

Estimação do Número de Fontes Utilizando Métodos Baseados na Teoria da Informação

Patrícia Maria Soares, Dayan Adionel Guimarães e Rausley Adriano Amaral de Souza

Abstract—This is a tutorial paper on methods that use information theoretic fundamentals for estimating the number of sources. Three methods are addressed: the Akaike information criterion (AIC), the minimum description length (MDL) and the Bayesian information criterion (BIC). The tutorial also provides a background material so that the understanding of the above methods is facilitated and the text becomes self-contained. Nevertheless, a large number of references are cited to allow for the reader to seek for more detailed information, if needed.

Index Terms—Akaike information criterion, Bayesian information criterion, estimation of the number of sources, maximum description length.

Resumo—Este tutorial discorre sobre métodos de estimação do número de fontes baseados na teoria da informação. Três métodos são abordados: o critério de informação de Akaike (AIC), o método do comprimento de mínima descrição (MDL) e o critério de informação Bayesiana (BIC). Alguns conceitos fundamentais também são fornecidos, objetivando facilitar o entendimento dos métodos supracitados e tornar o tutorial mais completo e até certo ponto independente da consulta a outras referências em um primeiro momento. Ainda assim, um grande número de referências é listado de forma que o leitor possa procurar por informações mais detalhadas, caso necessário.

Palavras chave—critério de informação de Akaike (AIC), critério de informação Bayesiana (BIC), estimação do número de fontes, método do comprimento de máxima descrição (MDL).

I. INTRODUÇÃO

Estimar o número de fontes significa determinar a quantidade de fontes geradoras dos estímulos recebidos por um arranjo de sensores. Trata-se de um problema clássico em processamento digital de sinais, encontrando aplicações em telecomunicações, em medicina, em sistemas de radares, em estatística e em muitas outras áreas [1]-[8].

No contexto de sistemas de telecomunicações, estimar o número de fontes significa estimar o número de transmissores ativos em uma determinada faixa de frequências, usando um arranjo de antenas como elemento sensor ou um conjunto de receptores com antenas simples compondo um arranjo virtual de antenas. No contexto de sensoriamento espectral em sistemas de rádio cognitivo (RC) [9], por exemplo, pode ser importante nas técnicas de detecção em que a rede secundária conheça o número de transmissores para que possa corretamente computar a estatística de teste que permitirá verificar o estado de ocupação do espectro, ou até mesmo como dado adicional de entrada de um processo cognitivo mais amplo. Neste caso as antenas dos rádios cognitivos compõem o arranjo virtual de antenas. Outra aplicação em sistemas de telecomunicações consiste da estimação da variância de ruído térmico na entrada de um receptor, na presença de sinal. Vários algoritmos de estimação desta variância necessitam conhecer o número de sinais recebidos de

forma que operem a contento [3]. Este também é o caso de vários algoritmos de estimação de direção de chegada (DoA, *direction of arrival*) de sinais em arranjos de antenas, os quais produzem estimativas mais precisas se é conhecido o número de transmissores [10]. Em sistemas de radar, a estimação do número de fontes pode auxiliar no conhecimento do número de ecos de um sinal refletido por um alvo, permitindo que etapas posteriores de processamento reduzam a influência de tais ecos na caracterização do alvo [2][11]. Em medicina, a estimação do número de fontes pode ser utilizada, por exemplo, na contagem do número de neurônios que respondem a estímulos controlados em cada parte do cérebro, permitindo que as áreas cerebrais sensíveis a cada tipo de estímulo sejam identificadas [8].

Muito se tem pesquisado sobre a estimação do número de fontes desde o trabalho seminal de M. Wax e T. Kailath [2], o qual discorre sobre os métodos fundamentados pela teoria da informação. Entre eles merecem destaque o método baseado no critério de informação de Akaike (AIC, *Akaike information criterion*), o método do comprimento de mínima descrição (MDL, *minimum description length*) e o método baseado no critério de informação Bayesiana (BIC, *Bayesian information criterion*) [12]. Dentre estes, o método MDL tem sido considerado referência ou elemento de comparação na maioria dos estudos de estimação [3].

Métodos baseados na teoria de matrizes aleatórias (RMT, *random matrix theory*) têm sido explorados como alternativos em relação àqueles baseados na teoria da informação. Em [3], por exemplo, são aplicados recentes resultados sobre a distribuição dos autovalores da matriz de covariância do sinal recebido e do ruído em um novo algoritmo baseado em uma sequência de testes de hipóteses. O algoritmo de [3] procura obter alta probabilidade de detecção correta em regimes de baixa relação sinal-ruído, característica marcante do estimador AIC, bem como boa consistência em regimes de grandes números de amostras do sinal recebido, característica relevante do estimador MDL. A maior desvantagem do método baseado na teoria de matrizes aleatórias consiste da necessidade de se estimar a variância de ruído térmico para que, em seguida, um limiar de decisão do teste de hipóteses seja computado. Isto pode representar um grande obstáculo de ordem prática, além de produzir variação de desempenho em função do grau de precisão do valor estimado de variância.

Nos métodos baseados na teoria da informação, assim como no método baseado na teoria de matrizes aleatórias, o número de fontes é estimado tomando por base uma classificação dos autovalores da matriz de covariância do sinal (estímulo) sensoriado. Tal classificação pode ser caracterizada como um problema de seleção de modelo (*model selection*) ou de

determinação de ordem de modelo (*model order determination*) [1]-[3][12] e busca identificar, dentre todos os autovalores, um subconjunto cujo número de elementos está associado ao número de sinais recebidos. Esta classificação dos autovalores corresponde a um dos maiores desafios destes métodos.

Neste tutorial apresenta-se uma introdução aos métodos de estimação do número de fontes baseados na teoria da informação, restringindo-se a abordagem àqueles que utilizam os autovalores da matriz de covariância do sinal recebido. Como preparação para o entendimento de tais métodos, a Seção II descreve a formulação do problema e na Seção III são apresentados alguns fundamentos relacionados à teoria da estimação de ordem de modelo. Nas seções IV, V e VI são apresentados os métodos de estimação AIC, BIC e MDL, respectivamente e também resultados de simulação e interpretações sobre o desempenho de cada um. A Seção VII apresenta resultados de simulação comparativos e os comentários finais estão presentes na Seção VIII.

II. FORMULAÇÃO DO PROBLEMA

Considere que o sinal incidente em um receptor com um arranjo de m antenas (ou em m receptores com uma antena cada um, formando um arranjo virtual) no i -ésimo instante de tempo discreto, $i = 1, 2, \dots, n$, possa ser representado por

$$\mathbf{x}_i = \mathbf{H}\mathbf{s}_i + \mathbf{n}_i, \quad (1)$$

em que $\{\mathbf{x}_i\} = \{\mathbf{x}(t_i)\}$ denota observações m -dimensionais do sinal recebido, \mathbf{H} é a matriz de canal de ordem $m \times p$, com p colunas linearmente independentes contendo m elementos. Os elementos $\{h_{kl}\}$ de \mathbf{H} , $k = 1, 2, \dots, m$ e $l = 1, 2, \dots, p$, representam o ganho do canal entre o l -ésimo transmissor (ou fonte) e o k -ésimo sensor (elemento em um arranjo de antenas ou receptor com antena simples), como ilustrado na Figura 1. O conjunto $\{\mathbf{s}_i\}$, de ordem $p \times 1$, representa sinais aleatórios estacionários de média nula. Por fim, $\{\mathbf{n}_i\}$, de ordem $m \times 1$, representa os vetores de ruído aditivo Gaussiano $\mathcal{N}(0, \sigma^2 \mathbf{I})$, com \mathbf{I} sendo uma matriz identidade de ordem m .

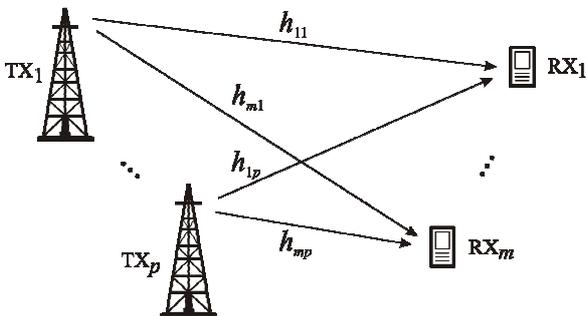


Figura 1: Representação dos ganhos na matriz de canal \mathbf{H} em um cenário com p transmissores e m receptores.

Seja \mathbf{R} a matriz de covariância do sinal recebido, a qual tem sua estimativa de máxima verossimilhança dada pela média amostral (*sample covariance matrix*)

$$\hat{\mathbf{R}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\dagger = \frac{1}{n} \mathbf{X} \mathbf{X}^\dagger, \quad (2)$$

em que \dagger denota conjugado transposto e \mathbf{X} , de ordem $m \times n$, é a matriz formada arranjando-se lado a lado os n vetores do conjunto $\{\mathbf{x}_i\}$. Seja \mathbf{S} , de ordem $p \times n$, a matriz formada arranjando-se lado a lado os n vetores $\{\mathbf{s}_i\}$. Se \mathbf{N} , de ordem $m \times n$, é a matriz formada arranjando-se lado a lado os n vetores $\{\mathbf{n}_i\}$, então se pode escrever que $\mathbf{X} = \mathbf{H}\mathbf{S} + \mathbf{N}$.

Devido à média nula do ruído e sua independência do sinal, pode-se facilmente verificar que \mathbf{R} pode ser assim escrita [2]:

$$\mathbf{R} = E[\mathbf{X}\mathbf{X}^\dagger] = \mathbf{H}\mathbf{R}_s\mathbf{H}^\dagger + \sigma^2\mathbf{I}, \quad (3)$$

em que \mathbf{R}_s é a matriz de covariância do sinal transmitido, ou seja, $\mathbf{R}_s = E[\mathbf{S}\mathbf{S}^\dagger]$. Da teoria de álgebra linear, se a matriz de canal \mathbf{H} tiver todas as p colunas linearmente independentes (*full rank*), o posto (*rank*) da matriz $\mathbf{H}\mathbf{R}_s\mathbf{H}^\dagger$ será p ou, equivalentemente, os $m - p$ menores autovalores de $\mathbf{H}\mathbf{R}_s\mathbf{H}^\dagger$ serão nulos. Como os autovalores de \mathbf{R} são os autovalores de $\mathbf{H}\mathbf{R}_s\mathbf{H}^\dagger$ mais os autovalores de $\sigma^2\mathbf{I}$, então os $m - p$ menores autovalores de \mathbf{R} serão iguais a σ^2 . Note então que é possível se estimar o número de fontes p determinando-se a multiplicidade dos menores autovalores de \mathbf{R} .

Na prática, entretanto, não se conhece a matriz de covariância \mathbf{R} , mas apenas sua estimativa $\hat{\mathbf{R}}$, a qual é obtida a partir de um número finito n de amostras, conforme [2]. Neste caso os menores autovalores de $\hat{\mathbf{R}}$ serão diferentes com probabilidade 1, e a estimação do número de fontes meramente “observando” os autovalores se torna tarefa bastante complexa [2].

A classificação dos autovalores de $\hat{\mathbf{R}}$ em dois grupos que permitam identificar quantos são os maiores (ou menores) corresponde justamente ao maior problema na estimação do número de fontes por meio da estimativa da matriz de covariância do sinal recebido. Para ilustrar esse problema, a Figura 2 mostra os gráficos dos autovalores $\{\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m\}$ da matriz de covariância do sinal recebido, $\hat{\mathbf{R}}$, normalizados em relação ao máximo autovalor, para valores de relação sinal-ruído (SNR, *signal-to-noise ratio*) de 0 dB e -15 dB. O número de fontes é $p = 5$, o número de sensores é $m = 30$, o número de amostras por receptor (sensor) é $n = 1000$ e a potência total do sinal transmitido é unitária. Perceba que os $m - p = 25$ menores autovalores estão nitidamente separados dos $p = 5$ maiores, possibilitando uma estimação correta do número de fontes p . Já para uma SNR = -15 dB, correspondente a uma relação sinal-ruído bem menor, é praticamente impossível notar a separação entre os p maiores e os $m - p$ menores autovalores, o que significa que a estimação do número de fontes nessa situação levaria a uma alta probabilidade de erro.

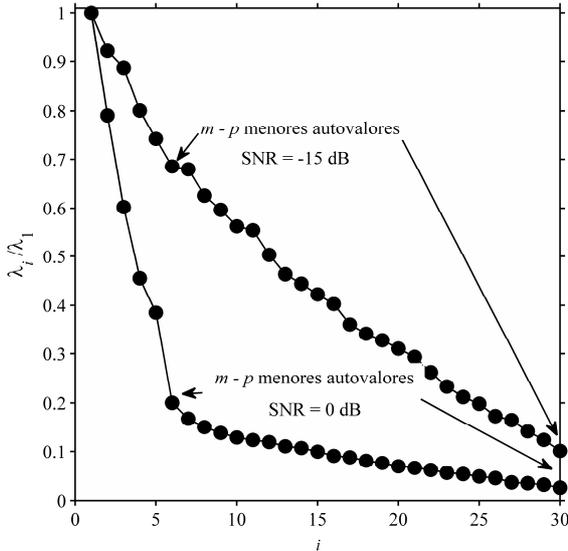


Figura 2: Classificação dos autovalores da matriz de covariância estimada para SNR = 0 dB e SNR = -15 dB .

III. FUNDAMENTOS

Nesta seção são apresentados alguns fundamentos sobre a teoria da estimação de ordem (ou seleção) de modelo. Não se pretende aqui cobrir com profundidade todos os conceitos e desenvolvimentos matemáticos, mas apenas estabelecer a conexão entre eles de forma que o leitor possa conseguir entender as origens de cada um dos métodos de estimação discutidos. Espera-se também que os fundamentos aqui apresentados sirvam para que o leitor tenha condições de estudar tais métodos de forma mais aprofundada, ou até mesmo estudar outros métodos não cobertos neste artigo.

A. Estimação de Ordem de Modelo pelo Critério de Máxima Verossimilhança

Considere um vetor m -dimensional \mathbf{x} observado em um determinado instante de tempo discreto. O critério de máxima verossimilhança pode ser utilizado para se obter a estimativa de um determinado parâmetro a partir desse vetor [12][13]. Considerando \mathbf{x} uma variável aleatória (multidimensional) associada a um sinal recebido e $\boldsymbol{\theta}$ um vetor de parâmetros desse sinal, a função densidade de probabilidade (FDP) de \mathbf{x} pode ser representada por $f(\mathbf{x}, \boldsymbol{\theta})$, conhecida como função de verossimilhança do vetor \mathbf{x} em função dos parâmetros em $\boldsymbol{\theta}$ [2][3][12]. Seja o vetor \mathbf{x} representado por [12]

$$\mathbf{x} = \boldsymbol{\mu}(\boldsymbol{\gamma}) + \boldsymbol{\eta}, \quad (4)$$

em que $\boldsymbol{\gamma}$ é um vetor desconhecido, $\boldsymbol{\mu}(\boldsymbol{\gamma})$ é uma função determinística de $\boldsymbol{\gamma}$ e $\boldsymbol{\eta}$ representa o ruído Gaussiano e branco (AWGN, *additive white Gaussian noise*) de média zero e matriz de covariância igual a $\sigma^2 \mathbf{I}$. Então se pode escrever $f(\mathbf{x}, \boldsymbol{\theta})$ como [12]

$$f(\mathbf{x}, \boldsymbol{\theta}) = \frac{1}{(2\pi)^{m/2} (\sigma^2)^{m/2}} e^{-\frac{\|\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\gamma})\|^2}{2\sigma^2}}, \quad (5)$$

em que

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\gamma} \\ \sigma^2 \end{bmatrix}. \quad (6)$$

Como $\boldsymbol{\theta}$, a princípio, é um vetor de parâmetros desconhecidos, representa-se por $\hat{\boldsymbol{\theta}}$ a sua estimativa. A aplicação do critério de máxima verossimilhança consiste em encontrar o conjunto de parâmetros que maximiza $f(\mathbf{x}, \boldsymbol{\theta})$ ou $\ln f(\mathbf{x}, \boldsymbol{\theta})$, já que a função logarítmica é uma função monotônica crescente. A estimativa de máxima verossimilhança do vetor $\boldsymbol{\theta}$ pode assim ser escrita [12]:

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \ln f(\mathbf{x}, \boldsymbol{\theta}). \quad (7)$$

Em muitas aplicações, não somente os valores dos parâmetros estimados a partir do vetor observado são de interesse, mas também a dimensão do vetor de parâmetros $\boldsymbol{\theta}$. Esta dimensão é usualmente denominada de *ordem do modelo* em análise e pode estar associada, por exemplo, ao número de fontes de sinais incidindo sobre um arranjo de sensores. É dessa forma que o critério de máxima verossimilhança está conectado aos métodos de seleção ou de determinação de ordem de modelo [12].

A seguir, o problema de seleção de ordem de modelo é formulado como um teste de hipóteses, para o qual se verifica que o critério do máximo *a posteriori* (MAP) pode ser considerado ótimo. Em seguida discute-se sobre o critério de informação de Kullback-Leibler (KL), o qual direciona para uma abordagem alternativa para a elaboração de regras de seleção de ordem de modelo baseadas na teoria da informação.

B. Estimação de Ordem de Modelo pelo Critério do Máximo a Posteriori (MAP)

Seja \mathcal{H}_k a hipótese de que a ordem do modelo em análise é k e seja K um limitante superior para essa ordem. Considere ainda que as hipóteses $\{\mathcal{H}_k\}_{k=1}^K$ são mutuamente excludentes, ou seja, apenas uma delas pode ser considerada como verdadeira por vez. O critério MAP é aquele que maximiza a probabilidade *a posteriori* da hipótese \mathcal{H}_k , $P(\mathcal{H}_k | \mathbf{x})$, e alternativamente pode ser assim descrito [12]:

$$\max_k f(\mathbf{x} | \mathcal{H}_k) P(\mathcal{H}_k), \quad (8)$$

em que $P(\mathcal{H}_k)$ é a probabilidade *a priori* de \mathcal{H}_k . Se as hipóteses são equiprováveis, o critério MAP se reduz a

$$\max_k f(\mathbf{x} | \mathcal{H}_k). \quad (9)$$

No contexto de estimação de parâmetros, quando o parâmetro estimado é um inteiro, se costuma dizer que se trata de um problema de *deteção*. Sendo assim, quando se estima o parâmetro de interesse atribui-se ao processo de estimação uma *probabilidade de detecção correta*. Considere que, para o problema de estimação de ordem de modelo, a probabilidade de detecção correta P_c seja aquela obtida pela média sobre

todas as hipóteses. Pode-se mostrar [12] que o critério MAP minimiza P_c e que, portanto, trata-se de um critério ótimo para determinação de ordem de modelo.

Aspectos de realização prática do critério MAP são discutidos mais adiante, na Subseção III-D.

C. Distância de Kullback-Leibler (KL)

A distância de Kullback-Leibler é uma medida de discrepância entre duas FDPs, a verdadeira e a estimada, por exemplo. Definindo a FDP verdadeira como $f(\mathbf{x}, \boldsymbol{\theta})$ e a FDP estimada como $\hat{f}(\mathbf{x}, \boldsymbol{\theta})$, a distância de Kullback-Leibler é definida como [12]

$$D[f(\mathbf{x}, \boldsymbol{\theta}), \hat{f}(\mathbf{x}, \boldsymbol{\theta})] = \int f(\mathbf{x}, \boldsymbol{\theta}) \ln \frac{f(\mathbf{x}, \boldsymbol{\theta})}{\hat{f}(\mathbf{x}, \boldsymbol{\theta})} d\mathbf{x}. \quad (10)$$

Pode-se mostrar [12] que $D[f(\mathbf{x}, \boldsymbol{\theta}), \hat{f}(\mathbf{x}, \boldsymbol{\theta})] \geq 0$ e que $D[f(\mathbf{x}, \boldsymbol{\theta}), \hat{f}(\mathbf{x}, \boldsymbol{\theta})] = 0$ se e somente se $f(\mathbf{x}, \boldsymbol{\theta}) = \hat{f}(\mathbf{x}, \boldsymbol{\theta})$.

A distância de KL é também conhecida como *entropia relativa*. Como, segundo a *teoria da informação*, a entropia é uma medida de quantidade de informação, então se pode associar a discrepância medida pela distância de KL à perda de informação pelo uso de $\hat{f}(\mathbf{x}, \boldsymbol{\theta})$ em vez de $f(\mathbf{x}, \boldsymbol{\theta})$. Por essa razão a distância de KL é às vezes denominada de função de *informação de KL* e os métodos de determinação de ordem de modelo dela derivados são usualmente denominados de *métodos baseados na teoria da informação*.

Aspectos de realização prática da informação de KL são discutidos na próxima subseção.

D. Aplicação do critério MAP e da informação de KL na estimação de ordem de modelo

É importante ressaltar que nem o critério MAP nem a informação de KL podem ser utilizados diretamente para estimação de ordem de modelo, pois as FDPs relacionadas ao vetor de dados sob as várias hipóteses, $f(\mathbf{x} | \mathcal{H}_k)$, ou a FDP verdadeira dos dados, $f(\mathbf{x}, \boldsymbol{\theta})$, são desconhecidas.

Uma alternativa para se utilizar a abordagem MAP consiste em se atribuir uma FDP *a priori* para o vetor desconhecido $\boldsymbol{\theta}$, integrando $f(\mathbf{x}, \boldsymbol{\theta} | \mathcal{H}_k)$ em relação a essa FDP para se encontrar a FDP associada a cada hipótese, $f(\mathbf{x} | \mathcal{H}_k)$.

Em relação à informação de KL, uma possibilidade de utilizá-la na estimação de ordem do modelo consiste em utilizar uma estimativa da distância de KL, denotada por $\hat{D}[f(\mathbf{x}, \boldsymbol{\theta}), \hat{f}(\mathbf{x}, \boldsymbol{\theta})]$, escolhendo-se uma FDP $\hat{f}(\mathbf{x}, \boldsymbol{\theta})$ que seja adequada ao modelo. A determinação da ordem do modelo é então realizada pela minimização de $\hat{D}[f(\mathbf{x}, \boldsymbol{\theta}), \hat{f}(\mathbf{x}, \boldsymbol{\theta})]$.

No caso particular da informação de KL, note que a expressão (10) pode ser reescrita como

$$\begin{aligned} D[f(\mathbf{x}, \boldsymbol{\theta}), \hat{f}(\mathbf{x}, \boldsymbol{\theta})] &= E \left[\ln \frac{f(\mathbf{x}, \boldsymbol{\theta})}{\hat{f}(\mathbf{x}, \boldsymbol{\theta})} \right] \\ &= E[\ln f(\mathbf{x}, \boldsymbol{\theta})] - E[\ln \hat{f}(\mathbf{x}, \boldsymbol{\theta})], \end{aligned} \quad (11)$$

em que o valor esperado é tomado com relação à FDP real dos dados observados, $f(\mathbf{x}, \boldsymbol{\theta})$. Note que somente a parcela da

direita em (11) é dependente do modelo. Então, maximizar $\hat{D}[f(\mathbf{x}, \boldsymbol{\theta}), \hat{f}(\mathbf{x}, \boldsymbol{\theta})]$ é equivalente a maximizar $E[\ln \hat{f}(\mathbf{x}, \boldsymbol{\theta})]$, a chamada informação relativa de Kullback-Leibler.

A escolha ideal para $\hat{f}(\mathbf{x}, \boldsymbol{\theta})$ de forma que $E[\ln \hat{f}(\mathbf{x}, \boldsymbol{\theta})]$ seja maximizada é $f(\mathbf{x}, \boldsymbol{\theta})$. Como esta função não está disponível, uma escolha razoável seria aquela obtida com os parâmetros estimados, ou seja, $f(\mathbf{x}, \hat{\boldsymbol{\theta}})$. No entanto, ainda não seria possível computar o valor esperado $E[\ln f(\mathbf{x}, \hat{\boldsymbol{\theta}})]$, já que $f(\mathbf{x}, \hat{\boldsymbol{\theta}})$, a FDP real dos dados sobre a qual o valor esperado seria tomado, não é conhecida. Como saída poderíamos utilizar o estimador não polarizado trivial de $E[\ln f(\mathbf{x}, \hat{\boldsymbol{\theta}})]$, que é $\ln f(\mathbf{x}, \hat{\boldsymbol{\theta}})$. Entretanto, como ressaltado em [12], a ordem de modelo obtida a partir da maximização de $\ln f(\mathbf{x}, \hat{\boldsymbol{\theta}})$ não terá propriedades interessantes, podendo falhar completamente em alguns casos, principalmente naqueles em que hipóteses estão aninhadas, ou seja, quando o modelo associado a uma hipótese é um caso particular de um modelo associado a outra hipótese.

Uma primeira abordagem para a solução do problema supracitado consiste em aproximar a FDP desconhecida do modelo, $f(\mathbf{x}, \boldsymbol{\theta})$, por uma expansão de Taylor em torno do vetor de parâmetros estimados, $\hat{\boldsymbol{\theta}}$, usando então essa aproximação como $\hat{f}(\mathbf{x}, \boldsymbol{\theta})$ no problema de maximização da informação relativa de KL, $E[\ln \hat{f}(\mathbf{x}, \boldsymbol{\theta})]$. Como resultado, a ordem de modelo k (dimensão do vetor $\hat{\boldsymbol{\theta}}$) estimada será aquela que minimiza [12]

$$-2 \ln f(\mathbf{x}, \hat{\boldsymbol{\theta}}) + k. \quad (12)$$

Verificou-se em [12] que a ordem do modelo obtida por meio da minimização de (12) tende a ser superestimada porque uma redução no primeiro termo tende a ser pouco compensada pelo aumento no segundo termo. Em outras palavras, o segundo termo em (12) penaliza a sobrestimação, mas não o faz a contento. Mais genericamente, o termo de penalização (ou termo de correção) é então responsável por tornar não polarizada a estimativa da distância média de KL entre a densidade real do modelo, $f(\mathbf{x}, \boldsymbol{\theta})$, e a densidade estimada $f(\mathbf{x}, \hat{\boldsymbol{\theta}})$.

Como visto na Subseção III-A, a densidade $f(\mathbf{x}, \hat{\boldsymbol{\theta}})$ pode ser determinada a partir da estimativa de máxima verossimilhança $\hat{\boldsymbol{\theta}}$ do vetor de parâmetros desconhecidos. Para o caso específico da estimação da ordem do modelo aplicada à estimação do número de fontes, $f(\mathbf{x}, \hat{\boldsymbol{\theta}})$, ou $\ln f(\mathbf{x}, \hat{\boldsymbol{\theta}})$, é uma função dos autovalores $\{\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m\}$ da matriz de covariância do sinal recebido. Admitindo ruído Gaussiano, sinais fonte com distribuição Gaussiana e matriz de canal com posto máximo (*full rank*), tal função de log-verossimilhança terá a forma [2]

$$\ln f(\mathbf{x}, \hat{\boldsymbol{\theta}}) = n \ln \frac{\prod_{i=k+1}^m \lambda_i}{\left(\frac{1}{m-k} \sum_{i=k+1}^m \lambda_i \right)^{m-k}}. \quad (13)$$

Pode-se verificar [2] que o vetor $\hat{\theta}$ terá $k+1+2mk$ parâmetros, dos quais $k(2m-k)+1$ são parâmetros livres para escolha. Portanto, de acordo com (12), o número de fontes estimado será o valor de k que minimiza

$$-2n \ln \frac{\prod_{i=k+1}^m \lambda_i}{\left(\frac{1}{m-k} \sum_{i=k+1}^m \lambda_i \right)^{m-k}} + \text{ termo de correção.} \quad (14)$$

A seguir verifica-se que os critérios AIC, BIC e MDL tem uma forma muito similar àquela representada pela expressão (14). Basicamente o que difere tais critérios é forma como cada um penaliza a sobrestimação da ordem do modelo, o que é determinado pelo termo de correção, que será uma função do número de parâmetros livres $k(2m-k)+1$.

IV. CRITÉRIO DE INFORMAÇÃO DE AKAIKE (AIC)

O método de estimação de ordem de modelo por meio do critério de informação de Akaike (AIC, *Akaike information criterion*) foi proposto por Akaike em 1973 [14][15]. Derivado dos conceitos de minimização da informação de Kullback-Leibler, o método AIC consiste em selecionar a ordem do modelo que minimiza [2]

$$-2 \ln f(\mathbf{x}, \hat{\theta}) + 2\kappa. \quad (15)$$

O primeiro termo do estimador AIC corresponde ao primeiro termo de (12) e o termo de correção κ corresponde ao número de parâmetros livres em θ . Substituindo em (15) a função de log-verossimilhança de (13) e o número de parâmetros livres $k(2m-k)+1$ de θ para o problema de estimação do número de fontes, o critério de informação de Akaike passa a ser escrito como

$$\text{AIC}(k) = -2n \ln \frac{\prod_{i=k+1}^m \lambda_i}{\left(\frac{1}{m-k} \sum_{i=k+1}^m \lambda_i \right)^{m-k}} + 2[k(2m-k)]. \quad (16)$$

O número de fontes estimado, \hat{p} , será o valor de k que minimiza $\text{AIC}(k)$, para $k = 0, 1, \dots, m-1$.

Todos os gráficos apresentados deste ponto em diante foram obtidos por meio de um mínimo de 3000 eventos de Monte Carlo utilizando o programa Matlab®. O código fonte desenvolvido está reproduzido no Apêndice.

O desempenho do estimador AIC é mostrado na Figura 3, considerando SNR variável, potência de transmissão unitária, número de fontes $p = 5$ e $p = 15$, $n = 500$ amostras por sensor e $m = 30$ sensores. As curvas P_c e P_{se} representam as probabilidades de detecção correta e de sobrestimação do número de fontes, respectivamente. Note que, como esperado, o aumento da SNR eleva a probabilidade de detecção correta. No entanto, verifica-se que é necessária uma SNR maior caso

se deseje alta probabilidade de detecção correta para um número maior de fontes. Tais comportamentos não são particulares do AIC, ressalta-se. Observe também na Figura 3 que, mesmo para altos valores de SNR, o estimador AIC não atinge $P_c = 1$, o que é uma característica intrínseca a este estimador (como mostrado mais adiante, esse comportamento é causado porque o estimador AIC não é *consistente*, apresentando sempre uma probabilidade não nula de sobrestimação). Note também que, para altos valores de SNR, a probabilidade de sobrestimação $P_{se} \rightarrow (1 - P_c)$, o que significa que, quando o estimador comete um erro, comete-o estimando um número de fontes maior que o real (p). Isto não ocorre para valores mais baixos de SNR, ou seja, embora haja erros de sobrestimação nesta situação, a maior parte deles é de subestimação: note na Figura 3 que em regimes de baixa SNR tem-se $(1 - P_c) \gg P_{se}$.

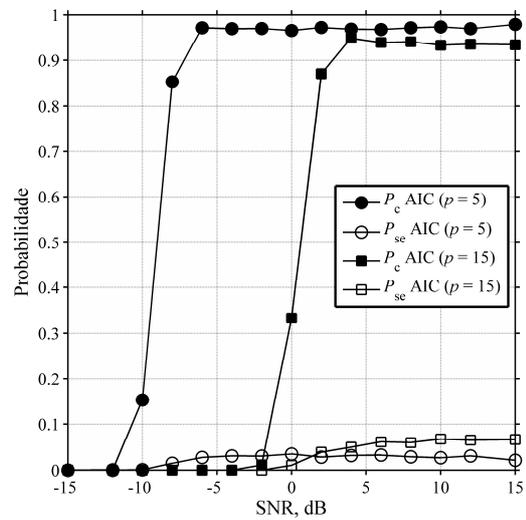


Figura 3: Desempenho do estimador AIC para SNR variável.

A Figura 4 mostra o desempenho do estimador AIC para o número de fontes variável, SNR = 0 dB e -5 dB, $n = 1000$ amostras por sensor e $m = 30$ sensores. Note que, para uma SNR fixa, a probabilidade de detecção correta decresce com o aumento do número de fontes. Note também que o máximo número de fontes estimado com alta probabilidade de detecção correta decresce à medida que a SNR é reduzida. Observe novamente que P_c não atinge o valor 1 devido à probabilidade não nula de sobrestimação produzida pelo estimador AIC.

A Figura 5 mostra o desempenho do estimador AIC para o número de amostras variável, SNR = -2 dB e -4 dB, número de fontes $p = 10$ e $m = 30$ sensores. Por meio desta figura é possível notar que, de fato, o estimador AIC é *inconsistente*, ou seja, o valor estimado não converge para o valor real do parâmetro estimado (p) à medida que se aumenta o número de amostras colhidas (n). Observe também que valores mais baixos de SNR pressupõem maior número de amostras para que se atinja uma determinada probabilidade de detecção correta. Esta não é, contudo, uma característica particular do estimador AIC.

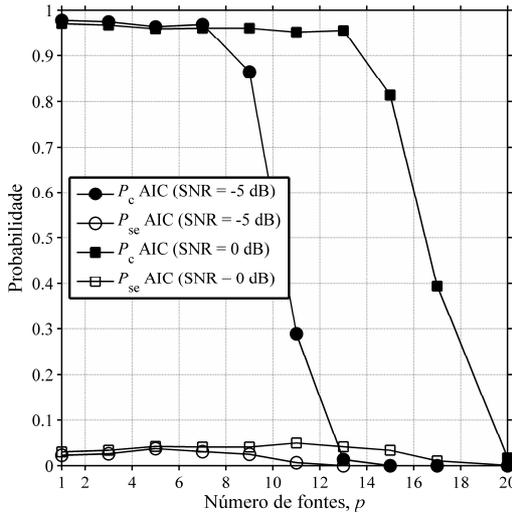


Figura 4: Desempenho do estimador AIC para número de fontes variável.

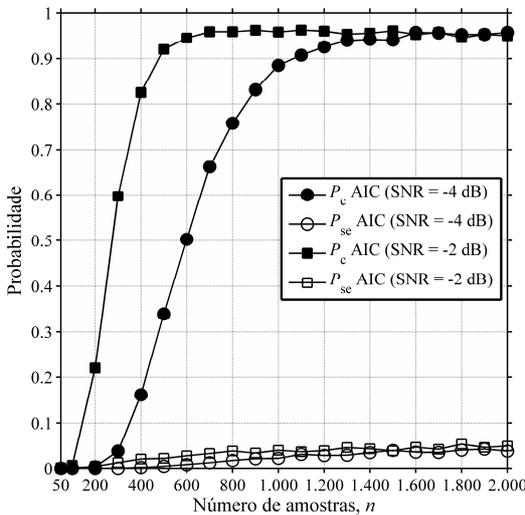


Figura 5: Desempenho do estimador AIC para número de amostras variável.

V. CRITÉRIO DE INFORMAÇÃO BAYESIANO (BIC)

O critério de informação Bayesiano (BIC, *Bayesian information criterion*), também conhecido como critério de Schwarz, foi proposto por Schwarz em 1978 [16].

Tanto o AIC quanto o BIC se baseiam na minimização da informação de KL. No entanto, diferentemente do AIC, em uma das formas de desenvolvimento matemático do BIC utiliza-se a FDP *a priori* do parâmetro a ser estimado, $f(\theta)$, de forma análoga ao que acontece no critério MAP com as probabilidades *a priori* das hipóteses. O uso de $f(\theta)$ se faz necessário devido à adoção de uma aproximação ligeiramente diferente daquela adotada no AIC na aproximação da FDP desconhecida do modelo em torno do vetor de parâmetros estimados.

Assim como no MAP, o teorema de Bayes é invocado no desenvolvimento matemático do BIC de forma a relacionar devidamente as probabilidades *a priori* e *a posteriori*. Essa é a razão da associação do nome “Bayesiano” ao BIC.

Contrastando com o estimador AIC, o estimador BIC é *consistente*, ou seja, a probabilidade de detecção correta tende

a 1 com o aumento no número de amostras. Matematicamente, $P_c \rightarrow 1$ se $n \rightarrow \infty$. Portanto, assim como o MAP, o estimador BIC é considerado ótimo para o problema de seleção de ordem de modelo.

O método BIC consiste em selecionar a ordem do modelo que minimiza [12]

$$-2 \ln f(\mathbf{x} | \hat{\theta}) + \kappa \ln n. \tag{17}$$

Note que o primeiro termo do estimador BIC também corresponde ao primeiro termo de (12) e o termo de correção que corresponde ao número de parâmetros livres em θ tem um fator $\ln n$ extra. Substituindo em (17) a função de log-verossimilhança de (13) e o número de parâmetros livres $k(2m-k)+1$ de θ para o problema de estimação no número de fontes, o critério de informação Bayesiano passa a ser escrito como

$$BIC(k) = -2n \ln \frac{\prod_{i=k+1}^m \lambda_i}{\left(\frac{1}{m-k} \sum_{i=k+1}^m \lambda_i \right)^{m-k}} + k(2m-k) \ln n. \tag{18}$$

O número de fontes estimado, \hat{p} , será o valor de k que minimiza $BIC(k)$, para $k = 0, 1, \dots, m - 1$.

O desempenho do estimador BIC é mostrado na Figura 6, considerando SNR variável, potência de transmissão unitária, $p = 5$ e $p = 15$, $n = 1000$ e $m = 30$. Observe novamente que o aumento da SNR traduz-se em aumento da probabilidade de detecção correta e que é necessária uma SNR maior caso se deseje alta probabilidade de detecção correta para um número maior de fontes. Em contraste com o estimador AIC, observe na Figura 6 que o estimador BIC atinge $P_c = 1$, o que ocorre porque o estimador BIC é *consistente*. Note também que a probabilidade de sobrestimação é sempre nula, ou seja, quando o estimador BIC comete um erro comete-o sempre estimando um número de fontes menor que p .

A Figura 7 mostra o desempenho do estimador BIC para o número de fontes variável, SNR = 0 dB e -5 dB, $n = 1000$ e $m = 30$. Assim como no caso do AIC, a probabilidade de detecção correta decresce com o aumento do número de fontes e o máximo número de fontes estimado com alta probabilidade de detecção correta decresce à medida que a SNR é reduzida. Observe novamente que P_c pode atingir o valor 1 e que $P_{se} = 0$ em qualquer situação.

A Figura 8 mostra o desempenho do estimador BIC para o número de amostras variável, SNR = 0 dB e 2 dB, $p = 10$ e $m = 30$. Por meio desta figura é possível notar que, de fato, o estimador BIC é *consistente*, ou seja, a probabilidade de detecção correta converge para 1 à medida que se aumenta o número de amostras. Observe também que, assim como no caso do AIC, valores mais baixos de SNR pressupõem maior número de amostras para que se atinja uma determinada probabilidade de detecção correta com o BIC.

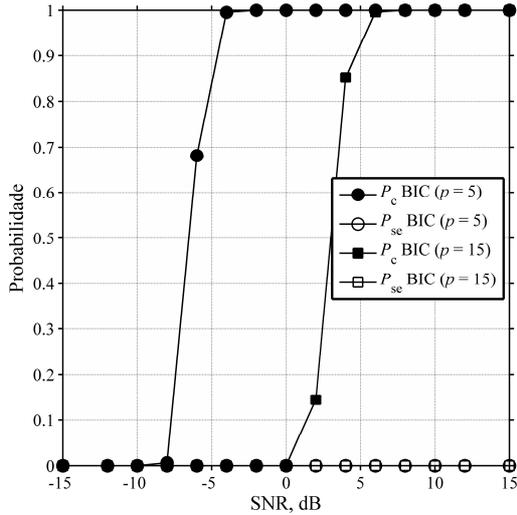


Figura 6: Desempenho do estimador BIC para SNR variável.

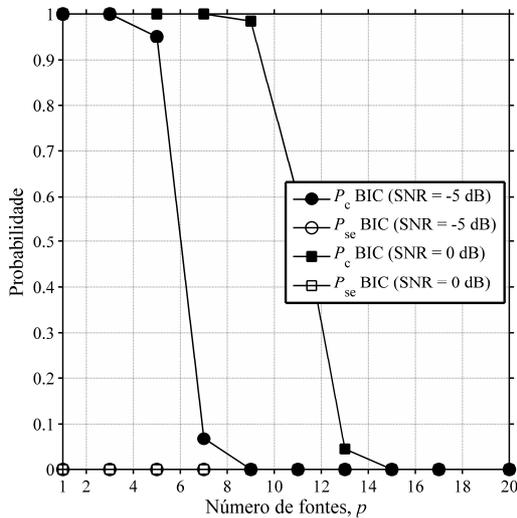


Figura 7: Desempenho do estimador BIC para número de fontes variável.

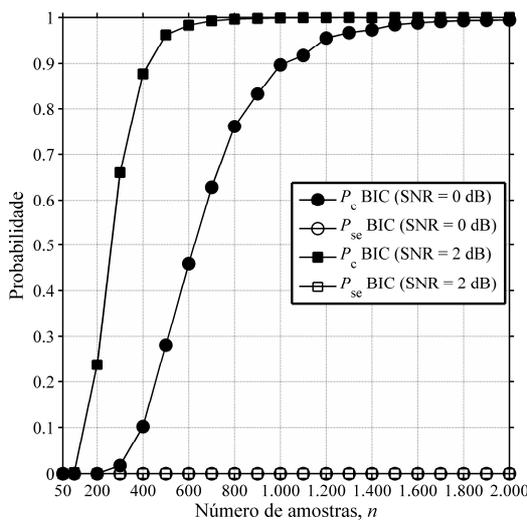


Figura 8: Desempenho do estimador BIC para número de amostras variável.

VI. MÉTODO DO COMPRIMENTO DE MÍNIMA DESCRIÇÃO (MDL)

O método do comprimento de mínima descrição (MDL, *minimum description length*) foi desenvolvido de forma independente por Schwarz [16] e Rissanen [17] em 1978. O desenvolvimento do método MDL também utiliza conceitos relacionados à teoria da informação. Particularmente, Schwarz adotou uma abordagem Bayesiana, similar àquela utilizada no desenvolvimento do BIC. Rissanen, no entanto, se baseou em aspectos de compressão de informação. Resumidamente, Rissanen notou que cada modelo em julgamento pode ser interpretado como uma forma de codificação de fonte (compressão) dos dados observados. O modelo selecionado é então aquele que mais comprime tais dados, ou seja, é aquele que minimiza o comprimento médio do código. O nome *minimum description length* advém dessa observação.

Embora derivado a partir de diferentes abordagens, o estimador MDL obtido por Schwarz e por Rissanen mostrou-se ser o mesmo. Mais que isso, em alguns casos o estimador MDL é equivalente ao BIC. Essa equivalência é válida somente se, no desenvolvimento de ambos os estimadores, considera-se que o número de parâmetros estimados é fixo. Caso seja variável ou crescente à medida que o número de amostras aumenta, os modelos para os estimadores BIC e MDL se tornam distintos, podendo estimar ordens de modelo bastante distintas [18].

No presente contexto os critérios para equivalência entre MDL e BIC são satisfeitos e, portanto, o estimador MDL tem a mesma expressão e o mesmo desempenho do BIC. Então, conforme (18),

$$\text{MDL}(k) = -2n \ln \frac{\prod_{i=k+1}^m \lambda_i}{\left(\frac{1}{m-k} \sum_{i=k+1}^m \lambda_i \right)^{m-k}} + k(2m-k) \ln n. \quad (19)$$

Assim, o número de fontes estimado, \hat{p} , será o valor de k que minimiza $\text{MDL}(k)$, para $k = 0, 1, \dots, m-1$.

VII. COMPARAÇÃO ENTRE AIC E BIC/MDL

Uma primeira comparação entre o AIC e o MDL/BIC é mostrada na Figura 9, considerando SNR variável, potência de transmissão unitária, $p = 5$, $n = 1000$ e $m = 30$. Nota-se que o MDL/BIC apresenta pior desempenho que o AIC em regime de baixa SNR. No entanto, a inferioridade do MDL/BIC deixa de existir em altos valores de SNR, pois mesmo produzindo altos valores de probabilidade de acerto, o AIC resulta em uma probabilidade de sobrestimação não desprezível, fazendo com que a probabilidade de acerto nunca seja 1.

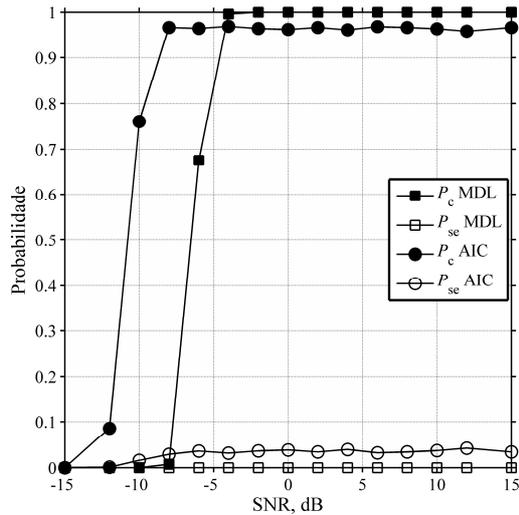


Figura 9: AIC versus MDL/BIC para SNR variável.

A Figura 10 mostra uma comparação entre o AIC e o MDL/BIC para o número de fontes variável, SNR = 0 dB, $n = 1000$ e $m = 30$. O AIC se mostra superior em termos do máximo número de fontes estimado a partir de $p = 10$, inclusive. No entanto, não consegue atingir probabilidade de detecção correta máxima para $p < 9$, também devido ao fato de apresentar uma probabilidade de sobrestimação não desprezível.

A Figura 11 mostra uma comparação entre o AIC e o MDL/BIC para o número de amostras variável, SNR = -5 dB, $p = 5$ e $m = 30$. Percebe-se que o AIC consegue proporcionar altas probabilidades de detecção correta com um número de amostras significativamente inferior ao necessário no MDL/BIC. No entanto, a probabilidade de detecção correta do estimador AIC nunca atinge o valor 1, posto que tal estimador não é consistente, apresentando uma probabilidade de sobrestimação irredutível mesmo com o aumento do número de amostras.

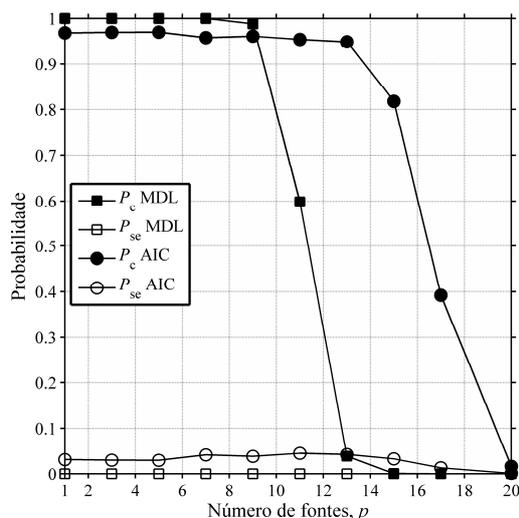


Figura 10: AIC versus MDL/BIC para número de fontes variável.

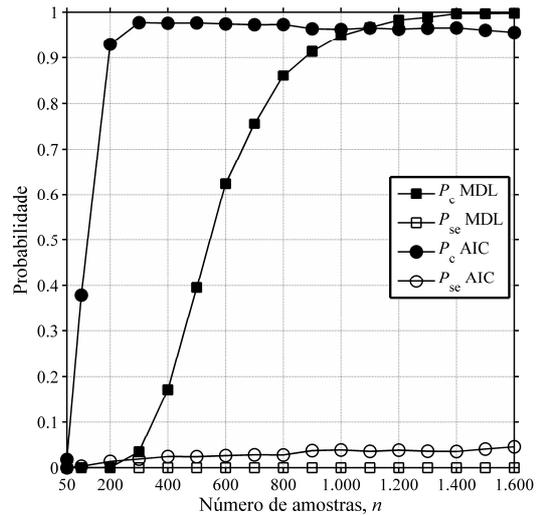


Figura 11: AIC versus BIC/MDL para o número de amostras variável.

VIII. CONCLUSÕES

Este tutorial abordou três métodos de estimação do número de fontes baseados na teoria da informação: o critério de informação de Akaike (AIC), o método do comprimento de mínima descrição (MDL) e o critério de informação Bayesiana (BIC). Alguns conceitos fundamentais também foram fornecidos, objetivando facilitar o entendimento desses métodos e tornar o tutorial mais completo e até certo ponto independente da consulta a outras referências em um primeiro momento. Ainda assim, um grande número de referências é listado a seguir, de forma que o leitor possa procurar por informações mais detalhadas caso assim deseje.

Como há outros métodos de estimação do número de fontes e este tutorial está restrito aos métodos AIC, BIC e MDL, pode-se considerar este trabalho como um primeiro contato do leitor com a formulação do problema e com algumas formas de resolvê-lo, servido assim como base para o entendimento de outros métodos.

Em termos teóricos, observou-se que, a despeito de utilizarem abordagens não totalmente iguais em seus modelos matemáticos, os estimadores AIC, BIC e MDL possuem expressões finais bastante parecidas: todas contêm uma parcela referente à estimação de máxima verossimilhança do vetor de parâmetros desconhecidos e uma parcela de correção cuja função é regular tendências de sobrestimação da ordem do modelo em análise. Os métodos diferem na forma como realizam tal correção.

Foi verificado que, para as condições adotadas na estimação do número de fontes, os modelos referentes ao MDL e ao BIC são idênticos, produzindo, portanto, desempenhos idênticos.

Comparando os estimadores AIC e MDL/BIC, observou-se que, por ser consistente, o MDL pode proporcionar probabilidade de detecção correta igual a 1, dependendo dos parâmetros sistêmicos. No entanto, percebeu-se que o AIC pode suplantá-lo em determinadas situações, principalmente em regime de baixa SNR, maior número de fontes e pequeno número de amostras.

Pode-se também concluir que o AIC pode ser preferido se a situação em que for aplicado permitir que a probabilidade de detecção correta seja um pouco menor que 1.

APÊNDICE

O programa em Matlab apresentado a seguir tem como objetivo simular a detecção do número de fontes utilizando as técnicas AIC, BIC e MDL descritas neste tutorial. Os parâmetros de entrada são m , n , N (eventos de Monte Carlo), p e SNR. Na linha de comando `sel = 'SNR'` selecione o parâmetro que se deseja variar. No exemplo, a SNR é o parâmetro variável.

```
m=30; % número de sensores
n=1000; % número de amostras por sensor
N=3000; % número de eventos de Monte Carlo
p=5; % número de fontes
SNR=0; % Relação sinal-ruído em dB

sel = 'SNR';

switch (sel)

    case 'SNR'
        Parametro = [-15 -12 -10 -8 -6 -4 -2 0 2 4 6
8 10 12 15]'; % para SNR %variável
    case 'Nr Sensores'
        Parametro=[1 5 10 15 20 25 30 35 40 45 50]';
% para m variável
    case 'Nr Fontes'
        Parametro=[1 3 5 7 9 11 13 15 17 20]'; %
para p variável
    case 'Nr Amostras'
        Parametro=[5 50 100 200 300 400 400 500 600
700 800]'; % para n variável
end

Npt=size(Parametro,1); % número de pontos %no
gráfico = número de valores do %parâmetro variável

for pt=1:Npt

    switch (sel)
        case 'SNR'
            SNR = Parametro(pt);
        case 'Nr Sensores'
            m = Parametro(pt);
        case 'Nr Fontes'
            p = Parametro(pt);
        case 'Nr Amostras'
            n = Parametro(pt);
    end

    CcAIC=0; % contador de acertos de estimação
    CseAIC=0; % contador de sobrestimação
    CcBIC=0; % contador de acertos de estimação
    CseBIC=0; % contador de sobrestimação
    CcMDL=0; % contador de acertos de estimação
    CseMDL=0; % contador de sobrestimação

    s2=1/(10^(SNR/10)); % variância de %ruído
    térmico

    for test=1:N

        if p==0
            X=0;
        else
            X=randn(p,n)*sqrt(1/(2*p))+1i*randn(p,n)*sqrt(1/(2*p
));
        end
    end
end
```

```
V=randn(m,n)*sqrt(s2/2)+1i*randn(m,n)*sqrt(s2/2);
if p==0
    H=0;
else
    H=randn(m,p)*sqrt(1/2)+1i*randn(m,p)*sqrt(1/2);
end

Y=H*X+V;
R=(Y*Y')/n;
lambda=sort(eig(R), 'descend');

% Algoritmo AIC -----
for i=0:m-1
    AIC(i+1)=-
2*n*log(prod(lambda(i+1:m)))/((1/(m-
i))*sum(lambda(i+1:m))^(m-i))+2*i*(2*m-i);
end
[minAIC, Iaic]=min(AIC);
NrAIC=Iaic-1;

% Algoritmo BIC -----
for i=0:m-1
    BIC(i+1)=-
2*n*log(prod(lambda(i+1:m)))/((1/(m-
i))*sum(lambda(i+1:m))^(m-i))+i*(2*m-i)*log(n);
end
[minBIC, Ibic]=min(BIC);
NrBIC=Ibic-1;

% Algoritmo MDL -----
for i=0:m-1
    MDL(i+1)=-
n*log(prod(lambda(i+1:m)))/((1/(m-
i))*sum(lambda(i+1:m))^(m-i))+1/2)*i*(2*m-
i)*log(n);
end
[minMDL, Imdl]=min(MDL);
NrMDL=Imdl-1;

if NrAIC==p
    CcAIC=CcAIC+1;
end
if NrAIC>p
    CseAIC=CseAIC+1;
end

if NrBIC==p
    CcBIC=CcBIC+1;
end
if NrBIC>p
    CseBIC=CseBIC+1;
end

if NrMDL==p
    CcMDL=CcMDL+1;
end
if NrMDL>p
    CseMDL=CseMDL+1;
end

end

PcAIC(pt)=CcAIC/Ntests;
PseAIC(pt)=CseAIC/Ntests;

PcBIC(pt)=CcBIC/Ntests;
PseBIC(pt)=CseBIC/Ntests;

PcMDL(pt)=CcMDL/Ntests;
PseMDL(pt)=CseMDL/Ntests;

end
plot(Parametro,PcAIC,'-
ok','MarkerFaceColor','k','MarkerEdgeColor','k');
hold on
plot(Parametro,PseAIC,'ko-');
hold on
```

```

plot(Parametro,PcBIC,'-
or','MarkerFaceColor','r','MarkerEdgeColor','r');
hold on
plot(Parametro,PseBIC,'ro-');
hold on
plot(Parametro,PcMDL,'-
ob','MarkerFaceColor','b','MarkerEdgeColor','b');
hold on
plot(Parametro,PseMDL,'bo-');

xlabel(sel, 'FontWeight','bold');
ylabel('Probabilidade', 'FontWeight','bold');
xlim([Parametro(1) Parametro(end)]);
grid on;
legend('PcAIC', 'PseAIC', 'PcBIC', 'PseBIC', 'PcMDL',
'PseMDL');

```

REFERÊNCIAS

- [1] B. Nadler, "Nonparametric Detection of signals by Information Theoretic Criteria: Performance Analysis and an Improved Estimator", *IEEE Trans. Signal Proc.*, vol. 58, 5, pp. 2746-2756, May. 2010.
- [2] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria", *IEEE Trans. Signal Process.*, vol. 33, 2, pp. 387-392, Apr. 1985.
- [3] E. S. Kritchman and B. Nadler, "Non-parametric Detection of the Number of Signals: Hypothesis Testing and Random Matrix Theory", *IEEE Trans. Signal Proc.*, vol. 57, 10, pp. 3930-3941, Oct. 2009.
- [4] S. Kritchman and B. Nadler, "Determining the number of components in a factor model from limited noisy data", *Elsevier, Chemometrics and Intelligent Laboratory Systems*, pp. 19-32, June 2008.
- [5] R. K. Swamy, S. R. Murty and B. Yegnanarayana, "Determining number of speakers from multispeaker speech signals using excitation source information", *IEEE Sig. Process.*, vol. 14, 7, pp. 481-484, July 2007.
- [6] M. Chiani and M. Z. Win, "Estimating the Number of Signals Observed by Multiple Sensors", *International Workshop on Cognitive Information Processing*, June 2010.
- [7] E. Fishler M. Grosman and H. Messer, "Determining the Number of Discrete Alphabet Sources from Sensor Data", *EURASIP Journal on Applied Signal Processing*, pp. 4-12, 1, Jan. 2005.
- [8] E. Fishler and H. V. Poor, "Estimation of the Number of Sources in Unbalanced Arrays via Information Theoretic Criteria", *IEEE Signal Proc.*, vol. 53, 9, pp. 3543-3553, Sept. 2005.
- [9] A. Ghasemi and E.S. Souza, "Spectrum Sensing in Cognitive Radio Networks: Requirements, Challenges and Design Trade-offs" *IEEE Communications Magazine*, vol. 46, 4, pp. 32-39, Apr. 2008.
- [10] N. Arkind and B. Nadler, "Parametric Joint Detection-Estimation of the Number of Sources in Array Processing", *Sensor Array and Multichannel Signal Process. Workshop, IEEE*, pp. 269-272, Oct. 2010.
- [11] A. Bruckstein, T. J. Shan, and T. Kailath, "The resolution of overlapping echoes", *IEEE Trans. Signal Proc.*, vol. 33, 6, pp. 1357-1367, Dec. 1985.
- [12] P. Stoica and Y. Selén, "Model-order selection: a review of information criterion rules", *IEEE Trans. Signal Proc.*, vol. 21, 4, pp. 36-47, July 2004.
- [13] C. W. Helstrom, "Elements of signal Detection & Estimation", Prentice-Hall, Inc., 1995.
- [14] H. Akaike, "Information theory and extension of the maximum likelihood principle", *Proc. 2nd Int. Symp. Inform. Theory, suppl. Problems of Control and Inform. Theory*, pp. 267-281, 1973.
- [15] H. Akaike, "A new look at the statistical model identification", *IEEE Control Systems Society*, vol.19, 6, pp. 716-723, Dec. 1974.
- [16] G. Schwarz, "Estimating the dimension of a model", *The Annals of statistics*, vol. 6, 2, pp. 461-464, 1978.
- [17] J. Rissanen, "Modeling by shortest data description", *Automatica*, vol.14, pp. 465-471, 1978.
- [18] P. Grünwald, "A Tutorial Introduction to the Minimum Description Length Principle", *Centrum voor Wiskunde en Informatica, Kruislaan, Amsterdam*, Apr. 2005.



Top Digital, elaborando relatórios de conformidade, e na CEMIG Telecom com redes SDH.

Patrícia Maria Soares nasceu em Belo Horizonte, MG, em novembro de 1986. Em maio de 2011 recebeu o título de Engenheira de Telecomunicações pela Fundação Mineira de Educação e Cultura (FUMEC). Em agosto de 2011 iniciou o Mestrado em Telecomunicações do Inatel e desde então realiza pesquisas em processamento de sinais para sistemas de Rádios Cognitivos, especificamente sobre algoritmos empíricos para estimação do número de fontes. Antes trabalhou na



Top Digital, elaborando relatórios de conformidade, e na CEMIG Telecom com redes SDH.

Dayan Adionel Guimarães nasceu em Carrancas, MG, em 01 de março de 1969. Possui os títulos: Técnico em Eletrônica (ETE FMC, 1987), Engenheiro Eletricista (Inatel, 1994), Especialista em Engenharia de Comunicação de Dados (Inatel, 2003), Especialista em Administração com ênfase em Gerência de RH (FAI, 1996), Mestre em Engenharia Elétrica (Unicamp, 1998) e Doutor em Engenharia Elétrica (Unicamp, 2003). De 1988 a 1993 desenvolveu sensores e equipamentos para instrumentação industrial e controle e também foi Supervisor de Produção e Supervisor de Engenharia de Produtos na SENSE Sensores e Instrumentos. Desde janeiro de 1995 é Professor do Inatel onde, por oito anos, foi responsável pela estrutura que dá apoio às atividades de ensino prático nas áreas de Telecomunicações, Eletrônica e Eletrotécnica. Suas pesquisas incluem aspectos gerais sobre Transmissão Digital e Comunicação Móvel. Seu mais recente interesse situa-se na aplicação de técnicas de Otimização Convexa em Telecomunicações, assunto de seu programa de Pós-Doutorado na UFSC, realizado em 2010. É membro do Conselho do Mestrado do Inatel e do Corpo Editorial da revista Telecomunicações (Inatel). Hoje ocupa a posição de Professor Titular do Inatel em regime de dedicação por tempo integral.

Dayan Adionel Guimarães nasceu em Carrancas, MG, em 01 de março de 1969. Possui os títulos: Técnico em Eletrônica (ETE FMC, 1987), Engenheiro Eletricista (Inatel, 1994), Especialista em Engenharia de Comunicação de Dados (Inatel, 2003), Especialista em Administração com ênfase em Gerência de RH (FAI, 1996), Mestre em Engenharia Elétrica (Unicamp, 1998) e Doutor em Engenharia Elétrica (Unicamp, 2003). De 1988 a 1993 desenvolveu sensores e equipamentos para instrumentação industrial e controle e também foi Supervisor de Produção e Supervisor de Engenharia de Produtos na SENSE Sensores e Instrumentos. Desde janeiro de 1995 é Professor do Inatel onde, por oito anos, foi responsável pela estrutura que dá apoio às atividades de ensino prático nas áreas de Telecomunicações, Eletrônica e Eletrotécnica. Suas pesquisas incluem aspectos gerais sobre Transmissão Digital e Comunicação Móvel. Seu mais recente interesse situa-se na aplicação de técnicas de Otimização Convexa em Telecomunicações, assunto de seu programa de Pós-Doutorado na UFSC, realizado em 2010. É membro do Conselho do Mestrado do Inatel e do Corpo Editorial da revista Telecomunicações (Inatel). Hoje ocupa a posição de Professor Titular do Inatel em regime de dedicação por tempo integral.



professor adjunto da Faculdade de Medicina de Itajubá. Já publicou artigos em revistas do IEEE e em congressos internacionais. Tem experiência na área de Engenharia Elétrica, com ênfase em Telecomunicações, atuando principalmente na área de comunicações sem fio. Tem experiência na indústria nas áreas produtiva e de suprimentos. Suas pesquisas incluem aspectos gerais sobre Transmissão Digital, Comunicação Móvel e canais com desvanecimento. Seu mais recente interesse situa-se em técnicas de sensoriamento espectral para sistemas de Rádio Cognitivo.

Rausley Adriano Amaral de Souza nasceu em Passos, MG, em 08 de março de 1972. Possui graduação em Engenharia Elétrica pelo Inatel - Instituto Nacional de Telecomunicações (1994), Mestrado em Engenharia Elétrica pelo Inatel (2002) e Doutorado em Engenharia Elétrica pela Unicamp - Universidade Estadual de Campinas (2009). Atualmente é professor adjunto do Inatel. Atuou por seis anos como coordenador adjunto para o programa de pós-graduação Lato Sensu do Inatel. Atuou como professor adjunto da Faculdade de Medicina de Itajubá. Já publicou artigos em revistas do IEEE e em congressos internacionais. Tem experiência na área de Engenharia Elétrica, com ênfase em Telecomunicações, atuando principalmente na área de comunicações sem fio. Tem experiência na indústria nas áreas produtiva e de suprimentos. Suas pesquisas incluem aspectos gerais sobre Transmissão Digital, Comunicação Móvel e canais com desvanecimento. Seu mais recente interesse situa-se em técnicas de sensoriamento espectral para sistemas de Rádio Cognitivo.