

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
MESTRADO EM ESTATÍSTICA

FILTROS DE PARTÍCULAS:
O ALGORITMO RESAMPLE-MOVE
Ana Flávia Cupertino Pinto

Orientador: Prof. Dr. Adrian Pablo Hinojosa Luna

Belo Horizonte
Outubro / 2007

RESUMO:

Neste trabalho fornecemos uma breve introdução aos Métodos Seqüenciais de Monte Carlo, abordamos, em particular, um sistema adaptativo conhecido como Filtro de Partículas. Apresentamos o filtro Bootstrap e também uma discussão mais formal do filtro de partículas. São discutidas algumas propriedades teóricas e práticas destes algoritmos.

Sumário

1	Introdução aos Métodos Seqüências de Monte Carlo	4
1.1	Apresentação do Problema	4
1.2	Métodos de Monte Carlo	8
1.2.1	Amostragens de Monte Carlo Perfeitas	9
1.2.2	Amostragem por Importância	11
1.2.3	Amostragem por Importância Seqüencial	12
1.3	O Filtro Bootstrap	13
2	Filtro de Partículas - Uma Perspectiva Teórica	16
2.1	Notações e definições	16
2.1.1	Cadeias de Markov e Núcleos Transição	17
2.1.2	O problema da filtragem	18
2.1.3	Convergência de Medidas Aleatórias	19
2.2	Teoremas de Convergência	22
2.2.1	O Caso da Observação Fixada	22
2.2.2	O Caso da Observação Aleatória	23
2.3	Exemplo de Filtro de Partículas	28
2.3.1	Descrição do Filtro de Partículas	28
2.3.2	Convergência do Algoritmo	31
2.3.3	Um Exemplo Markoviano	38
3	Reamostragem Móvel: Método de Filtragem com Saltos entre os Modelos	41
3.1	Apresentação do Problema	41
3.2	O Algoritmo RESAMPLE-MOVE	43
3.2.1	Um Exemplo Gaussiano	47
3.2.2	Comentários	50
A	A Fórmula de Recorrência	52
A.1	A fórmula de recorrência	53
A.2	Algoritmo do capítulo 2	59
A.3	Algoritmo do capítulo 3	64

Introdução

Os Métodos de Monte Carlo consistem em técnicas de simulação para a solução de problemas de estimação onde os estimadores têm uma distribuição complexa, como por exemplo quando a função de distribuição não tem uma solução analítica. Devido ao grande potencial da metodologia Monte Carlo, várias técnicas tem sido desenvolvidas. As técnicas de filtragem via sistemas de partículas, métodos Seqüenciais de Monte Carlo (MSMC), Monte Carlo Híbrido (HMC), etc., são alguns exemplos de avanços recentes nas técnicas de Monte Carlo. Esses avanços, de uma forma ou outra, são todos motivados pela necessidade de resolver problemas na amostragem de uma de uma distribuição de probabilidade, tais como por exemplo a tendência dos algoritmos Metropolis ficarem concentrados em uma região com bastante "massa", ver [5].

As técnicas de Monte Carlo Seqüencial são generalizações do Método Monte Carlo para o caso de problemas que evoluem com o tempo, isto é, na estimação de problemas relativos aos processos estocásticos. Na maioria destes problemas queremos técnicas adaptativas (amostragem por importância, reamostragem bootstrap) pois o Método de Monte Carlo Seqüencial simples não produz estimadores eficientes. O problema da filtragem é um caso particular dos MSMC. Ele consiste na estimação de um sinal desconhecido (oculto) a partir dos dados provenientes das observações que, por sua vez, são uma função aleatória dos sinais. O estimador é chamado de filtro. Se a função anterior é linear e gaussiana, então o modelo resultante é chamado espaço-estado linear ou modelo linear dinâmico, ver [5] seção 3.3. O algoritmo resultante é conhecido como filtro de Kalman. Estamos interessados no estudo de filtros não lineares.

Uma grande variedade de filtros de Kalman tem sido desenvolvidas desde a formulação original de Rudolf E. Kalman. Em 1960, R. E. Kalman publicou um artigo descrevendo uma solução recursiva para o problema da filtragem discreta para dados lineares. Por ser um estimador recursivo, somente o estado estimado no tempo anterior e a observação no tempo atual são necessárias para calcular a estimativa do estado atual. Uma introdução ao filtro de Kalman discreto básico e algumas discussões podem ser encontradas em [6].

O melhor filtro não linear é $\hat{X}_t = \mathbb{E}[X_t|Y_t]$, que é o estimador com menor variância. Em geral, não temos uma expressão fechada para a esperança condicional, essencialmente devido ao fato de que a distribuição condicional não é conhecida, exceto por uma constante normalizadora. Portanto, é necessário adotarmos técnicas numéricas para obtermos aproximações razoáveis. Os Métodos Seqüenciais de Monte Carlo (MSMC) são uma poderosa ferramenta que nos permite alcançar este objetivo. Na seção 4.4 de [7] é apresentado uma aplicação do MSMC para um problema que surge em comunicações digitais.

Os métodos de Monte Carlo usando Cadeias de Markov (MCMC) são uma aproximação comum

para amostragens de uma distribuição de probabilidade complexa. No entanto, os métodos MCMC são algoritmos iterativos, inadequados para problemas de estimação recursivos. Uma solução alternativa clássica consiste no uso dos métodos de amostragem por importância e amostragem por importância seqüencial. O método amostragem por importância, em sua forma simples, não é adequado para estimação recursiva (sua complexidade computacional aumenta com o tempo). O problema apresentado pelo método amostragem por importância seqüencial é que quando o tempo aumenta, a distribuição dos pesos torna-se cada vez mais assimétrica e conseqüentemente o algoritmo falha ao tentar representar a distribuição a posteriori adequadamente. A idéia principal do filtro bootstrap é eliminar as partículas tendo um pequeno peso da importância e multiplicar as partículas que possuem um grande peso da importância.

O problema da filtragem, apresentado no capítulo 2 de [1], consiste em calcular as distribuições condicionais do sinal dada a σ -álgebra gerada pela observação do processo do instante 0 até t . O filtro de partículas descrito envolve o uso de um sistema de n partículas que evoluem de acordo com um processo de Markov e, em tempos fixos, geram um certo número de descendentes. Este filtro de partículas geral é constituído basicamente por duas etapas: evolução e seleção. No passo de evolução, cada partícula no instante de tempo anterior é movida independentemente com o núcleo de transição do sinal, e assim, uma partícula candidata é obtida. No passo de seleção, cada partícula candidata é substituída por um número de descendentes, alguns possivelmente nulos (obtidos, em nosso caso, através de um mecanismo de ramificação multinomial), de tal forma que o número de partículas no instante atual permaneça constante e igual a n .

Filtros de partículas são também parte das técnicas Bayesianas dinâmicas, e visam a redução do tempo computacional de burn-in. Essas técnicas são especialmente úteis em problemas em que os dados observados são avaliados sequencialmente no tempo e estamos interessados em realizar inferência em uma modelagem on-line. Esses métodos requerem a geração de um conjunto inicial de partículas que então são progressivamente reamostrados levando em consideração os dados de entrada e os parâmetros. Em [4] é apresentado um caso particular de filtro de partículas, o algoritmo Resample-Move, proposto por Gilks e Berzuine em [3], para o contexto de sobrevivência de uma doença.

O algoritmo Resample-Move é um exemplo de uma classe de filtros de partículas chamado filtro de partículas híbrido. Embora esta técnica tenha as mesmas bases conceituais da amostragem-reamostragem por importância e amostragem MCMC, ela evita a degeneração dos métodos atuais (ver [3]).

Este trabalho está organizado da seguinte forma: o Capítulo 1 apresenta uma breve introdução aos Métodos Seqüenciais de Monte Carlo, abordando alguns sistemas adaptativos (amostragem por importância, amostragem por importância seqüencial). Também neste capítulo, consta a descrição de um filtro de partículas básico, o filtro bootstrap. O Capítulo 2 tem por objetivo dar um tratamento matemático mais rigoroso. Foi provada a consistência de um filtro de partículas genérico, apresentado na Seção 2.2. Também é apresentado neste capítulo uma implementação do algoritmo genérico para um exemplo simples de uma cadeia de Markov. O caso da normalidade assintótica não será apresentado neste trabalho e pode ser encontrado no Capítulo 3 de [1]. Finalmente, no Capítulo 3, descrevemos o algoritmo Resample-Move e apresentamos a implementação do algoritmo Resample-Move num exemplo simples de um processo gaussiano linear.

Capítulo 1

Introdução aos Métodos Seqüências de Monte Carlo

Métodos Seqüências de Monte Carlo são um conjunto de métodos de simulação que fornece uma conveniente e atrativa aproximação computacional para a distribuição a posteriori. Os Métodos SMC são muito flexíveis, fáceis para implementar e aplicáveis em muitos contextos gerais. Trabalhos interessantes em controle automático foram realizados durante os anos 60 e 70 baseados sobre métodos de integração SMC (ver [8]). No final dos anos 80 o grande aumento do poder computacional tornou possível rápidos avanços nos métodos de integração numerica para filtragem Bayesiana ([9]).

O problema da filtragem é um caso particular dos métodos SMC. Ele consiste na estimação de um sinal desconhecido (oculto) a partir dos dados provenientes das observações que, por sua vez, são uma função aleatória dos sinais. O estimador é chamado de filtro. A amostragem por importância sequencial (SIS) é um método Monte Carlo que forma as bases para a maior parte dos filtros SMC desenvolvidos sobre as décadas passadas; ver [1]. Esta aproximação SMC é conhecida diversamente como filtragem bootstrap, algoritmo de condensação, filtragem de partículas, interação de aproximação de partículas e sobrevivência do ajuste [10].

Neste capítulo apresentamos uma breve introdução aos métodos SMC, abordamos alguns sistemas adaptativos como, por exemplo, amostragem por importância e, amostragem por importância sequencial. Finalmente, consta a descrição de um filtro de partículas básico, o filtro bootstrap. Para mais detalhes ver Capítulo 1 de [1].

1.1 Apresentação do Problema

Por simplicidade, nos restringiremos a modelos de sinais quando o modelo de espaço de estados são não-gaussianos, não-lineares e Markovianos, embora os métodos SMC (Monte Carlo Seqüencial), possam ser aplicados em contextos mais gerais.

Modelo: $X_{t+1} = F(X_t, R_t)$; F é não linear e R_t ruído não-gaussiano.

Os sinais não observados (estados ocultos) $\{x_t; t \in \mathbb{N}\}$, $x_t \in X = \mathbb{R}^{n_x}$, são modelados como um processo de Markov (por exemplo, uma cadeia de Markov a tempo discreto) de distribuição inicial $p(x_0)$ e equação de transição $p(x_t|x_{t-1})$. As observações $\{y_t; t \in \mathbb{N}^*\}$, $y_t \in Y = \mathbb{R}^{n_y}$, são, por

hipótese, condicionalmente independentes dado o processo $\{x_t; t \in \mathbb{N}\}$ e a distribuição condicional marginal $p(y_t|x_{0:t}) = p(y_t|x_t)$.

Descrição do Modelo:

- $p(x_0)$ (distribuição inicial)
- $p(x_t|x_{t-1})$ (probabilidade de transição de um Processo de Markov: $p(x_{t+1}|x_{0:t}) = p(x_{t+1}|x_t)$)
- $p(y_t|x_t)$ (por hipótese, $p(y_t|x_{0:t}) = p(y_t|x_t)$)
- $p(y_s, y_u|x_{0:t}) = p(y_s|x_{0:t}) p(y_u|x_{0:t})$

Denotamos $x_{0:t} = \{x_0, \dots, x_t\}$ e $y_{1:t} = \{y_1, \dots, y_t\}$, respectivamente, os sinais e as observações até o tempo t .

Situação:

Espaço de Estados:

- $X_{t+1} = F(X_t, R_t)$ (sinal oculto)
- $Y_{t+1} = G(X_t, S_t)$ (sinal observado)

A proposta é estimar recursivamente no tempo a distribuição a posteriori, $p(x_{0:t}|y_{1:t})$, e características a ela associadas (incluindo distribuição marginal $p(x_t|y_{1:t})$ conhecida como distribuição filtrada). Para isto, basta estimar as esperanças

$$I(f_t) = E_{p(x_{0:t}|y_{1:t})}[f_t(x_{0:t})] = \int f_t(x_{0:t})p(x_{0:t}|y_{1:t})dx_{0:t} \quad (1.1)$$

(por hipótese, estaremos sempre supondo que $p(x_{0:t}|y_{1:t})$ é uma probabilidade condicional regular para todo $y_{1:t}$). Para alguma função de interesse $f_t : X^{(t+1)} \rightarrow \mathbb{R}^{n_{f_t}}$ integrável com respeito a $p(x_{0:t}|y_{1:t})$.

Distribuição filtrada: (a posteriori: $p(x_{0:t}|y_{1:t})$)

Em qualquer tempo t , a distribuição a posteriori é dada pelo Teorema de Bayes

$$p(x_{0:t}|y_{1:t}) = \frac{p(y_{1:t}|x_{0:t})p(x_{0:t})}{p(y_{1:t})} = \frac{p(y_{1:t}|x_{0:t})p(x_{0:t})}{\int p(y_{1:t}|x_{0:t})p(x_{0:t})dx_{0:t}}$$

É possível obter uma fórmula recursiva para a distribuição $p(x_{0:t}|y_{1:t})$,

Teorema 1.1.1. *Vale que:*

$$p(x_{0:t+1}|y_{1:t+1}) = p(x_{0:t}|y_{1:t}) \frac{p(y_{t+1}|x_{t+1})p(x_{t+1}|x_t)}{p(y_{t+1}|y_{1:t})}. \quad (1.2)$$

Demonstração. Temos que, pelo Teorema de Bayes

$$p(x_{0:t+1}|y_{1:t+1}) = \frac{p(y_{1:t+1}|x_{0:t+1})p(x_{0:t+1})}{p(y_{1:t+1})}. \quad (1.3)$$

Observe que,

$$\begin{aligned} p(y_{1:t+1}|x_{0:t+1}) &= p(y_{1:t}, y_{t+1}|x_{0:t+1}) \\ &= p(y_{1:t}|x_{0:t+1})p(y_{t+1}|x_{0:t+1}) \\ &= p(y_{1:t}|x_{0:t})p(y_{t+1}|x_{t+1}) \\ &= \frac{p(x_{0:t}|y_{1:t})p(y_{1:t})}{p(x_{0:t})}p(y_{t+1}|x_{t+1}) \end{aligned}$$

Substituindo em (1.3), obtemos

$$p(x_{0:t+1}|y_{1:t+1}) = p(x_{0:t}|y_{1:t})p(y_{t+1}|x_{t+1}) \frac{p(y_{1:t})}{p(x_{0:t})} \frac{p(x_{0:t+1})}{p(y_{1:t+1})}. \quad (1.4)$$

Note que:

$$p(y_{t+1}|y_{1:t}) = \frac{p(y_{t+1}, y_{1:t})}{p(y_{1:t})} = \frac{p(y_{1:t+1})}{p(y_{1:t})}. \quad (1.5)$$

Analogamente,

$$p(x_{t+1}|x_t) = p(x_{t+1}|x_{0:t}) = \frac{p(x_{t+1}, x_{0:t})}{p(x_{0:t})} = \frac{p(x_{0:t+1})}{p(x_{0:t})} \quad (1.6)$$

Assim, substituindo (1.5) e (1.6) em (1.4) resulta

$$p(x_{0:t+1}|y_{1:t+1}) = \frac{p(x_{0:t}|y_{1:t})p(y_{t+1}|x_{t+1})p(x_{t+1}|x_t)}{p(y_{t+1}|y_{1:t})}$$

Como queríamos provar. □

A distribuição marginal $p(x_t|y_{1:t})$ também satisfaz a seguinte recursão.

Teorema 1.1.2. *Vale que (Predição):*

$$p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1} \quad (1.7)$$

Demonstração. Observe que:

$$\begin{aligned} p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1}) &= \frac{p(x_t, x_{t-1})}{p(x_{t-1})} \frac{p(y_{1:t-1}|x_{t-1})p(x_{t-1})}{p(y_{1:t-1})} \\ &= p(x_t, x_{t-1}) \frac{p(y_{1:t-1}|x_{t-1})}{p(y_{1:t-1})} \end{aligned}$$

Logo tomando a integral de ambos os membros na expressão acima, obtemos

$$\int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1} = \int \frac{p(x_t, x_{t-1})p(y_{1:t-1}|x_{t-1})}{p(y_{1:t-1})}dx_{t-1} \quad (1.8)$$

Agora,

$$p(x_t|y_{1:t-1}) = \frac{p(y_{1:t-1}|x_t)p(x_t)}{p(y_{1:t-1})} = \frac{p(y_{1:t-1}, x_t)p(x_t)}{p(x_t)p(y_{1:t-1})} = \frac{p(y_{1:t-1}, x_t)}{p(y_{1:t-1})}$$

Note que:

$$\begin{aligned} p(y_{1:t-1}, x_t) &= \int p(y_{1:t-1}, x_t, x_{t-1})dx_{t-1} \\ &= \int p(y_{1:t-1}|x_t, x_{t-1})p(x_t, x_{t-1})dx_{t-1} \\ &= \int p(y_{1:t-1}|x_{t-1})p(x_t, x_{t-1})dx_{t-1} \end{aligned} \quad (1.9)$$

Desta forma, por (1.9)

$$\begin{aligned} p(x_t|y_{1:t-1}) &= \frac{1}{p(y_{1:t-1})} \int p(y_{1:t-1}|x_{t-1})p(x_t, x_{t-1})dx_{t-1} \\ &= \int \frac{p(y_{1:t-1}|x_{t-1})p(x_t, x_{t-1})}{p(y_{1:t-1})}dx_{t-1} \end{aligned} \quad (1.10)$$

De (1.8) e (1.10) obtemos a igualdade desejada. □

Teorema 1.1.3. (*Correção*) Vale que:

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{\int p(y_t|x_t)p(x_t|y_{1:t-1})dx_t} \quad (1.11)$$

Demonstração. Observe que,

$$\begin{aligned} p(y_t|x_t)p(x_t|y_{1:t-1}) &= \frac{p(y_t|x_t)p(y_{1:t-1}|x_t)p(x_t)}{p(y_{1:t-1})} \\ &= \frac{p(y_t, y_{1:t-1}|x_t)p(x_t)}{p(y_{1:t-1})} \\ &= \frac{p(y_{1:t}|x_t)p(x_t)}{p(y_{1:t-1})} \end{aligned}$$

Agora,

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

Desta forma,

$$\begin{aligned}
\frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{\int p(y_t|x_t)p(x_t|y_{1:t-1})dx_t} &= \frac{p(y_{1:t}|x_t)p(x_t)}{p(y_{1:t-1})} \frac{p(y_{1:t-1})}{p(y_{1:t})} \\
&= \frac{p(y_{1:t}|x_t)p(x_t)}{p(y_{1:t})}
\end{aligned}$$

Como pelo Teorema de Bayes,

$$p(x_t|y_{1:t}) = \frac{p(y_{1:t}|x_t)p(x_t)}{p(y_{1:t})}$$

chegamos ao resultado desejado. □

Essas expressões e recursões são enganosamente simples porque não podemos calcular tipicamente a constante normalizadora $p(y_{1:t})$, as marginais $p(x_t|y_t)$, e $I(f_t)$ pois elas requerem a avaliação de integrais complexas de grandes dimensões.

Em vista dessas dificuldades, vários artigos e livros tem sido dedicados a obtenção de métodos de aproximações dessas distribuições, incluindo, entre eles, o filtro de Kalman estendido, o filtro da soma gaussiana, métodos tipo malha (grid-based) e métodos de integração baseados sobre SMC.

O grande aumento do poder computacional no final dos anos 80 tornou possível rápidos avanços nos métodos de integração numérica por filtragem Bayesiana.

1.2 Métodos de Monte Carlo

Os métodos de integração de Monte Carlo apresentam grandes vantagens por não estarem sujeitos a qualquer limitação de linearidade e gaussianidade sobre o modelo e estes métodos também tem apresentado propriedades boas de convergência.

Começaremos mostrando que, quando temos um grande número de amostras de uma distribuição a posteriori, não é difícil aproximar as integrais intratáveis presentes nas equações , (1.2), (1.7) e

(1.11). No entanto, é raramente possível obter amostras dessas distribuições diretamente, porque falta a contante normalizadora.

Portanto temos que recorrer aos métodos alternativos de Monte Carlo, tal como a amostragem por importância. Fazendo uso desta técnica recursivamente obtemos o método de amostragem por importância sequencial (SIS).

Infelizmente, podemos mostrar que SIS com certeza vai falhar quando t aumenta, pois neste caso os erros cometidos vão sendo acumulados e logo aumentam quando t cresce.

Este problema pode ser superado incluindo um passo de seleção adicional. A introdução deste passo chave nos fornece o primeiro método *operacionalmente efetivo* (o passo chave consiste em minimizar o erro em relação a cada valor amostrado evitando assim o acúmulo excessivo dos erros). Desde então, resultados de convergência teórica para este algoritmo tem sido estabelecidas.

1.2.1 Amostragens de Monte Carlo Perfeitas

Suponha que seja possível obter n amostras aleatórias independentes e identicamente distribuídas, também chamadas partículas $\{x_{0:t}^{(i)}; i = 1, \dots, n\}$ de acordo com a distribuição a posteriori $p(x_{0:t}|y_{1:t})$. Uma estimativa empírica desta distribuição é dada por:

$$P_n(dx_{0:t}|y_{1:t}) := \frac{1}{n} \sum_{i=1}^n \delta_{x_{0:t}^{(i)}}(dx_{0:t}) \quad (1.12)$$

sendo que $dx_{0:t}$ é a densidade conjunta com respeito a medida de contagem e $\delta_{x_{0:t}^{(i)}}(dx_{0:t})$ é a função delta de Dirac ("massa" concentrada em $x_{0:t}^{(i)}$).

Considerando um conjunto $A \in \mathbb{R}^{t+1}$, temos

$$P_n(X_{0:t} \in A|y_{1:t}) = \int_A P_n(dx_{0:t}|y_{1:t})$$

Agora, por (1.12),

$$\begin{aligned} \int_A P_n(dx_{0:t}|y_{1:t}) &= \int_A \frac{1}{n} \sum_{i=1}^n \delta_{x_{0:t}^{(i)}}(dx_{0:t}) \\ &= \frac{1}{n} \sum_{i=1}^n \int_A \delta_{x_{0:t}^{(i)}}(dx_{0:t}) \\ &= \frac{1}{n} \sum_{i=1}^n \delta_{x_{0:t}^{(i)}}(A) \\ &= \frac{1}{n} \sum_{i=1}^n I_A(x_{0:t}^{(i)}) \\ &= \frac{\#\{x_{0:t}^{(i)} : x_{0:t}^{(i)} \in A, 1 \leq i \leq n\}}{\#\{x_{0:t}^{(i)} : 1 \leq i \leq n\}} \end{aligned} \quad (1.13)$$

O último termo acima é a distribuição empírica de $p(x_{0:t}|y_{1:t})$ para o caso em que estamos considerando o conjunto $A \in \mathbb{R}^{t+1}$.

Assim, obtemos a seguinte estimativa de $I(f_t)$

$$\begin{aligned} I_n(f_t) &= \int f_t(x_{0:t}) P_n(dx_{0:t}|y_{1:t}) = \int f_t(x_{0:t}) \frac{1}{n} \sum_{i=1}^n \delta_{x_{0:t}^{(i)}}(dx_{0:t}) \\ &= \frac{1}{n} \sum_{i=1}^n \int f_t(x_{0:t}) \delta_{x_{0:t}^{(i)}}(dx_{0:t}) \\ &= \frac{1}{n} \sum_{i=1}^n f_t(x_{0:t}^{(i)}) \end{aligned}$$

Esta estimativa é não viciada, pois

$$E(I_n(f_t)) = E\left(\frac{1}{n} \sum_{i=1}^n f_t(x_{0:t}^{(i)})\right) = E(f_t(x_{0:t})) = I(f_t)$$

Se a variância a posteriori de f_t satisfaz

$$\sigma_{f_t}^2 = E_{p(x_{0:t}|y_{1:t})}(f_t^2(x_{0:t})) - I^2(f_t) < \infty,$$

então a variância de $I_n(f_t)$ é igual a

$$\text{Var}(I_n(f_t)) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n f_t(x_{0:t}^{(i)})\right) = \frac{1}{n} \text{Var}(f_t(x_{0:t})) = \frac{1}{n} \sigma_{f_t}^2 < \infty$$

Pela Lei forte dos grandes números $I_n(f_t) \rightarrow I(f_t)$, q.c., quando $n \rightarrow \infty$. Temos ainda que, se $\sigma_{f_t}^2 < \infty$, então pelo Teorema Central do Limite

$$\frac{I_n(f_t) - I(f_t)}{\frac{\sigma_{f_t}}{\sqrt{n}}} \rightarrow N(0, 1), \text{ em distribuição, quando } n \rightarrow \infty$$

$$\therefore \sqrt{n}(I_n(f_t) - I(f_t)) \rightarrow N(0, \sigma_{f_t}^2), \text{ em distribuição, quando } n \rightarrow \infty$$

Vantagens do Método de Amostragem Perfeita de Monte Carlo

- O conjunto de amostras aleatórias $\{x_{0:t}^{(i)}, i = 1, \dots, n\}$ podem facilmente estimar qualquer quantidade $I(f_t)$,

$$I(f_t) \cong I_n(f_t) = \frac{1}{n} \sum_{i=1}^n f_t(x_{0:t}^{(i)})$$

$(I_n(f_t))$ é um estimador não viciado e assintoticamente consistente para $I(f_t)$ se a variância a posteriori de $f_t(x_{0:t})$, $\sigma_{f_t}^2$, é finita).

- A taxa de convergência desta estimativa é independente da dimensão do integrando, em contraste com alguns métodos de integração numérica que tem uma taxa de convergência que diminui quando a dimensão do integrando aumenta.

De fato, basta observarmos que

$$P(|I_n(f_t) - I(f_t)| > \epsilon) \leq \frac{\text{Var}(I_n(f_t))}{\epsilon} = \frac{\sigma_{f_t}^2}{n\epsilon} = cT(n),$$

sendo que $c = \frac{\sigma_{f_t}^2}{\epsilon}$ e $T(n) = \frac{1}{n}$. Observe que a taxa de convergência $cT(n)$ independe da dimensão do integrando, que em nosso caso é $(t+1)$, mas apenas do número de simulações n .

Em Estatística Aplicada, Métodos de Monte Carlo usando Cadeias de Markov (MCMC) são uma aproximação comum para amostragens de uma distribuição de probabilidade complexa. No entanto, os métodos MCMC são algoritmos iterativos, inadequados para problemas de estimação recursivos. Desta forma, métodos alternativos, como por exemplo, a amostragem por importância, tem sido desenvolvidos.

1.2.2 Amostragem por Importância

Uma solução alternativa clássica consiste no uso do método amostragem por importância. Introduzimos uma distribuição amostragem por importância arbitrária, $q(x_{0:t}|y_{1:t})$, também referida frequentemente como distribuição proposta ou função importância. Suponhamos que queremos avaliar $I(f_t)$, e sob a condição de que o suporte de $q(x_{0:t}|y_{1:t})$ contém o suporte de $p(x_{0:t}|y_{1:t})$, obtemos

$$I(f_t) = \frac{\int f_t(x_{0:t})w(x_{0:t})q(x_{0:t}|y_{1:t})dx_{0:t}}{\int w(x_{0:t})q(x_{0:t}|y_{1:t})dx_{0:t}} \quad (1.14)$$

sendo que,

$$w(x_{0:t}) = \frac{p(x_{0:t}|y_{1:t})}{q(x_{0:t}|y_{1:t})} \quad (1.15)$$

é chamado o *peso da importância*. Observe que substituindo (1.15) na expressão (1.14), obtemos a integral $I(f_t)$ como definida em (1.1).

Considerando uma aproximação da distribuição a posteriori, $p(x_{0:t}|y_{1:t})$, dada por

$$\hat{P}_n(dx_{0:t}|y_{1:t}) = \sum_{i=1}^n \tilde{w}_t^{(i)} \delta_{x_{0:t}^{(i)}}(dx_{0:t}) \quad (1.16)$$

então, se podemos simular n partículas i.i.d. $\{x_{0:t}^{(i)}, i = 1, \dots, n\}$ de acordo com a distribuição $q(x_{0:t}|y_{1:t})$, uma estimativa de Monte Carlo $I(f_t)$ é dada por

$$\begin{aligned} \hat{I}_n(f_t) &= \frac{\frac{1}{n} \sum_{i=1}^n f_t(x_{0:t}^{(i)})w(x_{0:t}^{(i)})}{\frac{1}{n} \sum_{j=1}^n w(x_{0:t}^{(j)})} \\ &= \sum_{i=1}^n f_t(x_{0:t}^{(i)})\tilde{w}_t^{(i)}, \end{aligned} \quad (1.17)$$

sendo que $\tilde{w}_t^{(i)} = \frac{w(x_{0:t}^{(i)})}{\sum_{j=1}^n w(x_{0:t}^{(j)})}$. Este método de integração pode também ser interpretado como um método de amostragem onde a distribuição a posteriori $p(x_{0:t}|y_{1:t})$ é aproximada por (1.16), e $\hat{I}_n(f_t)$ é a integral da função $f_t(x_{0:t})$ com respeito a medida empírica $\hat{P}_n(dx_{0:t}|y_{1:t})$, isto é

$$\hat{I}_n(f_t) = \int f_t(x_{0:t}) \hat{P}_n(dx_{0:t}|y_{1:t}).$$

Observe que para n finito, $\hat{I}_n(f_t)$ é viciado (razão de duas estimativas), mas assintoticamente, sob condições fracas (i.e., $f_t(x_{0:t}^{(i)})w(x_{0:t}^{(i)})$, $i = 1, \dots, n$ amostra aleatória, $\mathbb{E}[f_t(x_{0:t})w(x_{0:t})] < \infty$ e $\text{Var}(x_{0:t}) < \infty$) a aplicação da lei forte dos grandes números garante que

$$\begin{aligned} \lim_{n \rightarrow \infty} \hat{I}_n(f_t) &= \frac{\mathbb{E}[f_t(x_{0:t})w(x_{0:t})]}{\mathbb{E}[w(x_{0:t})]} \\ &= \frac{\int f_t(x_{0:t})w(x_{0:t})q(x_{0:t}|y_{1:t})dx_{0:t}}{\int w(x_{0:t})q(x_{0:t}|y_{1:t})dx_{0:t}} \\ &= I(f_t), \quad q.c. \end{aligned}$$

A Amostragem por importância é um método de integração Monte Carlo geral. No entanto, nesta forma simples, ele não é adequado para estimação recursiva. Isto é, precisamos obter todos os dados, $y_{1:t}$, antes de avaliarmos $p(x_{0:t}|y_{1:t})$. Em geral, a cada novo dado disponível, y_{t+1} , precisamos recalculamos o peso da importância sobre a seqüência inteira de estados. A complexidade computacional desta operação aumenta com o tempo, o que nos motiva o uso de uma outra forma de amostragem por importância. Na próxima seção apresentamos uma estratégia para a superação deste problema.

1.2.3 Amostragem por Importância Seqüencial

O método de amostragem por importância pode ser modificado de tal forma a tornar possível o cálculo da estimativa $\hat{P}_n(dx_{0:t}|y_{1:t})$ de $p(x_{0:t}|y_{1:t})$ sem modificar as trajetórias passadas simuladas $\{x_{0:t-1}^{(i)}; i = 1, \dots, n\}$.

Consideremos as funções importância $q(x_{0:t}|y_{1:t})$, que satisfazem a seguinte relação

$$q(x_{0:t}|y_{1:t}) = q(x_{0:t-1}|y_{1:t-1})q(x_t|x_{0:t-1}, y_{1:t}) \quad (1.18)$$

Iterando, obtemos

$$q(x_{0:t}|y_{1:t}) = q(x_0) \prod_{k=1}^t q(x_k|x_{0:k-1}, y_{1:k})$$

A função importância dada por (1.18) nos permite avaliar recursivamente no tempo o peso importância. De fato, observe que para $i = 1, \dots, n$,

$$p(x_{0:t}^{(i)}|y_{1:t}) = p(x_{0:t-1}^{(i)}|y_{1:t-1}) \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{p(y_t|y_{1:t-1})} \Rightarrow$$

$$\begin{aligned}
w(x_{0:t}^{(i)}) &= \frac{p(x_{0:t-1}^{(i)}|y_{1:t-1})}{q(x_{0:t-1}^{(i)}|y_{1:t-1})q(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{1:t})} \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{p(y_t|y_{1:t-1})} \Rightarrow \\
w(x_{0:t}^{(i)}) &= \frac{w(x_{0:t-1}^{(i)})}{p(y_t|y_{1:t-1})} \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{1:t})} \Rightarrow \\
\tilde{w}_t^{(i)} &= \frac{\tilde{w}_{t-1}^{(i)}}{p(y_t|y_{1:t-1})} \frac{\sum_{j=1}^n w(x_{0:t-1}^{(j)})}{\sum_{j=1}^n w(x_{0:t}^{(j)})} \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{1:t})} \Rightarrow \\
\tilde{w}_t^{(i)} &\propto \tilde{w}_{t-1}^{(i)} \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{1:t})}
\end{aligned}$$

sendo que a constante de proporcionalidade é $\frac{1}{p(y_t|y_{1:t-1})} \frac{\sum_{j=1}^n w(x_{0:t-1}^{(j)})}{\sum_{j=1}^n w(x_{0:t}^{(j)})}$.

Um caso especial da expressão (1.18) é considerar a distribuição a priori como distribuição importância

$$q(x_{0:t}|y_{1:t}) = p(x_{0:t}) = p(x_0) \prod_{k=1}^t p(x_k|x_{k-1})$$

pois, recordemos que, por hipótese, os sinais não observados $\{x_t; t \in \mathbb{N}\}$ são modelados por um processo de Markov de distribuição inicial $p(x_0)$ e probabilidade de transição $p(x_t|x_{t-1})$. Para este caso especial, os pesos importância satisfazem

$$\tilde{w}_t^{(i)} \propto \tilde{w}_{t-1}^{(i)} p(y_t|x_t^{(i)}).$$

De fato, basta observar que, para este exemplo, $q(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{1:t}) = p(x_t^{(i)}|x_{t-1}^{(i)})$.

1.3 O Filtro Bootstrap

O problema apresentado pelo método de amostragem por importância sequencial, é que quando t aumenta, a distribuição de $\tilde{w}_t^{(i)}$ torna-se cada vez mais assimétrica. Praticamente após um curto espaço de tempo (discreto), somente algumas partículas tem o peso da importância não-nulos.

O algoritmo, conseqüentemente, falha ao tentar representar a distribuição a posteriori adequadamente. Para evitar estes problemas, precisamos introduzir um passo adicional de seleção das partículas. O método descrito abaixo foi obtido do Capítulo 1 de [1].

Apresentação do Método

A idéia chave do filtro Bootstrap é eliminar as partículas tendo um pequeno peso da importância e multiplicar as partículas que possuem um grande peso da importância. Formalmente, substituímos a distribuição empírica com pesos

$$\hat{P}_n(dx_{0:t}|y_{1:t}) = \sum_{i=1}^n \tilde{w}_t^{(i)} \delta_{x_{0:t}^{(i)}}(dx_{0:t})$$

por uma distribuição empírica sem pesos

$$\hat{P}_n(dx_{0:t}|y_{1:t}) = \frac{1}{n} \sum_{i=1}^n n_t^{(i)} \delta_{x_{0:t}^{(i)}}(dx_{0:t})$$

onde $n_t^{(i)} \in \mathbb{N}$ é o número de nascimentos (partículas criadas) associados a trajetória $x_{0:t}^{(i)}$ e $\sum_{i=1}^n n_t^{(i)} = n$.

Se $n_t^{(j)} = 0$, então a j -ésima trajetória $x_{0:t}^{(j)}$ morreu. Os valores de $n_t^{(i)}$, $i = 1, \dots, n$, são escolhidos tais que $\hat{P}_n(dx_{0:t}|y_{1:t})$ estão próximos de $\hat{P}_n(dx_{0:t}|y_{1:t})$ tal que para qualquer função f_t ,

$$\int f_t(x_{0:t}) \hat{P}_n(dx_{0:t}|y_{1:t}) \approx \int f_t(x_{0:t}) \hat{P}_n(dx_{0:t}|y_{1:t}).$$

Depois do passo de seleção, as trajetórias sobreviventes $x_{0:t}^{(i)}$, isto é, aquelas tais que $n_t^{(i)} > 0$, são aproximadamente distribuídas de acordo com a distribuição a posteriori $p(x_{0:t}|y_{1:t})$. Existem diferentes maneiras de selecionar $n_t^{(i)}$, uma delas é obtida através de uma distribuição multinomial, $M(n, \tilde{w}_t^{(1)}, \dots, \tilde{w}_t^{(n)})$, tendo distribuição de probabilidade dada por,

$$P(M = (n_1, \dots, n_n)) = \frac{n!}{n_1! \dots n_n!} (\tilde{w}_t^{(1)})^{n_1} \dots (\tilde{w}_t^{(n)})^{n_n}$$

Descrição do Algoritmo

Filtro Bootstrap

Passo 1 : Distribuição inicial, $t = 0$.

- Iniciamos com uma amostra $x_0^{(i)} \sim p(x_0); i = 1, \dots, n$.
- Faça $t = 1$.

Passo 2 : Amostragem por importância.

Para $i = 1, \dots, n$,

- Amostramos $\tilde{x}_t^{(i)} \sim p(x_t|x_{t-1}^{(i)})$.
- Seja, $\tilde{x}_{0:t}^{(i)} = (x_{0:t-1}^{(i)}, \tilde{x}_t^{(i)})$.
- Para $i = 1, \dots, n$, calculamos o peso da importância,

$$\hat{w}_t^{(i)} = p(y_t|\tilde{x}_t^{(i)})$$

- Normalizamos o peso da importância,

$$\tilde{w}_t^{(i)} = c \hat{w}_t^{(i)} p(y_t|\tilde{x}_t^{(i)})$$

Passo 3 : Seleção.

- Reamostramos com reposição n processos(partículas) $(x_{0:t}^{(i)}; i = 1, \dots, n)$ do conjunto $\{\tilde{x}_{0:t}^{(i)}; i = 1, \dots, n\}$ de acordo com os pesos da importância.
- Fazemos $t = t + 1$ e voltamos ao passo 2.

Uma representação gráfica do algoritmo é mostrada na figura 1.1.

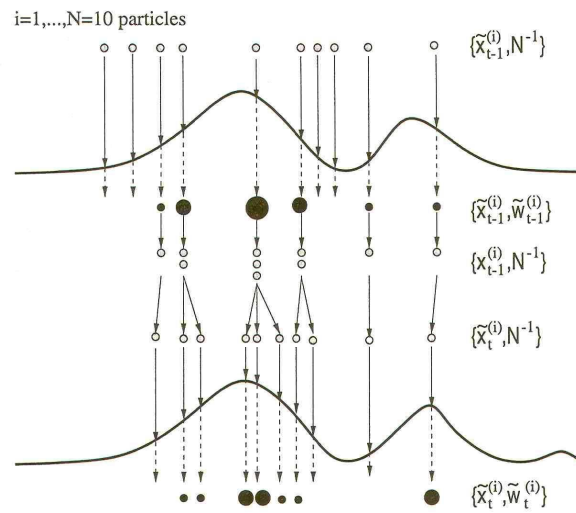


Figura 1.1: Representação gráfica do filtro bootstrap para um número de partículas $n = 10$. Gráfico extraído de [1], capítulo 1.

Capítulo 2

Filtro de Partículas - Uma Perspectiva Teórica

O objetivo deste capítulo é dar um tratamento matemático mais rigoroso da convergência dos Filtros de Partículas. O resultado sobre a consistência de um filtro de partículas genérico é apresentado na seção 2.2. Na seção 2.3 é dada a descrição do algoritmo Filtro de Partículas Espaço-Estado como uma aplicação dos resultados da seção 2.2. A convergência desta classe de Filtro de Partículas consta na subseção 2.3.2. Esta classe inclui vários filtros conhecidos, tais como os apresentados em [11], [12] e [13]. Para mais detalhes ver Capítulo 2 de [1].

Uma aplicação ilustrativa do Filtro de Partículas Espaço-Estado é apresentado na subseção 2.3.3. No apêndice é detalhado a demonstração das fórmulas de recorrência que surgem na seção 2.1.2, relação 2.4. Como referência para esta demonstração ver apêndice de [1].

2.1 Notações e definições

Seja \mathbb{R}^d o espaço Euclidiano d-dimensional e $\mathcal{B}(\mathbb{R}^d)$ a σ -álgebra de Borel de subconjuntos de \mathbb{R}^d . Usaremos as seguintes notações:

- $B(\mathbb{R}^d)$: o conjunto das funções limitadas, $\mathcal{B}(\mathbb{R}^d)$ -mensuráveis definidas sobre \mathbb{R}^d .
- $\mathcal{C}_b(\mathbb{R}^d)$: o conjunto das funções contínuas e limitadas definidas sobre \mathbb{R}^d .
- $\mathcal{C}_k(\mathbb{R}^d)$: o conjunto das funções contínuas com suporte compacto definidas sobre \mathbb{R}^d .
- $\mathcal{M}_F(\mathbb{R}^d)$: o conjunto das medidas finitas sobre $\mathcal{B}(\mathbb{R}^d)$.
- $\mathcal{P}(\mathbb{R}^d)$: o conjunto das medidas de probabilidade sobre $\mathcal{B}(\mathbb{R}^d)$.
- δ_a : a medida de Dirac concentrada em $a \in \mathbb{R}^d$.
- $\bar{1}$: a função constante 1.

Seguem algumas definições:

- Seja $f : \mathbb{R}^d \rightarrow \mathbb{R}$ uma função contínua. O suporte compacto de f é

$$\text{Supp}_k(f) = \overline{\{x : f(x) \neq 0\}}$$

- Se $f \in \mathcal{C}_b(\mathbb{R}^d)$,

$$\|f\| := \sup_{x \in \mathbb{R}^d} |f(x)|$$

é a norma do supremo de f .

- Se $\mu \in \mathcal{M}_F(\mathbb{R}^d)$ (ou $\mu \in \mathcal{P}(\mathbb{R}^d)$) e $f \in B(\mathbb{R}^{n_x})$, então

$$\mu f := \int_{\mathbb{R}^d} f(x) \mu(dx)$$

é a integral de f com respeito a medida μ .

- Seja $(\mu_n)_{n=1}^\infty$ uma seqüência de medidas finitas. Dizemos que μ_n converge fracamente ou na topologia fraca para $\mu \in \mathcal{M}_F(\mathbb{R}^d)$ (notação: $\mu_n \xrightarrow{w} \mu$) se

$$\lim_{n \rightarrow \infty} \mu_n f = \mu f, \forall f \in \mathcal{C}_b(\mathbb{R}^d),$$

a definição é análoga no caso de $\mu_n \in \mathcal{P}(\mathbb{R}^d), n \in \mathbb{N}$.

2.1.1 Cadeias de Markov e Núcleos Transição

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e $X = \{X_t, t \in \mathbb{N}\}$ um processo estocástico definido sobre (Ω, \mathcal{F}, P) com valores em \mathbb{R}^{n_x} . Seja \mathcal{F}_t^X a σ -álgebra gerada pelo processo, isto é, $\mathcal{F}_t^X := \sigma(X_s : s \in [0, t])$. Então X é uma cadeia de Markov, se $\forall t \in \mathbb{N}$ e $A \in \mathcal{B}(\mathbb{R}^{n_x})$,

$$P(X_{t+1} \in A | \mathcal{F}_t^X) = P(X_{t+1} \in A | X_t).$$

Segue a definição do Núcleo de Transição. Seja X uma cadeia de Markov. Uma função $K_t : \mathbb{R}^{n_x} \times \mathcal{B}(\mathbb{R}^{n_x}) \rightarrow \mathbb{R}$ tal que

$$K_t(x, A) = P(X_{t+1} \in A | X_t = x) = P_{X_{t+1}}(A | X_t = x)$$

é denominada o núcleo de transição da cadeia de Markov.

Propriedades do núcleo de transição K_t .

- $K_t(x, \cdot)$ é uma medida de probabilidade sobre $\mathbb{R}^{n_x}, \forall t \in \mathbb{N}$ e $x \in \mathbb{R}^{n_x}$ (K_t é uma probabilidade condicional regular)
- $K_t(\cdot, A) \in B(\mathbb{R}^{n_x}), \forall t \in \mathbb{N}$ e $A \in \mathcal{B}(\mathbb{R}^{n_x})$.

A distribuição de X é unicamente determinada pela distribuição inicial e o núcleo de transição. Seja q_t a distribuição da variável aleatória X_t ,

$$q_t(A) := P(X_t \in A).$$

Definimos a medida,

$$(q_t K_t)(A) := \int_{\mathbb{R}^{n_x}} K_t(x, A) q_t(dx); \quad A \in \mathcal{B}(\mathbb{R}^{n_x}).$$

Desta forma, obtemos a fórmula de recorrência $q_{t+1} = q_t K_t$. Observe que

$$(q_t K_t)(A) = \int_{\mathbb{R}^{n_x}} P(X_{t+1} \in A | X_t = x) q_t(dx) = P(X_{t+1} \in A) := q_{t+1}(A).$$

$$\therefore q_t = q_0 K_0 K_1 \dots K_{t-1}.$$

O núcleo de transição K_t satisfaz a propriedade de Feller se, para todo $t > 0$, a função $K_t f : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$ definida como

$$(K_t f)(x) := \int_{\mathbb{R}^{n_x}} f(y) K_t(x, dy)$$

é contínua, $\forall f \in \mathcal{C}_b(\mathbb{R}^d)$. Se K_t tem a propriedade de Feller, então $K_t f \in \mathcal{C}_b(\mathbb{R}^d)$, $\forall f \in \mathcal{C}_b(\mathbb{R}^d)$. Logo, dizemos que o processo $X = \{X_t; t \in \mathbb{N}\}$ é Feller se o núcleo de transição tem a propriedade de Feller.

2.1.2 O problema da filtragem

Informalmente, o problema de filtragem consiste em calcular as distribuições condicionais do sinal dada a σ -álgebra gerada pela observação do processo do instante 0 até t . Seja $X = \{X_t, t \in \mathbb{N}\}$ um processo de Markov com valores em \mathbb{R}^{n_x} (chamado o processo do sinal) com um núcleo de transição Feller, $K_t(x, dy)$. Seja também $Y = \{Y_t, t \in \mathbb{N}\}$ um processo estocástico assumindo valores em \mathbb{R}^{n_y} (chamado a observação do processo) definido como,

$$Y_t := h(t, X_t) + W_t, \quad t > 0 \text{ e } Y_0 = 0. \quad (2.1)$$

Em (2.1), temos que $h : \mathbb{N} \times \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_y}$ é uma função Borel-mensurável com a propriedade que $h(t, \cdot)$ é contínua sobre \mathbb{R}^{n_x} , $\forall t \in \mathbb{N}$. Temos ainda que, para todo $t > 0$, $W_t : \Omega \rightarrow \mathbb{R}^{n_y}$ são vetores aleatórios independentes e absolutamente contínuos com respeito a medida de Lebesgue λ sobre \mathbb{R}^{n_y} . Denotamos por $g(t, \cdot)$ a densidade de W_t com respeito a medida de Lebesgue λ , que existe pelo teorema de Radon-Nykodin, e supomos que $g(t, \cdot)$ é contínua e limitada. Como mencionamos no início desta seção, no problema da filtragem estaremos interessados em obter a medida de probabilidade \tilde{p}_t , tal que

$$\tilde{p}_t(A) := P(X_t \in A | \sigma(Y_{0:t})), \quad \tilde{p}_t f = \mathbb{E}[f(X_t) | \sigma(Y_{0:t})] \quad (2.2)$$

para toda $f \in B(\mathbb{R}^{n_x})$ e $A \in \mathcal{B}(\mathbb{R}^{n_x})$, onde $Y_{0:t} := (Y_0, Y_1, \dots, Y_t)$. Então $\tilde{p}_t = \tilde{p}_t^{Y_{0:t}}$, sendo que

$$\tilde{p}_t^{y_{0:t}}(A) := P(X_t \in A | Y_{0:t} = y_{0:t}), \quad \tilde{p}_t^{y_{0:t}} f = \mathbb{E}[f(X_t) | Y_{0:t} = y_{0:t}] \quad (2.3)$$

e $y_{0:t} := (y_0, y_1, \dots, y_t) \in (\mathbb{R}^{n_y})^{t+1}$. Observe que \tilde{p}_t é uma medida de probabilidade aleatória enquanto que $\tilde{p}_t^{y_{0:t}}$ é uma medida de probabilidade determinística. Definimos também p_t e $p_t^{y_{0:t-1}}$, como sendo as medidas de probabilidade condicional preditas para $t > 0$, onde $p_t = p_t^{Y_{0:t-1}}$ e

$$p_t^{y_{0:t-1}}(A) = P(X_t \in A | Y_{0:t-1} = y_{0:t-1}), \quad p_t^{y_{0:t-1}} f = \mathbb{E}[f(X_t) | Y_{0:t-1} = y_{0:t-1}]$$

As seguintes relações de recorrência se asseguram para \tilde{p}_t e $\tilde{p}_t^{y_{0:t}}$ respectivamente.

$$\begin{cases} \frac{d\tilde{p}_t}{dp_t} = \frac{g_t^{Y_t}}{\int g_t^{Y_t}(x)p_t(dx)} & , & \begin{cases} \frac{d\tilde{p}_t^{y_{0:t}}}{dp_t^{y_{0:t-1}}} = \frac{g_t^{y_t}}{\int g_t^{y_t}(x)p_t^{y_{0:t-1}}(dx)} & , \\ p_{t+1} = \tilde{p}_t K_t & = \tilde{p}_t^{y_{0:t}} K_t \end{cases} \end{cases} \quad (2.4)$$

onde $g_t^{y_t} := g(y_t - h(t, \cdot))$. Uma demonstração das relações de recorrência (2.4) encontra-se no apêndice. No caso geral, não existe uma solução fechada para o sistema (2.4). Na seção 2.3.1 apresentamos uma classe genérica de filtros de partículas que pode ser usado para resolver (2.4) numericamente.

2.1.3 Convergência de Medidas Aleatórias

Essencialmente, o resultado de qualquer algoritmo para resolver o problema da fitragem baseado no Método de Monte Carlo Sequencial é uma medida aleatória que aproxima \tilde{p}_t . De fato, para estabelecermos se o algoritmo é bom ou ruim, precisamos definir em que sentido uma medida, ou mais precisamente uma sequência de medidas aleatórias pode aproximar outra medida.

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e seja $(\mu^n)_{n=1}^\infty$ uma sequência de medidas aleatórias, $\mu^n : \Omega \rightarrow \mathcal{M}_F(\mathbb{R}^d)$ e $\mu \in \mathcal{M}_F(\mathbb{R}^d)$ uma medida finita determinística. Como veremos, no caso de aproximações obtidas usando o filtro de partículas, n representa o número de partículas usado no sistema de partículas para aproximação. Consideramos dois tipos de convergência:

1. $\lim_{n \rightarrow \infty} \mathbb{E}[|\mu^n f - \mu f|] = 0$, para toda $f \in \mathcal{C}_b(\mathbb{R}^d)$;
2. $\lim_{n \rightarrow \infty} \mu^n = \mu$, $P - q.t.p.$

Observações:

1. Observe que se existe uma função integrável $v : \Omega \rightarrow \mathbb{R}$ tal que $\mu^n \bar{1} \leq v$ para todo n , então (2) \Rightarrow (1). De fato, se existe tal função $v : \Omega \rightarrow \mathbb{R}$ com a propriedade acima, então para toda $f \in \mathcal{C}_b(\mathbb{R}^d)$,

$$|\mu^n(w)f| \leq \mu^n(w)|f| \leq c\mu^n(w)\bar{1} \leq cv,$$

sendo que c é uma constante tal que $|f| \leq c\bar{1}$. Logo, como $\mu^n(w)f \rightarrow \mu(w)f$, $w - q.t.p.$, pelo Teorema da Convergência Dominada, temos que,

$$\lim_{n \rightarrow \infty} \int_{\Omega} |\mu^n(w)f - \mu(w)f| dP(w) = \int_{\Omega} \lim_{n \rightarrow \infty} |\mu^n(w)f - \mu(w)f| dP(w) = 0.$$

$$\therefore \mathbb{E}[|\mu^n f - \mu f|] \rightarrow 0, \quad \text{quando } n \rightarrow \infty, \quad \forall f \in \mathcal{C}_b(\mathbb{R}^d).$$

Note que a condição adicional é satisfeita se $(\mu^n)_{n=1}^\infty$ é uma sequência aleatória de medidas de probabilidade pois, neste caso, $\mu^n \bar{1} = 1$, para todo n .

2. Se $\lim_{n \rightarrow \infty} \mathbb{E}[|\mu^n f - \mu f|] = 0$, para toda $f \in \mathcal{C}_b(\mathbb{R}^d)$, então existe uma subsequência $n(m)$ tal que $\lim_{m \rightarrow \infty} \mu^{n(m)} = \mu$, $P - q.t.p.$ Desde que \mathbb{R}^d é um espaço métrico, localmente compacto e separável então existe um conjunto enumerável \mathcal{M} que é denso em $\mathcal{C}_k(\mathbb{R}^d)$ (isto é, $\mathcal{C}_k(\mathbb{R}^d)$ é separável). Então,

$$\text{se } v_n f \rightarrow v f, \quad \forall f \in \mathcal{M} \Rightarrow v_n f \rightarrow v f, \quad \forall f \in \mathcal{C}_k(\mathbb{R}^d).$$

Portanto,

$$v_n f \rightarrow v f, \quad \forall f \in \mathcal{C}_b(\mathbb{R}^d).$$

Logo, $\mathcal{M} \cup \bar{\mathcal{I}}$ é um conjunto determinante da convergência, isto é, se $v_n, n = 1, 2, \dots$, e v são medidas finitas e $\lim_{n \rightarrow \infty} v_n f = v f$ para toda $f \in \mathcal{M} \cup \bar{\mathcal{I}}$, então $\lim_{n \rightarrow \infty} v_n = v$. Seja $A := \mathcal{M} \cup \bar{\mathcal{I}}$. Desde que,

$$\lim_{n \rightarrow \infty} \mathbb{E}[|\mu^n f - \mu f|] = 0, \quad \forall f \in A$$

e \mathcal{M} é enumerável, podemos obter uma subsequência $n(m)$, com probabilidade 1, tal que

$$\lim_{m \rightarrow \infty} \mu^{n(m)} f = \mu f, \quad \forall f \in A.$$

Para verificarmos este resultado, consideremos uma enumeração de A , isto é, $A = \{f_j\}_{j=1}^{\infty}$. Como,

$$\mu^n f_j \xrightarrow{L_1} \mu f_j, \quad \forall f_j \in A, \quad (2.5)$$

existe uma subsequência $\mu^{n_k^1}$ tal que,

$$\mu^{n_k^1} f_1 \rightarrow \mu f_1, \quad P - q.t.p.$$

Assim, sendo $\mu^{n_k^1}$ uma subsequência de μ^n e em vista de (2.5),

$$\mu^{n_k^1} f_2 \xrightarrow{L_1} \mu f_2$$

e portanto, existe uma subsequência n_k^2 de n_k^1 tal que

$$\mu^{n_k^2} f_2 \rightarrow \mu f_2, \quad P - q.t.p.$$

e

$$\mu^{n_k^2} f_1 \rightarrow \mu f_1, \quad P - q.t.p.$$

Seguindo analogamente, concluímos que para todo $m \in \mathbb{N}$ fixo, existe uma subsequência $\mu^{n_k^m}$ tal que

$$\mu^{n_k^m} f_m \rightarrow \mu f_m, \quad P - q.t.p.$$

Assim, definindo $n(k) := n_k^k$, segue que

$$\mu^{n(k)} f \rightarrow \mu f, \quad P - q.t.p., \quad \forall f \in A,$$

como queríamos mostrar.

Se $\mathcal{M} \in \mathcal{C}_k(\mathbb{R}^d)$ é o conjunto definido como acima, então

$$d_{\mathcal{M}} := |\mu\bar{1} - v\bar{1}| + \sum_{f_k \in \mathcal{M}} \frac{|\mu f_k - v f_k|}{2^k \|f_k\|},$$

é a distância em $\mathcal{M}_F(\mathbb{R}^d)$ (ou $\mathcal{P}(\mathbb{R}^d)$), que gera a topologia fraca

$$\lim_{n \rightarrow \infty} v_n = v \iff \lim_{n \rightarrow \infty} d_{\mathcal{M}}(v_n, v) = 0. \quad (2.6)$$

Note que:

1. Usando $d_{\mathcal{M}}$, a convergência quase certa do tipo 2 é equivalente a
2'. $\lim_{n \rightarrow \infty} d_{\mathcal{M}}(\mu^n, \mu) = 0$, $P - q.t.p.$

Sabemos que, se

$$\mu^n \rightarrow \mu, \quad P - q.t.p.,$$

existe um conjunto $\Omega' \subset \Omega$ tal que $P(\Omega') = 1$ e

$$\mu^n(w) \rightarrow \mu(w), \quad \forall w \in \Omega'.$$

Assim, por (2.6)

$$\begin{aligned} \lim_{n \rightarrow \infty} \mu^n(w) = \mu(w), \quad \forall w \in \Omega' &\iff \lim_{n \rightarrow \infty} d_{\mathcal{M}}(\mu^n(w), \mu(w)) = 0, \quad \forall w \in \Omega'; P(\Omega') = 1 \\ &\iff \lim_{n \rightarrow \infty} d_{\mathcal{M}}(\mu^n, \mu) = 0, \quad P - q.t.p. \end{aligned} \quad (2.7)$$

2. Se existe uma variável aleatória integrável $u : \Omega \rightarrow \mathbb{R}$ tal que $\mu^n \bar{1} \leq u$ para todo n , então a convergência do tipo 1. implica

$$1'. \lim_{n \rightarrow \infty} \mathbb{E}[d_{\mathcal{M}}(\mu^n, \mu)] = 0.$$

Para o caso de medidas aleatórias, devemos considerar a condição extra que $\mu \bar{1} \leq u$. Vamos demonstrar a afirmação acima para o caso em que as medidas são não determinísticas. Primeiramente, observe que

$$\begin{aligned}
\frac{|\mu^n f_k - \mu f_k|}{2^k \|f_k\|} &\leq \frac{|\mu^n f_k| + |\mu f_k|}{2^k \|f_k\|} \\
&\leq \frac{\|f_k\| \mu^n \bar{1} + \|f_k\| \mu \bar{1}}{2^k \|f_k\|} \\
&\leq \frac{2u}{2^k} = \frac{u}{2^{k-1}}
\end{aligned} \tag{2.8}$$

Como, por hipótese u é uma variável aleatória integrável, temos que pelo Teorema da Convergência Dominada

$$\begin{aligned}
\lim_{n \rightarrow \infty} \mathbb{E}[d_{\mathcal{M}}(\mu^n, \mu)] &= \lim_{n \rightarrow \infty} \mathbb{E}|\mu^n \bar{1} - \mu \bar{1}| + \lim_{n \rightarrow \infty} \sum_{f_k \in \mathcal{M}} \frac{\mathbb{E}|\mu^n f_k - \mu f_k|}{2^k \|f_k\|} \\
&= \sum_{f_k \in \mathcal{M}} \frac{1}{2^k \|f_k\|} \lim_{n \rightarrow \infty} \mathbb{E}|\mu^n f_k - \mu f_k| \\
&= 0,
\end{aligned} \tag{2.9}$$

o que conclui a demonstração.

2.2 Teoremas de Convergência

2.2.1 O Caso da Observação Fixada

Primeiramente, consideremos o caso em que a observação do processo tem um valor arbitrário mas fixo $y_{0:T}$, onde T é um tempo finito arbitrariamente grande. Suponhamos que a fórmula de recorrência para $\tilde{p}_t^{y_{0:t}}$ se assegure para esta observação particular, para todo $0 \leq t \leq T$ (recorde-mos que a relação de recorrência (2.4) é válida $P_{Y_{0:t}} - q.t.p.$). Observe de (2.4) para obtermos a distribuição condicional do sinal dado o evento $\{Y_{0:t} = y_{0:t}\}$, ou seja, $\tilde{p}_t^{y_{0:t}}$, precisamos fazer uso de um passo intermediário, que consiste em obtermos a medida de probabilidade condicional predita $p_t^{y_{0:t-1}}$:

$$\tilde{p}_t^{y_{0:t-1}} \longrightarrow p_t^{y_{0:t-1}} \longrightarrow \tilde{p}_t^{y_{0:t}}$$

sendo que a distribuição inicial do sinal \tilde{p}_0 é conhecida.

Portanto é natural estudarmos algoritmos que forneçam aproximações recursivas para $\tilde{p}_t^{y_{0:t}}$ usando aproximações intermediárias para $p_t^{y_{0:t-1}}$. Sejam $(\tilde{p}_t^n)_{n=1}^\infty$ e $(p_t^n)_{n=1}^\infty$ sequências aproximando $\tilde{p}_t^{y_{0:t}}$ e $p_t^{y_{0:t-1}}$, respectivamente. Suponhamos que \tilde{p}_t^n e p_t^n são medidas aleatórias, não necessariamente probabilidades, tais que $p_t^n \neq 0$, $\tilde{p}_t^n \neq 0$ e $p_t^n g_t^{y_t} > 0$, $\forall n > 0$, $0 \leq t \leq T$. Seja também \bar{p}_t^n definida como uma medida de probabilidade (aleatória) absolutamente contínua com respeito a p_t^n , $t \in \mathbb{N}$ e $n \geq 1$ tal que

O seguinte teorema fornece uma condição necessária e suficiente para as seguintes convergências: p_t^n para $p_t^{y_{0:t-1}}$ e \tilde{p}_t^n para $\tilde{p}_t^{y_{0:t}}$. Para simplificar a notação, suprimimos, nos dois teoremas abaixo a dependência de $y_{0:t}$ e denotamos $\tilde{p}_t^{y_{0:t}}$ por $\tilde{p}_t, p_t^{y_{0:t-1}}$ por p_t , e $g_t^{y_t}$ por g_t , mas sempre tendo em mente que a observação do processo é uma dada trajetória fixa $y_{0:T}$.

Teorema 2.2.1. *As sequências p_t^n e \tilde{p}_t^n convergem para p_t , respectivamente, para \tilde{p}_t convergência tomada como sendo do tipo 1. se, e somente se, as três condições seguintes são satisfeitas:*

- a1. Para toda $f \in \mathcal{C}_b(\mathbb{R}^{n_x})$, $\lim_{n \rightarrow \infty} \mathbb{E}[|\tilde{p}_0^n f - \tilde{p}_0 f|] = 0$
- b1. Para toda $f \in \mathcal{C}_b(\mathbb{R}^{n_x})$, $\lim_{n \rightarrow \infty} \mathbb{E}[|p_t^n f - \tilde{p}_{t-1}^n K_{t-1} f|] = 0$
- c1. Para toda $f \in \mathcal{C}_b(\mathbb{R}^{n_x})$, $\lim_{n \rightarrow \infty} \mathbb{E}[|\tilde{p}_t^n f - \tilde{p}_t f|] = 0$.

Teorema 2.2.2. *As sequências p_t^n, \tilde{p}_t^n convergem quase certamente para p_t , respectivamente para \tilde{p}_t , isto é, com convergência do tipo 2. se, e somente se, as três condições seguintes são satisfeitas:*

- a2. $\lim_{n \rightarrow \infty} \tilde{p}_0^n = \tilde{p}_0$, $P - q.t.p.$
- b2. $\lim_{n \rightarrow \infty} d_{\mathcal{M}}(p_t^n, \tilde{p}_{t-1}^n K_{t-1}) = 0$, $P - q.t.p.$
- c2. $\lim_{n \rightarrow \infty} d_{\mathcal{M}}(\tilde{p}_t^n, \tilde{p}_t^n) = 0$, $P - q.t.p.$

Uma demonstração dos Teoremas 2.2.1 e 2.2.2 encontra-se no Capítulo 2 de [1].

2.2.2 O Caso da Observação Aleatória

Na seção anterior tanto as sequências de medidas quanto a medida limite dependem de um valor fixo da observação. Explicitamente, escrevemos que

$$\lim_{n \rightarrow \infty} p_t^{n, y_{0:t-1}} = p_t^{y_{0:t-1}}, \quad \lim_{n \rightarrow \infty} \tilde{p}_t^{n, y_{0:t}} = \tilde{p}_t^{y_{0:t}},$$

sendo que os limites acima são ou do tipo 1, ou do tipo 2. Desde que $p_t = p_t^{Y_{0:t-1}}$ e $\tilde{p}_t = \tilde{p}_t^{y_{0:t}}$, esperamos que

$$\lim_{n \rightarrow \infty} p_t^{n, Y_{0:t-1}} = p_t, \quad \lim_{n \rightarrow \infty} \tilde{p}_t^{n, Y_{0:t}} = \pi_t. \quad (2.11)$$

Consideremos primeiramente a convergência do tipo 1. Temos que,

$$\begin{aligned} \mathbb{E} \left[\left| p_t^{n, Y_{0:t-1}} f - p_t f \right| \right] &= \int_{(\mathbb{R}^{n_y})^t} \mathbb{E} \left[\left| p_t^{n, y_{0:t-1}} f - p_t^{y_{0:t-1}} f \right| \right] P_{Y_{0:t-1}}(dy_{0:t-1}) \\ \mathbb{E} \left[\left| \tilde{p}_t^{n, Y_{0:t}} f - \tilde{p}_t f \right| \right] &= \int_{(\mathbb{R}^{n_y})^{t+1}} \mathbb{E} \left[\left| \tilde{p}_t^{n, y_{0:t}} f - \tilde{p}_t^{y_{0:t}} f \right| \right] P_{Y_{0:t}}(dy_{0:t}) \end{aligned}$$

Portanto, se

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left| p_t^{n, y_{0:t-1}} f - p_t^{y_{0:t-1}} f \right| \right] = 0, \quad P_{Y_{0:t-1}} - q.t.y_{0:t-1}$$

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left| \tilde{p}_t^{n, y_{0:t}} f - \tilde{p}_t^{y_{0:t}} f \right| \right] = 0, \quad P_{Y_{0:t}} - q.t.y_{0:t},$$

e existem funções $v_f(y_{0:t-1})$, $w_f(y_{0:t})$, tais que, para todo $n \geq 0$,

$$\mathbb{E} \left[\left| p_t^{n, y_{0:t-1}} f - p_t^{y_{0:t-1}} f \right| \right] \leq v_f(y_{0:t-1}), P_{Y_{0:t-1}} - q.t.p. \quad (2.12)$$

$$\mathbb{E} [|\tilde{p}_t^{n, Y_0:t} f - \tilde{p}_t^{Y_0:t} f|] \leq w_f(y_{0:t}), P_{Y_0:t} - q.t.p. \quad (2.13)$$

então, pelo Teorema da Convergência Dominada,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E} \left[\left| p_t^{n, Y_0:t-1} f - p_t f \right| \right] &= \lim_{n \rightarrow \infty} \int_{(\mathbb{R}^{n_y})^t} \mathbb{E} [|\tilde{p}_t^{n, Y_0:t-1} f - \tilde{p}_t^{Y_0:t-1} f|] P_{Y_0:t-1}(dy_{0:t-1}) \\ &= \int_{(\mathbb{R}^{n_y})^t} \lim_{n \rightarrow \infty} \mathbb{E} [|\tilde{p}_t^{n, Y_0:t-1} f - \tilde{p}_t^{Y_0:t-1} f|] P_{Y_0:t-1}(dy_{0:t-1}) = 0; \end{aligned}$$

e analogamente,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left| \tilde{p}_t^{n, Y_0:t} f - \tilde{p}_t f \right| \right] = 0.$$

As condições (2.12) e (2.13) são satisfeitas para o caso em que consideramos medidas de probabilidade, e neste caso, $v_f = w_f = 2\|f\|$. De fato, basta observar que

$$\begin{aligned} \left| p_t^{n, Y_0:t-1} f - p_t f \right| &= \left| \int_{(\mathbb{R}^{n_y})^t} f(y_{0:t-1}) p_t^{n, Y_0:t-1}(dy_{0:t-1}) - \int_{(\mathbb{R}^{n_y})^t} f(y_{0:t-1}) p_t^{Y_0:t-1}(dy_{0:t-1}) \right| \\ &\leq \int_{(\mathbb{R}^{n_y})^t} |f(y_{0:t-1})| p_t^{n, Y_0:t-1}(dy_{0:t-1}) + \int_{(\mathbb{R}^{n_y})^t} |f(y_{0:t-1})| p_t^{Y_0:t-1}(dy_{0:t-1}) \\ &\leq \|f\| \int_{(\mathbb{R}^{n_y})^t} p_t^{n, Y_0:t-1}(dy_{0:t-1}) + \|f\| \int_{(\mathbb{R}^{n_y})^t} p_t^{Y_0:t-1}(dy_{0:t-1}) \\ &= 2\|f\| \end{aligned} \quad (2.14)$$

Os próximos dois teoremas encontram-se no Capítulo 2 de [1] e as demonstrações foram realizadas pela autora do trabalho.

Teorema 2.2.3. *Desde que para todo $t \geq 0$, existe uma constante $c_t > 0$ tal que $p_t g_t \geq c_t$, as seguintes seqüências $p_t^{n, Y_0:t-1}, \tilde{p}_t^{n, Y_0:t}$ convergem para p_t e \tilde{p}_t com convergência do tipo 1. se, e somente se, para toda $f \in \mathcal{C}_b(\mathbb{R}^{n_x})$,*

- a3. $\lim_{n \rightarrow \infty} \mathbb{E} [|\tilde{p}_0^n f - \tilde{p}_0 f|] = 0$
- b3. $\lim_{n \rightarrow \infty} \mathbb{E} [|\tilde{p}_t^{n, Y_0:t-1} f - \tilde{p}_{t-1}^{n, Y_0:t-1} K_{t-1} f|] = 0$
- c3. $\lim_{n \rightarrow \infty} \mathbb{E} [|\tilde{p}_t^{n, Y_0:t} f - \tilde{p}_t^{n, Y_0:t} f|] = 0.$

Demonstração. (\Leftarrow) A prova será baseada no princípio de indução matemática. Pela condição a3. o teorema é verdadeiro para $t = 0$. Precisamos provar que se as seqüências $p_{t-1}^{n, Y_0:t-2}, \tilde{p}_{t-1}^{n, Y_0:t-1}$ convergem para p_{t-1} , respectivamente, para \tilde{p}_{t-1} então $p_t^{n, Y_0:t-1}, \tilde{p}_t^{n, Y_0:t}$ convergem para p_t , respectivamente, para \tilde{p}_t . Sendo $p_t = \tilde{p}_{t-1} K_{t-1}$, temos que para toda $f \in \mathcal{C}_b(\mathbb{R}^{n_x})$,

$$\left| p_t^{n, Y_0:t-1} f - p_t f \right| \leq \left| p_t^{n, Y_0:t-1} f - \tilde{p}_{t-1}^{n, Y_0:t-1} K_{t-1} f \right| + \left| \tilde{p}_{t-1}^{n, Y_0:t-1} K_{t-1} f - \tilde{p}_{t-1} K_{t-1} f \right|. \quad (2.15)$$

Tomando a esperança em ambos os membros da equação (2.15) e posteriormente o limite quando $n \rightarrow \infty$, segue que

$$\lim_{n \rightarrow \infty} \mathbb{E}[|p_t^{n, Y_{0:t-1}} f - p_t f|] = 0.$$

De fato, observe que, $\mathbb{E}|p_t^{n, Y_{0:t-1}} f - \tilde{p}_{t-1}^{n, Y_{0:t-1}} K_{t-1} f| \rightarrow 0$ por b3. e $\mathbb{E}|\tilde{p}_{t-1}^{n, Y_{0:t-1}} K_{t-1} f - \tilde{p}_{t-1} K_{t-1} f| \rightarrow 0$ pela hipótese de indução quando $K_{t-1} f \in \mathcal{C}_b(\mathbb{R}^{n_x})$, $\forall f \in \mathcal{C}_b(\mathbb{R}^{n_x})$, que é a propriedade de Feller do núcleo transição.

Observe que,

$$\bar{p}_t^{n, Y_{0:t}} f := \int_{\mathbb{R}^{n_x}} f(y) d\bar{p}_t^{n, Y_{0:t}}(y) = \int_{\mathbb{R}^{n_x}} \frac{f(y) g_t^{y_t} d p_t^{n, Y_{0:t-1}}(y)}{p_t^{n, Y_{0:t-1}} g_t^{y_t}} = \frac{p_t^{n, Y_{0:t-1}}(f g_t^{y_t})}{p_t^{n, Y_{0:t-1}}(g_t^{y_t})}.$$

Por simplicidade, denotamos $g_t^{y_t}$ por g_t e, analogamente, segue que

$$\tilde{p}_t f = \frac{p_t f g_t}{p_t g_t}$$

Logo,

$$\begin{aligned} |\bar{p}_t^{n, Y_{0:t}} f - \tilde{p}_t f| &= \left| \frac{p_t^{n, Y_{0:t-1}} f g_t}{p_t^{n, Y_{0:t-1}} g_t} - \frac{p_t f g_t}{p_t g_t} \right| \\ &\leq \left| \frac{p_t^{n, Y_{0:t-1}} f g_t}{p_t^{n, Y_{0:t-1}} g_t} - \frac{p_t^{n, Y_{0:t-1}} f g_t}{p_t g_t} \right| + \left| \frac{p_t^{n, Y_{0:t-1}} f g_t}{p_t g_t} - \frac{p_t f g_t}{p_t g_t} \right| \\ &\leq \frac{\|f\|}{p_t g_t} |p_t^{n, Y_{0:t-1}} g_t - p_t g_t| + \frac{1}{p_t g_t} |p_t^{n, Y_{0:t-1}} f g_t - p_t f g_t| \end{aligned} \quad (2.16)$$

Na última desigualdade o primeiro termo foi obtido da seguinte forma,

$$\begin{aligned} \frac{p_t^{n, Y_{0:t-1}} f g_t}{p_t^{n, Y_{0:t-1}} g_t} - \frac{p_t^{n, Y_{0:t-1}} f g_t}{p_t g_t} &= \int \frac{f g_t}{p_t^{n, Y_{0:t-1}} g_t} d p_t^{n, Y_{0:t-1}} - \int \frac{f g_t}{p_t g_t} d p_t^{n, Y_{0:t-1}} \\ &= \int f g_t \frac{(p_t g_t - p_t^{n, Y_{0:t-1}} g_t)}{p_t^{n, Y_{0:t-1}} g_t p_t g_t} d p_t^{n, Y_{0:t-1}} \\ &= \frac{(p_t g_t - p_t^{n, Y_{0:t-1}} g_t)}{p_t^{n, Y_{0:t-1}} g_t p_t g_t} \int f g_t d p_t^{n, Y_{0:t-1}} \end{aligned}$$

Agora, como $g_t \geq 0$, pois g_t é a função densidade de W_t , segue que

$$\begin{aligned} \left| \frac{p_t^{n, Y_{0:t-1}} f g_t}{p_t^{n, Y_{0:t-1}} g_t} - \frac{p_t^{n, Y_{0:t-1}} f g_t}{p_t g_t} \right| &\leq \frac{|p_t g_t - p_t^{n, Y_{0:t-1}} g_t|}{p_t^{n, Y_{0:t-1}} g_t p_t g_t} \|f\| \int g_t d p_t^{n, Y_{0:t-1}} \\ &= \frac{|p_t g_t - p_t^{n, Y_{0:t-1}} g_t|}{p_t^{n, Y_{0:t-1}} g_t p_t g_t} \|f\| p_t^{n, Y_{0:t-1}} g_t \\ &= \frac{|p_t g_t - p_t^{n, Y_{0:t-1}} g_t|}{p_t g_t} \|f\| \end{aligned}$$

Portanto por (2.16) e lembrando que $g_t \in \mathcal{C}_b(\mathbb{R}^{n_x})$,

$$\mathbb{E}[|\tilde{p}_t^{n, Y_{0:t}} f - \tilde{p}_t f|] \leq \frac{\|f\|}{p_t g_t} \mathbb{E}[|p_t^{n, Y_{0:t-1}} g_t - p_t g_t|] + \frac{1}{p_t g_t} \mathbb{E}[|p_t^{n, Y_{0:t-1}} f g_t - p_t f g_t|] \rightarrow 0 \quad (2.17)$$

Finalmente, como $\mathbb{E}[|\tilde{p}_t^{n, Y_{0:t}} f - \bar{p}_t^{n, Y_{0:t}} f|] \rightarrow 0$ por hipótese e $\mathbb{E}[|\bar{p}_t^{n, Y_{0:t}} f - \tilde{p}_t f|] \rightarrow 0$ por (2.17), resulta que

$$\mathbb{E}[|\tilde{p}_t^{n, Y_{0:t}} f - \tilde{p}_t f|] \leq \mathbb{E}[|\tilde{p}_t^{n, Y_{0:t}} f - \bar{p}_t^{n, Y_{0:t}} \tilde{p}_t f|] \rightarrow 0 \quad (2.18)$$

Portanto, a parte da suficiência está provada.

(\Rightarrow) Suponhamos que para todo $t \geq 0$ e para toda $f \in \mathcal{C}_b(\mathbb{R}^{n_x})$,

$$\lim_{n \rightarrow \infty} \mathbb{E}[|p_t^{n, Y_{0:t-1}} f - p_t f|] = 0 \quad (2.19)$$

e

$$\lim_{n \rightarrow \infty} \mathbb{E}[|\tilde{p}_t^{n, Y_{0:t}} f - \tilde{p}_t f|] = 0 \quad (2.20)$$

A expressão (2.20), para o caso particular em que $t = 0$, fornece

$$\lim_{n \rightarrow \infty} \mathbb{E}[|\tilde{p}_0^n f - \tilde{p}_0 f|] = 0,$$

e portanto, a condição a3. se verifica. Agora, em vista de (2.19), temos que a expressão (2.17) é satisfeita, i.e.,

$$\lim_{n \rightarrow \infty} \mathbb{E}[|\tilde{p}_t^{n, Y_{0:t}} f - \tilde{p}_t f|] = 0. \quad (2.21)$$

Portanto, por (2.20) e (2.21),

$$\mathbb{E}[|\tilde{p}_t^{n, Y_{0:t}} f - \bar{p}_t^{n, Y_{0:t}} f|] \leq \mathbb{E}[|\tilde{p}_t^{n, Y_{0:t}} f - \tilde{p}_t f|] + \mathbb{E}[|\tilde{p}_t f - \bar{p}_t^{n, Y_{0:t}} f|] \rightarrow 0,$$

o que resulta na condição c3. Sendo $p_t = \pi_{t-1} K_{t-1}$, segue que, para toda $f \in \mathcal{C}_b(\mathbb{R}^{n_x})$,

$$\mathbb{E}[|p_t^{n, Y_{0:t-1}} f - \tilde{p}_{t-1}^{n, Y_{0:t-1}} K_{t-1} f|] \leq \mathbb{E}[|p_t^{n, Y_{0:t-1}} f - p_t f|] + \mathbb{E}[|\tilde{p}_{t-1} K_{t-1} f - \tilde{p}_{t-1}^{n, Y_{0:t-1}} K_{t-1} f|] \rightarrow 0.$$

Logo, a condição b3. é satisfeita. \square

Teorema 2.2.4. *As sequências $p_t^{n, Y_{0:t-1}}$, $\tilde{p}_t^{n, Y_{0:t}}$ convergem quase certamente para p_t , respectivamente para \tilde{p}_t , isto é, com convergência do tipo 2., para todo $t \geq 0$ se, e somente se, as três condições seguintes são satisfeitas:*

- a4. $\lim_{n \rightarrow \infty} \tilde{p}_0^n = \tilde{p}_0$, $P - q.t.p.$
- b4. $\lim_{n \rightarrow \infty} d_{\mathcal{M}}(p_t^{n, Y_{0:t-1}}, \tilde{p}_{t-1}^{n, Y_{0:t-1}} K_{t-1}) = 0$, $P - q.t.p.$
- c4. $\lim_{n \rightarrow \infty} d_{\mathcal{M}}(\tilde{p}_t^{n, Y_{0:t}}, \bar{p}_t^{n, Y_{0:t}}) = 0$, $P - q.t.p.$

Demonstração. (\Leftarrow) Suponha que as relações a4., b4. e c4 sejam satisfeitas. Vamos provar, por indução matemática, que

$$p_t^{n, Y_{0:t-1}} \rightarrow p_t, \quad P - q.t.p.$$

e

$$\tilde{p}_t^{n, Y_{0:t}} \rightarrow \tilde{p}_t, \quad P - q.t.p.$$

Para $t = 0$, o teorema é verdadeiro pela condição a4. Por hipótese de indução, para toda $f \in \mathcal{C}_b(\mathbb{R}^{n_x})$

$$|\tilde{p}_{t-1}^{n, Y_{0:t-1}} K_{t-1} f - \tilde{p}_{t-1} K_{t-1} f| \rightarrow 0, \quad P - q.t.p.$$

Assim, pela condição b4, para toda $f \in \mathcal{C}_b(\mathbb{R}^{n_x})$

$$|p_t^{n, Y_{0:t-1}} f - p_t f| \leq |p_t^{n, Y_{0:t-1}} f - \tilde{p}_{t-1}^{n, Y_{0:t-1}} K_{t-1} f| + |\tilde{p}_{t-1}^{n, Y_{0:t-1}} K_{t-1} f - \tilde{p}_{t-1} K_{t-1} f| \rightarrow 0, \quad P - q.t.p.$$

$$\therefore p_t^{n, Y_{0:t-1}} \rightarrow p_t, \quad P - q.t.p.$$

Agora, por (2.16)

$$|\bar{p}_t^{n, Y_{0:t}} f - \tilde{p}_t f| \rightarrow 0, \quad P - q.t.p.$$

e portanto, pela relação c4,

$$|\tilde{p}_t^{n, Y_{0:t}} f - \tilde{p}_t f| \leq |\tilde{p}_t^{n, Y_{0:t}} f - \bar{p}_t^{n, Y_{0:t}} f| + |\bar{p}_t^{n, Y_{0:t}} f - \tilde{p}_t f| \rightarrow 0, \quad P - q.t.p.$$

$$\therefore \tilde{p}_t^{n, Y_{0:t}} \rightarrow \tilde{p}_t, \quad P - q.t.p.$$

(\Rightarrow) Primeiramente, vamos provar que se $\lim_{n \rightarrow \infty} p_t^{n, Y_{0:t-1}} = p_t$, $P - q.t.p.$ então

$$\lim_{n \rightarrow \infty} \bar{\pi}_t^{n, Y_{0:t}} = \pi_t, \quad P - q.t.p.$$

Seja \mathcal{M} um conjunto determinando a convergência das funções em $\mathcal{C}_b(\mathbb{R}^{n_x})$, por exemplo, o conjunto usado para construir a distância $d_{\mathcal{M}}$. Sabemos que $\mathcal{M} \subset \mathcal{C}_k(\mathbb{R}^{n_x})$ é enumerável e denso, e portanto basta provar que, para toda $f \in \mathcal{M}$,

$$\lim_{n \rightarrow \infty} \bar{p}_t^{n, Y_{0:t}(w)} f = \tilde{p}_t f(w), \quad w - q.t.p. \quad (2.22)$$

pois, neste caso, (2.22) se verifica, em particular, para toda $f \in \mathcal{C}_b(\mathbb{R}^{n_x})$. Consideremos um subconjunto $\Omega' \subset \Omega$ tal que $P(\Omega') = 1$ e para todo $w \in \Omega'$

$$\lim_{n \rightarrow \infty} p_t^{n, Y_{0:t-1}(w)} g_t = p_t g_t(w)$$

e

$$\lim_{n \rightarrow \infty} p_t^{n, Y_{0:t-1}(w)} (g_t f) = p_t (g_t f)(w)$$

para toda $f \in \mathcal{M}$.

Portanto, para todo $w \in \Omega'$

$$\lim_{n \rightarrow \infty} \bar{p}_t^{n, Y_{0:t}} f(w) = \lim_{n \rightarrow \infty} \frac{p_t^{n, Y_{0:t-1}(w)} (g_t f)}{p_t^{n, Y_{0:t-1}(w)} g_t} = \frac{p_t (g_t f)}{p_t g_t}(w) = \tilde{p}_t f(w), \quad \forall f \in \mathcal{M}$$

o que implica que,

$$\lim_{n \rightarrow \infty} \bar{p}_t^{n, Y_{0:t}} = \tilde{p}_t, \quad P - q.t.p. \quad (2.23)$$

Agora, por hipótese para todo $t \geq 0$, $P - q.t.p.$, temos que

$$\lim_{n \rightarrow \infty} d_{\mathcal{M}}(p_t^{n, Y_{0:t-1}}, p_t) = 0 \quad (2.24)$$

$$\lim_{n \rightarrow \infty} d_{\mathcal{M}}(\tilde{p}_t^{n, Y_{0:t}}, \tilde{p}_t) = 0 \quad (2.25)$$

e lembrando que $p_t = \tilde{p}_{t-1} K_{t-1}$, segue que

$$\lim_{n \rightarrow \infty} d_{\mathcal{M}}(\tilde{p}_{t-1}^{n, Y_{0:t-1}} K_{t-1}, p_t) = 0 \quad (2.26)$$

Assim, em vista das expressões (2.23) – (2.26), tomando o limite quando n tende a infinito, obtemos que

$$d_{\mathcal{M}}(p_t^{n, Y_{0:t-1}}, \tilde{p}_{t-1}^{n, Y_{0:t-1}} K_{t-1}) \leq d_{\mathcal{M}}(p_t^{n, Y_{0:t-1}}, p_t) + d_{\mathcal{M}}(p_t, \tilde{p}_{t-1}^{n, Y_{0:t-1}} K_{t-1}) \rightarrow 0 \quad (2.27)$$

$$d_{\mathcal{M}}(\tilde{p}_t^{n, Y_{0:t}}, \bar{\tilde{p}}_t^{n, Y_{0:t}}) \leq d_{\mathcal{M}}(\tilde{p}_t^{n, Y_{0:t}}, \tilde{p}_t) + d_{\mathcal{M}}(\tilde{p}_t, \bar{\tilde{p}}_t^{n, Y_{0:t}}) \rightarrow 0 \quad (2.28)$$

Logo, por (2.27) e (2.28) as expressões $b4$ e $c4$ se verificam. Agora, como $\tilde{p}_t^{n, Y_{0:t}} \rightarrow \tilde{p}_t$, $P - q.t.p.$, $\forall t \geq 0$, tomando $t = 0$, obtemos a expressão $a4$. □

2.3 Exemplo de Filtro de Partículas

2.3.1 Descrição do Filtro de Partículas

O algoritmo apresentado abaixo envolve o uso de um sistema de n partículas que evoluem de acordo com um processo de Markov e, em tempos fixos, geram um certo número de descendentes. Posteriormente, descrevemos o mecanismo de ramificação multinomial e veremos que impondo algumas condições sob esse mecanismo de ramificação, a medida empírica associada ao sistema de partículas converge, quando n tende a infinito, para a distribuição condicional do sinal, dadas as observações.

Descrição do Filtro de Partículas Espaço-Estado

1. Inicialização

- \tilde{p}_0 : distribuição inicial.
- \tilde{p}_0^n : distribuição empírica associada a uma amostra de tamanho n de \tilde{p}_0 .
- Para $i = 1, \dots, n$, amostramos $x_0^{(i)} \sim \tilde{p}_0$ e definimos

$$\tilde{p}_0^n = \frac{1}{n} \sum_{i=1}^n \delta_{\{x_0^{(i)}\}}.$$

2. Iteração

Descrevemos como obter \tilde{p}_t^n a partir de \tilde{p}_{t-1}^n .

- $\tilde{p}_{t-1}^n = \frac{1}{n} \sum_{i=1}^n \delta_{\{x_{t-1}^{(i)}\}}$.
- Mova cada partícula usando o núcleo transição do sinal

$$x_{t-1}^{(i)} \longrightarrow \bar{x}_t^{(i)} \sim K_{t-1} \left(x_{t-1}^{(i)}, \cdot \right).$$

Cada partícula se move independentemente das demais.

- A distribuição empírica associada a nova nuvem de partículas, $\bar{x}_t^{(i)}$, $i = 1, \dots, n$, será

$$p_t^n = \frac{1}{n} \sum_{i=1}^n \delta_{\{\bar{x}_t^{(i)}\}}.$$

- Calcule os pesos,

$$w_t^{(i)} = \frac{ng_t \left(\bar{x}_t^{(i)} \right)}{\sum_{j=1}^n g_t \left(\bar{x}_t^{(j)} \right)}.$$

- Observe que

$$\bar{p}_t^n = \frac{1}{n} \sum_{i=1}^n w_t^{(i)} \delta_{\{\bar{x}_t^{(i)}\}}.$$

é exatamente a medida \bar{p}_t^n como definida em (2.10)

- Cada partícula $\bar{x}_t^{(i)}$, $i = 1, \dots, n$, é substituída por um número de descendentes - $\xi_t^{(i)}$ - alguns possivelmente nulos, tal que

$$\sum_{i=1}^n \xi_t^{(i)} = n$$

- Denotamos as novas posições por $x_t^{(i)}$, $i = 1, \dots, N$, e definimos

$$\tilde{p}_t^n = \frac{1}{n} \sum_{i=1}^n \delta_{\{x_t^{(i)}\}}$$

Observações:

1. Vamos mostrar que a medida \bar{p}_t^n definida no passo de iteração é exatamente a medida dada por (2.10).

Por (2.10), temos que

$$\bar{p}_t^n = wp_t^n, \tag{2.29}$$

sendo que

$$w := \frac{g_t^{y_t}}{p_t^n g_t^{y_t}}.$$

Agora, para $i = 1, \dots, n$,

$$\begin{aligned} \frac{g_t^{y_t}(\bar{x}_t^{(i)})}{p_t^n g_t^{y_t}} &= \frac{g_t^{y_t}(\bar{x}_t^{(i)})}{\mathbb{E} g_t^{y_t}} \\ &= \frac{g_t^{y_t}(\bar{x}_t^{(i)})}{\sum_{j=1}^n g_t^{y_t}(\bar{x}_t^{(j)}) p_t^n(\bar{x}_t^{(j)})} \\ &= \frac{g_t^{y_t}(\bar{x}_t^{(i)})}{\frac{1}{n} \sum_{j=1}^n g_t^{y_t}(\bar{x}_t^{(j)})} := w_t^{(i)} \end{aligned} \quad (2.30)$$

Assim, substituindo $w_t^{(i)}$ em (2.29), obtemos que

$$\begin{aligned} \bar{p}_t^n &= w_t^{(i)} p_t^n = w_t^{(i)} \frac{1}{n} \sum_{i=1}^n \delta_{\{\bar{x}_t^{(i)}\}} \\ &= \frac{1}{n} \sum_{i=1}^n w_t^{(i)} \delta_{\{\bar{x}_t^{(i)}\}} \end{aligned} \quad (2.31)$$

O que mostra que a medida empírica para seleção das partículas, \bar{p}_t^n , está em concordância com a medida teórica dada por (2.10).

2. Seja A_t^n a matriz de variâncias e covariâncias do vetor aleatório $\xi_t := \left(\xi_t^{(i)} \right)_{i=1}^n$, ou seja, $A_t^n := \mathbb{E}[(\xi_t - w_t)^T (\xi_t - w_t)]$, sendo que $w_t := \left(w_t^{(i)} \right)_{i=1}^n$ é o vetor de pesos ou médias. Veremos posteriormente, na subseção *Convergência do Algoritmo*, deste capítulo, que para obtermos a convergência do algoritmo, devemos supor que existe uma constante c_t , tal que

$$q^T A_t^n q \leq n c_t, \quad (2.32)$$

para todo $q \in \mathbb{R}^{n \times 1}$; $q = \left(q^{(i)} \right)_{i=1}^n$ e $|q^{(i)}| \leq 1$, $i = 1, \dots, n$.

Mecanismo de Ramificação Multinomial

Para obtermos o número de descendentes $\xi_t^{(i)}$, $i = 1, \dots, n$, associados aos candidatos $\bar{x}_t^{(i)}$, no instante t , faremos uso do mecanismo de ramificação multinomial. Escolhemos,

$$\xi_t \sim \text{Multinomial} \left(n, \frac{w_t^{(1)}}{n}, \dots, \frac{w_t^{(n)}}{n} \right).$$

Neste caso,

1. $\mathbb{E} \left(\xi_t^{(i)} \right) = w_t^{(i)}$;
2. $Var \left(\xi_t^{(i)} \right) = \mathbb{E} \left[\left(\xi_t^{(i)} - w_t^{(i)} \right)^2 \right] = w_t^{(i)} \left(1 - \frac{w_t^{(i)}}{n} \right)$;
3. $cov(\xi_t^{(i)}, \xi_t^{(j)}) = \mathbb{E} \left[\left(\xi_t^{(i)} - w_t^{(i)} \right) \left(\xi_t^{(j)} - w_t^{(j)} \right) \right] = -\frac{w_t^{(i)} w_t^{(j)}}{n}$.

Observe também que,

$$\begin{aligned}
q^T A_t^n q &= \sum_{i=1}^n \mathbb{E} \left(\xi_t^{(i)} - w_t^{(i)} \right)^2 (q^{(i)})^2 + 2 \sum_{1 \leq i < j \leq n} \mathbb{E} \left[\left(\xi_t^{(i)} - w_t^{(i)} \right) \left(\xi_t^{(j)} - w_t^{(j)} \right) \right] q^{(i)} q^{(j)} \\
&= \sum_{i=1}^n w_t^{(i)} \left(1 - \frac{w_t^{(i)}}{n} \right) (q^{(i)})^2 - 2 \sum_{1 \leq i < j \leq n} \frac{w_t^{(i)} w_t^{(j)}}{n} q^{(i)} q^{(j)} \\
&= \sum_{i=1}^n w_t^{(i)} (q^{(i)})^2 - \frac{1}{n} \sum_{i=1}^n \left(w_t^{(i)} q^{(i)} \right)^2 - \frac{2}{n} \sum_{1 \leq i < j \leq n} w_t^{(i)} q^{(i)} w_t^{(j)} q^{(j)} \\
&= \sum_{i=1}^n w_t^{(i)} (q^{(i)})^2 - \frac{1}{n} \left(\sum_{i=1}^n w_t^{(i)} q^{(i)} \right)^2.
\end{aligned} \tag{2.33}$$

Agora, como $|q^{(i)}| \leq 1, i = 1, \dots, n$, segue que

$$q^T A_t^n q \leq \sum_{i=1}^n w_t^{(i)} - \frac{1}{n} \left(\sum_{i=1}^n w_t^{(i)} q^{(i)} \right)^2 \leq \sum_{i=1}^n w_t^{(i)}$$

e desde que $\sum_{i=1}^n w_t^{(i)} = n$, considerando $c_t = 1$, a condição (2.32), é satisfeita.

2.3.2 Convergência do Algoritmo

Fixemos uma observação arbitrária $y_{0:T}$, onde T é um tempo finito. Provaremos que as medidas aleatórias, resultantes das classes dos algoritmos descritas acima, convergem para $\tilde{p}_t^{y_{0:t}}$ (respectivamente $p_t^{y_{0:t-1}}$) para todo $0 \leq t \leq T$. Para isto, introduzimos as seguintes σ -álgebras

$$\mathcal{F}_t = \sigma \left(\bar{x}_s^{(i)}, x_s^{(i)}, s \leq t, \quad i = 1, \dots, n \right)$$

$$\bar{\mathcal{F}}_t = \sigma \left(\bar{x}_s^{(i)}, x_s^{(i)}, s < t, \quad \bar{x}_t^{(i)}, \quad i = 1, \dots, n \right)$$

Os próximos dois seguintes teoremas podem ser encontrados no Capítulo 2 de [1] e as demonstrações foram obtidas pela autora do trabalho.

Teorema 2.3.1. *Sejam $(p_t^n)_{n=1}^\infty$ e $(\tilde{p}_t^n)_{n=1}^\infty$ sequências de medidas produzidas pela classe dos algoritmos descrito acima. Então, para todo $0 \leq t \leq T$, temos que*

$$\lim_{n \rightarrow \infty} \mathbb{E}[|p_t^n f - \tilde{p}_t^{y_{0:t-1}} f|] = 0$$

e

$$\lim_{n \rightarrow \infty} \mathbb{E}[|\tilde{p}_t^n f - \tilde{p}_t^{y_0:t} f|] = 0$$

Demonstração. Pelo Teorema 2.2.1, basta verificarmos as expressões a1, b1 e c1. Seja $f \in \mathcal{C}_b(\mathbb{R}^{n_x})$. Observe que,

$$\tilde{p}_0^n f = \frac{1}{n} \sum_{i=1}^n \delta_{\{x_0^{(i)}\}} f = \frac{1}{n} \sum_{i=1}^n f(x_0^{(i)}) \quad (2.34)$$

Logo, desde que

$$\begin{aligned} \mathbb{E} \tilde{p}_0^n f &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} f(x_0^{(i)}) \\ &= \frac{1}{n} \sum_{i=1}^n \int_{\Omega} f(x_0^{(i)}) dP \\ &= \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}^{n_x}} f(t) dP_{\{x_0^{(i)}\}}(t) \\ &= \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}^{n_x}} f(t) d\pi_0 \\ &= \int_{\mathbb{R}^{n_x}} f(t) d\pi_0 \\ &= \mathbb{E} \pi_0 f \end{aligned} \quad (2.35)$$

a condição a1 é satisfeita. Mostraremos agora que a condição b1 se verifica. Para toda $f \in \mathcal{C}_b(\mathbb{R}^{n_x})$ vale que,

$$\mathbb{E} \left[f(\bar{x}_t^{(i)}) | \mathcal{F}_{t-1} \right] = K_{t-1} f(x_{t-1}^{(i)}), \quad \text{para todo } i = 1, \dots, n.$$

Portanto,

$$\mathbb{E} [p_t^n f | \mathcal{F}_{t-1}] = \tilde{p}_{t-1}^n K_{t-1} f. \quad (2.36)$$

De fato, temos que

$$p_t^n f = \frac{1}{n} \sum_{i=1}^n f(\bar{x}_t^{(i)})$$

e tomando a esperança condicional em \mathcal{F}_{t-1} , segue que

$$\begin{aligned} \mathbb{E} [p_t^n f | \mathcal{F}_{t-1}] &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[f(\bar{x}_t^{(i)}) | \mathcal{F}_{t-1} \right] \\ &= \frac{1}{n} \sum_{i=1}^n K_{t-1} f(x_{t-1}^{(i)}) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n} \sum_{i=1}^n \delta_{\{\bar{x}_{t-1}^{(i)}\}} K_{t-1} f \\
&= \tilde{p}_{t-1}^n K_{t-1} f
\end{aligned} \tag{2.37}$$

Agora, note que

$$(p_t^n f)^2 = \frac{1}{n^2} \sum_{i=1}^n f^2(\bar{x}_t^{(i)}) + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} f(\bar{x}_t^{(i)}) f(\bar{x}_t^{(j)})$$

o que implica que,

$$\mathbb{E}[(p_t^n f)^2 | \mathcal{F}_{t-1}] = \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}[f^2(\bar{x}_t^{(i)}) | \mathcal{F}_{t-1}] + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \mathbb{E}[f(\bar{x}_t^{(i)}) f(\bar{x}_t^{(j)}) | \mathcal{F}_{t-1}] \tag{2.38}$$

Observe que o primeiro termo do lado esquerdo de (2.38) é tal que,

$$\begin{aligned}
\frac{1}{n^2} \sum_{i=1}^n \mathbb{E}[f^2(\bar{x}_t^{(i)}) | \mathcal{F}_{t-1}] &= \frac{1}{n} \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \delta_{\{\bar{x}_t^{(i)}\}} f^2 | \mathcal{F}_{t-1}\right] \\
&= \frac{1}{n} \mathbb{E}[p_t^n f^2 | \mathcal{F}_{t-1}] = \frac{1}{n} \tilde{p}_{t-1}^n K_{t-1} f^2
\end{aligned} \tag{2.39}$$

Temos também que,

$$\begin{aligned}
(\mathbb{E}[p_t^n f | \mathcal{F}_{t-1}])^2 &= \left(\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n f(\bar{x}_t^{(i)}) | \mathcal{F}_{t-1}\right] \right)^2 \\
&= \frac{1}{n^2} \left(\sum_{i=1}^n \mathbb{E}[f(\bar{x}_t^{(i)}) | \mathcal{F}_{t-1}] \right)^2 \\
&= \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}^2[f(\bar{x}_t^{(i)}) | \mathcal{F}_{t-1}] \\
&\quad + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \mathbb{E}[f(\bar{x}_t^{(i)}) | \mathcal{F}_{t-1}] \mathbb{E}[f(\bar{x}_t^{(j)}) | \mathcal{F}_{t-1}]
\end{aligned} \tag{2.40}$$

Em vista de (2.38), (2.39) e (2.40), e lembrando que as partículas se movem independentemente, resulta que

$$\begin{aligned}
\mathbb{E}[(p_t^n f - \tilde{p}_{t-1}^n K_{t-1} f)^2 | \mathcal{F}_{t-1}] &= \mathbb{E}[(p_t^n f - \mathbb{E}[p_t^n f | \mathcal{F}_{t-1}])^2 | \mathcal{F}_{t-1}] \\
&= \mathbb{E}[(p_t^n f)^2 | \mathcal{F}_{t-1}] - (\mathbb{E}[p_t^n f | \mathcal{F}_{t-1}])^2 \\
&= \frac{1}{n} \tilde{p}_{t-1}^n K_{t-1} f^2 - \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}^2[f(\bar{x}_t^{(i)}) | \mathcal{F}_{t-1}]
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n} \tilde{p}_{t-1}^n K_{t-1} f^2 - \frac{1}{n^2} \sum_{i=1}^n \left(K_{t-1} f(x_{t-1}^{(i)}) \right)^2 \\
&= \frac{1}{n} \tilde{p}_{t-1}^n (K_{t-1} f^2 - (K_{t-1} f)^2)
\end{aligned} \tag{2.41}$$

Note que,

$$\begin{aligned}
\frac{1}{n^2} \sum_{i=1}^n \left(K_{t-1} f(x_{t-1}^{(i)}) \right)^2 &= \frac{1}{n^2} \sum_{i=1}^n (\delta_{\{x_{t-1}^{(i)}\}} K_{t-1} f)^2 \\
&= \frac{1}{n^2} \sum_{i=1}^n \delta_{\{x_{t-1}^{(i)}\}} (K_{t-1} f)^2 \\
&= \frac{1}{n} \tilde{p}_{t-1}^n (K_{t-1} f)^2
\end{aligned}$$

Agora,

$$(K_{t-1} f^2)(x) = \int_{\mathbb{R}^{n_x}} f^2(y) K_{t-1}(x, dy) \leq \|f\|^2 \int_{\mathbb{R}^{n_x}} K_{t-1}(x, dy) = \|f\|^2 \tag{2.42}$$

e pela desigualdade de Jensen,

$$(K_{t-1} f)^2(x) = \left(\int_{\mathbb{R}^{n_x}} f(y) K_{t-1}(x, dy) \right)^2 \leq \|f\|^2 \int_{\mathbb{R}^{n_x}} K_{t-1}(x, dy) = \|f\|^2 \tag{2.43}$$

Desta forma, por (2.42) e (2.43), segue que

$$\begin{aligned}
\mathbb{E}[(p_t^n f - \tilde{p}_{t-1}^n K_{t-1} f)^2] &\leq \frac{1}{n} (|\tilde{p}_{t-1}^n (K_{t-1} f^2)| + |\tilde{p}_{t-1}^n (K_{t-1} f)^2|) \\
&\leq \frac{2\|f\|^2}{n}
\end{aligned} \tag{2.44}$$

Tomando o limite quando n tende a infinito em (2.44) obtemos que a condição b1 é satisfeita. Finalmente, vamos mostrar que a condição c1 é verificada. Desde que,

$$\tilde{p}_t^n = \frac{1}{n} \sum_{i=1}^n \xi_t^{(i)} \delta_{\{\bar{x}_t^{(i)}\}}$$

e

$$\tilde{p}_t^n f = \frac{1}{n} \sum_{i=1}^n \xi_t^{(i)} f(\bar{x}_t^{(i)}),$$

temos que,

$$\mathbb{E}[\tilde{p}_t^n f | \bar{\mathcal{F}}_t] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\xi_t^{(i)} f(\bar{x}_t^{(i)}) | \bar{\mathcal{F}}_t]$$

$$\begin{aligned}
&= \frac{1}{n} \sum_{i=1}^n f(\bar{x}_t^{(i)}) \mathbb{E}[\xi_t^{(i)} | \mathcal{F}_t] \\
&= \frac{1}{n} \sum_{i=1}^n w_t^{(i)} f(\bar{x}_t^{(i)}) \\
&= \bar{p}_t^n f.
\end{aligned} \tag{2.45}$$

Temos também que,

$$\mathbb{E}[(\tilde{p}_t^n f - \bar{p}_t^n f)^2] = \frac{1}{n^2} (q_t^n)^T A_t^n q_t^n, \tag{2.46}$$

onde q_t^n é o vetor com entradas $(q_t^n)^{(i)} = f(\bar{x}_t^{(i)})$. Uma demonstração de (2.46) encontra-se no apêndice. De (2.32) e (2.46), resulta que

$$\begin{aligned}
\mathbb{E}[(\tilde{p}_t^n f - \bar{p}_t^n f)^2] &= \frac{1}{n^2} \|q_t^n\|^2 \left(\frac{(q_t^n)^T}{\|q_t^n\|} A_t^n \frac{q_t^n}{\|q_t^n\|} \right) \\
&\leq \frac{1}{n^2} \|q_t^n\|^2 n c_t = \frac{c_t \|f\|^2}{n}
\end{aligned}$$

Portanto,

$$\mathbb{E}[(\tilde{p}_t^n f - \bar{p}_t^n f)^2] \leq \frac{c_t \|f\|^2}{n}$$

e assim a condição $c1$ é satisfeita. □

Teorema 2.3.2. *Sejam $(p_t^n)_{n=1}^\infty$ e $(\tilde{p}_t^n)_{n=1}^\infty$ sequências de medidas produzidas pelo algoritmo com mecanismo de ramificação multinomial. Então, para todo $0 \leq t \leq T$, vale que*

$$\lim_{n \rightarrow \infty} p_t^n = p_t^{y_0:t-1}, \quad \lim_{n \rightarrow \infty} \tilde{p}_t^n = \tilde{p}_t^{y_0:t} \quad P - q.t.p.$$

Demonstração. A prova será baseada na aplicação do Teorema 2.2.2. Para isto, devemos verificar as expressões $a2$, $b2$ e $c2$. Seja $\mathcal{M} \in \mathcal{C}_b(\mathbb{R}^d)$ um conjunto enumerável determinando a convergência das funções definidas nas seções anteriores. Desde que $x_0^{(i)}$, $i = 1, \dots, n$, são independentes e distribuídas de acordo com π_0 , pela Lei dos Grandes Números, segue que

$$\tilde{p}_0^n f = \frac{1}{n} \sum_{i=1}^n f(x_0^{(i)}) \rightarrow \mathbb{E}f(x_0^{(1)}) \tag{2.47}$$

sendo que

$$\begin{aligned}
\mathbb{E}f(x_0^{(1)}) &= \int_{\Omega} f(x_0^{(1)}) dP \\
&= \int_{\mathbb{R}^{n_x}} f(t) dP_{\{x_0^{(1)}\}}(t) \\
&= \int_{\mathbb{R}^{n_x}} f(t) d\pi_0(t) \\
&= \tilde{p}_0 f
\end{aligned} \tag{2.48}$$

Portanto, a condição a_2 é verificada. Vamos mostrar agora que a condição b_2 é satisfeita.

Usando a independência de $\bar{x}_t^{(1)}, \dots, \bar{x}_t^{(n)}$ dado \mathcal{F}_t e desde que

$$\mathbb{E} \left[f \left(\bar{x}_t^{(i)} \right) \mid \mathcal{F}_{t-1} \right] = K_{t-1} f \left(x_{t-1}^{(i)} \right), \quad \text{para todo } i = 1, \dots, n$$

temos que

$$\begin{aligned} \mathbb{E}[(p_t^n f - \tilde{p}_{t-1}^n K_{t-1} f)^4 \mid \mathcal{F}_{t-1}] &= \mathbb{E} \left[\left(\frac{1}{n} \sum_{i=1}^n \left(f \left(\bar{x}_t^{(i)} \right) - K_{t-1} f \left(x_{t-1}^{(i)} \right) \right) \right)^4 \mid \mathcal{F}_{t-1} \right] \\ &= \frac{1}{n^4} \sum_{i=1}^n \mathbb{E} \left[\left(f \left(\bar{x}_t^{(i)} \right) - K_{t-1} f \left(x_{t-1}^{(i)} \right) \right)^4 \mid \mathcal{F}_{t-1} \right] + \\ &\quad \frac{2}{n^4} \sum_{1 \leq i < j \leq n} \mathbb{E} \left[\left(f \left(\bar{x}_t^{(i)} \right) - K_{t-1} f \left(x_{t-1}^{(i)} \right) \right)^2 \left(f \left(\bar{x}_t^{(j)} \right) - K_{t-1} f \left(x_{t-1}^{(j)} \right) \right)^2 \mid \mathcal{F}_{t-1} \right] \end{aligned}$$

Note que,

$$\begin{aligned} \frac{1}{n^4} \sum_{i=1}^n \mathbb{E} \left[\left(f \left(\bar{x}_t^{(i)} \right) - K_{t-1} f \left(x_{t-1}^{(i)} \right) \right)^4 \right] &\leq \frac{1}{n^4} \sum_{i=1}^n \mathbb{E} \left[\left(\left| f \left(\bar{x}_t^{(i)} \right) \right| + \left| K_{t-1} f \left(x_{t-1}^{(i)} \right) \right| \right)^4 \right] \\ &\leq \frac{1}{n^4} \sum_{i=1}^n \mathbb{E} [2 \|f\|^4] \\ &= \frac{16 \|f\|^4}{n^3} \\ &\leq \frac{16 \|f\|^4}{n^2} \end{aligned} \tag{2.49}$$

Note também que,

$$\begin{aligned} \frac{2}{n^4} \sum_{1 \leq i < j \leq n} \mathbb{E} \left[\left(f \left(\bar{x}_t^{(i)} \right) - K_{t-1} f \left(x_{t-1}^{(i)} \right) \right)^2 \left(f \left(\bar{x}_t^{(j)} \right) - K_{t-1} f \left(x_{t-1}^{(j)} \right) \right)^2 \right] &\leq \\ \frac{2}{n^4} \sum_{1 \leq i < j \leq n} \mathbb{E} \left[\left(\left| f \left(\bar{x}_t^{(i)} \right) \right| + \left| K_{t-1} f \left(x_{t-1}^{(i)} \right) \right| \right)^2 \left(\left| f \left(\bar{x}_t^{(j)} \right) \right| + \left| K_{t-1} f \left(x_{t-1}^{(j)} \right) \right| \right)^2 \right] &\leq \\ \frac{2}{n^4} \sum_{1 \leq i < j \leq n} \mathbb{E} [(2 \|f\|)^4] &= \\ \frac{32 \|f\|^4 n(n-1)}{n^4} &\leq \\ \frac{16 \|f\|^4}{n^2} &\end{aligned} \tag{2.50}$$

Assim, por (2.49) e (2.50), temos que

$$\mathbb{E} \left[(p_t^n f - \tilde{p}_{t-1}^n K_{t-1} f)^4 \right] \leq \frac{32 \|f\|^4}{n^2}$$

Agora, pela desigualdade de Markov,

$$\begin{aligned} \sum_{n=1}^{\infty} P \left[(p_t^n f - \tilde{p}_{t-1}^n K_{t-1} f)^4 \geq \epsilon \right] &\leq \sum_{n=1}^{\infty} \frac{\mathbb{E} \left[(p_t^n f - \tilde{p}_{t-1}^n K_{t-1} f)^4 \right]}{\epsilon} \\ &\leq \frac{1}{\epsilon} \sum_{n=1}^{\infty} \frac{32 \|f\|^4}{n^2} < +\infty \end{aligned} \quad (2.51)$$

Logo, pelo Lema de Borel Cantelli, a expressão (2.51) implica que

$$\begin{aligned} P \left(\limsup \left[(p_t^n f - \tilde{p}_{t-1}^n K_{t-1} f)^4 \geq \epsilon \right] \right) &= 0 \\ \Rightarrow P \left(\liminf \left[(p_t^n f - \tilde{p}_{t-1}^n K_{t-1} f)^4 \geq \epsilon \right]^c \right) &= 1 \\ \Rightarrow P \left(\liminf \left[(p_t^n f - \tilde{p}_{t-1}^n K_{t-1} f)^4 < \epsilon \right] \right) &= 1 \end{aligned} \quad (2.52)$$

Fazendo $A_n := \left[(p_t^n f - \tilde{p}_{t-1}^n K_{t-1} f)^4 < \epsilon \right]$, pela definição do limite inferior, para todo $w \in \liminf A_n$, existe $k_0 \in \mathbb{N}$ tal que $\forall n \geq k_0$,

$$(p_t^n f(w) - \tilde{p}_{t-1}^n K_{t-1} f(w))^4 < \epsilon.$$

Donde,

$$\lim_{n \rightarrow \infty} \left[(p_t^n f(w) - \tilde{p}_{t-1}^n K_{t-1} f(w))^4 \right] = 0, \quad w - q.t.p.$$

e portanto, para toda $f \in \mathcal{M}$

$$\lim_{n \rightarrow \infty} |p_t^n f - \tilde{p}_{t-1}^n K_{t-1} f| = 0, \quad P - q.t.p.$$

o que mostra que a expressão b2 é satisfeita. A expressão b3 é obtida de maneira analoga ao caso da expressão b2. Similarmente, para toda $f \in \mathcal{M}$,

$$\mathbb{E} \left[(\tilde{p}_t^n f - \bar{p}_t^n f)^4 | \mathcal{F}_t \right] \leq \frac{32 \|f\|^4}{n^2}$$

o que implica, como acima, que

$$\lim_{n \rightarrow \infty} |\tilde{p}_t^n f - \bar{p}_t^n f| = 0, \quad P - q.t.p.$$

e portanto, a condição c2 é verificada. \square

Para o caso em que as observações do processo são aleatórias, argumentos similares aos usados nos Teoremas 2.2.1 e 2.2.2 podem ser utilizados na demonstração dos seguintes Corolários.

Corolário 2.3.3. Desde que para todo $t \geq 0$, existe uma constante $c_t > 0$ tal que $p_t g_t \geq c_t$, então

$$\lim_{n \rightarrow \infty} \mathbb{E}[|p_t^{n, Y_{0:t-1}} f - p_t f|] = 0 \quad \lim_{n \rightarrow \infty} \mathbb{E}[|\tilde{p}_t^{n, Y_{0:t}} f - \tilde{p}_t f|] = 0.$$

Corolário 2.3.4. Sejam $(p_t^n)_{n=1}^\infty$ e $(\tilde{p}_t^n)_{n=1}^\infty$ duas seqüências de medidas produzidas pelo algoritmo com mecanismo de ramificação multinomial descrito acima. Então, para todo $0 \leq t \leq T$, temos que

$$\lim_{n \rightarrow \infty} p_t^{n, Y_{0:t-1}} = p_t \quad \lim_{n \rightarrow \infty} \tilde{p}_t^{n, Y_{0:t}} = \tilde{p}_t, \quad P - q.t.p.$$

2.3.3 Um Exemplo Markoviano

Nesta seção apresentaremos uma aplicação ilustrativa da implementação do filtro de partículas. Para isto escolhemos como modelo para o sinal uma cadeia de markov com 3 estados, se X_n é esta cadeia escolhemos a matriz de transição, $P = (p_{i,j})$ para X_n sendo:

$$\begin{aligned} \pi_1 &= 0.5; \pi_2 = 0.2; \pi_3 = 0.3; \\ p_{1,1} &= 0.5; p_{1,2} = 0.5; p_{1,3} = 0; \\ p_{2,1} &= 0.4; p_{2,2} = 0.2; p_{2,3} = 0.4; \\ p_{3,1} &= 0.1; p_{3,2} = 0.4; p_{3,3} = 0.5 \end{aligned}$$

onde $p_{i,j}$ é a probabilidade de transição de i para j .

As observações do modelo são Y_n que é uma perturbação do sinal X_n :

$$Y_n = X_n + U_n$$

onde U_n é uma seqüência de variáveis aleatórias uniformes independentes. Nosso objetivo é estimar X_n a partir das observações Y_n .

A figura a seguir mostra os seguintes valores da evolução de 10 partículas até o instante $n=15$. Temos uma descrição do processo de ramificação obtido na implementação do algoritmo. Este algoritmo foi implementado pela autora do trabalho.

Evolução de 10 partículas para um tempo de parada igual a 15															
Num.de partículas	Numero de descendentes														
1	0	0	0	2	1	2	1	2	0	1	0	2	2	0	2
2	4	0	0	2	2	2	2	0	0	0	0	0	2	3	2
3	1	2	0	0	0	0	3	0	3	0	0	0	3	0	0
4	0	0	0	0	0	0	0	4	0	5	0	1	0	1	0
5	1	3	0	1	0	3	1	0	1	0	2	0	0	2	2
6	0	0	0	2	2	1	0	0	5	0	3	2	0	1	1
7	0	0	0	2	1	0	0	0	1	0	0	2	0	0	0
8	0	0	10	0	2	0	1	4	0	3	2	3	0	0	3
9	2	3	0	1	0	0	2	0	0	1	3	0	3	2	0
10	2	2	0	0	2	2	0	0	0	0	0	0	0	1	0

Figura 2.1: Na figura, foram consideradas 10 partículas com tempo de parada igual a 15

Na seguinte tabela temos um resumo dos resultados obtidos. Optamos por tomar a média dos valores da nuvem de partículas, \bar{X}_n , como um preditor do sinal.

Nuvem de part.(tempo)	\bar{X}_n	$\hat{\sigma}_{X_n}$	$X_n(\text{signal})$	CV	EQM
1	2,20	0,63	1	0,29	1,80
2	1,80	0,78	2	0,43	0,60
3	1,50	0,71	3	0,47	2,70
4	1,70	0,82	2	0,48	0,70
5	1,90	0,57	2	0,30	0,40
6	1,90	0,99	1	0,52	1,70
7	1,90	0,74	2	0,39	0,50
8	2,10	0,88	3	0,42	1,50
9	2,10	0,87	2	0,41	0,70
10	2,10	0,83	3	0,40	1,50
11	2,10	0,84	3	0,40	1,50
12	2,10	0,74	2	0,35	0,50
13	2,00	0,94	3	0,47	1,80
14	1,60	0,84	2	0,53	0,80
15	1,70	0,82	2	0,48	0,70

Tabela 2.1: Sinal filtrado.

A seguir temos o gráfico para a média, \bar{X}_n e X_n juntos. O filtro, \bar{X}_n , não acompanha perfeitamente o sinal em média. Devido ao fato de estarmos trabalhando apenas um exemplo ilustrativo optamos por usar um número pequeno de partículas.

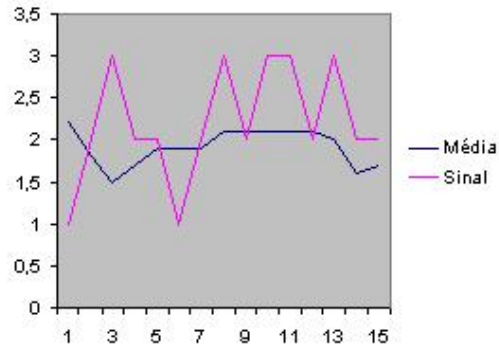


Figura 2.2: Gráfico para a Média e o Sinal

De acordo com o gráfico, a variância tende a estabilizar.

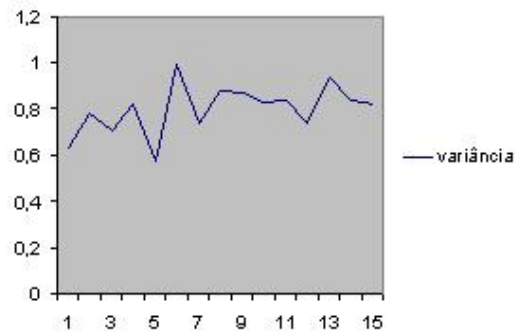


Figura 2.3: Gráfico para a variância

Capítulo 3

Reamostragem Móvel: Método de Filtragem com Saltos entre os Modelos

O algoritmo Resample-Move é um exemplo de uma classe de filtros de partículas chamado filtro de partículas híbrido. Embora esta técnica tenha as mesmas bases conceituais da amostragem-reamostragem por importância e amostragem MCMC, ela evita a degeneração dos métodos atuais (ver [3]).

Em [3] é proposto um novo método para análise bayesiana dinâmica, chamado o algoritmo Resample-Move, que evita a degeneração dos métodos existentes. O algoritmo Resample-Move combina SIR e iterações de cadeia de Markov. O objetivo é obter uma distribuição evoluindo por periódicos passos de reamostragem e por ocasionais movimentos de cadeias de Markov, que deixam cada partícula individual da posição corrente para uma nova posição aleatória do espaço de parâmetros. Ao contrário dos métodos MCMC padrão, o número de iterações da cadeia é arbitrário e nenhum tempo de burn-in é requerido [4].

Neste capítulo descrevemos o algoritmo Resample-Move e apresentamos sua implementação para o caso de um processo gaussiano linear. Como referência ver Capítulo 6 de [1].

3.1 Apresentação do Problema

Suponha que tenhamos uma sequência de observações $y_1, y_2, \dots, y_t, \dots$ avaliadas no tempo $1, 2, \dots, t, \dots$. No que segue, o subscrito t indexa o tempo discreto $t = 1, 2, \dots$. Supomos, a princípio, que as observações são geradas de um modelo paramétrico, posteriormente consideraremos situações mais gerais com modelos múltiplos. Seja a distribuição de probabilidade de $y_{1:t} := (y_1, \dots, y_t)$ sobre um tal modelo dada por,

$$p(y_{1:t}|\theta_t),$$

onde $\theta_t \in \Omega_t \subseteq \mathbb{R}^t$, representa um vetor de dados ausentes ou desconhecidos do modelo. Esses dados ausentes ou desconhecidos são chamados de *parâmetros*. Supomos que θ_t pode desenvolver-se aumentado de tamanho a todo tempo que uma nova observação surge. Considerando θ_t^+ o conjunto de parâmetros entrando no modelo no instante t , temos que

$$\theta_{t+1} = (\theta_t, \theta_t^+),$$

que podemos reescrever como

$$\theta_{t+1} = (\theta_{0:t+1}).$$

Dentro de uma aproximação Bayesiana para o problema, θ_t tem uma distribuição a priori

$$p(d\theta_t) = p(d\theta_{t-1})p(d\theta_{t-1}^+|\theta_{t-1})$$

Do ponto de vista das cadeias de Markov, consideramos $\theta_t^+ = x_{t+1}$ e $\theta_{t+1} = (x_{0:t}, x_{t+1})$, e neste caso

$$p(dx_{0:t}) = p(dx_{0:t-1})p(dx_t|x_{0:t-1}) = p(dx_{0:t-1})p(dx_t|x_{t-1})$$

Consideraremos, do ponto de vista Bayesiano, problemas em que inferências são requeridas on-line, como ocorre por exemplo em uma unidade de tratamento intensivo, onde estamos interessados a todo instante na detecção de anormalidades em pacientes baseado em um fluxo contínuo de dados gerados pelo monitoramento de aparelhos. Em problemas deste tipo, no tempo t , a distribuição de interesse, é a distribuição a posteriori Bayesiana,

$$\pi_t(d\theta_t) = p(d\theta_t|y_{1:t})$$

Esta expressão, no contexto Markoviano, é dada por

$$\pi_{0:t}(dx_{0:t}) = p(dx_{0:t}|y_{1:t})$$

representando nosso estado de incerteza do sinal após a obtenção da observação y_t . Esta distribuição é em geral conhecida a menos de uma constante de normalização dada pela integral

$$\int p(y_{1:t}|x_{0:t})p(dx_{0:t}),$$

que, em geral, é intratável. Devido ao aumento de informação dos dados, a distribuição alvo $p(\cdot|y_{1:t})$ desenvolverá geralmente com o tempo, gerando uma sequência suave

$$\pi_0(d\theta_0), \pi_1(d\theta_1), \dots, \pi_t(d\theta_t), \dots \quad (3.1)$$

que no contexto das cadeias de Markov é representada por

$$p(dx_0|y_{1:t}), p(dx_{0:1}|y_{1:t}), \dots, p(dx_{0:t}|y_{1:t}), \dots \quad (3.2)$$

onde $dx_{0:t} := (dx_0, \dots, dx_t)$. Temos a seguinte recursão Bayesiana,

$$\pi_t(d\theta_t) \propto \pi_{t-1}(d\theta_{t-1})p(d\theta_{t-1}^+|\theta_{t-1}, y_{1:t-1})p(y_t|y_{1:t-1}, \theta_t).$$

Esta relação, por simplicidade, será verificada para o caso Markoviano. Observe que seu análogo Markoviano é dado por,

$$p(dx_{0:t}|y_{1:t}) \propto p(dx_{0:t-1}|y_{1:t-1})p(dx_t|x_{0:t-1}, y_{1:t-1})p(y_t|y_{1:t-1}, x_{0:t})$$

De fato,

$$\begin{aligned}
p(dx_{0:t}|y_{1:t}) &= \frac{p(y_{1:t}|x_{0:t})p(dx_{0:t})}{\int p(y_{1:t}|x_{0:t})p(dx_{0:t})} \\
&\propto p(y_{1:t}|x_{0:t})p(dx_{0:t}) \\
&\propto p(y_{1:t}|x_{0:t})p(x_{0:t-1})p(dx_t|x_{0:t-1}) \\
&\propto p(y_{1:t-1}|x_{0:t-1})p(dx_{0:t-1})\frac{p(dx_t|x_{0:t-1})p(y_{1:t}|x_{0:t})}{p(y_{1:t-1}|x_{0:t-1})} \\
&\propto p(dx_{0:t-1}|y_{1:t-1})\frac{p(y_{1:t-1}, x_{0:t})p(dx_t|x_{0:t-1})p(y_t, y_{1:t-1}, x_{0:t})p(x_{0:t-1})}{p(x_{0:t-1}, y_{1:t-1})p(y_{1:t-1}, x_{0:t})p(x_{0:t})} \\
&\propto p(dx_{0:t-1}|y_{1:t-1})\frac{p(y_t|y_{1:t-1}, x_{0:t})p(y_{1:t-1}|x_t, x_{0:t-1})p(dx_t|x_{0:t-1})p(x_{0:t-1})}{p(x_{0:t-1}, y_{1:t-1})} \\
&\propto p(dx_{0:t-1}|y_{1:t-1})p(y_t|y_{1:t-1}, x_{0:t})\frac{p(x_t, x_{0:t-1}, y_{1:t-1})}{p(x_{0:t-1}, y_{1:t-1})} \\
&\propto p(dx_{0:t-1}|y_{1:t-1})p(y_t|y_{1:t-1}, x_{0:t})p(dx_t|x_{0:t-1}, y_{1:t-1})
\end{aligned}$$

Nosso objetivo, a grosso modo, está na evolução da sequência alvo. Mais formalmente, para qualquer função de interesse v_t dos parâmetros desconhecidos, definida sobre o suporte de $p(\cdot|y_{1:t})$. No contexto Bayesiano, queremos estimar a esperança condicional a posteriori para toda função $v_t : \Omega_t \rightarrow \mathbb{R}$ limitada e mensurável,

$$\mathbb{E}_{\pi_t} v_t(\theta_t) := \int v_t(\theta_t) \pi_t(d\theta_t) \quad (3.3)$$

onde $v_t(\theta_t)$ é o valor de v_t em um subconjunto menor $d\theta_t$ no tempo t , e a integral é calculada sobre tais subconjuntos. Note que de (3.3) decorre a unicidade em distribuição de π_t . Se Ω_t é discreto e $p(\cdot|y_{1:t})$ é a massa de probabilidade definida neste espaço, a integral (3.3) tornar-se

$$\mathbb{E}_{\pi_t} v_t(\theta_t) := \sum v_t(\theta_t) \pi_t(\theta_t) \quad (3.4)$$

No caso em que Ω_t é contínuo com densidade $\pi_t(\theta_t)$, então (3.3) se reduz a

$$\mathbb{E}_{\pi_t} v_t(\theta_t) := \int v_t(\theta_t) \pi_t(\theta_t) d\theta_t \quad (3.5)$$

3.2 O Algoritmo RESAMPLE-MOVE

Filtros de partículas para a trajetória da sequência alvo (3.1) produzem em cada instante de tempo inteiro t uma representação discreta de $p(\cdot|y_{1:t})$ através de uma coleção n_t de partículas aleatórias:

$$\Theta_t := (\theta_t^1, \theta_t^2, \dots, \theta_t^{n_t})$$

onde θ_t^j , a j -ésima partícula no instante de tempo t , é uma realização do vetor de parâmetros θ_t . O conjunto Θ_t de partículas é chamado a geração de ordem t . Consideremos filtros de partículas em que as gerações são atualizadas recursivamente, sendo a geração $(t - 1)$ atualizada na geração

t , em um tempo inteiro t . Os filtros são construídos de tal forma que a distribuição empírica das partículas na geração t converge para $p(\cdot|y_{1:t})$ quando o número de partículas contidas aumenta.

Isto significa que para qualquer função de interesse $v_t(\theta_t)$ a estimativa Monte Carlo

$$V_t(\theta_t) = \frac{1}{n_t} \sum_{j=1}^{n_t} v_t(\theta_t^j) \rightarrow \mathbb{E}_{\pi_t} v_t(\theta_t),$$

quando $n_t \rightarrow \infty$, sendo que $\mathbb{E}_{\pi_t} v_t(\theta_t)$ é a esperança a posteriori definida em (3.3). O análogo, para o caso das cadeias de Markov, é considerar a convergência

$$V_t(x_{0:t}) = \frac{1}{n_t} \sum_{j=1}^{n_t} v_t(x_{0:t}^{(j)}) \rightarrow \mathbb{E}_{\pi_{0:t}} v_t(x_{0:t}),$$

quando $n_t \rightarrow \infty$.

O algoritmo abaixo foi obtido do Capítulo 6 de [1].

Descrição do Filtro de Partículas Bayesiano

1. Inicialização, $t = 0$.

Crie a geração 0 amostrando, independentemente, para $j = 1, \dots, n_0$:

$$\theta_0^{(j)} \sim \pi_0(d\theta_0)$$

2. Inclusão e evolução

Para $t = 1, 2, \dots$, criamos a geração t realizando os dois seguintes passos:

2.1. Inclusão:

- para cada $i = 1, \dots, n_{t-1}$, amostre

$$\theta_{t-1}^{+, (i)} \sim f_t^{(i)}(d\theta_{t-1}^+)$$

com a f.d.p. $f_t^{(i)}$ definida abaixo.

- defina uma nova partícula

$$\theta_{t-1}^{*, (i)} = (\theta_{t-1}^{(i)}, \theta_{t-1}^{+, (i)}) \tag{3.6}$$

e seja o novo conjunto de partículas denotado por

$$\Theta_{t-1}^* := (\theta_{t-1}^{*, (1)}, \dots, \theta_{t-1}^{*, (n_{t-1})})$$

2.2. Evolução:

- as partículas da geração t são criadas por amostragem, independentemente, para $j = 1, \dots, n_t$:

$$\theta_t^{(j)} \sim Q_t(d\theta_t | \Theta_{t-1}^*), \tag{3.7}$$

sendo a distribuição condicional Q_t discutida posteriormente nesta seção.

De acordo com o algoritmo descrito acima, em um tempo inteiro t a geração anterior ($t - 1$) é atualizada em uma nova geração t por um par de passos: inclusão e evolução. A forma é requerida somente quando os novos parâmetros são introduzidos no modelo no instante t , causando assim que o espaço associado a θ_t tenha um maior tamanho do que o espaço associado a θ_{t-1} . Nessas situações, nos aumentamos cada i -ésima partícula da geração ($t - 1$) por uma componente extra apropriada, $\theta_{t-1}^{+, (i)}$, depois que essas são amostradas de uma distribuição $f_t^{(i)}$, como por exemplo, da distribuição a priori condicional

$$f_t^{(i)}(d\theta_{t-1}^{+, (i)}) = p(d\theta_{t-1}^{+, (i)} | \theta_{t-1}^{(i)}, y_{1:t-1}) \quad (3.8)$$

ou da distribuição a posteriori condicional,

$$f_t^{(i)}(d\theta_{t-1}^{+, (i)}) = p(d\theta_{t-1}^{+, (i)} | \theta_{t-1}^{(i)}, y_{1:t}) \quad (3.9)$$

Uma vez que cada partícula da geração ($t - 1$) tem sido melhorada pela inclusão de novas partículas, a geração ($t - 1$) é evoluída na geração t de acordo com (3.7).

O algoritmo resample-move, no contexto Bayesiano, é obtido como um caso especial do esquema descrito acima aplicando duas restrições. A *primeira*, supor que $\pi_{t-1} f_t^{(i)}$ é estritamente positiva para todo ponto no suporte de π_t , tal que depois de acrescentarmos a i -ésima partícula da geração ($t - 1$) podemos fixar seu peso como sendo

$$\begin{aligned} w_{t-1}^{(i)} &\propto \frac{\pi_t(d\theta_{t-1}^{*, (i)})}{\pi_{t-1}(d\theta_{t-1}^{(i)}) f_t^{(i)}(d\theta_{t-1}^{+, (i)})}, & \text{para } t > 1, \\ w_{t-1}^{(i)} &= 1, & \text{para } t = 1. \end{aligned}$$

Em particular, para o caso das cadeias de Markov, e em vista da relação para os pesos dada pela equação (2.10) do Capítulo 2, uma demonstração para este caso particular pode ser obtida.

No contexto Bayesiano, $w_{t-1}^{(i)}$ é dado de forma simplificada por $p(y_t | \theta_{t-1}^{*, (i)}, y_{1:t-1})$, quando $f_t^{(i)}$ é escolhido de acordo com a f.d.p. a priori condicional dada por (3.8). De fato, observe que

$$\begin{aligned} p(y_t | \theta_{t-1}^{*, (i)}, y_{1:t-1}) &= \frac{p(\theta_{t-1}^{*, (i)} | y_{1:t}) p(y_{1:t})}{p(\theta_{t-1}^{(i)}, \theta_{t-1}^{+, (i)}, y_{1:t-1})} \\ &= \frac{p(\theta_{t-1}^{*, (i)} | y_{1:t}) p(y_{1:t})}{p(\theta_{t-1}^{+, (i)} | \theta_{t-1}^{(i)}, y_{1:t-1}) p(\theta_{t-1}^{(i)} | y_{1:t-1}) p(y_{1:t-1})} \\ &\propto \frac{\pi_t(\theta_{t-1}^{*, (i)})}{\pi_{t-1}(\theta_{t-1}^{(i)}) f_t^{(i)}(\theta_{t-1}^{+, (i)})} \\ &\propto w_{t-1}^{(i)}. \end{aligned}$$

Para discutirmos a segunda restrição, introduzimos a notação $q_t(dz|x)$ para um núcleo transição no espaço Ω_t . Para todo x , o núcleo q_t especifica uma distribuição de probabilidade sobre Ω_t , e assim descreve uma maneira de mover de um dado ponto x de Ω_t a um novo ponto z do mesmo espaço. Suponhamos que q_t é invariante com respeito a π_t (ou seja, $\pi_t q_t = \pi_t$) tal que se o ponto inicial x é distribuído de acordo com π_t então o destino z também será distribuído de acordo com π_t . A *segunda* restrição ao Filtro de Partículas Bayesiano é considerar a distribuição Q_t em (3.7)

tendo uma forma de mistura (de densidades)

$$Q_t(d\theta_t|\Theta_{t-1}^*) := \sum_{i=1}^{n_{t-1}} \frac{w_{t-1}^{(i)}}{\sum_{i=1}^{n_{t-1}} w_{t-1}^{(i)}} q_t(d\theta_t|\theta_{t-1}^{*,(i)}) \quad (3.10)$$

De acordo com (3.10), devemos criar cada j -ésima partícula da geração t aleatoriamente selecionando uma partícula da geração $(t-1)$, com probabilidade $\frac{w_{t-1}^{(i)}}{\sum_{i=1}^{n_{t-1}} w_{t-1}^{(i)}}$, e então movendo a partícula selecionada de acordo com o núcleo q_t . A seleção será uma reamostragem por importância, com pesos e com reposição, tal que a probabilidade de uma dada partícula da geração $(t-1)$ ser selecionada deve ser proporcional a seu peso. Infelizmente, tal procedimento consome muito tempo computacional e em vista disto, várias propostas de esquemas alternativos de reamostragem tem sido propostos com atenção ao custo computacional do procedimento e seu impacto sobre a variância do estimador. Uma aproximação útil é implementar (3.10) por um procedimento consistindo de um passo de reamostragem seguido de um passo móvel. No passo de reamostragem, cada i -ésima partícula da geração $(t-1)$ se ramifica em um número m_i de cópias de si mesma, este número de cópias será proporcional a seu peso (ou seja, partículas com um maior peso dão origem a um maior número de descendentes). No passo móvel, cada uma das cópias geradas é movida de acordo com o núcleo q_t , condicionalmente independente de cada outra sobre Θ_{t-1}^* . Se nós permitimos que o número total de partículas no instante t , $n_t = \sum_{i=1}^{n_{t-1}} m_i$, seja aleatório, então podemos calcular m_i , $i = 1, \dots, n_{t-1}$, como a seguir:

1. Calcule o número de descendentes da i -ésima partícula no instante $(t-1)$, $\theta_{t-1}^{(i)}$, por

$$k_i = \frac{n_{t-1} w_{t-1}^{(i)}}{\sum_{i=1}^{n_{t-1}} w_{t-1}^{(i)}} \quad (3.11)$$

2. Considere m_i igual a parte inteira de $(k_i + 1)$ com probabilidade $frac(k_i)$, ou caso contrário, igual a parte inteira de k_i . Onde $frac(k)$ representa a parte fracionária de k .

Observações:

1. A expressão para o número de descendentes, k_i , é obtida como segue. Por hipótese da construção do esquema,

$$\begin{cases} k_i & \propto w_{t-1}^{(i)} \\ k_i & = \mathbb{E}(m_i) \end{cases}$$

Assim, $k_i = c w_{t-1}^{(i)}$, e $k_i = n_{t-1} p(m_i)$. Portanto,

$$c w_{t-1}^{(i)} = n_{t-1} p(m_i) \Rightarrow c \sum_{i=1}^{n_{t-1}} w_{t-1}^{(i)} = n_{t-1}$$

que fornece, $c = \frac{n_{t-1}}{\sum_{i=1}^{n_{t-1}} w_{t-1}^{(i)}}$. Desta forma, substituindo o valor da constante c , obtemos

$$\frac{n_{t-1}}{\sum_{i=1}^{n_{t-1}} w_{t-1}^{(i)}} w_{t-1}^{(i)} = n_{t-1} p(m_i) \Rightarrow p(m_i) = \frac{w_{t-1}^{(i)}}{\sum_{i=1}^{n_{t-1}} w_{t-1}^{(i)}}$$

$$\therefore k_i = n_{t-1} \frac{w_{t-1}^{(i)}}{\sum_{i=1}^{n_{t-1}} w_{t-1}^{(i)}}$$

2. Neste procedimento n_t tem valor esperado n_{t-1} . Basta observarmos que,

$$\mathbb{E}(n_t) = \sum_{i=1}^{n_{t-1}} \mathbb{E}(m_i) = \sum_{i=1}^{n_{t-1}} n_{t-1} p(m_i) = n_{t-1}.$$

3. Infelizmente, o esquema acima não assegura que o número de partículas seja constante, e não podemos portanto prever a extinção ou explosão do algoritmo. Isto pode ser evitado por um mecanismo de ramificação multinomial.

4. Na prática, é comum considerar, para todo t , $n_{t-1} = n_t = n$, sendo $n \in \mathbb{R}$ fixo, e de preferência não muito grande, por simplicidade computacional.

3.2.1 Um Exemplo Gaussiano

Vamos aplicar o algoritmo Reamostragem-Móvel a um problema simples envolvendo um único parâmetro desconhecido θ com valores em \mathbb{R} , sendo que a sequência alvo é uma sequência de distribuições Gaussianas univariadas com variância unitária e uma média que evolui com o tempo t de acordo com $0.5t$, isto é

$$\theta_t = 0.5t + N(0, 1).$$

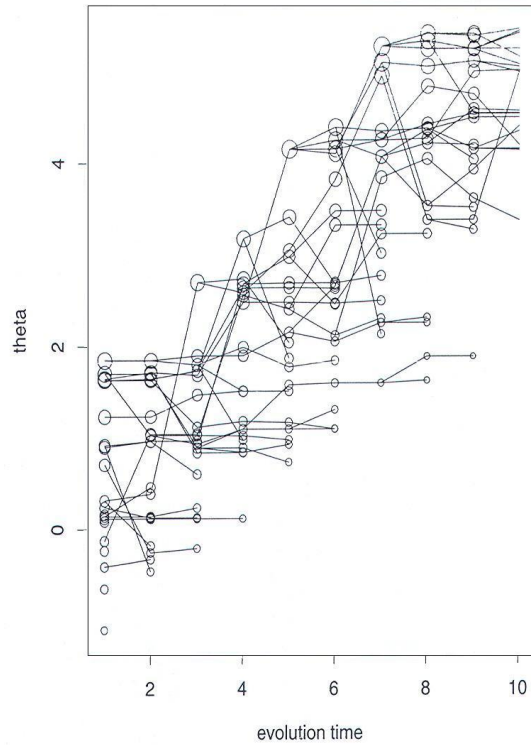


Figura 3.1: Representação gráfica do algoritmo RESAMPLE-MOVE em um exemplo univariado, esta figura foi retirada de [1]. Na figura, θ é um vetor de parâmetros desconhecidos com valores em \mathbb{R} representado pelo eixo vertical. Cada t -ésima geração de partículas fornece um conjunto de realizações de θ . Essas realizações são apresentadas na forma de um conjunto de círculos com abscissa t e raio proporcional ao peso da partícula.

Cada t -ésima geração de partículas é representada na figura como um conjunto de círculos com abscissa t e raio proporcional ao peso da partícula. Toda partícula importante da geração $(t - 1)$ tende a dar origem a uma numerosa família de partículas descendentes na geração t . Na figura, partículas de uma mesma família são ligadas a seus parentes comuns por uma reta, e seu número e localização são variáveis aleatórias que dependem do peso e localização de seus parentes comuns. O processo periodicamente redistribui as partículas tal que sua densidade tende a aumentar em regiões do espaço de parâmetros tendo uma maior probabilidade. O processo de seleção e movimento cooperam para obter uma adaptação enquanto ao mesmo tempo evitam a redução da diversidade das partículas. Partículas com pouca importância em cada geração tendem a permanecerem não selecionadas e morrem. O tamanho esperado de cada família é proporcional ao peso de seus parentes comuns.

- Implementação do algoritmo pela autora do trabalho.

Considere

$$\theta_t = 0,5t + N(0, 1).$$

Obtivemos os seguintes resultados que estão na tabela 3.1.

Tempo	$\bar{\theta}_t$	$\hat{\sigma}_{\theta_t}$	$\theta_t(sinal)$
1	1,29	1,18	0,72
2	0,64	0,14	0,28
3	1,37	0,28	1,84
4	0,94	0,66	2,01
5	2,08	0,65	1,71
6	3,53	1,13	2,72
7	5,76	1,62	5,51
8	5,10	0,25	3,97
9	2,23	0,09	2,23
10	1,92	0,24	3,97
11	7,30	0,84	6,12
12	5,95	1,30	6,16
13	6,81	0,28	6,46
14	8,53	2,41	8,89
15	8,62	0,48	9,01
16	8,31	0,61	8,21
17	7,54	1,03	8,49
18	9,14	0,88	10,01
19	11,75	0,25	12,74
20	8,69	1,09	7,78

Tabela 3.1: Sinal filtrado.

A aproximação é muito boa devido ao fato de que o sinal é linear. Nos esperamos uma flutuação maior no caso de um problema não linear.

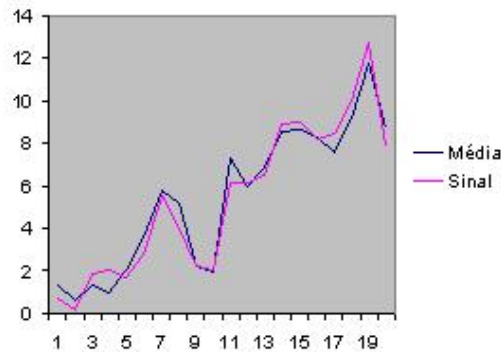


Figura 3.2: Gráfico para a Média e o Sinal

A distribuição a priori para o filtro deveria ter uma maior variância para que acompanhe melhor o sinal. Observe que o sinal é linear, o efeito de aumentar a variância não vai ser um problema, mas se o sinal for não linear o efeito de aumentar a variância não é bom. Isso ocorre porque os efeitos combinados do sinal e da nuvem de partículas vão fazer com que o filtro se afaste do sinal.

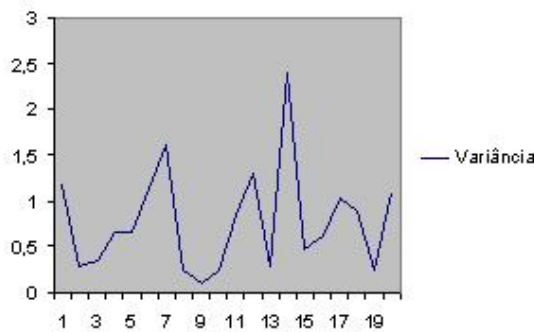


Figura 3.3: Gráfico para a variância

3.2.2 Comentários

1. Estamos admitindo, que o núcleo transição $q_t(.|x)$, que usamos para mover as partículas no tempo t , tem uma distribuição invariante π_t . Portanto, se uma partícula no tempo t antes de ser movida é distribuída de acordo com π_t , então ela será distribuída de acordo com π_t também depois de ser movida. Formalmente,

$$\pi_t(dz) = \int q_t(dz|x)\pi_t(dx) \quad (3.12)$$

Dois métodos conhecidos para a construção de núcleos de transição com a propriedade (3.12) são o Gibbs-Sampler (amostrador de Gibbs) e o método Metropolis-Hastings. Em particular, consideremos o método Metropolis-Hastings. Em um instante de tempo t , seja x a posição atual da partícula que desejamos mover. De acordo com o método Metropolis-Hastings devemos:

- escolher uma distribuição de referência,

$$m_t(\cdot|x)$$

- gerar uma nova posição candidata z' da distribuição de referência, isto é

$$z' \sim m_t(z|x)$$

sendo que \sim significa amostrado de.

- calcular

$$r = \min \left(1, \frac{\pi_t(z')m_t(x|z')}{\pi_t(x)m_t(z'|x)} \right)$$

- gere $u \in U(0, 1)$.
Se $u \leq r \Rightarrow$ faça $z = x$
Senão \Rightarrow faça $z = z'$

Se a distribuição de referência m_t é simétrica, $m_t(z|x) = m_t(x|z)$, como ocorre por exemplo no caso em que consideramos a distribuição de referência como sendo uma Normal, a probabilidade de aceitação α se reduz a

$$\alpha(x, z') = \min \left(1, \frac{\pi_t(dz')}{\pi_t(dx)} \right)$$

2. Em [3] é apresentado um teorema central do limite para V_t , que garante a consistência e normalidade assintótica quando o número de partículas cresce para infinito. Sobre condições de regularidade adequadas, o teorema afirma que

$$\frac{V_t - \mathbb{E}_{\pi_t} v(\theta_t)}{\sqrt{Var(V_t)}} \rightarrow N(0, 1)$$

quando $N_1, N_2, \dots, N_t \rightarrow \infty$, no sentido que $N_1 \rightarrow \infty, N_2 \rightarrow \infty$ e assim por diante. A variância do estimador, $Var(V_t)$, envolve integrais intratáveis e por [3] podem ser aproximadas pela história da partícula até o instante t .

Em alguns contextos estamos interessados na convergência da medida empírica das partículas em um tempo específico t , em outros podemos querer provar a convergência dos conjuntos de caminhos empíricos da distribuição a posteriori de toda a trajetória $(\theta_1, \theta_2, \dots, \theta_t)$.

Apêndice A

A Fórmula de Recorrência

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e $\mathcal{G} \in \mathcal{F}$ uma sub- σ -álgebra de \mathcal{F} . A esperança condicional de uma variável aleatória integrável, \mathcal{F} -mensurável, ξ dado \mathcal{G} é definida como uma variável aleatória integrável, \mathcal{G} -mensurável, denotada por $\mathbb{E}[\xi|\mathcal{G}]$, com a seguinte propriedade

$$\int_A \xi dP = \int_A \mathbb{E}[\xi|\mathcal{G}] dP,$$

para todo $A \in \mathcal{G}$. Então $\mathbb{E}[\xi|\mathcal{G}]$ existe e é única quase certamente. Isto é, se $\bar{\xi}$ é outra variável aleatória integrável, \mathcal{G} -mensurável, tal que $\int_A \xi dP = \int_A \bar{\xi} dP$, para todo $A \in \mathcal{G}$, então $\mathbb{E}[\xi|\mathcal{G}] = \bar{\xi}$, $P - q.t.p.$.

Algumas propriedades da esperança condicional

a. Se $\alpha_1, \alpha_2 \in \mathbb{R}$ e ξ_1, ξ_2 são \mathcal{F} -mensurável, então

$$\mathbb{E}[\alpha_1 \xi_1 + \alpha_2 \xi_2 | \mathcal{G}] = \alpha_1 \mathbb{E}[\xi_1 | \mathcal{G}] + \alpha_2 \mathbb{E}[\xi_2 | \mathcal{G}], \quad P - q.t.p.$$

b. Se $\xi \geq 0$, então $\mathbb{E}[\xi|\mathcal{G}] \geq 0$, $P - q.t.p.$.

c. Se $0 \leq \xi_n \nearrow \xi$, então $\mathbb{E}[\xi_n | \mathcal{G}] \nearrow \mathbb{E}[\xi | \mathcal{G}]$, $P - q.t.p.$.

d. Se \mathcal{H} é uma sub- σ -álgebra de \mathcal{G} , então $\mathbb{E}[\mathbb{E}[\xi|\mathcal{G}] | \mathcal{H}] = \mathbb{E}[\xi | \mathcal{H}]$, $P - q.t.p.$.

e. Se ξ é \mathcal{G} -mensurável, então $\mathbb{E}[\xi\eta | \mathcal{G}] = \xi \mathbb{E}[\eta | \mathcal{G}]$, $P - q.t.p.$.

f. Se \mathcal{H} é independente de $\sigma(\xi, \mathcal{G})$, então $\mathbb{E}[\xi | \sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[\xi | \mathcal{G}]$, $P - q.t.p.$.

A probabilidade condicional de um conjunto $A \in \mathcal{F}$ com respeito a σ -álgebra \mathcal{G} é uma variável aleatória denotada por $P(A|\mathcal{G})$, definida como $P(A|\mathcal{G}) := \mathbb{E}[I_A | \mathcal{G}]$, sendo que I_A é a função indicadora do conjunto A . Sejam $\eta_1, \eta_2, \dots, \eta_k$ variáveis aleatórias \mathcal{F} -mensurável. A esperança condicional de ξ com relação a η_1, \dots, η_k , $\mathbb{E}[\xi | \eta_1, \dots, \eta_k]$, é por definição a esperança condicional de ξ com respeito a σ -álgebra gerada por η_1, \dots, η_k , isto é, $\mathbb{E}[\xi | \eta_1, \dots, \eta_k] = \mathbb{E}[\xi | \sigma(\eta_1, \dots, \eta_k)]$. De maneira análoga, temos a definição de $P(A|\eta_1, \dots, \eta_k)$, a probabilidade condicional de A com relação a η_1, \dots, η_k . Em particular, para um evento $\{\eta = y\}$, $P(A|\eta = y)$ é uma variável aleatória $\mathcal{B}(\mathbb{R})$ -mensurável ($P(A|\eta = y)$ é por definição $\mathbb{E}[I_A | \eta = y]$) tal que

$$P(A \cap \{\eta \in B\}) = \int_B P(A|\eta = y) P_\eta(dy),$$

para todo $B \in \mathcal{B}(\mathbb{R})$.

A.1 A fórmula de recorrência

Primeiramente, vamos mostrar o seguinte lema

Lema A.1.1. *Seja $P_{Y_{s:t}} \in \mathcal{P}((\mathbb{R}^{n_y})^{t-s+1})$ a distribuição de probabilidade de $Y_{s:t}$ e seja λ a medida de Lebesgue sobre $((\mathbb{R}^{n_y})^{t-s+1}, \mathcal{B}((\mathbb{R}^{n_y})^{t-s+1}))$. Então, para todo $0 < s \leq t < \infty$, $P_{Y_{s:t}}$ é absolutamente contínua com respeito a λ e a derivada de Radon-Nikodym é*

$$\frac{dP_{Y_{s:t}}}{d\lambda} = \Upsilon(y_{s:t}) := \int_{(\mathbb{R}^{n_x})^{t-s+1}} \prod_{i=s}^t g_i(y_i - h(i, x_i)) P_{X_{s:t}}(dx_{s:t}).$$

Demonstração. Seja $C_{s:t} = C_s \times \dots \times C_t$, onde C_r são conjuntos borelianos arbitrários, $C_r \in \mathcal{B}(\mathbb{R}^{n_y}), \forall s \leq r \leq t$. Precisamos provar que

$$P_{Y_{s:t}}(C_{s:t}) = P(\{Y_{s:t} \in C_{s:t}\}) = \int_{C_{s:t}} \Upsilon(y_{s:t}) dy_s \dots dy_t$$

pois, como $P_{Y_{s:t}} \ll \lambda$, a derivada de Radon-Nikodym é a única função com a propriedade acima.

Temos que, pelo Teorema da Probabilidade Total,

$$\begin{aligned} P(\{Y_{s:t} \in C_{s:t}\}) &= P(\{Y_{s:t} \in C_{s:t}\} \cap \{X_{s:t} = x_{s:t}\}) \\ &= \int_{(\mathbb{R}^{n_x})^{t-s+1}} P(Y_{s:t} \in C_{s:t} | X_{s:t} = x_{s:t}) P_{X_{s:t}}(dx_{s:t}). \end{aligned} \quad (\text{A.1})$$

Como para todo i , X_i e W_i são independentes e também $W_s \dots W_t$ são independentes, obtemos que

$$\begin{aligned} P(Y_{s:t} \in C_{s:t} | X_{s:t} = x_{s:t}) &= \mathbb{E}(I_{C_1 \times \dots \times C_t}(Y_s, \dots, Y_t) | X_{s:t} = x_{s:t}) \\ &= \mathbb{E}\left(\prod_{i=s}^t I_{C_i}(h(i, X_i) + W_i) | X_{s:t}\right) \\ &= \mathbb{E}\left(\prod_{i=s}^t I_{C_i}(h(i, x_i) + W_i) | X_{s:t}\right) \\ &= \mathbb{E}\left(\prod_{i=s}^t I_{C_i}(h(i, x_i) + W_i)\right) \\ &= \prod_{i=s}^t \mathbb{E}(I_{C_i}(h(i, x_i) + W_i)) \\ &= \prod_{i=1}^t \int_{C_i} g_i(y_i - h(i, x_i)) dy_i. \end{aligned} \quad (\text{A.2})$$

Observe que na última passagem acima usamos o fato de que $W_t = Y_t - h(t, X_t)$ e que g_t é a função de densidade de W_t .

Combinando (A.1) e (A.2), e pelo teorema de Fubini, segue que

$$\begin{aligned} P(\{Y_{s:t} \in C_{s:t}\}) &= \int_{(\mathbb{R}^{n_x})^{t-s+1}} \prod_{i=s}^t \int_{C_i} g_i(y_i - h(i, x_i)) dy_i P_{X_{s:t}}(dx_{s:t}) \\ &= \int_{(\mathbb{R}^{n_x})^{t-s+1}} \int_{C_s \times \dots \times C_t} \prod_{i=s}^t g_i(y_i - h(i, x_i)) dy_s \dots dy_t P_{X_{s:t}}(dx_{s:t}) \\ &= \int_{C_{s:t}} \int_{(\mathbb{R}^{n_x})^{t-s+1}} \prod_{i=s}^t g_i(y_i - h(i, x_i)) P_{X_{s:t}}(dx_{s:t}) dy_s \dots dy_t. \end{aligned}$$

Logo, $\Upsilon(y_{s:t}) := \int_{(\mathbb{R}^{n_x})^{t-s+1}} \prod_{i=s}^t g_i(y_i - h(i, x_i)) P_{X_{s:t}}(dx_{s:t})$ é a derivada de Radon-Nikodym de $P_{Y_{s:t}}$ com relação a medida de Lebesgue λ . \square

Proposição A.1.2. *A densidade condicional do sinal satisfaz a seguinte relação de recorrência, para $t \geq 0$:*

$$\begin{cases} \frac{d\tilde{p}_t^{y_{0:t}}}{dp_t^{y_{0:t-1}}} = \frac{g_t^{y_t}}{\int_{\mathbb{R}^{n_x}} g_t^{y_t}(x_t) p_t^{y_{0:t-1}}(dx_t)} \\ p_{t+1}^{y_{0:t}} = \tilde{p}_t^{y_{0:t}} K_t \end{cases}$$

onde $g_t^{Y_t} := g(Y_t - h(t, \cdot))$ e a recorrência é satisfeita $P_{Y_{0:t}} - q.t.p.$, ou equivalentemente, $\lambda - q.t.p.$.

Demonstração. Primeiramente, vamos mostrar a segunda relação. Para toda $f \in B(\mathbb{R}^{n_x})$, observemos que, fixado x ,

$$K_t f(X_t = x) = \int_{\mathbb{R}^{n_x}} f(y) K_t(x, dy)$$

Logo, pela definição de esperança condicional,

$$K_t f(X_t) = \mathbb{E}[f(X_{t+1}) | X_t]$$

Usando a propriedade de Markov do processo X , obtemos que

$$\mathbb{E}[f(X_{t+1}) | \mathcal{F}_t^X] = \mathbb{E}[f(X_{t+1}) | X_t] = K_t f(X_t)$$

Agora, desde que $W_{0:t}$ é independente de $X_{0:t}$, temos que, pela propriedade f. da esperança condicional

$$\mathbb{E}[f(X_{t+1}) | \sigma(\mathcal{F}_t^X, \sigma(W_{0:t}))] = \mathbb{E}[f(X_{t+1}) | \mathcal{F}_t^X] = K_t f(X_t)$$

Portanto, pela propriedade d. da esperança condicional e a definição de \tilde{p}_t , segue que

$$\begin{aligned} p_{t+1} f &:= \mathbb{E}[f(X_{t+1}) | Y_{0:t}] \\ &= \mathbb{E}[\mathbb{E}[f(X_{t+1}) | \sigma(\mathcal{F}_t^X, \sigma(W_{0:t}))] | \sigma(Y_{0:t})] \\ &= \mathbb{E}[K_t f(X_t) | \sigma(Y_{0:t})] \\ &= \tilde{p}_t K_t f. \end{aligned}$$

O que implica que para observações do processo até o instante $t, y_{0:t}$, temos $p_{t+1}^{y_{0:t}} = \tilde{p}_t^{y_{0:t}} K_t$.

Provaremos agora a primeira relação. Seja $C_{0:t} = C_0 \times \dots \times C_t$, onde C_r são conjuntos borelianos arbitrários, $C_r \in \mathcal{B}(\mathbb{R}^{n_y})$, para todo $0 \leq r \leq t$. Precisamos provar que

$$\int_{C_{0:t}} \tilde{p}_t^{y_{0:t}} P_{Y_{0:t}}(dy_{0:t}) = \int_{C_{0:t}} \frac{\int_A g_t^{y_t}(x_t) p_t^{y_{0:t}}(dx_t)}{\int_{\mathbb{R}^{n_x}} g_t^{y_t}(x_t) p_t^{y_{0:t-1}}(dx_t)} P_{Y_{0:t}}(dy_{0:t}) \quad (\text{A.3})$$

De fato, observe que

$$(\text{A.3}) \Rightarrow \tilde{p}_t^{y_{0:t}}(A) = \int_A \frac{g_t^{y_t}(x_t) p_t^{y_{0:t}}(dx_t)}{\int_{\mathbb{R}^{n_x}} g_t^{y_t}(x_t) p_t^{y_{0:t-1}}(dx_t)}, \quad qtp$$

e usando a notação $d\lambda = p_t^{y_{0:t}}(dx_t) = P(X_t \in dx_t | Y_{0:t-1} = y_{0:t-1})$, temos que se $\lambda(A) = 0$ então $\tilde{p}_t^{y_{0:t}}(A) = 0$. Desta forma,

$$\tilde{p}_t^{y_{0:t}}(A) = 0 \ll \lambda \quad \text{e} \quad \frac{g_t^{y_t}(x_t)}{\int_{\mathbb{R}^{n_x}} g_t^{y_t}(x_t) p_t^{y_{0:t-1}}(dx_t)} := \frac{d\tilde{p}_t^{y_{0:t}}}{dp_t^{y_{0:t-1}}}$$

é a derivada de Radon-Nikodyn de $\tilde{p}_t^{y_{0:t}}$ com relação a medida $p_t^{y_{0:t-1}}$. Temos que,

$$\begin{aligned} P(\{X_t \in A\} \cap \{Y_{0:t} \in C_{0:t}\}) &= \int_{C_{0:t}} P(X_t \in A | Y_{0:t} = y_{0:t}) P_{Y_{0:t}}(dy_{0:t}) \\ &:= \int_{C_{0:t}} \tilde{p}_t^{y_{0:t}}(A) P_{Y_{0:t}}(dy_{0:t}) \end{aligned}$$

Assim precisamos provar que o segundo termo em (A.3) é igual a

$$P(\{X_t \in A\} \cap \{Y_{0:t} \in C_{0:t}\}).$$

Desde que $\sigma(X_{0:t}, W_{0:t-1}) \supset \sigma(X_t, Y_{0:t-1})$, pela propriedade d. da esperança condicional,

$$\mathbb{E}(I_{A_t}(Y_t) | X_t, Y_{0:t-1}) = \mathbb{E}(\mathbb{E}(I_{A_t}(Y_t) | X_{0:t}, W_{0:t-1}) | X_t, Y_{0:t-1})$$

Usando a propriedade f. da esperança condicional e a expressão (A.2) do Lema,

$$\begin{aligned} \mathbb{E}(I_{A_t}(Y_t) | X_{0:t}, W_{0:t}) &= \mathbb{E}(I_{A_t}(Y_t) | X_{0:t}) \\ &= P(Y_t \in A_t | X_{0:t}) = P(Y_{0:t} \in (\mathbb{R}^{n_y})^t \times A_t | X_{0:t}) \\ &= \prod_{i=0}^{t-1} \int_{(\mathbb{R}^{n_y})} g_i(y_i - h(i, X_i)) dy_i \int_{A_t} g_t(y_t - h(t, X_t)) dy_t \\ &= \int_{A_t} g_t(y_t - h(t, X_t)) dy_t := H(X_t). \end{aligned}$$

Assim, obtemos que

$$\begin{aligned} P(Y_t \in A_t | X_t, Y_{0:t-1}) &= \mathbb{E}(I_{A_t}(Y_t) | X_t, Y_{0:t-1}) \\ &= \mathbb{E}(\mathbb{E}(I_{A_t}(Y_t) | X_{0:t}, W_{0:t-1}) | X_t, Y_{0:t-1}) \\ &= \mathbb{E}(H(X_t) | X_t, Y_{0:t-1}) \\ &= \int_{A_t} g_t(y_t - h(t, X_t)) dy_t. \end{aligned}$$

$$\therefore P(Y_t \in A_t | X_t, Y_{0:t}) = \int_{A_t} g_t(y_t - h(t, X_t)) dy_t.$$

Assim, usando a notação, $g_t^{y_t} = g_t(y_t - h(t, \cdot))$, temos

$$P(Y_t \in A_t | X_t = x_t, Y_{0:t-1} = y_{0:t-1}) = \int_{A_t} g_t(y_t - h(t, x_t)) dy_t := \int_{A_t} g_t^{y_t}(x_t) dy_t. \quad (A.4)$$

Portanto,

$$\begin{aligned} P_{Y_{0:t}}(A_{0:t}) &:= P(\{Y_{0:t} \in A_{0:t}\}) = P(\{Y_{0:t} \in A_{0:t}\} \cap \{X_t \in \mathbb{R}^{n_x}\}) \\ &= P(\{Y_t \in A_t\} \cap \{X_t \in \mathbb{R}^{n_x}\} \cap \{Y_{0:t-1} \in A_{0:t-1}\}) \\ &= \int_{\mathbb{R}^{n_x} \times A_{0:t-1}} P(Y_t \in A_t | X_t = x_t, Y_{0:t-1} = y_{0:t-1}) P_{X_t, Y_{0:t-1}}(dx_t dy_{0:t-1}) \\ &= \int_{\mathbb{R}^{n_x} \times A_{0:t-1}} \int_{A_t} g_t^{y_t}(x_t) dy_t p_t^{y_{0:t-1}} P_{Y_{0:t-1}}(dy_{0:t-1}). \end{aligned} \quad (A.5)$$

Observe que,

$$P_{X_t, Y_{0:t-1}}(dx_t dy_{0:t-1}) = p_t^{y_{0:t-1}} P_{Y_{0:t-1}}(dy_{0:t-1}). \quad (A.6)$$

De fato, $\forall A \in \mathcal{B}(\mathbb{R}^{n_x})$,

$$\begin{aligned} P((X_t, Y_{0:t-1}) \in A \times C_{0:t-1}) &= \int_{C_{0:t-1}} P(X_t \in A | Y_{0:t-1} = y_{0:t-1}) P_{Y_{0:t-1}}(dy_{0:t-1}) \\ &= \int_{C_{0:t-1}} \int_A P(X_t \in dx | Y_{0:t-1} = y_{0:t-1}) P_{Y_{0:t-1}}(dy_{0:t-1}) \\ &= \int_{A \times C_{0:t-1}} p_t^{y_{0:t-1}}(dx_t) P_{Y_{0:t-1}}(dy_{0:t-1}) \end{aligned}$$

Logo,

$$P_{X_t, Y_{0:t-1}}(dx_t dy_{0:t-1}) := dP_{X_t, Y_{0:t-1}}(x_t, y_{0:t-1}) = p_t^{y_{0:t-1}}(dx_t) P_{Y_{0:t-1}}(dy_{0:t-1})$$

ou seja, $p_t^{y_{0:t-1}}$ é a derivada de Radon-Nikodym de $P_{X_t, Y_{0:t-1}}$ com relação a $P_{Y_{0:t-1}}$. De (A.5) segue que

$$P_{Y_{0:t}}(dy_{0:t}) = \int_{\mathbb{R}^{n_x}} g_t^{y_t} p_t^{y_{0:t-1}}(dx_t) dy_t P_{Y_{0:t-1}}(dy_{0:t-1}).$$

Portanto, o segundo termo em (A.3) é igual a

$$\begin{aligned} \Xi &:= \frac{\int_{C_{0:t}} \left(\int_A g_t^{y_t}(x_t) p_t^{y_{0:t-1}}(dx_t) \int_{\mathbb{R}^{n_x}} g_t^{y_t} p_t^{y_{0:t-1}}(dx_t) \right) dy_t P_{Y_{0:t-1}}(dy_{0:t-1})}{\int_{\mathbb{R}^{n_x}} g_t^{y_t}(x_t) p_t^{y_{0:t-1}}(dx_t)} \\ &= \int_{C_{0:t}} \int_A g_t^{y_t}(x_t) p_t^{y_{0:t-1}}(dx_t) dy_t P_{Y_{0:t-1}}(dy_{0:t-1}) \end{aligned}$$

Agora, por (A.4) e (A.6), obtemos que

$$\begin{aligned} \Xi &= \int_{A \times C_{0:t-1}} \left(\int_{C_t} g_t^{y_t}(x_t) dy_t \right) p_t^{y_{0:t-1}}(dx_t) P_{Y_{0:t-1}}(dy_{0:t-1}) \\ &= \int_{A \times C_{0:t-1}} P(Y_t \in C_t | X_t = x_t, Y_{0:t-1} = y_{0:t-1}) P_{X_t, Y_{0:t-1}}(dx_t dy_{0:t-1}) \\ &= P(\{X_t \in A\} \cap \{Y_{0:t} \in C_{0:t}\}). \end{aligned}$$

□

Vamos demonstrar agora, a relação,

$$\mathbb{E}[(\tilde{p}_t^n f - \bar{p}_t^n f)^2] = \frac{1}{n^2} (q_t^n)^T A_t^n q_t^n, \quad (\text{A.7})$$

onde q_t^n é o vetor com entradas $(q_t^n)^{(i)} = f(\bar{x}_t^{(i)})$, apresentada no Capítulo 2, Subseção Convergência do Algoritmo.

Temos que,

$$\begin{aligned} \mathbb{E}[(\tilde{p}_t^n f - \bar{p}_t^n f)^2 | \bar{\mathcal{F}}_t] &= \mathbb{E}[(\tilde{p}_t^n f - \mathbb{E}[\tilde{p}_t^n f | \bar{\mathcal{F}}_t])^2 | \bar{\mathcal{F}}_t] \\ &= \mathbb{E}[(\tilde{p}_t^n f)^2 | \bar{\mathcal{F}}_t] - \mathbb{E}[\tilde{p}_t^n f | \bar{\mathcal{F}}_t]^2 \end{aligned} \quad (\text{A.8})$$

O primeiro termo de (A.8) pode ser reescrito como,

$$\begin{aligned} \mathbb{E}[(\tilde{p}_t^n f)^2 | \bar{\mathcal{F}}_t] &= \mathbb{E} \left[\frac{1}{n^2} \left(\sum_{i=1}^n \xi_t^{(i)} f(\bar{x}_t^{(i)}) \right)^2 \middle| \bar{\mathcal{F}}_t \right] \\ &= \frac{1}{n^2} \mathbb{E} \left[\sum_{i=1}^n (\xi_t^{(i)})^2 f^2(\bar{x}_t^{(i)}) \middle| \bar{\mathcal{F}}_t \right] \\ &\quad + \frac{2}{n^2} \mathbb{E} \left[\sum_{1 \leq i < j \leq n} \xi_t^{(i)} f(\bar{x}_t^{(i)}) \xi_t^{(j)} f(\bar{x}_t^{(j)}) \middle| \bar{\mathcal{F}}_t \right] \\ &= \frac{1}{n^2} \sum_{i=1}^n f^2(\bar{x}_t^{(i)}) \mathbb{E} \left[(\xi_t^{(i)})^2 \middle| \bar{\mathcal{F}}_t \right] \\ &\quad + \frac{2}{n^2} \mathbb{E} \left[\sum_{1 \leq i < j \leq n} \xi_t^{(i)} f(\bar{x}_t^{(i)}) \xi_t^{(j)} f(\bar{x}_t^{(j)}) \middle| \bar{\mathcal{F}}_t \right] \end{aligned} \quad (\text{A.9})$$

O segundo termo de (A.8) pode ser reescrito como,

$$\begin{aligned} \mathbb{E}[\tilde{p}_t^n f | \bar{\mathcal{F}}_t]^2 &= \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \xi_t^{(i)} f(\bar{x}_t^{(i)}) \middle| \bar{\mathcal{F}}_t \right]^2 \\ &= \frac{1}{n^2} \left[\sum_{i=1}^n \mathbb{E}[\xi_t^{(i)} f(\bar{x}_t^{(i)}) | \bar{\mathcal{F}}_t] \right]^2 \\ &= \frac{1}{n^2} \left[\sum_{i=1}^n f(\bar{x}_t^{(i)}) \mathbb{E}[\xi_t^{(i)} | \bar{\mathcal{F}}_t] \right]^2 \\ &= \frac{1}{n^2} \left[\sum_{i=1}^n f(\bar{x}_t^{(i)}) w_t^{(i)} \right]^2 \\ &= \frac{1}{n^2} \left[\sum_{i=1}^n f^2(\bar{x}_t^{(i)}) (w_t^{(i)})^2 \right] \\ &\quad + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} f(\bar{x}_t^{(i)}) w_t^{(i)} f(\bar{x}_t^{(j)}) w_t^{(j)} \end{aligned} \quad (\text{A.10})$$

Definindo $q_t^{(i)} := f(\bar{x}_t^{(i)})$, por (A.9) e (A.10), temos que

$$\begin{aligned}
\mathbb{E}[(\tilde{p}_t^n f - \bar{p}_t^n f)^2 | \bar{\mathcal{F}}_t] &= \frac{1}{n^2} \sum_{i=1}^n f^2(\bar{x}_t^{(i)}) \left\{ \mathbb{E} \left[(\xi_t^{(i)})^2 | \bar{\mathcal{F}}_t \right] - (w_t^{(i)})^2 \right\} \\
&+ \frac{2}{n^2} \sum_{1 \leq i < j \leq n} f(\bar{x}_t^{(i)}) f(\bar{x}_t^{(j)}) \mathbb{E} \left[\xi_t^{(i)} \xi_t^{(j)} | \bar{\mathcal{F}}_t \right] - \frac{2}{n^2} \sum_{1 \leq i < j \leq n} f(\bar{x}_t^{(i)}) w_t^{(i)} f(\bar{x}_t^{(j)}) w_t^{(j)} \\
&= \frac{1}{n^2} \sum_{i=1}^n (q_t^{(i)})^2 \left\{ \mathbb{E} \left[(\xi_t^{(i)})^2 | \bar{\mathcal{F}}_t \right] - (w_t^{(i)})^2 \right\} \\
&+ \frac{2}{n^2} \sum_{1 \leq i < j \leq n} (q_t^{(i)}) (q_t^{(j)}) \mathbb{E} \left[\xi_t^{(i)} \xi_t^{(j)} | \bar{\mathcal{F}}_t \right] - \frac{2}{n^2} \sum_{1 \leq i < j \leq n} (q_t^{(i)}) (q_t^{(j)}) w_t^{(i)} w_t^{(j)} \\
&= \frac{1}{n^2} \sum_{i=1}^n (q_t^{(i)})^2 \mathbb{E} \left[(\xi_t^{(i)})^2 | \bar{\mathcal{F}}_t \right] - \frac{1}{n^2} \left(\sum_{i=1}^n q_t^{(i)} w_t^{(i)} \right)^2 \\
&+ \frac{2}{n^2} \sum_{1 \leq i < j \leq n} q_t^{(i)} q_t^{(j)} \mathbb{E} \left[\xi_t^{(i)} \xi_t^{(j)} | \bar{\mathcal{F}}_t \right] \tag{A.11}
\end{aligned}$$

Agora observe que,

$$\text{cov}(\xi_t^{(i)}, \xi_t^{(j)}) = -\frac{w_t^{(i)} w_t^{(j)}}{n} \Rightarrow \mathbb{E} \left[\xi_t^{(i)} \xi_t^{(j)} \right] = -\frac{w_t^{(i)} w_t^{(j)}}{n} + w_t^{(i)} w_t^{(j)} \tag{A.12}$$

Assim, tomando a esperança em ambos os membros de (A.11) e em vista de (A.12), obtemos que

$$\begin{aligned}
\mathbb{E}[(\tilde{p}_t^n f - \bar{p}_t^n f)^2] &= \frac{1}{n^2} \sum_{i=1}^n (q_t^{(i)})^2 \mathbb{E} \left[(\xi_t^{(i)})^2 \right] - \frac{1}{n^2} \left(\sum_{i=1}^n q_t^{(i)} w_t^{(i)} \right)^2 \\
&+ \frac{2}{n^2} \sum_{1 \leq i < j \leq n} q_t^{(i)} q_t^{(j)} \mathbb{E} \left[\xi_t^{(i)} \xi_t^{(j)} \right] \\
&= \frac{1}{n^2} \sum_{i=1}^n w_t^{(i)} (q_t^{(i)})^2 - \frac{1}{n^3} \sum_{i=1}^n (q_t^{(i)})^2 (w_t^{(i)})^2 + \frac{1}{n^2} \sum_{i=1}^n (q_t^{(i)})^2 (w_t^{(i)})^2 \\
&+ \frac{2}{n^2} \sum_{1 \leq i < j \leq n} q_t^{(i)} q_t^{(j)} w_t^{(i)} w_t^{(j)} - \frac{2}{n^3} \sum_{1 \leq i < j \leq n} q_t^{(i)} q_t^{(j)} w_t^{(i)} w_t^{(j)} - \frac{1}{n^2} \left(\sum_{i=1}^n q_t^{(i)} w_t^{(i)} \right)^2 \\
&= \frac{1}{n^2} \left(\sum_{i=1}^n w_t^{(i)} (q_t^{(i)})^2 - \frac{1}{n} \sum_{i=1}^n (q_t^{(i)})^2 (w_t^{(i)})^2 - \frac{2}{n} \sum_{1 \leq i < j \leq n} q_t^{(i)} q_t^{(j)} w_t^{(i)} w_t^{(j)} \right) \\
&= \frac{1}{n^2} \left(\sum_{i=1}^n w_t^{(i)} (q_t^{(i)})^2 - \frac{1}{n} \left(\sum_{i=1}^n q_t^{(i)} w_t^{(i)} \right)^2 \right) \\
&= \frac{1}{n^2} (q_t)^T A_t^n q_t. \tag{A.13}
\end{aligned}$$

Finalmente, para mantermos a coerência de notação com o Capítulo 2, consideramos $(q_t^n)^{(i)} := q_t^{(i)} := f(\bar{x}_t^{(i)})$, e portanto

$$\mathbb{E}[(\tilde{p}_t^n f - \bar{p}_t^n f)^2] = (q_t^n)^T A_t^n (q_t^n).$$

A.2 Algoritmo do capítulo 2

Exemplo de uma aplicação do filtro de partículas do capítulo 2. Linhas de comando em C.

```
#include <stdio.h>
#include <stdlib.h> /* for malloc & free */
#include "memory.h"

#include <math.h>
#include <iostream>
#include <R.h>
//#include <Rmath.h>

#define IA 16807
#define IM 2147483647
#define AM (1.0/IM)
#define IQ 127773
#define IR 2836
#define MASK 123459876

long sem=48123;
float ran1(long sem);

double pi[100], qi[100], p[100][100], q[1000][1000], U, V, Y[1000],
dens[1000][1000];
double aux, pmult[1000][1000], pmultacum[1000][1000];
int i, j, k, s, tam=3, parad, n=10, XXX[1000], X[1000][1000], m,
gauX[1000][1000], somag[1000];
int qsi[1000][1000], cont[1000];
// n:numero de part. , parad:tempo de parada
//XXX: sinal, X: nuvem

char quit;
using namespace std;

FILE *SAIDA;
main (){
    SAIDA=fopen("SAIDA.txt", "w");
    printf("Digite o tempo de parada:");
```

```

scanf("%d",&parad);

//valores iniciais
pi[1]=0.5; pi[2]=0.2, pi[3]=0.3;
p[1][1]=0.5; p[1][2]=0.5; p[1][3]=0;
p[2][1]=0.4; p[2][2]=0.2; p[2][3]=0.4;
p[3][1]=0.1; p[3][2]=0.4; p[3][3]=0.5;

//dist. inicial acum.
qi[0]=0; qi[1]=pi[1];
for (j=2; j<=tam; j++) qi[j]=qi[j-1]+pi[j];

//dist. acum. por linha para a matriz de transição
for (i=1; i<=tam; i++){
    j=1;
    q[i][j]=p[i][j];
    for (j=2; j<=tam; j++){
        q[i][j]=q[i][j-1]+p[i][j];
        //printf("q[%d][%d]=%f\n",i,j,q[i][j]);
    }
}
fprintf(SAIDA, "\n");
//for(j=1; j<=tam; j++) printf("qi[%d]=%f\n",j,qi[j]);
//for(j=1; j<=tam; j++) printf("q[%d][%d]=%f\n",j,tam);

//gerando a Cadeia de Markov

//valor inicial gerado
U=ran1(sem);
j=0;
do{
    j++;
}while (qi[j]<U);
XXX[0]=j;

//demais valores gerados ate o instante t=parad
for (i=1; i<=parad; i++){
    U=ran1(sem);
    j=0;
    do{
        j++;
    }
    while(q[XXX[i-1]][j]<U);
    XXX[i]=j;
}

```

```

        if(XXX[i]<=2) Y[i]=U;
        else Y[i]=1+U;
    }

    for(i=0; i<=parad; i++)
        fprintf(SAIDA, "X[%d]=%d , Y[%d]=%f\n",i,XXX[i],i,Y[i]);
// usar fprintf para ver a saida no arquivo SAIDA
    fprintf(SAIDA, "\n");

//criando a nuvem de particulas em t=0

    for (i=1;i<=n;i++){
        U=ran1(sem);
        //printf("%f\n",U);
        j=0;
        do{
            j++;
        }while (qi[j]<U);
        X[0][i]=j;
        dens[0][j]+=1/(float)n; //fornece a dist. empirica no tempo
zero

    }
    m=0;
    fprintf(SAIDA, "Nuvem de particulas em t=0:\n");
    for(i=1; i<=n; i++) fprintf(SAIDA, "X[%d] [%d]=%d\n",m,i,X[m][i]);

//printf("Distribuição empirica em t=0:\n");
//for(i=1;i<=3;i++) printf("pi [%d] [%d]=%f\n",0,i,dens[0][i]);

//Movendo as particulas com a matriz de transição
    for (i=1; i<=parad; i++){
        for (k=1; k<=n; k++){
            U=ran1(sem);
            j=0;
            do{
                j++;
            }while(q[X[i-1][k]][j]<U);
            X[i][k]=j;

            //calculando a soma no denominador dos pesos, w, em cada
inst i
            if(X[i][k]==3) aux=Y[i]-1;

```

```

else aux=Y[i];

if((aux>=0)&&(aux<=1)) gaux[i][k]=1;
else gaux[i][k]=0;
somag[i]+=gaux[i][k];
};//fim k

//Criando as probabilidades da multinomial em cada instante
de tempo i
for(s=1;s<=n;s++){
pmult[i][s]=(float)gaux[i][s]/(float)somag[i];
pmultacum[i][s]=pmult[i][s]+pmultacum[i][s-1];
}

//obtendo o número de descendentes em cada instante de tempo
i
cont[i]=0;
for(s=1;s<=n;s++){
    U=ran1(sem);
    j=0;
    do{
        j++;
    }while(pmultacum[i][j]<U);
    //qsi[i][s]=j;
    //cont[i]+=qsi[i][s];
    qsi[i][j]++;
}
for (s=1; s<=n; s++) cont[i]+=qsi[i][s];

};//fim i

fprintf(SAIDA, "Nuvem de particulas para t>=1:\n");
for (i=1; i<=parad; i++){
    for (k=1; k<=n; k++){
        fprintf(SAIDA, "X[%d] [%d]=%d\n",i,k,X[i][k]);
    }
    fprintf(SAIDA, "\n");
}

```



```

// printf("\n");
// for (i=1; i<=parad; i++) printf("somag[%d]=%d\n",i,somag[i]);

for (i=1; i<=parad; i++){
    for (s=1; s<=n; s++){
        // printf("pmultacum[%d] [%d]=%f\n",i,s,pmultacum[i][s]);
    }
// printf("\n");
}

for (i=1; i<=parad; i++){
    fprintf(SAIDA, "Numero de descendentes no instante %d\n",i);
    for(s=1;s<=n;s++){
        fprintf(SAIDA, "qsi[%d] [%d]=%d\n",i,s,qsi[i][s]);
    }
    fprintf(SAIDA, "Numero total de descendentes no instante
%d:%d\n",i,cont[i]);
    fprintf(SAIDA, "\n");
}

cin >> quit;

fclose(SAIDA);
system("PAUSE");
return 0;
} // fim main

float ran1(long idum0){// Press et al.,Numerical Recipes in C
long k;
float ans;
do{
    idum0 ^= MASK;
    k=(idum0)/IQ;
    idum0=IA*(idum0-k*IQ)-IR*k;
    if(idum0<0) idum0+=IM;
    ans=AM*(idum0);
    idum0 ^=MASK;
    sem=idum0;
}while(ans>0.999999);
return(ans);
}

```

A.3 Algoritmo do capítulo 3

Código em R utilizado na implementação do exemplo do capítulo 3.

```
n<- 10 ##numero de particulas
nsim <- 20 ## (Tempo)

## valores dos nsim  thetas (theta=sinal, observado ate o instante de tempo nsim)

## acrescento uma observacao no instante inicial (este valor nao interessa)

tempo<-nsim+1
theta <- (.5)*(1:tempo)+rnorm(1:tempo,0,1)-0.5

## nsim+1 y's (Observe que simula ate o instante nsim=20)

y<-theta + rnorm(tempo,0,1)

### matriz para guardar os valores do filtro

### a primeira linha contem os valores iniciais

th.mat <- rbind(rnorm(n,0,1),matrix(0,nsim,n))

### faz nsim iterações

for (i in 1:nsim){

th.mais <- th.mat[i,] + .5 + rnorm(n,0,2)

yi<- rep(y[i+1], each=n)

## yi é a observacao no instante i que sera comparada

## com o valor de cada particula

w <- dnorm(yi, th.mais , 1)
prob <- w/sum(w)
th1 <- rnorm(n, th.mais, 1)

raz <- exp(-(yi-th1)^2/2)/exp(-(yi-th.mais)^2/2)
```

```

uns <- rep(1, each=n)

alpha <- pmin(raz,uns)

## alpha <- sapply((raz,s), pmin), ou a correcao com "sapply"

u <- runif(n)
new1 <- th.mais

new1[u<alpha] <- th1[u<alpha]

p.ac <- cumsum(prob)

u2 <- runif(n)

## movendo as particulas

## guardando na matriz

th.mat[i+1,]<- new1[findInterval(u2, c(0,p.ac))]

}

options(width=1000)

round(th.mat,2)

#theta
th.mais

##Médias

filtro<-rowMeans(th.mat)

##Variâncias

var<-apply(th.mat[,], 1, var)

Tempo<- rep(1:tempo, 1)

plot(Tempo,filtro,col="red",type="l",main="Filtro de um sinal", sub="sinal em azul e filt

lines(y,col="blue")

```

```
plot(var, type="l")
```

Referências Bibliográficas

- [1] Doucet, A.; de Freitas, N.; Gordon, N., *Sequential Monte Carlo in Practice*, Springer-Verlag New York, 2001.
- [2] Gilks, W. R.; Berzuini, C., *RESAMPLE-MOVE Filtering with Cross-Model Jumps*, em *Sequential Monte Carlo in Practice*, Springer-Verlag New York, 2001, pg. 117-138.
- [3] Gilks, W. R.; Berzuini, C., *Following a moving target - Monte Carlo inference for dynamic Bayesian models*, J. R. Statist. B, 2001, pg. 127-146.
- [4] Rodeiro, C. L. V.; Lawson, A. B., *Online updating of space disease surveillance models via particle filters*, Statistical Methods in Medical Research, 2006, v. 15, pg. 423-444.
- [5] Liu, J. S., *Monte Carlo Strategies in Scientific Computing*, Springer-Verlag New York, 2001.
- [6] Welch, G.; Bishop, G., *An Introduction to the Kalman Filter*, Department of Computer Science, University of North Carolina at Chapel Hill, 2006.
- [7] Andrieu, C.; Doucet, A.; Punskeya, E., *Sequential Monte Carlo Methods for Optimal Filtering*, em *Sequential Monte Carlo in Practice*, Springer-Verlag New York, 2001, pg. 79-95.
- [8] Handschin, J.E.;Mayne, D.Q., *Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering*, International Journal of Control, 1969, pg. 547-559.
- [9] Kitagawa, G., *Non-Gaussian state-space modeling of nonstationary time series*, Journal of the American Statistical Association, 1987, v. 82(400), pg. 1032-1063.
- [10] Arulampalam, M.S;Maskell, S.; Gordon,N.; Clapp,T., *A Tutorial on Particle Filters for Online Nonlinear Non-Gaussian Bayesian Tracking*, IEEE Transactions on Signal Processing, 2002, v.50, n.2, pg. 174-188.
- [11] Carpenter, J.; Clifford,P.; Fearnhead, P.; *An improved particle filter for nonlinear problems*, IEE Proceedings-Radar Sonar e Navigation, 1999, v.146(1), pg. 2-7.
- [12] Crisan, D.; Del Moral, P.; Lyons, T.; *Discrete filtering using branching and interacting particle systems*, Markov Processes an Related Fieds, 1999, v. 5(3), pg. 293-318.
- [13] Gordon, N.J.; Salmond, D.J.; Smith, A.F.M.; *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*, IEE Proceedings-F, 1993, v. 140(2), pg. 107-113.