Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Matemática

Estabilidade do Filtro de Partículas e Estimação da Distribuição de Filtragem

por

Walter Batista dos Santos

Brasília
2010

Resumo

Filtros de Partículas constituem uma classe de algoritmos sequenciais que permitem, recursivamente em cada passo, atualizações das estimativas das características do processo oculto. A questão da estabilidade do filtro, independência da distribuição inicial, tem sido pouco abordada na literatura. Na maioria dos trabalhos assume-se que é uma propriedade intrínsica do filtro herdada da ergodicidade do processo oculto. Neste trabalho, estudamos a estabilidade do filtro analisando a ergodicidade fraca de cadeias de Markov associadas e convenientemente selecionadas. Como aplicação apresentamos o uso do Filtro de Bootstrap na solução do problema do controle de qualidade "on-line" num processo de produção.

Palavras Chaves: filtros de partículas; distribuição de filtragem; estabilidade do filtro; ergodicidade; cadeia não-homogênea.

Abstract

Particle Filters Algorithms constitute a class of sequential MCMC methods that produce and, recusively in time, update distributional estimates for the non-observable signal process. Filter stability issues, independence of initial distribution, have been either assumed or overlooked by most of the authors. In this work we address the stability question by studying the weak ergodicity of a properly associated non-homogeneous Markov chain. Applications includes the use of Bootstrap Filter to solve an on-line quality control process.

Key Words and Phrases: Particle Filters; filtering distribution; filter stability; ergodicity; non-homogeneous chain.

Aos meus pais, Pedro
e Aparecida, com todo
carinho e amor.

Que proveito tira o
trabalhador de sua
obra?
Ecle 3,9.

Agradecimentos

Primeiramente a Deus, pois tudo o que conquistei e o que ainda hei de conquistar foi e será, sem dúvida alguma, pelas mãos Dele.

À Professora Chang Dorea, pela orientação nestes quatro anos, em especial agradeço pela paciência, atenção e dedicação que demonstrou durante todo o doutorado. A Professora Chang é um exemplo de professor e pesquisador que terei como referência em minha vida profissional. Também, ao Professor Jorge Lucero que foi meu co-orientador.

À minha família, de maneira especial meus pais, pelas muitas orações e por passarem ao meu lado tudo o que precisei passar.

Aos professores, André Pereira, Ary Medino, Cátia Gonçalves, Cira Otiniano e Silvia Lopes que aceitaram participar da banca. Apesar de suas ocupações diárias em suas instituições leram a tese e contribuíram com correções e sugestões que valorizaram o trabalho.

Aos amigos, colegas, professores e funcionários do Departamento de Matemática da Universidade de Brasília. De uma forma ou de outra deram sua parte de contribuição na realização deste trabalho.

A todos os amigos e colegas que não são do "Mundo da Matemática". Estes também apresentaram suas contribuições nos momentos oportunos.

Ao CNPq pela bolsa durante todo o doutorado, uma vez que sem ela certamente não teria as vinte e quatro horas que tive para me dedicar aos estudos.

Sumário

Notação

iid - independente e identicamente distribuída

$Y_{1:n} = (Y_1, Y_2, \ldots, Y_n)$ e, analogamente, $y_{1:n} = (y_1, y_2, \ldots, y_n)$

HMM - Modelo de Markov Oculto

g - densidade genérica

$p_{n-1}^{y_{0:n}}(.|.)$ - densidade de transição do processo filtrado X^f (pág. 29)

$\tilde{p}_k^{y_{0:n}}(.|.)$ - densidade de transição da cadeia condicional X^c (pág. 18)

$\tilde{P}_k^{y_{0:n}}$ - núcleo de transição da cadeia condicional X^c (pág. 18)

$\pi_n^{y_{0:n}}(.)$ - densidade de $X_n|Y_{0:n} = y_{0:n}$, $n = 0, 1, \ldots$ (pág. 8)

$\pi_n^{y_{0:n}}(.)$ - distribuição de $X_n|Y_{0:n} = y_{0:n}$; $\pi_n^{y_{0:n}}(A) = \int_A \pi_n^{y_{0:n}}(x)d\mu(x)$ (pág. 7)

$\tilde{\pi}_k^{y_{0:n}}(.)$ - densidade de $X_k|Y_{0:n} = y_{0:n}$, $k = 0, 1, \ldots, n$ (pág. 18)

$q(.|x)$ - densidade de $Y_n|X_n = x$ (pág. 5)

$p(.|.)$ - densidade de transição da cadeia de Markov $X = \{X_n\}_{n\geq 0}$ (pág. 5)

$P_n(x, A) = P(X_{n+1} \in A|X_n = x)$ - notação para cadeias não-homogêneas (pág. 10)

$P_{m,m+n}(x, A) = P(X_{m+n} \in A|X_m = x)$ - notação para cadeias não-homogêneas (pág. 10)

$p(j|i) = P(j|i) = P(X_{n+1} = j|X_n = i)$

$P(x, A) = P(X_{n+1} \in A|X_n = x)$

$P^n(x, A) = P(X_{k+n} \in A|X_k = x)$

$\mathcal{X}, \mathcal{Y}, \ldots$ - espaço de estados

$\mathbb{X}, \mathbb{Y}, \ldots$ - σ-álgebra de subconjuntos de $\mathcal{X}, \mathcal{Y}, \ldots$

$\int_{\mathcal{X}^{n-k+1}} f(x_{k:n})d\mu(x_{k:n}) = \int_{\mathcal{X}} \ldots \int_{\mathcal{X}} f(x_{k:n})d\mu(x_{k:n})$

$d\mu(x_{k:n}) = d\mu(x_k)d\mu(x_{k+1}) \ldots d\mu(x_n)$

$1_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases}$ - função indicadora do conjunto A

$\delta_{\{x\}}(.) := \delta_x(.)$ - massa de probabilidade concentrada em $\{x\}$ (função delta Dirac)

Introdução

Filtros de Partículas constituem uma classe de algoritmos MCMC sequenciais e com estrutura Bayesiana que permitem, recursivamente a cada passo, atualizações das estimativas das características de interesse. A sua formulação tem como base o modelo de uma cadeia de Markov oculta (HMM: Hidden Markov Model). Assim, dispomos de um processo Markoviano bivariado $\{(X_n, Y_n)\}_{n\geq 0}$ onde a cadeia $X = \{X_n\}_{n\geq 0}$ não é observável e a inferência sobre suas características só podem ser derivadas mediante o processo observável $Y = \{Y_n\}_{n\geq 0}$ e este se relaciona com o sinal oculto X de forma condicionalmente independente. A sua aplicabilidade teve origem na busca de soluções alternativas para problemas de filtragem não-linear em processamento de sinais e, hoje, se diversifica nos mais variados campos da ciência. A idéia básica do Filtro de Partículas foi apresentada em [21]. Ainda que vários trabalhos surgiram na década 1965-1975, sua formulação precisa foi apresentada em 1993 [16].

Quanto à estabilidade do Filtro, em 1957 em [3] foi conjecturado, para $\mathcal{X}$ finito, que a ergodicidade do processo sinal X era suficiente para garantir a independência do Filtro com relação a distribuição inicial. O interesse em [3] não era a estabilidade do Filtro mas sim obter uma fórmula simples para a taxa de entropia do processo de observações Y. Em 1975, em [19], foi apresentado um exemplo em que o sinal X é ergódico mas o Filtro depende da distribuição inicial. O exemplo utilizado em [19] é de uma cadeia de Markov ergódica com $\mathcal{X} = \{1, 2, 3, 4\}$, que já se encontrava em [3], que tem a particularidade que as observações são dadas por uma função determinística do sinal (sem ruído). Ainda em [19], foi obtido um resultado em que a independência da distribuição do Filtro é assegurada para o caso em que o processo sinal X é ergódico mas com hipóteses adicionais sobre a matriz de transição de X e sobre a função que relaciona as observações com o sinal. Aparentemente de forma independente, em 1971 em [20] foi demonstrado que a ergodicidade do sinal é suficiente para que o Filtro seja independente da distribuição inicial. Em [20] o modelo utilizado é a tempo contínuo e com o ruído sendo o Movimento Browniano. Acontece que em [20] foi assumida, sem prova, uma igualdade entre σ-álgebras que posteriormente foi mostrada não ser válida em geral. O contra-exemplo para invalidar a igualdade entre as σ-álgebras assumida em [20] é o próprio exemplo que [19] utilizou para refutar a conjectura em [3]. Entretanto, o contra-exemplo apresenta a particularidade de não conter ruído na expressão das observações, diferentemente da situação em [20]. Dessa forma, a igualdade entre σ-álgebras utilizada em [20] não vale em geral mas permanece em aberto a questão de que tal igualdade seja válida na presença de ruído. Com o tempo apareceram outros trabalhos com outros contra-exemplos de que a igualdade entre as σ-álgebras que [20] utilizou não é válida mas todos têm em comum o fato que as observações são dadas por uma função determinística do sinal. Nosso exemplo 2.2 exibe esta situação pois temos o processo sinal X ergódico com $\mathcal{X} = \{1, 2, 3\}$ e o Filtro depende da distribuição inicial quando $p_1 < p_2$. As observações em nosso exemplo 2.2 são da forma $Y_k = f(X_k)$ exatamente da forma em [19].

O nosso interesse pelo estudo do Filtro de Partículas foi motivado, inicialmente, pela tentativa do seu uso no problema de monitoramento e controle de qualidade "on-line" de um processo de produção. Neste caso, o estado interno do sistema não é observável, porém, podemos obter as características dos itens produzidos que são intrinsicamente relacionados aos estados internos. Maiores detalhes deste processo podem ser encontrados em [11] e [17]. Também, no Capítulo 4 indicamos o uso do Filtro de Bootstrap em problemas desta natureza.

Uma questão delicada na análise do Filtro de Partículas é a sua estabilidade relativa a distribuição inicial. Mais especificamente, se $\pi_{n,\,\nu}^{y_{0:n}}(.)$ representa a distribuição de filtragem, isto é, a distribuição condicional do sinal X_n dadas as observações $Y_{0:n} = y_{0:n} = (y_0, y_1, \ldots, y_n)$ e distribuição inicial ν de X_0, queremos garantir que $\pi_{n,\,\nu}^{y_{0:n}}$ independe de ν quando $n \to \infty$. Isto é, para alguma norma apropriada queremos garantir que

$$\|\pi_{n,\,\nu}^{y_{0:n}}(.) - \pi_{n,\,\nu'}^{y_{0:n}}(.)\| \xrightarrow{n} 0,\ \forall\, \nu\, \forall\, \nu'.$$

Ainda que a questão da estabilidade seja central para a convergência do Filtro, a mesma tem passado desapercebida por um longo tempo. Na literatura, a quase totalidade dos artigos e textos sobre os Filtros assume-se que a estabilidade é garantida pelas hipóteses de ergodicidade do sinal oculto X (ver, por exemplo, a coletânea dos artigos reunidos em

[15]). O primeiro a apontar inconsistência dos resultados devido à falta de estabilidade foi Kaijser em 1975 [19]. Mais recentemente, em 2004, Chigansky e Lipster [7] retomam o estudo desta questão fazendo uso de condições tipo "mixing" sobre o sinal X. Alternativamente, neste trabalho exploramos as propriedades de cadeias de Markov não-homogêneas associadas ao Filtro. Veremos que a estabilidade pode ser garantida pela ergodicidade fraca da cadeia condicionada $X^c = \{X_k^{y_{0:n}}\}_{0 \leq k \leq n}$, onde $X_k^{y_{0:n}} \equiv (X_k | Y_{0:n} = y_{0:n})$. De fato, os Teoremas 3.1 e 3.3 exibem condições suficientes para a estabilidade desejada.

No Capítulo 1, introduzimos a notação e terminologia a serem utilizadas além de alguns resultados sobre cadeias de Markov não-homogêneas. O Capítulo 2 é dedicado à motivação do estudo da estabilidade. Os exemplos 2.1 e 2.2, derivados do problema de controle "on-line", ilustram situações de estabilidade e de não-estabilidade. Também, no Capítulo 2 formulamos as cadeias associadas ao Filtro e derivamos as distribuições de interesse. Os resultados de estabilidade estão reunidos no Capítulo 3. Os Corolários 3.1 e 3.3 mostram que a condição de positividade e limitação da densidade do processo sinal X garantem a estabilidade do Filtro. O nosso Teorema 3.2 mostra que o processo condicionado X^c e o processo filtrado $X^f = \{X_n^{y_{0:n}}\}_{n \geq 0}$ possuem o mesmo comportamento ergódico, sob apropriadas condições. Também, o Teorema 3.3 mostra que a hipótese de limitação e positividade sobre a densidade de transição da cadeia oculta X, comumente utilizada na literatura, pode ser enfraquecida. Finalmente, no Capítulo 4 indicamos o uso do Filtro de Bootstrap para o problema de monitoramento "on-line" e propomos estimadores dos parâmetros do sinal X e estimador tipo núcleo para a distribuição de equilíbrio do processo Y ([11]).

Capítulo 1

Preliminares

Seja $\mathcal{X}$ um conjunto mensurável e $\mathbb{X}$ os subconjuntos de Borel de $\mathcal{X}$. O processo $X = \{X_n\}_{n\geq 0}$ é uma cadeia de Markov homogênea com espaço de estados $\mathcal{X}$ e núcleo de transição (de probabilidade)

$$P = \{P(x, A) : x \in \mathcal{X}, A \in \mathbb{X}\}. \tag{1.1}$$

Assumimos que $P(., .)$ possui densidade $p(.|.)$ relativa à alguma medida σ-finita μ,

$$P(x, A) = \int_A p(x'|x)d\mu(x'). \tag{1.2}$$

O processo X na terminologia de Filtro de Partículas é comumente denominado processo sinal e não é observável. Associado a X temos o processo observável $Y = \{Y_n\}_{n\geq 0}$ com valores em $\mathcal{Y} \subset \mathbb{R}^d$ para algum $d \geq 1$. A sequência $\{Y_n\}$ satisfaz

$$P(Y_n \in B|Y_{0:n-1}, X_{0:n}) = P(Y_n \in B|X_n)$$

e é condicionalmente independente dado $\{X_n\}$, isto é,

$$P(Y_{0:n} \in B_{0:n}|X_{0:n}) = \prod_{j=0}^{n} P(Y_j \in B_j|X_j). \tag{1.3}$$

Usamos a notação $X_{0:n} = (X_0, X_1, \ldots, X_n)$, $B_{0:n} = (B_0, \ldots, B_n)$, etc; $B_j \in \mathbb{Y}$ e $B \in \mathbb{Y}$. Assumimos que Y_n, para $n = 1, 2, \ldots$, satisfaz

$$P(Y_n \in B|X_n = x) = \int_B q(y|x)d\lambda(y) \tag{1.4}$$

onde λ é alguma medida σ-finita. E, por conveniência, tomamos $P(Y_0 = 0) = 1$.

Um processo bivariado (X, Y) onde X é uma cadeia de Markov e Y satisfaz a relação (1.3) é denominado um modelo de cadeia de Markov oculta (HMM: Hidden Markov Model). O Filtro de Partículas tem por base um HMM. Note que o processo conjunto $(X, Y) = \{(X_n, Y_n)\}_{n\geq 0}$ também é uma cadeia de Markov. Usando a notação $g(.)$ para denotar qualquer densidade relativa a alguma medida σ-finita temos,

$$g(x_{n+1}, y_{n+1}|x_{0:n}, y_{0:n}) = \frac{g(x_{0:n+1}, y_{0:n+1})}{g(x_{0:n}, y_{0:n})}$$

e pelas relações (1.2)-(1.4) segue que

$$\begin{aligned} g(x_{n+1}, y_{n+1}|x_{0:n}, y_{0:n}) &= q(y_{n+1}|x_{n+1})p(x_{n+1}|x_n) \\ &= g(x_{n+1}, y_{n+1}|x_n, y_n). \end{aligned} \tag{1.5}$$

Na seção 1.1, introduzimos os algoritmos tipo Filtro de Partículas onde o problema central da filtragem consiste em calcular a distribuição de filtragem $P(X_n \in A|Y_{0:n} = y_{0:n})$ ou, equivalentemente, alguma característica da distribuição do tipo

$$E(f(X_n)|Y_{0:n} = y_{0:n}) = \int_{\mathcal{X}} f(x_n)dP(x_n|y_{0:n}).$$

Destacamos o interesse no comportamento assintótico da equação de filtragem e detalhamos o Filtro de Bootstrap, algoritmo popularmente utilizado para se estimar a distribuição de filtragem de forma recursiva no tempo.

Na seção 1.2, apresentamos resultados conhecidos sobre cadeias de Markov não-homogêneas e que serão utilizados na análise da estabilidade dos filtros. Coeficiente δ de Dobrushin e os conceitos de ergodicidade fraca e forte serão introduzidos. Condições de ergodicidade D (de Doeblin) e as extensões D^* e $\{D_k^*\}$ de [13] e [24] também serão detalhadas.

Uma listagem das principais notações a serem utilizadas pode ser encontrada logo após o índice.

1.1 Filtro de Partículas

Filtros de Partículas constituem uma classe de algoritmos Monte Carlo sequencial formulados, originalmente, para solucionar numericamente o problema de ótima filtragem em modelos de Equações de Estados não-linear. Tipicamente, estes modelos de estado são descritos pelas equações

$$\begin{aligned} x_{t+1} &= f_t(x_t, \varepsilon_t) \\ y_t &= h_t(x_t, \eta_t), \end{aligned} \tag{1.6}$$

onde os sinais $\{x_t\}$ são não observáveis, os ruídos aleatórios $\{\varepsilon_t\}$ e $\{\eta_t\}$ são mutuamente independentes com as variáveis ε_t´s e η_t´s iid e dispomos das observações $\{y_t\}$. Exceto em situações particulares o sistema (1.6) não possui solução explícita. O Filtro de Partículas propõe soluções para aproximar a distribuição a posteriori do sinal $g(x_t|y_{0:t})$ por distribuições empíricas que evoluem aleatoriamente no tempo conforme a dinâmica do modelo e das observações $y_0, y_1, y_2, \ldots$.

Na nossa formulação os sinais são representados pelo processo sinal $\{X_n\}$ e as observações pelo processo $\{Y_n\}$. Neste caso, se as Equações de Estados (1.6) tiverem a representação

$$\begin{aligned} X_{n+1} &= f_n(X_n) + \varepsilon_n \\ Y_n &= h_n(X_n) + \eta_n, \end{aligned}$$

temos a estrutura de um HMM. A distribuição a posteriori, também conhecida como distribuição de filtragem, é dada por

$$\begin{aligned} \pi_n^{y_{0:n}}(A) &= \int_A \pi_n^{y_{0:n}}(x_n) d\mu(x_n) \\ &= P(X_n \in A | Y_{0:n} = y_{0:n}). \end{aligned} \tag{1.7}$$

Ainda nesta formulação mais simples o Filtro de Kalman só proporciona solução para o caso de f_n e h_n serem lineares e os ruídos $\{\varepsilon_n\}$ e $\{\eta_n\}$ Gaussianos. Neste caso, a distribuição de filtragem é Gaussiana multivariada. A idéia básica do Filtro de Partículas é aproximar, a cada passo, a distribuição de filtragem pela distribuição empírica ponderada da coleção de partículas $\{x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(N)}\}$

$$\hat{\pi}_{n,N}^{y_{0:n}}(x) = \sum_{j=1}^{N} w_n^{(j)} \delta_{x_n^{(j)}}(x) = \sum_{j=1}^{N} w_n^{(j)} 1_{x_n^{(j)}}(x),$$

onde $w_n^{(1)}, \ldots, w_n^{(N)}$ são pesos convenientemente selecionados e δ_x é a distribuição massa de probabilidade em x (função delta Dirac). Alguns algoritmos incluem uma etapa de reamostragem, por importância por exemplo, e ao final é obtida uma aproximação onde as partículas são igualmente pesadas. Um exemplo de algoritmo que faz uso da etapa de reamostragem é o Filtro de Bootstrap como veremos abaixo.

Existe uma vasta literatura sobre Filtros de Partículas com variantes envolvendo diferentes estratégias de reamostragem/função-peso ou técnicas alternativas nas etapas de predição e atualização. Como referência básica sugerimos [4], [5], [6], [7] e [15] não por serem os principais trabalhos na área mas por serem mais relacionados com os problemas abordados nesta tese. Os vários algoritmos propostos para estimar $\pi_n^{y_{0:n}}(.)$ exploram as equações recursivas de filtragem:

$$\pi_n^{y_{0:n}}(x_n) = \frac{q(y_n|x_n)\int_{\mathcal{X}} p(x_n|x_{n-1})\pi_{n-1}^{y_{0:n-1}}(x_{n-1})d\mu(x_{n-1})}{\int_{\mathcal{X}^2} q(y_n|x'_n)p(x'_n|x'_{n-1})\pi_{n-1}^{y_{0:n-1}}(x'_{n-1})d\mu(x'_{n-1:n})}, \tag{1.8}$$

onde $\pi_0^{y_0} = \nu$. A obtenção de (1.8) é consequência imediata do Teorema de Bayes e da relação (1.3). Estas equações recursivas (1.8), via de regra, não possuem soluções explícitas e seu uso é subdividido em duas etapas: predição e atualização. Na etapa de predição temos

$$\begin{aligned} g(x_k|y_{0:k-1}) &= \int_{\mathcal{X}} g(x_k|x_{k-1})g(x_{k-1}|y_{0:k-1})d\mu(x_{k-1}) \\ &= \int_{\mathcal{X}} p(x_k|x_{k-1})\pi_{k-1}^{y_{0:k-1}}(x_{k-1})d\mu(x_{k-1}) \end{aligned} \tag{1.9}$$

e na etapa de atualização

$$g(x_k|y_{0:k}) = \frac{q(y_k|x_k)g(x_k|y_{0:k-1})}{\int_{\mathcal{X}} q(y_k|x_k)g(x_k|y_{0:k-1})d\mu(x_k)}. \tag{1.10}$$

Observe que a equação de predição fornece uma maneira de aproximar a atualização quando temos disponível o estimador $\hat{\pi}_{k-1,N}^{y_{0:k-1}}$ de $\pi_{k-1}^{y_{0:k-1}}$. Um algoritmo que faz uso do recurso predição/atualização é o algoritmo Filtro de Bootstrap, bastante popular devido as suas boas propriedades computacionais. O algoritmo é chamado Filtro de Bootstrap porque sua fase de atualização, equação (1.10), é implementada como um "Bootstrap pesado", ou seja, através de reamostragem com pesos.

Algoritmo Filtro de Bootstrap

1. $k = 0$
   - Para $i = 1, \ldots, N$, amostrar $x_0^{(i)} \sim \nu$ e definir $k = 1$
2. $k \geq 1$
   - Para $i = 1, \ldots, N$, amostrar $\tilde{x}_k^{(i)} \sim p(x_k | x_{k-1}^{(i)})$
   - Para $i = 1, \ldots, N$, calcular os pesos

$$\tilde{w}_k^{(i)} = q(y_k | \tilde{x}_k^{(i)})$$

   - Normalizar os pesos

$$w_k^{(i)} = \frac{\tilde{w}_k^{(i)}}{\displaystyle\sum_{j=1}^{N} \tilde{w}_k^{(j)}}$$

3. Reamostrar, com reposição, N partículas $\{x_k^{(i)} : i = 1, \ldots, N\}$ a partir do conjunto $\{\tilde{x}_k^{(i)} : i = 1, \ldots, N\}$ de acordo com os pesos $\{w_k^{(i)} : i = 1, \ldots, N\}$
   - $k \longleftarrow k + 1$ e ir para o passo 2

A dinâmica deste algoritmo é bastante simples. Para compilar o algoritmo é necessário uma amostra da distrituição inicial ν, conhecer a densidade $q(.|.)$ e ser possível gerar amostras da densidade do ruído $\{\varepsilon_n\}$ (geralmente esta densidade é conhecida).

Do passo $k-1$ tem-se disponível as observações $y_{0:k-1}$ e o conjunto de partículas $\{x_{k-1}^{(1)}, x_{k-1}^{(2)}, \ldots, x_{k-1}^{(N)}\}$ que foi obtido no final do passo 3. O conjunto destas partículas possui peso uniforme e sua distribuição empírica aproxima $\pi_{k-1}^{y_{0:k-1}}(x_{k-1}) = g(x_{k-1}|y_{0:k-1})$. No passo 2, cada partícula é passada pelo sistema, ou seja,

$$\tilde{x}_k^{(i)} = f_{k-1}(x_{k-1}^{(i)}) + \varepsilon_{k-1}^{(i)}$$

e obtém-se o conjunto de partículas $\{\tilde{x}_k^{(1)}, \tilde{x}_k^{(2)}, \ldots, \tilde{x}_k^{(N)}\}$ igualmente pesadas. A distribuição empírica deste conjunto de partículas fornece uma aproximação para $g(x_k|y_{0:k-1})$. Ainda no passo 2, tem-se acesso à observação y_k e, como $q(.|.)$ é conhecida, os pesos de importância $\{w_k^{(1)}, w_k^{(2)}, \ldots, w_k^{(N)}\}$ podem ser calculados. No passo 3, reamostra-se, com reposição, N partículas do conjunto $\{\tilde{x}_k^{(1)}, \tilde{x}_k^{(2)}, \ldots, \tilde{x}_k^{(N)}\}$ de acordo com os pesos $\{w_k^{(1)}, w_k^{(2)}, \ldots, w_k^{(N)}\}$ para obter $\{x_k^{(1)}, x_k^{(2)}, \ldots, x_k^{(N)}\}$. A reamostragem é feita de forma que

$$P\left(x_k^{(j)} = \tilde{x}_k^{(i)}\right) = w_k^{(i)}, \quad i, j = 1, 2, \ldots, N.$$

É atribuído peso uniforme para as partículas obtidas no final do passo 3, $\{x_k^{(1)}, x_k^{(2)}, \ldots, x_k^{(N)}\}$ e o resultado é que a distribuição empírica deste conjunto de partículas fornece aproximação para $\pi_k^{y_{0:k}}(x_k) = g(x_k|y_{0:k})$.

1.2 Cadeias de Markov Não-Homogêneas

O conteúdo desta seção pode ser encontrado em [8], [10], [18], [22] e [24]. Entretanto, os trabalhos [8] e [24] são as referências mais indicadas para o nosso tipo de abordagem.

Consideraremos cadeias de Markov sobre um conjunto arbitrário $\mathcal{X}$ equipado com uma σ-álgebra $\mathbb{X}$ contavelmente gerada, isto é, gerada por uma coleção contável de subconjuntos de $\mathcal{X}$. No texto deixaremos claro o espaço $(\mathcal{X}, \mathbb{X})$ sobre o qual as cadeias estão definidas. Também, estaremos interessados em cadeias de Markov não-homogêneas. Neste caso, temos uma sequência de núcleos de transição de $\{P_n(x, A) \ : \ x \in \mathcal{X}, \ A \in \mathbb{X}\}_{n \geq 0}$,

$$P_n(x, A) = P(X_{n+1} \in A | X_n = x) = P_{n,n+1}(x, A),$$

onde a m-ésima transição da cadeia é dada pela convolução dos núcleos

$$P_n * P_{n+1} * \ldots * P_{n+m-1},$$

com

$$\begin{aligned} P_{n,m+n}(x, A) &= P(X_{n+m} \in A | X_n = x) \\ &= P_n * P_{n+1} * \ldots * P_{n+m-1}(x, A) \\ &= \int_{\mathcal{X}} \ldots \int_{\mathcal{X}} P_n(x, dy_1) P_{n+1}(y_1, dy_2) \ldots P_{n+m-1}(y_{m-1}, A). \end{aligned}$$

A convolução de dois núcleos de probabilidade P e Q em $(\mathcal{X}, \mathbb{X})$ é dada por

$$P * Q(x, A) = \int_{\mathcal{X}} P(x, dy) Q(y, A).$$

E a correspondente equação de Chapman-Kolmogorov é dada por

$$P_{n,n+m} = P_{n,n+r} * P_{n+r,n+m}.$$

Para uma cadeia homogênea P, sua ergodicidade é garantida pela existência de uma probabilidade π_∞ tal que

$$\lim_{n\to\infty} P^n(x,A) = \pi_\infty(A),\ \forall\, x \in \mathcal{X} \text{ e } \forall\, A \in \mathbb{X}.$$

Já para uma cadeia não-homogênea $\{P_n\}$, dois tipos de ergodicidade são estudados: a ergodicidade fraca e a ergodicidade forte. A ergodicidade fraca é baseada no coeficiente ergódico de Dobrushin $\delta(.)$ e a ergodicidade forte na existência de um núcleo de transição limite constante.

Para o conceito de ergodicidade fraca (forte) a distância a ser usada é a da norma da variação total. Para um medida sinalada μ sobre $(\mathcal{X},\mathbb{X})$ definimos a variação total (VT) por

$$VT(\mu) = \sup_{A\in\mathbb{X}} \mu(A) - \inf_{B\in\mathbb{X}} \mu(B).$$

E para dois núcleos de transição P e Q em $(\mathcal{X},\mathbb{X})$

$$\begin{aligned} \|P-Q\| &= \sup_{x\in\mathcal{X}} VT(P(x,.)-Q(x,.)) \\ &= 2\sup_{x\in\mathcal{X}}\sup_{A\in\mathbb{X}} |P(x,A)-Q(x,A)|, \end{aligned} \tag{1.11}$$

que no caso enumerável se reduz a

$$\|P-Q\| = \sup_{i\in\mathcal{X}} \sum_{j\in\mathcal{X}} |P(j|i)-Q(j|i)|.$$

Definição 1.1. Para um núcleo de transição P o coeficiente ergódico de Dobrushin é dado por

$$\delta(P) = \sup_{\substack{x,\,y\,\in\mathcal{X} \\ A\,\in\,\mathbb{X}}} |P(x,A)-P(y,A)|. \tag{1.12}$$

Definição 1.2. A cadeia $\{P_n\}$ é dita ergódica fraca se satisfizer (a), (b) ou (c)

(a) para todo m

$$\lim_{n\to\infty} \delta(P_{m,m+n}) = 0$$

(b) para quaisquer probabilidades μ_0 e μ_1 em $(\mathcal{X},\mathbb{X})$

$$\lim_{n\to\infty} \|\mu_0 P_{m,m+n} - \mu_1 P_{m,m+n}\| = 0,\ \forall\, m$$

(c) para todo m, existe uma sequência de núcleos de transição constante $\{R_{mn}\}_{n\geq 0}$ tal que

$$\lim_{n\to\infty} \|P_{m,m+n} - R_{mn}\| = 0.$$

Na definição acima, um núcleo de transição R é dito constante se $R(x, A) = R(y, A)$ $\forall\, x, y \in \mathcal{X}$ e $\forall\, A \in \mathbb{X}$. Também, (a), (b) e (c) são equivalentes.

Definição 1.3. A cadeia $\{P_n\}$ é dita ergódica forte se existe um núcleo de transição constante P_∞ tal que

$$\lim_{n\to\infty} \|P_{m,m+n} - P_\infty\| = 0,\ \forall\, m.$$

Definição 1.4. A cadeia P é uniformemente ergódica se existe uma medida de probabilidade π_∞ tal que

$$\lim_{n\to\infty} \sup_{x\in\mathcal{X}} \|P^n(x, .) - \pi_\infty(.)\| = 0.$$

Observações

(a) Se a cadeia é homogênea, $P_n = P$, podemos mostrar que: (i) se a cadeia for fracamente ergódica ela será ergódica; (ii) a cadeia será fracamente ergódica se, e somente se, for fortemente ergódica.

(b) O seguinte exemplo, extraído de [24], exibe uma situação em que P é ergódica porém não é fracamente ergódica,

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots \end{bmatrix}$$

Na análise da ergodicidade de cadeias de Markov várias condições são comumente utilizadas. Historicamente, a condição de Doeblin é uma das mais importantes condições que implicam a ergodicidade de cadeias.

Condição de Doeblin (D):

Existe uma probabilidade ϕ, um inteiro $m_0 \geq 1$ e constantes $\varepsilon < 1$ e $\delta > 0$ tal que para todo $A \in \mathbb{X}$,

$$\text{se } \phi(A) > \varepsilon \text{ então } \inf_{x\,\in\mathcal{X}} P^{m_0}(x, A) \geq \delta. \tag{1.13}$$

Ou, equivalentemente, a hipótese de Doob:

$$\text{se } \ \phi(A) < \varepsilon \ \text{ então } \ \sup_{x \in \mathcal{X}} P^{m_0}(x, A) \leq 1 - \delta. \tag{1.14}$$

Em [13] foi proposta uma variação da condição de Doeblin que é equivalente a ergodicidade uniforme. Tal condição é

Condição (D^*):
Existe uma probabilidade ϕ, um inteiro $m_0 \geq 1$ e constantes $\varepsilon < \dfrac{1}{2}$ e $\delta > 0$ tal que para todo $A \in \mathbb{X}$,

$$\text{se } \phi(A) > \varepsilon \text{ então } \inf_{x \in \mathcal{X}} P^{m_0}(x, A) \geq \delta. \tag{1.15}$$

Ainda em [13] foi proposta uma condição que, para cadeias não-homogêneas, é equivalente a ergodicidade fraca.

Condição (D_k^*):
Para cada k, existe uma probabilidade ϕ_k, um inteiro $m_k \geq 1$ e constantes $\varepsilon_k < \dfrac{1}{2}$ e $\delta_k > 0$ tal que para cada $A \in \mathbb{X}$,

$$\text{se } \ \phi_k(A) > \varepsilon_k \ \text{ então } \ \inf_{x \in \mathcal{X}} P_{k,k+m_k}(x, A) > \delta_k \tag{1.16}$$

e $\displaystyle\sum_{k'} \delta_{k'} = \infty$ para qualquer subsequência $\{k'\}$ de $\{k\}$.

Extraímos de [8] e [24] os seguintes resultados sobre ergodicidade de cadeias não-homogêneas necessários para o desenvolvimento do Capítulo 3.

Proposição 1.1. *Sejam P, Q e R núcleos de transição. Então*
(a) $\|P * (Q - R)\| \leq \|Q - R\|$;
(b) $\|(P - Q) * R\| \leq \|P - Q\|\delta(R) \leq \|P - Q\|$;
(c) $|\delta(P) - \delta(Q)| \leq 2\|P - Q\|$;
(d) $\delta(PQ) \leq \delta(P)\delta(Q)$.

Proposição 1.2. *Seja $\{P_n\}$ uma cadeia de Markov não-homogênea e P um núcleo de transição. Então*
a) se para todo n, $P_n = P$, então a cadeia é ergódica se, e somente se, é fracamente ergódica;
b) se P satisfaz a Condição D, então P é uniformemente ergódica;

c) P é uniformemente ergódica se, e somente se, satisfaz a Condição D^*;

d) A cadeia $\{P_n\}$ é fracamente ergódica se, e somente se, satisfaz a Condição D_k^*.

Capítulo 2

Algoritmo Filtro de Partículas e Motivação

Neste capítulo, através de exemplos de Filtros de Partículas com espaço de estados finito, ilustraremos situações de não-estabilidade. Mais especificamente, situações em que a distribuição de filtragem $\pi_n^{y_{0:n}}$ depende da distribuição inicial ν do sinal não observado X,

$$\|\pi_{n,\,\nu}^{y_{0:n}} - \pi_{n,\,\nu'}^{y_{0:n}}\| = \sum_{x\in\mathcal{X}} |P_\nu(X_n = x|Y_{0:n} = y_{0:n}) - P_{\nu'}(X_n = x|Y_{0:n} = y_{0:n})| \overset{n\to\infty}{\nrightarrow} 0,$$

onde ν e ν' são duas distribuições iniciais distintas de X_0.

Na seção 2.1, introduzimos o Filtro de Partículas que tem sido estudado no contexto de canal de ruído. Neste Filtro o sinal observado consiste do sinal oculto perturbado por um ruído ξ_n, independente do sinal oculto. Nesta seção estudamos a estrutura particular de HMM onde $\mathcal{X}$ e $\mathcal{Y}$ onde são finitos com o objetivo de motivar o estudo da seção 2.2 e da seção 2.3.

Na seção 2.2, apresentamos dois exemplos onde verificamos a estabilidade ou não estabilidade do Filtro. No exemplo 2.1, no contexto de controle on-line, a cadeia oculta X representa o estado de funcionamento de um equipamento em um processo de produção. A partir do exemplo 2.1 verificamos que, no caso em que $\mathcal{X}$ é finito com 2 estados, o Filtro associado ao HMM é sempre estável no sentido que a distribuição de filtragem não depende da distribuição inicial. O exemplo 2.2, também considerado no contexto de controle on-line, é importante para mostrar que a propriedade desejável de estabilidade do Filtro nem sempre é válida. Para o caso de $\mathcal{X}$ finito com 3 estados, verificamos no exemplo 2.2 que o Filtro pode ou não ser estável. A estabilidade ou não do Filtro depende da relação entre as probabilidades p_1 e p_2. Mostramos que, quando $p_1 > p_2$, o Filtro é estável e quando $p_1 < p_2$ o Filtro não é estável. Para mostrar que o Filtro não é estável fixamos uma dada observação $Y_{0:n} = y_{0:n}$ e, para condições iniciais apropriadas ν e ν', mostramos que $\|\pi^{y_{0:n}}_{n,\,\nu} - \pi^{y_{0:n}}_{n,\,\nu'}\| \overset{n\to\infty}{\nrightarrow} 0$. Em cada exemplo obtemos a distribuição de $Y_n|X_n = x$, exibimos a matriz de transição do processo conjunto $(X, Y) = \{(X_n, Y_n)\}$, a distribuição do processo sinal filtrado X^f (que é a distribuição de filtragem) e a matriz de transição da cadeia não-homogênea X^c (cadeia condicionada).

Na seção 2.3, tratamos das cadeias não-homogêneas associadas a um HMM. Voltando para o caso geral, analisamos as cadeias estendida $\{(X_n, Y_{0:n})\}_{n\geq 0}$ e condicionada $X^c = \{X^{y_{0:n}}_k\}_{0\leq k\leq n}$ e derivamos as respectivas densidades de transição. Também, mostramos que a distribuição de filtragem $\pi^{y_{0:n}}_n$ pode ser obtida a partir da cadeia estendida $\{(X_n, Y_{0:n})\}_{n\geq 0}$. Observamos ainda a grande dificuldade em se trabalhar com o processo $X^f = \{X^{y_{0:n}}_n\}_{n\geq 0}$ devido ao fato que suas transições estão realacionadas a um processo de espaços de estados variantes (veja [12] para um estudo de processos com espaços de estados variantes no contexto do algoritmo EM - Expectation Maximization). Novamente considerando o caso de equações de estados (2.20), obtemos a densidade de transição do processo X^f, que neste caso será uma cadeia de Markov.

2.1 Filtro de Partículas: Caso Particular

Em modelagem de “canal de ruído” é comum a variável observável Y_n ser o próprio sinal X_n acrescido de um ruído multiplicativo. Assim, quando $X_n = i$ temos $Y_n = 1_{(X_n=i)}\xi_n(i)$, onde o ruído $\xi_n(i)$ é provocado pelo canal quando o sinal i é transmitido. Neste caso, se o espaço de estados de X for finito, $\mathcal{X} = \{1, \ldots, K\}$, podemos escrever

$$Y_n = \sum_{i=1}^{K} 1_{(X_n=i)}\xi_n(i), \tag{2.1}$$

onde $\{(\xi_n(1), \ldots, \xi_n(K))\}_{n\geq 0}$ é uma sequência de vetores aleatórios iid e independentes de X. Assume-se também que os componentes $\xi_n(1), \ldots, \xi_n(K)$ são independentes e possuem, respectivamente, densidades $g_i(.)$, $i = 1, \ldots, K$. Maiores detalhes deste modelo podem ser encontrados em [6].

Estudaremos a estabilidade do Filtro associado ao HMM através da análise da ergodicidade fraca de cadeias de Markov não-homogêneas associadas ao HMM e convenientemente escolhidas. Assim, aplicando a nossa abordagem ao modelo (2.1) derivamos as várias cadeias de Markov associadas bem como a distribuição de filtragem. Naturalmente, o sinal $\{X_n\}$ é uma cadeia de Markov com probabilidades de transição $p(.|.)$. O processo observável $\{Y_n\}$ satisfaz

$$\begin{aligned} q(y_n|i) &= P(Y_n \in dy_n|X_n = i) \\ &= P(\xi_n(i) \in dy_n) \\ &= g_i(y_n)d\lambda(y_n). \end{aligned}$$

O processo conjunto $\{(X_n, Y_n)\}$ é Markoviano e possui núcleo de transição dado por

$$P((X_n, Y_n) \in \{x_n\} \times B|X_{n-1} = x_{n-1}, Y_{n-1} = y_{n-1}) = \int_B g_{x_n}(y)p(x_n|x_{n-1})d\lambda(y). \quad (2.2)$$

Para (2.2) basta notar que

$$\begin{aligned} g(x_{n-1:n}, y_{n-1:n}) &= q(y_n|x_n)p(x_n|x_{n-1})g(x_{n-1}, y_{n-1}) \\ &= g_{x_n}(y_n)p(x_n|x_{n-1})g(x_{n-1}, y_{n-1}). \end{aligned}$$

O processo conjunto estendido $\{(X_n, Y_{0:n})\}$ é uma cadeia de Markov e o seu núcleo de transição é dado por

$$P((X_n, Y_{0:n}) \in \{x_n\} \times B_{0:n}|x_{n-1}, y_{0:n-1}) = 1_{(y_{0:n-1} \in B_{0:n-1})} \left[\int_{B_n} g_{x_n}(y)p(x_n|x_{n-1})d\lambda(y) \right]. \quad (2.3)$$

Trata-se de uma cadeia de difícil trato, visto que os espaços de estados são variantes, pois, $(X_{n-1}, Y_{0:n-1})$ toma valores em $\mathcal{X} \times \mathcal{Y}^n$ e $(X_n, Y_{0:n})$ em $\mathcal{X} \times \mathcal{Y}^{n+1}$. No entanto, o processo tem utilidade para analisar propriedades de distribuição de filtragem. De fato, podemos derivar a distribuição de filtragem

$$\pi_n^{y_{0:n}}(i) = P(X_n = i|Y_{0:n} = y_{0:n}).$$

Note que de (2.1) temos

$$P(Y_n \in B) = \sum_{i=1}^{K} P(X_n = i) P(\xi_n(i) \in B),$$

de modo que a densidade de Y_n é dada por

$$g(y_n) = \sum_{i=1}^{K} P(X_n = i) g_i(y_n).$$

E como $P(X_n = i | Y_{0:n} = y_{0:n}) = P(X_n = i | Y_n = y_n)$, temos

$$\pi_n^{y_{0:n}}(i) = \frac{P(X_n = i) g_i(y_n)}{\sum_{j=1}^{K} P(X_n = j) g_j(y_n)}. \tag{2.4}$$

O processo do sinal condicionado $X^c = \{X_k^{y_{0:n}}\}_{0 \le k \le n}$ é uma cadeia de Markov não-homogênea e a distribuição condicionada, por (2.1), é dada por

$$\begin{aligned} \tilde{\pi}_k^{y_{0:n}}(x_k) &= P(X_k^{y_{0:n}} = x_k) = P(X_k = x_k | Y_{0:n} = y_{0:n}) \\ &= P(X_k = x_k | Y_k = y_k) = \frac{g(x_k, y_k)}{g(y_k)} \\ &= \frac{P(X_k = x_k) g_{x_k}(y_k)}{\sum_{j=1}^{K} P(X_k = j) g_j(y_k)}. \end{aligned} \tag{2.5}$$

Para o cálculo das matrizes de transição $\tilde{P}_k^{y_{0:n}} = (\tilde{p}_k^{y_{0:n}}(x_{k+1}|x_k))$,

$$\begin{aligned} P(X_{k+1} = x_{k+1}, X_k = x_k | Y_{0:n} = y_{0:n}) &= P(X_{k+1} = x_{k+1}, X_k = x_k | Y_{0:k+1} = y_{0:k+1}) \\ &= g_{x_{k+1}}(y_{k+1}) p(x_{k+1}|x_k) \frac{P(X_k = x_k | Y_{0:k} = y_{0:k})}{P(Y_{0:k+1} = y_{0:k+1})} \end{aligned}$$

e

$$P(X_k = x_k | Y_{0:n} = y_{0:n}) = \sum_{x'_{k+1}} P(X_{k+1} = x'_{k+1}, X_k = x_k | Y_{0:n} = y_{0:n}).$$

Segue que, para $k < n$,

$$\begin{aligned} \tilde{p}_k^{y_{0:n}}(x_{k+1}|x_k) &= P(X_{k+1}^{y_{0:n}} = x_{k+1} | X_k^{y_{0:n}} = x_k) \\ &= \frac{g_{x_{k+1}}(y_{k+1}) p(x_{k+1}|x_k)}{\sum_{j=1}^{K} g_j(y_{k+1}) p(j|x_k)}. \end{aligned} \tag{2.6}$$

Para o processo do sinal filtrado $X^f = \{X_k^{y_{0:k}}\}_{k\geq 0}$, note que a estrutura (2.1) garante que é uma cadeia de Markov. Como para $n > k$ temos

$$P(X_k = x_k | Y_{0:k} = y_{0:k}) = P(X_k = x_k | Y_{0:n} = y_{0:n}),$$

segue

$$\pi_k^{y_{0:k}} = \tilde{\pi}_k^{y_{0:k}} \quad \text{e} \quad P_k^{y_{0:k+1}} = \tilde{P}_k^{y_{0:n}}. \tag{2.7}$$

2.2 Motivação

Nesta seção motivaremos a questão da estabilidade do Filtro. Afim de ilustrar situações em que a distribuição de filtragem depende da distribuição inicial faremos uso de exemplos relacionados com o modelo de monitoramento de qualidade "on-line". A cadeia de Markov X representa o estado de funcionamento de algum equipamento eletrônico, por exemplo uma máquina em um processo industrial, onde o estado 1 representa que o processo de produção está *sob controle* (ou que a máquina está funcionando bem) enquanto o estado 2 representa que o processo está *fora de controle* (ou que a máquina não está funcionando bem). Ao se considerar vários estados $\{1, \ldots, K\}$ podemos interpretar os estados $2, 3, \ldots, K$ como sendo subsequentes estados de deterioração. Neste sentido, não observamos o estado de funcionamento da máquina mas temos acesso aos itens por ela produzidos. A conclusão que a máquina está funcionando bem ou mal deve ser tirada a partir das observações $\{Y_n\}$ que representam as observações de alguma característica dos itens produzidos. Dessa forma, temos uma estrutura de HMM onde X é a cadeia de Markov oculta e Y é a sequência de observações.

Sendo a cadeia de Markov X oculta, na maioria das vezes não dispomos da distribuição inicial. Como nosso interesse é obter a distribuição de X_n dadas as observações Y_k até o tempo n, seria desejável que tal distribuição não dependesse da distribuição inicial da cadeia. Verificaremos através dos exemplos a seguir que isso nem sempre acontece.

Exemplo 2.1. *Assume-se que a cadeia X toma valores em $\mathcal{X} = \{1, 2\}$ e possui matriz de transição*

$$P = \begin{bmatrix} 1-p_1 & p_1 \\ \varepsilon & 1-\varepsilon \end{bmatrix}, \tag{2.8}$$

onde $0 < p_1 < 1$ *e* $0 < \varepsilon < 1$. *No contexto de controle "on-line"* p_1 *é grande em relação a* ε. *O processo* Y *toma valores em* $\mathcal{Y} = \{0, 1\}$ *e satisfaz* $Y_n = 1_{(X_n=1)}$. *Note que*

$$Y_n = 1_{(X_n=1)}\xi_n(1) + 1_{(X_n=2)}\xi_n(2) = 1_{(X_n=1)},$$

onde $\xi_n = (\xi_n(1), \xi_n(2))$ *e* $P(\xi_n(1) = 1) = P(\xi_n(2) = 0) = 1$.

Como $P(X_n = 1, Y_n = 1) = 1$ *e* $P(X_n = 1, Y_n = 0) = 0$, *facilmente obtemos as distribuições condicionais*

$$\begin{array}{rcl} Y_n|X_n = 1 & \sim & q(y|1) = y \\ Y_n|X_n = 2 & \sim & q(y|2) = 1 - y. \end{array} \tag{2.9}$$

Claramente, temos também a relação

$$P(Y_{0:n} = y_{0:n}|X_{0:n} = x_{0:n}) = \prod_{j=0}^{n} P(Y_j = y_j|X_j = x_j) = \prod_{j=0}^{n} q(y_j|x_j).$$

Existem vários processos associados ao modelo. O processo sinal X *é ergódico e possui distribuição limite* $\left(\dfrac{\varepsilon}{p_1+\varepsilon}, \dfrac{p_1}{p_1+\varepsilon}\right)$ *independente da distribuição inicial* ν *de* X_0. *O processo observável* Y, *neste exemplo particular, é uma cadeia de Markov. O processo conjunto* $(X, Y) = \{(X_n, Y_n)\}_{n\geq 0}$ *é também uma cadeia de Markov com a mesma matriz de transição de (2.8) e distribuição inicial* $(X_0, Y_0) \sim \nu(x_0)1_{\{0\}}(y_0)$.

Embora as relações (2.3)-(2.6) possam ser utilizadas, neste exemplo o cálculo direto é mais fácil.
(*a*) *Processo conjunto estendido* $\{(X_n, Y_{0:n})\}_{n\geq 0}$.
É um cadeia de Markov com espaço de estados variantes. Note que $(X_{n-1}, Y_{0:n-1})$ *toma valores em* $\{1, 2\} \times \{0, 1\}^n$ *e* $(X_n, Y_{0:n})$ *em* $\{1, 2\} \times \{0, 1\}^{n+1}$. *Como*

$$P(X_n = 1, Y_{0:n} = y_{0:n}) = \begin{cases} 0 & se \;\; y_n = 0 \\ 1 & se \;\; y_n = 1 \end{cases}$$

e $P(X_n = 2, Y_{0:n} = y_{0:n}) = 1 - P(X_n = 1, Y_{0:n} = y_{0:n})$, *segue que*

$$g(1, y_{0:n}) = y_n \quad e \quad g(2, y_{0:n}) = 1 - y_n.$$

Mais ainda, as probabilidades de transição são dadas por

$$\begin{array}{rcl} g(1, y_{0:n}|x_{n-1}, y_{0:n-1}) & = & y_n \\ g(2, y_{0:n}|x_{n-1}, y_{0:n-1}) & = & 1 - y_n. \end{array}$$

Daí, temos, independência da distribuição inicial,

$$\pi_n^{y_{0:n}}(1) = y_n \quad e \quad \pi_n^{y_{0:n}}(2) = 1 - y_n. \tag{2.10}$$

(b) Processo do sinal condicionado $X^c = \{X_k^{y_{0:n}}\}_{0 \le k \le n}$.
Temos

$$\begin{aligned} \tilde{p}_k^{y_{0:n}}(x_{k+1}|x_k) &= P(X_{k+1}^{y_{0:n}} = x_{k+1}|X_k^{y_{0:n}} = x_k) \\ &= \frac{P(X_{k+1} = x_{k+1}, X_k = x_k|Y_{0:n} = y_{0:n})}{P(X_k = x_k|Y_{0:n} = y_{0:n})}. \end{aligned}$$

Usando o fato que $X_{k+1} = 1_{(Y_{k+1}=1)} + 2\ 1_{(Y_{k+1}=0)}$ *temos* $\tilde{p}_k^{y_{0:n}}(1|x_k) = y_{k+1}$, $\tilde{p}_k^{y_{0:n}}(2|x_k) = 1 - y_{k+1}$ *e*

$$\tilde{P}_k^{y_{0:n}} = \begin{bmatrix} y_{k+1} & 1 - y_{k+1} \\ y_{k+1} & 1 - y_{k+1} \end{bmatrix}.$$

Se a distribuição inicial for ν *temos também*

$$\tilde{P}_0^{y_{0:n}} = \begin{bmatrix} \nu(1) & \nu(2) \\ \nu(1) & \nu(2) \end{bmatrix}$$

e para $0 \le k < n$ *e* $k + m < n$ *temos*

$$\tilde{P}_{k,k+m}^{y_{0:n}} = \tilde{P}_k^{y_{0:n}} * \ldots * \tilde{P}_{k+m-1}^{y_{0:n}} = \begin{bmatrix} y_{k+m} & 1 - y_{k+m} \\ y_{k+m} & 1 - y_{k+m} \end{bmatrix}.$$

O coeficiente ergódico de Dobrushin (1.12) é dado por

$$\begin{aligned} \delta(\tilde{P}_{k,k+m}^{y_{0:n}}) &= \delta(\tilde{P}_{k+m-1}^{y_{0:n}}) \\ &= \frac{1}{2} \sup_{i,j} \sum_{\ell=1}^{2} |\tilde{p}_{k+m-1}^{y_{0:n}}(\ell|i) - \tilde{p}_{k+m-1}^{y_{0:n}}(\ell|j)| \\ &= 0. \end{aligned}$$

Assim, a cadeia X^c *é fracamente ergódica e também podemos verificar diretamente que*

$$\begin{aligned} \tilde{\pi}_k^{y_{0:n}}(1) &= P(X_k = 1|Y_{0:n} = y_{0:n}) &= y_k \\ \tilde{\pi}_k^{y_{0:n}}(2) &= P(X_k = 2|Y_{0:n} = y_{0:n}) &= 1 - y_k. \end{aligned}$$

(c) O processo sinal filtrado $X^f = \{X_n^{y_{0:n}}\}_{n \ge 0}$.
Por (2.7) e (2.10) temos

$$\pi_n^{y_{0:n}} = (\pi_n^{y_{0:n}}(1), \pi_n^{y_{0:n}}(2)) = (y_n, 1 - y_n)$$

e X^f é uma cadeia de Markov não-homogênea com

$$P_{n-1}^{y_{0:n}} = \begin{bmatrix} y_n & 1-y_n \\ y_n & 1-y_n \end{bmatrix}.$$

Claramente temos estabilidade do Filtro com

$$\|\pi_{n,\,\nu}^{y_{0:n}} - \pi_{n,\,\nu'}^{y_{0:n}}\| = 0, \ \ \forall\, \nu, \ \ \forall\, \nu'.$$

O próximo exemplo mostra que ainda que X seja ergódica não podemos garantir a estabilidade do Filtro.

Exemplo 2.2. *Assuma que X toma valores em $\mathcal{X} = \{1, 2, 3\}$ com matriz de transição*

$$P = \begin{bmatrix} 1-p_1 & p_1 & 0 \\ 0 & 1-p_2 & p_2 \\ \varepsilon & 0 & 1-\varepsilon \end{bmatrix}. \tag{2.11}$$

onde $0 < p_1 < 1$, $0 < p_2 < 1$ e $0 < \varepsilon < 1$. O processo observável $Y = \{Y_n\}_{n\geq 0}$ satisfaz a relação

$$Y_n = 1_{(X_n=1)} + 1_{(X_n=2)}.$$

Note que temos a estrutura (2.1)

$$Y_n = 1_{(X_n=1)}\xi_n(1) + 1_{(X_n=2)}\xi_n(2) + 1_{(X_n=3)}\xi_n(3),$$

onde $P(\xi_n(1) = 1) = P(\xi_n(2) = 1) = P(\xi_n(3) = 0) = 1$.

É fácil verificar que

$$q(y_n|x_n) = \begin{cases} y_n & ,se \quad x_n = 1, 2 \\ 1-y_n & ,se \quad x_n = 3 \end{cases}$$

e que X é ergódica com distribuição limite

$$\left(\frac{\varepsilon p_2}{\varepsilon(p_1+p_2)+p_1p_2}, \frac{\varepsilon p_1}{\varepsilon(p_1+p_2)+p_1p_2}, \frac{p_1p_2}{\varepsilon(p_1+p_2)+p_1p_2}\right).$$

Também, o processo conjunto $\{(X_n, Y_n)\}_{n\geq 0}$ tem a mesma matriz de transição (2.11) e é ergódico.

(a) Processo sinal condicionado $X^c = \{X_k^{y_{0:n}}\}_{0\le k\le n}$.
É uma cadeia de Markov não-homogênea especificada por

$$\begin{aligned}\tilde{p}_k^{y_{0:n}}(x_{k+1}|x_k) &= P(X_{k+1}=x_{k+1}|X_k=x_k, Y_{0:n}=y_{0:n}) \\ &= \frac{q(y_{k+1}|x_{k+1})p(x_{k+1}|x_k)Q_{k+1}(x_{k+1})}{\sum\limits_{x'_{k+1}} q(y_{k+1}|x'_{k+1})p(x'_{k+1}|x_k)Q_{k+1}(x'_{k+1})},\end{aligned}$$

com,
$Q_n(x) = 1,\ \forall\, x \in \mathcal{X}$,
$Q_k(x) = \sum\limits_{x_{k+1}} q(y_{k+1}|x_{k+1})p(x_{k+1}|x)Q_{k+1}(x_{k+1})$.

Portanto, para $k = n-1$ *e* $y_n = 1$ *temos,*

$$\tilde{P}_{n-1}^{y_{0:n}} = \begin{bmatrix} 1-p_1 & p_1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

e, para $y_n = 0$,

$$\tilde{P}_{n-1}^{y_{0:n}} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Escrevendo, como no exemplo 2.1, em função de y_n *temos*

$$\tilde{P}_{n-1}^{y_{0:n}} = \begin{bmatrix} (1-p_1)y_n & p_1 y_n & 1-y_n \\ 0 & y_n & 1-y_n \\ y_n & 0 & 1-y_n \end{bmatrix}. \tag{2.12}$$

(b) Processo sinal filtrado $X^f = \{X_n^{y_{0:n}}\}_{n\ge 0}$.
Também é uma cadeia de Markov não-homogênea com matrizes de transição $P_{n-1}^{y_{0:n}} = \tilde{P}_{n-1}^{y_{0:n}}$ *dada por (2.12).*
(c) Distribuição de filtragem $\pi_n^{y_{0:n}}$.
$X_n|Y_{0:n} = y_{0:n} \sim \pi_{n,\nu}^{y_{0:n}}$, *onde* $\pi_{0,\nu}^{y_0} \sim \nu$. *Aqui utilizaremos as equações de filtragem recursivas*

$$\pi_{n,\,\nu}^{y_{0:n}}(i) = \frac{g_i(y_n)\sum\limits_{j=1}^{K} P(i|j)\pi_{n-1}^{y_{0:n-1}}(j)}{\sum\limits_{i,j=1}^{K} g_i(y_n)P(i|j)\pi_{n-1}^{y_{0:n-1}}(j)},$$

em que $\pi_0^{y_0} = \nu$.

$$\pi_{n,\,\nu}^{y_{0:n}}(1) = \frac{y_n[(1-p_1)\pi_{n-1,\,\nu}^{y_{0:n-1}}(1) + \varepsilon\pi_{n-1,\,\nu}^{y_{0:n-1}}(3)]}{\alpha_n},$$

$$\pi_{n,\,\nu}^{y_{0:n}}(2) = \frac{y_n[p_1\pi_{n-1,\,\nu}^{y_{0:n-1}}(1) + (1-p_2)\pi_{n-1,\,\nu}^{y_{0:n-1}}(2)]}{\alpha_n},$$

$$\pi_{n,\,\nu}^{y_{0:n}}(3) = \frac{(1-y_n)[p_2\pi_{n-1,\,\nu}^{y_{0:n-1}}(2) + (1-\varepsilon)\pi_{n-1,\,\nu}^{y_{0:n-1}}(3)]}{\alpha_n},$$

onde,

$$\begin{aligned}\alpha_n &= y_n[(1-p_1)\pi_{n-1,\,\nu}^{y_{0:n-1}}(1) + \varepsilon\pi_{n-1,\,\nu}^{y_{0:n-1}}(3) + p_1\pi_{n-1,\,\nu}^{y_{0:n-1}}(1) + (1-p_2)\pi_{n-1,\,\nu}^{y_{0:n-1}}(2)] + \\ &\quad + (1-y_n)[p_2\pi_{n-1,\,\nu}^{y_{0:n-1}}(2) + (1-\varepsilon)\pi_{n-1,\,\nu}^{y_{0:n-1}}(3)].\end{aligned}$$

(Caso 1) se $y_n = 0$

Neste caso temos,

$$\pi_{n,\,\nu}^{y_{0:n}}(1) = \pi_{n,\,\nu}^{y_{0:n}}(2) = 0 \ e \ \pi_{n,\,\nu}^{y_{0:n}}(3) = 1, \ \forall\, \nu \ \Rightarrow \|\pi_{n,\,\nu}^{y_{0:n}} - \pi_{n,\,\nu'}^{y_{0:n}}\| = 0, \ \forall\, n.$$

(Caso 2) se $y_n = 1$

(2*a*) $y_{n-1:n} = 01$
Neste caso temos, $\pi_{n-1,\,\nu}^{y_{0:n-1}}(1) = \pi_{n-1,\,\nu}^{y_{0:n-1}}(2) = 0$ *e* $\pi_{n-1,\,\nu}^{y_{0:n-1}}(3) = 1$. *Também,* $\pi_{n,\,\nu}^{y_{0:n}}(3) = \pi_{n,\,\nu}^{y_{0:n}}(2) = 0$ *e* $\pi_{n,\,\nu}^{y_{0:n}}(1) = 1$. *Logo,*

$$\|\pi_{n,\,\nu}^{y_{0:n}} - \pi_{n,\,\nu'}^{y_{0:n}}\| = 0, \ \forall\, n.$$

(2*b*) $y_{n-1:n} = 11$ *e* $p_1 > p_2$
Aqui, como $y_n = 1 \Rightarrow \pi_{n-1,\,\nu}^{y_{0:n-1}}(3) = \pi_{n,\,\nu}^{y_{0:n}}(3) = 0$ *e* $\pi_{n-1,\,\nu}^{y_{0:n-1}}(1) + \pi_{n-1,\,\nu}^{y_{0:n-1}}(2) = 1 \ \Rightarrow \ \alpha_n = 1 - p_2\pi_{n-1,\,\nu}^{y_{0:n-1}}(2)$.

$$\begin{aligned}\pi_{n,\,\nu}^{y_{0:n}}(1) &= \frac{(1-p_1)\pi_{n-1,\,\nu}^{y_{0:n-1}}(1)}{1-p_2\pi_{n-1,\,\nu}^{y_{0:n-1}}(2)} = \frac{(1-p_1)\pi_{n-1,\,\nu}^{y_{0:n-1}}(1)}{1-p_2+p_2\pi_{n-1,\,\nu}^{y_{0:n-1}}(1)}, \\ \pi_{n,\,\nu}^{y_{0:n}}(2) &= \frac{p_1\pi_{n-1,\,\nu}^{y_{0:n-1}}(1) + (1-p_2)\pi_{n-1,\,\nu}^{y_{0:n-1}}(2)}{1-p_2\pi_{n-1,\,\nu}^{y_{0:n-1}}(2)} = \frac{p_1 + (1-p_1-p_2)\pi_{n-1,\,\nu}^{y_{0:n-1}}(2)}{1-p_2\pi_{n-1,\,\nu}^{y_{0:n-1}}(2)}.\end{aligned}$$

Assim,

$$\begin{aligned}
\pi^{y_{0:n}}_{n,\,\nu}(1) - \pi^{y_{0:n}}_{n,\,\nu'}(1) &= \frac{(1-p_1)\pi^{y_{0:n-1}}_{n-1,\,\nu}(1)}{(1-p_2)+p_2\pi^{y_{0:n-1}}_{n-1,\,\nu}(1)} - \frac{(1-p_1)\pi^{y_{0:n-1}}_{n-1,\,\nu'}(1)}{(1-p_2)+p_2\pi^{y_{0:n-1}}_{n-1,\,\nu'}(1)} \\
&= \frac{(1-p_1)(1-p_2)(\pi^{y_{0:n-1}}_{n-1,\,\nu}(1) - \pi^{y_{0:n-1}}_{n-1,\,\nu'}(1))}{[(1-p_2)+p_2\pi^{y_{0:n-1}}_{n-1,\,\nu}(1)][(1-p_2)+p_2\pi^{y_{0:n-1}}_{n-1,\,\nu'}(1)]} \\
&= \frac{(1-p_1)(1-p_2)(\pi^{y_{0:n-1}}_{n-1,\,\nu}(1) - \pi^{y_{0:n-1}}_{n-1,\,\nu'}(1))}{[1-p_2\pi^{y_{0:n-1}}_{n-1,\,\nu}(2)][1-p_2\pi^{y_{0:n-1}}_{n-1,\,\nu'}(2)]}. \\
\pi^{y_{0:n}}_{n,\,\nu}(2) - \pi^{y_{0:n}}_{n,\,\nu'}(2) &= \frac{(1-p_1)(1-p_2)(\pi^{y_{0:n-1}}_{n-1,\,\nu}(2) - \pi^{y_{0:n-1}}_{n-1,\,\nu'}(2))}{[1-p_2\pi^{y_{0:n-1}}_{n-1,\,\nu}(2)][1-p_2\pi^{y_{0:n-1}}_{n-1,\,\nu'}(2)]}. \ \textit{Também,} \\
\pi^{y_{0:n}}_{n,\,\nu}(2) - \pi^{y_{0:n}}_{n,\,\nu'}(2) &= (1-\pi^{y_{0:n}}_{n,\,\nu}(1)) - (1-\pi^{y_{0:n}}_{n,\,\nu'}(1)) = \pi^{y_{0:n}}_{n,\,\nu'}(1) - \pi^{y_{0:n}}_{n,\,\nu}(1).
\end{aligned}$$

Note que,

$$\frac{(1-p_1)(1-p_2)}{[(1-p_2)+p_2\pi^{y_{0:n-1}}_{n-1,\,\nu}(1)][(1-p_2)+p_2\pi^{y_{0:n-1}}_{n-1,\,\nu'}(1)]} \leq \frac{1-p_1}{1-p_2} = \delta < 1, \ \textit{pois } p_1 > p_2.$$

Logo,

$$|\pi^{y_{0:n}}_{n,\,\nu}(1) - \pi^{y_{0:n}}_{n,\,\nu'}(1)| \leq \delta|\pi^{y_{0:n-1}}_{n-1,\,\nu}(1) - \pi^{y_{0:n-1}}_{n-1,\,\nu'}(1)| \leq \ldots \leq \delta^n|\nu(1)-\nu'(1)| \to 0, \ \textit{quando } n \to \infty,$$

e, com isso, $\|\pi^{y_{0:n}}_{n,\,\nu} - \pi^{y_{0:n}}_{n,\,\nu'}\| \to 0$, *quando* $n \to \infty$.

(2c) *O mesmo não acontece quando* $(p_1 < p_2)$ *e* $y_{0:n} = 011\ldots1$.

Seja $p_2 = 2p_1$ *com* $0 < p_1 < \dfrac{1}{2}$.

Para $\nu(1) = \nu(3) = 0$ *e* $\nu(2) = 1$ *temos* $\pi^{y_{0:n}}_{n,\,\nu}(1) = \pi^{y_{0:n}}_{n,\,\nu}(3) = 0$ *e* $\pi^{y_{0:n}}_{n,\,\nu}(2) = 1$.

Para $\nu'(3) = 0$ *e* $\nu'(1) = \nu'(2) = \dfrac{1}{2}$ $\left(\textit{ ou podemos tomar } p_2 = \dfrac{p_1}{\nu'(2)}\right)$ *obtemos*

$$\pi^{y_{0:1}}_{1,\,\nu'}(1) = \frac{(1-p_1)\nu'(1)}{1-2p_1\nu'(2)} = \frac{1-p_1}{1-p_1}\nu'(1) = \frac{1}{2} \Rightarrow \pi^{y_{0:1}}_{1,\,\nu'}(2) = \frac{1}{2}.$$

Em geral,

$$\pi^{y_{0:n}}_{n,\,\nu'}(1) = \pi^{y_{0:n}}_{n,\,\nu'}(2) = \frac{1}{2} \Rightarrow \|\pi^{y_{0:n}}_{n,\,\nu} - \pi^{y_{0:n}}_{n,\,\nu'}\| = |\pi^{y_{0:n}}_{n,\,\nu}(1) - \pi^{y_{0:n}}_{n,\,\nu'}(1)| + |\pi^{y_{0:n}}_{n,\,\nu}(2) - \pi^{y_{0:n}}_{n,\,\nu'}(2)| = 1, \ \forall\, n.$$

Portanto, $\|\pi^{y_{0:n}}_{n,\,\nu} - \pi^{y_{0:n}}_{n,\,\nu'}\| \nrightarrow 0.$
Com isso vemos que, neste exemplo, para $p_1 > p_2$ *o Filtro é estável enquanto que, para* $p_1 < p_2$ *o Filtro não é estável.*

2.3 Cadeias de Markov Associadas

Retomamos o Filtro geral apresentado no Capítulo 1. Assim, o sinal $\{X_n\}$ tem as transições $X_n \longrightarrow X_{n+1}$ determinadas pela densidade de transição $p(x_{k+1}|x_k)d\mu(x_{k+1})$, a observação Y_n condicionada ao X_n é determinada por $q(y_n|x_n)d\lambda(y_n)$ além de satisfazer a estrutura HMM dada por (1.3). Derivamos abaixo as densidades de transição do processo conjunto estendido $\{(X_n, Y_{0:n})\}_{n\geq 0}$ e do sinal condicionado $X^c = \{X^{y_{0:n}}_k\}_{0\leq k\leq n}$, a densidade de filtragem $\pi^{y_{0:n}}_n$ e, no caso do modelo de equação de estados, a densidade de transição do processo do sinal filtrado $X^f = \{X^{y_{0:n}}_n\}_{n\geq 0}$.

Proposição 2.1. *O processo* $\{(X_n, Y_{0:n})\}_{n\geq 0}$ *é uma cadeia de Markov e possui densidade de transição dada por*

$$g(x_n, y_{0:n}|x_{n-1}, y_{0:n-1}) = q(y_n|x_n)p(x_n|x_{n-1}). \tag{2.13}$$

O processo X^c *também é uma cadeia de Markov e sua densidade de transição, para* $0 \leq k \leq n$*, é dada por*

$$\tilde{p}^{y_{0:n}}_{k-1}(x_k|x_{k-1}) = \frac{q(y_k|x_k)p(x_k|x_{k-1})Q_k(x_k)}{\int_{\mathcal{X}} q(y_k|x'_k)p(x'_k|x_{k-1})Q_k(x'_k)d\mu(x'_k)}, \tag{2.14}$$

onde $\{Q_k\}$ *satisfaz*

$$\begin{aligned} Q_k(x_k) &= \int_{\mathcal{X}} q(y_{k+1}|x_{k+1})p(x_{k+1}|x_k)Q_{k+1}(x_{k+1})d\mu(x_{k+1}) \\ Q_n(x_n) &= 1. \end{aligned} \tag{2.15}$$

Demonstração: Primeiro, note que a existência das densidades (2.13) e (2.14) é garantido pela existência de densidade de (X_n, Y_n),

$$g(x_n, y_n) = q(y_n|x_n)\nu_n(x_n),$$

onde $\nu_n(x_n) = \displaystyle\int_{\mathcal{X}^n} \prod_{j=1}^{n} p(x_j|x_{j-1})\nu(x_0)d\mu(x_{0:n-1})$, sendo ν alguma distribuição de X_0. A estrutura HMM garante que ambas são cadeias de Markov.

(a) Para $\{(X_n, Y_{0:n})\}$ temos

$$g(x_n, y_{0:n}|x_{n-1}, y_{0:n-1}) = \frac{g(x_{n-1:n}, y_{0:n})}{g(x_{n-1}, y_{0:n-1})}.$$

Para qualquer distribuição inicial $\nu(x_0)d\mu(x_0)$ de X_0 e $P(Y_0 = 0) = 1$ temos por (1.3) e o fato de $\{X_n\}$ ser uma cadeia de Markov

$$\begin{aligned}
g(x_{0:n}, y_{0:n}) &= \left[\prod_{j=0}^{n} q(y_j|x_j)p(x_j|x_{j-1})\right]\nu(x_0)\delta_0(y_0) \\
&= q(y_n|x_n)p(x_n|x_{n-1})\left[\prod_{j=0}^{n-1} q(y_j|x_j)p(x_j|x_{j-1})\right]\nu(x_0)\delta_0(y_0) \\
&= q(y_n|x_n)p(x_n|x_{n-1})g(x_{0:n-1}, y_{0:n-1}).
\end{aligned}$$

Assim,

$$\begin{aligned}
g(x_{n-1:n}, y_{0:n}) &= q(y_n|x_n)p(x_n|x_{n-1})\int_{\mathcal{X}^{n-1}} g(x_{0:n-1}, y_{0:n-1})d\mu(x_{0:n-2}) \\
&= q(y_n|x_n)p(x_n|x_{n-1})g(x_{n-1}, y_{0:n-1})
\end{aligned}$$

e temos (2.13).

(b) Para X^c temos para a transição $X_{k-1}^{y_{0:n}} \longrightarrow X_k^{y_{0:n}}$,

$$\begin{aligned}
\tilde{p}_{k-1}^{y_{0:n}}(x_k|x_{k-1}) &= \frac{g(x_{k-1:k}|y_{0:n})}{g(x_{k-1}|y_{0:n})} \\
&= \frac{g(x_{k-1:k}, y_{0:n})}{g(x_{k-1}, y_{0:n})}.
\end{aligned}$$

Note que para $0 < k \leq n$ vale

$$g(x_k, y_{0:n}) = g(x_k, y_{0:k})\int_{\mathcal{X}^{n-k}} \prod_{j=k+1}^{n} q(y_j|x_j)p(x_j|x_{j-1})d\mu(x_{k+1:n}).$$

Vamos definir

$$Q_k(x_k) = \frac{g(x_k, y_{0:n})}{g(x_k, y_{0:k})} = \int_{\mathcal{X}^{n-k}} \prod_{j=k+1}^{n} q(y_j|x_j)p(x_j|x_{j-1})d\mu(x_{k+1:n}),$$

onde, para $k = n$, definiremos $Q_n(x_n) = 1$. Segue diretamente desta definição que

$$Q_k(x_k) = \int_{\mathcal{X}} Q_{k+1}(x_{k+1})q(y_{k+1}|x_{k+1})p(x_{k+1}|x_k)d\mu(x_{k+1}) \tag{2.16}$$

e

$$g(x_k, y_{0:n}) = Q_k(x_k)g(x_k, y_{0:k}).$$

Também,

$$\begin{aligned} g(x_{k:k+1}, y_{0:n}) &= q(y_{k+1}|x_{k+1})p(x_{k+1}|x_k)g(x_k, y_{0:k})Q_{k+1}(x_{k+1}) \\ &= g(x_{k:k+1}, y_{0:k+1})Q_{k+1}(x_{k+1}) \end{aligned}$$

onde, para $k = n-1$, adotamos

$$\int_{\mathcal{X}^2} \prod_{j=n+1}^{n} q(y_j|x_j)p(x_j|x_{j-1})d\mu(x_{n+1:n}) := Q_n(x_n) = 1.$$

E (2.14) segue.

■

A cadeia $\{(X_n, Y_{0:n})\}$ possui espaços de estados variantes $\mathcal{X} \times \mathcal{Y}^{n+1}$ e, claramente, a cadeia X^c é uma cadeia não-homogênea pois, para $k \neq k'$, temos $Q_k \neq Q'_k$. A distribuição da cadeia condicionada $\tilde{\pi}_k^{y_{0:n}}$ pode ser obtida observando que para $0 \leq k \leq n$,

$$\begin{aligned} g(y_{0:n}) &= \int_{\mathcal{X}} \left[\int_{\mathcal{X}^{n-k}} \prod_{j=k+1}^{n} q(y_j|x_j)p(x_j|x_{j-1})d\mu(x_{k+1:n}) \right] g(x_k, y_{0:k})d\mu(x_k) \\ &= \int_{\mathcal{X}} Q_k(x_k)g(x_k, y_{0:k})d\mu(x_k). \end{aligned} \tag{2.17}$$

Assim temos

$$\begin{aligned} \tilde{\pi}_k^{y_{0:n}}(x_k) &= \frac{g(x_k, y_{0:n})}{g(y_{0:n})} \\ &= \frac{Q_k(x_k)g(x_k, y_{0:k})}{\int_{\mathcal{X}} Q_k(x'_k)g(x'_k, y_{0:k})d\mu(x'_k)} \\ &= \frac{Q_k(x_k)\pi_k^{y_{0:k}}(x_k)}{\int_{\mathcal{X}} Q_k(x'_k)\pi_k^{y_{0:k}}(x'_k)d\mu(x'_k)}. \end{aligned} \tag{2.18}$$

Também, a distribuição de filtragem $\pi_n^{y_{0:n}}$ pode ser derivada da cadeia $\{(X_n, Y_{0:n})\}$, pois, para $\pi_{n-1}^{y_{0:n-1}}(x_{n-1}) = \dfrac{g(x_{n-1}, y_{0:n-1})}{g(y_{0:n-1})}$, temos

$$g(x_n, y_{0:n}) = g(y_{0:n-1}) \int_{\mathcal{X}} q(y_n|x_n)p(x_n|x_{n-1})\pi_{n-1}^{y_{0:n-1}}(x_{n-1})d\mu(x_{n-1})$$

e

$$
\begin{aligned}
\pi_n^{y_{0:n}}(x_n) &= \frac{g(x_n, y_{0:n})}{g(y_{0:n})} = \frac{g(y_{0:n-1})}{g(y_{0:n})} \int_{\mathcal{X}} q(y_n|x_n) p(x_n|x_{n-1}) \pi_{n-1}^{y_{0:n-1}}(x_{n-1}) d\mu(x_{n-1}) \\
&= \frac{g(y_{0:n-1})}{g(y_{0:n})} q(y_n|x_n) \int_{\mathcal{X}} p(x_n|x_{n-1}) \pi_{n-1}^{y_{0:n-1}}(x_{n-1}) d\mu(x_{n-1}),
\end{aligned}
$$

o que, por (2.17), resulta na equação da filtragem

$$
\pi_n^{y_{0:n}}(x_n) = \frac{q(y_n|x_n) \int_{\mathcal{X}} p(x_n|x_{n-1}) \pi_{n-1}^{y_{0:n-1}} d\mu(x_{n-1})}{\int_{\mathcal{X}^2} q(y_n|x'_n) p(x'_n|x'_{n-1}) \pi_{n-1}^{y_{0:n-1}}(x'_{n-1}) d\mu(x'_{n-1:n})}. \tag{2.19}
$$

Proposição 2.2. *A densidade de filtragem $\pi_n^{y_{0:n}}(.)$ é dada pela equação de filtragem (2.19) onde $\pi_0^{y_0}(x_0) = \nu(0)$, sendo ν a distribuição inicial de X_0. A densidade condicionada $\tilde{\pi}_k^{y_{0:n}}$, para $0 \leq k \leq n$ é dada por (2.18) com $\{Q_k\}$ definidos por (2.16). Mais ainda, $\pi_n^{y_{0:n}} = \tilde{\pi}_n^{y_{0:n}}(.)$.*

Quanto ao processo do sinal filtrado X^f as suas densidades de transição são de difícil cálculo, pois $X_{n-1}^{y_{0:n-1}} \sim X_{n-1}|Y_{0:n-1} = y_{0:n-1}$ e $X_n^{y_{0:n}} \sim X_n|Y_{0:n} = y_{0:n}$. No entanto, para o modelo de Equações de estados introduzido no Capítulo 1 podemos explicitar as densidades de transição. Temos

$$
Y_n = h_n(X_n) + \eta_n \tag{2.20}
$$

onde o ruído $\{\eta_n\}$ consiste de variáveis aleatórias iid e $\{\eta_n\}$ é independente de $\{X_n\}$.

Neste caso, usando os mesmos tipos de argumentos da seção 2.1, temos

Proposição 2.3. *Assumindo a estrutura de Equações de Estados (2.20) a cadeia de Markov $X^f = \{X_n^{y_{0:n}}\}_{n \geq 0}$ possui densidades de transição dadas por*

$$
p_{n-1}^{y_{0:n}}(x_n|x_{n-1}) = \frac{q(y_n|x_n) p(x_n|x_{n-1})}{\int_{\mathcal{X}} q(y_n|x'_n) p(x_n|x'_{n-1}) d\mu(x'_n)}. \tag{2.21}
$$

Capítulo 3

Estabilidade do Filtro

Em 1957, em [3], foi conjecturado que o filtro seria independente da distribuição inicial se o processo sinal X fosse ergódico (X com espaço de estados finito). No modelo estudado em [3] as observações são dadas por $Y_k = f(X_k)$, ou seja, sem dependência de ruído. Em 1975, em [19], foi apresentado um exemplo de cadeia ergódica em que o Filtro depende da distribuição inicial mostrando que a conjectura em [3] não é verdadeira. Ainda em [19], foi demonstrado que o Filtro é independente da distribuição inicial, sob a hipótese de ergodicidade do processo sinal mas com hipóteses adicionais sobre a função determinística que relacionava o processo sinal com as observações e sobre a matriz de transição de X (X com espaço de estados finito).

Nas últimas décadas resultados de independência da distribuição inicial do Filtro foram obtidos por vários autores sob variadas condições que incluem ou não a ergodicidade do processo sinal e o espaço de estados do sinal sendo enumerável ou não. Atualmente, exibir condições necessárias e suficientes para assegurar a independência do Filtro com relação a distribuição inicial permanece um problema em aberto. Dessa forma há um esforço em se conseguir exigências mínimas sobre o modelo para que o Filtro seja independente da distribuição inicial. As exigências comumente impostas ao modelo dizem respeito à densidade de transição do processo sinal X (denotada por $p(.|.)$), à densidade condicional de Y_k dado X_k (denotada por $q(.|.)$), ao núcleo de transição do processo sinal (denotado por $P(.\,,.)$) e/ou às distribuições iniciais ν e ν'.

Em [1] foi proposto o índice de estabilidade

$$\gamma = \limsup_n \frac{1}{n} \log \|\pi_{n,\,\nu} - \pi_{n,\,\nu'}\|$$

para estudar a estabilidade do Filtro. Caso o limite seja negativo, se existir, implicará a estabilidade do Filtro. Este índice foi estudado em [1] e [2] utilizando técnicas de expoentes de Lyapunov e distância projetiva de Hilbert, respectivamente. Em [1] foi introduzida e utilizada a condição tipo "mixing"

$$0 < \lambda_* \leq p(y|x) \leq \lambda^* < \infty \tag{3.1}$$

no estudo de γ. Esta condição também foi utilizada em [9] para estudar a estabilidade do Filtro através de técnicas de semi-grupos. Em [7] o coeficiente de estabilidade γ também foi utilizado para estudar a estabilidade do Filtro.

Outra condição relacionada a (3.1) que aparece em [5] é

$$\varepsilon\varphi_k(A) \leq \tilde{P}_k^{y_{0:k}}(x, A) \leq \frac{1}{\varepsilon}\varphi_k(A), \tag{3.2}$$

onde $\{\varphi_k\}$ é uma sequência de medidas σ-finitas. Em [5] a estabilidade do Filtro é obtida via cópulas e técnicas de cadeia dividida (veja [22] e [23] para teoria de cadeias divididas).

Em nosso trabalho evitamos as técnicas acima descritas embora utilizamos as condições do tipo (3.1) e (3.2). No caso da condição (3.1) nossa técnica evita impor condições sobre as condições iniciais ou sobre $q(.|.)$. A exigência básica que utilizamos é a ergodicidade fraca da cadeia condicionada. O Teorema 3.1 estende os resultados em [5]. Em [7] a condição utilizada sobre a densidade de transição do processo sinal é mais fraca que (3.1) mas existem condições adicionais sobre as distribuições iniciais do Filtro e é exigida a existência de medida invariante para o núcleo de transição do processo sinal. Em nosso trabalho evitamos tais hipóteses.

3.1 Resultados de Estabilidade

Nesta seção enunciamos os resultados quanto a estabilidade do Filtro. Os resultados se baseiam, essencialmente, na análise do comportamento ergódico (fraco e forte) dos processos X^c e X^f. As provas serão feitas na seção 3.2.

Da seção 2.3 derivamos os núcleos de transição. Para o processo condicionado X^c temos para $0 < k \leq n$ e $A \in \mathbb{X}$

$$\tilde{P}_{k-1}^{y_{0:n}}(x_{k-1}, A) = \int_A \tilde{p}_{k-1}^{y_{0:n}}(x_k|x_{k-1})d\mu(x_k), \tag{3.3}$$

onde $\tilde{p}_{k-1}^{y_{0:n}}$ é dado por (2.14) e (2.15).

O processo X^f com $\{Y_n\}$ obedecendo as Equações de Estados (2.20) tem núcleos de transição dados por

$$P_{n-1}^{y_{0:n}}(x_{n-1}, A) = \int_A p_{n-1}^{y_{0:n}}(x_n|x_{n-1})d\mu(x_n), \tag{3.4}$$

onde $p_{n-1}^{y_{0:n}}$ é dado por (2.21).

Com abuso de notação, escreveremos para $A \in \mathbb{X}$

$$\pi_{n,\,\nu}^{y_{0:n}}(A) = \int_A \pi_{n,\,\nu}^{y_{0:n}}(x_n)d\mu(x_n)$$

e

$$\tilde{\pi}_{k,\,\nu}^{y_{0:n}}(A) = \int_A \tilde{\pi}_{k,\,\nu}^{y_{0:n}}(x_k)d\mu(x_k),$$

onde ν é uma distribuição inicial de X_0.

Lema 3.1. *Assuma que o coeficiente ergódico da cadeia $\{\tilde{P}_k^{y_{0:n}}\}$ satisfaz*

$$\lim_{n\to\infty} \delta(\tilde{P}_{m,m+k_n}^{y_{0:n}}) = 0, \; m = 0, 1, \ldots, \tag{3.5}$$

onde $k_n \uparrow \infty$ com $m + k_n \leq n$. Então a cadeia X^c é fracamente ergódica e o Filtro é estável, isto é, para quaisquer distribuições iniciais ν e ν'

$$\lim_{n\to\infty} \|\pi_{n,\,\nu}^{y_{0:n}} - \pi_{n,\,\nu'}^{y_{0:n}}\| = 0. \tag{3.6}$$

Mais ainda, temos

$$\|\pi_{n,\,\nu}^{y_{0:n}} - \pi_{n,\,\nu'}^{y_{0:n}}\| \leq 2\delta(\tilde{P}_{0,n-1}^{y_{0:n}}) \leq 2\prod_{j=0}^{n-1} \delta(\tilde{P}_j^{y_{0:n}}). \tag{3.7}$$

Exemplo 3.1. *O seguinte exemplo ilustra o cálculo da taxa de convergência (3.7). Considere o modelo $\{(X_n, Y_n)\}_{n\geq 2}$ onde $\{X_n\}$ tem espaço de estados $\mathcal{X} = \{1, 2, 3\}$ e matriz* *de transição*

$$P = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

e o processo $\{Y_n\}$ *com valores em* $\mathcal{Y} = \{0, 1\}$ *é dado por*

$$Y_n = \sum_{i,j=1}^{3} 1_{(X_{n-1}=i, X_n=j)} \xi_n(i,j)$$

com $\xi_n(i,j)$*´s independentes tomando valores* 0 *ou* 1*. Assuma que*

$$P(\xi_n(1,j) = 1) = \begin{cases} \frac{1}{n} & , j = 1 \\ 1 - \frac{1}{n} & , j = 2 \\ 0 & , j = 3 \end{cases},$$

$$P(\xi_n(2,j) = 1) = \begin{cases} \frac{1}{n} & , j = 1 \\ 0 & , j = 2 \\ 1 - \frac{1}{n} & , j = 3 \end{cases}$$

e

$$P(\xi_n(3,j) = 1) = \begin{cases} \frac{1}{n} & , j = 1 \\ \frac{1}{2}\left(1 - \frac{1}{n}\right) & , j = 2 \\ \frac{1}{2}\left(1 - \frac{1}{n}\right) & , j = 3 \end{cases}.$$

Dado $y_n = 1$*, temos*

$$1_{(Y_n=1)} = \sum_{i,j} 1_{(X_{n-1}=i, X_n=j)} 1_{(\xi_n(i,j)=1)}.$$

Assim, se $X_{n-1}^{y_{2:n}} = i$ *com* $y_n = 1$ *temos* $X_n^{y_{2:n}} = j$ *se* $\xi_n(i,j) = 1$*. Temos*

$$\begin{aligned} \tilde{p}_{n-1}^{y_{2:n}}(j|1) &= \frac{p(j|1)P(\xi_n(1,j)=1)}{\sum_{\ell} p(\ell|1)P(\xi_n(1,\ell)=1)} \\ &= P(\xi_n(1,j)=1), \end{aligned}$$

pois $p(j|1) = \dfrac{1}{3}$, $\forall\ j$ *e* $\sum_{\ell} P(\xi_n(1,\ell) = 1) = 1$. *Similarmente, temos* $\tilde{p}^{y_{2:n}}_{n-1}(j|2) = P(\xi_n(2,j) = 1)$ *e* $\tilde{p}^{y_{2:n}}_{n-1}(j|3) = P(\xi_n(3,j) = 1)$. *Segue que para* $y_n = 1$

$$\tilde{P}^{y_{2:n}}_{n-1} = \begin{bmatrix} \dfrac{1}{n} & 1 - \dfrac{1}{n} & 0 \\[2ex] \dfrac{1}{n} & 0 & 1 - \dfrac{1}{n} \\[2ex] \dfrac{1}{n} & \dfrac{1}{2}\left(1 - \dfrac{1}{n}\right) & \dfrac{1}{2}\left(1 - \dfrac{1}{n}\right) \end{bmatrix}$$

e

$$\begin{aligned} \delta(\tilde{P}^{y_{2:n}}_{n-1}) &= \frac{1}{2}\sup_{i,j}\sum_k |\tilde{p}^{y_{2:n}}_{n-1}(k|i) - \tilde{p}^{y_{2:n}}_{n-1}(k|j)| \\ &= \left(1 - \frac{1}{n}\right), \quad y_n = 1. \end{aligned}$$

Similarmente, se $y_n = 0$ *então* $X^{y_{2:n}}_{n-1} = i \longrightarrow X^{y_{2:n}}_{n} = j$ *se* $\xi_n(i,j) = 0$, *e usando o fato que* $p(j|i) = \dfrac{1}{3}$, $\forall\ i$, $\forall\ j$, *temos*

$$\tilde{p}^{y_{2:n}}_{n-1}(j|i) = \frac{P(\xi_n(i,j) = 0)}{\sum_{\ell} P(\xi_n(i,\ell) = 0)}.$$

Note que, para todo i, *temos* $\sum_{\ell} P(\xi_n(i,\ell) = 0) = 2$. *Segue que, para* $y_n = 0$, *temos*

$$\tilde{P}^{y_{2:n}}_{n-1} = \begin{bmatrix} \dfrac{1}{2}\left(1 - \dfrac{1}{n}\right) & \dfrac{1}{2n} & \dfrac{1}{2} \\[2ex] \dfrac{1}{2}\left(1 - \dfrac{1}{n}\right) & \dfrac{1}{2} & \dfrac{1}{2n} \\[2ex] \dfrac{1}{2}\left(1 - \dfrac{1}{n}\right) & \dfrac{1}{4}\left(1 + \dfrac{1}{n}\right) & \dfrac{1}{4}\left(1 + \dfrac{1}{n}\right) \end{bmatrix}$$

e

$$\delta(\tilde{P}^{y_{2:n}}_{n-1}) = \frac{1}{2}\left(\frac{n-1}{n}\right) \leq \left(1 - \frac{1}{n}\right), \quad y_n = 0.$$

Como neste exemplo temos $\tilde{P}_k^{y_{2:n}} = \tilde{P}_k^{y_{2:k+1}}$ *para* $k < n$, *segue por (3.7)*

$$\|\pi_{n,\,\nu}^{y_{2:n}} - \pi_{n,\,\nu'}^{y_{2:n}}\| \leq 2\prod_{k=2}^{n-1}\delta(\tilde{P}_k^{y_{2:n}}) \leq 2\prod_{k=2}^{n-1}\left(1 - \frac{1}{k+1}\right).$$

Como $\sum_k \frac{1}{k} = \infty$, *não só temos a taxa de convergência como também a estabilidade do Filtro* $\forall\ y_{2:n}$.

Lema 3.2. *Assuma que para algum núcleo de transição constante* P_∞ *temos*

$$\lim_{n\to\infty}\|\tilde{P}_{m,m+k_n}^{y_{0:n}} - P_\infty\| = 0,\ m = 0, 1, \ldots\ . \tag{3.8}$$

Então X^c *é fortemente ergódica e existe uma probabilidade* π_∞ *em* $(\mathcal{X}, \mathbb{X})$ *tal que*

$$\lim_{n\to\infty}\|\pi_{n,\,\nu}^{y_{0:n}} - \pi_\infty\| = 0,\ \forall\ \nu. \tag{3.9}$$

Teorema 3.1. *Assuma que existam uma sequência de constantes* $\{\alpha_k\}$, $\alpha_k > 0$, $\forall\ k$ *e uma sequência de medidas* σ*-finitas* $\{\varphi_k\}$ *em* $(\mathcal{X}, \mathbb{X})$ *tais que*

$$\sum_{k'}\alpha_{k'}^3 = \infty,\ \forall\ \{k'\} \subset \{k\} \tag{3.10}$$

e para $0 \leq k < n$

$$\alpha_k\varphi_k(A) \leq \tilde{P}_k^{y_{0:n}}(x, A) \leq \frac{1}{\alpha_k}\varphi_k(A),\ \forall\ x \in \mathcal{X},\ \forall\ A \in \mathbb{X}. \tag{3.11}$$

Então X^c *é fracamente ergódica e o Filtro é estável, isto é, vale (3.6)*

Corolário 3.1. *Sob as condições do Teorema 3.1, se* $\alpha_k = \alpha$, $\forall\ k$, *então temos que*

$$\|\pi_{n,\,\nu}^{y_{0:n}} - \pi_{n,\,\nu'}^{y_{0:n}}\| \leq 2\left(1 - \frac{\alpha^3}{2}\right)^n,\ \forall\ n. \tag{3.12}$$

O resultado acima representa extensão dos resultados de [5]. Observe que (3.11) pode ser verificado por condições tipo positividade e limitação.

Condição 1

Assuma que a densidade de transição do sinal X satisfaz

$$0 < \lambda_* \leq p(v|u) \leq \lambda^* < \infty,\ \forall\ u, v \in \mathcal{X}. \tag{3.13}$$

Corolário 3.2. *Sob a Condição 1 temos a estabilidade do Filtro com a validade de (3.12).*

Teorema 3.2. *Sejam $\{P^{y_{0:n}}_{n-1}\}_{n\geq 0}$ os núcleos de transição de $X^f = \{X^{y_{0:n}}_n\}_{n\geq 0}$ e assuma que*

$$\limsup_{n\to\infty} \sum_{k=0}^{n} \|P^{y_{0:k}}_{k-1} - \tilde{P}^{y_{0:n}}_{k-1}\| < \infty. \tag{3.14}$$

Então as cadeias X^f e X^c possuem o mesmo comportamento ergódico, isto é, X^f é fracamente (fortemente) ergódica se, e somente se, X^c for fracamente (fortemente) ergódica.

Note que para o modelo de Equações de Estados (2.20) temos $P^{y_{0:k}}_{k-1} = \tilde{P}^{y_{0:n}}_{k-1}$. Assim, (3.14) é trivialmente satisfeita e podemos estudar a estabilidade do Filtro analisando a ergodicidade de X^f.

Condição 2

Para o modelo (2.20) assuma que existam constantes $\rho > 0$, $p_* > 0$ e $p^* < \infty$ tais que para todo $y_n \in \mathcal{Y}$ existe $C_{y_n} \subset \mathcal{X}$ satisfazendo

$$p_* \leq p(v|u) \leq p^*,\ \forall\, v \in C_{y_n} \text{ e } \forall\, u \in \mathcal{X} \tag{3.15}$$

e

$$P^{y_{0:n}}_{n-1}(x_{n-1}, C_{y_n}) \geq \rho,\ \forall\, x_{n-1} \in \mathcal{X}. \tag{3.16}$$

Teorema 3.3. *Se a Condição 2 for satisfeita então as cadeias X^f e X^c são fracamente ergódicas e o Filtro é estável.*

Observações 3.1

(a) Note que, se temos Condição 1 então, tomando-se $C_{y_n} = \mathcal{X}$, $\forall\, y_n \in \mathcal{Y}$ e $\rho = 1$, temos a Condição 2.

(b) Pela prova do Teorema 3.3 é fácil concluir que podemos relaxar (3.16), substituindo por: existe $\{\rho_n\}$, $\rho_n > 0$, $\forall\, n$, satisfazendo $\sum_{n'} \rho_{n'} = \infty$, $\forall\, \{n'\} \subset \{n\}$ e

$$P^{y_{0:n}}_{n-1}(x, C_{y_n}) \geq \rho_n,\ \forall\, x \in \mathcal{X}.$$

(c) Se $\mathcal{X}$ e $\mathcal{Y}$ são finitos podemos substituir (3.15) e (3.16) por: para todo y_n existe x^*_n tal que

$$q(y_n|x^*_n)p(x^*_n|x_{n-1}) > 0,\ \forall\, x_{n-1} \in \mathcal{X}. \tag{3.17}$$

Corolário 3.3. *Sob o modelo (2.20), se $\mathcal{X}$ e $\mathcal{Y}$ são espaços finitos então (3.17) garante a estabilidade do Filtro.*

Exemplo 3.2. (a) *Para o exemplo 2.1 temos a Condição 2 satisfeita. Se $y_n = 0$ tomamos $x_n^* = 2$ e temos*

$$q(y_n|x_n^*)p(x_n^*|x_{n-1}) = \begin{cases} p_1 & se\ x_{n-1} = 1 \\ 1-\varepsilon & se\ x_{n-1} = 2. \end{cases}$$

Similarmente, se $y_n = 1$ tomamos $x_n^* = 1$.
(b) *Para o exemplo 2.2 não temos a Condição 2 satisfeita, pois, não existe x^* tal que $p(x^*|x) > 0,\ \forall\, x \in \mathcal{X}$.*

O exemplo seguinte exibe uma situação em que vale a Condição 2 mas não vale a Condição 1. Note que neste exemplo temos que o ruído é uma sequência de variáveis aleatórias iid normais truncadas.

Exemplo 3.3. *Vamos considerar a cadeia oculta $X = \{X_n\}$ com espaço de estados $\mathcal{X} = \{1, 2, 3\}$ e matriz de transição*

$$P_\delta = \begin{bmatrix} 1-p_1 & p_1 & 0 \\ 0 & 1-p_2 & p_2 \\ \varepsilon-\delta & \delta & 1-\varepsilon \end{bmatrix}.$$

Observe que a matriz P_δ é uma perturbação, por $0 < \delta < \varepsilon$, da matriz P do exemplo 2.2. Aqui assumimos também $0 < p_1 < 1$, $0 < p_2 < 1$ e $0 < p_1 < \varepsilon$. Vamos considerar as observações $Y = \{Y_n\}$ dadas por

$$Y_n = h(X_n) + \eta_n',$$

onde $\eta_n' = \frac{1}{c}\eta_n 1_{([-d,d])}$ com $\eta_n \sim N(0, \sigma^2)$, $c = \int_{-d}^{d} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}} dy$, h uma função conhecida e d tal que $\max\{|h(i)|; i = 1, 2, 3\} < d$. O espaço de estados das observações Y é $\mathcal{Y} = \cap_{i=1}^{3}[-d + h(i), d + h(i)]$.

Nestas condições, $Y_n|X_n = i \sim h(i) + \eta_n'$ e

$$q(y|i) = \frac{1}{c_i\sqrt{2\pi}} e^{-\frac{(y-h(i))^2}{2\sigma^2}},\quad y \in \mathcal{Y},$$

onde $c_i = \displaystyle\int_{\mathcal{Y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-h(i))^2}{2\sigma^2}} dy.$

Claramente a Condição 1 não é satisfeita. Entretanto, para qualquer $y \in \mathcal{Y}$ *existe* $C_y = \{2\} \subset \mathcal{X}$ *tal que*

$$p_* \leq p(2|i) \leq p^*, \ \ \forall\, i \in \mathcal{X},$$

onde $p_* = \min\{p_1, 1-p_2, \delta\} > 0$ *e* $p^* = \max\{p_1, 1-p_2, \delta\}$.

Sejam $a_* = \min\{c_1, c_2, c_3\}$ *e* $e_* = \min\left\{ e^{-\frac{(y-h(2))^2}{2\sigma^2}} ; y \in \mathcal{Y} \right\}$. *Note que,*

$$\begin{aligned} P^{y_{0:n}}_{n-1}(i,2) &= \frac{q(y|2)p(2|i)}{\sum_{j=1}^{3} q(y|j)p(j|i)} \\ &\geq a_* p_* \frac{1}{c_2} e^{-\frac{(y-h(2))^2}{2\sigma^2}} \\ &\geq \frac{a_*}{c_2} p_* e_* = \rho > 0. \end{aligned}$$

Portanto, vale a Condição 2, e temos a estabilidade do Filtro.

3.2 Demonstrações dos Resultados

Lema 3.1

Demonstração: Note que, de (3.5) temos: dado $\varepsilon > 0$ existe $k_n^m(\varepsilon)$ tal que para $k_n \geq k_n^m(\varepsilon)$

$$\delta(\tilde{P}^{y_{0:n}}_{m,m+k_n}) < \varepsilon,$$

e a ergodicidade fraca segue da Definição 1.2. Sejam ν e ν' duas distribuições iniciais quaisquer, temos pela Proposição 1.1 (b)

$$\begin{aligned} \|\nu \tilde{P}^{y_{0:n}}_{0,n-1} - \nu' \tilde{P}^{y_{0:n}}_{0,n-1}\| &= \|(\nu - \nu')\tilde{P}^{y_{0:n}}_{0,n-1}\| \\ &\leq \|\nu - \nu'\| \delta(\tilde{P}^{y_{0:n}}_{0,n-1}) \\ &\leq 2\delta(\tilde{P}^{y_{0:n}}_{0,n-1}) \xrightarrow{n} 0. \end{aligned} \tag{3.18}$$

Por outro lado, temos

$$\begin{aligned}
\pi_n^{y_{0:n}}(A) &= \tilde{\pi}_n^{y_{0:n}}(A) = P(X_n^{y_{0:n}} \in A) = \int_{\mathcal{X}} P(X_n^{y_{0:n}} \in A | X_{n-1}^{y_{0:n}} = x_{n-1}) P(X_{n-1}^{y_{0:n}} \in dx_{n-1}) \\
&= \int_{\mathcal{X}} \int_{\mathcal{X}} \tilde{P}_{n-1}^{y_{0:n}}(x_{n-1}, A) P(X_{n-1}^{y_{0:n}} \in dx_{n-1} | X_{n-2}^{y_{0:n}} = x_{n-2}) P(X_{n-2}^{y_{0:n}} \in dx_{n-2}) \\
&= \int_{\mathcal{X}} \int_{\mathcal{X}} \tilde{P}_{n-1}^{y_{0:n}}(x_{n-1}, A) \tilde{P}_{n-2}^{y_{0:n}}(x_{n-2}, dx_{n-1}) P(X_{n-2}^{y_{0:n}} \in dx_{n-2}) = \ldots = \\
&= \int_{\mathcal{X}} \ldots \int_{\mathcal{X}} P(X_0^{y_{0:n}} \in dx_0) \tilde{P}_0^{y_{0:n}}(x_0, dx_1) \ldots \tilde{P}_{n-1}^{y_{0:n}}(x_{n-1}, A) \\
&= \int_{\mathcal{X}} P(X_0^{y_{0:n}} \in dx_0) \tilde{P}_{0,n-1}^{y_{0:n}}(x_0, A).
\end{aligned}$$

Assim, para qualquer distribuição inicial ν, isto é, $X_0^{y_{0:n}} \sim \nu$ temos

$$\pi_{n,\,\nu}^{y_{0:n}}(A) = \int_{\mathcal{X}} \nu(dx_0) \tilde{P}_{0,n-1}^{y_{0:n}}(x_0, A). \tag{3.19}$$

E, por (3.18)

$$\lim_{n\to\infty} \|\pi_{n,\,\nu}^{y_{0:n}} - \pi_{n,\,\nu'}^{y_{0:n}}\| = 0.$$

Também, de (3.18) e Proposição 1.1 (d),

$$\|\pi_{n,\,\nu}^{y_{0:n}} - \pi_{n,\,\nu'}^{y_{0:n}}\| \leq 2 \prod_{j=0}^{n-1} \delta(\tilde{P}_j^{y_{0:n}}).$$

■

Lema 3.2

Demonstração: Por (3.8), dado $\varepsilon > 0$ existe $k_n^m(\varepsilon)$ tal que para $k_n \geq k_n^m(\varepsilon)$

$$\|\tilde{P}_{m,m+k_n}^{y_{0:n}} - P_\infty\| < \varepsilon.$$

Pela Definição 1.3 temos a ergodicidade forte. Em particular, temos

$$\lim_{n\to\infty} \|\tilde{P}_{0,n-1}^{y_{0:n}} - P_\infty\| = 0. \tag{3.20}$$

Seja ν uma distribuição inicial qualquer e defina $\pi_\infty(A) = P_\infty(x, A)$. Note que $P_\infty(x, A) = P_\infty(x', A)$, $\forall\ x$, $\forall\ x'$, pois P_∞ é um núcleo de transição constante. Além disso, $\nu * P_\infty = \pi_\infty$ e por (3.19)

$$\|\nu\tilde{P}_{0,n-1}^{y_{0:n}} - \nu P_\infty\| = \|\pi_{n,\,\nu}^{y_{0:n}} - \pi_\infty\|.$$

Usando a Proposição 1.1 (a) temos

$$\|\pi_{n,\,\nu}^{y_{0:n}} - \pi_\infty\| \leq \|\tilde{P}_{0,n-1}^{y_{0:n}} - P_\infty\|$$

e o resultado (3.9) segue de (3.20).

■

Teorema 3.1

Demonstração: Faremos a prova em duas etapas. Primeiro, mostraremos que temos a Condição D_k^*, (1.16), satisfeita. Em seguida, mostraremos que a Condição D_k^* assegura as hipóteses do Lema 3.1, do qual obtemos a ergodicidade fraca de X^c e a estabilidade do Filtro.

(i) Note que se temos (3.11) então o mesmo vale para qualquer $\alpha'_k < \alpha_k$. Assim, sem perda de generalidade, podemos assumir que $\alpha_k < 1$, $\forall\, k$. Por (3.11) temos

$$\alpha_k \leq \varphi_k(\mathcal{X}) \leq \frac{1}{\alpha_k}. \tag{3.21}$$

Podemos definir as probabilidades $\phi_k(A) = \dfrac{\varphi_k(A)}{\varphi_k(\mathcal{X})}$. Sejam $m_k = 1$, $\varepsilon_k = \dfrac{\alpha_k}{2} < \dfrac{1}{2}$ e $\delta_k = \dfrac{\alpha_k^3}{2} > 0$.

Para $A \in \mathbb{X}$, se $\phi_k(A) > \varepsilon_k$ então por (3.21)

$$\varphi_k(A) > \varepsilon_k \varphi_k(\mathcal{X}) \geq \frac{\alpha_k^2}{2}.$$

Por (3.11) temos

$$\tilde{P}_k^{y_{0:n}}(x, A) \geq \alpha_k \varphi_k(A) > \frac{\alpha_k^3}{2} = \delta_k$$

e

$$\phi_k(A) > \varepsilon_k \Rightarrow \inf_{x\in\mathcal{X}} \tilde{P}_k^{y_{0:n}}(x, A) > \delta_k. \tag{3.22}$$

Como, por (3.10), temos $\displaystyle\sum_{k'} \alpha_{k'}^3 = \infty$, $\forall\, \{k'\} \subset \{k\}$, a Condição D_k^* está satisfeita.

(ii) Mostraremos que vale (3.5) e a ergodicidade fraca é obtida do Lema 3.1. Dado $A \in \mathbb{X}$ então $\phi_k(A) > \varepsilon_k$ ou $\phi_k(A) \leq \varepsilon_k$. Se $\phi_k(A) > \varepsilon_k$ então por (3.22) temos $\forall\, x$, $\forall\, y$

$$|\tilde{P}_k^{y_{0:n}}(x, A) - \tilde{P}_k^{y_{0:n}}(y, A)| < 1 - \delta_k.$$

Se $\phi_k(A) \leq \varepsilon_k$ então, como $\varepsilon_k < \dfrac{1}{2}$, obtemos $\phi_k(A^c) > \varepsilon_k$. Por (3.22) temos para $\forall\ x,\ \forall\ y$

$$\begin{aligned}
|\tilde{P}_k^{y_{0:n}}(x, A) - \tilde{P}_k^{y_{0:n}}(y, A)| &= |(1 - \tilde{P}_k^{y_{0:n}}(x, A^c)) - (1 - \tilde{P}_k^{y_{0:n}}(y, A^c))| \\
&= |\tilde{P}_k^{y_{0:n}}(y, A^c) - \tilde{P}_k^{y_{0:n}}(x, A^c)| \\
&< 1 - \delta_k.
\end{aligned}$$

Concluímos que $\forall\ x, y \in \mathcal{X}$ e $\forall\ A \in \mathbb{X}$

$$|\tilde{P}_k^{y_{0:n}}(x, A) - \tilde{P}_k^{y_{0:n}}(y, A)| < 1 - \delta_k$$

ou seja,

$$\delta(\tilde{P}_k^{y_{0:n}}) = \sup_{\substack{x,\, y\, \in \mathcal{X} \\ A \in \mathbb{X}}} |\tilde{P}_k^{y_{0:n}}(x, A) - \tilde{P}_k^{y_{0:n}}(y, A)| < 1 - \delta_k. \tag{3.23}$$

Por outro lado, para $m = 0, 1, \ldots$ e $m + k_n \leq n$ temos

$$\tilde{P}_{m,m+k_n}^{y_{0:n}} = \tilde{P}_m^{y_{0:n}} * \tilde{P}_{m+1}^{y_{0:n}} * \ldots * \tilde{P}_{m+k_n-1}^{y_{0:n}}$$

e pela Proposição 1.1 (d)

$$\delta(\tilde{P}_{m,m+k_n}^{y_{0:n}}) \leq \prod_{j=m}^{m+k_n-1} \delta(\tilde{P}_j^{y_{0:n}}).$$

Usando (3.23) temos

$$\delta(\tilde{P}_{m,m+k_n}^{y_{0:n}}) \leq \prod_{j=m}^{m+k_n-1} (1 - \delta_j)$$

e fazendo $k_n \uparrow \infty$ com $m + k_n \leq n$ obtemos (3.5), pois, por (3.10) e o fato de $\delta_j = \dfrac{\alpha_j^3}{2}$ temos $\displaystyle\sum_{j=m}^{m+k_n-1} \delta_j = \infty$.

■

Observação 3.2

O fato da Condição D_k^* assegurar a ergodicidade fraca da cadeia pode ser encontrado em [13]. Entretanto, na demonstração do teorema acima apresentamos uma prova mais simples e direta.

Corolário 3.1

Demonstração: De fato, obtivemos na demonstração do Teorema 3.1 que

$$\delta(\tilde{P}^{y_{0:n}}_{m,m+k_n}) \leq \prod_{j=m}^{m+k_n-1} (1-\delta_j).$$

Usando (3.7) e o fato que $\delta_j = \dfrac{\alpha^3}{2}$, obtemos

$$\|\pi^{y_{0:n}}_{n,\,\nu} - \pi^{y_{0:n}}_{n,\,\nu'}\| \leq 2\delta(\tilde{P}^{y_{0:n}}_{0,n-1}) \leq 2\prod_{k=0}^{n-1}(1-\delta_j) = 2\left(1-\frac{\alpha^3}{2}\right)^n, \ \forall\, n.$$

■

Corolário 3.2

Demonstração: Basta verificar que temos (3.11) $\alpha_k = \dfrac{\lambda_*}{\lambda^*}$, $\forall\, k$. De (2.14) e (3.3) temos

$$\begin{aligned} \tilde{P}^{y_{0:n}}_{k-1}(x_{k-1}, A) &= \int_A \tilde{p}^{y_{0:n}}_{k-1}(x_k|x_{k-1})\mu(dx_k) \\ &= \frac{\int_A q(y_k|x_k)p(x_k|x_{k-1})Q_k(x_k)\mu(dx_k)}{\int_{\mathcal{X}} q(y_k|x_k)p(x_k|x_{k-1})Q(x_k)\mu(dx_k)}. \end{aligned}$$

Por (3.13) temos $\forall\, x_{k-1} \in \mathcal{X}$ e $A \in \mathbb{X}$,

$$\frac{\lambda_*}{\lambda^*}\varphi_{k-1}(A) \leq \tilde{P}^{y_{0:n}}_{k-1}(x_{k-1}, A) \leq \frac{\lambda^*}{\lambda_*}\varphi_{k-1}(A),$$

onde

$$\varphi_{k-1}(A) = \frac{\int_A q(y_k|x_k)Q_(x_k)\mu(dx_k)}{\int_{\mathcal{X}} q(y_k|x_k)Q_k(x_k)\mu(dx_k)}.$$

■

Teorema 3.2

Demonstração: No que segue fica subtendido que $m+k_n \leq n$ e $k_n \uparrow \infty$ quando $n \to \infty$. Também, afim de simplificar a notação, escreveremos

$$P^{y_{0:m}}_m * P^{y_{0:m+1}}_{m+1} * \ldots * P^{y_{0:m+k_n-1}}_{m+k_n-1} := P^{y_{0:m};y_{0:m+k_n}}_{m,m+k_n} := P_{m,m+k_n},$$

$$\tilde{P}^{y_{0:n}}_m * \tilde{P}^{y_{0:n}}_{m+1} * \ldots * \tilde{P}^{y_{0:n}}_{m+k_n-1} := \tilde{P}^{y_{0:n}}_{m,m+k_n} := \tilde{P}_{m,m+k_n}.$$

De modo que (3.14) se reduz a

$$\limsup_{n\to\infty} \sum_{k=0}^{n} \|P_k - \tilde{P}_k\| < \infty. \tag{3.24}$$

Note que, usando (a) e (b) da Proposição 1.1, obtemos

$$\begin{aligned} \|P_{m,m+k_n} - \tilde{P}_{m,m+k_n}\| &\leq \|P_{m,m+k_n-1} * (P_{m+k_n-1} - \tilde{P}_{m+k_n-1})\| + \\ &+ \|(P_{m,m+k_n-1} - \tilde{P}_{m,m+k_n-1}) * \tilde{P}_{m+k_n-1}\| \\ &\overset{(a),(b)}{\leq} \|P_{m+k_n-1} - \tilde{P}_{m+k_n-1}\| + \|P_{m,m+k_n-1} - \tilde{P}_{m,m+k_n-1}\|. \end{aligned}$$

Repetindo o processo para $\|P_{m,m+k_n-1} - \tilde{P}_{m,m+k_n-1}\|$ e assim sucessivamente, obtemos

$$\|P_{m,m+k_n} - \tilde{P}_{m,m+k_n}\| \leq \sum_{j=m}^{m+k_n-1} \|P_j - \tilde{P}_j\|. \tag{3.25}$$

(i) Vamos supor que $X^f = \{P_k\}$ não é fracamente ergódica e que $X^c = \{\tilde{P}_k\}$ é fracamente ergódica. Pela Proposição 1.1 (d) temos $\delta(P_{m,m+k_n})$ não-crescente com $n \to \infty$, assim temos

$$\exists\, m(a) \text{ tal que } \lim_{n\to\infty} \delta(P_{m(a),m(a)+k_n}) = a > 0. \tag{3.26}$$

Assim, segue de (3.25) que, dado $a > 0$ existe $m_0(a)$ tal que

$$\|P_{m,m+k_n} - \tilde{P}_{m,m+k_n}\| < \frac{a}{4}, \ \ \forall\, m \geq m_0(a). \tag{3.27}$$

Note que para $m < m^* < m + k_n$ temos

$$\|P_{m,m+k_n} - \tilde{P}_{m,m+k_n}\| \leq \|P_{m^*,m+k_n} - \tilde{P}_{m^*,m+k_n}\| + 2\delta(\tilde{P}_{m^*,m+k_n}). \tag{3.28}$$

Como $k_n \to \infty$ podemos tomar k_n de forma que $\max\{m(a), m_0(a)\} < m^* < m(a) + k_n$ e, usando (c) da Proposição 1.1, (3.27) e (3.28) obtemos

$$|\delta(P_{m(a),m(a)+k_n}) - \delta(\tilde{P}_{m(a),m(a)+k_n})| < 2\left(\frac{a}{4} + 2\delta(\tilde{P}_{m^*,m(a)+k_n})\right).$$

Pela ergodicidade fraca de $\{\tilde{P}_k\}$ segue que

$$\lim_{n\to\infty} \delta(P_{m(a),m(a)+k_n}) < \frac{a}{2},$$

contrariando (3.26). Logo, as cadeias X^f e X^c ou são fracamente ergódicas ou nenhuma delas é fracamente ergódica.

(ii) Vamos tratar da ergodicidade forte. Vamos supor que $X^c = \{\tilde{P}_k\}$ é fortemente ergódica e que $X^f = \{P_k\}$ não é fortemente ergódica. Assim existe um núcleo de transição constante P_∞ tal que

$$\lim_{n\to\infty} \|\tilde{P}_{m,m+k_n} - P_\infty\| = 0, \;\; \forall\, m. \tag{3.29}$$

Por hipótese, para qualquer núcleo de transição constante Q existe $m(Q)$ tal que

$$\lim_{n\to\infty} \|P_{m(Q),m(Q)+k_n} - Q\| = a(Q) > 0. \tag{3.30}$$

Logo, dado P_∞ núcleo de transição constante, existe $m(P_\infty) := m_0$ tal que

$$\lim_{n\to\infty} \|P_{m_0,m_0+k_n} - P_\infty\| = a_0 > 0. \tag{3.31}$$

Sendo $X^c = \{\tilde{P}_k\}$ fortemente ergódica, também é fracamente ergódica, assim

$$\lim_{n\to\infty} \delta(\tilde{P}_{m,m+k_n}) = 0, \; \forall\, m. \tag{3.32}$$

Por (3.25), dado $a_0 > 0$ existe $m(a_0)$ tal que

$$\|P_{m,m+k_n} - \tilde{P}_{m,m+k_n}\| < \frac{a_0}{2}, \;\; \forall\, m \geq m(a_0). \tag{3.33}$$

Como $k_n \to \infty$, podemos tomar $\max\{m_0, m(a_0)\} < m^* < m_0 + k_n$, e observando que

$$\|P_{m_0,m_0+k_n} - P_\infty\| \leq \|P_{m_0,m_0+k_n} - \tilde{P}_{m_0,m_0+k_n}\| + \|\tilde{P}_{m_0,m_0+k_n} - P_\infty\|$$

usando (3.28) e (3.33) obtemos

$$\|P_{m_0,m_0+k_n} - P_\infty\| \leq \Big(\frac{a_0}{2} + 2\delta(\tilde{P}_{m^*,m_0+k_n})\Big) + \|\tilde{P}_{m_0,m_0+k_n} - P_\infty\|.$$

Então, usando (3.29) e (3.32) obtemos

$$\lim_{n\to\infty} \|P_{m_0,m_0+k_n} - P_\infty\| \leq \frac{a_0}{2}$$

contrariando (3.31). Assim, ou as duas cadeias são fortemente ergódicas ou nenhuma delas é fortemente ergódica.

Portanto, $X^f = \{P_k\}$ e $X^c = \{\tilde{P}_k\}$ possuem o mesmo comportamento ergódico.

■

Teorema 3.3

Demonstração: Pela prova do Teorema 3.1 basta mostrar que temos a Condição D_k^* satisfeita. Por (2.21) temos

$$P_{n-1}^{y_{0:n}}(x_{n-1}, A) = \frac{\int_A q(y_n|x_n)p(x_n|x_{n-1})d\mu(x_n)}{\int_{\mathcal{X}} q(y_n|x'_n)p(x'_n|x_{n-1})d\mu(x'_n)}.$$

Seja C_{y_n} dado por (3.15) então por (3.15) e (3.16)

$$\frac{\int_{A\cap C_{y_n}} q(y_n|x_n)p(x_n|x_{n-1})d\mu(x_n)}{\int_{C_{y_n}} q(y_n|x'_n)p(x'_n|x_{n-1})d\mu(x'_n)} \geq \frac{p_*}{p^*}\phi_{n-1}(A)$$

onde $\phi_{n-1}(A) = \dfrac{\int_{A\cap C_{y_n}} q(y_n|v)d\mu(v)}{\int_{C_{y_n}} q(y_n|v')d\mu(v')}$.

Segue que $\forall\ x \in \mathcal{X}$ e $A \in \mathbb{X}$ temos por (3.16)

$$\begin{aligned} P_{n-1}^{y_{0:n}}(x, A) &\geq P_{n-1}^{y_{0:n}}(x, C_{y_n})\frac{p_*}{p^*}\phi_{n-1}(A) \\ &\geq \rho\frac{p_*}{p^*}\phi_{n-1}(A). \end{aligned}$$

Dado $\varepsilon_k = \varepsilon < \dfrac{1}{2}$ temos que se $\phi_{n-1}(A) > \varepsilon_k$, então $\inf\limits_{x\in\mathcal{X}} P_{n-1}^{y_{0;n}}(x, A) \geq \delta_k$ onde $\delta_k = \rho\dfrac{p_*}{p^*}\varepsilon$, $\forall\ k$. Assim, temos a Condição D_k^*, pois $\sum\limits_{k'} \delta_{k'} = \infty$, $\forall\ \{k'\} \subset \{k\}$.

■

Corolário 3.3

Demonstração: A prova é imediata.

■

Capítulo 4

Aplicações

Neste Capítulo, faremos alguns comentários sobre a relação entre a estabilidade do Filtro e a estimativa da distribuição de filtragem obtida via Filtro de Bootstrap. Também, faremos alguns comentários necessários sobre a estimação da densidade de equilíbrio do processo observável afim de obtermos estimativas para as probabilidades de transição da cadeia oculta, quando $\mathcal{X}$ é finito.

Na seção 4.1, analisamos a estimativa da distribuição de filtragem obtida via Filtro de Bootstrap sob a condição de estabilidade do Filtro. Mostramos que a estimativa da distribuição de filtragem, obtida pelo Filtro de Bootstrap, depende da distribuição inicial quando o Filtro não é estável. No caso do Filtro ser estável qualquer distribuição inicial pode ser utilizada para inicializar o algoritmo Filtro de Bootstrap. Utilizamos os exemplos do Capítulo 2 para mostrar que o resultado obtido via Filtro de Bootstrap pode depender da condição inicial adotada.

Na seção 4.2, fazemos comentários sobre como obter a densidade de equilíbrio de Y_n. Utilizamos a hipótese que a cadeia oculta sendo ergódica tem distribuição limite π_∞ e, com isso, podemos obter a densidade de equilíbrio de Y_n. Nesta seção trabalhamos com $\mathcal{X}$ geral, $\mathcal{Y} = \mathbb{R}$ e utilizamos a teoria de estimadores de núcleo para uma amostra de observações $Y_0, \ldots, Y_n$. A densidade de equilíbrio de Y_n será importante para estimar a distribuição limite da cadeia oculta, quando $\mathcal{X}$ for finito. Esta, por sua vez, será importante para estimar as transições da cadeia oculta.

Na seção 4.3, utilizamos os comentários da seção 4.2 e o Teorema 2 de [14] para obtermos a Proposição 4.1. Para o contexto de HMM, é mostrado no Teorema 2 de [14], que existem constantes $c_1 = c_1(\varepsilon) > 0$ e $c_2 = c_2(\varepsilon) > 0$ independentes de n tais que para qualquer condição inicial ν,

$$P_\nu\left(\int |\hat{f}_n(y) - f_\infty(y)|dy \geq \delta\right) \leq c_1 e^{-c_2 n}, \tag{4.1}$$

onde $f_n(y)$ é obtida via estimadores de núcleo e f é a densidade de equilíbrio de Y_n. Como consequência obtemos uma estimativa para probabilidade de transição da cadeia oculta do exemplo 2.1, ou seja, estimaremos p_1.

4.1 Estimação da Distribuição de Filtragem

Conforme comentamos na seção 1.2, existem algoritmos para estimar a distribuição de filtragem, sendo que o Filtro de Bootstrap é um dos mais conhecidos. Para que possamos compilar o algoritmo a densidade de $Y_k|X_k = x_k$ deve ser conhecida e deve ser possível gerar amostras do ruído $\{\varepsilon_k\}$ (veja (1.6)), comumente é assumido que sua densidade g_ε é conhecida. Além disso, para inicializar o algoritmo é necessária a distribuição inicial da cadeia oculta $X = \{X_n\}$. Entretanto, como a cadeia X é não-observável, geralmente não temos disponível a verdadeira distribuição inicial da cadeia. Assim, devemos escolher uma distribuição inicial para inicializar o algoritmo.

Quando a distribuição inicial que escolhemos é a verdadeira distribuição inicial da cadeia oculta o algoritmo fornece uma estimação para a distribuição de filtragem e, assim, o problema de estimação da distribuição de filtragem está resolvido. No entanto, quando a distribuição inicial que escolhemos não é a verdadeira distribuição inicial da cadeia oculta, o Filtro de Bootstrap pode não estimar a distribuição de filtragem. De fato, neste caso, para cada distribuição inicial diferente que escolhermos o Filtro de Bootstrap pode retornar uma estimativa diferente e não teremos como decidir qual é a estimativa para a distribuição de filtragem. Este seria o caso, por exemplo, se o algoritmo Filtro de Bootstrap fosse utilizado para estimar a distribuição de filtragem do exemplo 2.2, pois mostramos que a distribuição de filtragem depende da condição inicial. Na verdade, no exemplo 2.2 caso 2, $(2c)$, mostramos que

$$\lim_{n\to\infty} \|\pi^{y_{0:n}}_{n,\,\nu} - \pi^{y_{0:n}}_{n,\,\nu'}\| = 1$$

para duas condições iniciais diferentes ν e ν'. Assim, inicializando o algoritmo Filtro de Bootstrap pela condição inicial ν obteríamos um conjunto de partículas $\{x_{n,\nu}^{(i)} : i = 1, \ldots, N\}$ tal que

$$\frac{1}{N}\sum_{i=1}^{N}\delta_{x_{n.\nu}^{(i)}} \overset{N\to\infty}{\longrightarrow} \pi_{n,\nu}^{y_{0:n}}$$

e, analogamente, inicializando por ν' obteríamos $\{x_{n,\nu'}^{(i)} : i = 1, \ldots, N\}$ tal que

$$\frac{1}{N}\sum_{i=1}^{N}\delta_{x_{n,\nu'}^{(i)}} \overset{N\to\infty}{\longrightarrow} \pi_{n,\nu'}^{y_{0:n}}.$$

Portanto,

$$\left\| \frac{1}{N}\sum_{i=1}^{N}\delta_{x_{n,\nu}^{(i)}} - \frac{1}{N}\sum_{i=1}^{N}\delta_{x_{n,\nu'}^{(i)}} \right\| \overset{N\to\infty}{\longrightarrow} 1.$$

Uma situação desejável seria que, independente da distribuição incial escolhida, o algoritmo Filtro de Bootstrap retornasse, para N grande, a mesma (ou muito próxima) estimativa para a distribuição de filtragem. Esta situação desejável ocorre quando o Filtro é estável, isto é, quando o Filtro é independente da distribuição inicial. Isto pode ser observado no exemplo 2.1, onde a distribuição de filtragem independe da distribuição inicial. Neste caso, se o algoritmo Filtro de Bootstrap fosse utilizado para estimar a distribuição de filtragem então poderíamos inicializar o algoritmo a partir de qualquer distribuição inicial que o algoritmo retornaria a mesma estimativa para a distribuição de filtragem. De fato, mostramos no exemplo 2.1 que, para quaisquer distribuições iniciais ν e ν'

$$\lim_{n\to\infty} \|\pi_{n,\nu}^{y_{0:n}} - \pi_{n,\nu'}^{y_{0:n}}\| = 0.$$

Assim, inicializando o algoritmo Filtro de Bootstrap por ν e ν' obteríamos conjuntos de partículas $\{x_{n,\nu}^{i} : i = 1, \ldots, N\}$ e $\{x_{n,\nu'}^{(i)} : i = 1, \ldots, N\}$, respectivamente, tais que

$$\left\| \frac{1}{N}\sum_{i=1}^{N}\delta_{x_{n,\nu}^{(i)}} - \frac{1}{N}\sum_{i=1}^{N}\delta_{x_{n,\nu'}^{(i)}} \right\| \overset{N\to\infty}{\longrightarrow} 0.$$

Portanto, vemos que o algoritmo Filtro de Bootstrap pode depender da distribuição inicial escolhida para inicialização. Com isso, considerando que o algoritmo seja eficiente quando independe da distribuição inicial utilizada, temos que sua eficiência está dependendo da estabilidade do Filtro.

Quando $\mathcal{X}$ é finito é possível obter a sequência de estados $X_0, \ldots, X_n$ da cadeia oculta a partir das observações, conforme [4] Capítulo 5. Um problema associado com a recuperação dos estados da cadeia oculta é o de estimar as transições da cadeia oculta. Por exemplo, estimar p_1 e ε no exemplo 2.1. No contexto de controle on-line o ε pode ser estimado a partir do histórico de falhas no processo de produção (por exemplo, como a proporção de vezes que o processo ficou sob controle imediatamente após a intervenção de um técnico). Assim, o problema de interesse seria estimar a probabilidade de falha p_1.

Associado ao problema de se estimar as transições da cadeia oculta, quando $\mathcal{X}$ é finito, há o problema de se estimar a densidade de equilíbrio do processo de observações $Y = \{Y_n\}$. Trataremos deste problema na próxima seção.

4.2 Estimação da Densidade de Equilíbrio

Nesta seção vamos considerar o HMM $(X, Y) = \{(X_n, Y_n)\}$ onde a cadeia não-observável tem espaço de estados $\mathcal{X}$ e o processo de observações tem espaço de estados $\mathbb{R}$.

Assumiremos que a cadeia X tenha distribuição limite π_∞, de forma que

$$\pi_\infty(A) = \int_A \pi_\infty(x)dx = \lim_{n\to\infty} P(X_n \in A), \ \forall\ A \in \mathbb{X}.$$

A distribuição de $Y_n|X_n$ é conhecida e vamos considerar que

$$P(Y_n \in B|X_n = x) = \int_B q(y|x)dy, \ \forall\ B \in \mathcal{B}(\mathbb{R}),$$

onde $\mathcal{B}(\mathbb{R})$ é a σ-álgebra dos Borelianos em $\mathbb{R}$.

Supondo que desejamos obter a distribuição limite para $\{Y_n\}$, isto é, a densidade de Y_n para n arbitrariamente grande (quando o processo Y atinge um regime estável) podemos observar que

$$\begin{aligned}
\lim_{n\to\infty} P(Y_n \in B) &= \lim_{n\to\infty} \int_{\mathcal{X}} P(Y_n \in B|X_n = x)\pi_\infty(dx) \\
&= \int_{\mathcal{X}} \int_B q(y|x)dy\pi_\infty(dx) \\
&= \int_B \left[\int_{\mathcal{X}} q(y|x)\pi_\infty(dx)\right] dy.
\end{aligned}$$

Logo, a densidade de equilíbrio de Y_n é dada por

$$f_\infty(y) = \int_{\mathcal{X}} q(y|x)\pi_\infty(dx).$$

Como é conhecido que núcleos estimadores constituem uma ferramenta para aproximar densidades, podemos lançar mão desta teoria para estimar f_∞.

Dada uma amostra $Y_1, \ldots, Y_n$ de observações, definimos para uma densidade K sobre $\mathbb{R}$,

$$\hat{f}_n(y) = \frac{1}{nh}\sum_{k=1}^{n} K\left(\frac{Y_k - y}{h}\right),\ h = h_n \downarrow 0,\ nh_n \stackrel{n\to\infty}{\longrightarrow} \infty. \tag{4.2}$$

Em geral, nem mesmo o conhecimento de $f_\infty(y)$ para todos os valores de y permitem conhecer π_∞. Entretanto, em casos especiais, como por exemplo $\mathcal{X}$ finito, é possível obter π_∞ a partir do conhecimento de $f_\infty(y)$. Assim, uma estimativa de f_∞ pode ser utilizada para obter uma estimativa para π_∞. Isso é o que mostraremos na próxima seção.

4.3 O Problema de Controle "On-line"

Nesta seção, utilizaremos uma estimativa de f_∞ para estimar a distribuição ergódica (que é estacionária), π_∞, da cadeia oculta do exemplo 2.1. Como a distribuição π_∞, no exemplo 2.1, depende de p_1 e ε, e como podemos estimar ε a partir do histórico das falhas, teremos um mecanismo para estimar p_1 a partir da estimação de π_∞. Os detalhes dos resultados desta seção podem ser vistos em [11].

A Proposicão a seguir é uma consequência do Teorema 2 de [14] e será utilizada para obter uma estimativa para as transições da cadeia não-observável X. A Proposição será mostrada para $\mathcal{X} = \{a_0, a_1, \ldots\}$ e no Corolário utilizaremos $\mathcal{X} = \{1, 2\}$.

Proposição 4.1. *Suponha o HMM $\{(X_n, Y_n)\}$ com a cadeia $\{X_n\}$ uniformemente ergódica. Então, dado $\delta > 0$ existem constantes $c_1 = c_1(\delta) > 0$ e $c_2 = c_2(\delta) > 0$ tais que, para qualquer distribuição inicial de X_0 vale*

$$P_\nu\left(\int |\hat{f}_n(y) - f_\infty(y)|dy \geq \delta\right) \leq c_1 e^{-c_2 n} \tag{4.3}$$

e, quase certamente,

$$\lim_{n\to\infty}\int_{\mathcal{X}} |\hat{f}_n(y) - f_\infty(y)|dy = 0. \tag{4.4}$$

Demonstração: A demonstração desta proposição, conforme [14], se resume em verificar a validade da condição: X tem distribuição estacionária π e satisfaz a condição ϕ-mixing: $\forall A \in \mathcal{F}_k,\ B \in \mathcal{F}_{k+n}^{\infty}$

$$|P_\pi(A \cap B) - P_\pi(A)P_\pi(B)| \leq \phi(n)P_\pi(A)$$

com $\sum_{n\geq 1} \phi(n) = M_1 < \infty$, onde as σ-álgebras são definidas por

$$\mathcal{F}_k = \sigma(X_0, X_1, \ldots, X_k) \text{ e } \mathcal{F}_{k+n}^{\infty} = \sigma(X_{k+n}, \ldots).$$

Como estamos assumindo que X é uniformemente ergódica segue que existe uma distribuição limite π_∞, que é estacionária. Logo, tomando $A \in \mathcal{F}_k$ com $P_{\pi_\infty}(A) > 0$ e $B \in \mathcal{F}_{k+n}^{\infty}$ temos $A = (X_0 = a_0, \ldots, X_k = a_k)$ e $B = (X_{k+n} = b_{k+n}, \ldots)$ donde segue que

$$\begin{aligned} P_{\pi_\infty}(B|A) &= P_{\pi_\infty}^n(b_{k+n}|a_k)P_{\pi_\infty}(X_{k+n+1} = b_{k+n+1}, \ldots |X_{k+n} = b_{k+n}) \text{ e} \\ P_{\pi_\infty}(B) &= P_{\pi_\infty}(X_{k+n} = b_{k+n})P_{\pi_\infty}(X_{k+n+1} = b_{k+n+1}, \ldots |X_{k+n} = b_{k+n}) \\ &= \pi_\infty(b_{k+n})P_{\pi_\infty}(X_{k+n+1} = b_{k+n+1}, \ldots |X_{k+n} = b_{k+n}). \end{aligned}$$

Assim,

$$|P_{\pi_\infty}(B|A) - P_{\pi_\infty}(B)| \leq |P_{\pi_\infty}^n(b_{k+n}|a_k) - \pi_\infty(b_{k+n})| \leq \rho^n = \phi(n)$$

e $\sum_{n\geq 1} \phi(n) = M_1 < \infty$, pois $0 < \rho < 1$. A última desigualdade é válida por causa da ergodicidade uniforme da cadeia X.

■

Para o problema de controle "on-line" na forma original em [17], a estrutura HMM considerada é composta pela cadeia não-observável $X = \{X_n\}_{n\geq 0}$ com espaço de estados $\mathcal{X} = \{1, 2\}$, matriz de transição

$$Q = \begin{bmatrix} 1-p & p \\ 0 & 1 \end{bmatrix}$$

e pelo processo de obseravções $Y = \{Y_n\}_{n\geq 0}$, onde $Y_n|X_n = x_n \sim N(x_n, \sigma^2)$. Aqui, consideramos uma pequena perturbação na matriz de transição Q e a denominamos P onde

$$P = \begin{bmatrix} 1-p & p \\ \varepsilon & 1-\varepsilon \end{bmatrix}.$$

Nossa perturbação na matriz Q, do ponto de vista teórico, nos fornece uma matriz ergódica que é extremamente desejável. Do ponto de vista prático, essa perturbação pode ser interpretada como uma pequena probabilidade que o processo fique sob controle, sem intervenção de um técnico, após ficar fora de controle. Por exemplo, em uma queda de energia alguns equipamentos podem voltar a funcionar imediatamente após a restauração da energia. Também, consideramos as observações como

$$Y_n = 1_{(X_n=1)}.$$

Neste contexto, podemos estimar a probabilidade de falha através da estimativa da distribuição estacionária da cadeia X. Temos que

$$P = \begin{bmatrix} 1-p & p \\ \varepsilon & 1-\varepsilon \end{bmatrix}, \tag{4.5}$$

onde $0 < p < 1$ e $0 < \varepsilon < 1$. Aqui assumiremos ainda que $p + \varepsilon < 1$. Logo, como

$$P^n = \begin{bmatrix} \dfrac{\varepsilon}{p+\varepsilon} + p\dfrac{(1-p-\varepsilon)^n}{p+\varepsilon} & \dfrac{p}{p+\varepsilon} - p\dfrac{(1-p-\varepsilon)^n}{p+\varepsilon} \\ \\ \dfrac{\varepsilon}{p+\varepsilon} - \varepsilon\dfrac{(1-p-\varepsilon)^n}{p+\varepsilon} & \dfrac{p}{p+\varepsilon} + \varepsilon\dfrac{(1-p-\varepsilon)^n}{p+\varepsilon} \end{bmatrix}$$

e a cadeia X é ergódica com distribuição limite $\pi_\infty = \left(\dfrac{\varepsilon}{\varepsilon+p}, \dfrac{p}{\varepsilon+p}\right)$, um simples cálculo mostra que X é uniformemente ergódica com $\rho^n = (1-p-\varepsilon)^n$.

Também, temos que $f_\infty(y) = \pi(1)q(y|1) + \pi(2)q(y|2)$, donde segue que

$$\pi(1) = \frac{f_\infty(y) - q(y|2)}{q(y|1) - q(y|2)},$$

desde que $q(y|1) - q(y|2) \neq 0$.

Portanto, para y_* tal que $0 < |\hat{f}_n(y_*) - q(y_*|2)| < |q(y_*|1) - q(y_*|2)|$ podemos definir

$$\hat{\gamma}_n(1) = \left|\frac{\hat{f}_n(y_*) - q(y_*|2)}{q(y_*|1) - q(y_*|2)}\right| \text{ e } \hat{\gamma}_n(2) = 1 - \hat{\gamma}_n(1).$$

Segue assim o seguinte corolário

Corolário 4.1. *Nas condições de (4.5) obtemos $\hat{\gamma}_n(1) \to \pi(1)$ e $\hat{\gamma}_n(2) \to \pi(2)$, quando $n \to \infty$.*

Como o ε pode ser estimado a partir do histórico de falhas do processo de produção e o p pode ser estimado via o Corolário 4.1, temos que as transições da cadeia não-observável podem ser estimadas a partir das observações.

Neste contexto, calculamos $\pi_n^{y_{0:n}}(x_n)$ no exemplo 2.1 e obtivemos

$$\pi_n^{y_{0:n}}(x_n) = \begin{cases} y_n & \text{se } x_n = 1 \\ 1 - y_n & \text{se } x_n = 2 \end{cases} . \tag{4.6}$$

Podemos utilizar o algoritmo Filtro de Bootstrap para estimar $\pi_n^{y_{0:n}}$. Como já mostramos que o Filtro é estável, podemos inicializar o algoritmo Filtro de Bootstrap a partir de qualquer condição inicial da cadeia oculta X que obtemos um conjunto de partículas $\{x_n^{(i)} : i = 1, 2, \ldots, N\}$ tal que

$$\frac{1}{N}\sum_{i=1}^{N} \delta_{x_n^{(i)}} \overset{N\to\infty}{\longrightarrow} \pi_n^{y_{0:n}}.$$

No caso em que o Filtro é estável podemos utilizar o algoritmo Filtro de Bootstrap para estimar a distribuição de filtragem sem preocupação com a condição inicial a ser utilizada. Vamos observar a situação de se estimar a distribuição de filtragem, via algoritmo Filtro de Bootstrap, para o problema de controle "on-line" do exemplo 2.2, onde o Filtro não necessariamente é estável, ou seja, quando a cadeia X tem espaço de estados $=\{1, 2, 3\}$, matriz de transição

$$P = \begin{bmatrix} 1-p_1 & p_1 & 0 \\ 0 & 1-p_2 & p_2 \\ \varepsilon & 0 & 1-\varepsilon \end{bmatrix}$$

e o processo das observações é dado por

$$Y_n = 1_{(X_n=1)} + 1_{(X_n=2)}.$$

Consideramos $0 < p_i < 1$, para $i = 1, 2$, e $0 < \varepsilon < 1$.

Nestas condições já mostramos que a estabilidade ou não do Filtro depende da relação entre p_1 e p_2, sendo estável para $p_1 > p_2$ e não-estável para $p_1 < p_2$. No entanto, sem o conhecimento de p_1 e p_2 não podemos assegurar a estabilidade do Filtro de forma que inviabiliza o uso do algoritmo Filtro de Bootstrap para estimar a distribuição de filtragem. Podemos contornar esta dificuldade adotando o procedimento de parar o algoritmo quando ocorrer a primeira falha no processo. De fato, mostramos que para $(y_n = 0)$ ou $(y_{n-1:n} =$

01) o Filtro é estável sem considerações sobre p_1 e p_2. Também, mostramos que quando $y_{n-1:n} = 11$ precisamos da condição $p_1 > p_2$ para garantir a estabilidade do Filtro. Logo, enquanto não ocorrer falha no processo o Filtro é estável, ou seja, podemos utilizar o algoritmo Filtro de Bootstrap para estimar a distribuição de filtragem até o momento em que ocorrer a primeira falha. Portanto, esta constitui uma alternativa para estimar a distribuição de filtragem mesmo quando o Filtro não é necessariamente estável.

Referências Bibliográficas

[1] Atar, R., Zeitouni, O., *Lyapunov Exponents for Finite State Nonlinear Filtering*, SIAM Journal on Control and Optimization **35**: 36-55 (1997).

[2] Atar, R., Zeitouni, O., *Exponential Stability for Nonlinear Filtering*, SIAM Journal on Control and Optimization **33**: 697-725 (1997).

[3] Blackwell, D., *The Entropy of Functions of Finite-State Markov Chains*, Transactions of First Prague Conference in Information Theory, Statistical Decision Functions, Random Processes, 13-20, Prague, (1957).

[4] Cappé, O., Moulines, E., Rydén, T., *Inference in Hidden Markov Models*, Springer, New York (2005).

[5] Chigansky, P., Lipster, R., Van Hndel, R., *Intrinsic Methods in Filter Stability*, in Oxford University Handbook of Nonlinear Filtering, Oxford University Press (to appear).

[6] Chigansky, P., *Stability of Nonlinear Filters: A Survey*, Lecture notes on mine course at Laboratório Nacional de Computação Científica, Petrópolis (2006).

[7] Chigansky, P., Lipster, R., *Stability of Nonlinear Filters in Nonmixing Case*, The Annals of Applied Probability **14**: 2038-2056 (2004).

[8] Cruz, J., *Convergência de Cadeias de Markov Não-Homogêneas: Ergodicidade Fraca e Forte*, Tese de doutorado, Brasília (1998).

[9] Del Moral, P., Guionnet, A., *On the Stability of Interacting Processes With Applications to Filtering and Genetic Algorithms*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques **37**: 155-194 (2001).

[10] Dorea, C., Cruz, J., *Approximation Results for Non-Homogeneous Markov Chains and Some Applications*, Sankhyã: The Indian Journal of Statistics **66**: 243-252 (2004).

[11] Dorea, C., Gonçalves, C., Medeiros, P., Santos, W., *Modelling On-line Quality Control Analysis by Hidden Markov Model* (a aparecer).

[12] Dorea, C., Martins Neto, D., *Convergence of Non-Homogeneous Versions of the MCEM and StEM Algorithms*, Advances and Applications in Statistics **15**: 101-113 (2010).

[13] Dorea, C., Pereira, A., *A Note on a Variation of Doeblin´s Condition for Uniform Ergodicity of Markov Chains*, Acta Mathematica Hungarica **110**: 287-292 (2006).

[14] Dorea, C., Zhao, L., *Nonparametric Density Estimation in Hidden Markov Models*, Statistical Inference for Stochastic Processes **5**: 55-64 (2002).

[15] Doucet, A., Freitas, N., Gordon, N., *Sequential Monte Carlo Methods in Practice*, Springer, New York (2001).

[16] Gordon, N., Salmond, D., Smith, A., *Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation*, IEE Proceedings-F **40**: 107-113 (1993).

[17] Ho, L., Medeiros, P., Borges, W., *An Alternative Model for On-Line Quality Monitoring for Variables*, International Journal of Production Economics **107**: 202-222 (2007).

[18] Isaacson, D., Madsen, R., *Markov Chains, Theory and Applications*, Wiley, New York (1976).

[19] Kaijser, T., *A Limit Theorem for Partially Observed Markov Chains*, The Annals of Probabiblity **3**: 677-696 (1975).

[20] Kunita,H., *Asymptotic Behavior of the Nonlinear Filtering Erros of Markov Processes*, Journal of Multivariate Analysis **1**: 365-393 (1971).

[21] Metropolis, N., Ulam, S., *The Monte Carlo Method*, Journal of the American Statistical Association **44**: 335-341 (1949).

[22] Meyn, S., Tweedie, R., *Markov Chains and Stochastic Stability*, Springer, Berlin (1993).

[23] Nummelin, E., *General Irreducible Markov Chains and Non-Negative Operators*, Cambridge University Press, Cambridge (1984).

[24] Pereira, A., *Análise da Ergodicidade das Cadeias de Markov Não-Homogêneas Via Condição do Tipo Doeblin*, Tese de doutorado, Brasília (2001).