

**PROPOSTA DE UMA DISTRIBUIÇÃO  
BIMODAL SEM MISTURA: UMA  
APLICAÇÃO EM DADOS DE TEOR DE  
SÓLIDOS SOLÚVEIS DE CANA-DE-AÇÚCAR**

**MARCELO LEMOS DE MEDEIROS**

**2009**

**MARCELO LEMOS DE MEDEIROS**

**PROPOSTA DE UMA DISTRIBUIÇÃO BIMODAL SEM MISTURA:  
UMA APLICAÇÃO EM DADOS DE TEOR DE SÓLIDOS SOLÚVEIS DE  
CANA-DE-AÇÚCAR**

Tese apresentada à Universidade Federal de Lavras,  
como parte das exigências do Programa de Pós-  
Graduação em Estatística e Experimentação  
Agropecuária, para a obtenção do título de “Doutor”.

Orientador

Prof. Dr. Mário Javier Ferrua Vivanco

LAVRAS  
MINAS GERAIS - BRASIL  
2009

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da  
Biblioteca Central da UFLA**

Medeiros, Marcelo Lemos de.

Proposta de uma distribuição bimodal sem mistura: uma aplicação em dados de teor de sólidos solúveis de cana-de-açúcar / Marcelo Lemos de Medeiros. – Lavras : UFLA, 2009.

54 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2009.

Orientador: Mário Javier Ferrua Vivanco.

Bibliografia.

1. Mistura de distribuições. 2. Família exponencial de probabilidades. 3. Distribuição bimodal de probabilidades. I. Universidade Federal de Lavras. II. Título.

CDD – 519.24

**MARCELO LEMOS DE MEDEIROS**

**PROPOSTA DE UMA DISTRIBUIÇÃO BIMODAL SEM MISTURA:  
UMA APLICAÇÃO EM DADOS DE TEOR DE SÓLIDOS SOLÚVEIS DE  
CANA-DE-AÇÚCAR**

Tese apresentada à Universidade Federal de Lavras,  
como parte das exigências do Programa de Pós-  
Graduação em Estatística e Experimentação  
Agropecuária, para a obtenção do título de “Doutor”.

APROVADA em 31 de julho de 2009

Prof. Dr. Eric Batista Ferreira	UNIFAL-MG
Prof. Dr. Joel Augusto Muniz	UFLA
Prof. Dr. Telde Natel Custódio	UFSJ
Prof. <sup>a</sup> Dr. <sup>a</sup> Thelma Sáfadi	UFLA

Prof. Dr. Mário Javier Ferrua Vivanco  
UFLA  
(Orientador)

LAVRAS  
MINAS GERAIS - BRASIL

A Deus,

Por nunca me abandonar nos momentos mais difíceis de minha humilde caminhada e dar-me forças para prosseguir e jamais desistir.

**Ofereço**

A minha mãe, Nadir Lemos, por seu exemplo de perseverança e apoio irrestrito.

Ao meu pai, Marcos Antônio Medeiros, pela vida.

Ao meu irmão Marcos, por seu exemplo de lutas e conquistas.

Aos meus filhos, Larissa, Heitor e Thayssa, estímulos para a minha vida.

**Dedico**

## AGRADECIMENTOS

À Universidade Federal de Lavras e ao Departamento de Ciências Exatas, pela oportunidade de realizar este curso.

Ao professor Mário, pela orientação e paciência nos momentos oportunos.

Ao professor Eric, pelas sugestões.

Ao professor Joel, pelo incentivo ao meu trabalho desde o início de minha caminhada nesta instituição.

Aos professores participantes da banca de qualificação e defesa, pelas valiosas contribuições dadas ao trabalho.

Aos professores do Departamento de Ciências Exatas da UFLA, pelos ensinamentos transmitidos.

Aos funcionários do Departamento de Ciências Exatas da UFLA, pela presteza.

Aos professores do Departamento de Estatística da Universidade Federal de Viçosa, pelo estímulo inicial à minha paixão pela estatística.

Ao amigo Fernando Ewald, pelo apoio e fornecimento dos dados.

Ao amigo e colega de curso Aladir, pelo companheirismo e apoio nos momentos difíceis dessa jornada.

Ao colega Anderson, pela indispensável ajuda no R.

Aos demais colegas de curso, que por muito tempo ou por apenas alguns momentos, permitiram a troca de experiências pessoais e profissionais.

A toda minha família, pelo apoio e confiança no sucesso dos meus desafios.

Enfim, a todos aqueles que contribuíram, de forma direta ou indireta, para a realização de mais essa etapa de minha vida.

“Da mesma forma que você não precisa saber ler uma partitura musical para apreciar uma música ou uma sinfonia de Beethoven, você também não precisa saber escrever equações matemáticas para apreciar a beleza da ciência.”

Marcelo Gleiser  
Físico e Astrônomo brasileiro (Dartmouth College-USA)

## **BIOGRAFIA**

Marcelo Lemos de Medeiros, filho de Marcos Antônio Medeiros (sempre presente) e Nadir Lemos de Medeiros, nascido em Paranavaí, estado do Paraná, em 15 de setembro de 1972.

Em 1990, ingressou no Curso de Agronomia da Universidade Federal de Viçosa (UFV), em Viçosa, MG, tendo atuado como Monitor I de Estatística, de 1991 a 1993 e como bolsista de iniciação científica no Departamento de Solos de 1993 a 1994, graduando-se em fevereiro de 1995

Em março de 1995, iniciou o Curso de Mestrado em Solos e Nutrição de Plantas na UFV, onde atuou como Monitor II de Estatística em 1996, submetendo-se à defesa de dissertação em fevereiro de 1999.

Foi professor da Sociedade Educacional III Milênio em Curitiba, PR, no ano de 1998.

Foi professor de matemática do pré-vestibular Pitágoras em Belo Horizonte, MG, no período de fevereiro de 2001 a julho de 2003.

Contratado pela Fundação Educacional de Divinópolis, unidade agregada à Universidade do Estado de Minas Gerais (UEMG), em abril de 1999, onde trabalha com atividades de docência superior e pesquisa até os dias atuais.

Em março de 2004, ingressou no Doutorado em Estatística e Experimentação Agropecuária na Universidade Federal de Lavras (UFLA), em Lavras, MG, submetendo-se à defesa de tese em julho de 2009.



## SUMÁRIO

LISTA DE FIGURAS.....	i
RESUMO.....	iii
ABSTRACT .....	iv
1 INTRODUÇÃO.....	1
2 REFERENCIAL TEÓRICO.....	4
2.1 Modelos de distribuições de probabilidade.....	4
2.2 Estimação de parâmetros .....	5
2.2.1 Método da máxima verossimilhança .....	6
2.2.2 Método dos momentos.....	8
2.3 Família exponencial de probabilidades.....	9
2.4 Misturas de distribuições de probabilidade.....	10
2.5 Aplicações de misturas de distribuições .....	11
2.6 Teor de sólidos solúveis da cana-de-açúcar.....	16
3 A DISTRIBUIÇÃO BIMODAL SEM MISTURA .....	19
3.1 Construção do modelo .....	19
3.2 Classificação dos parâmetros.....	23
3.3 Família de probabilidades do modelo proposto .....	27
4 APLICAÇÃO EM CANA-DE-AÇÚCAR.....	29
4.1 Definição do modelo bimodal.....	30
4.2 Estimação dos parâmetros .....	31
4.3 Pontos críticos.....	40
4.4 Desempenho do modelo.....	42
5 CONCLUSÕES .....	45
6 REFERÊNCIAS BIBLIOGRÁFICAS .....	46
ANEXOS .....	50

## LISTA DE FIGURAS

	<b>Página</b>
FIGURA 1	Densidades não paramétricas para os dados da velocidade máxima do vento no mês de abril (linha pontilhada) e distribuição ajustada pela mistura de distribuições Gumbel (linha contínua).....12
FIGURA 2	Imagem de uma amostra de juntas coladas em madeira utilizando técnicas de visão artificial.....14
FIGURA 3	Perfis das médias do número de fibras mielínicas do nervo hipoglosso direito nos grupos <60 anos e >= 60 anos, de acordo com o diâmetro das fibras.....15
FIGURA 4	Histograma dos valores de impedância acústica, mostrando a distribuição bimodal.....16
FIGURA 5	Ilustração gráfica da equação (1).....20
FIGURA 6	Ilustração gráfica da equação (2).....21
FIGURA 7	Ilustração gráfica da equação (3).....22
FIGURA 8	Gráficos da densidade de probabilidade obtida para valores variados do parâmetro $a$ e valores fixos para os parâmetros $b$ , $c$ e $d$ . ....24
FIGURA 9	Gráficos da densidade de probabilidade obtida para valores variados do parâmetro $b$ e valores fixos para os parâmetros $a$ , $c$ e $d$ . ....25
FIGURA 10	Gráficos da densidade de probabilidade obtida para valores variados do parâmetro $c$ e valores fixos para os parâmetros $a$ , $b$ e $d$ . ....26
FIGURA 11	Gráficos da densidade de probabilidade obtida para valores variados do parâmetro $d$ e valores fixos para os parâmetros $a$ , $b$ e $c$ . ....27

FIGURA 12	Distribuição dos valores obtidos para o teor de sólidos solúveis (TSS), em °Brix, para 620 observações dos cortes de cana-de-açúcar da safra de 2007, de uma usina da região de Araraquara, SP. ....	30
FIGURA 13	Distribuição dos valores obtidos para $k$ e $G(x_i; a, b, c, d)$ em função das 12 combinações utilizadas para valores atribuídos aos parâmetros $a, b, c$ e $d$ . ....	33
FIGURA 14	Gráfico das funções de densidades de probabilidade e função de distribuição acumulada para o modelo bimodal proposto ajustado para as 620 observações do teor de sólidos solúveis. ....	43

## RESUMO

MEDEIROS, Marcelo Lemos de. **Proposta de uma distribuição bimodal sem mistura**: uma aplicação em dados de teor de sólidos solúveis de cana-de-açúcar. 2009. 54p. Tese (Doutorado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras.<sup>1</sup>

Nos trabalhos científicos, em várias áreas do conhecimento, diante de problemas de modelagem com características bimodais em variáveis aleatórias contínuas, geralmente se opta por utilizar distribuições mistas ou mistura de distribuições. Esses procedimentos consistem, de maneira geral, no ajuste de dois modelos unimodais distintos ou não, com fatores de ponderação  $p$  e  $1-p$  ( $0 < p < 1$ ). Isso é feito para cada modelo, no respectivo intervalo de domínio da variável aleatória estudada, de modo que fiquem combinados em todo o domínio real. Em geral, as misturas de distribuições usuais apresentam muitos parâmetros e, portanto, a solução do sistema de equações que compõem esses parâmetros não é analítica e requer o uso de algoritmos iterativos pouco comuns para os profissionais. O presente trabalho foi realizado com o objetivo construir um modelo de densidade de probabilidade bimodal com lei matemática única e estimadores explícitos para os seus parâmetros. Foi obtido um modelo que pertence à família exponencial com quatro parâmetros: dois de locação, um de forma e um de escala. Os estimadores de máxima verossimilhança dos parâmetros de posição foram obtidos por iteração simples e os demais apresentaram formas explícitas. O método dos momentos não permitiu a estimação dos parâmetros de posição. Uma amostra de 620 observações do teor de sólidos solúveis (em graus Brix) da cana-de-açúcar colhida na safra de 2007, numa usina de açúcar e álcool da cidade de Araraquara, SP, que apresentou características de bimodalidade, foi utilizada para aplicação do modelo proposto. Os resultados obtidos permitiram explicar satisfatoriamente o fenômeno em estudo. Sugere-se o uso do modelo de densidade bimodal proposto como alternativa aos modelos de misturas de distribuições para se tratar as variáveis aleatórias com comportamento bimodal e contribuir para popularizar a modelagem em áreas científicas com menor afinidade para a estatística computacional avançada.

Palavras-chave: Distribuição bimodal. Mistura de distribuições. Família exponencial. Teor de sólidos solúveis da cana-de-açúcar.

---

<sup>1</sup> Comitê Orientador: Prof. Dr. Mário Javier Ferrua Vivanco – UFLA (orientador), Prof. Dr. Eric Batista Ferreira – UNIFAL-MG (coorientador).

## ABSTRACT

MEDEIROS, Marcelo Lemos de. **Propose of a bimodal distribution without mixture**: an application on soluble solids content of sugar cane. 2009. 54p. Thesis (Doctorate in Statistics and Agricultural Experimentation) - Universidade Federal de Lavras, Lavras.<sup>2</sup>

In the scientific papers in various areas of knowledge, when they face modeling problems with bimodal characteristics in random variables continuous, usually opt to use mixed distributions or mixture of distributions. These procedures are generally in the adjustment of two distinct or not models with weighting factors  $p$  and  $1-p$  ( $0 < p < 1$ ). This is done for each model, in its range of random variable domain studied so that the transceivers are combined into the real domain. In General, the usual distribution mixtures presents many parameters, and therefore the solution of the equations system that make up these parameters is not analytical and requires use of unusual iterative algorithms for professionals. This paper aimed to build a model of bimodal probability density with single mathematical law and explicit estimators for their parameters. It was obtained a model which belongs to the exponential family with four parameters: two of the location, one of shape and one of scale. The maximum likelihood estimators of positional parameters were obtained by simple iteration and others presented explicit forms. The method of moments did not allow the estimation of position parameters. A sample of 620 observations of soluble solids content (in degrees Brix) sugar cane harvested in 2007 a crop in a sugar factor and distillery in Araraquara city (São Paulo State - Brazil) which has presented bimodal characteristics, was used to the application of the model proposed. The results obtained have enabled explain satisfactorily the phenomenon in study. It's suggested the use of bimodal density model proposed as an alternative to the models proposed of mixed distributions to work the random variables with bimodal behavior and to contribute to popularize science areas modeling with less affinity for advanced computing statistics.

Key-words: Bimodal distribution. Mixture of distributions. Exponential family, Soluble solids content of sugar cane.

---

<sup>2</sup> Guidance Committe: Prof. DSc. Mário Javier Ferrua Vivanco – UFLA (Adviser) e Prof. Dr. Eric Batista Ferreira – UNIFAL-MG (Co-Adviser)

## 1 INTRODUÇÃO

A cana-de-açúcar é uma planta de grande importância na produção de açúcar (sacarose) para o consumo humano e a produção de álcool combustível, fonte de energia renovável, pouco poluente e de baixo custo.

Na década de 1970, após as crises do petróleo mundial em que o preço do barril de petróleo chegou a patamares de preços insustentáveis para a época, foi criado o Programa Nacional do Álcool, ou Proálcool. Esse programa incentivou o desenvolvimento de novas tecnologias na produção e uso do álcool, derivado da fermentação do caldo da cana-de-açúcar, nos veículos nacionais. Sendo assim, o Proálcool alavancou a expansão das áreas produtivas no país e construção de novas usinas principalmente no interior do estado de São Paulo.

Após o declínio dessa cultura, ocorrido nas décadas de 1980 e 1990, com a extinção do programa por pressão de outros interesses econômicos, essa cultura voltou a se destacar. As usinas de cana-de-açúcar atualmente não só produzem o açúcar e o álcool a partir de açúcares extraídos do colmo da planta, como também geram energia elétrica suficiente para sustentar suas atividades a partir da queima do bagaço resultante da moagem e, adicionalmente, vendem a energia excedente para concessionárias estaduais de energia elétrica, que a repassam para suas redes de distribuição.

Tendo em vista que a sacarose (principal açúcar existente na planta da cana-de-açúcar) é a principal matéria prima para a produção de açúcar e álcool, há grande interesse das usinas no controle de sua concentração por ocasião da colheita. Sendo assim, na chegada de um carregamento de cana à usina, colhe-se uma amostra de material e, dentre outras medidas de interesse, mede-se, em graus Brix, o teor de sólidos solúveis, cujo componente mais abundante é a sacarose.

Os graus Brix representam o teor de sólidos solúveis (TSS) obtido por meio de um equipamento denominado refratômetro. Esse valor, obtido a partir de várias amostras dos carregamentos de colmos que chegam à usina, é uma das variáveis mais importantes para o cálculo do açúcar total recuperável (ATR), base para o pagamento de cana ao produtor.

Como existe uma grande variabilidade na cana entregue para suprir as necessidades de uma usina de grande porte, há relevante variação dos TSS obtidos e que refletem diretamente no rendimento industrial.

O TSS durante a safra varia, dentre outros fatores, com a fertilidade do solo, o regime pluviométrico, variedade de cana plantada e a sua maturação no momento de cada corte. A distribuição de frequências dos valores do TSS obtidos ao longo da safra apresenta características bimodais em alguns casos. Isso sugere que, para determinadas condições numa dada safra, há um grupo de cortes de cana em que a combinação desses fatores gera rendimento superior (do ponto de vista do colmo da cana), com distribuição dos resultados em torno de um TSS médio maior e outro grupo de cortes com rendimento inferior, em torno de um TSS menor. Portanto, torna-se importante um modelo de previsibilidade do TSS para uma safra vindoura, que esteja sob as mesmas condições, para planejamento do rendimento industrial na produção de açúcar e álcool combustível e, conseqüentemente, no resultado financeiro para o produtor e a usina numa dada safra.

A presença de distribuições de probabilidades que apresentam duas ou mais populações independentes dentro de uma mesma variável aleatória  $X$  é pouco comum. Na maior parte dos casos, são definidas (de forma empírica ou não) essas populações presentes e são tratadas isoladamente com distribuições de probabilidade que melhor se ajustam em cada caso. No entanto, seja para simplificação do estudo de algum fenômeno ou por julgar desnecessária sua estratificação na obtenção de seus objetivos, existem situações em que há

interesse do pesquisador em modelar os dados em uma distribuição de probabilidade única e os modelos unimodais não descrevem de forma satisfatória esses fenômenos.

Considerando-se o caso específico da bimodalidade, vários trabalhos científicos, em várias áreas do conhecimento, mostram que diante de problemas de modelagem com características bimodais em variáveis aleatórias contínuas, geralmente opta-se por utilizar distribuições mistas ou mistura de distribuições. Esses procedimentos consistem, de maneira geral, no ajuste de dois modelos unimodais distintos ou não (uniforme, exponencial, normal, gama, Beta, Weibull, Gumbel, etc.) com fatores de ponderação  $p$  e  $1-p$ . Isso é feito para cada modelo, no respectivo intervalo de domínio da variável aleatória  $X$  estudada, para que assim fiquem combinados em todo o domínio real de  $X$ . A carência, na literatura, de modelos que expressem a bimodalidade com uma lei matemática única também corroboram para o uso de mistura de distribuições nesses casos.

O uso da mistura de distribuições é bastante vantajoso, à medida que se trabalha com modelos unimodais de características bastante exploradas e que se ajustam aos dados com facilidade. A desvantagem é que, em geral, as misturas de distribuições usuais apresentam muitos parâmetros e, portanto, a solução do sistema de equações que compõem esses parâmetros não é analítica e requer a aplicação de algoritmos iterativos complexos que limitam seu uso pelos profissionais responsáveis por essas análises.

Em razão do exposto, objetivou-se com este trabalho, propor e estudar um modelo de distribuição de probabilidade de variável aleatória contínua, com características de bimodalidade, que apresente lei matemática única para todo o domínio real da variável aleatória  $X$ , tenha maior praticidade para manuseio computacional e aplicá-lo num banco de dados do TSS de cana-de-açúcar, colhida no ano de 2007, numa usina localizada no município de Araraquara, SP.



## **2 REFERENCIAL TEÓRICO**

Para atender aos propósitos do presente trabalho, se faz necessário definir alguns conceitos de forma mais detalhada, bem como abordar trabalhos que ilustram a problemática envolvida em modelagem bimodal. A seguir, serão relatados tópicos pertinentes ao entendimento das propostas a serem apresentadas posteriormente na construção do modelo e sua aplicação em dados de cana-de-açúcar.

### **2.1 Modelos de distribuições de probabilidade**

Muitas variáveis aleatórias se adaptam a uma série de problemas práticos e aparecem, com bastante frequência, no cotidiano científico. Um estudo pormenorizado dessas variáveis permite a determinação de seus principais parâmetros. Essas construções recebem o nome de modelos de probabilidades. Esses modelos tornaram-se úteis para a descrição de várias situações, quando se atendem às suas condições, e assim, facilitam o trabalho de construção da distribuição de probabilidades.

De acordo com Mood et al. (1974), os modelos mais conhecidos são divididos de acordo com o tipo de variável aleatória que descrevem. Para as variáveis aleatórias discretas, os modelos mais conhecidos são o de Bernoulli, o binomial, o hipergeométrico e o de Poisson, dentre outros. Já para as variáveis aleatórias contínuas, destacam-se os modelos exponencial, normal, gamma, beta, Gumbel e Weibull. O modelo de distribuição uniforme tem formato adaptado para as variáveis aleatórias discreta e contínua.

O detalhamento dos modelos citados não é objetivo do presente trabalho e, portanto não será feito nesta seção. No entanto, convém ressaltar que os mesmos são amplamente explorados pela literatura especializada em vários

livros didáticos, utilizados nos cursos de graduação e de forma mais aprofundada por outros autores, como Mood et al. (1974) e Lindsey (1996).

Em situações em que os modelos citados não atendem as expectativas do pesquisador quanto à forma ou flexibilidade, por exemplo, é possível que se possa gerar um novo modelo adequado aos interesses do problema. No entanto, esse tipo de objetivo requer uma familiaridade do pesquisador com modelagem matemática, criatividade e um bom editor gráfico para visualização das eventuais propostas de funções que atendam aos propósitos desejados.

Mesmo depois de escolhido o modelo que apresente as características desejáveis de ajustamento, é importante a inferência estatística para definir estimadores de seus parâmetros, bem como as propriedades ótimas desses estimadores.

## **2.2 Estimação de parâmetros**

Segundo Mood et al. (1974), os processos de estimação permitem obter aproximações numéricas para os parâmetros. Existem, basicamente, dois processos de estimação: a pontual e a intervalar. A estimação pontual ou por ponto é aquela em que se obtém um único valor amostral para estimar o parâmetro populacional por meio de um estimador. Porém, esse processo de estimação não fornece uma ideia da margem de erro que é cometida ao se estimar um determinado parâmetro. A estimação intervalar consiste em se construir um intervalo com probabilidade pré-fixada, em geral de 95% ou 99%, de conter o verdadeiro valor do parâmetro conhecido.

Algumas propriedades dos estimadores pontuais são desejáveis na teoria da inferência e a escolha de um estimador de um parâmetro  $\theta$  qualquer em relação a outro depende de avaliação dessas propriedades (Mood et al., 1974). As principais propriedades são citadas a seguir:

- a) **não tendenciosidade:** um estimador  $\hat{\theta}$  é dito um estimador não tendencioso do parâmetro  $\theta$  se a esperança matemática de  $\hat{\theta}$  for igual a  $\theta$ , ou seja,  $E(\hat{\theta}) = \theta$ . Vale lembrar que os termos não tendencioso, não viciado, não viesado e imparcial são sinônimos;
- b) **consistência:** um estimador é consistente se, além de não viesado, com o aumento do tamanho da amostra  $n$ , seu valor converge para o parâmetro e sua variância tende a zero;
- c) **eficiência:** o estimador de maior eficiência, dentre todos os estimadores não viesados de  $\theta$ , é aquele que possui a menor variância;
- d) **suficiência:** um estimador tem uma estatística suficiente quando essa estatística consegue resumir toda informação sobre o parâmetro  $\theta$  presente na amostra sem necessariamente possuir a amostra completa.

Dentre os métodos de estimação existentes, destacam-se o método da máxima verossimilhança e o método dos momentos.

### 2.2.1 Método da máxima verossimilhança

O método da máxima verossimilhança é um método de estimação pontual amplamente utilizado na inferência estatística. Esse método foi introduzido pela primeira vez por Fisher (1912) e é mostrado a seguir, de forma resumida.

Seja uma variável aleatória  $X$ , cuja distribuição é dada pela função de densidade (ou de probabilidade)  $f(x, \theta_1, \theta_2, \dots, \theta_k)$  em que  $\theta_1, \theta_2, \dots, \theta_k$  são  $k$

parâmetros dessa função. Seja  $\{x_1, x_2, \dots, x_n\}$  uma amostra aleatória de  $n$  elementos dessa população. A função de verossimilhança ( $L$ ) é dada por:

$$L(\theta_1, \theta_2, \dots, \theta_k | x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta_1, \theta_2, \dots, \theta_k)$$

Em geral, por facilidade algébrica, usa-se o logaritmo de  $L$ , chamado de função suporte, que é expressa por:

$$\ln L(\theta_1, \theta_2, \dots, \theta_k | x_1, x_2, \dots, x_n) = \sum_{i=1}^n \ln f(x_i; \theta_1, \theta_2, \dots, \theta_k)$$

Os estimadores de máxima verossimilhança (EMV) de  $\theta_1, \theta_2, \dots, \theta_k$  são obtidos pelos valores máximos de  $L$  ou de  $\ln L$ . Utilizando a função suporte e considerando que o valor máximo está no ponto de derivada nula, pode-se dizer que  $\theta_1, \theta_2, \dots, \theta_k$  formam a solução de um sistema de  $k$  equações, tal que

$$\frac{\partial \ln L}{\partial \theta_j} = 0 \text{ para } j = 1, 2, \dots, k$$

Segundo Bolfarine & Sandoval (2001), os EMV são funções de estatísticas suficientes e são invariantes. O princípio da invariância define que se  $\hat{\theta}$  é um EMV do parâmetro  $\theta$  e  $g(\hat{\theta})$  é uma função real inversível, então,  $g(\hat{\theta})$  é um EMV de  $g(\theta)$ . Os mesmos autores definem também que, para grandes amostras, esses estimadores são assintoticamente não viciados com distribuição aproximadamente normal e suas variâncias coincidem com os correspondentes limites inferiores de Cramer-Rao das variâncias desses estimadores, e por consequência, são eficientes.

Com base na distribuição assintótica dos EMV, podem-se construir intervalos de confiança aproximados, no caso de grandes amostras, para os parâmetros e que são expressos por

$$\frac{\hat{\theta}_j - \theta_j}{\sqrt{(nI_F(\theta_j))^{-1}}} \underset{a}{\approx} N(0,1) \text{ para } j = 1, 2, \dots, k$$

em que  $I_F(\theta_j)$  é a matriz de informação de Fisher de  $\theta_j$ , dada por:

$$nI_F(\theta_j) = -\frac{\partial^2 \ln L}{\partial \theta_j^2}$$

### 2.2.2 Método dos momentos

De acordo com Mood et al. (1974), definem-se como momentos populacionais simples e centrados de ordem  $k$  (caso existam), respectivamente, por:

$$E[X^k] \text{ e } E[(X - \mu)^k] \quad \text{para } k = 1, 2, 3, \dots$$

Dada uma amostra aleatória  $X = \{x_1, x_2, \dots, x_n\}$ , definem-se como momentos amostrais simples e centrados de ordem  $k$ , respectivamente, por:

$$\frac{1}{n} \sum_{i=1}^n X_i^k \text{ e } \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \quad \text{para } k = 1, 2, 3, \dots$$

Uma alternativa para obtenção de momentos populacionais é o uso da função geradora de momentos. Seja uma variável aleatória  $X$ , cuja distribuição é dada pela função de densidade (ou de probabilidade)  $f(x)$ . O valor  $m(t)$  esperado para  $e^{tx}$ , ou seja,  $m(t) = E[e^{tx}]$ , é definido como sendo a função geradora de momentos de  $X$ , se o valor esperado existir para todo valor de  $t$ . Essas expressões são dadas por:

$$m(t) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx \text{ para o caso de variável aleatória contínua e}$$

$$m(t) = \sum_{i=1}^n e^{tx_i} f(x_i) \text{ para o caso de variável aleatória discreta.}$$

O método dos momentos é um dos métodos de estimação mais simples e tem sido utilizado desde o século XVIII (Bolfarine & Sandoval, 2001). Consiste na obtenção de estimadores para  $\theta = (\theta_1, \theta_2, \dots, \theta_k)$  igualando-se os momentos amostrais ordinários aos momentos populacionais ordinários, ou seja, resolvendo-se as equações  $m_r = \mu_r$ , com  $r = 1, 2, \dots, k$ .

### 2.3 Família exponencial de probabilidades

De acordo com Mood et al. (1974) e Lindsey (1996), uma função densidade de probabilidade (fdp) é definida como uma função de família exponencial de probabilidades, se puder ser expressa na forma

$$f(x; \theta_1, \theta_2, \dots, \theta_k) = a_{\theta}(\theta_1, \theta_2, \dots, \theta_k) \cdot b(x) \cdot e^{\sum_{j=1}^k c_j(\theta_1, \theta_2, \dots, \theta_k) \cdot T_j(x)} \quad (8)$$

em que

$f(x; \theta_1, \theta_2, \dots, \theta_k)$  é uma fdp que pertence à família exponencial de probabilidades;

$a_\theta(\theta_1, \theta_2, \dots, \theta_k)$  é uma função dos parâmetros  $\theta_1, \theta_2, \dots, \theta_k$ ;

$b(x)$  é uma estatística;

$e = 2,71828\dots$  é a constante de Nepper e ;

$\sum_{j=1}^k c_j(\theta_1, \theta_2, \dots, \theta_k) \cdot T_j(x)$  é um somatório de funções dos parâmetros  $\theta_1, \theta_2, \dots, \theta_k$  como coeficientes associados a  $T_j(x)$ .

$T_j(x)$  é um vetor de estatísticas suficientes para  $f(x; \theta_1, \theta_2, \dots, \theta_k)$ .

A família exponencial inclui muitas das distribuições de probabilidade mais comumente utilizadas em estatística, tanto contínuas quanto discretas. Uma característica essencial desta família é que existe uma estatística suficiente com dimensão fixa.

## 2.4 Misturas de distribuições de probabilidade

De acordo com Silva (2008), citando McLachlan & Peel (2000), os modelos de misturas finitas de distribuições são utilizados em casos em que se suspeita que a amostra observada seja proveniente de duas ou mais populações independentes e não seja possível identificar a qual das populações pertence cada elemento dessa amostra.

No entanto, esses modelos também são úteis em situações em que o ajuste de um componente simples dessa mistura não é capaz de descrever o fenômeno estudado de forma adequada. As misturas de modelos permitem melhor flexibilidade nos ajustes de populações heterogêneas.

Para modelar os casos em que os dados sugerem uma distribuição bimodal em variáveis aleatórias contínuas, é comum uma mistura de duas distribuições dada por

$$g(x) = p f_1(x) + (1-p) f_2(x)$$

em que

$p$  e  $1-p$  (com  $0 < p < 1$ ) são fatores de ponderação e  $f_1(x)$  e  $f_2(x)$  são funções densidades de probabilidade.

Essas distribuições podem ser, por exemplo, duas distribuições normais com parâmetros  $N_1(\mu_1, \sigma_1^2)$  e  $N_2(\mu_2, \sigma_2^2)$ , respectivamente ou, ainda, duas distribuições distintas com respectiva parametrização.

Todavia, sabe-se que a obtenção das estimativas dos parâmetros é trabalhosa e requer a implementação de algoritmos iterativos, como, por exemplo, o algoritmo *expectation maximization*, ou EM, proposto por Dempster et al. (1977) e uma variante sua, o algoritmo *expectation-conditional maximization*, ou ECM, proposto por Meng & Rubin (1993).

## 2.5 Aplicações de misturas de distribuições

Vários trabalhos em ecléticas áreas do conhecimento utilizam da mistura de distribuições. Silva (2008) utilizou uma mistura de distribuições Gumbel para modelar dados com bimodalidade sugeridos por Silva & Zocchi (2005) na distribuição da velocidade máxima do vento em Piracicaba, SP, como mostrado na Figura 1. Os parâmetros foram obtidos pelo uso do método ECM, com a técnica de reamostragem “bootstrap”.

Silva (2003) utilizou estimadores de máxima verossimilhança em misturas de densidades normais, aplicando-os na genética do fenômeno de partenocarpia em abobrinha. Em seus estudos, concluiu, dentre outras coisas, que os estimadores de máxima verossimilhança dos componentes de média e de variância obtidos foram expressões recursivas que requereram o uso de métodos numéricos iterativos.



Morais et al. (2001) para estimar precipitação em Lavras, MG e Silva et al. (1998), em estudos da disponibilidade hídrica climática no estado de Goiás, utilizaram uma mistura de distribuições para descrever esses fenômenos. Esses autores utilizaram uma distribuição uniforme para a probabilidade de precipitação nula com fator de ponderação  $p$  para a função  $f_1(x)$  e a distribuição gama para a probabilidade de precipitação não-nula como função  $f_2(x)$  da mistura com fator de ponderação  $1 - p$ .

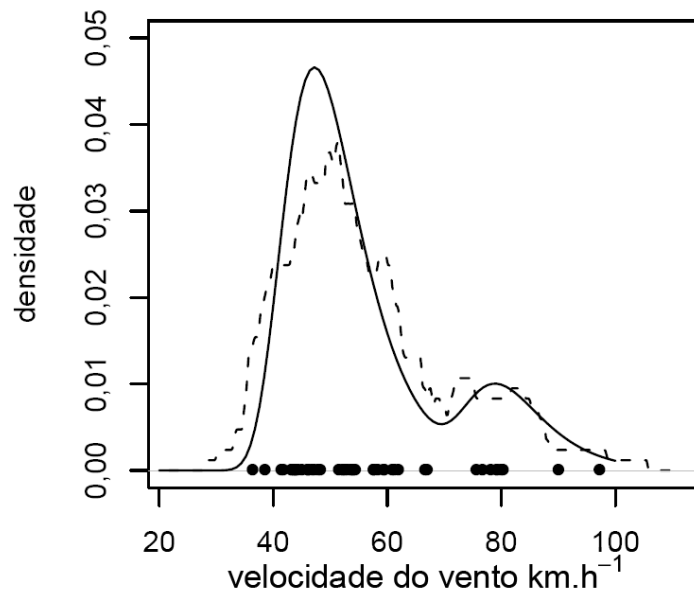


FIGURA 1 Densidades não paramétricas para os dados da velocidade máxima do vento no mês de abril (linha pontilhada) e distribuição ajustada pela mistura de distribuições Gumbel (linha contínua).

Fonte: Adaptado de Silva (2008, p. 65).

Observam-se outros trabalhos na literatura com enfoque na mistura de distribuições. Os artigos tratam de assuntos como: estimação de parâmetros de escala sob função de perda e uso de método dos momentos (Dey, 1990);

caracterização paramétrica da função de taxa de falhas em misturas de distribuições Weibull (Jiang & Murthy, 1998); testes para tendências em proporções de misturas de distribuições beta (Tiwari & Zalkikar, 1999); aplicações de misturas finitas de distribuições em mercados futuros aplicados à economia (Ané & Labidi, 2001); cálculo de tamanho de amostra para médias (Chu et al., 2004) e também aplicações de inferência bayesiana para misturas de distribuições (Choy & Chan, 2008).

Esse levantamento da literatura mostra que há uma maior preocupação natural dos pesquisadores em aprofundar e aperfeiçoar o estudo das misturas de distribuições em detrimento de pesquisas que possam, eventualmente, romper esse paradigma e propor modelos de lei matemática única com tratamento analítico menos complexo para casos mais simples de modelagem bimodal.

Em contrapartida, em escala não desprezível, há trabalhos que relatam distribuições com características bimodais dos dados avaliados e não são submetidos à modelagem estatística.

Abrahão et. al. (2003) quantificaram falhas na madeira em juntas coladas utilizando técnicas de visão artificial com digitalizadores. Neste trabalho, foi elaborada uma distribuição de frequências do número de pixels em função do valor do brilho. Esse histograma, apresentado na Figura 2, é uma distribuição bimodal que é considerada como soma de normais. O limiar ótimo de interesse dos pesquisadores é um ponto situado no vale entre os picos (ponto de mínimo). Para detectá-lo, utilizaram o método iterativo de Yang et al. (2001). Os autores relatam que esse método tem a vantagem de apresentar o tempo de processamento bem menor que o método bayesiano, porém, não é ótimo porque não minimiza o erro de classificação.

A conclusão dos autores supracitados sugere que, se houvesse possibilidade de um ajuste de um modelo de distribuição de probabilidade com lei matemática única, poder-se-ia obter esse limiar ótimo apenas com o uso de

uma raiz da derivada de primeira ordem do modelo que possuísse segunda derivada negativa.

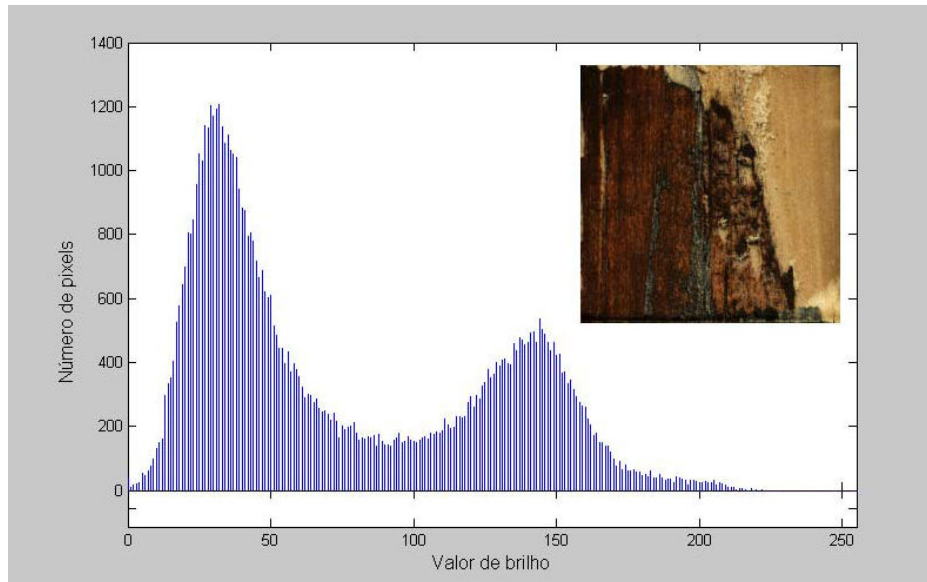


FIGURA 2 Imagem de uma amostra de juntas coladas em madeira utilizando técnicas de visão artificial.

Fonte: Adaptado de Abrahão et al. (2003, p. 74).

Tiago et al. (2005) relatam, nas conclusões de seu trabalho, a presença de uma distribuição de frequências bimodal, mostrada na Figura 3, em fibras mielínicas do nervo hipoglossal do olho, independente dos grupos etários estudados, mas não utilizaram nenhuma técnica estatística mais complexa do que a análise de variância para comparar as médias dos grupos etários. Talvez, a complexidade do conhecimento específico em estatística para modelagem, mistura de distribuições e estatística computacional, nesse caso, não estimule a busca e aplicação em modelagem de probabilidade.

Vidal et al. (2007) verificaram distribuição bimodal da impedância acústica 3D para a caracterização do Campo de Namorado, na bacia de Campos,

RJ, a 80 km do litoral. Apesar de os autores estabelecerem cenários probabilísticos com base nas distribuições de impedância acústica, mostrado na Figura 4, também não existiu uma preocupação com a modelagem estatística além do histograma descritivo, como no estudo anterior.

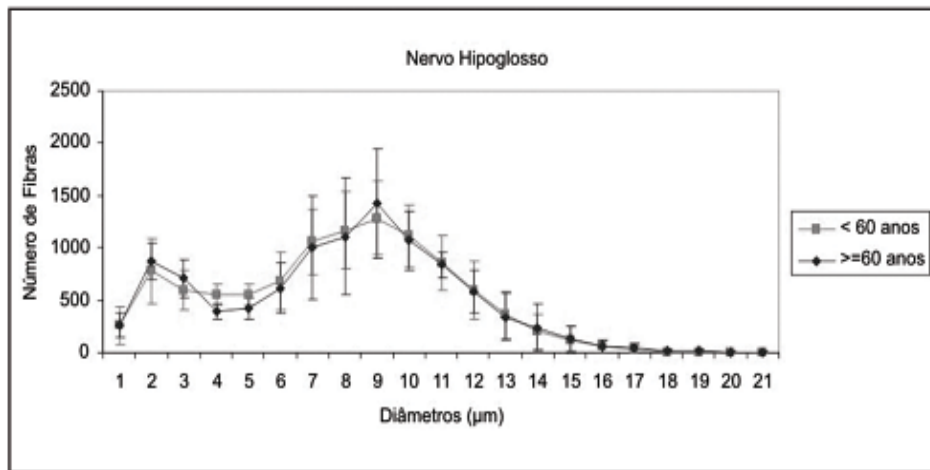


FIGURA 3 Perfis das médias do número de fibras mielínicas do nervo hipoglosso direito nos grupos <60 anos e >= 60 anos, de acordo com o diâmetro das fibras.

Fonte: Adaptado de Tiago et al. ( 2005, p. 557).

Enfim, em vista da literatura recém-citada, pode-se verificar que as pesquisas com distribuições de probabilidade bimodais estão polarizadas entre as que envolvem desenvolvimento de técnicas computacionais avançadas em modelagem matemática complexa e as que, de alguma maneira, não tratam os dados muito além da estatística descritiva.

A preocupação deste trabalho em desenvolver alternativas mais simples para tratar as variáveis bimodais deve contribuir sobremaneira para popularizar a modelagem em áreas científicas com menor afinidade para a estatística computacional avançada.

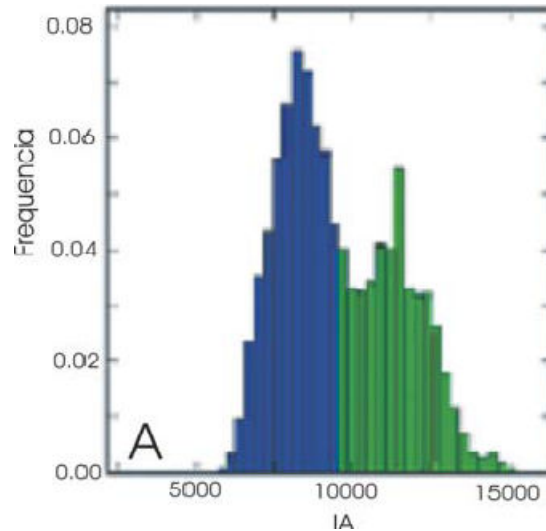


FIGURA 4 Histograma dos valores de impedância acústica, mostrando a distribuição bimodal.

Fonte: Adaptado de Vidal et al. (2007, p. 301).

## 2.6 Teor de sólidos solúveis da cana-de-açúcar

A cana-de-açúcar é uma planta reconhecida mundialmente pela grande importância na produção de açúcar (sacarose) para o consumo humano e na produção de álcool combustível, fonte de energia renovável, pouco poluente e de baixo custo. Segundo Taupier & Rodrigues (1999), tratando-se da alimentação, a cana-de-açúcar tem o maior rendimento em matéria verde e oferece o maior número de megacalorias metabolizáveis que qualquer outra cultura, constituindo, ainda, o alimento energético mais completo para o ser humano e de consumo mais geral.

De acordo com Caliaro et al. (2004), uma maior recuperação do açúcar contido na cana-de-açúcar é um assunto bastante pesquisado pelos técnicos ligados ao setor de produção de álcool e açúcar. O seu aumento se deve à evolução tecnológica dos equipamentos fabricados e às melhorias feitas nos

processos clássicos de preparo da cana, de extração e clarificação de caldo, e de fabricação de álcool e açúcar. No entanto, aumentos globais na eficiência são muito pequenos, uma vez que as eficiências industriais alcançadas nas diversas etapas já são bastante expressivas.

Tendo em vista que a sacarose (principal açúcar existente na planta da cana-de-açúcar) é a principal matéria prima para a produção de açúcar e álcool, há grande interesse das usinas no controle de sua concentração por ocasião da colheita. Sendo assim, na chegada de um carregamento de cana à usina, colhe-se uma amostra de material e, dentre outras medidas de interesse, mede-se, em graus Brix, o teor de sólidos solúveis e o Pol do caldo em %. Ambas as medidas são as mais importantes para a previsão do rendimento final de açúcar e álcool da cana colhida que chega à usina com o nome de açúcar total recuperável (ATR). Para esse trabalho, a medida de maior interesse é o teor de sólidos solúveis totais (TSS).

Segundo Assis et al. (2004), o teor de sólidos solúveis totais, que é medido em graus Brix (°Brix em %), é a porcentagem, em gramas, de sólidos dissolvidos na água presente em um produto (cana-de-açúcar). A determinação do Brix é feita a partir do caldo extraído da cana-de-açúcar efetuada em refratômetro digital, provido de correção automática de temperatura e ajuste de campo, com saída para impressora e/ou registro magnético, devendo o valor final ser expresso a 20°C. O índice de refração ( $I_r$ ), que é proporcional ao índice de sólidos solúveis, ou Brix, é obtido pela expressão

$$I_r = \frac{\text{Seno do ângulo de incidência}}{\text{Seno do ângulo de refração}}$$

Segundo Caliari et al. (2004), a colheita de cana-de-açúcar é uma “administração de sacrifícios”. Não se pode colher todo contingente de matéria-

prima no ponto de máxima maturação de cada variedade e, ainda, fatores edafoclimáticos, como tipo de solo, regime de chuvas, temperatura e variedades plantadas também influenciam no processo de maturação da cana. A safra pode se estender por até oito meses e os picos de riqueza da cana se concentram em apenas dois ou três meses, mas podem variar conforme a interação dos fatores anteriores. De acordo com Silveira Neto (1992), essa característica da planta determinou o procedimento usual, à maior parte das empresas do setor, de administrar o processo de colheita sacrificando determinadas variedades em benefício de outras.

Em vista dos aspectos citados, pode-se concluir que fatores edafoclimáticos e de logística interferem na composição final do TSS da cana colhida que chega à usina e, assim, uma forma alternativa de modelagem probabilística para essa variável, que em alguns casos pode recair na bimodalidade, deve contribuir para alimentar os modelos utilizados para planejamento dos cortes e maximização do açúcar total recuperável.

### 3 A DISTRIBUIÇÃO BIMODAL SEM MISTURA

Nesta seção, pretende-se mostrar a lógica empregada para a construção do modelo de densidade bimodal sem mistura a ser proposto.

#### 3.1 Construção do modelo

Inicialmente, o modelo a ser proposto deve atender a algumas características desejáveis para sua empregabilidade em ajustes de dados que, a priori, atendam a condições de bimodalidade.

Seja  $f(x; \theta_1, \theta_2, \dots, \theta_k)$  uma função de densidade de probabilidade bimodal com o vetor de parâmetros  $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ . As condições a serem satisfeitas são:

- (i)  $f(x; \theta_1, \theta_2, \dots, \theta_k)$  não seja uma mistura de distribuições;
- (ii)  $f(x; \theta_1, \theta_2, \dots, \theta_k) \geq 0$  para todo  $x$  pertencente ao domínio real;
- (iii)  $\lim_{x \rightarrow -\infty} f(x; \theta_1, \theta_2, \dots, \theta_k) = \lim_{x \rightarrow +\infty} f(x; \theta_1, \theta_2, \dots, \theta_k) = 0$ ;
- (iv)  $\int_{-\infty}^{+\infty} f(x; \theta_1, \theta_2, \dots, \theta_k) dx = 1$ ;
- (v)  $f(x; \theta_1, \theta_2, \dots, \theta_n)$  deve possuir três pontos críticos: duas modas, representadas por dois máximos locais e uma antimoda representada por um mínimo local entre os máximos locais.

Pode-se observar que a condição (i) é o principal objetivo do trabalho. As condições (ii) e (iv), definidas por Mood et al. (1974), são necessárias para que  $f$  seja uma função densidade de probabilidade (fdp) legítima. Em (iii), garante-se a convergência da área sob a curva de  $f$  para todo o domínio real e (v) estabelece a bimodalidade.

O ponto de partida para atender às condições propostas é a expressão



$$f(x; \theta_1) = k \cdot e^{-\frac{1}{2}(x-\theta_1)^2} \quad (1)$$

em que:

$\theta_1$  = parâmetro real e

$k$  = constante normalizadora positiva (Lindsey, 1996).

A equação (1) atende às condições (i), (ii) e (iii). Na Figura 5 pode-se observar que a função se maximiza à medida que se tende o valor de X ao parâmetro  $\theta_1$ . Verifica-se também que  $f$  é unimodal e tende a zero quando X tende ao infinito negativo ou positivo.

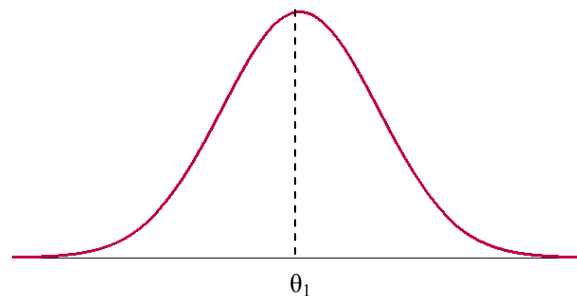


FIGURA 5 Ilustração gráfica da equação (1).

Observa-se que, para se maximizar  $f$ , a expressão  $(x - \theta_1)$  deve ser nula. Sendo assim, para tornar  $f$  bimodal, é necessário que esta seja maximizada em dois valores distintos de X. Isso pode ser conseguido com expressão modificada

$$f(x; \theta_1, \theta_2) = k \cdot e^{-\frac{1}{2}[(x-\theta_1)(x-\theta_2)]^2} \quad (2)$$

É importante ressaltar que esta etapa do processo garante a condição (v) para o modelo proposto. Para a equação (2), esboçada na Figura 6,  $f$  é

maximizada em  $x = \theta_1$  e  $x = \theta_2$ . Todavia, nesta condição, ambas as regiões em torno dos máximos locais apresentam a mesma densidade.

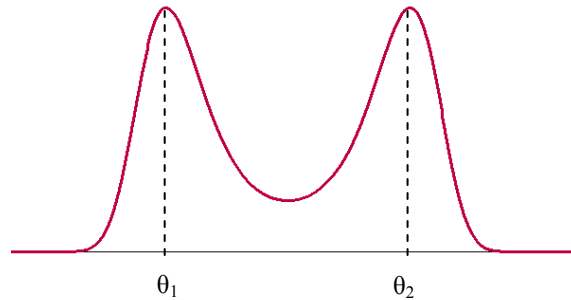


FIGURA 6 Ilustração gráfica da equação (2).

Para se ponderar as densidades atribuídas a cada moda de  $f$ , optou-se por acrescentar um novo parâmetro que permita ponderar as modas e, assim tem-se a equação

$$f(x; \theta_1, \theta_2, \theta_3) = k \cdot e^{-\frac{1}{2}[(x-\theta_1)(x-\theta_2)]^2 + (x-\theta_3)^2} \quad (3)$$

A equação (3) é ilustrada na Figura 7 e mostra a influência do parâmetro  $\theta_3$ . Este parâmetro permite a ponderação das densidades atribuídas às regiões das modas de  $f$ . Nota-se que, quando  $\theta_3$  está localizada no ponto médio de  $\theta_1$  e  $\theta_2$ , as respectivas regiões modais apresentam a mesma densidade (FIGURA 6b). Quando ocorre o contrário, a região modal mais próxima do parâmetro  $\theta_3$  recebe maior densidade (FIGURAS 7a e 7c). A inclusão do parâmetro  $\theta_3$  no modelo proposto altera a posição dos pontos críticos de  $f$  anteriormente localizados em  $x = \theta_1$ ,  $x = \theta_2$  e  $x = (\theta_1 + \theta_2)/2$ , obrigando ao uso de métodos de diferenciação para a sua obtenção.

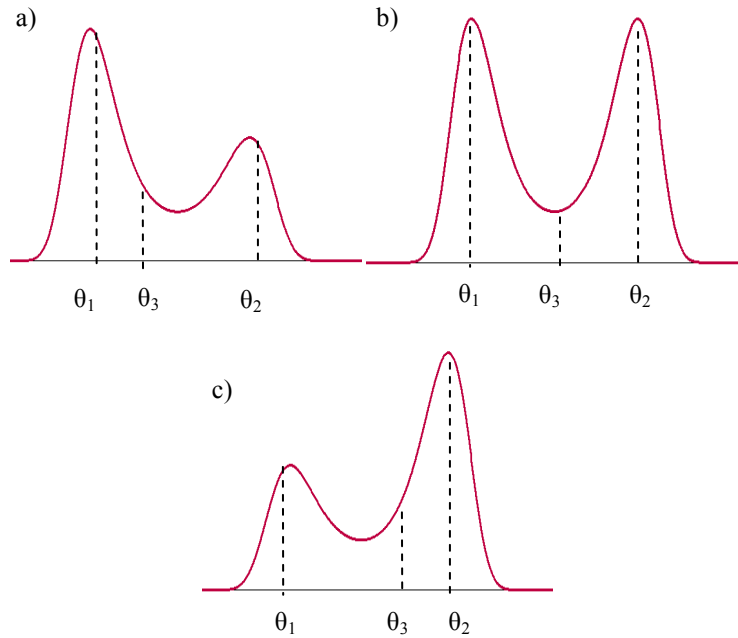


FIGURA 7 Ilustração gráfica da equação (3).

Tendo em vista que o modelo deve permitir o ajustamento a um conjunto de dados amostrais  $X = (x_1, x_2, \dots, x_n)$ , é necessário que o modelo apresente alguma flexibilidade. Para isso, deve-se acrescentar um parâmetro de escala e, assim, obter a expressão

$$f(x; \theta_1, \theta_2, \theta_3, \theta_4) = k \cdot \frac{1}{\theta_4^2} \cdot e^{-\frac{1}{2\theta_4^2} \{[(x-\theta_1)(x-\theta_2)]^2 + (x-\theta_3)^2\}} \quad (4)$$

que é definida como a expressão bimodal desejada para esse estudo.

Para que a expressão (4) seja uma fdp legítima, a condição (iv) proposta no início dessa seção deve ser atendida. Utilizando-se a condição (iv), tem-se a equação

$$\int_{-\infty}^{+\infty} f(x; \theta_1, \theta_2, \theta_3, \theta_4) dx = \int_{-\infty}^{+\infty} k \cdot \frac{1}{\theta_4^2} \cdot e^{-\frac{1}{2\theta_4^2} \{[(x-\theta_1)(x-\theta_2)]^2 + (x-\theta_3)^2\}} dx = 1 \quad (5)$$

Sendo  $k$  uma constante, seu valor é dado pela expressão definida por

$$k = \frac{1}{\int_{-\infty}^{+\infty} \frac{1}{\theta_4^2} \cdot e^{-\frac{1}{2\theta_4^2} \{[(x-\theta_1)(x-\theta_2)]^2 + (x-\theta_3)^2\}} dx} \quad (6)$$

ou, ainda,

$$k = \theta_4^2 \cdot \left[ \int_{-\infty}^{+\infty} e^{-\frac{1}{2\theta_4^2} \{[(x-\theta_1)(x-\theta_2)]^2 + (x-\theta_3)^2\}} dx \right]^{-1} \quad (7)$$

cuja forma explícita ainda é desconhecida. No caso em que os parâmetros são conhecidos, pode ser obtida por métodos numéricos.

### 3.2 Classificação dos parâmetros

Considerando-se o modelo proposto expresso em (4) e substituindo-se os parâmetros  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  e  $\theta_4$  por  $a$ ,  $b$ ,  $c$  e  $d$ , respectivamente, pode-se observar, na Figura 8, o comportamento gráfico para o parâmetro  $a$ . Considerando-se que os parâmetros  $b$ ,  $c$  e  $d$  foram mantidos constantes e que as variações do parâmetro  $a$  foram sob a condição de  $a < b$ , verifica-se que o parâmetro  $a$  se comporta como um parâmetro de posição para o primeiro máximo local da distribuição bimodal obtida.

No gráfico da Figura 9, pode-se observar o comportamento gráfico para o parâmetro  $b$ . Considerando-se que os parâmetros  $a$ ,  $c$  e  $d$  foram mantidos constantes e que as variações do parâmetro  $b$  foram sob a condição de  $b > a$ ,

verifica-se que o parâmetro  $b$  se comporta como um parâmetro de posição para o segundo máximo local da distribuição bimodal obtida.

Do exposto, conclui-se que os parâmetros  $a$  e  $b$  são de posição e identificam os máximos locais da distribuição bimodal.

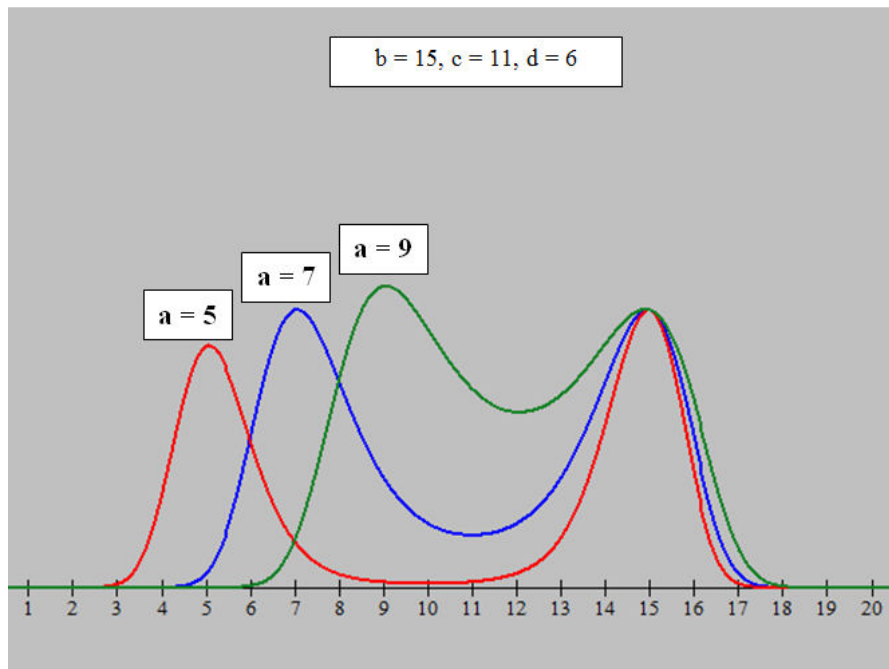


FIGURA 8 Gráficos da densidade de probabilidade obtida para valores variados do parâmetro  $a$  e valores fixos para os parâmetros  $b$ ,  $c$  e  $d$ .

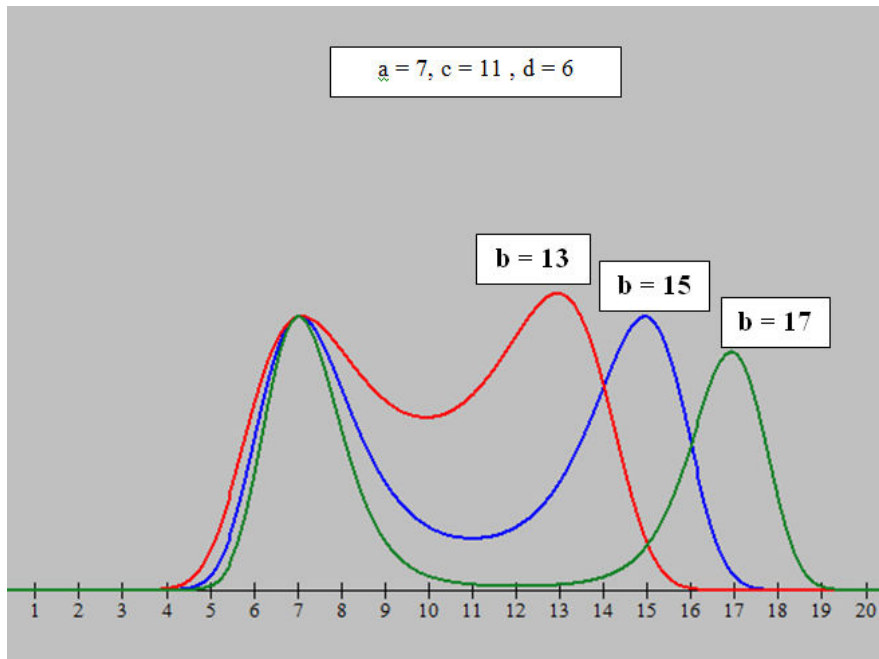


FIGURA 9 Gráficos da densidade de probabilidade obtida para valores variados do parâmetro  $b$  e valores fixos para os parâmetros  $a$ ,  $c$  e  $d$ .

No gráfico da Figura 10, pode-se observar o comportamento gráfico para o parâmetro  $c$ . Considerando-se que os parâmetros  $a$ ,  $b$  e  $d$  foram mantidos constantes e que as variações do parâmetro  $c$  foram sob a condição de  $a < c < b$ , verifica-se que o parâmetro  $c$  se comporta como um parâmetro de forma que pondera a densidade da distribuição obtida para os máximos locais. Quanto mais próximo de um dos parâmetros  $a$  ou  $b$ , maior a densidade atribuída a ele em relação ao outro parâmetro. Quando localizado no ponto médio entre  $a$  e  $b$ , atribui densidades equivalentes aos máximos da distribuição bimodal obtida. Esse resultado sugere que o parâmetro  $c$  substitua o fator de ponderação  $p$  utilizado nas misturas de distribuições.

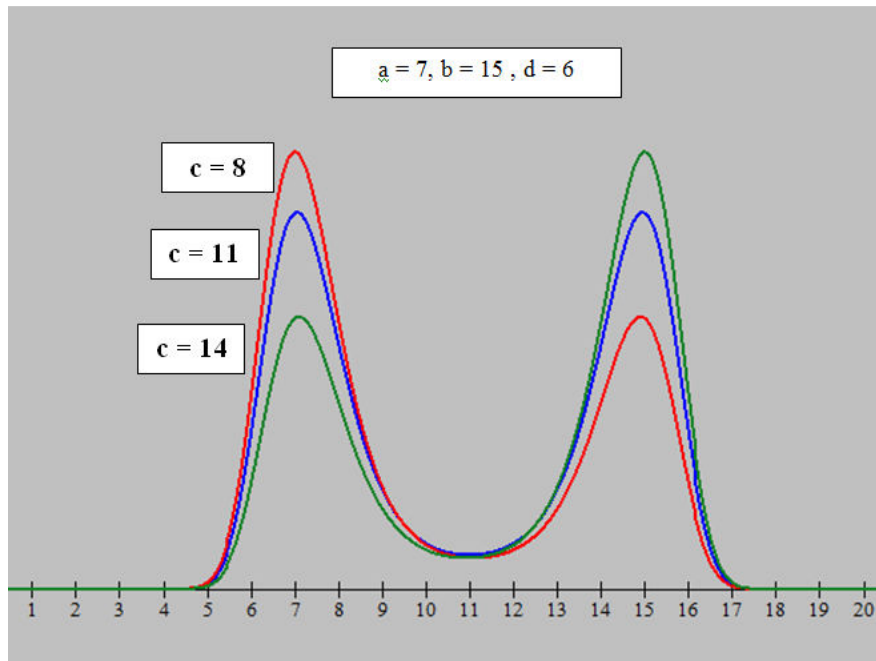


FIGURA 10 Gráficos da densidade de probabilidade obtida para valores variados do parâmetro  $c$  e valores fixos para os parâmetros  $a$ ,  $b$  e  $d$ .

No gráfico da Figura 11, observa-se o comportamento gráfico para o parâmetro  $d$ . Considerando-se que os parâmetros  $a$ ,  $b$  e  $c$  foram mantidos constantes, verifica-se que o parâmetro  $d$  se comporta como um parâmetro de escala. Quanto maior o valor do parâmetro  $d$ , mais achatada encontra-se a distribuição bimodal obtida e vice-versa. Além disso, o valor da antimoda cresce, ou seja, a concentração de dados se reduz nas modas locais e aumenta na antimoda.

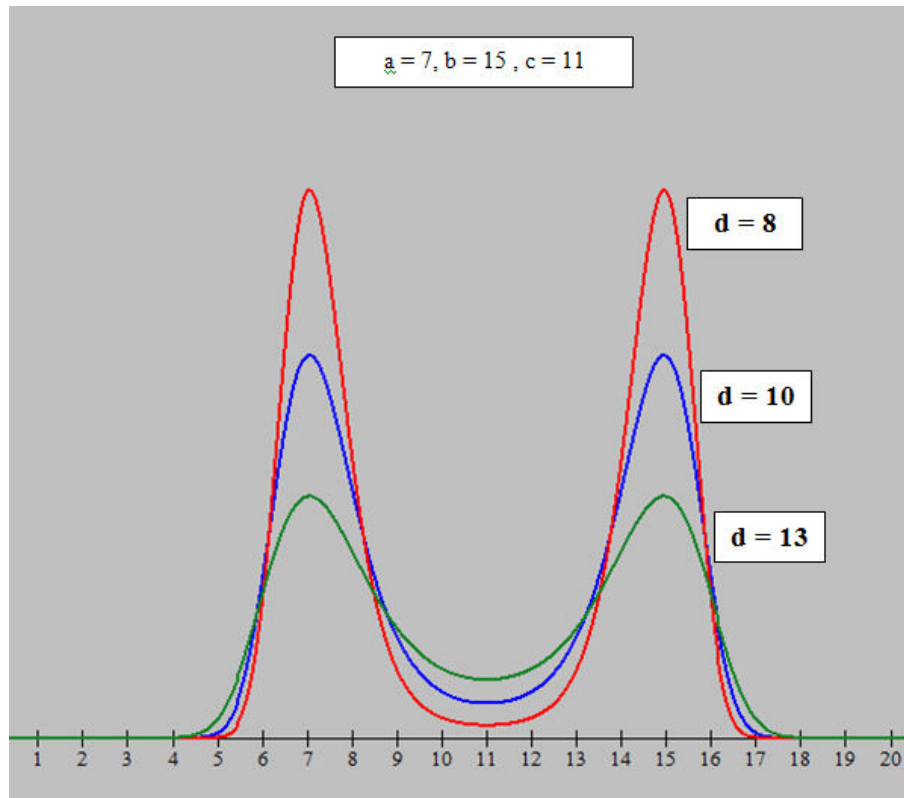


FIGURA 11 Gráficos da densidade de probabilidade obtida para valores variados do parâmetro  $d$  e valores fixos para os parâmetros  $a, b$  e  $c$ .

### 3.3 Família de probabilidades do modelo proposto

Para se verificar se o modelo bimodal proposto expresso em (4) pertence a uma família exponencial de probabilidades, foram também substituídos os parâmetros  $\theta_1, \theta_2, \theta_3$  e  $\theta_4$  por  $a, b, c$  e  $d$ , respectivamente, e, assim, a equação

$$f(x; a, b, c, d) = k \cdot \frac{1}{d^2} \cdot e^{-\frac{1}{2d^2}\{[(x-a)(x-b)]^2 + (x-c)^2\}} \quad (8)$$

tem o desenvolvimento do expoente resultando na expressão



$$f(x; a, b, c, d) = k \cdot \frac{1}{d^2} \cdot e^{-\frac{1}{2d^2}\{x^4 - 2(a+b)x^3 + (a^2+b^2+4ab+1)x^2 - 2(a^2b+ab^2+c)x + (a^2b^2+c^2)\}} \quad (9)$$

E, assim,

$$a_\theta(a, b, c, d) = k \cdot \frac{1}{d^2} \cdot e^{-\frac{1}{2d^2}(a^2b^2+c^2)} \quad b(x) = 1 \quad (10)$$

$$C_1(a, b, c, d) = -\frac{1}{2d^2}; \quad T_1(x) = x^4 \quad (11)$$

$$C_2(a, b, c, d) = +\frac{(a+b)}{d^2}; \quad T_2(x) = x^3 \quad (12)$$

$$C_3(a, b, c, d) = -\frac{(a^2+b^2+4ab+1)}{2d^2}; \quad T_3(x) = x^2 \quad (13)$$

$$C_4(a, b, c, d) = +\frac{(a^2b+ab^2+c)}{d^2}; \quad T_4(x) = x \quad (14)$$

Portanto, como as equações de (10) a (14) atendem aos pressupostos da equação (8), demonstra-se que a função de densidade bimodal proposta pertence à família exponencial de probabilidades. Esse fato permite concluir que, pelo critério da fatorização (Bolfarine & Sandoval, 2001), o vetor de estatísticas para uma amostra aleatória de X extraídos das expressões (11), (12), (13) e (14) que é dado por

$$T_k(X) = (T_1(X), T_2(X), T_3(X), T_4(X)) = \left( \sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i^3, \sum_{i=1}^n X_i^4 \right) \quad (15)$$

é conjuntamente suficiente para o vetor de parâmetros  $(a, b, c, d)$ .

#### 4 APLICAÇÃO EM CANA-DE-AÇÚCAR

Para verificar a performance do modelo com dados reais, foram obtidos dados de 620 observações, correspondentes ao teor de sólidos solúveis (TSS), medido em graus Brix, da safra de cana de 2007, cujo histograma está representado na Figura 12, de uma usina de açúcar e álcool de grande porte da região do município de Araraquara, SP. A partir das informações existentes na literatura sobre os fatores edafoclimáticos influentes no valor final do TSS e análise do histograma dos dados, foi admitido que esse fenômeno tem bimodalidade na sua distribuição e, assim, o modelo proposto está apto a se candidatar a esse ajuste. Nesta seção verifica-se apenas a capacidade do modelo proposto em explicar um fenômeno considerado, *a priori*, bimodal. Portanto, não se pretende aqui compará-lo com outras distribuições.

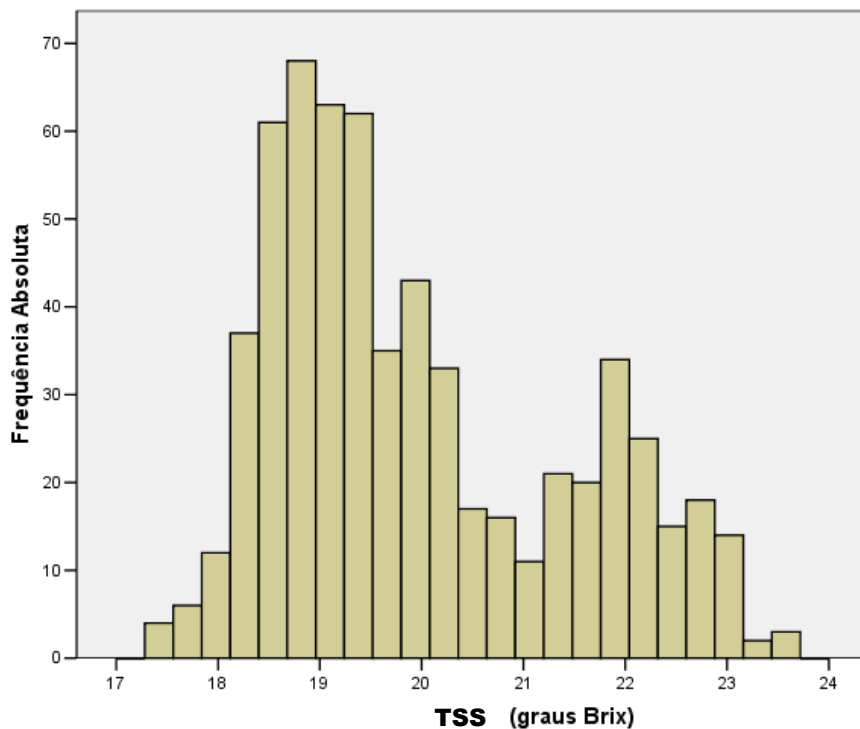


FIGURA 12 Distribuição dos valores obtidos para o teor de sólidos solúveis (TSS), em °Brix, para 620 observações dos cortes de cana-de-açúcar da safra de 2007, de uma usina da região de Araraquara, SP.

#### 4.1 Definição do modelo bimodal

Seja  $X$  a variável aleatória que expressa o teor de sólidos solúveis (graus Brix) da cana-de-açúcar colhida e entregue na usina. Define-se a função densidade de probabilidade bimodal expressa por:

$$f(x; a, b, c, d) = f(x) = k \cdot \frac{1}{d^2} \cdot e^{-\frac{1}{2d^2}\{(x-a)(x-b)\}^2 + (x-c)^2} \quad (16)$$

em que  $a$ ,  $b$ ,  $c$  e  $d$  são, respectivamente, os parâmetros de posição, posição, forma e escala, sendo  $(a, b, c) \in \mathbb{R}$  com  $a < b < c$ ,  $d^2 > 0$ ,  $-\infty < x < +\infty$  e  $k$ , a priori constante, dada por:

$$k = d^2 \cdot \left[ \int_{-\infty}^{+\infty} e^{-\frac{1}{2d^2}\{(x-a)(x-b)\}^2 + (x-c)^2} dx \right]^{-1}. \quad (17)$$

#### 4.2 Estimação dos parâmetros

Para o modelo proposto na equação (16), foram obtidos os estimadores pelo método da máxima verossimilhança, explicitado a seguir.

Seja a função de densidade bimodal dada por:

$$f(x; a, b, c, d) = f(x) = k \cdot \frac{1}{d^2} \cdot e^{-\frac{1}{2d^2}\{(x-a)(x-b)\}^2 + (x-c)^2}$$

A função de verossimilhança é expressa por:

$$\begin{aligned} L &= L(a, b, c, d) = \prod_{i=1}^n f(x_i; a, b, c, d) = \\ &\prod_{i=1}^n \{k \cdot G(x_i; a, b, c, d)\} = \prod_{i=1}^n k \cdot \frac{1}{d^2} \cdot e^{-\frac{1}{2d^2}\{(x_i-a)(x_i-b)\}^2 + (x_i-c)^2} \end{aligned} \quad (18)$$

$$L = \prod_{i=1}^n f(x_i; a, b, c, d) = [k]^n \cdot d^{-2n} \cdot e^{\sum_{i=1}^n -\frac{1}{2d^2}\{(x_i-a)(x_i-b)\}^2 + (x_i-c)^2} \quad (19)$$

Aplicando-se o logarítmo em (19), tem-se

$$\ln L = n \cdot \ln[k] - n \cdot \ln d^2 - \frac{1}{2d^2} \cdot \sum_{i=1}^n [(x_i - a)(x_i - b)]^2 - \frac{1}{2d^2} \cdot \sum_{i=1}^n (x_i - c)^2 \quad (20)$$

Para obter os estimadores de máxima verossimilhança, deve-se encontrar a solução que anula o sistema das derivadas parciais de primeira ordem, também conhecida como vetor de escores dos parâmetros, expresso por:

$$\left\{ \begin{array}{l} \frac{\partial \ln L}{\partial a} = \frac{\partial n \cdot \ln[k]}{\partial a} + d^{-2} \cdot \sum_{i=1}^n [(x_i - a)(x_i - b)]^2 = 0 \\ \frac{\partial \ln L}{\partial b} = \frac{\partial n \cdot \ln[k]}{\partial b} + d^{-2} \cdot \sum_{i=1}^n [(x_i - a)^2(x_i - b)] = 0 \\ \frac{\partial \ln L}{\partial c} = \frac{\partial n \cdot \ln[k]}{\partial c} + d^{-2} \cdot \sum_{i=1}^n (x_i - c) = 0 \\ \frac{\partial \ln L}{\partial d^2} = \frac{\partial n \cdot \ln[k]}{\partial d^2} - n \cdot (d^2)^{-1} + (2d^2)^{-1} \cdot \sum_{i=1}^n [(x_i - a)(x_i - b)]^2 + (2d^2)^{-1} \cdot \sum_{i=1}^n (x_i - c)^2 = 0 \end{array} \right. \quad (21)$$

Observa-se que o sistema de equações (21) não é explícito. Estas equações são dependentes da expressão  $\ln[k]$  que, como visto na equação (17), também não tem solução explícita e é dependente dos parâmetros  $a$ ,  $b$ ,  $c$  e  $d$ . Sendo assim, foram simulados valores para os parâmetros  $a$ ,  $b$ ,  $c$  e  $d$  com base na análise visual do histograma da Figura 12, para verificar o comportamento de  $k$ . Foram utilizadas 12 combinações de valores para os parâmetros para obtenção de valores de  $k$  e da expressão  $G(x_i; a, b, c, d)$  com  $x_i = 20,5$  e são mostradas na Figura 13. Foram utilizados valores de 18,5; 19,0; 19,5 e 20,0 para o parâmetro  $a$ ; 21,0; 21,5 e 22,0 para o parâmetro  $b$ ; 19,0; 19,5; 20,0; 20,5; 21,0 e 21,5 para o parâmetro  $c$  e 2,0; 2,5; 3,0; 3,5 e 4,0 para o parâmetro  $d$ . As combinações foram elaboradas atendendo à condição de que  $a < c < b$  e estão disponíveis no Anexo A.

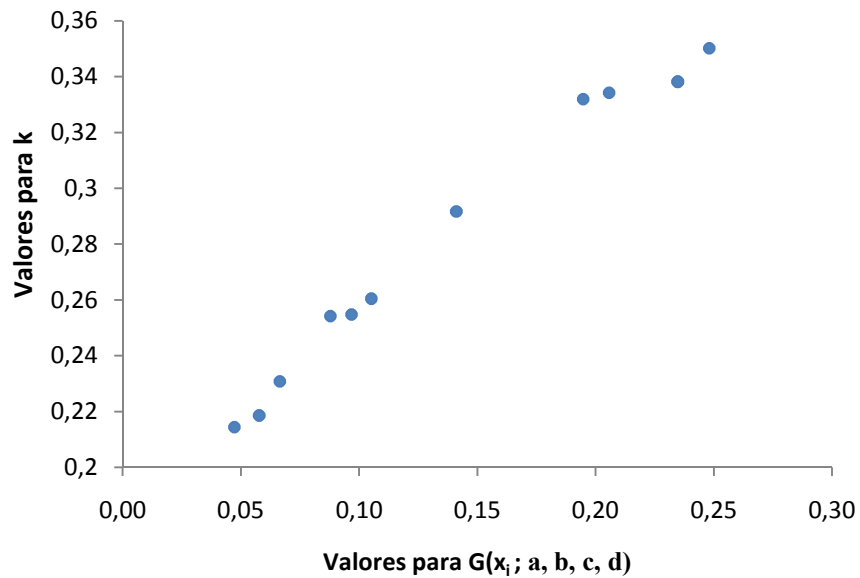


FIGURA 13 Distribuição dos valores obtidos para  $k$  e  $G(x_i; a, b, c, d)$  em função das 12 combinações utilizadas para valores atribuídos aos parâmetros  $a$ ,  $b$ ,  $c$  e  $d$ .

Os valores obtidos mostram que  $k$  não é constante e depende dos parâmetros do modelo. No entanto, observa-se que os valores simulados, estabelecidos a partir da análise visual dos dados, apresentam uma relação monotônica estritamente crescente. Isso mostra que, quando se maximiza a expressão  $G(x_i; a, b, c, d)$  para um dado valor  $x_i$ , maximiza-se também o valor de  $k$  e, assim, por consequência, maximizam-se o produto  $k \cdot G(x_i; a, b, c, d)$  e a função de verossimilhança da equação (18). Sendo assim, a sistema de equações (21) pode ser maximizado apenas com o sistema dado por:

$$\left\{ \begin{array}{l} \frac{\partial \ln L}{\partial a} = d^{-2} \cdot \sum_{i=1}^n [(x_i - a)(x_i - b)^2] = 0 \\ \frac{\partial \ln L}{\partial b} = d^{-2} \cdot \sum_{i=1}^n [(x_i - a)^2(x_i - b)] = 0 \\ \frac{\partial \ln L}{\partial c} = d^{-2} \cdot \sum_{i=1}^n (x_i - c) = 0 \\ \frac{\partial \ln L}{\partial d^2} = -n \cdot (d^2)^{-1} + (2d^2)^{-1} \cdot \sum_{i=1}^n [(x_i - a)(x_i - b)]^2 + (2d^2)^{-1} \cdot \sum_{i=1}^n (x_i - c)^2 = 0 \end{array} \right. \quad (22)$$

cuja solução, agora explícita, é dada por:

$$\hat{a} = \frac{\sum_{i=1}^n [x_i(x_i - \hat{b})^2]}{\sum_{i=1}^n (x_i - \hat{b})^2} = \frac{\sum_{i=1}^n x_i^3 - 2\hat{b} \sum_{i=1}^n x_i^2 + \hat{b}^2 \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - 2\hat{b} \sum_{i=1}^n x_i + n\hat{b}^2} \quad (23)$$

$$\hat{b} = \frac{\sum_{i=1}^n [x_i(x_i - \hat{a})^2]}{\sum_{i=1}^n (x_i - \hat{a})^2} = \frac{\sum_{i=1}^n x_i^3 - 2\hat{a} \sum_{i=1}^n x_i^2 + \hat{a}^2 \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - 2\hat{a} \sum_{i=1}^n x_i + n\hat{a}^2} \quad (24)$$

$$\hat{c} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x} \quad (25)$$

$$\begin{aligned} \hat{d}^2 &= \frac{\sum_{i=1}^n \{[(x_i - \hat{a})(x_i - \hat{b})]^2 + (x_i - \hat{c})^2\}}{2n} = \\ &= \frac{\sum_{i=1}^n x^4 - 2(\hat{b} + \hat{a}) \sum_{i=1}^n x^3 + (\hat{b}^2 + 4\hat{a}\hat{b} + \hat{a}^2) \sum_{i=1}^n x^2 - 2\hat{a}\hat{b}(\hat{b} + \hat{a}) \sum_{i=1}^n x + n\hat{a}^2\hat{b}^2}{n} \\ &+ \frac{n \sum_{i=1}^n x^2 - (\sum_{i=1}^n x)^2}{2n^2} \end{aligned} \quad (26)$$

que são os estimadores de máxima verossimilhança dos parâmetros do modelo proposto.

É importante ressaltar que a derivada segunda dada por:

$$\left\{ \begin{array}{l} \frac{\partial \ln L}{\partial a^2} = -d^{-2} \cdot \sum_{i=1}^n [(x_i - b)^2] < 0 \\ \frac{\partial \ln L}{\partial b^2} = -d^{-2} \cdot \sum_{i=1}^n [(x_i - a)^2] < 0 \\ \frac{\partial \ln L}{\partial c^2} = -n \cdot d^{-2} < 0 \\ \frac{\partial \ln L}{\partial (d^2)^2} = n \cdot (d^2)^{-2} - (2d^2)^{-2} \cdot \sum_{i=1}^n [(x_i - a)(x_i - b)]^2 - (2d^2)^{-2} \cdot \sum_{i=1}^n (x_i - c)^2 < 0 \end{array} \right. \quad (27)$$

garante a maximização da solução do sistema de equações (22).

Para as expressões (23) e (24), pode-se observar que é possível, mas não trivial, explicitar os estimadores de a e b apenas em função dos dados. Sendo assim, devem ser obtidos por iteração simples. O estimador para o parâmetro c é a média amostral e o estimador para d<sup>2</sup> é uma expressão obtida com o vetor de estatística suficiente.

O método dos momentos também foi utilizado para obter estimadores dos parâmetros do modelo bimodal e compará-los com os estimadores de máxima verossimilhança.

A função geradora de momentos é dada pela expressão

$$M_x(t) = E[e^{tx}] \quad (28)$$

Aplicando-se esta expressão à função densidade bimodal proposta, tem-se



$$\begin{aligned}
M_x(t) &= E[e^{tx}] = E[e^{t(x-c+c)}] = E[e^{tx-tc+tc}] = E[e^{tc} \cdot e^{t(x-c)}] = e^{tc} \cdot E[e^{t(x-c)}] \\
&= e^{tc} \cdot \int_{-\infty}^{+\infty} e^{t(x-c)} \cdot k \cdot \frac{1}{d^2} \cdot e^{-\frac{1}{2d^2}\{(x-a)(x-b)\}^2 + (x-c)^2} dx \\
&= e^{tc} \cdot \int_{-\infty}^{+\infty} k \cdot \frac{1}{d^2} \cdot e^{-\frac{1}{2d^2}\{(x-a)(x-b)\}^2 + (x-c)^2 + 2d^2t(x-c)} dx \\
&= e^{tc} \cdot \int_{-\infty}^{+\infty} k \cdot \frac{1}{d^2} \cdot e^{-\frac{1}{2d^2}\{(x-a)(x-b)\}^2 + (x-c)^2 + 2d^2t(x-c) + d^4t^2 - d^4t^2} dx \\
&= e^{tc} \cdot \int_{-\infty}^{+\infty} k \cdot \frac{1}{d^2} \cdot e^{-\frac{1}{2d^2}\{(x-a)(x-b)\}^2 + (x-c+d^2t)^2 - d^4t^2} dx \\
&= e^{tc} \cdot \int_{-\infty}^{+\infty} k \cdot \frac{1}{d^2} \cdot e^{-\frac{1}{2d^2}\{(x-a)(x-b)\}^2 + (x-c+d^2t)^2 - d^4t^2} dx \\
&= e^{tc} \cdot \int_{-\infty}^{+\infty} e^{\frac{d^2t^2}{2}} \cdot k \cdot \frac{1}{d^2} \cdot e^{-\frac{1}{2d^2}\{(x-a)(x-b)\}^2 + (x-c+d^2t)^2} dx \\
&= e^{tc} \cdot e^{\frac{d^2t^2}{2}} \cdot \int_{-\infty}^{+\infty} k \cdot \frac{1}{d^2} \cdot e^{-\frac{1}{2d^2}\{(x-a)(x-b)\}^2 + (x-c+d^2t)^2} dx \\
&= e^{tc} \cdot e^{\frac{d^2t^2}{2}} \cdot \int_{-\infty}^{+\infty} f(a, b, c - d^2t, d) dx = e^{tc + \frac{d^2t^2}{2}} \cdot 1
\end{aligned}$$

ou seja,

$$M_x(t) = e^{tc + \frac{d^2t^2}{2}} \quad (29)$$

As derivadas de primeira, segunda, terceira e quarta ordem são dadas pelas expressões:

$$M_x^I(t) = (c + d^2t) \cdot e^{tc + \frac{d^2t^2}{2}} \quad (30)$$

$$M_x^{II}(t) = d^2 \cdot e^{tc + \frac{d^2t^2}{2}} + (c + d^2t)^2 \cdot e^{tc + \frac{d^2t^2}{2}} \quad (31)$$

$$M^{III}_x(t) = d^2 \cdot (c + d^2t) \cdot e^{tc + \frac{d^2t^2}{2}} + 2d^2(c + d^2t) \cdot e^{tc + \frac{d^2t^2}{2}} + (c + d^2t)^3 \cdot e^{tc + \frac{d^2t^2}{2}} \quad (32)$$

$$M^{IV}_x(t) = 3d^4 \cdot e^{tc + \frac{d^2t^2}{2}} + 6d^2(c + d^2t)^2 \cdot e^{tc + \frac{d^2t^2}{2}} + (c + d^2t)^4 \cdot e^{tc + \frac{d^2t^2}{2}} \quad (33)$$

Por meio das equações (29) a (32) obtêm-se os momentos populacionais, em  $t = 0$ , que resultam em

$$E(X) = M^I_x(0) = c \quad (34)$$

$$E(X^2) = M^{II}_x(0) = c^2 + d^2 \quad (35)$$

$$E(X^3) = M^{III}_x(0) = c^3 + 3d^2c \quad (36)$$

$$E(X^4) = M^{IV}_x(0) = 3d^4 + 6d^2c^2 + c^4 \quad (37)$$

De acordo com Mood et al. (1974), o primeiro momento populacional dado por  $\mu_1 = E(X)$  é a média populacional e os demais momentos centrais são dados pela expressão  $\mu_r = E[(X - E(X))^r]$  para  $r > 1$  sendo  $\mu_2 =$  variância.

O primeiro momento populacional simples e o segundo, o terceiro e o quarto momentos populacionais centrais, respectivamente, são

$$\mu_1 = E(X) = c$$

$$\begin{aligned}\mu_2 &= E \left[ (X - E(X))^2 \right] = E[X^2 - 2X \cdot E(X) + [E(X)]^2] \\ &= E(X^2) - 2E(X) \cdot E(X) + [E(X)]^2 \\ &= c^2 + d^2 - 2c^2 + c^2 \\ &= d^2\end{aligned}$$

$$\begin{aligned}\mu_3 &= E \left[ (X - E(X))^3 \right] = E(X^3 - 3X^2 \cdot E(X) + 3X[E(X)]^2 - [E(X)]^3) \\ &= E(X^3) - 3E(X^2) \cdot E(X) + 3E(X) \cdot [E(X)]^2 - [E(X)]^3 \\ &= (c^3 + 3d^2c) - 3(c^2 + d^2)c + 3c[c]^2 - c^3 \\ &= 0\end{aligned}$$

$$\begin{aligned}\mu_4 &= E \left[ (X - E(X))^4 \right] \\ &= E(X^4 - 4X^3 \cdot E(X) + 6X^2[E(X)]^2 - 4X \cdot [E(X)]^3 + [E(X)]^4) \\ &= E(X^4) - 4E(X^3) \cdot E(X) + 6E(X^2)[E(X)]^2 - 4E(X) \cdot [E(X)]^3 + [E(X)]^4 \\ &= 3d^4 + 6d^2c^2 + c^4 - 4(c^3 + 3d^2c) \cdot c + 6(c^2 + d^2)c^2 - 4c(c)^3 + c^4 \\ &= 3d^4\end{aligned}$$

De acordo com os resultados, pode-se observar que o método dos momentos não permite determinar os estimadores para os parâmetros  $a$  e  $b$ . O estimador do parâmetro  $c$  foi o mesmo para os dois métodos e o estimador do parâmetro  $d^2$  apresentou diferenças pelos dois métodos. Pelo método dos momentos, o parâmetro  $d^2$  foi equivalente ao segundo momento amostral, definido como a variância populacional, diferindo da expressão encontrada pelo método da máxima verossimilhança.

Com base nos EMV, as estimativas para os parâmetros  $a$  e  $b$  foram obtidas com as equações (23) e (24). Como essas expressões para  $\hat{a}$  e  $\hat{b}$  não são explícitas, seus valores foram obtidos por um processo iterativo simples com valor inicial (semente)  $b_0 = 0$  na equação (23) seguida da equação (24).

Na Tabela 1 são apresentados os resultados das etapas desse cálculo e se observa que a convergência com erro menor que  $10^{-10}$  ocorreu rapidamente após cinco iterações resultando nas estimativas de  $\hat{a} \cong 18,911025$  e  $\hat{b} \cong 21,934243$ . Foram colocados outros valores para a semente  $b_0$  que variaram entre  $-10^6$  e  $10^6$ , bem como valores da ordem de  $10^{-5}$ , e a convergência foi imediata e idêntica aos valores anteriores. Isso mostra que, nesta estimação, mesmo com a necessidade da iteração, o valor da semente inicial para a obtenção destas estimativas pode ser aleatório.

TABELA 1 Valores obtidos no algoritmo de convergência para as estimativas de  $\hat{a}$  e  $\hat{b}$ .

$b_0$	$\hat{a}$	$\hat{b}$
0	20,172413665906000	20,450996909645000
	19,910275105443800	20,984553786911000
	19,214636730885700	21,865064190015500
	18,911907056386800	21,934242265344800
	18,911024945430000	21,934242749250400
	18,911024945429200	21,934242749250400
	<b>18,911024945429300</b>	<b>21,934242749250700</b>

As estimativas para  $\hat{c}$  e  $\hat{d}^2$  foram obtidas diretamente com o uso das equações (25) e (26). O valor de  $\hat{c} = 19,9611$  corresponde à média dos dados amostrais. Para  $\hat{d}^2$  foi obtido o valor de 2,92627 que corresponde à aplicação na equação (26). O valor de  $k = 0,780345$  corresponde à aplicação de métodos numéricos na equação (17). Sendo assim, a função de densidade de probabilidade bimodal proposta para os dados amostrais é dada por

$$f(x) = 0,780345 \cdot \frac{1}{2,92627} \cdot e^{-\frac{1}{5,85254}\{(x-18,911025)(x-21,93424)\}^2 + (x-19,96111)^2}$$

e resulta em:

$$f(x) = 0,266669 \cdot e^{-\frac{1}{5,85254}\{(x-18,911025)(x-21,93424)\}^2 + (x-19,96111)^2}$$

### 4.3 Pontos críticos

Os pontos críticos de máximos e mínimos locais da função bimodal proposta são obtidos pelos valores que anulam a primeira derivada de  $f$  e o sinal de sua derivada de segunda ordem para os pontos críticos determina a classificação em máximo ou mínimo local. A seguir, é mostrado o desenvolvimento da expressão que permite obtê-los.

Seja a função de densidade bimodal dada por:

$$f(x; a, b, c, d) = k \cdot \frac{1}{d^2} \cdot e^{-\frac{1}{2d^2}\{(x-a)(x-b)\}^2 + (x-c)^2} \quad (38)$$

Para facilitar o cálculo da primeira derivada ( $f'(x)$ ), faz-se

$$g(x) = -\frac{1}{d^2} \cdot [(x-a)(x-b)]^2 \quad e \quad t(x) = e^{g(x)} \quad (39)$$

$$u(x) = -\frac{1}{d^2} \cdot (x-c)^2 \quad e \quad m(x) = e^{u(x)} \quad (40)$$

E, assim,

$$\begin{aligned} g'(x) &= -\frac{1}{d^2} \cdot [2(x-a)(x-b)^2 + 2(x-a)^2(x-b)] \\ &= -\frac{1}{d^2} \cdot [2(x-a)(x-b) \cdot (2x-a-b)] \end{aligned} \quad (41)$$

$$u'(x) = -\frac{1}{d^2} \cdot 2(x-c) \quad (42)$$

Da equação (34) tem-se

$$f(x; a, b, c, d) = k \cdot \frac{1}{d^2} \cdot e^{-\frac{1}{2d^2}\{(x-a)(x-b)\}^2} \cdot e^{-\frac{1}{2d^2}(x-c)^2} = k \cdot \frac{1}{d^2} \cdot t(x) \cdot m(x) \quad (43)$$

cuja derivada é

$$f'(x) = k \cdot \frac{1}{d^2} \cdot [t'(x) \cdot m(x) + t(x) \cdot m'(x)] \quad (44)$$

Substituindo-se (35) e (36) em (40) resulta em

$$\begin{aligned}
 f'(x) &= k \cdot \frac{1}{d^2} \cdot [g'(x) \cdot e^{g(x)} \cdot e^{u(x)} + e^{g(x)} \cdot u'(x) \cdot e^{u(x)}] \\
 &= k \cdot \frac{1}{d^2} \cdot [e^{g(x)} \cdot e^{u(x)} \cdot [g'(x) + u'(x)]] \\
 &= k \cdot \frac{1}{d^2} \cdot [g'(x) + u'(x)] \cdot e^{g(x) + u(x)}
 \end{aligned} \tag{45}$$

Fazendo  $f'(x) = 0$  para encontrar os pontos críticos e sabendo-se que as expressões  $k$ ,  $\frac{1}{d^2}$  e  $e^{g(x)+u(x)}$  são sempre positivas, tem-se que a solução da equação

$$g'(x) + u'(x) = 0 \tag{46}$$

fornece os pontos críticos para a função de densidade bimodal proposta. Portanto, substituindo-se (41) e (42) em (46), tem-se

$$g'(x) + u'(x) = -\frac{1}{d^2} \cdot [2(x-a)(x-b) \cdot (2x-a-b)] - \frac{1}{d^2} \cdot 2(x-c) = 0$$

que, simplificada, resulta na equação polinomial

$$2x^3 - 3(a+b)x^2 + [(a+b)^2 + 2ab + 1]x - (a^2b + ab^2 + c) = 0 \tag{47}$$

cujas raízes reais  $\alpha$ ,  $\gamma$  e  $\beta$  são os pontos críticos de função bimodal proposta.

Na equação (47) observa-se que os pontos críticos não dependem do parâmetro de escala e utilizam-se os estimadores  $\hat{a}$ ,  $\hat{b}$  e  $\hat{c}$  para a determinação dos pontos críticos  $\alpha$ ,  $\gamma$  e  $\beta$ .

Em vista dos resultados supracitados, os valores dos pontos críticos para os dados utilizados foram obtidos substituindo-se os parâmetros  $a$ ,  $b$  e  $c$  pelas suas estimativas obtidas em  $\hat{a}$ ,  $\hat{b}$  e  $\hat{c}$  no polinômio da equação (47) e, assim, obteve-se

$$2x^3 - 122,5358x^2 + 2.498,9339x - 16.962,5378 = 0$$

cujas raízes são  $\alpha \cong 19,0261$ ,  $\gamma \cong 20,5532$  e  $\beta \cong 21,6886$ .

Adicionalmente, com a segunda derivada de  $f$  expressa por

$$f''(x) = k \cdot \{g''(x) + u''(x) + [g'(x) + u'(x)]^2\} \cdot e^{g(x)+u(x)} \quad (48)$$

é possível verificar que, para  $\alpha \cong 19,0261$  e  $\beta \cong 21,6886$ , tem-se  $f''(x) < 0$  e, assim, são pontos de máximos locais. Por conseguinte, para  $\gamma \cong 20,5532$  tem-se  $f''(x) > 0$  que o torna um ponto de mínimo local. Os valores de  $\alpha$  e  $\beta$  são as modas enquanto  $\gamma$  é a antimoda da distribuição bimodal proposta.

#### 4.4 Desempenho do modelo

Com a amostra utilizada e as estimativas obtidas foram construídos o histograma, a função de densidade de probabilidade e a função acumulada de probabilidade apresentadas na Figura 14. Foi utilizado, para a construção do histograma, o software livre R e a rotina elaborada para os cálculos está disponível no Anexo B.

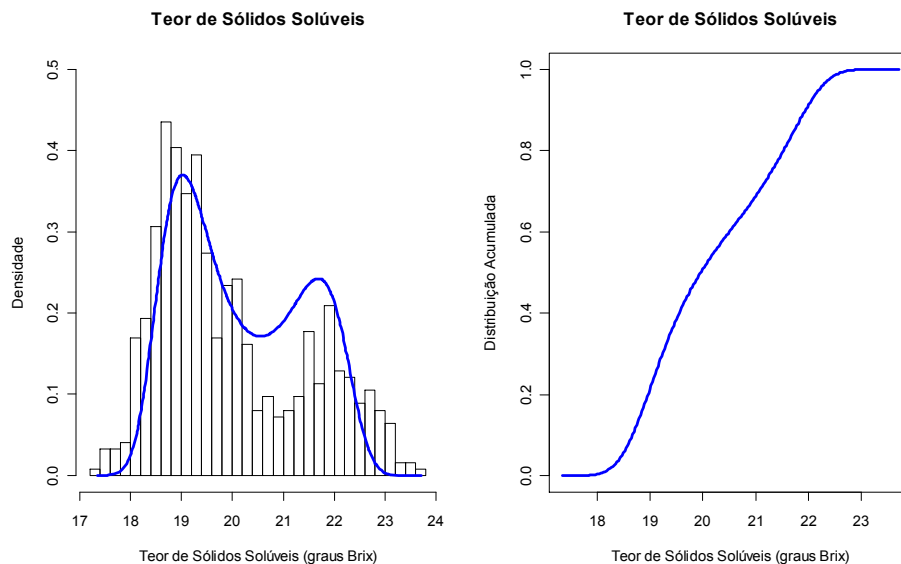


FIGURA 14 Gráfico das funções de densidades de probabilidade e função de distribuição acumulada para o modelo bimodal proposto ajustado para as 620 observações do teor de sólidos solúveis.

Os resultados mostram que o modelo proposto atendeu ao objetivo de explicar a bimodalidade do TSS na cana-de-açúcar colhida na safra de 2007 na usina. Visualmente, é possível verificar que houve um bom ajustamento da curva ao histograma.

Outro ponto relevante associado ao modelo estimado é seu domínio da variável aleatória que apresenta probabilidade não nula. O ajuste ocorreu numa amplitude de domínio menor do que os dados porque o modelo apresenta uma dupla forma quadrática, que o faz convergir rapidamente nas extremidades. No caso particular, pode-se observar que o modelo se ajusta no domínio do TSS entre 18 e 23°Brix e não sofre muita influência dos valores extremos dos dados. Adicionalmente, o modelo se ajusta bem à prática das usinas considerando que os administradores têm interesse que a colheita ocorra sempre a partir de uma



medida de TSS de 18°Brix e, conseqüentemente, a produção de açúcar e álcool na usina tenha viabilidade econômica.

Utilizando-se o valor obtido para a antimoda dado por  $\gamma = 20,55^\circ\text{Brix}$ , pode-se dividir o TSS em duas populações mais homogêneas. Uma população com TSS médio mais baixo concentrada em torno do valor  $\alpha = 19,03^\circ\text{Brix}$  e outra com TSS médio mais alto concentrada em torno do valor  $\beta = 21,69^\circ\text{Brix}$ .

Foi obtido  $P(X < 20,55) = 0,609$  e, portanto, pode-se inferir que, para essa usina, 60,9% da cana colhida na safra de 2007 chegou à usina com um valor médio inferior de TSS da ordem de 19,0°Brix, enquanto os demais 39,1% tiveram um valor médio superior de TSS da ordem de 21,7°Brix. Com essas informações e de posse dos dados das covariáveis associadas ao TSS, por exemplo, é possível avaliar e comparar características que distinguem essas duas populações e intervir para a melhoria do TSS que chega à usina para processamento em safras vindouras.

Enfim, sugere-se o uso do modelo de densidade bimodal proposto, quando possível, como alternativa aos modelos de misturas de distribuições por se tratar de uma forma mais simples para se tratar as variáveis bimodais e, assim, contribuir para popularizar a modelagem em áreas científicas com menor afinidade para a estatística computacional avançada.

## 5 CONCLUSÕES

Foi possível ajustar o modelo proposto a dados reais com comportamento bimodal, sem a necessidade de mistura de distribuições e com estimadores explícitos e, em poucas ocasiões, com uso de métodos numéricos simples.

Com o uso do modelo bimodal proposto e a disponibilidade de dados de covariáveis edafoclimáticas associadas ao teor de sólidos solúveis da cana-de-açúcar que chega para processamento, podem-se promover estudos comparativos para se avaliar a influência desses fatores e, assim, planejar intervenções que melhorem a qualidade da cana-de-açúcar colhida que chega à usina nas safras futuras.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

ABRAHÃO, C. P.; VARELLA, C. A. A.; PINTO, F. A. C.; KHOURY JÚNIOR, J. K. Quantificação da falha na madeira em juntas coladas utilizando técnicas de visão artificial. **Revista Árvore**, Viçosa, MG, v. 27, n. 1, p. 71-78, jan. 2003.

ANÉ, T.; LABIDI, C. Revisiting the finite mixture of Gaussian distributions with application to futures markets. **Journal of Futures Markets**, New York, v. 21, n. 4, p. 347-376, Apr. 2001.

ASSIS, P. C. O.; LACERDA, R. D.; AZEVEDO, H. M.; DANTAS NETO, J.; FARIAS, C. H. A. Resposta dos parâmetros tecnológicos da cana-de-açúcar a diferentes lâminas de irrigação e adubação. **Revista de Biologia e Ciências da Terra**, João Pessoa, v. 4, n. 2, p. 1-12, jul. 2004.

BOLFARINI, H.; SANDOVAL, M. C. **Introdução à inferência estatística**. Rio de Janeiro: Sociedade Brasileira de Matemática, 2001. 125 p.

CALIARI, M.; SOARES JÚNIOR, M. S.; SERRA, G. E. Programação matemática aplicada à colheita de cana-de-açúcar com ênfase na recuperação máxima de açúcar. **Pesquisa Agropecuária Tropical**, Goiânia, v. 35, n. 1, p. 1-8, jan. 2004.

CHOY, S. T. B.; CHAN, J. S. K. Scale mixtures distributions in statistical modeling. **Australian & New Zealand Journal of Statistics**, Oxford, v. 50, n. 2, p. 135-146, June 2008.

CHU, H.; REN, S.; COLE, S. R. Sample size calculations for the mean in a two component nonstandard mixture distribution. **Biometrical Journal**, Berlin, v. 46, n. 5, p. 565-571, Oct. 2004.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum Likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society B**, London, v. 39, n. 1, p. 1- 38, Jan. 1977.

DEY, D. K. Estimation of scale parameters in mixture distributions. **Canadian Journal of Statistics**, Montreal, v. 18, n. 2, p. 171-178, June 1990.

FISHER, R. A. On an absolute criterion for fitting frequency curves. **Messenger of Mathematics**, London, v. 41, p. 155-160, 1912.

JIANG, R.; MURTHY, D. N. P. Mixture of Weibull distributions-parametrics characterization of failure rate function. **Applied Stochastic Models and Data Analysis**, West Sussex, v. 14, n. 1, p. 47-65, Mar. 1998.

LINDSEY, J. K. **Parametric statistical inference**. Oxford: Clarendon, 1996. 490 p.

MCLACHLAN, G. J.; PEEL, D. **Finite mixture models**. New York: J. Wiley, 2000. 419 p.

MENG, X. L.; RUBIN, D. B. Maximum likelihood estimation via the ECM algorithm: a general framework. **Biometrika**, Cambridge, v. 80, n. 2, p. 267-278, June 1993.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. New York: McGraw-Hill, 1974. 564 p.

MORAIS, A. R.; BOTELHO, V. A.; CARVALHO, L. G.; MUNIZ, J. A.; LAGE, G. Estimativa da precipitação provável em Lavras (MG) através da distribuição Gama. **Revista Brasileira de Agrometeorologia**, Santa Maria, v. 9, n. 2, p. 305-310, set. 2001.

R DEVELOPMENT CORE TEAM. R: a language and environment for statistical computing. Viena: R Foundation for Statistical Computing, 2008. Disponível em: <<http://www.R-project.org>>. Acesso em: 30 jun. 2009.

SILVA, F. A. M. da; ASSAD, E. D.; MATTOS, A.; LUIZ, A. J. B. Variação espaço-temporal da disponibilidade hídrica climática no estado de Goiás. **Pesquisa Agropecuária Brasileira**, Brasília, v. 33, n. 5, p. 605-612, maio 1998.

SILVA, R. R. **A distribuição generalizada de Pareto e mistura de distribuições de Gumbel no estudo da vazão e da velocidade máxima do vento em Piracicaba, SP.** 2008. 100 p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) – Universidade de São Paulo, Escola Superior de Agricultura Piracicaba, “Luiz de Queiróz”, São Paulo.

SILVA, R. R.; ZOCCHI, S. S. Densidades não paramétricas no estudo de velocidades máximas de vento em Piracicaba-SP. In: REUNIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 50., 2005; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 11., 2005, Londrina...**Anais...** Londrina: UEL, 2005. 1 CD-ROM.

SILVA, W. P. da. **Estimadores de máxima verossimilhança em misturas de normais:** uma aplicação em genética. 2003. 60 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras.

SILVEIRA NETO, T. Software especial para programação de colheita. **Álcool & Açúcar**, São Paulo, v. 12 n. 61, p. 38-41, 1992.

TAUPIER, L. O. G.; RODRÍGUES, G. G. A. Cana-de-açúcar. In: INSTITUTO CUBANO DE PESQUISAS DOS DERIVADOS DE CANA-DE-AÇÚCAR - ICIDCA. **Manual dos derivados da cana-de-açúcar:** diversificação, matérias-primas, derivados do bagaço, derivados do melaço, outros derivados, resíduos, energia. Brasília: ABIPTI, 1999. cap. 21, p. 21-27.

TIAGO, R. S. L.; FARIA, F. P.; PONTES, P. A. L.; BRASIL, O. O. C. Aspectos morfométricos do nervo hipoglosso humano em adultos e idosos. **Revista Brasileira de Otorrinolaringologia**, Rio de Janeiro, v. 71, n. 5, p. 554-558, set./out. 2005.

TIWARI, R. C.; ZALKIKAR, J. N. Tests for a trend in proportion based on mixtures of beta distributions incorporating historical controls. **Environmetrics**, Chichester, v. 10, n. 1, p. 1-22, Jan./Feb. 1999.

VIDAL, A. C.; SANCEVERO, S. S.; REMACRE, A. Z.; COSTANZO, C. P. Modelagem geoestatística 3D da impedância acústica para a caracterização do campo de namorado. **Revista Brasileira de Geofísica**, São Paulo, v. 25, n. 3, p. 295-305, jul./set. 2007.

YANG, X.; BEYENAL, H.; HARKIN, G.; LEWANDOWSKI, Z. Evaluation of biofilm image thresholding methods. **Water Research**, New York, v. 35, n. 5, p. 1149-1158, Apr. 2001.

## ANEXOS

	<b>Página</b>
ANEXO A	
TABELA 1A Valores das combinações utilizadas para simular valores de $k$ e $G(x_i, a, b, c, d)$ .....	51
ANEXO B	
Rotina para gerar a função de densidade bimodal proposta.....	52

## ANEXO A

TABELA 1A Valores das combinações utilizadas para simular valores de  $k$  e  $G(x_i, a, b, c, d)$ .

Combinação	Parâmetros				$x_i$	$k$	$G(x_i, a, b, c, d)$
	$a$	$b$	$c$	$d$			
1	18,5	21,0	19,5	2,0	20,5	0,3320	0,1947
2	18,5	21,5	20,0	3,0	20,5	0,2543	0,0877
3	18,5	22,0	20,5	4,0	20,5	0,2145	0,0472
4	19,0	21,0	19,5	2,0	20,5	0,3343	0,2056
5	19,0	21,5	20,0	3,0	20,5	0,2548	0,0967
6	19,0	22,0	20,5	3,5	20,5	0,2309	0,0664
7	19,5	21,0	20,0	2,0	20,5	0,3381	0,2349
8	19,5	21,5	20,5	3,0	20,5	0,2604	0,1051
9	19,5	22,0	21,0	4,0	20,5	0,2186	0,0578
10	20,0	21,0	20,5	2,0	20,5	0,3502	0,2481
11	20,0	21,5	21,0	2,0	20,5	0,3381	0,2349
12	20,0	22,0	21,5	2,5	20,5	0,2917	0,1412



## ANEXO B

### Rotina para gerar a função de densidade bimodal proposta

```
##Rotina para gerar a função densidade Bimodal##
##Arquivo de dados##
setwd("/Documents and Settings/Usuario/Desktop/R-Rotinas Tese/")
dados<-read.table("dadosbrixtese.txt", h=T)
#O arquivo dados brixtese.txt é um arquivo de 620 linhas e 1
coluna      com      os      dados      do      vetor
c(17.35,17.42,17.45,17.55,17.57,17.68,17.78,17.78,17.78,17.84,17.9
2,17.95,17.97,18.00,18.01,18.02,18.02,18.04,18.04,18.05,18.05,18.0
7,18.12,18.12,18.13,18.14,18.15,18.16,18.16,18.16,18.17,18.19,18.1
9,18.19,18.20,18.21,18.22,18.24,18.25,18.25,18.26,18.27,18.28,18.29,18.3
0,18.33,18.35,18.35,18.35,18.36,18.36,18.36,18.36,18.36,18.37,18.3
7,18.39,18.40,18.40,18.40,18.41,18.42,18.44,18.44,18.44,18.45,18.4
5,18.46,18.47,18.48,18.48,18.48,18.48,18.50,18.50,18.51,18.52,18.5
3,18.54,18.54,18.55,18.55,18.55,18.56,18.56,18.56,18.56,18.56,18.5
7,18.58,18.58,18.59,18.59,18.59,18.59,18.60,18.60,18.60,18.61,18.6
1,18.62,18.62,18.62,18.62,18.63,18.63,18.63,18.63,18.63,18.64,18.6
4,18.64,18.64,18.65,18.65,18.65,18.65,18.66,18.66,18.66,18.67,18.6
9,18.69,18.70,18.70,18.71,18.72,18.72,18.72,18.73,18.73,18.73,18.7
3,18.73,18.74,18.74,18.75,18.75,18.75,18.75,18.76,18.77,18.77,18.7
7,18.78,18.78,18.78,18.79,18.79,18.79,18.80,18.80,18.82,18.82,18.8
4,18.84,18.84,18.85,18.85,18.87,18.88,18.89,18.89,18.89,18.89,18.8
9,18.89,18.89,18.90,18.91,18.91,18.91,18.91,18.93,18.93,18.93,18.9
3,18.93,18.94,18.94,18.94,18.94,18.94,18.94,18.94,18.94,18.95,18.9
5,18.95,18.96,18.96,18.96,18.97,18.98,18.99,18.99,19.00,19.00,19.0
0,19.00,19.00,19.00,19.01,19.02,19.02,19.03,19.03,19.04,19.04,19.0
4,19.04,19.04,19.05,19.06,19.06,19.07,19.07,19.08,19.09,19.09,19.1
0,19.10,19.10,19.11,19.11,19.11,19.11,19.11,19.12,19.12,19.13,19.1
3,19.13,19.13,19.15,19.15,19.16,19.16,19.17,19.17,19.17,19.18,19.1
8,19.19,19.20,19.21,19.22,19.23,19.23,19.23,19.23,19.24,19.25,19.2
5,19.26,19.26,19.27,19.27,19.27,19.27,19.27,19.28,19.28,19.28,19.2
9,19.29,19.30,19.30,19.31,19.32,19.32,19.32,19.33,19.34,19.35,19.3
5,19.35,19.35,19.36,19.36,19.36,19.36,19.37,19.37,19.37,19.38,19.3
8,19.38,19.38,19.38,19.38,19.39,19.40,19.40,19.41,19.41,19.41,19.4
1,19.42,19.42,19.42,19.44,19.45,19.46,19.46,19.47,19.48,19.50,19.5
0,19.50,19.51,19.51,19.52,19.52,19.53,19.53,19.53,19.53,19.55,19.5
5,19.56,19.57,19.57,19.58,19.59,19.60,19.60,19.60,19.61,19.62,19.6
3,19.63,19.64,19.64,19.67,19.68,19.69,19.70,19.70,19.71,19.71,19.7
1,19.72,19.73,19.74,19.74,19.75,19.76,19.78,19.82,19.83,19.85,19.8
6,19.86,19.88,19.89,19.90,19.90,19.90,19.91,19.91,19.92,19.92,19.9
3,19.94,19.96,19.96,19.96,19.96,19.96,19.97,19.97,19.97,19.99,19.9
9,20.00,20.00,20.00,20.01,20.01,20.01,20.03,20.03,20.03,20.03,20.0
5,20.05,20.06,20.06,20.06,20.08,20.08,20.09,20.10,20.12,20.13,20.1
3,20.14,20.14,20.15,20.16,20.16,20.17,20.18,20.19,20.19,20.20,20.2
0,20.22,20.23,20.24,20.24,20.25,20.26,20.28,20.29,20.30,20.31,20.3
1,20.33,20.34,20.35,20.35,20.36,20.36,20.37,20.37,20.40,20.41,20.4
3,20.43,20.46,20.47,20.49,20.53,20.55,20.56,20.59,20.61,20.63,20.6
3,20.63,20.68,20.70,20.73,20.74,20.76,20.76,20.77,20.78,20.81,20.8
```

```

3,20.83,20.87,20.88,20.90,20.90,20.91,20.99,21.02,21.12,21.12,21.1
3,21.13,21.15,21.18,21.18,21.19,21.20,21.22,21.22,21.23,21.24,21.2
4,21.27,,21.29,21.29,21.33,21.34,21.34,21.37,21.41,21.41,21.43,21.
43,21.43,21.44,21.45,21.46,21.46,21.48,21.50,21.50,21.51,21.52,21.
55,21.56,21.57,21.58,21.58,21.59,21.59,21.60,21.62,21.63,21.63,21.
64,21.64,21.74,21.74,21.76,21.77,21.77,21.79,21.79,21.80,21.80,21.
81,21.81,21.83,21.84,21.84,21.84,21.85,21.85,21.85,21.86,21.86,21.
88,21.89,21.92,21.93,21.93,21.94,21.94,21.96,21.96,21.97,21.97,21.
97,21.98,22.00,22.00,22.04,22.08,22.08,22.09,22.11,22.11,22.11,22.
12,22.12,22.12,22.12,22.13,22.16,22.16,22.18,22.19,22.22,22.23,22.
25,22.25,22.26,22.26,22.27,22.27,22.29,22.29,22.32,22.35,22.37,22.
37,22.38,22.41,22.46,22.47,22.49,22.50,22.50,22.50,22.52,22.53,22.
55,22.60,22.62,22.62,22.65,22.65,22.66,22.67,22.67,22.68,22.69,22.
71,22.74,22.80,22.80,22.81,22.83,22.84,22.85,22.90,22.92,22.93,22.
96,22.99,23.00,23.01,23.01,23.01,23.03,23.06,23.06,23.11,23.12,23.
25,23.40,23.45,23.47,23.71)#

```

```

attach(dados)
##Estimativas dos Parâmetros##
estimacao=function(x,fator)
{
x=as.vector(x)
x2=x^2
x3=x^3
x4=x^4
s1=sum(x)
s2=sum(x2)
s3=sum(x3)
s4=sum(x4)
c=as.numeric(mean(x))
d2=sum((x-a)^2*(x-b)^2+(x-c)^2)/(2*n)
b=0
erro=10000
n=nrow(x)
i=1
A=0
B=0
while(erro>fator)
{
if(i>1)
{
A=a
B=b
}
a=(s3-2*b*s2+(b^2)*s1)/(s2-2*b*s1+n*(b^2))
b=(s3-2*a*s2+(a^2)*s1)/(s2-2*a*s1+n*(a^2))
Erroa=abs(A-a)
Errob=abs(B-b)
erro=max(Eroa,Errob)
i=i+1
}
a1=min(a,b)

```

```

b1=max(a,b)
a=a1
b=b1

##Cálculo do valor k##
d2=sum((x-a)^2*(x-b)^2+(x-c)^2)/(2*n)
d1=1/d2
fd=function(y){
(1/d2)*exp((-1/(2*d2))*((y-a)^2*(y-b)^2+(y-c)^2))}
k=integrate(fd,min(x)-100,max(x)+100,subdivisions=100000)
k1=k$value
fdp=(1/k1)*(1/d2)*exp((-1/(2*d2))*((x-a)^2*(x-b)^2+(x-c)^2))
x1=seq(min(x),max(x),0.025)
n1=length(x1)
##Função Acumulada##
acumulada=matrix(0,n1,2)
md=0
for(i in 1:n1)
{
kk=integrate(fd,min(x)-100,x1[i],subdivisions=100000)
acumulada[i,2]=kk$value
acumulada[i,1]=x1[i]
#if(acumulada[i,2]>=0.5 && md==0) md=x1[i]
}
acumulada[,2]=acumulada[,2]/k1
return(a,b,c,d2,erro,k,fdp,acumulada,md)
}
Es=estimacao(dados,10^-10)
##Plotagem dos Gráficos##
x=dados
par(mfrow=c(1,2))
hist(Brix,freq=FALSE,xlab="Teor de Sólidos Solúveis (graus Brix)",ylab="Densidade",ylim=c(0,0.5),nclass=sqrt(Brix),main="Teor de Sólidos Solúveis")
lines(Brix,Es$fdp,lwd=3,col="blue")
x1=seq(min(Brix),max(Brix),0.025)
plot(Es$acumulada,xlab="Teor de Sólidos Solúveis (graus Brix)",ylab="Distribuição Acumulada",type="l",lwd=3,main="Teor de Sólidos Solúveis",col="blue")
##Cálculo dos Pontos Críticos##
polyroot(c(-
(Es$a^2*Es$b+Es$a*Es$b^2+Es$c),(Es$a+Es$b)^2+2*Es$a*Es$b+1,-
3*(Es$a+Es$b),2))

```