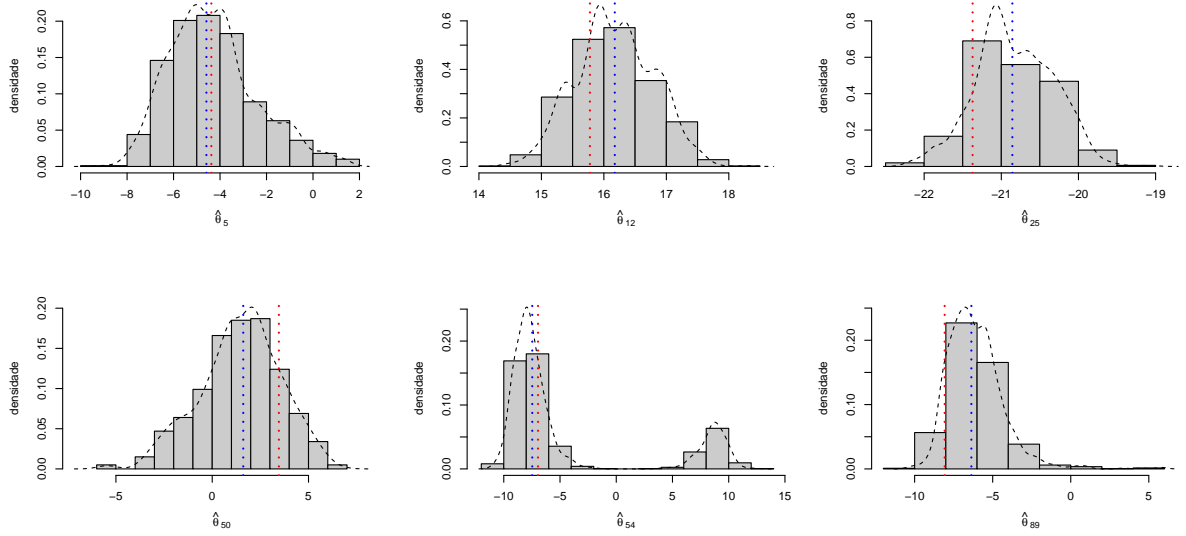


Figura 4.9: Densidades do processo de estimação BF

4.3.1 Descrição matemática

Considere um processo que pode ser representado por dois sub-processos, um observável e outro não-observável (latente). O processo latente evolui segundo um kernel de transição entre instantes no tempo t e $t - 1$ descrito pela função $f(\theta_t|\theta_{t-1})$. O processo observável é dependente do estado latente e é descrito por meio da função $f(y_t|\theta_t)$. Seja $f(\theta_0)$ a distribuição inicial desse sistema recursivo.

Considere, ainda, a representação $f(\theta_t|Y_t) = f(\theta_t|(y_t, \dots, y_1))$, como a distribuição do estado latente no instante t condicionado a toda informação até tal instante.

O processo de filtragem pode ser obtido repetindo dois passos importantes. O primeiro deles consistem em **propagar** a densidade relativa ao estado atual θ_t via densidade de transição $f(\theta_t|\theta_{t-1})$, produzindo:

$$f(\theta_{t+1}|Y_t) = \int f(\theta_{t+1}|\theta_t)dF(\theta_t|Y_t).$$

Posteriormente, com a chegada da nova observação y_{t+1} , deve-se obter a densidade filtrada utilizando-se o teorema de Bayes:

$$f(\theta_{t+1}|Y_{t+1}) = \frac{f(y_{t+1}|\theta_{t+1})f(\theta_{t+1}|Y_t)}{f(y_{t+1}|Y_t)}, \quad (4.21)$$

sendo

$$f(y_{t+1}|Y_t) = \int f(y_{t+1}|\theta_{t+1})dF(\theta_{t+1}|Y_t), \quad (4.22)$$

a densidade preditiva.

O FP pode, então, ser definido como o processo que aproxima a distribuição $(\theta_t|Y_t = (y_1, \dots, y_t))$ por meio de pontos de tal densidade, conhecidos como **partículas**, $(\theta_t^1, \dots, \theta_t^M)$, que possuem massa de probabilidade $(\pi_t^1, \dots, \pi_t^M)$.

Essa forma de representação discreta do suporte da distribuição de interesse é útil para aproximar a distribuição preditiva:

$$\hat{f}(\theta_{t+1}|Y_t) = \int f(\theta_{t+1}|\theta_t)f(\theta_t|Y_t)d\theta_t, \quad (4.23)$$

por meio de

$$\hat{f}(\theta_{t+1}|Y_t) = \sum_j^M f(\theta_{t+1}|\theta_t^j)\pi_t^j, \quad (4.24)$$

com $\pi_t^i = p(\theta_t = \theta_t^i|Y_t)$. Este resultado pode ser combinado com a equação (4.21) para produzir:

$$\hat{f}(\theta_{t+1}|Y_{t+1}) \propto f(y_{t+1}|\theta_{t+1}) \sum_i^M f(\theta_{t+1}|\theta_t^i)\pi_t^i. \quad (4.25)$$

A partir da distribuição proposta gera-se novas partículas $(\theta_{t+1}^1, \dots, \theta_{t+1}^M)$ que possuem massas de probabilidades $(\pi_{t+1}^1, \dots, \pi_{t+1}^M)$.

É possível amostrar partículas de $\hat{f}(\theta_{t+1}|Y_{t+1})$ de três maneiras: SIR, Amostragem por aceitação e MCMC. Nesse trabalho, o processo de amostragem utilizará

a metodologia SIR, que foi apresentada na seção 3.3. Mais detalhes das demais metodologias podem ser encontrados no trabalho de Pitt e Shephard (1999).

4.3.2 SIR

O método SIR (Rubin, 1992) é um procedimento com o qual é possível amostrar $\theta^1, \dots, \theta^R$ a partir da *priori* $f(\theta)$ e, então, associar, a cada um dos valores selecionados θ^i , um peso π_i . Essa técnica é utilizada para calcular a distribuição alvo $f(\theta|y)$, (*posteriori* de θ em relação a y), que pode ser expressa, via teorema de Bayes, como:

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{\int f(y|\theta)f(\theta)d\theta}.$$

O denominador é a verossimilhança dos dados, isto é:

$$f(y) = \int f(y|\theta)f(\theta)d\theta.$$

Dessa forma, tem-se que:

$$f(\theta|y) = \left[\frac{f(y|\theta)}{f(y)} \right] f(\theta)$$

O peso π_i pode então ser calculado por:

$$\pi_i = \frac{w_i}{\sum_j^R w_j}, j = 1, \dots, R, \quad (4.26)$$

com

$$w_j = f(y|\theta^j). \quad (4.27)$$

Como apresentado na seção 3.3, utilizando uma aproximação $\hat{f}(\theta)$, obtida por meio das amostras θ^i geradas a partir da *priori*, e o resultado dos pesos obtidos pela equação (4.26), para obter uma estimativa da distribuição a *posteriori* de θ :

$$f(\theta|y) \approx \pi_i \times \hat{f}(\theta). \quad (4.28)$$

A amostra ponderada pelos w_j 's converge, quando $R \rightarrow \infty$, para uma amostra não aleatória da distribuição alvo $f(\theta|y)$, uma vez que (Pitt e Shephard, 1999):

$$\frac{1}{R} \sum_{i=1}^R w_i \xrightarrow{p} f(y). \quad (4.29)$$

Essa convergência em probabilidade é obtida pela aproximação da seguinte média:

$$f(y) = \int f(y|\theta)f(\theta)d\theta = E_{\theta}[f(y|\theta)] \approx \frac{1}{R} \sum_{j=1}^R f(y|\theta^j). \quad (4.30)$$

Para compreender a eficiência do SIR, é útil pensar neste método como sendo uma aproximação para o amostrador de importância do primeiro momento *a posteriori* de θ :

$$\begin{aligned} E(h(\theta)) &= \int h(\theta)f(\theta|y)d\theta \\ &= \int h(\theta) \frac{f(y|\theta)f(\theta)}{f(y)} d\theta \\ &= \int h(\theta) \left[\frac{f(y|\theta)}{f(y)} \right] f(\theta) d\theta \\ &= E_{f(\theta)} \left[h(\theta) \left(\frac{f(y|\theta)}{f(y)} \right) \right] \\ &\approx \frac{1}{R} \sum_j^R h(\theta^j) \frac{f(y|\theta^j)}{f(y)} = \frac{\frac{1}{R} \sum_j^R h(\theta^j) f(y|\theta^j)}{\int f(y|\theta) f(\theta) d\theta}, \end{aligned}$$

desta forma, tem-se a seguinte relação

$$\approx \frac{1}{R} \frac{\sum_j^R h(\theta^j) f(y|\theta^j)}{\left[\frac{1}{R} \sum_i^R f(y|\theta^i) \right]} = \frac{\sum_j^R h(\theta^j) f(y|\theta^j)}{\sum_i^R f(y|\theta^i)}. \quad (4.31)$$

Na aproximação (4.31), o numerador e o denominador são médias e a convergência se dá em probabilidade. Assim, o primeiro momento *a posteriori*, $E(h(\theta))$, é aproximado por meio de:

$$\widehat{E(h(\theta))} = \frac{\sum_j^R h(\theta^j) f(y|\theta^j)}{\sum_i^R f(y|\theta^i)}.$$

Dessa forma, utilizando o resultado (4.26), tem-se:

$$\widehat{E(h(\theta))} = \frac{\sum_j^R h(\theta^j)w_j}{\sum_i^R w_i} = \sum_{i=1}^R h(\theta^i)\pi^i.$$

O valor π^i estima a razão $f(y|\theta^i)/f(y)$. Então, para obter uma amostra de $f(\theta|y)$, gera-se amostras $\theta^i \sim f(\theta)$, a *priori*, e **reamostra-se** $(\theta_1, \dots, \theta_R)$ com pesos (π_1, \dots, π_R) .

Liu e West (2001) sugerem que a variância de $\widehat{E(h(\theta))}$ é aproximadamente proporcional a $E_f(\pi(\theta)^2)/R$, isto é:

$$V(\widehat{E(h(\theta))}) \propto E_f(\pi(\theta)^2)/R.$$

Então, o SIR é um método que se torna muito impreciso na estimação de $E(h(\theta))$ (e de outras características de $f(\theta|y)$, assim como as próprias amostras selecionadas, de forma aproximada, de tal distribuição) quando π_j torna-se muito variável. Isto ocorre se a verossimilhança $f(y|\theta)$ é muito leptocúrtica em comparação à priori $f(\theta)$, isto é, tem caudas mais pesadas (Pitt e Shephard, 1999).

4.3.3 SIR Adaptativo

O método SIR faz propostas “às cegas” por meio de uma *priori* $f(\theta)$, ignorando o fato as informações de y . Diz-se que um FP é adaptativo quando há a possibilidade de incorporar o conhecimento sobre y , sendo, então, um método “não as cegas”.

Um FP adaptativo baseado no SIR segue a seguinte estrutura:

1. Amostra-se $\theta^1, \dots, \theta^R \sim g(\theta|y)$;
2. Calcula-se $w_j = \frac{f(y|\theta^j)f(\theta^j)}{g(\theta^j|y)}$, $j = 1, \dots, R$;

3. Reamostra entre os θ^j utilizando pesos proporcionais w^j para produzir uma amostra de tamanho M .

Isso implica que deve-se calcular pelo menos $M \times R$ densidades de modo a gerar M amostras a partir de $f(\theta|y)$. Dado que M e R são muito grandes, isto implica que essa adaptação requer um grande esforço computacional. Há alguns anos, esse esforço seria inviável, porém com o desenvolvimento de computadores com alto desempenho, essa dificuldade é mais facilmente contornada.

4.3.4 SIR Auxiliar

No contexto do SIR, é possível reduzir a alta demanda computacional devida à magnitude de M e R quando os problemas são tratados em dimensões superiores, isto é, deixando de lado a análise via distribuições **marginais** e lidando com as distribuições **conjuntas** das variáveis envolvidas no processo (Pitt e Shephard, 1999).

O objetivo do FP auxiliar é amostrar a partir da distribuição conjunta $f(\theta_{t+1}, k|Y_{t+1})$,

$$f(\theta_{t+1}, k|Y_{t+1}) \propto f(y_{t+1}|\theta_{t+1})f(\theta_{t+1}|\theta_t^k)\pi^k, k = 1, \dots, M, \quad (4.32)$$

sendo k um índice na mistura na equação (4.24).

Seja:

$$f(\theta_{t+1}|Y_{t+1}) = \int f(\theta_{t+1}, k|Y_{t+1})dk, \quad (4.33)$$

em que

$$\begin{aligned} f(\theta_{t+1}, k|Y_{t+1}) &= \frac{f(y_{t+1}|\theta_{t+1}, k, Y_t)f(\theta_{t+1}, k|Y_t)}{f(y_{t+1}|Y_t)} \\ &= \frac{f(y_{t+1}|\theta_{t+1})f(\theta_{t+1}, k|Y_t)}{f(y_{t+1}|Y_t)}. \end{aligned} \quad (4.34)$$

Sendo,

$$f(\theta_{t+1}, k|Y_t) = \int f(\theta_{t+1}|\theta_t)f(\theta_t, k|Y_t)d\theta_t \quad (4.35)$$

e

$$f(y_{t+1}|Y_t) = \int \int f(y_{t+1}|\theta_{t+1})f(\theta_{t+1}, k|Y_t)dkd\theta_{t+1}. \quad (4.36)$$

Em (4.35), para

$$f(\theta_t, k|Y_t) \approx p(\theta_t = \theta_t^k|Y_t) = \pi^k, \quad (4.37)$$

então, para um dado valor θ_t^k de θ_t , tem-se:

$$f(\theta_{t+1}, k|Y_t) \approx f(\theta_{t+1}|\theta_t^k)p(\theta_t = \theta_t^k|Y_t) = f(\theta_{t+1}|\theta_t^k)\pi^k. \quad (4.38)$$

Defina, de (4.34),

$$f(\theta_{t+1}, k|Y_{t+1}) \approx f(y_{t+1}|\theta_{t+1})f(\theta_{t+1}, k|Y_t) \quad (4.39)$$

e de (4.38),

$$\begin{aligned} f(\theta_{t+1}, k|Y_{t+1}) &\propto f(y_{t+1}|\theta_{t+1})f(\theta_{t+1}|\theta_t^k)p(\theta_t = \theta_t^k|Y_t) \\ &\propto f(y_{t+1}|\theta_{t+1})f(\theta_{t+1}|\theta_t^k)\pi^k, k = 1, \dots, M \end{aligned} \quad (4.40)$$

Portanto, com base nas expressões (4.38) e (4.40), amostrar-se partículas para estimar a *posteriori*, apresentada na equação (4.25), por meio de um **variável auxiliar** (índice) k .

Denota-se k por **auxiliar**, uma vez que ela está presente, apenas, para auxiliar na tarefa de simulação e geração das partículas. Os filtros que utilizam essa variável auxiliar são conhecidos como **Filtros de Partículas Auxiliares**.

Assim, pode-se amostrar a partir de $f(\theta_{t+1}, k|Y_t)$ utilizando a técnica SIR. A ideia de amostragem pela metodologia SIR é fazer R propostas para $(\theta_{t+1}^j, k^j) \sim g(\theta_{t+1}, k|Y_{t+1})$, e reconstruir os pesos por meio de:

$$w_j = \frac{f(y_{t+1}|\theta_{t+1}^j)f(\theta_{t+1}^j|\theta_t^{k^j})}{g(\theta_{t+1}^j, k^j|Y_{t+1})}, \quad \pi_j = \frac{w_j}{\sum_i^R w_i}, j = 1, \dots, R.$$

É importante resaltar que a utilização da variável auxiliar para geração dos candidatos não é suficiente para considerar o método SIR auxiliar como um processo "não às cegas", Pitt e Shephard (1999).

Tem-se total controle no delineamento e definição da distribuição $g(\cdot)$, que pode depender tanto de y_{t+1} como de θ_t^k . Pode-se aproximar a equação (4.25) por:

$$g(\theta_{t+1}, k|Y_{t+1}) \propto f(y_{t+1}|\mu_{t+1}^k)f(\theta_{t+1}|\theta_t^k)\pi^k, k = 1, \dots, M,$$

sendo μ_{t+1}^k uma medida representativa da distribuição $f(\theta_{t+1}|\theta_t^k)$: média, moda, mediana ou uma amostra a partir de $f(\theta_{t+1}^j|\theta_t^k)$.

A forma da densidade marginal em relação a k é descrita como

$$\begin{aligned} g(k|Y_{t+1}) &= \int g(\theta_{t+1}, k|Y_{t+1})d\theta_{t+1} \\ g(k|Y_{t+1}) &\propto \pi^k \int f(y_{t+1}|\mu_{t+1}^k)dF(\theta_{t+1}|\theta_t^k) \\ g(k|Y_{t+1}) &\propto \pi^k f(y_{t+1}|\mu_{t+1}^k). \end{aligned}$$

A amostra de $g(\theta_{t+1}, k|Y_{t+1})$ é simulada amostrando, primeiramente, o indexador k , com probabilidade $\lambda_k \propto g(k|Y_{t+1})$. Posteriormente, amostra-se θ_{t+1} a partir de $f(\theta_{t+1}^j|\theta_t^k)$. O peso λ_k é chamado de primeiro estágio de ponderação.

Tendo amostrado R vezes da densidade conjunta $g(\theta_{t+1}, k|Y_{t+1})$, as partículas (θ_{t+1}^j, k^j) são reponderadas com peso proporcional ao segundo estágio:

$$w_j = \frac{f(y_{t+1}|\theta_{t+1}^j)}{f(y_{t+1}|\mu_{t+1}^{k^j})}, \pi_j = \frac{w_j}{\sum_i^R w_i}, j = 1, \dots, R. \quad (4.41)$$

Com esse método, faz-se propostas que tenham alta verossimilhança condicional, reduzindo o desperdício de se gerar partículas provenientes de regiões com baixa representatividade na distribuição *a posteriori* de interesse.

4.3.5 Algoritmo relativo ao FP auxiliar

Seja $k = 1, \dots, M$ o número de partículas e $i = 1, \dots, R$ o número de reamostragens realizadas nas partículas.

- 1.1 Obter propostas de partículas($t = 1$): $\theta_1^* \sim p_0(\cdot)$;
- 1.2 Obter propostas de partículas($t > 1$): $\theta_t^* | \theta_{t-1}^k \sim f_t(\cdot | \theta_{t-1}^k)$;
- 2 Obter valores representativos da distribuição $\mu_t^k \leftarrow \theta_t^* | \theta_{t-1}^k$;
- 3 calcular os pesos de $\lambda_k \propto w_{t-1}^k f_t(y_t | \mu_t^k)$;
- 4 Reamostrar R partículas com pesos λ_k : $\theta_{t-1}^{k^i} \sim \{\lambda_k, \theta_{t-1}^k\}$;
- 5 Obter propostas: $\theta_t^i | \theta_{t-1}^{k^i} \sim f_t(\cdot | \theta_{t-1}^{k^i})$;
- 7 Obter valores representativos da distribuição $\mu_t^{k^i} \leftarrow \theta_t^i | \theta_{t-1}^{k^i}$;
- 8 Atualizar os pesos das partículas $w_t^i = f_t(y_t | \theta_t^i) / f_t(y_t | \mu_t^{k^i})$;
- 9 Normalizar os pesos θ_t^i usando:

$$\pi_t^i = \frac{w_t^i}{\sum_j^R w_t^j};$$

- 10 Reamostrar M partículas θ_t^k de θ_t^i com peso igual a π_t^i
- 11 Normalizar os pesos θ_t^k usando:

$$\pi_t^k = \frac{w_t^k}{\sum_j^M w_t^j}.$$

4.3.6 Exemplos de aplicação do FP auxiliar

Modelos não-lineares

Utilizando o mesmo modelo não linear proposto na seção 4.2.6 e replicando o algoritmo para filtro de partícula auxiliar, tem-se o resultado representado pelo Gráfico 4.10 para as estimativas dos estados latentes ao longo do processo.

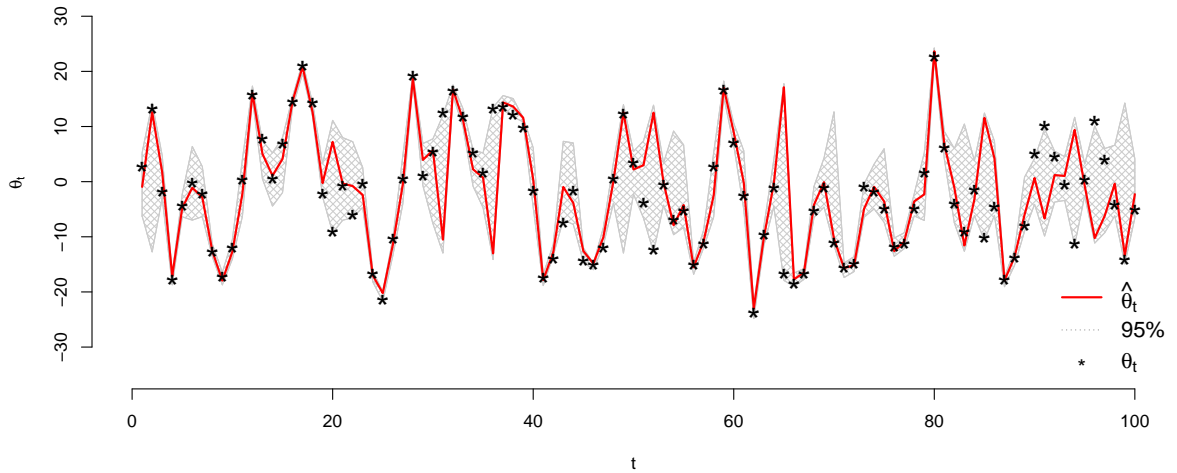


Figura 4.10: FP Auxiliar: Estimativas dos estados latentes

O Filtro de Partículas auxiliar apresentou excelentes estimações para os estados latentes do modelo não-linear. É possível perceber que as estimativas foram precisas, mesmo com a alta variabilidade dos estados latentes, com valores entre $[-30, 30]$.

A figura 4.11 mostra a distribuição das partículas ao longo do processo de estimação do modelo. É possível perceber que o Filtro de Partículas gera pontos no espaço paramétrico do estado latente, e, para cada ponto na distribuição gerada, atribui uma **massa de probabilidade empírica**, baseada na verossimilhança da parte observável do processo, y_t .

Na figura, é possível notar as distribuições empíricas geradas pelas partículas para os instantes $t = \{5, 12, 25, 50, 54, 89\}$. São apresentadas as partículas simuladas (apresentados pelos histogramas), a suavização por kernel gaussiano para o histograma (representado pela linha na cor preta), a posição no espaço paramétrico para a estimativa do estado latente $\hat{\theta}_t$ (representado pelo ponto no espaço paramétrico na

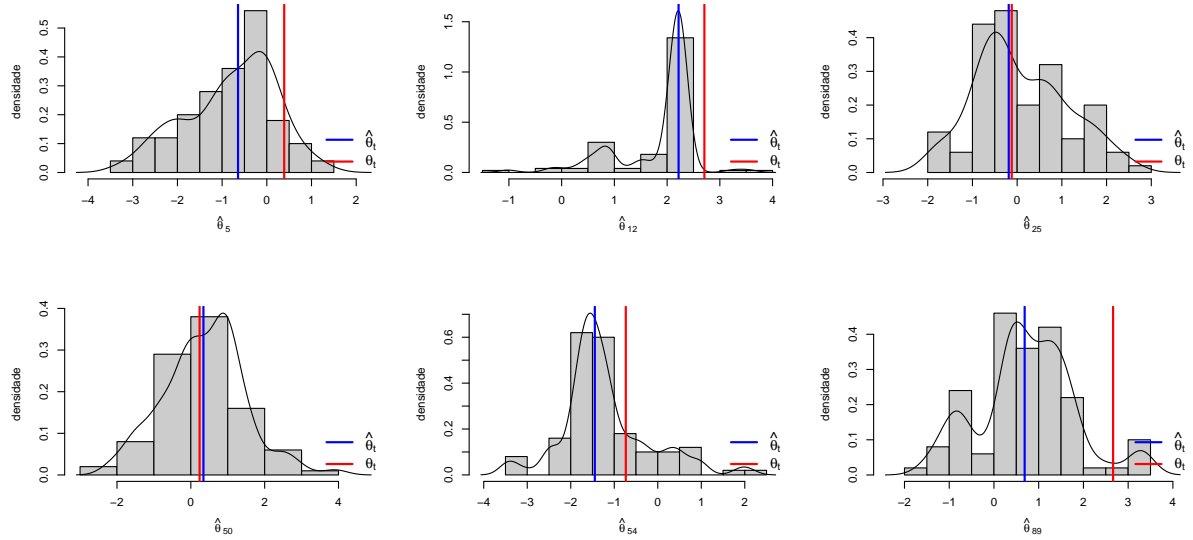


Figura 4.11: Densidades do processo de estimação FP auxiliar

cor azul) e para o valor real do estado latente θ_t (linha na cor vermelha).

ARCH

Uma grande classe de modelos de séries temporais são os *Autoregressive conditional heteroskedasticity*, **ARCH** (Bollerslev, 1986). A figura 4.12 mostra um processo ARCH com erros independentes gaussianos.

Para esse modelo, a equação de observação foi gerada por

$$y_t \sim N(\theta_t, \sigma^2),$$

e a equação de evolução dos estados por

$$\theta_t \sim N(0, \beta_0 + \beta_1 \theta_{t-1}^2).$$

A figura 4.13 mostra o resultado da aplicação do FP auxiliar para a estimação dos estados latente do processo dinâmico ARCH apresentado. São apresentados as estimativas pontuais (linha na cor vermelha) e intervalares (região apresentada na cor cinza) para para o estado latente θ_t (apresentada pelo símbolo asterisco).

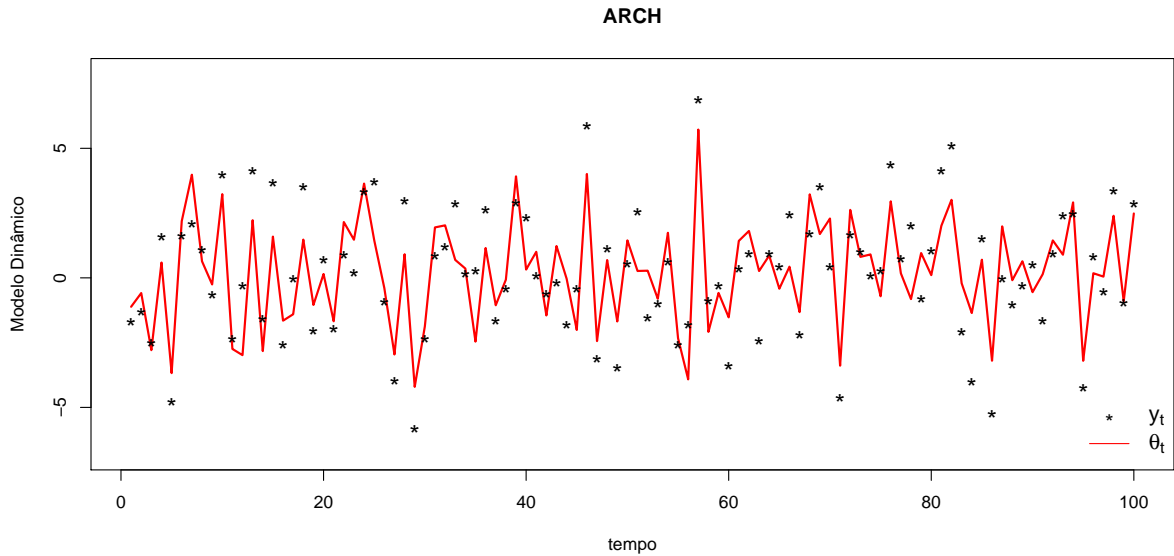


Figura 4.12: Processo ARCH

As estimativas via Filtro de Partículas apresentaram boas estimativas para os estados latentes. No entanto, para esta simulação, as estimativa intervalares de 95%, feita por meio dos quantis $\{2.5\%, 97.5\%\}$ das partículas, apresentou uma cobertura para 89% dos estados latentes.

Para verificar a real cobertura das estimativas intervalares para os estados latentes do modelo apresentado, foram realizadas 1.000 simulação do processo de estimação via FP auxiliar. A simulação apontou que as estimativas intervalares apresentam uma cobertura, em média, de 90.88%.

4.4 Filtro de Partículas Básico

Nesta seção, será apresentado uma definição mais ampla dos recursos metodológicos utilizados na técnica **Filtro de Partículas**. Tal definição incorpora as ideias e os desenvolvimentos dos artigos seminais apresentados nesta dissertação, complementando as discussões sobre a técnica. Petris et al. (2009) considera tal

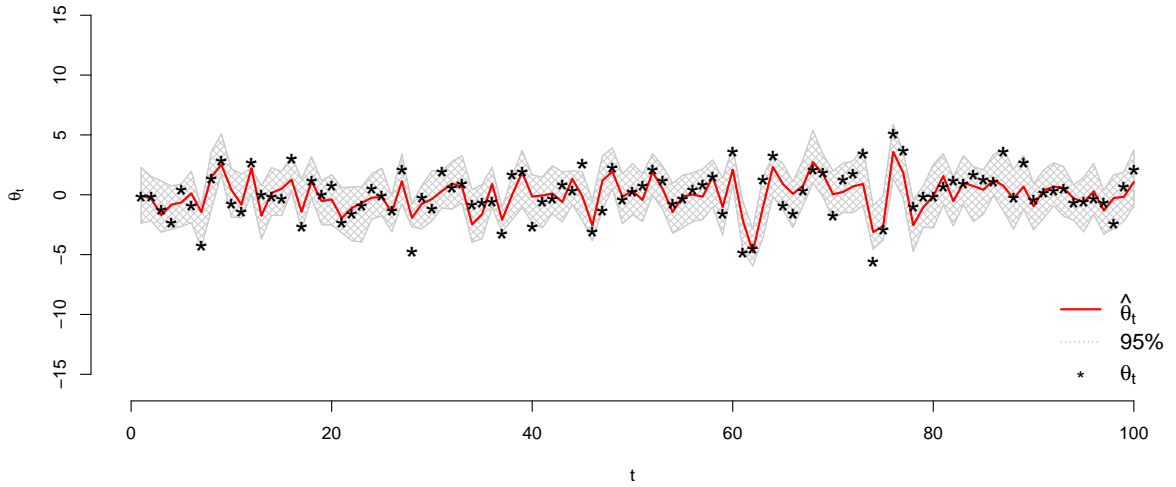


Figura 4.13: Estimação para o modelo ARCH

definição como **Filtro de Partículas Básico**.

As probabilidades associadas às partículas são a base para o processo de dependência da cadeia. Como visto anteriormente, a *posteriori* do estado $\pi(\theta_{0:t}|y_{1:t})$ pode ser decomposta, em cada instante t , pelo método de **Amostragem de Importância Sequencial**. A cada instante, os pesos para as partículas são representados de forma recursiva por meio de uma **função auxiliar de importação**, $g(\theta_{0:t}|y_{1:t})$, (Petris et al., 2009):

$$w_t \propto \frac{\pi(\theta_{0:t}|y_{1:t})}{g(\theta_{0:t}|y_{1:t})} \propto \frac{\pi(\theta_{0:t}, y_t|y_{1:t-1})}{g(\theta_{0:t}|y_{1:t})}. \quad (4.42)$$

De forma semelhante à apresentada na Seção 3.2, pode-se reescrever w_t , utilizando a estrutura recursiva da função auxiliar e do processo dinâmico, como:

$$w_t \propto \frac{\pi(\theta_t, y_t|\theta_{0:t-1}, y_{1:t-1})\pi(\theta_{0:t-1}|y_{1:t-1})}{g_{t-1}(\theta_t|\theta_{0:t-1}, y_{1:t})g_{t-1}(\theta_{0:t-1}|y_{1:t-1})}. \quad (4.43)$$

A última parte da equação (4.43) é exatamente a representação para os pesos no tempo $t - 1$ (w_{t-1}). Além disso, rescrevendo $\pi(\theta_t, y_t|\theta_{0:t-1}, y_{1:t-1})$ utilizando as

propriedades recursivas de processos Markovianos, pode-se representar a equação dos pesos por:

$$w_t \propto \frac{\pi(y_t|\theta_t)\pi(\theta_t|\theta_{t-1})}{g_{t|t-1}(\theta_t|\theta_{0:t-1}, y_{1:t})} w_{t-1}. \quad (4.44)$$

Com a função $g_{t|t-1}(\theta_t|\theta_{0:t-1}, y_{1:t})$, conhecida como **função de transição de importância**, gera-se as propostas de partículas. Os tipos de FP são definidos pelo tipo de equação $g_{t|t-1}(\cdot)$ escolhida.

Ao longo do processo *online* de estimação, a distribuição dos pesos pode degenerar-se. Isso acontece quando existe alguma partícula com peso próximo de 1. Considerando w_t^i o peso associado à partícula i no instante t , adota-se, então, um critério para controlar o fenômeno de tal processo degenerativo. Este consiste em monitorar, ao longo do tempo, o “tamanho efetivo da amostra”, N_{eff}^t , definido por:

$$N_{eff}^t = \frac{1}{\sum_{i=1}^N (w_t^i)^2}.$$

Esta quantidade varia entre N , quando todos os pesos são iguais, e 1, quando um dos pesos é próximo de 1. Quando N_{eff} atinge um valor inferior a um determinado valor N_0 , é recomendável realizar o processo de renovação do vetor de partículas, isto é, proceder à **reamostragem** (vide seção 3.3). O passo da reamostragem não modifica o valor esperado da distribuição aproximada $\hat{\pi}_t$, mas aumenta sua variância de Monte Carlo. Petris et al. (2009) indicam a utilização de $n_0 = N/2$ como critério para proceder ao processo de regeneração das partículas. Tal regeneração é realizada reamostrando-se as partículas utilizando os pesos w_t^i . Em seguida, os pesos das partículas reamostradas são redefinidos para $w_t^i = 1/N$.

Como mencionado, o processo de reamostragem não muda o valor pontual da

estimativa de π_t e proporciona a diversificação desejada. Como penalidade para essa técnica, a variância da distribuição do espaço de estado aproximada pelas partículas é inflacionada (Petris et al., 2009).

O processo a seguir refere-se ao modelo apresentado na seção 2.8, isto é, ao modelo dinâmico Linear Gaussiano de primeira ordem. A equação das observações é:

$$y_t \sim N(\theta_t, V), \quad (4.45)$$

e a equação dos estados latentes é:

$$\theta_t \sim N(\theta_{t-1}, W). \quad (4.46)$$

A figura 4.14 mostra uma simulação do modelo descrito acima.

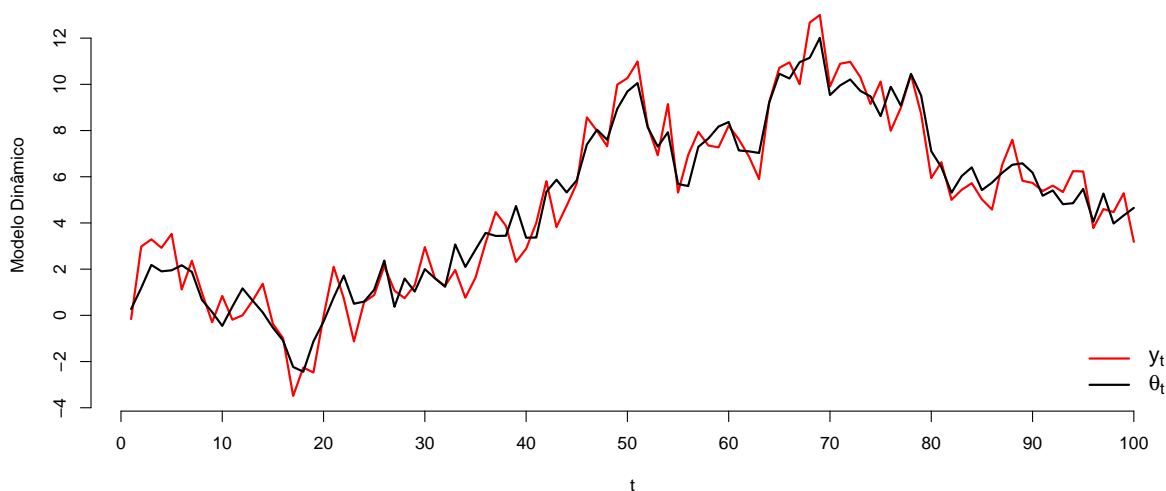


Figura 4.14: Processo dinâmico: Gaussiano linear de ordem 1

A seguir serão apresentados dois filtros de partícula, segundo a escolha da **função de transição de importância** ($g_{t|t-1}(\cdot)$), para esse modelo. Como tal modelo possui solução analítica, os resultados podem ser comparados com o filtro de Kalman.

4.4.1 FP Simples

O primeiro filtro a ser apresentado soluciona da forma mais simples a equação (4.44). Esse filtro simples utiliza como função de transição de importância a equação de evolução dos estados latentes. Dessa forma, como descrito na seção 3.3, a equação de ponderação das partículas é calculada mais facilmente. É importante notar que a estrutura de dependência do modelo dinâmico apresentado é incorporado pelo processo iterativo de atualização dos pesos.

Dessa forma, seleciona-se a função de transição de importância:

$$g_{t|t-1}(\theta_t|\theta_{0:t-1}, y_{1:t}) = q_{opt}(\theta_t|\theta_{t-1}) = \pi(\theta_t|\theta_{t-1}).$$

De forma similar à apresentada por Gordon et al. (1993), esse é um filtro com equação de transição “às cegas”. Substituindo tal escolha na equação dos pesos, tem-se:

$$w_t \propto \frac{\pi(y_t|\theta_t)\pi(\theta_t|\theta_{t-1})}{g_{t|t-1}(\theta_t|y_{1:t}, \theta_{1:t-1})} w_{t-1}.$$

A atualização dos pesos é feita por:

$$w_t \propto \frac{\pi(y_t|\theta_t)\pi(\theta_t|\theta_{t-1})}{\pi(\theta_t|\theta_{t-1})} w_{t-1}.$$

Tem-se, então, que a atualização dos pesos da cadeia de partículas é feita por meio da equação (4.45), ou seja, a equação das observações. Dessa forma, os pesos em t são calculados por:

$$w_t^i = \pi(y_t|\theta_t^i)w_{t-1}^i.$$

A principal diferença entre o FP mais geral, apresentado por Petris et al. (2009), e filtro proposto por Gordon et al. (1993), mostrado na seção 4.2, é que o processo

recursivo de cálculo dos pesos é substituído pelo processo de **reamostragem**.

Algoritmo relativo ao FP simples

- 1.1 Gerar distribuição inicial do estado latente, com $1 < i < N$:

$$\theta_0^i \sim \pi_0;$$

- 1.2 Definir os pesos:

$$w_0^i = \frac{1}{N};$$

- 2 Amostrar partículas da distribuição auxiliar, com $1 < t < T$:

$$\theta_t^i \sim f(\cdot | \theta_{t-1}^i) \sim N(\theta_{t-1}^i, W);$$

- 3 Atualizar pesos com a nova observação:

$$w_t^i = w_{t-1}^i f(y_t | \theta_t^i) = w_{t-1}^i f_N(y_t; \theta_t^i, V);$$

- 4 Normalizar os pesos:

$$\tilde{w}_t^i = \frac{w_t^i}{\sum_i w_t^i};$$

- 5 Reamostrar as partículas utilizando \tilde{w}_t^i , nos casos de valores N_{eff}^t inferiores a $N_0 = N/2$, sendo:

$$N_{eff}^t = \frac{1}{\sum (\tilde{w}_t^i)^2};$$

- 5.1 Nesses casos, redefinir os pesos $\tilde{w}_t^i = 1/N$.

- 6 Estimação:

$$\hat{\theta}_t = \sum_i^N f(\theta_t^i) \tilde{w}_t^i.$$

O processo é repetido até a conclusão da filtragem, em $t = T$.

A figura 4.15 mostra as estimativas para o estado latente do modelo proposto pelo FP simples e as estimativas do Filtro de Kalman. É possível perceber bom ajuste do Filtro de Partículas, e a proximidade do resultado analítico do Filtro de Kalman.

4.4.2 FP ótimo

Escolhendo a distribuição de importância como:

$$g_{t|t-1}(\theta_t | \theta_{0:t-1}, y_{1:t}) = q_{opt}(\theta_t | \theta_{t-1}) = \pi(\theta_t | y_t, \theta_{t-1}) \quad (4.47)$$

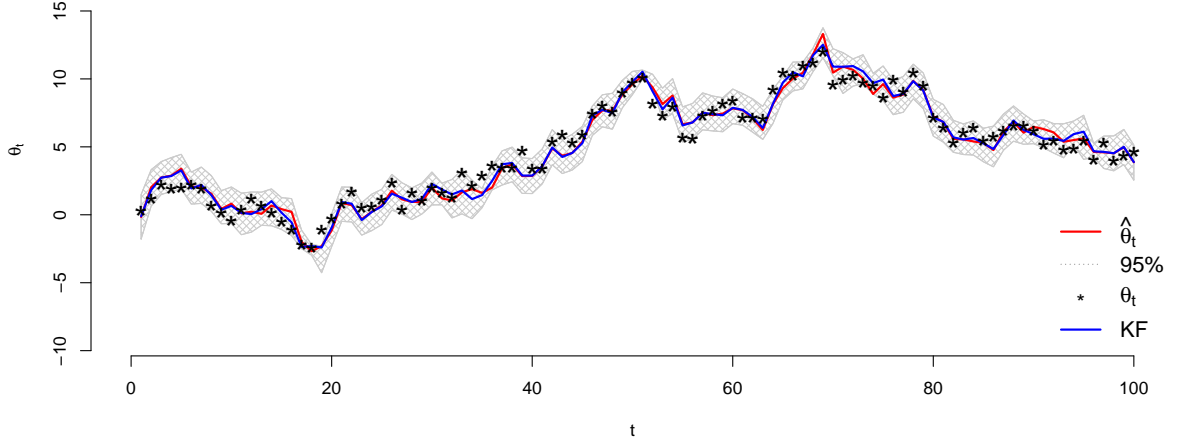


Figura 4.15: FP simples: Gaussiano linear de ordem 1

e substituindo tal distribuição na equação de atualização dos pesos, tem-se:

$$w_t \propto \frac{\pi(y_t|\theta_t)\pi(\theta_t|\theta_{t-1})}{g_{t|t-1}(\theta_t|y_{1:t}, \theta_{1:t-1})} w_{t-1}.$$

Sabe-se que:

$$\pi(\theta_t|y_t, \theta_{t-1}) = \frac{\pi(y_t|\theta_t)\pi(\theta_t|\theta_{t-1})}{\pi(y_t|\theta_{t-1})}.$$

Dessa forma, tem-se que a atualização dos pesos é feito por:

$$w_t \propto \frac{\pi(y_t|\theta_t)\pi(\theta_t|\theta_{t-1})}{\frac{\pi(y_t|\theta_t)\pi(\theta_t|\theta_{t-1})}{\pi(y_t|\theta_{t-1})}},$$

em que

$$w_t = \pi(y_t|\theta_{t-1})w_{t-1},$$

isto é, a atualização dos pesos é feita utilizando a equação das observações no tempo t utilizando o estado latente, θ_{t-1} no tempo $t - 1$.

Algoritmo relativo ao FP ótimo

Utilizando o modelo dinâmico linear gaussiano de primeira ordem apresentado na primeira parte do capítulo, tem-se o seguinte algoritmo:

1.1 Gerar distribuição inicial do estado latente, com $1 < i < N$:

$$\theta_0^i \sim \pi_0;$$

1.2 Definir os pesos:

$$w_0^i = \frac{1}{N};$$

2 Amostrar partículas da distribuição auxiliar, com $1 < t < T$:

$$\theta_t^i \sim f(\cdot | \theta_{t-1}^i) = \pi(\theta_t | y_t, \theta_{t-1}^i).$$

2.1 Isso e feito reescrevendo a distribuição auxiliar como:

$$\pi(\theta_t | y_t, \theta_{t-1}^i) \propto \pi(y_t | \theta_t) \pi(\theta_t | \theta_{t-1}^i).$$

2.2 Dessa forma, se tem que:

$$\mu_t^i = \theta_{t-1}^i + \frac{W}{W+V}(y_t - \theta_{t-1}^i).$$

2.3 Amostrar partículas da distribuição:

$$\pi(\theta_t | y_t, \theta_{t-1}^i) \stackrel{d}{=} N\left(\mu_t^i, \frac{WV}{W+V}\right).$$

3 Atualizar pesos com a nova observação:

$$w_t^i = w_{t-1}^i f(y_t | \theta_t^i) = w_{t-1}^i f_N(y_t; \theta_t^i, V);$$

4 Normalizar os pesos:

$$\tilde{w}_t^i = \frac{w_t^i}{\sum_i w_t^i};$$

5 Reamostrar as partículas utilizando \tilde{w}_t^i , nos casos de valores N_{eff}^t inferiores a $N_0 = N/2$, sendo:

$$N_{eff}^t = \frac{1}{\sum(\tilde{w}_t^i)^2};$$

5.1 Nesses casos, redefinir os pesos $\tilde{w}_t^i = 1/N$.

6 Estimação:

$$\hat{\theta}_t = \sum_i^N f(\theta_t^i) \tilde{w}_t^i.$$

O processo é repetido até a conclusão da filtragem, em $t = T$.

A figura 4.16 compara as estimativas para o estado latente utilizando o FP ótimo e o filtro de Kalman. É possível perceber a proximidade das estimativas calculadas pelos dois métodos.

Para verificar a proximidade de tais estimativas e compará-las com as estimativas encontradas pelo FP básico, a figura 4.17 apresenta uma comparação dos desvio

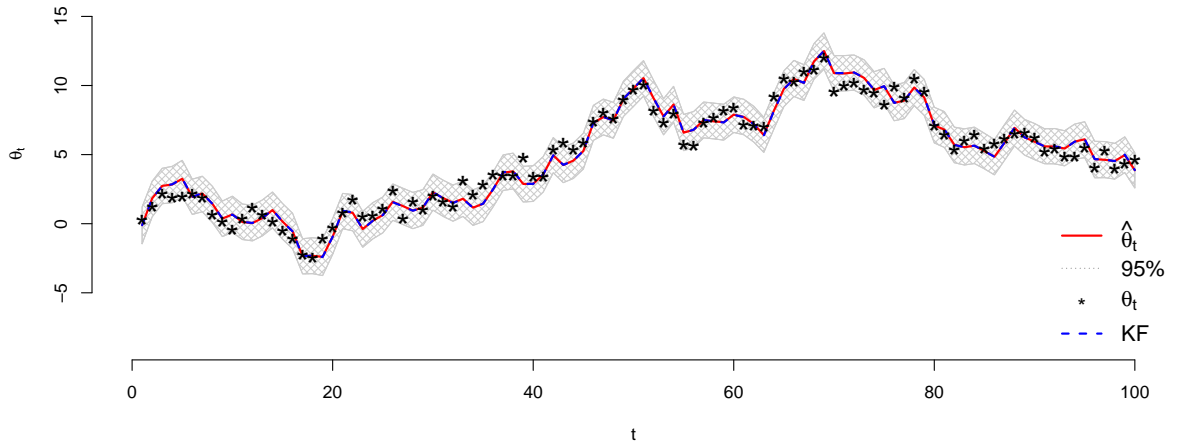


Figura 4.16: FP ótimo: Gaussiano linear de ordem 1

quadráticos, com relação aos valores reais dos estados latentes, durante o processo de estimação de $\theta_{1:T}$. Para colaborar com a comparação das metodologias, os três boxplots apresentam os valores dispostos na tabela de forma visual.

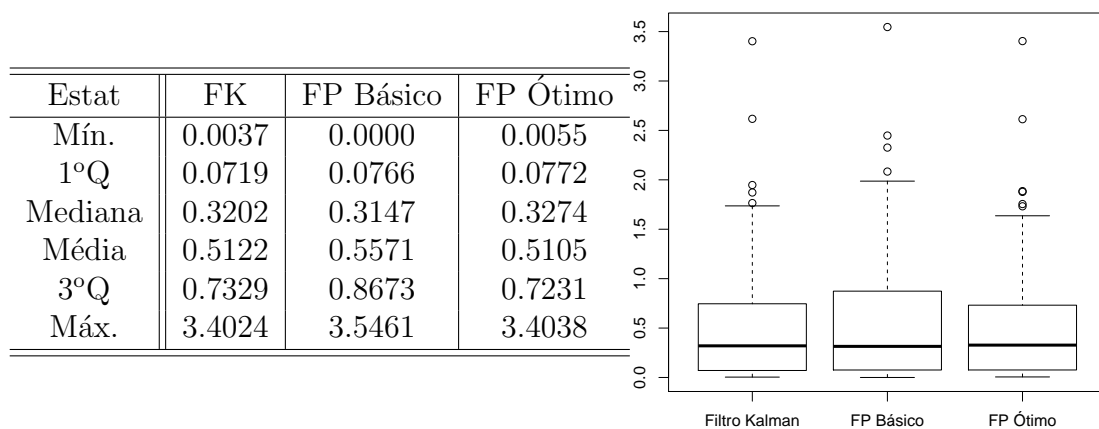


Figura 4.17: Tabela e Boxplot dos desvios quadráticos, com relação ao valor verdadeiro dos estados latentes $\theta_{1:T}$, das metodologias Filtro de Kalman, FP básico e FP ótimo.

Com os recursos apresentados na figura 4.17, é possível perceber a proximidade entre as estimativas do Filtro de Kalman e do FP ótimo, com média e mediana

muito próximas. Para a metodologia do FP básico é possível perceber uma maior imprecisão das estimativas, isto é, maior variabilidade dos desvios quadráticos. Isso é notado observado estimativas do 1º e 3º quartil, com valores superiores às demais metodologias.

4.5 Suavização

Ao final do processo numérico *online* proposto pelo Filtro de Partículas, pode-se realizar uma estimação *offline* utilizando o conhecimento de todas as informações da amostra. Como abordado na seção 2.8.2, esse processo é conhecido como suavização.

Ao final do processo de filtragem, considere a seguinte decomposição da distribuição conjunta $p(x_{1:T}|y_{1:T})$.

$$\begin{aligned} p(x_{1:T}|y_{1:T}) &= p(x_T|y_{1:T}) \prod_{n=1}^{T-1} p(x_n|x_{n+1}, y_{1:T}) \\ &= p(x_T|y_{1:T}) \prod_{n=1}^{T-1} p(x_n|x_{n+1}, y_{1:n}), \end{aligned} \tag{4.48}$$

esse resultado é garantido pelas propriedades de independência Markovianas A.1 e A.2 (vide seção 2.2). Por meio delas, pode-se escrever:

$$\begin{aligned} p(x_n|y_{1:T}, x_{n+1:T}) &= p(x_n|y_{1:n}, x_{n+1}) \\ &= \frac{p(x_{n+1}|x_n)p(x_n|y_{1:n})}{p(x_{n+1}|y_{1:n})} \\ &\propto p(x_n|y_{1:n})p(x_{n+1}|x_n) \end{aligned} \tag{4.49}$$

Utilizando a estimativa para a *posteriori* $\hat{p}(x_n|y_{1:n})$, obtida no processo de filtragem, a expressão (4.49) pode ser obtida pela aproximação

$$p(x_n|x_{n+1}, y_{1:T}) \approx \sum_{i=1}^N w_{n|n+1}^i \delta_{x_n^i}(x_n), \tag{4.50}$$

sendo $w_{n|n+1}^i$ os pesos de suavização, expressos pela equação (4.51), (Doucet et al., 2004).

$$w_{n|n+1}^i = \frac{w_n^i p(x_{n+1}|x_n^i)}{\sum_{j=1}^N w_n^j p(x_{n+1}|x_n^j)}. \quad (4.51)$$

Os pesos w_n^i são os valores obtidos no processo de filtragem.

Com isso, podem ser obtidas novas estimativas para o estado latente x_n por meio de $\tilde{x}_n \sim \hat{p}(x_n|\tilde{x}_{n+1:T}, y_{1:T})$, em um processo recursivo na direção oposta à do processo de filtragem inicial.

Mais detalhes do processo de suavização são encontrados em Doucet e Johansen (2008) e Doucet et al. (2004).

Capítulo 5

Estimação de parâmetros

Nesta dissertação, até este ponto, discutiu-se a técnica de Filtro de Partículas para a estimação dos estados latentes no caso de parâmetros fixos e conhecidos. Neste capítulo serão discutidas as principais técnicas para o Filtro de Partículas que propiciam a estimação de parâmetros estáticos no modelo dinâmico.

5.1 Liu e West (2001)

Liu e West (2001) propõem uma metodologia para a estimação de parâmetros estáticos, no modelo dinâmico, utilizando a ideia de suavização pelo método de kernel ((West, 1993a) e (West, 1993b)). Para a estimação dos estados latentes, o autor utiliza o Filtro de Partículas Auxiliar de Pitt e Shephard (1999).

Considere uma modelagem para a série $\{y_t : j = 1, \dots, N\}$ por modelos dinâmicos com estimação via Filtro de Partículas, respectivamente, do vetor de estados latentes e dos parâmetros, $\{x_t^j, \theta_t^j : j = 1, \dots, N\}$, associados ao peso $\{w_t^j : j = 1, \dots, N\}$. Com tais partículas constrói-se a aproximação para a distribuição a *posteriori* no tempo t , $p(x_t, \theta | D_t)$. Para o processo de estimação, a *posteriori* conjunta no tempo $t + 1$ é

representada por:

$$\begin{aligned} p(x_{t+1}, \theta | D_{t+1}) &= p(y_{t+1} | x_{t+1}, \theta) p(x_{t+1}, \theta | D_t) \\ &= p(y_{t+1} | x_{t+1}, \theta) p(x_{t+1} | \theta, D_t) p(\theta | D_t). \end{aligned} \tag{5.1}$$

Considere a substituição de θ por θ_t para o processo numérico de estimação. Apesar de ser estático, tal representação do vetor de parâmetros auxilia no processo estimação ao longo do tempo. Considere a seguinte representação artificial da evolução de t para $t + 1$ das estimativas dos parâmetros:

$$\theta_{t+1} = \theta_t + \zeta_{t+1},$$

com

$$\zeta_{t+1} \sim N(0, W_{t+1}),$$

sendo W_{t+1} uma matriz de variância e θ_t e ζ_{t+1} condicionamente independentes a D_t . Considerando tal forma de representar o vetor de parâmetros, a estimação da densidade a *posteriori* $p(\theta | D_t)$ é feita por densidade suavizada via kernel, (Liu e West, 2001):

$$p(\theta | D_t) \approx \sum_{j=1}^N w_t^j N(\theta | m_t^j, h^2 V_t). \tag{5.2}$$

Dessa forma, a distribuição artificial dos parâmetros θ é representada pela mistura ponderada por w_t^j de normais com médias m_t^j e matriz de variância $h^2 V_t$. Liu e West (2001) sugerem que a constante h (parâmetro de suavização) seja escolhida como uma função suave e decrescente de N .

Liu e West (2001) consideram:

$$m_t^j = a\theta_t^j + (1 - a)\bar{\theta}_t, \tag{5.3}$$

sendo

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

$$\bar{\theta}_t = \sum_{j=1}^N \frac{\theta_t^j}{N}$$

$$V_t = \frac{1}{N} \sum_{j=1}^N (\theta_t^j - \bar{\theta}_t)(\theta_t^j - \bar{\theta}_t)'$$

As constantes a e h medem, respectivamente, o grau de “encolhimento” (shrinkage) e superdispersão da mistura. Com tal método de “encolhimento”, a média $\bar{\theta}_t$ é mantida, mas a correspondente variância V_t corrige o problema de superdispersão, que é típico dos métodos de kernel convencionais.

Algoritmo de Liu e West (2001)

Assumindo $\{(x_{t-1}, \theta_{t-1}, w_{t-1})^i, i = 1, \dots, N\}$ a aproximação para $p(x_{t-1}, \theta | D_{t-1})$ no tempo $t - 1$, a estimação dos parâmetros e filtragem é feita por meio do seguinte algoritmo:

- 1 Identificar os pontos

$$\mu_t^i = E(x_t | x_{t-1}^i, \theta),$$

$$m_t^i = a\theta_{t-1}^i + (1 - a)\bar{\theta}_{t-1}$$

- 2 Amostrar uma partícula auxiliar k^i com probabilidade proporcional:

$$p(k^i = k) \propto w_{t-1}^i p(y_t | \mu_t^k, m_{t-1}^k)$$

- 3 Amostre um novo vetor de parâmetros θ_t^i :

$$\theta_t^i \sim N(m_{t-1}^{k^i}, (1 - a^2)V_{t-1}),$$

sendo $V_t = \sum_{i=1}^N (\theta_{t-1}^i - \bar{\theta}_{t-1})(\theta_{t-1}^i - \bar{\theta}_{t-1})' w_{t-1}^i$.

- 4 Obtenha uma amostra x_t^i dos estados latentes a partir da equação do sistema $p(\cdot | x_{t-1}^{k^i}, \theta_t^i)$:

- 5 Compute os pesos:

$$w_t^i \propto \frac{p(y_t | x_t^i, \theta_t^i)}{p(y_t | \mu_t^{k^i}, m_{t-1}^{k^i})}$$

- 6 Gerar $p(x_t, \theta | D_t)$ e amostrando de $\{(x_t, \theta_t, w_t)^i, i = 1, \dots, N\}$:

5.2 Storvik (2002)

Storvik (2002) assume que a distribuição a *posteriori* do vetor de parâmetros estáticos θ , condicionados a x_t e y_t dependem de um conjunto de estatísticas suficientes de baixa densidade que são atualizadas recursivamente.

Storvik (2002) acrescenta uma equação de **atualização do vetor de parâmetros** ao processo de estimação dos estados latentes. Dessa forma, em cada iteração do processo de estimação, realiza-se a atualização dos parâmetros estáticos.

Em casos mais específicos, em que há forma fechada para a distribuição conjunta a *posteriori* dos parâmetros estáticos, West e Harrison (1997) apresentam as equações do processo de estimação. Quando não há, a técnica apresentada por Storvik pode ser utilizada caso se conheça as distribuições marginais dos parâmetros a *posteriori*.

O processo de atualização é feito por meio da técnica conhecida como *Rao-Blackwellization*, utilizando estatísticas suficientes para encontrar a distribuição a *posteriori* dos parâmetros estáticos (Doucet et al., 2004).

Para entender a metodologia apresentada, considere um modelo dinâmico geral contendo uma equação do processo latente

$$x_k \sim p(x_k | x_{k-1}; \theta)$$

e uma equação que modele os valores observados

$$y_k \sim p(y_k | x_k; \theta).$$

O parâmetro θ é um vetor de constantes que auxiliam a modelagem dos estados

latentes. Como apresentado no Capítulo 3, com a técnica filtro de partícula simulam-se valores representantes da distribuição a *priori* no tempo k , que, por sua vez, são evoluídas para $k+1$ por uma probabilidade proporcional à verossimilhança. Também foi visto que um dos ganhos no uso do Filtro de Partículas está na possibilidade de realizar o processo de estimação, de forma *on-line*, por meio dos processos de atualização e evolução.

Considere $S_k = \{(x_{1:k}^i, w_{1:k}^i)\}$ a representação da partícula i e seu respectivo peso na distribuição aproximada gerada, sendo:

$$w_k = \frac{p(x_{0:k}|y_{1:k})}{f_0(x_0) \prod f(x_l|x_{l-1}, y_{1:l})} \quad (5.4)$$

Dessa forma, pode-se representar $p(x_{1:k-1}|y_{1:k-1})$ por meio do par mencionado, fazendo:

$$\hat{p}(x_{1:k-1}|y_{1:k-1}) = \frac{\sum_i^N w_{k-1}^i I(x_{1:k-1} = x_{1:k-1}^i)}{\sum_i^N w_{k-1}^i}. \quad (5.5)$$

Com os parâmetros estáticos conhecidos, a distribuição a *posteriori* era obtida pelas seguintes distribuições:

$$\begin{aligned} p(x_{1:k}|y_{1:k}) &\propto p(x_{1:k-1}|y_{1:k-1})p(x_k|x_{1:k-1})p(y_k|x_k) \\ &= \hat{p}(x_{1:k-1}|y_{1:k-1})p(x_k|x_{1:k-1})p(y_k|x_k). \end{aligned} \quad (5.6)$$

Um novo conjunto de partículas S_k pode ser obtido ao simularmos de tal distribuição aproximada. Um procedimento de rejeição possível para simular dados a partir de (5.6) consiste em amostrar $x_{1:k}$ a partir de $\hat{p}(x_{1:k}|y_{1:k-1}) = \hat{p}(x_{1:k-1}|y_{1:k-1})p(x_k|x_{k-1})$ e aceitar a amostra com probabilidade proporcional a $p(y_k|x_k)$. Este procedimento pode ser repetido N vezes de modo a se obter um novo conjunto de partículas $S_k = \{x_k^i, \dots, x_k^N\}$.

Quando há parâmetros estáticos desconhecidos, acrescenta-se ao processo acima a distribuição a *posteriori* relativa a θ . A metodologia proposta por Storvik (2002) sugere que as estimativas de θ no tempo k dependam apenas de estatísticas suficientes de $x_{1:k}$ e $y_{1:k}$, representadas por T_k , que são atualizadas a cada k .

Considerando C a constante normalizadora, a posteriori dos parâmetros e estados latentes pode ser obtida por meio de:

$$\begin{aligned} p(x_{1:k}, \theta | y_{1:k}) &= C \times p(x_{1:k}, \theta, y_k | y_{1:k-1}) \\ &\propto p(x_{1:k-1} | y_{1:k-1}) \times p(\theta | x_{1:k-1}, y_{1:k-1}) \times p(x_k | x_{1:k-1}, \theta) \times p(y_k | x_k, \theta) \end{aligned} \quad (5.7)$$

As simulações a partir de (5.7) podem ser realizadas ao simular $x_{1:k-1}$ a partir de $\hat{p}(x_{1:k-1} | y_{1:k-1})$, θ a partir de $p(\theta | T_{k-1})$, x_k a partir de $p(x_k | x_{1:k-1}, \theta)$ e aceitar com probabilidade proporcional a $p(y_k | x_k, \theta)$.

Essa técnica tem como principal característica fazer com que a distribuição de atualização para θ , $p(\theta | y_{1:k}, x_{1:k})$, dependa do conjunto $(y_{1:k}, x_{1:k})$, exclusivamente, por meio das estatísticas T_k . A redução da dimensão e complexidade envolvida em $(y_{1:k}, x_{1:k})$ representa ganhos computacionais para o processo de estimação, (Storvik, 2002).

Algoritmo de Storvik (2002)

- 1 Amostrar $\theta \sim g_{k,1}(\theta | x_{1:k-1}^i, D_k)$.
- 2 Amostrar $x_k^{*i} \sim g_{k,2}(x_k | x_{1:k-1}^i, y_k, \theta)$.
- 3 Computar os pesos amostrais de importância:

$$w_k^i \propto w_{k-1}^i \frac{p(\theta | T_k^i) p(x_k^{*i} | x_{k-1}^i, \theta) p(y_k | x_k^{*i}, \theta)}{g_{k,1}(\theta | x_{1:k-1}^i, D_k) g_{k,2}(x_k | x_{1:k-1}^i, y_k, \theta)}$$

- 4 Reamostrar as partículas x^{*i} com probabilidade proporcional a w_k^i .

As distribuições $g_{k,1}(\theta|x_{1:k-1}^i, D_k)$ e $g_{k,2}(x_k|x_{1:k-1}^i, y_k, \theta)$ são distribuições geradoras de candidatos de θ e x_k , respectivamente. Tipicamente escolhe-se $g_{k,1}(\theta|x_{1:k-1}^i, D_k) = p(\theta|T_{k-1})$ e $g_{k,2}(x_k|x_{1:k-1}^i, y_k, \theta) = p(x_t|x_{t-1}, \theta)$. Nesse caso, tem-se um processo de filtragem de Gordon et al. (1993).

5.3 Exemplo de aplicação do método de Storvik (2002)

Para comparar as metodologias de filtragem dos estados latentes e estimação de parâmetros desconhecidos, apresentadas nessa dissertação, considere o modelo dinâmico apresentado por Storvik (2002), com pequenas alterações:

$$\begin{aligned} x_t &\sim N(\alpha + \beta x_{t-1}, \sigma^2) \\ z_t &\sim \text{Binom}(r, \text{logit}^{-1}(x_t)). \end{aligned} \tag{5.8}$$

Na equação de evolução, o parâmetro x_t representa o estado latente do processo dinâmico, e os parâmetros α , β e σ^2 são desconhecidos e estáticos. A parte observável do processo é representada por z_t , e o parâmetro r , presente na equação de observação, é conhecido.

Nesse modelo, a distribuição de interesse em cada t é a *posteriori* $p(\alpha, \beta, \sigma^2|D_t)$. Para utilizar a metodologia proposta por Storvik (2002) é necessário que as distribuições marginais da *posteriori* de interesse sejam conhecidas. Para a metodologia de Liu e West (2001), a única restrição é que os espaços paramétricos dos parâmetros tenham o mesmo espaço paramétrico de uma variável com distribuição gaussiana.

Considere as *prioris*, $\sigma^2 \sim IG(v_0/2, d_0/2)$, $\alpha \sim N(0, \sigma_\alpha^2)$ e $\beta \sim N(0, \sigma_\beta^2)$. Utili-

zando as propriedades Bayesianas, a *posteriori* de interesse pode ser reescrita como:

$$\begin{aligned}
\pi(\sigma^2, \beta, \alpha, x_{1:t} | z_{1:t}) &\propto p(z_{1:t} | \sigma^2, \beta, \alpha, x_{1:t}) \times p(x_{1:t} | \beta, \alpha, \sigma^2) \times p(\beta) p(\alpha) p(\sigma^2) \\
&\propto \left[\prod_{i=1}^t p(z_i | x_i, \sigma^2, \beta, \alpha) \right] \left[\prod_{i=1}^t p(x_i | x_{i-1}, \beta, \alpha, \sigma^2) \right] \times p(\beta) p(\alpha) p(\sigma^2) \\
&\propto \left[\prod_{i=1}^t \text{Bin}(z_i | r, \text{logit}^{-1}(x_t)) \right] \left[\prod_{i=1}^t N(x_i | \alpha + \beta x_{i-1}, \sigma^2) \right] \times N(0, \sigma_\beta^2) \\
&\quad \times N(0, \sigma_\alpha^2) \times IG\left(\frac{v_0}{2}, \frac{d_0}{2}\right).
\end{aligned}$$

Para esse modelo, as distribuições marginais dos parâmetros no tempo t são:

$$\begin{aligned}
p(\alpha | \beta, \sigma^2, x_{0:t}, z_{1:t}) &= p(\alpha | \beta, \sigma^2, x_{0:t}) \stackrel{d}{=} N(m_\alpha; \sigma_\alpha^2), \\
p(\beta | \alpha, \sigma^2, x_{0:t}, z_{1:t}) &= p(\beta | \alpha, \sigma^2, x_{0:t}) \stackrel{d}{=} N(m_\beta; \sigma_\beta^2), \\
p(\sigma^2 | \beta, \alpha, x_{0:t}, z_{1:t}) &= p(\sigma^2 | \beta, \alpha, x_{0:t}) \stackrel{d}{=} IG(v_t/2, d_t/2),
\end{aligned} \tag{5.9}$$

sendo

$$m_\alpha = \frac{\sigma_\alpha^2 (\sum_{i=1}^t x_i - \beta \sum_{i=1}^t x_{i-1})}{\sigma^2 + \sigma_\alpha^2 t},$$

$$\sigma_\alpha^2 = \frac{\sigma^2 \sigma_\alpha^2}{\sigma^2 + \sigma_\alpha^2},$$

$$m_\beta = \frac{(\sum_{i=1}^t x_i x_{i-1} - \sum_{i=1}^t x_{i-1}) \sigma_\beta^2}{(\sigma^2 + \sigma_\beta^2 \sum_{i=1}^t x_{i-1}^2)},$$

$$\sigma_\beta^2 = \frac{\sigma_\beta^2 \sigma^2}{(\sigma^2 + \sigma_\beta^2 \sum_{i=1}^t x_{i-1}^2)},$$

$$v_t = v_0 + t,$$

$$d_t = \sum_{i=1}^t x_i^2 + t\alpha^2 + \beta^2 \sum_{i=1}^t x_{i-1}^2 - 2 \left(\alpha \sum_{i=1}^t x_i + \beta \sum_{i=1}^t x_i x_{i-1} \right) + 2\beta\alpha \sum_{i=1}^t x_{i-1}.$$

Tais hiperparâmetros são recalculados a cada nova observação. Dessa forma, são obtidas novas estimativas para os parâmetros α , β e σ^2 , e para o estado latente x_t .

As análises referentes a este problema serão discutidas na próxima seção. Será feita, também, uma comparação dos resultados inferenciais com aqueles obtidos com o método proposto por Liu e West (2001).

5.4 Comparando os métodos

Para realizar a comparação dos métodos de estimação, foram simulados 100 instantes de observação do modelo proposto, tendo, assim, 100 pares de observações e espaços de estados, $\{y_t, x_t\}$, com $t = 1, \dots, 100$. Para a simulação, o estado latente inicial da cadeia foi definido como $x_0 = 0$, foram geradas $N = 1.000$ partículas para cada ponto de estimação e o processo de estimação foi replicado 100 vezes. A figura 5.1 apresenta o modelo binomial simulado para a comparação dos métodos.

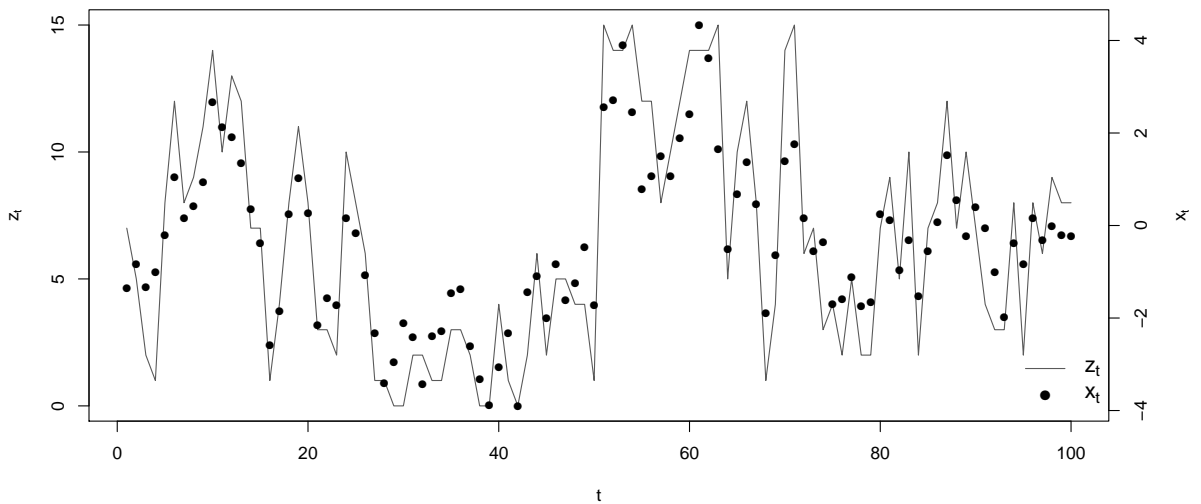


Figura 5.1: Modelo dinâmico binomial simulado (vide equação 5.8).

As figuras a seguir mostram os resultados das estimações para o modelo 5.8 utilizando os métodos de filtragem de Gordon et al. (1993) e Pitt e Shephard (1999) (vide capítulo 4), combinados aos métodos de estimação dos parâmetros de Storvik (2002) e Liu e West (2001) (vide capítulo 5). Para estimação dinâmica dos parâmetros estáticos utilizando o método proposto por Liu e West (2001), em que os parâmetros são estimados via aproximação da distribuição normal, é necessário que

os parâmetros estejam bem definidos na escala \mathfrak{R} , espaço amostral de uma variável com distribuição de probabilidade gaussiana. No modelo proposto, o parâmetro σ^2 , definido em \mathfrak{R}^+ , é o único que necessita de uma transformação para se adequar às restrições do método.

Na tabela 5.1 apresentam-se algumas estatísticas de sumarização, tais como a média *a posteriori*, erro quadrático médio (EQM) e alguns percentis das distribuições para as estimativas de cada parâmetro do modelo dinâmico binomial (vide a expressão (5.8)). São apresentadas as combinações das 2 metodologias de estimação de parâmetros estáticos (Liu e West (2001) e Storvik (2002)), aplicadas aos filtros de partículas de Gordon et al. (1993) e Pitt e Shephard (1999), um total de 4 resultados.

Tabela 5.1: Estimativas para os parâmetros estáticos do modelo dinâmico binomial utilizando os métodos de Storvik (2002) e Liu e West (2001) e com filtragem utilizando os métodos de Gordon et al. (1993) e Pitt e Shephard (1999).

Parâmetros	Método de estimação										
	Storvik (2002)					Liu e West (2001)					
	\bar{x}	$P_{50\%}$	$P_{2.5\%}$	$P_{97.5\%}$	EQM	\bar{x}	$P_{50\%}$	$P_{2.5\%}$	$P_{97.5\%}$	EQM	
Gordon et al. (1993)	$\alpha = 0, 0$	-0,035	-0,034	-0,271	0,200	0,016	-0,129	-0,133	-0,321	0,067	0,029
	$\beta = 0, 9$	0,848	0,849	0,697	1,001	0,009	0,724	0,723	0,559	0,872	0,039
	$\sigma^2 = 1, 1$	1,320	1,295	0,907	1,916	0,125	1,297	1,284	0,99	1,717	0,085
Pitt e Shephard (1999)	$\alpha = 0, 0$	-0,0855	-0,0854	-0,3701	0,204	0,0282	-0,139	-0,144	-0,346	0,079	0,033
	$\beta = 0, 9$	0,7303	0,7301	0,5733	0,8834	0,0351	0,726	0,727	0,577	0,868	0,037
	$\sigma^2 = 1, 1$	2,0412	2,0107	1,5179	2,7751	0,9918	1,288	1,277	0,94	1,678	0,086

O parâmetro σ^2 , cujo verdadeiro valor é 1.1, foi melhor estimado utilizando a técnica de estimação proposta por Liu e West (2001), nas duas metodologias de filtragem, 1.288 (Pitt e Shephard, 1999) e 1.297 (Gordon et al., 1993). No entanto, o resultado utilizando Storvik (2002) via filtro de Gordon et al. (1993) foi bem próximo dos melhores resultados (1.320). Para todos os resultados observa-se EQM próximos de zero. O pior resultado observado para esse critério ocorreu com o método de estimação de Storvik (2002) e filtragem via Pitt e Shephard (1999), observando-se grande viés e um EQM próximo a 1. Para os parâmetros α e β , os

melhores resultados foram encontrados utilizando o método de estimação Storvik (2002) com a filtragem feita por meio do método de Gordon et al. (1993), com estimativas pontuais mais precisas, segundo o critério de EQM.

O gráfico 5.2 apresenta o resultado encontrado para a estimação do parâmetro α por meio do método de Storvik (2002) e utilizando o método de filtragem de Gordon et al. (1993). É possível perceber uma boa convergência para o valor real do parâmetro, representada pela linha em vermelho, a partir de $t = 20$.

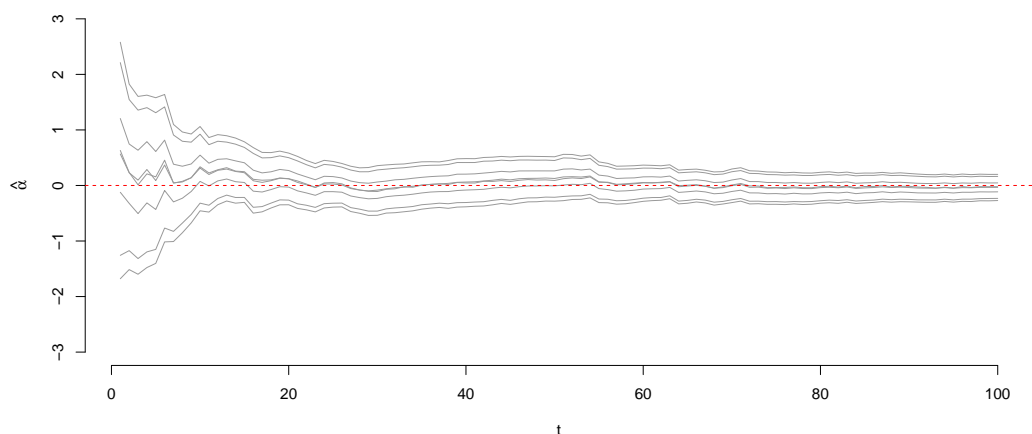


Figura 5.2: Estimativas para os percentis do parâmetro α no modelo Binomial utilizando o processo de estimação dos parâmetros via Storvik (2002) e o método de filtragem de Gordon et al. (1993).

Para o parâmetro β (5.3), há estabilidade das estimativas a partir da 60^a estimação. O gráfico 5.4 apresenta o resultado observado para o parâmetro σ^2 . Como apresentado na tabela resumo 5.1, percebe-se uma superestimação com relação ao verdadeiro, 1.1.

O gráfico 5.5 apresenta o resultado para a estimativa dos estados latentes do modelo dinâmico binomial. Percebe-se uma boa estimação para o vetor $x_{1:100}$.

Ainda utilizando o método de filtragem de Gordon et al. (1993), porém utilizando

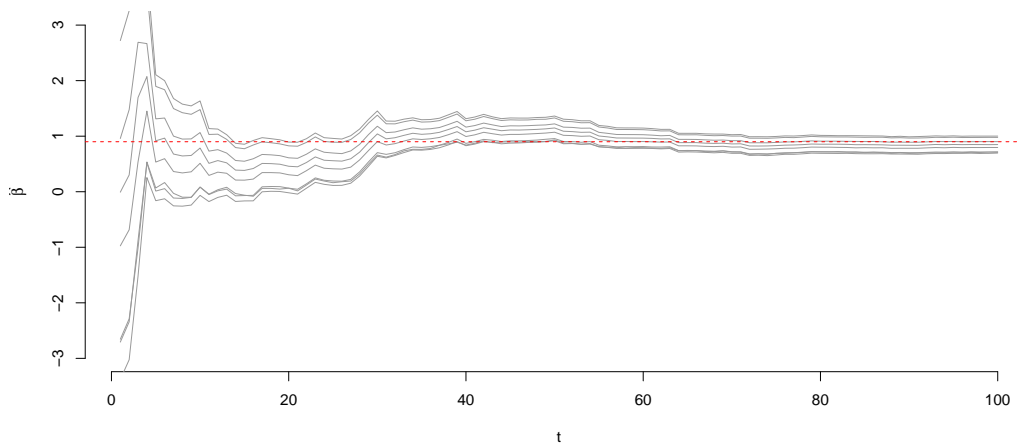


Figura 5.3: Estimativas para os percentis do parâmetro β no modelo Binomial utilizando o processo de estimação dos parâmetros via Storvik (2002) e o método de filtragem de Gordon et al. (1993).

o método de estimação dos parâmetros estáticos de Liu e West (2001), o resultado para o parâmetro α (vide figura 5.6) apresentou um estimativa inferior ao valor verdadeiro. O mesmo acontece com o parâmetro β (vide figura 5.7).

Com relação ao parâmetro σ^2 , o resultado foi melhor do que o encontrado utilizando o método Storvik (2002) de estimação.

Utilizando o método de filtragem de Pitt e Shephard (1999) e Storvik (2002), foi obtido um bom resultado para o parâmetro α (vide figura 5.10). No entanto, o parâmetro β foi subestimado (vide figura 5.11).

O resultado encontrado para o parâmetro σ^2 foi o pior entre as combinações de método de filtragem e estimação de parâmetro estáticos, com um estimativa acima do valor verdadeiro (vide figura 5.12).

Utilizando o método de estimação de Liu e West (2001) e filtragem de Pitt e Shephard (1999), obteve-se bom resultado para a estimação do parâmetro α (vide figura 5.14). Para o parâmetro β , houve uma subestimação (vide figura 5.15).

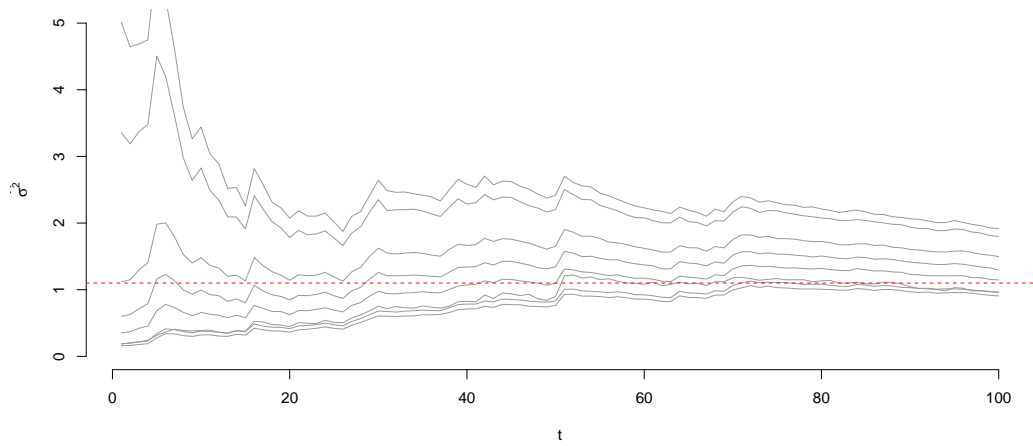


Figura 5.4: Estimativas para os percentis do parâmetro σ^2 no modelo Binomial utilizando o processo de estimação dos parâmetros via Storvik (2002) e o método de filtragem de Gordon et al. (1993). A transformação escolhida para esse parâmetro foi a logaritmica.

Para o parâmetro σ^2 , a figura 5.16, referente aos métodos de Pitt e Shephard (1999) e Liu e West (2001), apresenta o melhor resultado, segundo o critério de EQM no final da estimação.

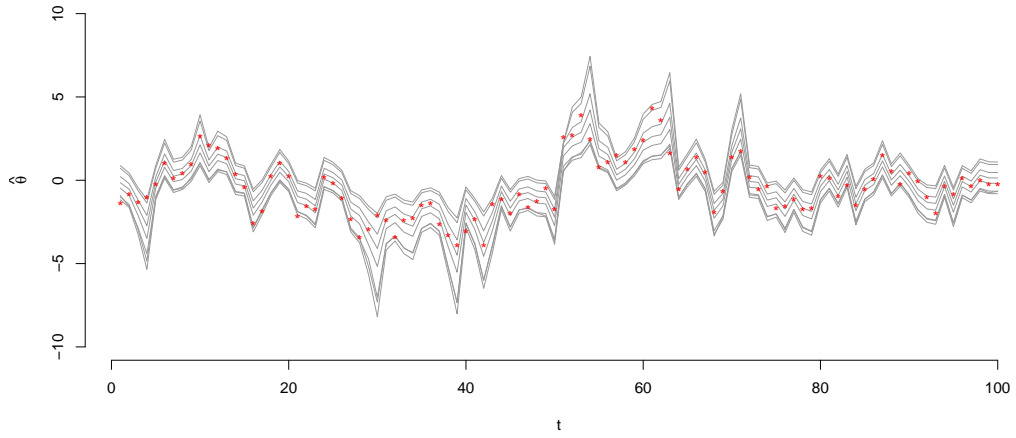


Figura 5.5: Estimativas para os percentis dos estados latentes $x_{1:100}$ no modelo Binomial utilizando o processo de estimação dos parâmetros via Storvik (2002) e o método de filtragem de Gordon et al. (1993).

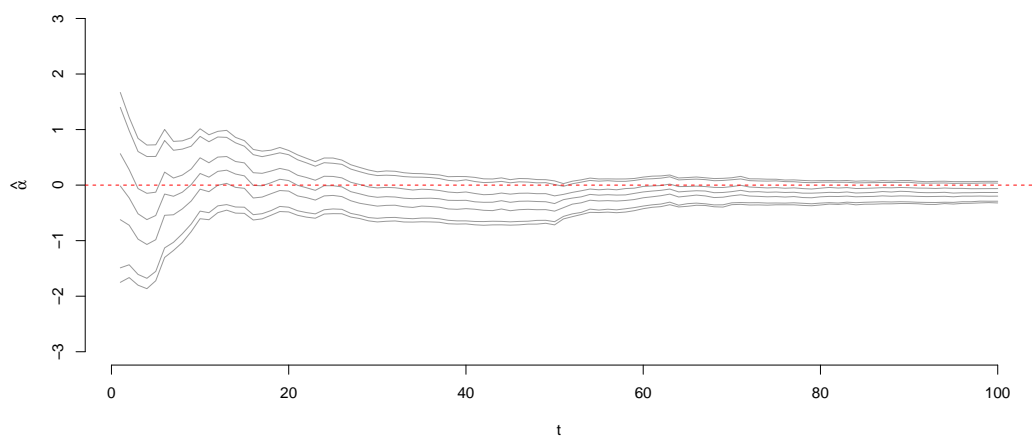


Figura 5.6: Estimativas para os percentis do parâmetro α no modelo Binomial utilizando o processo de estimação dos parâmetros via Liu e West (2001) e o método de filtragem de Gordon et al. (1993).

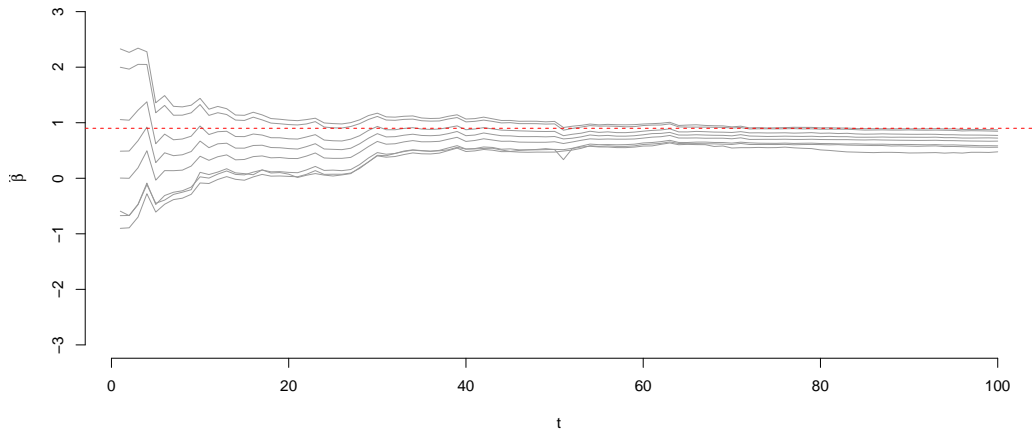


Figura 5.7: Estimativas para os percentis do parâmetro β no modelo Binomial utilizando o processo de estimação dos parâmetros via Liu e West (2001) e o método de filtragem de Gordon et al. (1993).

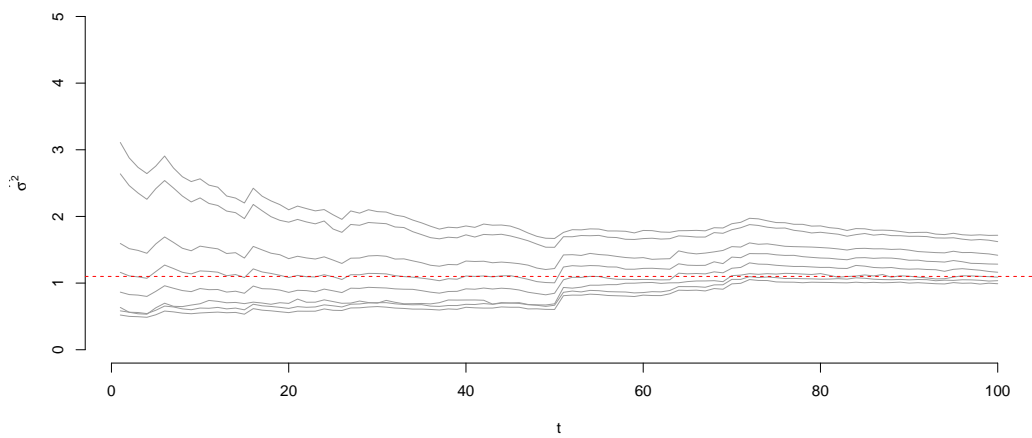


Figura 5.8: Estimativas para os percentis do parâmetro σ^2 no modelo Binomial utilizando o processo de estimação dos parâmetros via Liu e West (2001) e o método de filtragem de Gordon et al. (1993). A transformação escolhida para esse parâmetro foi a logaritmica.

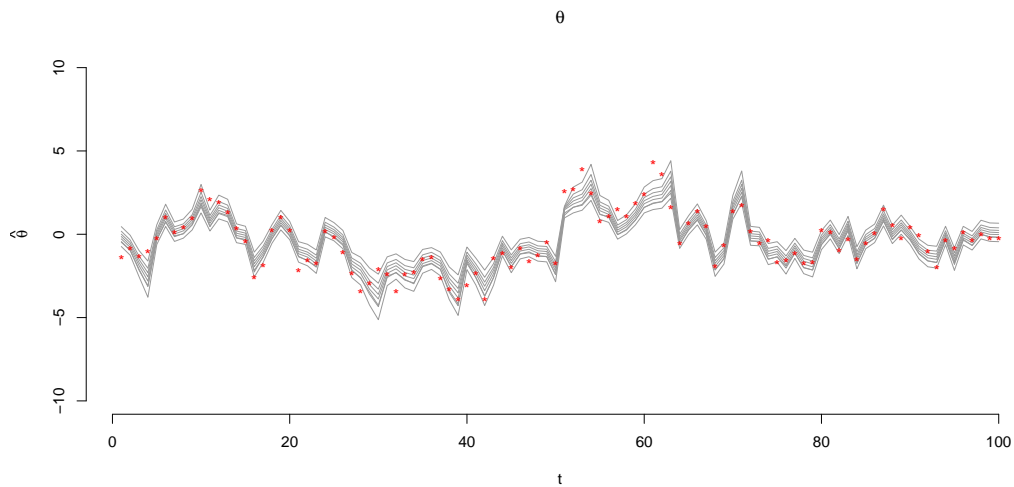


Figura 5.9: Estimativas para os percentis dos estados latentes $x_{1:100}$ no modelo Binomial utilizando o processo de estimação dos parâmetros via Liu e West (2001) e o método de filtragem de Gordon et al. (1993).

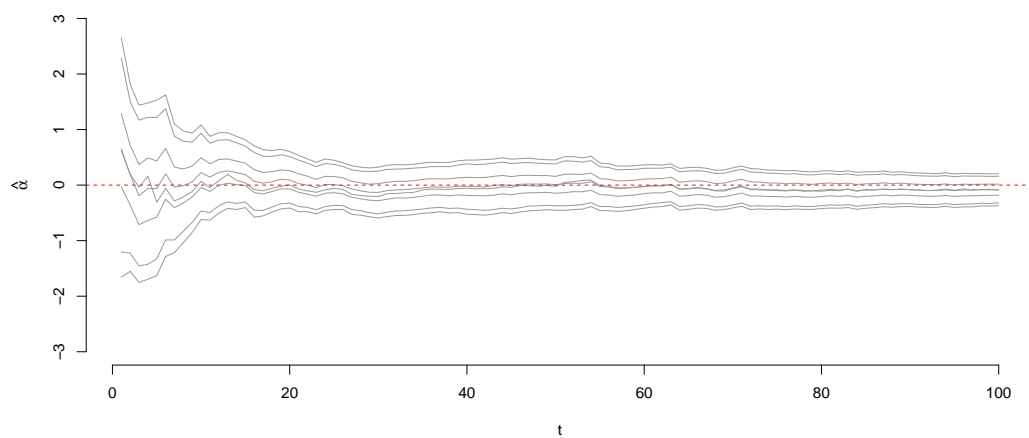


Figura 5.10: Estimativas para os percentis do parâmetro α no modelo Binomial utilizando o processo de estimação dos parâmetros via Storvik (2002) e o método de filtragem de Pitt e Shephard (1999).

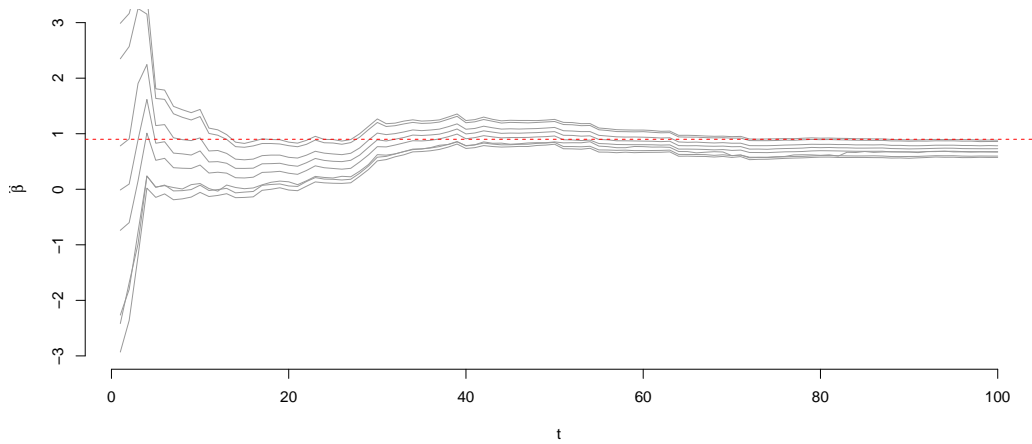


Figura 5.11: Estimativas para os percentis do parâmetro β no modelo Binomial utilizando o processo de estimação dos parâmetros via Storvik (2002) e o método de filtragem de Pitt e Shephard (1999).

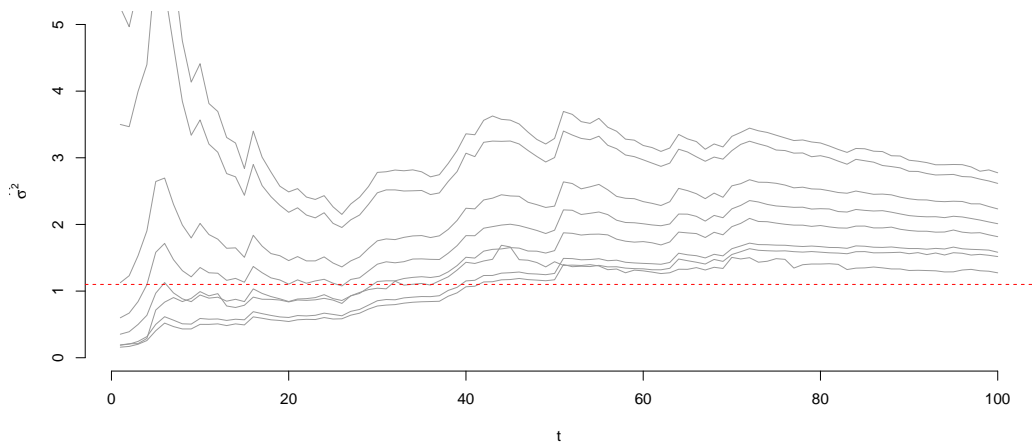


Figura 5.12: Estimativas para os percentis do parâmetro σ^2 no modelo Binomial utilizando o processo de estimação dos parâmetros via Storvik (2002) e o método de filtragem de Pitt e Shephard (1999). A transformação escolhida para esse parâmetro foi a logarítmica.

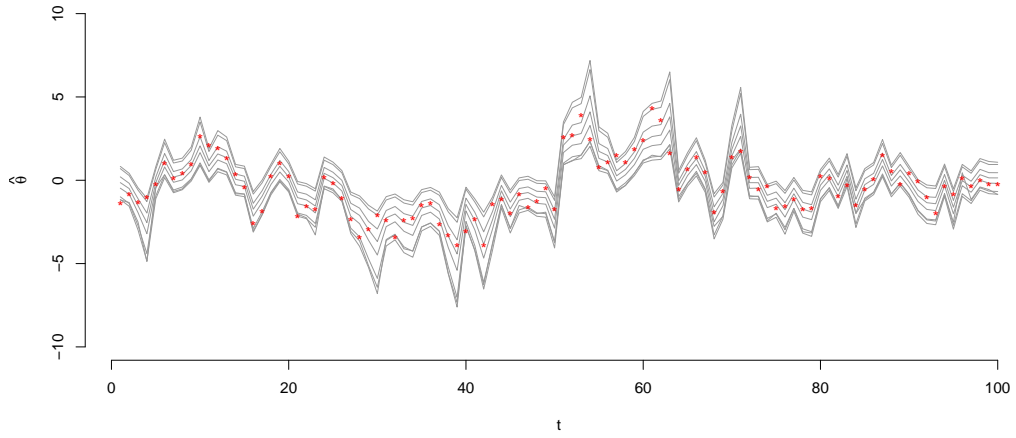


Figura 5.13: Estimativas para os percentis dos estados latentes $x_{1:100}$ no modelo Binomial utilizando o processo de estimação dos parâmetros via Storvik (2002) e o método de filtragem de Pitt e Shephard (1999).

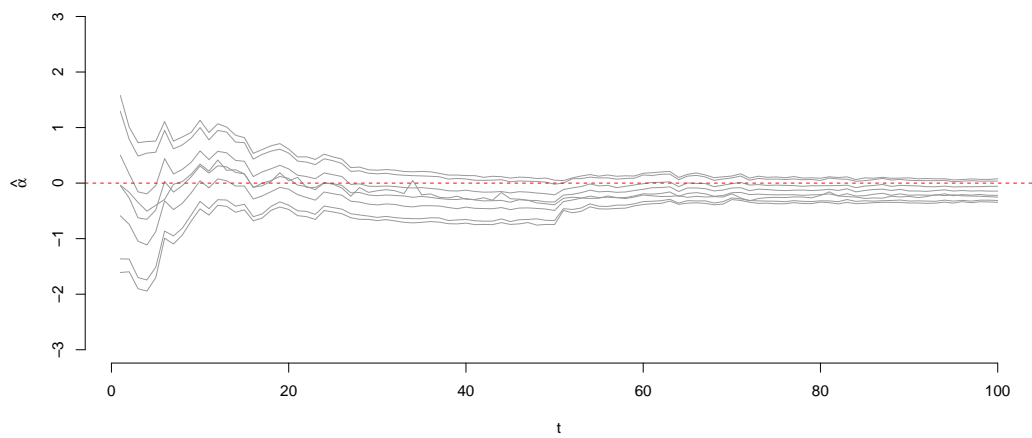


Figura 5.14: Estimativas para os percentis do parâmetro α no modelo Binomial utilizando o processo de estimação dos parâmetros via Liu e West (2001) e o método de filtragem de Pitt e Shephard (1999).

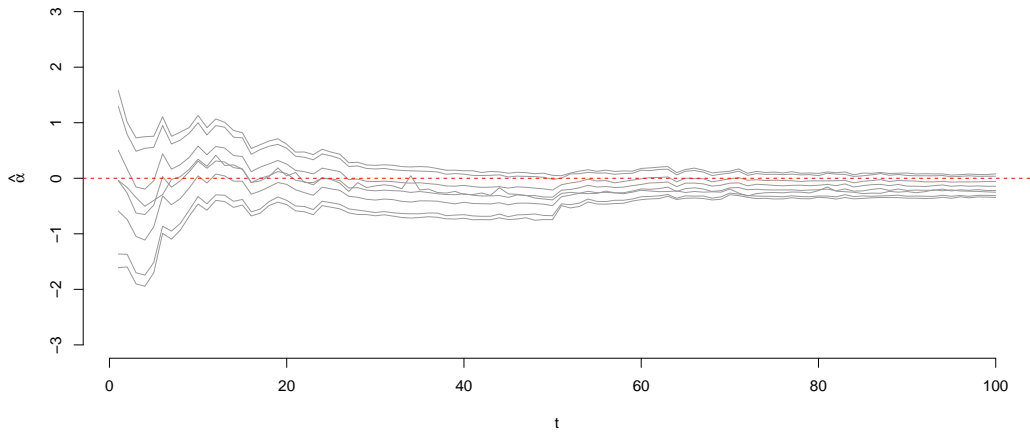


Figura 5.15: Estimativas para os percentis do parâmetro β no modelo Binomial utilizando o processo de estimação dos parâmetros via Liu e West (2001) e o método de filtragem de Pitt e Shephard (1999).

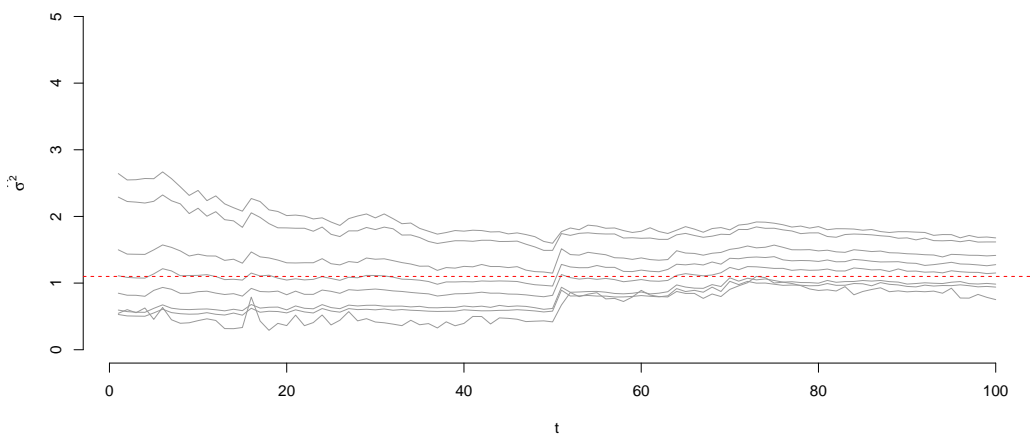


Figura 5.16: Estimativas para os percentis do parâmetro σ^2 no modelo Binomial utilizando o processo de estimação dos parâmetros via Liu e West (2001) e o método de filtragem de Pitt e Shephard (1999). A transformação escolhida para esse parâmetro foi a logaritmica.

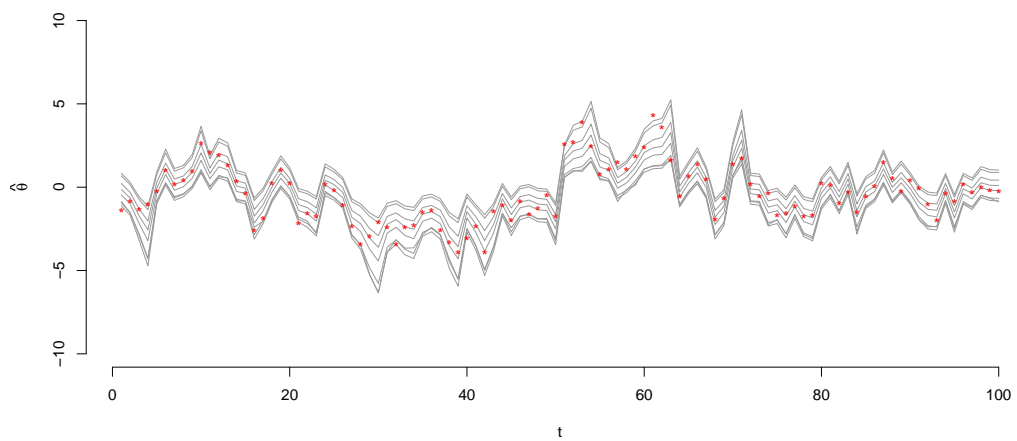


Figura 5.17: Estimativas para os percentis dos estados latentes $x_{1:100}$ no modelo Binomial utilizando o processo de estimação dos parâmetros via Liu e West (2001) e o método de filtragem de Pitt e Shephard (1999).

Capítulo 6

Considerações finais

6.1 Sumário Filtro de Partículas

Na seção 4.4 apresenta-se a equação geral para o processo de **evolução incremental** (EI), que representa o peso na distribuição das partículas.

$$w_t \propto w_{t-1} \frac{\pi(y_t|x_t)\pi(\theta_t|\theta_{t-1})}{g_{t|t-1}(\theta_t|\theta_{0:t-1}, y_{1:t})} \quad (6.1)$$

Como visto, o processo de reamostragem ajuda a solucionar o problema de degeneração da distribuição das partículas no processo de estimação (Storvik, 2002). As partículas reamostradas têm o peso amostral de importância redefinido, com valor de $w_t = 1/N$, sendo N o número de partículas. Dessa forma, a distribuição atual dos pesos w_t não mais é recursiva, de modo que, as informação anteriores w_{t-1} não são mais necessárias:

$$w_t \propto \frac{\pi(y_t|x_t)\pi(\theta_t|\theta_{t-1})}{g_{t|t-1}(\theta_t|\theta_{0:t-1}, y_{1:t})}. \quad (6.2)$$

Na equação 6.2, a expressão $g_{t|t-1}(\theta_t|\theta_{0:t-1}, y_{1:t})$ é a **função de transição de importância** (FTI), e a escolha define as principais características do Filtro de Partículas.

Para a classificação dos filtros de partículas, o processo de **propagação** pode ser considerado como *blind* (B) ou *unblind* (UB), Pitt e Shephard (1999), e como *sampling-resampling* (S-R) ou *resampling-sampling* (R-S) (Lopes e Tsay, 2011).

A tabela 6.1 apresenta a FTI nas principais referências desta dissertação:

Tabela 6.1: Resumo para estimação via Filtro de Partículas.

Referência	(FTI)	(EI)	Tipo FP	Nome FP
Gordon et al. (1993)	$f(x_t x_{t-1})$	$\propto \pi(y_t x_t)$	(B) e (S-R)	Bootstrap
Pitt e Shephard (1999)	$f(y_t \mu_t^k)f(x_t x_{t-1}^k)\pi^k$	$\propto \frac{f(y_t x_t^j)}{f(y_t \mu_t^{kj})}$	(B) e (R-S)	Auxiliar
Petris et al. (2009)	$f(x_t y_t, x_{t-1})$	$\propto \pi(y_t x_{t-1})$	(UB) e (S-R)	Ótimo

6.2 Conclusão

Ao longo da dissertação, foram discutidas técnicas necessárias para o estudo do Filtro de Partículas. Temas como modelos dinâmicos, amostragem de importância e Monte Carlo Sequencial foram apresentados para estudar os trabalhos seminais do Filtro de Partículas. Tais temas requereram recursos de programação, estimação numérica e teoria Bayesiana.

Apresentou-se os desenvolvimentos seminais do Filtro de Partículas, com o intuito de iniciar estudos futuros na área. Foram apresentados os resultados do Filtro de Partículas para a estimação dos estados latentes e encontraram-se bons resultados, com estimativas próximas de soluções analíticas do Filtro de Kalman.

Comparou-se os resultados encontrados da combinação entre os artigos seminais para o processo de estimação dos estados latentes, Gordon et al. (1993) e Pitt e Shephard (1999), e para a estimação dos parâmetros estáticos, Liu e West (2001) e Storvik (2002). Os melhores resultados foram obtidos com as combinações de

técnicas apresentadas por Liu e West (2001), que utilizam no processo de filtragem a metodologia de Pitt e Shephard (1999), e Storvik (2002), que utiliza a metodologia proposta por Gordon et al. (1993). No entanto, foram encontradas boas estimativas para a combinação das metodologias propostas por Gordon et al. (1993) e por Liu e West (2001). O pior resultado, segundo o critério EQM, foi para a combinação de Pitt e Shephard (1999) e Storvik (2002).

O Filtro de Partículas conseguiu apresentar bons resultados por meio de estimativas *online*, nos modelos dinâmicos, utilizando aproximações numéricas. Mostrou-se uma excelente alternativa para problemas de difícil estimação e tratamento analítico.

Referências Bibliográficas

- Bollerslev, T. (1986). Generalized autoregressive condicional heteroskedastic. *Journal of Econometrics*, (31):307–327.
- Douc, R. e Cappe, O. (2006). Comparison of resampling scheme for particle filter. *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, pages 64–69.
- Doucet, A. e Johansen, A. M. (2008). A tutorial on particle filtering and smoothing:fifteen years later. *Technical report (Department of Statistics, University of British Columbia)*.
- Doucet, A., West, M., e Godsill, S. J. (2004). Monte carlo smoothing for non-linear time series. *Journal of the American Statistical Association*, 99:156–168.
- Gordon, N. J., Salmond, D. J., e Smith, A. F. M. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEEE Proceedings F on Radar and Signal Processing*, (140):107–113.
- Kitagawa, G. (1996). Monte carlo filter and smoother for non-gaussian nonlinear state models. *Journal of Computational and Graphical Statistics*, (1):1–25.
- Liu, J. e West, M. (2001). *Combined parameter and state estimation in simulation-based Filtering*. Springer-Verlag.
- Lopes, H. (2009). Tutorial: An introduction to particle filters. *XIII Brazilian School of Time Series and Econometrics*.
- Lopes, H. e Tsay, R. (2011). Particle filters and bayesian inference in financial econometrics. *Journal of Forecasting*, (30):168–209.
- Petris, G., Petrone, S., e Campagnoli, P. (2009). *Dynamic Linear Models with R*. Springer.

- Pitt, M. K. e Shephard, N. (1999). Filter via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, (94):590–599.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- Ross, S. M. (2007). *Introduction a probability Models*, (9th ed.). University of California.
- Rubin, D. B. (1992). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*.
- Smith, F. M. e Gelfand, A. E. (1992). Bayesian statistic without tears: A sampling-resampling perspective. *The American Statistician*, (46):84–88.
- Storvik, G. (2002). Particle filters for state-space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing*, (50):281–298.
- West, M. (1993a). Approximating posterior distributions by mixtures. *Journal of Royal Statistical Society*, (55):409–422.
- West, M. (1993b). Mixture models, monte carlo, bayesian updating and dynamic models. *Computing Science and Statistics: Proceedings of the 24th Symposium on the Interface*, pages 325–333.
- West, M. e Harrison, J. (1997). *Bayesian Forecast and Dynamic Models*. Springer-Verlag.
- Wichern, D. e Johnson, R. A. (2007). *Applied Multivariate Statistical Analysis*, (6th ed.). Statistical Analysis.

Apêndice A

Código em R para estimação via Filtro de Partículas Gordon et al. (1993) e Liu e West (2001)

```
#####
# Model:
#       x[t] = alpha + beta *x[t-1] + e[t] , e[t] ~ N(0,Sigma^2)
#       y[t] ~ Binomial(r,invlogit(x[t]))
#####
# unknown static parameters : alpha, beta e sigma^2
# state : x
# observed series : y
#####
rm(list=ls())
#Functions
percent.function=function(x){
result=c(quantile(x,c(.025,.05,0.25,.5,.75,.95,.975)),mean(x))
return(result)
}
invlogit=function(x){
z=exp(x) / (1+ exp(x))
return(z)
}
LW = function(x,y,alphas,betas,sig2s,r,xs,a){
result=array(0,c(T,8+1,4))
n = length(y)
N = length(xs)
Matriz.alpha=matrix(0,ncol=n,nrow=N)
Matriz.beta=matrix(0,ncol=n,nrow=N)
Matriz.sig2=matrix(0,ncol=n,nrow=N)
Matriz.theta=matrix(0,ncol=n,nrow=N)
h2 = 1-a^2
pars = cbind(alphas,betas,log(sig2s))
for (t in 1:n){
#Parameters update
mpar = apply(pars,2,mean)
vpar = var(pars)
ms = a*pars+(1-a)*matrix(mpar,N,3,byrow=T)
means1 =alphas+betas*xs
propost=rmnorm(N,means1,sqrt(sig2s))
p1=dbinom(y[t],r,invlogit(propost))
pesos = p1
k = sample(1:N,size=N,replace=T,prob=pesos)
pars = ms[k,] + matrix(rnorm(3*N),N,3)*%chol(h2*vpar)
alphas = pars[,1]
betas = pars[,2]
sig2s = exp(pars[,3])
Matriz.alpha[,t]=alphas
Matriz.beta[,t]=betas
Matriz.sig2[,t]=sig2s
# states latentes estimates
# Propagation
means1=alphas+betas*xs[k]
propost=propost[k]
p1=dnorm(propost,means1,sqrt(sig2s))
p2=dbinom(y[t],r,invlogit(propost))
# Resampling
p3=p1*p2
x.simul=sample(propost,N,replace=T,prob=p3)
xs = x.simul
Matriz.theta[,t]=xs
}
#Storing percents of estimates alpha,beta,sig2,theta
res1=t(apply(Matriz.alpha,2,percent.function))
res2=t(apply(Matriz.beta,2,percent.function))
res3=t(apply(Matriz.sig2,2,percent.function))
res4=t(apply(Matriz.theta,2,percent.function))
#Storing EQM
res1=cbind(res1,apply(((Matriz.alpha - alpha)^2)/N,2,sum))
res2=cbind(res2,apply(((Matriz.beta - beta)^2)/N,2,sum))
res3=cbind(res3,apply(((Matriz.sig2 - sig2)^2)/N,2,sum))
res4=cbind(res4,apply(((Matriz.theta - x)^2)/N,1,sum))
result[,1]=res1;result[,2]=res2;result[,3]=res3
result[,4]=res4
return(result)
}
# data
set.seed(231006);n = 100;alpha = 0.0;beta = 0.9;sig2 = 1.1
r = 15;sig = sqrt(sig2);x = rep(0,n+1)
true = c(alpha,beta,sig2)
names = c("alpha","beta","sig2")
for(t in 1:n){
x[t+1] = rnorm(1,alpha+beta*x[t],sig)
}
x = x[-1]
y = rbinom(n,r,invlogit(x))
# hiperparameters
mx0 = 0.0;vx0 = 10;vsigs0 = 10;pars = c(alpha,beta);
vpars = c(1,1) /sig;
N = 1000;nsim = 100;a = 0.95;T=length(y)
result=array(0,c(nsim,T,9,4))
for (s in 1:nsim){
xs = rnorm(N,mx0,vx0)
sig2s = 1/rgamma(N,vsigs0/2,vsigs0*sig2/2)
sigs = sqrt(sig2s)
alphas = rnorm(N,pars[1],vpars[1]^ .5)
betas = rnorm(N,pars[2],vpars[2]^ .5)
result[s,,] = LW(x,y,alphas,betas,sig2s,r,xs,a)
}
}
```

```

#ALPHA
kpar=1
res1=result[,,kpar]
plot(res1[1,,4],type="l",col="gray60",ylim=c(-3,3),
main=NULL,xlab=expression(t),ylab=expression(hat(alpha)),bty="n")
for(j in 1:7){
results2=apply(result[,j,kpar],2,median)
points(results2,type="l",col="gray60")
}
abline(h=true[kpar],col=2,lty=2)

#BETA
kpar=2
res1=result[,,kpar]
plot(res1[1,,4],type="l",col="gray60",ylim=c(-3,3),
main=NULL,xlab=expression(t),ylab=expression(hat(alpha)),bty="n")
for(j in 1:7){
results2=apply(result[,j,kpar],2,median)
points(results2,type="l",col="gray60")
}
abline(h=true[kpar],col=2,lty=2)

#SIGMA
kpar=3
res1=result[,,kpar]
plot(res1[1,,4],type="l",col="gray60",ylim=c(-3,3),main=NULL,
xlab=expression(t),ylab=expression(hat(alpha)),bty="n")
for(j in 1:7){
results2=apply(result[,j,kpar],2,median)
points(results2,type="l",col="gray60")
}
abline(h=true[kpar],col=2,lty=2)

#THETA
kpar=4
res1=result[,,kpar]
plot(res1[1,,4],type="l",col="gray60",ylim=c(-3,3),main=NULL,
xlab=expression(t),ylab=expression(hat(alpha)),bty="n")
for(j in 1:7){
results2=apply(result[,j,kpar],2,median)
points(results2,type="l",col="gray60")
}
points(x,col=2,lty=2,pch="*",type="p")

```

Apêndice B

Código em R para estimação via Filtro de Partículas Pitt e Shephard (1999) e Liu e West (2001)

```
LW.P = function(x,y,alphas,betas,sig2s,r,xs,a){
result=array(0,c(T,9,4))
n = length(y)
N = length(xs)
Matriz.alpha=matrix(0,ncol=n,nrow=N)
Matriz.beta=matrix(0,ncol=n,nrow=N)
Matriz.sig2=matrix(0,ncol=n,nrow=N)
Matriz.theta=matrix(0,ncol=n,nrow=N)
h2 = 1-a^2
pars = cbind(alphas,betas,log(sig2s))
for (t in 1:n){
#Parameters update
mpar = apply(pars,2,mean)
vpar = var(pars)
ms = a*pars+(1-a)*matrix(mpar,N,3,byrow=T)
means1 =ms[,1]+ms[,2]*xs
propost=rnorm(N,means1,sqrt(exp(ms[,3])))
weight=dbinom(y[t],r,invlogit(propost))
k = sample(1:N,size=N,replace=T,prob=weight)
pars = ms[k,] + matrix(rnorm(3*N),N,3)%*%chol(h2*vpar)
alphas = pars[,1]
betas = pars[,2]
sig2s = exp(pars[,3])
theta1 = rnorm(N,alphas+betas*xs[k],sqrt(sig2s))
weight = dbinom(y[t],r,invlogit(theta1))/
dbinom(y[t],r,invlogit(propost[k]))
ind = sample(1:N,size=N,replace=T,prob=weight)
xs = theta1[ind]
alphas = alphas[ind]
betas = betas[ind]
sig2s = sig2s[ind]

Matriz.theta[,t]=xs
Matriz.alpha[,t]=alphas
Matriz.beta[,t]=betas
Matriz.sig2[,t]=sig2s

}

#Storing percents of estimates alpha,beta,sig2,theta
res1=t(apply(Matriz.alpha,2,percent.function))
res2=t(apply(Matriz.beta,2,percent.function))
res3=t(apply(Matriz.sig2,2,percent.function))
res4=t(apply(Matriz.theta,2,percent.function))

#Storing EQM
res1=cbind(res1,apply(((Matriz.alpha - alpha)^2)/N,2,sum))
res2=cbind(res2,apply(((Matriz.beta - beta)^2)/N,2,sum))
res3=cbind(res3,apply(((Matriz.sig2 - sig2)^2)/N,2,sum))
res4=cbind(res4,apply(((Matriz.theta - x)^2)/N,1,sum))

result[,1]=res1
result[,2]=res2
result[,3]=res3
result[,4]=res4
return(result)
}

result=array(0,c(nsim,T,9,4))
for (s in 1:nsim){
xs = rnorm(N,mx0,vx0)
sig2s = 1/rgamma(N,vsigs0/2,vsigs0*sig2/2)
sigs = sqrt(sig2s)
alphas = rnorm(N,pars[1],vpars[1]^ .5)
betas = rnorm(N,pars[2],vpars[2]^ .5)
result[s,,] = LW.P(x,y,alphas,betas,sig2s,r,xs,a)
}

#ALPHA
kpar=1
res1=result[, ,kpar]
plot(res1[,1,,4],type="l",col="gray60",ylim=c(-3,3),
main=NULL,xlab=expression(t),ylab=expression(hat(alpha)),bty="n")
for(j in 1:7){
results2=apply(result[, ,j,kpar],2,median)
points(results2,type="l",col="gray60")
}
abline(h=true[kpar],col=2,lty=2)

#BETA
kpar=2
res1=result[, ,kpar]
plot(res1[,1,,4],type="l",col="gray60",ylim=c(-3,3),
main=NULL,xlab=expression(t),ylab=expression(hat(alpha)),bty="n")
for(j in 1:7){
results2=apply(result[, ,j,kpar],2,median)
points(results2,type="l",col="gray60")
}
```



```

}
abline(h=true[kpar],col=2,lty=2)

#SIGMA
kpar=3
res1=result[,,kpar]
plot(res1[1,,4],type="l",col="gray60",ylim=c(-3,3),main=NULL,
xlab=expression(t),ylab=expression(hat(alpha)),bty="n")
for(j in 1:7){
results2=apply(result[,j,kpar],2,median)
points(results2,type="l",col="gray60")
}
abline(h=true[kpar],col=2,lty=2)

#THETA
kpar=4
res1=result[,,kpar]
plot(res1[1,,4],type="l",col="gray60",ylim=c(-3,3),main=NULL,
xlab=expression(t),ylab=expression(hat(alpha)),bty="n")
for(j in 1:7){
results2=apply(result[,j,kpar],2,median)
points(results2,type="l",col="gray60")
}
points(x,col=2,lty=2,pch="*",type="p")

```

Apêndice C

Código em R para estimação via Filtro de Partículas Gordon et al. (1993) e Storvik (2002)

```
#####
#Processo de estimação de parâmetros via Storvik.G(2002)
# Modelo:
#      x[t] = alpha + beta *x[t-1] + e[t] , e[t] ~ N(0,Sigma^2)
#      y[t] ~ Binomial(r,invlogit(x[t]))
#####
# Parâmetros estáticos desconhecidos: Alpha, Beta e Sigma^2
# Parâmetros dinâmicos desconhecidos: x
# Observações do sistema      : y
# t                            : índice tempo
#####
rm(list=ls())

#Funções criadas

percent=function(x){
  result=c(quantile(x,c(.025,.05,0.25,.5,.75,.95,.975)),mean(x))
  return(result)
}

invlogit=function(x){
  z=exp(x) / (1+ exp(x))
  return(z)
}

Storvik.G=function(x,y,N,m,alpha0,est.sigma.alpha0,m.a0,est.sigma.a0,v0,d0,mk,D,sig.sm.prior){xs1
  T=length(y)
  matrix.alpha =matrix(1,ncol=T,nrow=N)
  matrix.a     =matrix(1,ncol=T,nrow=N)
  matrix.sig.hat =matrix(1,ncol=T,nrow=N)
  X            =matrix(1,ncol=T,nrow=N)

  #prior
  alpha.est =rnorm(N,m.alpha0,est.sigma.alpha0^.5)
  a.est =rnorm(N,m.a0,est.sigma.a0^.5)
  sigma.hat.est =1/(rgamma(N,shape=v0/2,scale=1/(d0/2)))
  x0 =rnorm(N,mx.0,sig.x.prior^.5)
  x00 =x0

  #initial statistic sufficient
  s = matrix(0,N,5)
  s[,1] = 0.0
  s[,2] = 0.0
  s[,3] = 0.0
  s[,4] = 0.0
  s[,5] = 0.0

  k=1
  for(k in 1:T){
    #Filter
    propostaX =rnorm(N,alpha.est+ a.est*(x0),sigma.hat.est^.5)
    summary(propostaX)

    q =dbinom(y[k],r,invlogit(propostaX))
    q =q + 0.00001 #ajuste computacional
    q =q/sum(q)
    ind=sample(1:N,N,replace=T,prob=q)

    X[,k]=propostaX[ind]
    alpha.est=alpha.est[ind]
    a.est=a.est[ind]
    sigma.hat.est=sigma.hat.est[ind]

    xs1 = x0[ind]
    xs = X[,k]
    so = s[ind,]

    # update statistic sufficient
    s[,2] = so[,2] + xs
    s[,3] = so[,3] + xs1^2
    s[,4] = so[,4] + xs^2
    s[,5] = so[,5] + xs*xs1

    #Gibbs Sampler
    for( i in 1:GB){
      #beta
      den=(sigma.hat.est + est.sigma.a0*(s[,3]))
      summary(den)
      m.a=(s[,5] - s[,1]) * est.sigma.a0 + sigma.hat.est * m.a0
      summary(m.a)
      m.a=m.a/den
      est.sigma.a=est.sigma.a0*sigma.hat.est /den
    }
  }
}
```

```

a.est = rnorm(N,m.a,est.sigma.a^0.5)
summary(m.a)
  matrix.a[,k]=a.est
#alpha
den =(sigma.hat.est + est.sigma.alpha0*k)
m.alpha =(est.sigma.alpha0 * s[,2] + sigma.hat.est*m.alpha0 - a.est * s[,1]*est.sigma.alpha0)
  m.alpha=m.alpha/den
summary(m.alpha)
est.sigma.alpha =(est.sigma.alpha0 * sigma.hat.est) /den
alpha.est = rnorm(N,m.alpha,est.sigma.alpha^0.5)
  matrix.alpha[,k]=alpha.est

#sigma
m.v= v0 + k
square = s[,4] + k*alpha.est^2 + a.est^2*s[,3] - 2 * (alpha.est * s[,2] + sigma.hat.est * s[,1])
m.d = d0 + square
  sigma.hat.est = (1/(gamma(N,m.v/2,m.d/2)))
matrix.sig.hat[,k] = sigma.hat.est
}

  x0=X[,k]
}
result=array(0,c(T,8+1,4))
#Storing percents of estimates alpha,beta,sig2,theta
res1=t(apply(matrix.alpha,2,percent.function))
res2=t(apply(matrix.a,2,percent.function))
res3=t(apply(matrix.sig.hat,2,percent.function))
res4=t(apply(X,2,percent.function))
#Storing EQM
res1=cbind(res1,apply(((matrix.alpha - alpha)^2)/N,2,sum))
res2=cbind(res2,apply(((matrix.a - beta)^2)/N,2,sum))
res3=cbind(res3,apply(((matrix.sig.hat - sigma2)^2)/N,2,sum))
res4=cbind(res4,apply(((t(X) - x)^2)/N,1,sum))

result[,,1]=res1
result[,,2]=res2
result[,,3]=res3
result[,,4]=res4
return(result)
}

# data
set.seed(231006)
n = 100
T = n
alpha = 0
beta = 0.9
sigma2 = 1.1
r = 15
sigma = sqrt(sigma2)
GB=10
x = 0:n
true = c(alpha,beta,sigma2)
names = c("alpha","beta","sigma2")
for (t in 1:n){
x[t+1] = rnorm(1,alpha+beta*x[t],sigma)
}
x = x[-1]
y = rbinom(n,r,invlogit(x))

N = 1000
nsim = 100
a = 0.95
T=length(y)
result=array(0,c(nsim,T,8+1,4))
s=1
for (s in 1:nsim){
  print(s)
  est.sigma.alpha0 =4
  m.alpha0 =0
  est.sigma.a0 =4
  m.a0 =0
  d0 =1.5
  v0 =3
  mx.0 =0
  sig.x.prior =3
  result[s,,] = StorviK.G(x,y,N,m.alpha0,est.sigma.alpha0,m.a0,est.sigma.a0,v0,d0,mx.0,
}

#ALPHA
kpar=1
res1=result[, ,kpar]
plot(res1[1,,4],type="l",col="gray60",ylim=c(-3,3),
main=NULL,xlab=expression(t),ylab=expression(hat(alpha)),bty="n")
for(j in 1:7){
  results2=apply(result[, ,j,kpar],2,median)
  points(results2,type="l",col="gray60")
}
abline(h=true[kpar],col=2,lty=2)

#BETA
kpar=2
res1=result[, ,kpar]
plot(res1[1,,4],type="l",col="gray60",ylim=c(-3,3),
main=NULL,xlab=expression(t),ylab=expression(hat(alpha)),bty="n")
for(j in 1:7){
  results2=apply(result[, ,j,kpar],2,median)
  points(results2,type="l",col="gray60")
}
abline(h=true[kpar],col=2,lty=2)

#SIGMA
kpar=3
res1=result[, ,kpar]
plot(res1[1,,4],type="l",col="gray60",ylim=c(-3,3),main=NULL,
xlab=expression(t),ylab=expression(hat(alpha)),bty="n")
for(j in 1:7){
  results2=apply(result[, ,j,kpar],2,median)
  points(results2,type="l",col="gray60")
}
abline(h=true[kpar],col=2,lty=2)

#THETA
kpar=4
res1=result[, ,kpar]
plot(res1[1,,4],type="l",col="gray60",ylim=c(-3,3),main=NULL,
xlab=expression(t),ylab=expression(hat(alpha)),bty="n")
for(j in 1:7){
  results2=apply(result[, ,j,kpar],2,median)
  points(results2,type="l",col="gray60")
}
points(x,col=2,lty=2,pch="*",type="p")

```

Apêndice D

Código em R para estimação via Filtro de Partículas Pitt e Shephard (1999) e Storvik (2002)

```
Storvik.P=function(x,y,N,
m.alpha0,est.sigma.alpha0,
m.a0,est.sigma.a0,
v0,d0,
mx.0,sig.x.prior){
T=length(y)
Matriz.alpha =matrix(1,ncol=T,nrow=N)
Matriz.a =matrix(1,ncol=T,nrow=N)
Matriz.sig.hat =matrix(1,ncol=T,nrow=N)
X =matrix(1,ncol=T,nrow=N)
#prior
alpha.est =rnorm(N,m.alpha0,est.sigma.alpha0^.5)
a.est =rnorm(N,m.a0,est.sigma.a0^.5)
sigma.hat.est =1/(rgamma(N,shape=v0/2,scale=1/(d0/2)))
x0 =rnorm(N,mx.0,sig.x.prior^.5)
x00 =x0

#initial statistic sufficient
s = matrix(0,N,5)
s[,1] = 0.0
s[,2] = 0.0
s[,3] = 0.0
s[,4] = 0.0
s[,5] = 0.0

for(k in 1:T){
#Filter
means1 = alpha.est+a.est*x0
propost = rnorm(N,means1,sqrt(sigma.hat.est))
weight = dbinom(y[k],r,invlogit(propost))
ind = sample(1:N,size=N,replace=T,prob=weight)
theta1 = rnorm(N,alpha.est+a.est*x0[ind],sqrt(sigma.hat.est))
weight = dbinom(y[k],r,invlogit(theta1))/
dbinom(y[k],r,invlogit(propost[ind]))
ind = sample(1:N,size=N,replace=T,prob=weight)
X[,k] = theta1[ind]
alpha.est = alpha.est[ind]
a.est = a.est[ind]
sigma.hat.est = sigma.hat.est[ind]
xs1 = x0[ind]
xs = X[,k]
so = s[ind,]

# update statistic sufficient
s[,1] = so[,1] + xs1
s[,2] = so[,2] + xs
s[,3] = so[,3] + xs1^2
s[,4] = so[,4] + xs^2
s[,5] = so[,5] + xs*xs1

#Gibbs Sampler
for( i in 1:GB){
#beta
den=(sigma.hat.est + est.sigma.a0*(s[,3]))
summary(den)
m.a=(s[,5] - s[,1]) * est.sigma.a0 + sigma.hat.est * m.a0
summary(m.a)
m.a=m.a/den
est.sigma.a=est.sigma.a0*sigma.hat.est /den
a.est = rnorm(N,m.a,est.sigma.a^0.5)
summary(m.a)
Matriz.a[,k]=a.est

#alpha
den =(sigma.hat.est + est.sigma.alpha0*k)
m.alpha =(est.sigma.alpha0 * s[,2] + sigma.hat.est*m.alpha0 -
a.est * s[,1]*est.sigma.alpha0)
m.alpha=m.alpha/den
summary(m.alpha)
est.sigma.alpha =(est.sigma.alpha0 * sigma.hat.est) /den
alpha.est = rnorm(N,m.alpha,est.sigma.alpha^0.5)
Matriz.alpha[,k]=alpha.est

#sigma
m.v= v0 + k
square = s[,4] + k*alpha.est^2 + a.est^2*s[,3] -
2 * (alpha.est * s[,2] + a.est*s[,5]) + 2* a.est * alpha.est * s[,1]
m.d = d0 + square
sigma.hat.est = (1/(rgamma(N,m.v/2,m.d/2)))
Matriz.sig.hat[,k] = sigma.hat.est

}

result=array(0,c(T,9,4))

#Storing percents of estimates alpha,beta,sig2,theta
res1=t(apply(Matriz.alpha,2,percent.function))
res2=t(apply(Matriz.a,2,percent.function))
res3=t(apply(Matriz.sig.hat,2,percent.function))
res4=t(apply(X,2,percent.function))
#Storing EQM
res1=cbind(res1,apply(((Matriz.alpha - alpha)^2)/N,2,sum))
res2=cbind(res2,apply(((Matriz.a - beta)^2)/N,2,sum))
res3=cbind(res3,apply(((Matriz.sig.hat - sigma2)^2)/N,2,sum))
res4=cbind(res4,apply(((t(X) - x)^2)/N,1,sum))
result[, ,1]=res1
}
```

```

result[,2]=res2
result[,3]=res3
result[,4]=res4
return(result)
}
result=array(0,c(nsim,T,9,4))
for (s in 1:nsim){
est.sigma.alpha0 =4 ;
m.alpha0 =0
est.sigma.a0 =4
m.a0 =0
d0 =1.5
v0 =3
mx.0 =0
sig.x.prior =3
  result[s,,] = Storvik.P(x,y,N,
  m.alpha0,est.sigma.alpha0,
  m.a0,est.sigma.a0,
  v0,d0,
  mx.0,sig.x.prior)
}
#ALPHA
kpar=1
res1=result[, ,kpar]
plot(res1[1,,4],type="l",col="gray60",ylim=c(-3,3),
main=NULL,xlab=expression(t),ylab=expression(hat(alpha)),bty="n")
for(j in 1:7){
results2=apply(result[, ,j,kpar],2,median)
points(results2,type="l",col="gray60")
}
abline(h=true[kpar],col=2,lty=2)

#BETA
kpar=2
res1=result[, ,kpar]
plot(res1[1,,4],type="l",col="gray60",ylim=c(-3,3),
main=NULL,xlab=expression(t),ylab=expression(hat(alpha)),bty="n")
for(j in 1:7){
results2=apply(result[, ,j,kpar],2,median)
points(results2,type="l",col="gray60")
}
abline(h=true[kpar],col=2,lty=2)

#SIGMA
kpar=3
res1=result[, ,kpar]
plot(res1[1,,4],type="l",col="gray60",ylim=c(-3,3),main=NULL,
xlab=expression(t),ylab=expression(hat(alpha)),bty="n")
for(j in 1:7){
results2=apply(result[, ,j,kpar],2,median)
points(results2,type="l",col="gray60")
}
abline(h=true[kpar],col=2,lty=2)

#THETA
kpar=4
res1=result[, ,kpar]
plot(res1[1,,4],type="l",col="gray60",ylim=c(-3,3),main=NULL,
xlab=expression(t),ylab=expression(hat(alpha)),bty="n")
for(j in 1:7){
results2=apply(result[, ,j,kpar],2,median)
points(results2,type="l",col="gray60")
}
points(x,col=2,lty=2,pch="*",type="p")

```