TEKA Kom. Mot. Energ. Roln., 2006, 6, 55–66

SHOULD WE ALWAYS USE THE MEAN VALUE?

Krzysztof S. Kubacki

Department of Applied Mathematics, Agricultural University of Lublin

Summary. In many books on statistics for practitioners the expectation μ (sample mean $\overline{x}$) is said to be a measure of the central tendency in the distribution (of the sample). But it is not always the case as the concept of expectation is analogous to the physical concept of the *center of gravity* of a distribution of mass. The aim of the paper is to influence the practitioners to consider the use of a larger variety of models., e.g. normal mixtures.

Key words: estimator, robustness, normal mixture, Eisenberger's criterion; Behboodian's criterion

INTRODUCTION

A random vector $(X_1,...,X_n)$ is called a *random sample* of size n if $X_1,...,X_n$ are independent and identically (as P_X) distributed. A *probability distribution* P_X is a mathematical model that relates the value of the variable X with the probability of occurrence of that value in the population. Such a random sample is said to be from a distribution with mean μ and variance σ^2 if each X_i has mean μ and variance σ^2. For a random sample $(X_1,...,X_n)$, the *sample mean* is $\overline{X}_n = (1/n)\sum_{j=1}^{n} X_j$, and the *sample variance* is $S_n^2 = (1/n)\sum_{j=1}^{n}(X_j - \overline{X}_n)^2$. These are one-dimensional random variables with the following properties (cf. e.g. Lehmann [1991]):

a) $E\overline{X}_n = \mu$, $Var(\overline{X}_n) = \sigma^2/n$ [Niedokos 1995],
b) $ES_n^2 = (n-1)\sigma^2/n$ [Niedokos 1995], $Var(S_n^2) = 2(n-1)\sigma^4/n^2$,

c) the sample mean $\overline{X}_n$ and sample variance S_n^2 are independent for a random sample from a normal population; this independence holds *only* for normal populations,

d) $\overline{X}_n$ (respectively S_n^2) converges with probability 1 to μ (respectively σ^2) as $n \to \infty$, which means that $\overline{X}_n$ (respectively S_n^2) is a *strongly consistent* estimator of the parameter μ (respectively σ^2).

The last property d), follows from the well known strong law of large numbers: If $\xi, \xi_1, \xi_2, \ldots$ are independent and identically distributed random variables with finite expectation $E\xi = M$ and variance $Var(\xi) = D^2$, then $\xi_n \to M$ with probability 1.

In most textbooks only the consistency (which means the stochastic convergence) of $\overline{X}_n$ (respectively S_n^2) is mentioned. The consistency follows from (a) (respectively (b)) and the famous inequality of Chebyshev: If ξ is a random variable with finite expectation $E\xi$ and variance $Var(\xi)$, then for any real $\lambda > Var(\xi)$, $P(|\xi - E\xi| > \lambda) \le Var(\xi)/\lambda^2$.

Since $Var(\overline{X}_n) \to 0$ (respectively $Var(S_n^2) \to 0$) as $n \to \infty$, the stochastic convergence follows: $\overline{X}_n \to \mu$ and $S_n^2 \to \sigma^2$ (with probability).

In spite of good properties of $\overline{X}_n$ it is not always the best possible estimator of the unknown parameter μ (cf. e.g.[in:] Lehmann [1991]).

When n is small, or we do have any additional information on X, such as e.g. information on distribution function of X, we can, use estimators of the unknown parameter θ which are better than $\overline{X}_n$, Let $X_j = I(Z_j > L)$, where $Z_j \sim N(\mu, \sigma^2)$. Then $EX_j = P(Z_j > L) \equiv p$ and $\overline{X} = (1/n)\sum_{j=1}^{n} I(Z_j > L) = \#\{j \le n : Z_j > L\}/n \equiv \tilde{p}$. Tarasińska discusses the estimates of a small fraction p under normality of Z; she compares $\tilde{p}$, $\hat{p} \equiv \Phi(\frac{\overline{X}-L}{S})$, and other unbiased estimator $\hat{\hat{p}}$ which has conditionally a linear function of $(\overline{X} - L)/S$ the beta distribution.

Suppose X denotes the average monthly rainfall in Honolulu, Hawaii (1941--1980). Due to [Brase Ch.H and Brase C.P. 1987] we have the following data:

Table 1. Average Monthly Rainfall, Honolulu, Hawaii (1941-1980)

Month	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Rainfall (in)	4.40	2.46	3.18	1.36	0.96	0.32	0.60	0.76	0.67	1.51	2.99	3.64

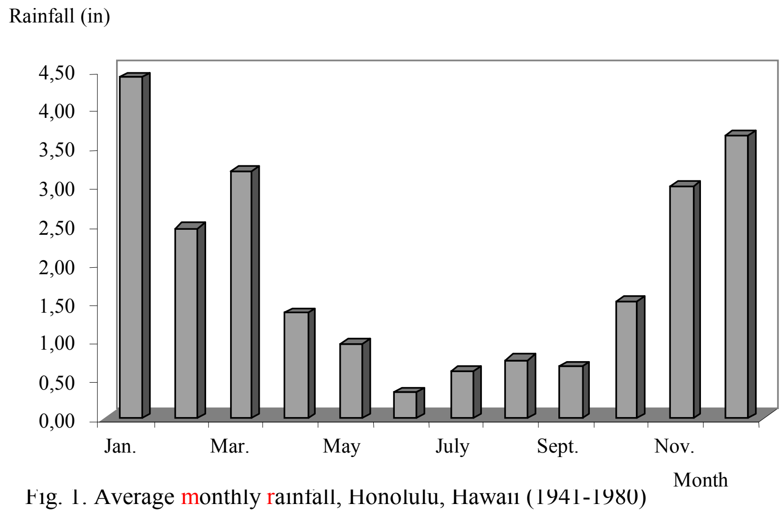

Fig. 1. Average monthly rainfall, Honolulu, Hawaii (1941-1980)

Is it informative if we say that the mean value of X equals to 1.90 (in)? What if me calculate standard deviation S and obtain 1.37 (in)? And what if we calculate the *coefficient of variation* and obtain more than 72%?

MIXTURE OF DISTRIBUTIONS

Fortunately, several subjects can be modeled by contaminated distributions (called mixtures), see e.g. Johnson *et al*. [1995] and Everitt and Hand [1981].

Imagine that a *population* is formed by combining two given populations in a given proportion. The distribution of this third population may be derived from the two given distributions and the proportion mentioned. This distribution is termed in the literature a *mixture distribution* or *compound distribution* or *contaminated distribution*.

The problems of central interest arise when data are not available for each conditional distribution separately, but only for the overall mixture distribution. Often such situations arise because it is impossible to observe some underlying variable which splits the observations into groups – only the combined distribution can be studied. In these circumstances interest often focuses on estimating *mixing proportions* and on estimating the parameters in the conditional distributions. Everitt and Hand [1981] concentrate in these areas.

The applications of finite mixture distributions are treated in Everitt and Hand [1981] usually from the mathematical statistics point of view (estimation of parameters by the method of moments or maximum likelihood estimation). We prefer a graphical method here.

Table 2. Dissection of a heterogeneous distribution.
Distribution of 430 samples of peat according to ash content. Class length: 0.5%

Ash (%)	Frequency	f	F_1	F_2
0.25	1	1	1	
0.75	1	1	1	
1.25	11	2	2	
1.75	11111	5	5	
2.25	111111111111	12	12	
2.75	111111111111111111	18	17	1
3.25	11111111111111111111	20	18	2
3.75	1111111111111111111	19	16	3
4.25	1111111111111111	16	10	6
4.75	11111111111111	14	3	11
5.25	11111111111111111111	20	1	19
5.75	1111111111111111111111111	25		25
6.25	11111111111111111111111111111111111	35		35
6.75	1111111111111111111111111111111111111111111	43		43
7.25	111111111111111111111111111111111111111111111111	48		48
7.75	111111111111111111111111111111111111111111111	45		45
8.25	11111111111111111111111111111111111	35		35
8.75	11111111111111111111111111	26		26
9.25	11111111111111111	17		17
9.75	1111111111111	13		
10.25	111111111	9		9
10.75	1111	4		4
11.25	11	2		2
	Total:	**430**	**86**	**344**

These data are the ash content of 430 peat samples given originally by Hald [1952]; Everitt and Hand [1981] who examined how the method of moments estimation procedure performs in practice. The method of moments gives:

$$\begin{aligned} \hat{p} &= 0.24; & \hat{\mu}_1 &= 3.42; & \hat{\sigma}_1 &= 1.14; \\ 1-\hat{p} &= 0.76; & \hat{\mu}_2 &= 7.41; & \hat{\sigma}_2 &= 1.46, \end{aligned} \tag{1}$$

whereas maximum likelihood estimates (taken from Hasselblad [1966]) are as follows:

$$\begin{aligned} \hat{p} &= 0.22; & \hat{\mu}_1 &= 3.21; & \hat{\sigma}_1 &= 1,00; \\ 1-\hat{p} &= 0.78; & \hat{\mu}_2 &= 7.34; & \hat{\sigma}_2 &= 1.49. \end{aligned} \tag{2}$$

Dissection of the observed values into two groups, each corresponding to a conditional normal distribution as given in Table 2, will be discussed latter on. In some experiments the analysis of a sample (histogram or frequency polygon) seems to suggest an existence of a mixture of two populations, as shown in Table 2. In some cases the data can be shown by one of the most useful graphical techniques called *stem-and-leaf display* [Brase Ch.H and Brase C.P. 1987].

MIXTURE OF TWO NORMAL DISTRIBUTIONS

A *normal mixture* occurs when a population is made up of two or more individual sub-populations (components), each of which is distributed normally, but with different parameters values.

1. Suppose Φ_i denotes the normal distribution function with mean μ_i and variance σ_i^2 $(i=1,2)$, and let $F = p\cdot\Phi_1 + (1-p)\cdot\Phi_2$. Then F is a normal mixture distribution. It is easy to see that

$$\begin{aligned}\mu^* \triangleq EX &= \int_{-\infty}^{+\infty} x\,dF(x) = \int_{-\infty}^{+\infty} x\,d(p\Phi_1(x)+(1-p)\Phi_2(x))\\ &= p\int_{-\infty}^{+\infty} x\,d\Phi_1(x) + (1-p)\int_{-\infty}^{+\infty} x\,d\Phi_2(x)\\ &= p\mu_1 + (1-p)\mu_2,\end{aligned} \tag{3}$$

and similarly, for any r>0,

$$EX^r = p\int_{-\infty}^{+\infty} x^r\,d\Phi_1(x) + (1-p)\int_{-\infty}^{+\infty} x^r\,d\Phi_2(x). \tag{4}$$

Clearly, as $Var(Y) = EY^2 - (EY)^2$, we have

$$\begin{aligned}Var(X) \triangleq \sigma^2 &= pE(\Phi_1)^2 + (1-p)E(\Phi_2)^2 - (p\mu_1 + (1-p)\mu_2)^2\\ &= p(\sigma_1^2+\mu_1^2) + (1-p)(\sigma_2^2+\mu_2^2) - (p\mu_1+(1-p)\mu_2)^2\\ &= p\sigma_1^2 + (1-p)\sigma_2^2 + p\mu_1^2(1-p)\\ &\qquad + (1-p)\mu_2^2(1-(1-p)) - 2p(1-p)\mu_1\mu_2\\ &= p\sigma_1^2 + (1-p)\sigma_2^2 + p(1-p)(\mu_1-\mu_2)^2.\end{aligned} \tag{5}$$

We can also write

$$Var(X) = p\sigma_1^2 + (1-p)\sigma_2^2 + p(\mu_1-\mu^*)^2 + (1-p)(\mu_2-\mu^*)^2. \tag{6}$$

2. Let ϕ_i denote the density function of the normal $N(\mu_i,\sigma_i^2)$ distribution, $i=1,2$. Then $f(x) = p\phi_1(x) + (1-p)\phi_2(x)$ is called a *normal mixture* density. It is easy seen from the plot of densities ϕ_1 and ϕ_2 that f is symmetrical if

a) $\mu_1 = \mu_2$ or

b) $p = 1/2$ and $\sigma_1 = \sigma_2$.

3. The third view on normal mixture is the following. Suppose Λ denotes a discrete random variable, assuming two distinct positive values: λ_1 and λ_2 with positive probabilities p and *1-p*. Let X be a random variable which is defined as follows: conditionally on $\Lambda = \lambda_j$, $X \sim N(\mu_j, \sigma_j^2)$, $j = 1, 2$. Then the unconditional distribution of X is that of F, and the unconditional density function of X is that of f, i.e. X is a normal mixture (with *mixing* random variable Λ).

Finite mixture distributions, and more specifically normal mixtures, have an important applications in genetics (see e.g. Lin and Biswas [2004] and references therein) and have been used in a wide variety of biomedical and other scientific fields [Everitt and Hand 1981, Johnson *et al.* 1995]. For example Pearson attempted to solve the parameter estimation problem using the method of moments [Everitt and Hand 1981].

Some of the more well known applications of normal mixtures include an analysis of fish lengths (Bhattacharya, Hosmar, MacDonald), botany (e.g., Fisher's iris data) and zoology (e.g., Pearson's trypanosome data), cf. [Everitt and Hand 1981] and references therein.

CLASSIFICATION OF NORMAL MIXTURES

Normal mixtures can be classified according to whether the individual components have unequal means and/or variances. The normal mixture with unequal component means and variances (location and scale mixture) of normals (LSMN) is given by

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}}\left[\frac{p}{\sigma_1}\exp\{\frac{1}{2}(\frac{x-\mu_1}{\sigma_1})^2\} + \frac{1-p}{\sigma_2}\exp\{\frac{1}{2}(\frac{x-\mu_2}{\sigma_2})^2\}\right] \\ &= \frac{p}{\sigma_1}\cdot\phi(\frac{x-\mu_1}{\sigma_1}) + \frac{1-p}{\sigma_2}\cdot\phi(\frac{x-\mu_2}{\sigma_2}), \end{aligned} \tag{7}$$

where ϕ stands for the standard normal density function (cf. [Everitt and Hand 1981, Ravishanker and Dey 2002]). Parameters which are of interest include p, the ratio σ_2/σ_1, and the distance between the component means $D_S = |\mu_2 - \mu_1|/\sigma_1$.

The normal mixture with unequal component means and common component variance (location mixture) of normals (LMN) is given by

$$f(x) = \frac{p}{\sigma}\cdot\phi(\frac{x-\mu_1}{\sigma}) + \frac{1-p}{\sigma}\cdot\phi(\frac{x-\mu_2}{\sigma}). \tag{8}$$

The scale mixture of normals (SMN) has a common mean and is given by

$$f(x) = \frac{p}{\sigma_1}\cdot\phi(\frac{x-\mu}{\sigma_1}) + \frac{1-p}{\sigma_2}\cdot\phi(\frac{x-\mu}{\sigma_2}). \tag{9}$$

The SMN was Tukey's initial model for his development of *robust* methods (see [Tukey 1960]).

One might expect a mixture to be multimodal. However, mixtures can also be unimodal and symmetric or unimodal and skewed, thereby making it difficult to distinguish between non-normal and non-mixture distributions, such as beta or gamma distributions. A sufficient condition that a normal mixture is unimodal for any value of p is that

$$(\mu_2-\mu_1)^2<\frac{27\sigma_1^2\sigma_2^2}{4(\sigma_1^2+\sigma_2^2)} \qquad \text{(or, letting } \sigma_2=a\sigma_1\text{)} \quad D_S^2<\frac{27a^2}{4(1+a^2)}; \tag{10}$$

a sufficient condition that there exist values of p for which the mixture is bimodal is that

$$(\mu_2-\mu_1)^2>\frac{8\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} \qquad \text{(or, letting } \sigma_2=a\sigma_1\text{)} \quad D_S^2>\frac{8a^2}{1+a^2} \tag{11}$$

(Eisenberger [1964]). For LMN mixtures, the sufficient condition for unimodality (10) reduces to $D_S<1,837$ while for bimodality (for some p) the sufficient condition (11) is $D_S>2$. For LSMN, the unimodality and bimodality conditions are respectively.

$$D_S<2,588\sqrt{\frac{a^2}{1+a^2}} \qquad \text{and} \qquad D_S>2,82\sqrt{\frac{a^2}{1+a^2}}.$$

Behboodian [1970] derives the following condition for a normal mixture to be unimodal:

$$|\mu_2-\mu_1|\le 2\min(\sigma_1,\sigma_2) \qquad \text{(or, equivalently)} \quad D_S\le 2\min(1,\sigma_2/\sigma_1). \tag{12}$$

His criterion is better than (10) for $a \equiv \sigma_2/\sigma_1 \in (\sqrt{11/16},\sqrt{16/11})$. In other case, Eisenberger's criterion (10) is better. In the following picture, Fig. 12 one can see the boundaries for D_S given by inequalities (10) (plot F_1), (12) (plot F_2) and (11) (plot F_3), for $0\le a\le 2$.

Note that when $p=0.5$ and $\sigma_1=\sigma_2=\sigma$, then (12) (i.e. inequality $0\le D_S\le 2$) becomes a necessary and sufficient condition for a unimodal distribution with the mode $(\mu_1+\mu_2)/2$ – it follows since the normal mixture density is symmetrical. Behboodian also shows that a sufficient condition for unimodality of LMN is given by

$$D_S\le 2\sqrt{1+|\log p-\log q|/2}. \tag{13}$$

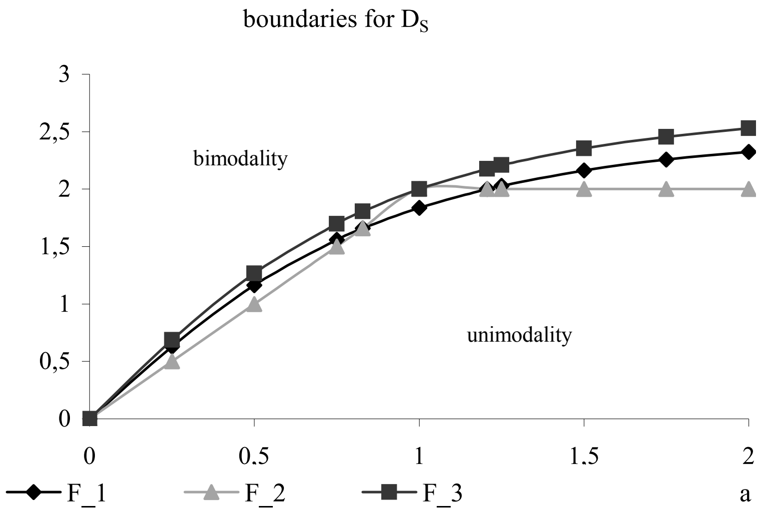

Fig. 2. Boundaries for D_S given by Eisenberger [1964] and Behboodian [1970]

This is also a necessary condition when $p = 1/2$ (from the same reason as above) [Everitt and Hand 1981].

Note that letting $\mu_1 = 0,\ \sigma_1 = 1$, and $\mu_2 = 3,\ \sigma_2 = 1$, we get $a = 1$ and $D_S = 3$. Then from the above Eisenberger's criterions it follows that the mixture is not unimodal for all values of p; and that there exist p for which this mixture is bimodal. Moreover, it follows from (13) that the mixture is unimodal for all p such that $|\log p - \log(1-p)| \geq 2,5$.

Examples of bimodal and unimodal distributions are shown in Figs. 3 to 8. The first three illustrate the dependence on p of the bimodality property when condition (11) holds; we put $a = 1$ and $D_S = 3$ as in the above consideration (note that with e.g. $p = 0,999$ the mixture is unimodal). Figs. 6–8 illustrate the unimodality of a mixture of two normal distributions independent of the value of p, since in this case Behboodian's criterion (12) holds whereas Eisenberger's criterion (10) does not. This proves that Behboodian's criterion (12) is sharper than Eisenberger's. Fig. 9 shows the density functions of some SMN mixtures compared to a single normal distribution.

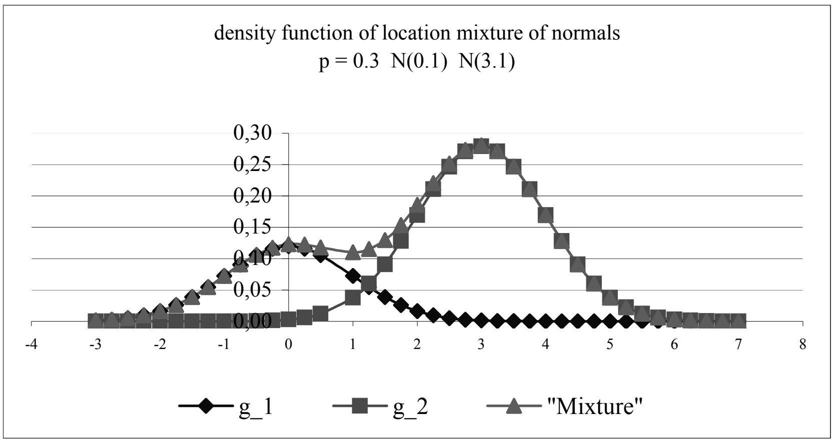

Fig. 3. Density function of LMN with p = 0.3; N(0.1); N(3.1); (D = 3 and a = 1)

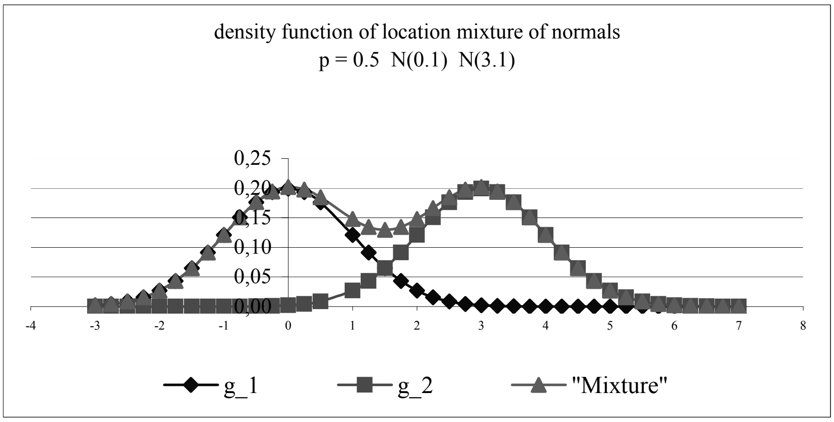

Fig. 4. Density function of LMN with p = 0.5; N(0.1); N(3.1)

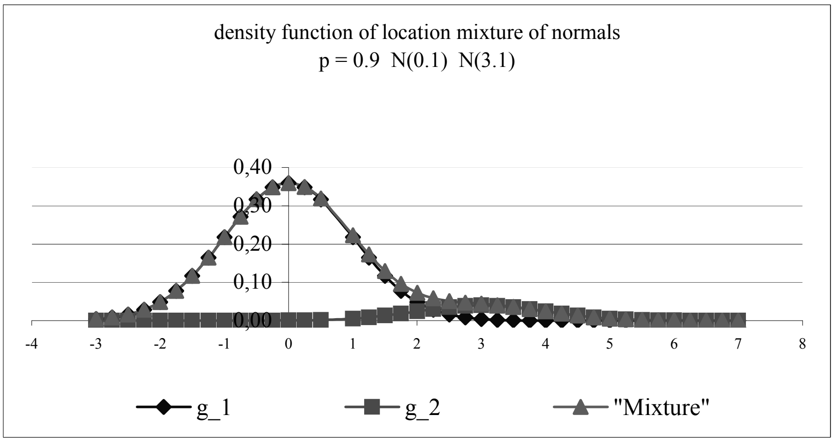

Fig. 5. Density function of LMN with p = 0.9; N(0.1); N(3.1)

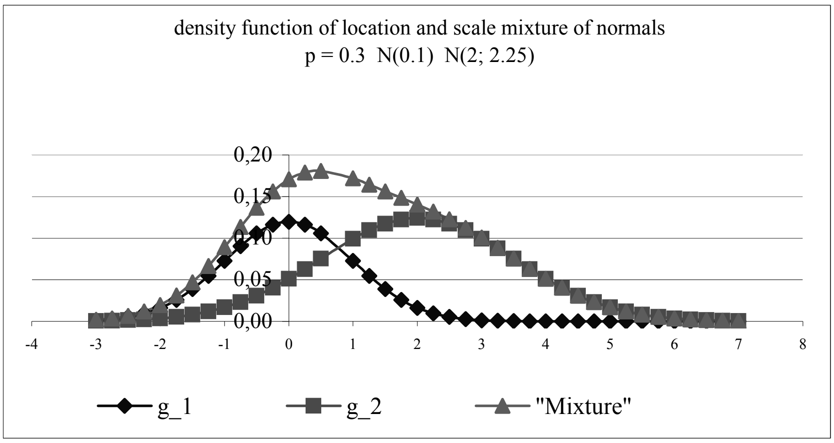

Fig. 6. Density function of LSMN with p = 0.3; N (0.1); N (2; 2.25); (12) holds whereas (10) does not

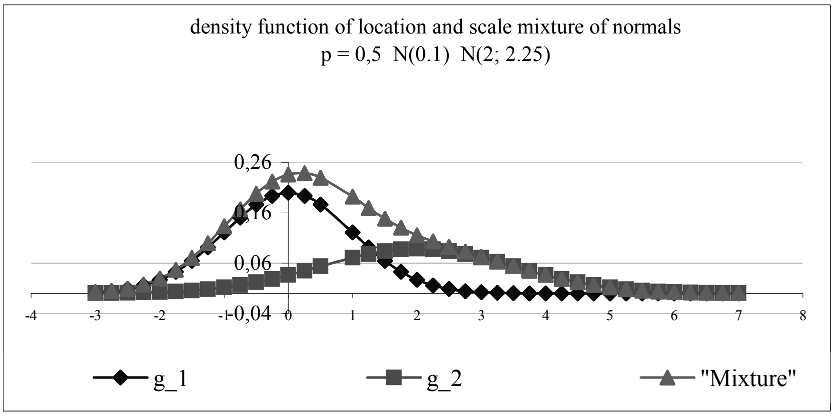

Fig. 7. Density function of LSMN with p = 0,5; N(0,1); N(2; 2,25); (12) holds whereas (10) does not

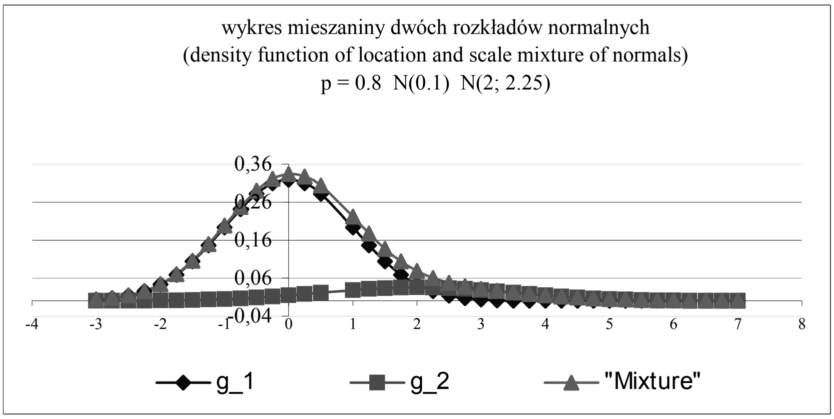

Fig. 8. Density function of LSMN with p = 0.8; N (0.1); N (2; 2.25); (12) holds whereas (10) does not

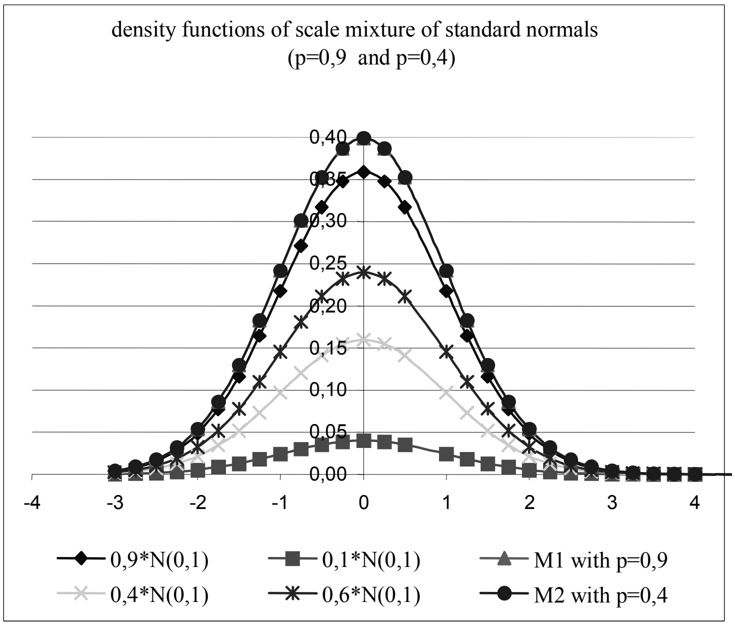

Fig. 9. Density functions of selected SMN mixtures, each with N (0.1) multiplying by p or 1-p

Dissection of a heterogeneous distribution from Table 2

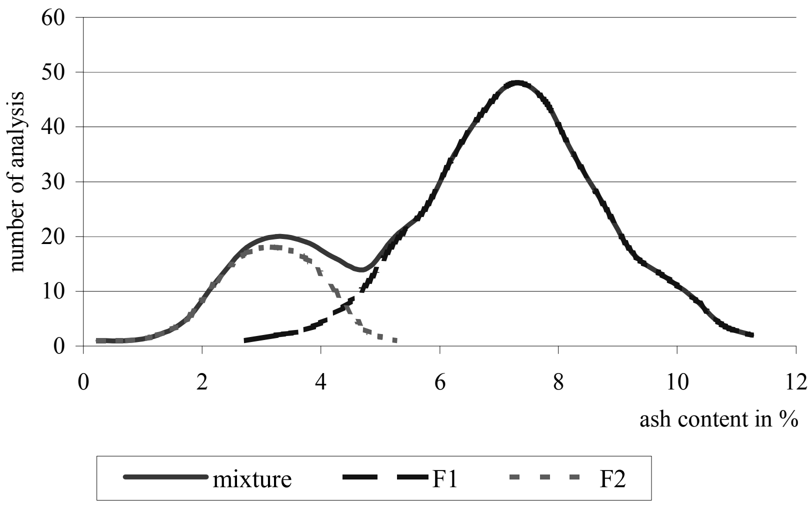

Fig. 10. Dissection of a heterogeneous distribution from Table 1

We end the paper with some numerical characteristics of the data shown in Table 2. The sample moments for these data are also calculated:

sample mean $= 6.45$; sample variance $= 4.86$; sample standard deviation $= 2.20$; moments for smaller component:

$$\text{mean}_1 = 3.10\,;\ \text{variance}_1 = 0.87\,;\ \text{standard\_dev}_1 = 0.93\,;$$

moments for bigger component:

$$\text{mean}_2 = 7.28\,;\ \text{variance}_2 = 2.35\,;\ \text{standard\_dev}_2 = 1.53\,;$$

proportion:

$p = 0.20$; $1 - p = 0.80$.

Using (3) and (5) we can calculate that:

$$EX = 0.2 \cdot 3.10 + 0.8 \cdot 7.28 = 6.45$$

and

$$Var(X) = 0.2 \cdot 0.87 + 0.8 \cdot 2.35 + 0.2 \cdot 0.8 \cdot (7.28 - 3.10)^2 = 4.85,$$

which are very close to sample moments given above.

REFERENCES

Behboodian J. 1970: On the modes of a mixture of two normal distributions. Technometrics 12, 131-139.

Brase Ch.H, Brase C.P. 1987: Understandable Statistics. 3rd ed. D.C.Heath and Comp. Lexington.

Eisenberger I. 1964: Genesis of bimodal distributions. Technometrics 6, 357-363.

Everitt B.S., Hand D.J. 1981: Finite Mixture Distributions. Chapman and Hall, London.

Hald A. 1952: Statistical theory with engineering applications. John Wiley & Sons, New York.

Hasselblad V. 1966: Estimation of parameters for a mixture of normal distributions. Technometrics 8, 431-444.

Johnson N.L., Kotz S., Balakrishnan N. 1995: Continuous Univariate Distributions. John Wiley & Sons, New York.

Lehmann E.L. 1991: Theory of point estimation. PWN, Warszawa.

Lin S., Biswas S. 2004: On modeling locus heterogeneity using mixture distributions. BMC Genetics, 5:29. The electronic version of this article is the complete one and can be found online at, http://www.biomedcentral.com/1471-2156/5/29

Niedokos E. 1995: Zastosowania Rachunku Prawdopodobieństwa i statystyki matematycznej. Wydawnictwo Akademii Rolniczej, Lublin (Wydanie bez zmian oparte na pierwszej edycji PWN).

Ravishanker N., Dey D.K. 2002: A First Course in Linear Model Theory. Chapman & Hall/CRC, Boca Raton.

Tarasińska J.: Estimation of a small fraction under normality (Submitted for publication).

Tukey J.W. 1960: A survey of sampling from contaminated distributions. [In:] Olkin I., ed., Contributions to Probability and Statistics. Stanford University Press, Stanford CA.