Available online at www.sciencedirect.com
SCIENCE @ DIRECT®
ELSEVIER
C. R. Acad. Sci. Paris, Ser. I 336 (2003) 89–94
COMPTES RENDUS MATHEMATIQUE

Statistique/Probabilités

# Unicité dans la méthode des moments pour les mélanges de deux distributions normales

# Uniqueness in the method of moments for mixtures of two normal distributions

Emmanuel Monfrini

*ISFA, 43, boulevard du 11 novembre 1918, 69622 Villeurbanne, France*

Reçu le 10 juillet 2002 ; accepté après révision le 13 novembre 2002

Présenté par Paul Deheuvels

**Résumé**

A travers l'étude de l'unicité des solutions du système classique des moments du mélange de deux distributions normales, nous mettons en évidence la nécessité d'élargir la méthode de Pearson. Nous considérons, alors, un second système que nous montrons être complémentaire du premier, et que nous inversons. Une utilisation combinée de ces deux systèmes permet de rendre stable la méthode des moments jusqu'ici réputée trop instable. ***Pour citer cet article : E. Monfrini, C. R. Acad. Sci. Paris, Ser. I 336 (2003).***


**Abstract**

By studying uniqueness, we show that Pearson's method of moments for mixtures of two normal distributions must be completed. We then invert a second set of moment equations, which is constructed to complete the classical system of moments. We are thus able to stabilize the method. ***To cite this article: E. Monfrini, C. R. Acad. Sci. Paris, Ser. I 336 (2003).***


## Abridged English version

The purpose of this article is to present an original method of moments for mixtures of two unknown univariate Gaussian distributions thanks to a new approach of the moment problem put forward by Karl Pearson (cf. [10]). The method of moments mainly consists in solving a system of algebraic equations with statistical constraints (cf. [10] or [1]). For a $n$-sample of random variables identically distributed with cumulative distribution function $F$, let $\vec{M}$ and $\vec{M}_n$ denote the vectors (which can possibly be of infinite dimensions) for which the $i$-th component is

*Adresse e-mail :* monfrini@univ-lyon1.fr (E. Monfrini).

1631-073X/03/$ – see front matter 
doi:10.1016/S1631-073X(02)00011-0

respectively the $i$-th theoretical and the $i$-th empirical moment. If $\theta$ denotes the vector of all unknown parameters, there is a function noted $\mathfrak{F}$, so that $\vec{M} = \mathfrak{F}(\theta)$. We then obtain a moment estimator $\hat{\theta}$ of $\theta$ by solving the equation $\vec{M}_n = \mathfrak{F}(\hat{\theta})$. Indeed, we have to find when possible the inverse function of $\mathfrak{F}$ considering a finite number of moment equations, which will have to be determined cautiously. As a matter of fact, estimating a $k$-dimensionnal vector of parameters generally requires to resort to $k$ moment equations that must be chosen with consideration for the growing uncertainty of the sample values of the empirical moments when increasing the order of those moments. We aim here at having a more precise look at this question for a mixture of two Gaussian distributions, a case for which Pearson's method of moments has turned out to be unstable, without this phenomena being explained (cf. [5] and [6]).

In the case of mixtures of two Gaussian distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ in some proportions $p_1$ and $p_2$, let $S_1$ and $S_2$ be the two systems consisting in the first four moment equations, to which is respectively attached the fifth and the sixth moment equation. The well-known identifiability of the class of finite mixtures of univariate normal distributions (cf. [12] or [13]) makes it possible to ensure that there exists one solution only to the moment problem $\vec{M} = \mathfrak{F}(\theta)$. Considering the following developments of our work, it is worth pointing out that the notion of identifiability of a class of finite mixtures is defined up to a permutation of the symmetrical parameters of the mixture distribution (in this case $\mu_1 \leftrightarrow \mu_2$, $\sigma_1^2 \leftrightarrow \sigma_2^2$ et $p_1 \leftrightarrow p_2$, cf. [7]), which is bound to have the same consequences on the global uniqueness of the solutions of $S_1$ and $S_2$. Moreover it is worth noticing that the uniqueness of the solution of the moment problem is not sufficient to ensure the uniqueness of the solutions of $S_1$ and $S_2$, since they both consist in five moment equations only.

This is leading to the following results. We demonstrate the existence of two varieties, $V_1$ and $V_2$, sub-sets of the parameter space $\Theta = \mathbb{R}^2 \times \mathbb{R}^{+2} \times \,]0,1[$ for which the local inversion of $S_1$ cannot be achieved. $V_1$ and $V_2$ are directly responsible for the unstability noticed when using Pearson's method of moment. Bringing the existence of $V_1$ and $V_2$ to the fore naturally leads us to resort to $S_2$, which would be liable to make up for $S_1$. We then demonstrate that there is not a unique local solution to $S_2$ on a third sub-variety of $\Theta$ which will be denoted $V_3$ and for which $V_3 \cap [V_1 \cup V_2] = \emptyset$. It establishes, in particular, the local uniqueness of the solution of the system $S_1 \cup S_2$ of the first six moment equations. We then invert $S_1$ and $S_2$ according to an approach derived from the method of substitution used by Pearson which leads us to solve polynomial equations (cf. [10]). For $S_1$ it leads to Pearson's nonic, $P_{\text{Pearson}}$, and for $S_2$ it leads to $P_{12}$ which is a twelfth degree polynomial equation.

The existence of $V_1$ and $V_2$ makes it intractable to bring a solution to the moment problem for the mixture of two Gaussian distributions only with $S_1$. Let us recall that, in his founder article [10], Pearson has pointed out the inadequacies of $S_1$ for vectors of parameters belonging to $V_1$ that lead to a symmetrical mixture distribution. Nevertheless the problem related to $V_2$ is neither mentionned by Pearson, nor after him. We are here to demonstrate that in the case of $V_2$ the root of $P_{\text{Pearson}}$ enabling us to invert $S_1$ is a double root.

Thanks to our results we are able to propound a stabilized method of moments, avoiding problems of local and global ununiqueness of the solutions of the moment problem, by exploiting complementarity of $S_1$ and $S_2$ (cf. [9]). As a matter of fact, unstable cases described in [5] and [6] are now under control.

## 1. Introduction

Nous nous intéressons ici à l'estimation des paramètres d'un mélange de deux distributions gaussiennes inconnues par la méthode des moments. La méthode des moments se résume principalement à l'inversion d'un système d'équations algébriques avec contraintes (cf. [10] ou [1]). Ainsi, pour un $n$-échantillon, notons $\vec{M}$ et $\vec{M}_n$, les vecteurs (éventuellement de dimension infinie) dont la $i$-ème composante est le moment d'ordre $i$, respectivement le moment empirique d'ordre $i$, d'une distribution $F$. En notant $\theta$ le vecteur des paramètres de $F$, il existe une fonctionnelle notée $\mathfrak{F}$ telle que $\vec{M} = \mathfrak{F}(\theta)$. Nous définissons, alors, un estimateur des moments $\hat{\theta}$ de $\theta$ en résolvant l'équation $\vec{M}_n = \mathfrak{F}(\hat{\theta})$. Il s'agit donc d'inverser $\mathfrak{F}$ en considérant la restriction de cette fonctionnelle à un système constitué d'un nombre fini d'équations, dont la détermination mérite quelques précautions. En effet, l'estimation d'un nombre fini, noté $k$, de paramètres nécessite, en général, le recours à un même nombre $k$ d'équations de moment, qu'il faut choisir en tenant compte de la perte de précision de l'estimation liée à l'augmentation de l'ordre des moments. Nous cherchons, à préciser cette problématique pour un mélange de deux distributions gaussiennes, mélange ayant révélé des cas d'instabilité jusqu'ici inexpliqués (voir sur ce point les deux articles de référence [5] et [6]).

Fixons les notations. Nous considérons le cas du mélange des distributions gaussiennes $N(\mu_1,\sigma_1^2)$ et $N(\mu_2,\sigma_2^2)$ en proportions $p_1 = p$ et $p_2 = 1 - p$. Nous notons $M_1$ la moyenne et $M_i^c$ le $i$-ème moment centré du mélange. Les six premières équations de moment sont :

$$\begin{aligned}
&(E_1): && 0 = px_1 + (1-p)x_2,\\
&(E_2): && M_2^c = px_1^2 + (1-p)x_2^2 + \big(p\sigma_1^2 + (1-p)\sigma_2^2\big),\\
&(E_3): && M_3^c = px_1^3 + (1-p)x_2^3 + 3\big(p\sigma_1^2x_1 + (1-p)\sigma_2^2x_2\big),\\
&(E_4): && M_4^c = px_1^4 + (1-p)x_2^4 + 6\big(p\sigma_1^2x_1^2 + (1-p)\sigma_2^2x_2^2\big) + 3\big(p\sigma_1^4 + (1-p)\sigma_2^4\big),\\
&(E_5): && M_5^c = px_1^5 + (1-p)x_2^5 + 10\big(p\sigma_1^2x_1^3 + (1-p)\sigma_2^2x_2^3\big) + 15\big(p\sigma_1^4x_1 + (1-p)\sigma_2^4x_2\big),\\
&(E_6): && M_6^c = px_1^6 + (1-p)x_2^6 + 15\big(p\sigma_1^2x_1^4 + (1-p)\sigma_2^2x_2^4\big) + 15\big(p\sigma_1^6 + (1-p)\sigma_2^6\big)\\
& && \qquad + 45\big(p\sigma_1^4x_1^2 + (1-p)\sigma_2^4x_2^2\big),
\end{aligned}$$

où, pour des raisons de symétrie, nous avons posé : $x_1 = \mu_1 - M_1$ et $x_2 = \mu_2 - M_1$. Soient, alors, les deux systèmes $S_1$ et $S_2$, formés des quatre premières équations $(E_1)$–$(E_4)$, auxquelles on adjoint respectivement la cinquième $(E_5)$ et la sixième $(E_6)$ équation. Remarquons que le fait de travailler sur un mélange fini de distributions normales nous assure de l'identifiabilité de cette classe de mélanges finis (cf. [12] ou [13]) et soulignons pour la suite que la notion d'identifiabilité d'une classe de mélanges finis est définie à une permutation des indices près ($\mu_1 \leftrightarrow \mu_2$, $\sigma_1^2 \leftrightarrow \sigma_2^2$ et $p_1 \leftrightarrow p_2$, cf. [7]). Le même problème survient dans l'étude de l'unicité des solutions des systèmes $S_1$ et $S_2$. De plus, il convient de signaler que l'unicité de la solution du problème des moments $\vec{M} = \mathfrak{F}(\theta)$, donnée par l'identifiabilité, ne suffit pas à assurer l'unicité des solutions des systèmes $S_1$ et $S_2$ qui ne sont composés que de cinq équations de moments.

Nos résultats sont alors les suivants. Dans la Section 2 nous définissons deux variétés, notées $V_1$ et $V_2$, sur lesquelles l'inversion du système $S_1$ n'est pas possible. Nous indiquons aussi, que le système $S_2$ ne s'inverse pas sur une variété, notée $V_3$, dont l'intersection avec $V_1 \cup V_2$ est vide. Dans la Section 3 nous inversons, en suivant la méthode de Pearson [10], ces deux systèmes par une méthode de substitution. Nous les ramenons, ainsi, à la résolution d'équations polynomiales. Pour $S_1$, nous retrouvons le polynôme de Pearson défini par :

$$\begin{aligned}
P_{\mathrm{Pearson}}(x) = {} & 24x^9 + 84k_4x^7 + 36M_3^{c2}x^6 + \big(90k_4^2 + 72k_5M_3^c\big)x^5 + \big(444k_4M_3^{c2} - 18k_5^2\big)x^4\\
& + \big(288M_3^{c4} - 108M_3^ck_4k_5 + 27k_4^3\big)x^3 - \big(63M_3^{c2}k_4^2 + 72M_3^{c3}k_5\big)x^2 - 96M_3^{c4}k_4x - 24M_3^{c6},
\end{aligned}$$

où $k_4$ et $k_5$ sont les quatrième et cinquième cumulants du mélange.

Dans son article fondateur Pearson signale les insuffisances de $S_1$ sur la variété $V_1$. Ce cas correspond à une distribution symétrique où $(E_3)$ et $(E_5)$ sont toujours vérifiées. Le probléme lié à $V_2$ n'est mentionné ni par Pearson, ni après lui. Nous nous sommes attachés à donner une interprétation de la variété $V_2$ liée au polynôme de Pearson (cf. la Proposition 3.1 ci-dessous). Notons, par ailleurs, que les problèmes d'unicité étudiés pour les systèmes de moments théoriques $S_1$ et $S_2$ se compliquent lors du passage aux systèmes de moments empiriques $\vec{M}_n = \mathfrak{F}(\hat{\theta})$. En effet, il n'y a pas continuité, en fonction des moments du mélange, des solutions des systèmes théoriques au voisinage de $V_1 \cup V_2$ ou $V_3$ suivant les cas. Cependant, une étude approfondie au voisinage de $V_1$, $V_2$ et $V_3$, nous a permis de proposer une méthode stable d'estimation des paramètres d'un mélange de deux distributions normales, exploitant la complémentarité des systèmes $S_1$ et $S_2$. Pour l'énoncé et la mise en oeuvre de cette méthodologie, nous renvoyons à [8] et [9], où sont aussi développées les preuves des résultats énoncés ici.

Considérons enfin les points suivants qui permettent de situer notre travail par rapport à d'autres méthodes d'estimation utilisables dans le contexte des mélanges. Mentionnons, d'abord, la méthode $EM$ qui est une méthode générale intéressante pour maximiser la vraisemblance et qui peut être adaptée au cas du mélange de deux distributions normales où la vraisemblance est infinie (cf. [3,7] et [11]). Cette technique itérative d'estimation est souvent considérée comme meilleure que la méthode des moments, historiquement la première (cf. [10]). Notons, cependant, que la méthode des moments est souvent utilisée pour initialiser cette méthode itérative, et que cette phase est déterminante pour pouvoir éviter les maxima locaux (cf. [4]) et améliorer la vitesse de convergence (cf. [7]). De nombreuses autres méthodes ont aussi été proposées (cf., par exemple, [7]), parmi lesquelles celles du $\chi^2$ ou d'approximation du maximum de la vraisemblance proposées dans [6].

Dans le prolongement de ce travail, une étude du système sur-déterminé constitué par les six premières équations de moments, rendue possible grâce aux outils les plus récents du calcul formel, a permis d'aborder les questions d'existence et d'unicité des solutions du problème statistique sous un angle plus général et fera l'objet d'une prochaine publication.

## 2. Unicité des solutions des systèmes $S_1$ et $S_2$

Posons $D = \mu_1 - \mu_2$ et $Z = \sigma_1^2 - \sigma_2^2$, et définissons les trois sous-variétés $V_1$, $V_2$ et $V_3$, de $\mathbb{R}^2 \times \mathbb{R}^{+2} \times ]0,1[$ par :

$$V_1 = \{(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p),\ D = 0\},$$
$$V_2 = \{(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p),\ |D| = (6\sqrt{6} - 9)^{1/4}\sqrt{|Z|}\},$$
$$V_3 = \{(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p),\ (2p-1)D^2(D^8 + 18D^4Z^2 - 135Z^4) + 3Z(D^8 + 10D^4Z^2 - 15Z^4) = 0\}.$$

Remarquons que l'intersection de deux quelconques des variétés $V_1$, $V_2$ ou $V_3$ est réduite à la variété définie par $D = Z = 0$, cas où les deux distributions composantes sont égales. Le cas des mélanges correspondant à des distributions symétriques, cas équivalent ici à $M_3^c = M_5^c = 0$, coïncide avec $V_1$ lorsque $D = 0$ (cas $k_4 > 0$) et avec $V_3$ lorsque $Z = 0$ et $p = 0.5$ (cas $k_4 < 0$). Nous établissons les résultats suivants.

**Théorème 2.1.** *Il y a unicité locale des solutions de $S_1$ si et seulement si $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p)$ n'appartient pas à $V_1 \cup V_2$, et il y a unicité locale des solutions de $S_2$ si et seulement si $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p)$ n'appartient pas à $V_3$.*

**Eléments de preuve.** La preuve utilise le théorème des fonctions implicites. □

## 3. Résolution algébrique des systèmes $S_1$ et $S_2$

Nous nous restreignons, ici, au cas $D \neq 0$ (le développement du cas $D = 0$ figure dans [2]). Nous commençons par effectuer les changements de variables $\sigma = p\sigma_1^2 + (1-p)\sigma_2^2$ et $\sigma' = p\sigma_1^2 x_1 + (1-p)\sigma_2^2 x_2$. En notant $X = M_2^c - \sigma$ et $Y = M_3^c - 3\sigma'$, et en notant que $X > 0$, nous tirons, des trois premières équations, communes aux deux systèmes, les expressions :

$$\mu_1 = M_1 + \frac{Y - \operatorname{sgn}(Y)\sqrt{4X^3 + Y^2}}{2X}, \qquad \mu_2 = M_1 + \frac{Y + \operatorname{sgn}(Y)\sqrt{4X^3 + Y^2}}{2X}$$

et

$$p = \frac{1}{2} + \frac{|Y|}{2\sqrt{4X^3 + Y^2}}.$$

Notons qu'ici $p \in [0.5, 1[$. Nous obtenons, alors, pour $S_1$, respectivement pour $S_2$ lorsque $M_3^c \neq 0$,

$$\sigma' = \frac{2M_3^c X^3 + k_5 X^2 - 3k_4 X M_3^c + 2M_3^{c3}}{-2X^3 - 3k_4 X + 4M_3^{c2}},$$

et

$$\sigma' = \frac{4X^6 + 8X^4k_4 + (k_6 + 2M_3^{c2})X^3 - k_4^2X^2 - 4k_4M_3^{c2}X + 4M_3^{c4}}{M_3^c(-2X^3 - 7Xk_4 + 10M_3^{c2})},$$

dont les dénominateurs ne s'annulent pas pour un mélange de deux distributions normales.

Le résultat suivant est obtenu par substitution de $\sigma'$ dans la dernière équation de $S_1$ et $S_2$ respectivement.

**Théorème 3.1.** *Lorsque $M_3^c$ et $M_5^c$ ne sont pas simultanément nuls, la résolution de $S_1$ se ramène à la recherche des racines strictement positives du polynôme $P_\sigma(X) = -P_{\text{Pearson}}(-X)/3$, et, lorsque $M_3^c \neq 0$, la résolution de $S_2$ se ramène à la recherche des racines strictement positives du polynôme* :

$$\begin{aligned} P_{12}(X) = {} & 96X^{12} + 384k_4X^{10} + 8\bigl(6k_6 + 13M_3^{c2}\bigr)X^9 + 336k_4^2X^8 + 12k_4\bigl(8k_6 + 5M_3^{c2}\bigr)X^7 \\ & + 6\bigl(-16k_4^3 + k_6^2 + 4k_6M_3^{c2} + 22M_3^{c4}\bigr)X^6 \\ & - 6k_4^2\bigl(2k_6 + 47M_3^{c2}\bigr)X^5 + 6k_4\bigl(k_4^3 - 2\bigl(4k_6M_3^{c2} + 5M_3^{c4}\bigr)\bigr)X^4 \\ & + \bigl(97k_4^3M_3^{c2} + 48M_3^{c4}\bigl(k_6 + 7M_3^{c2}\bigr)\bigr)X^3 - 141k_4^2M_3^{c4}X^2 + 48k_4M_3^{c6}X - 4M_3^{c8}. \end{aligned}$$

**Remarque 1.** Les cas symétriques sont traités avec le sous système de $S_1$ ou $S_2$ approprié.

**Remarque 2.** Il existe des solutions de $S_2$ lorsque $M_3^c = 0$ et $M_5^c \neq 0$. Elles sont aussi solution de $S_1$.

**Remarque 3.** Il existe souvent plusieurs racines positives de $P_\sigma$ ou de $P_{12}$ qui engendrent des ensembles de paramètres qui ne sont pas tous dans $\mathbb{R}^2 \times \mathbb{R}^{+2} \times [0.5, 1[$. Le cas de plusieurs ensembles de paramètres différents et solutions de $S_1$, ou de $S_2$, se présente (cf. [8]). Nous n'avons, par contre, jamais observé plusieurs solutions au système des six premières équations de moments. Ce point sera développé dans une prochaine publication.

Enfin, l'explication du comportement de $S_1$ sur $V_2$ réside dans le résultat suivant.

**Proposition 3.1.** *Sur la variété $V_2$, l'ensemble des solutions est fini et la racine positive du polynôme $P_\sigma$ qui permet d'estimer les paramètres est une racine double.*

Pour être complet, indiquons que parmi les neuf exemples traités dans [5], les trois cas instables sont dans un voisinage de $V_2$, et sur les neuf exemples traités dans [6], deux se révèlent être dans un voisinage de $V_1$ et deux autres dans un voisinage de $V_2$. Les autres exemples traités dans ces deux articles n'étant pas défavorables à la méthode des moments nous trouvons là l'explication du biais constaté lors de l'utilisation de la méthode de Pearson. La méthode d'estimation que nous proposons dans [9] favorise l'équivalent empirique de $S_1$ pour lequel l'estimation des moments théoriques est meilleure, et propose le recours au système empirique complémentaire dans les voisinages critiques de $V_1 \cup V_2$. Elle s'est avérée stable sur tous les exemples rencontrés, y compris ceux présentés dans [5] et [6].

## Remerciements

J'adresse mes sincères remerciements à Monsieur Pierre-Loti-Viaud et à Monsieur Lazard.

## Références

[1] A.A. Borovkov, Statistique Mathématique, Mir, Moscou, 1987.

[2] A.C. Cohen, Estimation in mixtures of two normal distributions, Technometrics 9 (1) (1967) 15–28.
[3] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. Roy. Statist. Soc. Ser. B 39 (1977) 1–38.
[4] S. Fiorin, Inconsistency for roots of likelihood equations which are relative maxima of the likelihood function, Rapport technique ISUP-LSTA, 2001-6, 2001.
[5] J.G. Fryer, C.A. Robertson, The bias and accuracy of moment estimators, Biometrika 57 (1970) 57–65.
[6] J.G. Fryer, C.A. Robertson, A comparison of some methods for estimating mixed normal distributions, Biometrika 59 (1972) 639–648.
[7] G. McLachlan, D. Pell, Finite Mixture Models, Wiley, New York, 2000.
[8] E. Monfrini, Identifiabilité et méthode des moments pour les mélanges de distributions du système de Pearson, Thèse de doctorat, Université Paris 6, 2002.
[9] E. Monfrini, Une méthode des moments stable pour le mélange de deux distributions normales, Rapport technique, LSTA, 2002.
[10] K. Pearson, Contribution to the mathematical theory of evolution, Philos. Trans. Roy. Soc. London Ser. A 185 (1894) 71–110.
[11] R. Redner, H.F. Walker, Mixture densities, maximum likelihood and the E.M. algorithm, SIAM Rev. 26 (1984) 195–239.
[12] H. Teicher, Identifiability of finite mixtures, Ann. Math. Statist. 34 (1963) 1265–1269.
[13] D.M. Titterington, A.F.M. Smith, U.E. Makov, Statistical Analysis of Finite Mixture Distributions, Wiley, New York, 1985.