

# Pesquisa em andamento

Número 32

3p.

100 exemplares

dez./1999

ISSN 1517-4921

## IDENTIFICAÇÃO DE MISTURAS DE DISTRIBUIÇÕES DE BINGHAM NA SUPERFÍCIE DA HIPERESFERA $S_n$

Antonio Carlos Gomes<sup>1</sup>; Luis Hernan Rodriguez Castro<sup>2</sup>; Ives Escoufier<sup>3</sup>

O objetivo deste trabalho é identificar os componentes da mistura de distribuições de BINGHAM e os caracterizar. Neste trabalho, estudou-se a mistura de distribuições de BINGHAM, na hiperesfera  $S_n$ . A identificação dos componentes da mistura é feita pelo algoritmo de nuvens dinâmicas e a caracterização de cada uma das classes, pela estimação dos parâmetros da distribuição de BINGHAM utilizando o método de máxima verossimilhança (EMV) dentro de cada classe considerada.

As diferentes etapas deste trabalho permitem o estudo:

- Da lei de BINGHAM, da estimação dos parâmetros e sua simulação.
- Das misturas de distribuições de BINGHAM e sua identificação.
- Dos algoritmos de estimação das misturas de distribuições de BINGHAM e seus parâmetros.

Na fase final, pretende-se apresentar os resultados obtidos e os procedimentos de obtenção no uso da metodologia utilizada.

### DEFINIÇÃO

Seja  $L$  uma variável aleatória com  $n$  componentes tomando seus valores 1 na superfície da esfera  $S_n$  no espaço  $R^n$ :

$$l' l = 1$$

Seja  $L$  uma matriz diagonal não negativa  $n \times n$  e  $U$  uma matriz  $n \times n$  cujas  $n$  colunas são vetores ortonormais de  $R^n$ :

$$U' U = I$$

Diz-se que  $L$  segue uma lei de BINGHAM de parâmetros  $\Lambda$  e  $U$  se, e somente se, a função de densidade de probabilidade de  $L$  for dada por:

$$\Psi(l; \Lambda, U) = A_n^{-1} F(\Lambda)^{-1} \exp[\text{Tr}(\Lambda U' l U)]$$

$$A_n = \text{CSTE} = \frac{\pi^{\frac{1}{2}} \Gamma\left(\frac{(n-1)}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$$

<sup>1</sup> Pesquisador da Embrapa Cerrados.

<sup>2</sup> Professor-Adjunto da Universidade de Brasília.

<sup>3</sup> Professor de l'Université de Montpellier II - França.

onde  $F(\Lambda)$  é uma constante de normalização que depende de  $L$ ,

$dK = {}_1F_1\left(\frac{1}{2}; \frac{n}{2}; \Lambda\right)$  é a função confluyente hipergeométrica de argumento matricial.

Essa distribuição foi introduzida por (Bingham, 1964), para  $S_n$  e redefinida por Gomes (1987) para o caso  $S_n$ .

Para o caso particular da distribuição de BINGHAM com um parâmetro de concentração diferente de zero tem-se que,

$$f(l; u_1, K_1) = \{A_n d(K_1)\}^{-1} \exp \{K_1(l'u_1)^2\}$$

onde,

$K_1$  é o parâmetro de concentração,

$u_1$  é a média direcional,

$d(K_1)$  é a constante de normalização, a qual depende do valor de  $K_1$ .

### Resultados da simulação de misturas de duas leis de Bingham na hipersfera $S_n$ .

#### Comentário:

Nas Tabelas 1 e 2, são apresentados os resultados obtidos por simulação com base nos valores teóricos dos parâmetros da mistura. Os valores das estimações são médias de 68 simulações.

**TABELA 1. Resultados obtidos por simulação.**

*Classe = I, (n = 10, p = 5,  $\theta = 90, \phi = 0, \lambda = 10$ ),  
Classe = II, (n = 10, p = 5,  $\theta = 90, \phi = 60, \lambda = 20$ ).*

Classe	$\lambda$	$\phi$	L	M	p	Inércia
I	10,11	0,19	0,47	0,00	5,62	0,49
II	21,00	80,67	0,04	0,24	4,38	0,73

**TABELA 2. Resultados obtidos por simulação.**

*Classe = I, (n = 10, p = 5,  $\theta = 0, \phi = 0, \lambda = 10$ ),  
Classe = II, (n = 10, p = 5,  $\theta = 0, \phi = 0, \lambda = 20$ ).*

Classe	$\lambda$	$\phi$	L	M	p	Inércia
I	9,80	9,55	0,53	0,09	5,58	0,48
II	22,88	1,37	0,79	0,02	4,42	0,75

Neste trabalho, mostra-se como os algoritmos de identificação e as técnicas de estimação dos parâmetros da função de máxima verossimilhança fornecem um quadro geral para abordar o aspecto estatístico dos métodos de identificação de misturas da distribuição de BINGHAM.

Dos muitos exemplos de aplicação, é possível citar os procedimentos para estudos de variáveis que indicam a direção média angular e os parâmetros de concentração de cada tipo de solo, podendo assim auxiliar o pesquisador em tomadas de decisão quanto à correção da fertilidade do solo para fins agrícolas. Em estudos biológicos, é possível estudar o comportamento animal, indicando a direção média angular e o parâmetro de concentração de grupos de espécies dentro de uma região.

O método matemático permite resolver com eficácia, os problemas de identificação de misturas, bem como os problemas de estimação dos parâmetros das funções de densidade de probabilidade de BINGHAM. No caso geral, quando não se dispõe de outra informação que se pode extrair da amostragem, essa técnica fornece uma descrição da distribuição das observações sob a forma de identificação de grupo de variáveis.

Praticamente, a análise das funções de densidades multidimensionais permite colocar em evidência os componentes da mistura de distribuições de BINGHAM e determinar, para cada uma delas, os valores aproximados da média direcional  $\theta$ , o parâmetro de concentração  $\lambda$  e a proporção estimada  $p$  da mistura.

Este trabalho mostra que é possível pelo estudo prévio das variáveis de Bingham, melhor simplificar e interpretar os fenômenos. O reconhecimento das classes de variáveis permite estabelecer um compromisso entre os custos das variáveis a observar e a necessidade de manter boa descrição do fenômeno em estudo. Espera-se que, o pesquisador que queira utilizar no todo, ou em parte a metodologia encontre no desenvolvimento deste trabalho uma descrição das diferentes etapas.

## REFERÊNCIAS BIBLIOGRÁFICAS

- DEMPSTER A.P.; LAIRD N.M.; RUBIN D.B. Detailed account of the properties of the E-M algorithm. *Journal Royal Statistics Society, B*, London, v.39, p. 1-38, 1977.
- ERDELYI, A. *Higher transcendental functions*. New York: Mc Graw-Hill Book, 1953. 3v.
- GOMES A.C. *Reconnaissance de melanges de lois de bingham: application a la classification de variables*. Montpellier: Université des Montpellier II. 1993. 130p. Doctorat Nouveau Regime.
- GOMES P.M.P. *Distribuição de Bingham sur n-sphere: Une nouvelle approche de l'analyse factorielle*. Montpellier: Université des Montpellier II. 1987. 179p. Doctorat d'Etat.
- HUO V. *Small samples from Bingham distributions*. Minnesota: University of Minnesota. 1984. PhDThesis.
- MARDIA K.V. *Statistics of directional data*. London: Academic Press, 1972.
- RUSHTON S.; LANG E.D. *Tables of the confluent hypergeometric function*. The Indian Journal Statistics. v. 13, part 4, p. 377-411, 1954.
- TEICHER H. *Identifiability of finite mixtures*. Annals Mathematics Statistics: London, v. 34, p.1265-9, 1963.
- TEICHER H. *Identifiability of mixtures*. Annals Mathematics Statistics: London, v. 32, p. 244-8, 1961.
- WATSON G.S. *Statistics on spheres*. John Wiley e Sons, 1983. v.6.



*Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Cerrados*

*Ministério da Agricultura e do Abastecimento*  
BR 020, km 18, Rodovia Brasília/Fortaleza, Caixa Postal 08223  
CEP 73301-970, Planaltina, DF  
Telefone: (61) 388-9898 FAX: (61) 388-9879