ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XII, 3 (1971)

W. KLONECKI (Wrocław)

ON IDENTIFIABILITY OF MIXTURES OF COMPOSED POISSON DISTRIBUTIONS *

1. Introduction. In this paper we investigate conditions under which various subfamilies of a family of mixtures (to be defined in section 2) of composed Poisson distributions are identifiable in the sense of Teicher [7]. The theorem asserts that every subfamily consisting of mixtures of composed Poisson distributions of the same degree is identifiable while the whole family is not. This theorem is an extention of a result of Feller [1], stating that the family of mixtures of Poisson distributions is identifiable. We also present a complete proof of a theorem already published [2], [3] that a subfamily restricted to mixtures generated by distributions with entire characteristic functions is identifiable.

The problem of identifiability of mixtures of composed Poisson distributions originated from studying the possibility of distinguishing between two categories of hypothetical chance mechanisms of carcinogenesis considered by Neyman and Scott [4].

2. Preliminaries and summary. Let S with or without affixes stand for the class of all distributions $S(y)$ satisfying the conditions $S(0) = 0$ and $S(0+) < 1$. Let S_E and S_A denote the subclasses of S of distributions with entire and analytic characteristic functions, respectively. Finally, the symbol C will be used to denote the class of all sequences of non-negative numbers $c = \{c_k\}$, with $0 < \sum_{k=1}^{\infty} c_k < \infty$. Moreover, let C_N stand for the subclass of C composed of sequences $c = \{c_k\}$ such that $c_N > 0$ and $c_k = 0$ for $k > N$ and C_F for the subclass of C composed of finite sequences.

The family of mixtures considered is defined as follows. Let $c = \{c_k\} \in C$. Then

$$P(u\,|\,c) = \exp\Big\{\sum_{k=1}^{\infty} c_k(u^k - 1)\Big\}, \qquad |u| \leqslant 1,$$

* This investigation was supported in part by research grant GM 10525-07 from the National Institute of Health Service, U.S.A.

represents a generating function of a probability distribution, say $\{p(k\,|\,\boldsymbol{c})\}$, called *composed Poisson distribution* [5]. If $\boldsymbol{c}\in \boldsymbol{C}_N$, then the composed Poisson distribution corresponding to $\boldsymbol{c}$ is said to be *of degree* N. Let

$$\boldsymbol{F}_{\boldsymbol{c}} = \{\{p(k\,|\,y\boldsymbol{c})\} : y\in[0,\infty)\}, \qquad \text{where } y\boldsymbol{c} = \{yc_k\}.$$

The distribution

$$p(k\,|\,S,\boldsymbol{c}) = \int_0^\infty p(k\,|\,y\boldsymbol{c})\,dS(y), \qquad k = 0, 1, \ldots, \tag{1}$$

where $S\in\boldsymbol{S}$, is called a *mixture* (or an *S-mixture* of $\boldsymbol{F}_{\boldsymbol{c}}$), and S — the *mixing distribution*. Now for $\boldsymbol{S}_0 \subset \boldsymbol{S}$ and $\boldsymbol{C}_0 \subset \boldsymbol{C}$ we define

$$\boldsymbol{K}(\boldsymbol{S}_0, \boldsymbol{C}_0) = \{\{p(k\,|\,S,\boldsymbol{c})\} : S\in\boldsymbol{S}_0, \boldsymbol{c}\in\boldsymbol{C}_0\}.$$

Clearly, $\boldsymbol{K}(\boldsymbol{S}_0, \boldsymbol{C}_0)$ is the class of, say, $\boldsymbol{S}_0$-mixtures of all composed Poisson distributions corresponding to all $\boldsymbol{c}\in\boldsymbol{C}_0$. In particular, $\boldsymbol{K}(\boldsymbol{S}, \boldsymbol{C}_1)$ is the family of mixtures of Poisson distributions.

Following Teicher [7], the family $\boldsymbol{K}(\boldsymbol{S}_0, \boldsymbol{C}_0)$ is said to be *identifiable* (with respect to $\boldsymbol{S}_0$ and $\boldsymbol{C}_0$) if mapping (1) of $\boldsymbol{S}_0\times\boldsymbol{C}_0 \to \boldsymbol{K}(\boldsymbol{S}_0, \boldsymbol{C}_0)$ is one-to-one.

The subject of study is to determine subsets $\boldsymbol{S}_0 \subset \boldsymbol{S}$ and $\boldsymbol{C}_0 \subset \boldsymbol{C}$ so that the resulting family of mixtures $\boldsymbol{K}(\boldsymbol{S}_0, \boldsymbol{C}_0)$ is identifiable in the above-mentioned sense.

To exclude a trivial case of non-identifiability we assume throughout the paper that $\boldsymbol{C}$ consists of sequences $\boldsymbol{c} = \{c_k\}$ such that $\sum_{k=1}^{\infty} c_k = 1$.

An intuitive discription of the considered problem of identifiability may be as follows. Consider chance occurrences, to be called "arrivals", occurring in accordance with a Poisson process with expectation $y \geqslant 0$. Suppose that y is subject to a distribution $S(y)\in\boldsymbol{S}_0 \subset \boldsymbol{S}$. Next suppose that at each arrival come into existence $1, 2, \ldots$ particles with probabilities $c_1, c_2, \ldots$, respectively, where $\{c_i\}\in\boldsymbol{C}_0 \subset \boldsymbol{C}$. Then the number of particles $X(T)$ produced within time interval $[0, T)$ has a composed Poisson distribution. The question asked is under what assumptions on $\boldsymbol{S}_0$ and $\boldsymbol{C}_0$ the distribution of $X(T)$ determines uniquely $S(y)$ and $\{c_k\}$, given an arbitrary $T > 0$.

In this paper the following two theorems are proved:

THEOREM 1. *For* $N = 1, 2, \ldots$ *the families* $\boldsymbol{K}(\boldsymbol{S}, \boldsymbol{C}_N)$, *consisting of mixtures of composed Poisson distributions of degree* N, *are identifiable.*

THEOREM 2. *The family* $\boldsymbol{K}(\boldsymbol{S}_E, \boldsymbol{C}_F)$, *consisting of* $\boldsymbol{S}_E$ *mixtures of composed Poisson distributions of finite degree, is identifiable.*

For $N = 1$, Theorem 1 states that the family of mixtures of Poisson distributions is identifiable. This result is due to Feller [1]. The assertion of Theorem 2 fails if the restriction to S_E-mixtures is omitted. This is shown in Example 2. The family $K(S_E, C)$ consisting of S_E-mixtures of composed Poisson distributions of finite and infinite degree is non-identifiable [2].

3. The lemmas. The chief instrument in the proof of Theorems 1 and 2 is the following lemma:

LEMMA 1. *Let* $a_{jk}+i\beta_{jk}$, *where* $j, k = 1, 2$, *be complex numbers such that*

$$\beta_{11} = \beta_{22} = 0 \tag{2}$$

and

$$a_{21} \leqslant a_{22}. \tag{3}$$

Then

$$\int_0^\infty e^{y(a_{1j}+i\beta_{1j})}dS_1(y) = \int_0^\infty e^{y(a_{2j}+i\beta_{2j})}dS_2(y), \quad j = 1, 2, \tag{4}$$

implies

$$a_{11} \leqslant a_{12} \tag{5}$$

provided that the integrals appearing in (4) *exist.*

Proof. From (2) and (4) it follows that

$$\int_0^\infty e^{ya_{11}}dS_1(y) = \int_0^\infty e^{y[a_{21}+i\beta_{21}]}dS_2(y).$$

Hence

$$\int_0^\infty e^{ya_{11}}dS_1(y) \leqslant \int_0^\infty e^{ya_{21}}dS_2(y). \tag{6}$$

Similarly,

$$\int_0^\infty e^{ya_{22}}dS_2(y) \leqslant \int_0^\infty e^{ya_{12}}dS_1(y). \tag{7}$$

Combining (3), (6) and (7), we get

$$\int_0^\infty e^{ya_{11}}dS_1(y) \leqslant \int_0^\infty e^{ya_{12}}dS_1(y),$$

which implies (5).

We shall also need Lemmas 2-6.

LEMMA 2. *If $N_1 < N_2$ and $-\pi < x_2 \leqslant \pi$, then it is possible to find a value y such that*

$$(8) \qquad \sin(x_2 + N_2 y) = 0$$

and

$$(9) \qquad \varepsilon\cos(x_2 + N_2 y) > 0, \quad \cos N_1 y < 0,$$

where $\varepsilon = 1$ if $-\pi < x_2 < \pi$ and $\varepsilon = -1$ if $x_2 = \pi$.

Proof. We give the following rule for the selection of y: take $y = (2s\pi - x_2)/N_2$. For s select an integer so that

$$(10) \qquad \left(2v + \frac{1}{2}\right)\pi < \frac{N_1}{N_2}(2s\pi - x_2) < \left(2v + \frac{3}{2}\right)\pi,$$

while v is an integer.

In case $-\pi < x_2 < \pi$, the rules for the selection of s are as follows:

(I) $2N_1 < N_2$. Put $v = 0$. Then there exists an integer s satisfying (10).

(II) $2N_1 = N_2$. Put $v = 1$. Then $s = 3$ satisfies (10).

(III) $2N_1 > N_2$. Let

$$a_w = \frac{4(w+3)\pi - 2x_2}{(4w+3)\pi}, \quad w = 0, 1, \ldots$$

If $a_{w+1} < N_2/N_1 \leqslant a_w$, then $v = w+1$ and $s = w+4$ satisfy (10).

In case $x_2 = \pi$, the corresponding conditions (8) and (9) are equivalent to $\sin N_2 y = 0$, $\cos N_2 y > 0$ and $\cos N_1 y < 0$, which are conditions (8) and (9) with 0 substituted for x_2. Consequently, in case $x_2 = \pi$, the assertion follows from the above.

LEMMA 3. *Let $N \geqslant 2$ and let $1 \leqslant k \leqslant N-1$. Further, let $\varepsilon_j = 1$ or -1, while $j = 1, 2$. Then there exists a value y such that*

$$(11) \qquad \sin Ny = 0,$$

$$(12) \qquad \cos Ny = 1,$$

$$(13) \qquad \varepsilon_1 \cos ky > 0,$$

$$(14) \qquad \varepsilon_2 \sin ky \geqslant 0.$$

Proof. Put $y = 2s\pi/N$, where s is an integer. Then (11) and (12) hold. To show that (13) and (14) can be simultaneously satisfied, it is sufficient to show that there exist integers s and v such that:

(i) if $\varepsilon_1 = \varepsilon_2 = 1$, then $0 \leqslant sk/N - v < \frac{1}{4}$:

(ii) if $\varepsilon_1 = -\varepsilon_2 = -1$, then $\frac{1}{4} < sk/N - v \leqslant \frac{1}{2}$:

(iii) if $\varepsilon_1 = \varepsilon_2 = -1$, then $\frac{1}{2} \leqslant sk/N - v < \frac{3}{4}$;

(iv) if $\varepsilon_1 = -\varepsilon_2 = 1$, then $\frac{3}{4} < sk/N - v \leqslant 1$.

In cases (i) and (iv), the existence of the integers s and v is obvious. In the two other cases and $N \geqslant 4$, the assertion follows from a well known theorem on number theory [6]. For $N = 2$ and $N = 3$ the existence of integers s and v is easily checked.

LEMMA 4. *Let $N > 2$ and let $1 \leqslant k \leqslant N-1$. Further, let $\varepsilon = 1$ or -1. Then there exists a value y such that*

$$\sin Ny = 0, \tag{15}$$

$$\cos Ny = 1, \tag{16}$$

$$\varepsilon \sin ky > 0. \tag{17}$$

Proof. If we select $y = 2s\pi/N$, where s is an integer, then conditions (15) and (15) are satisfied. Moreover, (17) also holds provided we can select s so that if $\varepsilon = 1$, then $0 < sk/N - v < \frac{1}{2}$, and if $\varepsilon = -1$, then $\frac{1}{2} < sk/N - v < 1$, where v is an arbitrary integer. As in Lemma 3, the existence of integers s and v, satisfying the above inequalities, follows from the well-known theorem on number theory mentioned in the proof of Lemma 3.

LEMMA 5. *Let $N = 2m+1$ and let $1 \leqslant k \leqslant N-1$. Further, let $\varepsilon = 1$ or -1. Then one can select a value y so that*

$$\sin Ny = 0, \tag{18}$$

$$\cos Ny = -1, \tag{19}$$

$$\varepsilon \cos ky > 0. \tag{20}$$

Proof. In case $\varepsilon = -1$, the rule for the selection of y is as follows. For $1 \leqslant k \leqslant m$ put $y = s\pi/N$, where s is an odd integer such that $N/2k < s < 3N/2k$.

Because $3N/2k - N/2k > 2$ for $1 \leqslant k \leqslant m$, there must be at least one odd integer s lying between these limits. For $m+1 \leqslant k \leqslant 2m$, take $y = \pi/N$. For $\varepsilon = 1$, the existence of a value y satisfying conditions (18)-(19) follows from the above and from the relation

$$\cos\frac{sk\pi}{N} = -\cos\frac{s(N-k)\pi}{N},$$

which holds whenever s is odd.

The next lemma, which is of some independent interest, will be used to prove Theorem 2.

LEMMA 6. *For $j = 1, 2$ let $P_j(z)$ stand for a polynomial of degree at least 1 with arbitrary, real or complex coefficients, subject to the restriction $P_j(1) = 0$. Let $S_j \in \mathbf{S}_E$, $j = 1, 2$.*

Then

$$\int_0^\infty e^{yP_1(z)}dS_1(y) \equiv \int_0^\infty e^{yP_2(z)}dS_2(y) \tag{21}$$

if and only if there exists a positive number a such that

$$P_1(z) \equiv aP_2(z) \tag{22}$$

and

$$S_1(y/a) \equiv S_2(y). \tag{23}$$

Proof. Identities (22) and (23) imply (21) immediately. To prove the converse proposition we first show that (21) and (22) imply (23). Indeed, (21) reduces then to

$$\int_0^\infty e^{yP_2(z)}dS_1(y/a) \equiv \int_0^\infty e^{yP_2(z)}dS_2(y).$$

Consequently, for all complex z,

$$\int_0^\infty e^{yz}dS_1(y/a) = \int_0^\infty e^{yz}dS_2(y).$$

By the uniqueness theorem on characteristic functions we conclude that (23) holds.

To show that (21) implies (22) we suppose that there exist polynomials $P_1(z)$ and $P_2(z)$ both of degree at least 1 and both vanishing at $z = 1$, and that there exist distributions S_1 and S_2 belonging to $\mathbf{S}_E$ such that (21) holds but not (22).

For $j = 1, 2$ let

$$P_j(z) = \sum_{k=0}^{N_j}(a_{jk}+ib_{jk})z^k.$$

Substituting $a_{jk}+ib_{jk} = r_{jk}(\cos x_{jk}+i\sin x_{jk})$ and $z = z(r, y) = r(\cos y+ +i\sin y)$, we obtain

$$P_j(z) = \sum_{k=0}^{N_j} r^k r_{jk}\cos(x_{jk}+ky)+i\sum_{k=0}^{N_j} r^k r_{jk}\sin(x_{jk}+ky)$$

or, for short,

$$P_j(z) = A_j(z)+iB_j(z), \quad j = 1, 2,$$

where $A_j(z)$ and $B_j(z)$ represent the real and the imaginary part of $P_j(z)$, respectively. For further use let us note that for $j = 1, 2$

$$A_j(z) = r_{jN_j}r^{N_j}\cos(x_{jN_j}+N_jy)+o(r^{N_j}).$$

For the sake of brevity we shall write $x_j = x_{jN_j}$, $j = 1, 2$. Without loss of generality we may assume that $N_1 \leqslant N_2$, $x_1 = 0$, $-\pi < x_2 \leqslant \pi$ and $r_{jN_j} = 1$, $j = 1, 2$.

First we show that (21) implies $N_1 = N_2$. This will be proved by showing that in case $N_1 < N_2$ there would exist two numbers z_1 and z_2 such that $\beta_{jk} = B_j(z_k)$ and $a_{jk} = A_j(z_k)$, where $j, k = 1, 2$, would satisfy all assumptions of Lemma 1 and the inequality $a_{12} < a_{11}$ which is opposite to the assertion of Lemma 1.

We shall show this by using the fact that the asymptotes of $B_1(z) = 0$ and $B_2(z) = 0$ are parallel to the corresponding asymptotes of $\operatorname{Im}(z^{N_1}) = 0$ and $\operatorname{Im}[(\cos x_2 + i \sin x_2) z^{N_2}] = 0$, respectively. Thus, in particular, one of the asymptotes of $B_1(z) = 0$ is parallel to the asymptote $y = 0$ and one of the asymptotes of $B_2(z) = 0$ is parallel to the asymptote $y = y_0$, where y_0 is a value that satisfies conditions (8) and (9) of Lemma 2. Denoting, for $j = 1, 2$, by $z_j = z_j\big(r(y_j), y_j\big)$ points satisfying equation $B_j(z) = 0$, we conclude from the above that by selecting y_1 and y_2 sufficiently close to 0 and y_0, respectively, the radius $r(y_j)$ may be as large as we please.

Since $x_1 = 0$ and $r_1 = r_2 = r$, we have

$$\begin{aligned} A_1(z) &= r^{N_1} \cos N_1 y + o(r^{N_1}), \\ A_2(z) &= r^{N_2} \cos (x_2 + N_2 y) + o(r^{N_2}). \end{aligned} \tag{24}$$

In view of (24) it follows from the above that for y_1 sufficiently close to 0 there exist solutions $z_1 = z_1(r(y_1), y_1)$ of $B_1(z) = 0$ such that $0 < A_1(z_1)$.

Moreover, it follows from Lemma 2 that in case $-\pi < x_2 < \pi$, there exist solutions $z_2 = z_2\big(r(y_2), y_2\big)$ of $B_2(z) = 0$ such that $A_1(z_2) < 0$ and $A_2(z_1) < A_2(z_2)$, whatever z_1 might be.

On the other hand, it follows from Lemma 2 that in case $x_2 = \pi$, there exist solutions z_1 of $B_1(z) = 0$ and z_2 of $B_2(z) = 0$ such that $A_1(z_1)$ may be arbitrarily large and $A_2(z_1)$, $A_2(z_2)$ and $A_1(z_2)$ may be arbitrarily small.

This implies that one can select two numbers z_1 and z_2 so that $B_j(z_j) = 0$, $j = 1, 2$, and $A_2(z_1) \leqslant A_2(z_2)$, while $A_1(z_2) < A_1(z_1)$, which gives the desired contradiction with the assertion of Lemma 1. Thus we conclude that the polynomials $P_1(z)$ and $P_2(z)$ must be of the same degree, say N. Next we show that relation (21) implies $x_2 = 0$.

Let, respectively, $z_1 = (r_1(y_1), y_1)$ and $z_2 = (r_2(y_2), y_2)$ be solutions of $B_1(z) = 0$ and $B_2(z) = 0$ such that $r = r_1(y_1) = r_2(y_2)$. Here r may be as large as we please by selecting y_1 and y_2 sufficiently close to, say, 0 and $-x_2/N$, respectively.

Clearly,

$$A_j(z_2) - A_j(z_1) = r^N [\cos(x_j + N y_2) - \cos(x_j + N y_1)] + o(r^N).$$

Introducing the notation $y_j = (-x_j + \varepsilon_j)/N$, $j = 1, 2$, and using the fact that $x_1 = 0$, we obtain

$$A_1(z_1) - A_1(z_2) = r^N[\cos\varepsilon_1 - \cos(x_2 - \varepsilon_2)] + o(r^N)$$

and

$$A_2(z_2) - A_2(z_1) = r^N[\cos\varepsilon_2 - \cos(x_2 + \varepsilon_1)] + o(r^N).$$

If $x_2 \neq 0$, then for sufficiently small values ε_1 and ε_2 the expressions in the parentheses are positive and, at the same time, r may be arbitrarily large. This implies that there exist z_1 and z_2 such that $B_j(z_j) = 0$, $j = 1, 2$, while $A_1(z_2) < A_1(z_1)$ and $A_2(z_1) \leqslant A_2(z_2)$.

Putting $\alpha_{jk} = A_j(z_k)$ and $\beta_{jk} = B_j(z_k)$, where $j, k = 1, 2$, we note a contradiction with Lemma 1 again. Thus $P_1(z)$ and $P_2(z)$ are polynomials with coefficients of the highest power of z equal to 1.

Finally, to prove that $P_1(z) \equiv P_2(z)$, suppose, to the contrary, that the first $N-k$ coefficients of $P_1(z)$ and $P_2(z)$ are identical, where $k = 1, \ldots$ $\ldots, N-1$, but $a_{1k} + ib_{1k} \neq a_{2k} + ib_{2k}$.

To show that $a_{1k} = a_{2k}$, we substitute $r(\cos y + i\sin y)$ instead of z and observe that

$$(25) \qquad A_1(z) - A_2(z) = r^k[(a_{1k} - a_{2k})\cos ky - (b_{1k} - b_{2k})\sin ky] + o(r^k).$$

To arrive at a contradiction with Lemma 1 we need to show that there exist numbers y_0 and y_{00} depending upon the a_{jk}'s and the b_{jk}'s such that

$$(26) \qquad \sin Ny_0 = \sin Ny_{00} = 0,$$

$$(27) \qquad (a_{1k} - a_{2k})\cos ky_0 - (b_{1k} - b_{2k})\sin ky_0 > 0$$

and

$$(28) \qquad (a_{1k} - a_{2k})\cos ky_{00} - (b_{1k} - b_{2k})\sin ky_{00} < 0.$$

Suppose first that $a_{1k} \neq a_{2k}$. The existence of y_0 and y_{00} follows from Lemma 3. The ε's, appearing in Lemma 3, are selected depending upon the a's and the b's as shown in Table 1.

TABLE 1

	$a_{1k} > a_{2k}$				$a_{1k} < a_{2k}$			
	$b_{1k} \geqslant b_{2k}$		$b_{1k} < b_{2k}$		$b_{1k} \geqslant b_{2k}$		$b_{1k} < b_{2k}$	
	y_0	y_{00}	y_0	y_{00}	y_0	y_{00}	y_0	y_{00}
ε_1	1	-1	1	-1	-1	1	-1	1
ε_2	-1	1	1	-1	-1	1	1	-1

This in view of (25) implies that under the assumption $a_{1k} \neq a_{2k}$ there exist solutions of $B_1(z) = 0$ such that $0 < A_2(z) < A_1(z)$ and solutions of $B_2(z) = 0$ such that $0 < A_1(z) < A_2(z)$.

Consequently, there exist numbers z_1 and z_2 such that $\alpha_{jk} = \mathrm{Re}[P_j(z_k)]$ and $\beta_{jk} = \mathrm{Im}[P_j(z_k)]$, $j = 1, 2$, satisfy conditions (2) and (3) of Lemma 1 and, moreover, the inequality $\alpha_{11} > \alpha_{12}$, which is opposite to assertion (6) of Lemma 1. This contradiction leads to the conclusion that $a_{1k} = a_{2k}$.

Now suppose that $b_{1k} \neq b_{2k}$, while $a_{1k} = a_{2k}$. In this case and for $N > 2$, the existence of numbers y_0 and y_{00} satisfying inequalities (26) to (28) follows from Lemma 4. Here ε must be selected as shown in Table 2.

TABLE 2

	$b_{1k} > b_{2k}$		$b_{1k} < b_{2k}$	
	y_0	y_{00}	y_0	y_{00}
ε	-1	1	1	-1

Now, similarly as above, we can show that in case $N > 2$ the assumption $b_{1k} \neq b_{2k}$ leads to a contradiction with assertion (5) of Lemma 1. For $N = 2$, we easily conclude by differentiating both sides of (21) with respect to z and by using the assumption $P_1(1) = P_2(1) = 0$ that $b_{11} = b_{21}$. Thus, for $k = 1, \ldots, N$, we have

$$(29) \qquad a_{1k} + ib_{1k} = a_{2k} + ib_{2k}.$$

Since by assumption $P_1(1) = P_2(1)$, relation (29) holds also for $k = 0$. This completes the proof of Lemma 6.

If the assumption that $S_j \in \boldsymbol{S}_E$, $j = 1, 2$, is omitted, then Lemma 6 is not valid as is shown in the following example:

Example 1. Let S_1 and S_2 be distributions corresponding to the characteristic functions $f_1(z) = (1-z)^{-3}$ and $f_2(z) = (1-z)^{-5}$, respectively. Clearly, $S_j \in \boldsymbol{S}_A - \boldsymbol{S}_E$, where $j = 1, 2$.

Putting $P_1(z) = (z-2)^5 + 1$ and $P_2(z) = (z-2)^3 + 1$, we obtain $f_1[P_1(z)] \equiv f_2[P_2(z)]$, which shows that (21) holds without (22) and (23) being true.

4. Proofs of the theorems.

(i) Theorem 1. Suppose that for some $N > 1$ the family $\boldsymbol{K}(\boldsymbol{S}, \boldsymbol{C}_N)$ is non-identifiable, i.e. there exist two sequences $\boldsymbol{c}_1 = \{c_{1k}\}$ and $\boldsymbol{c}_2 = \{c_{2k}\}$ belonging to $\boldsymbol{C}_N$ and two distributions S_1 and S_2 belonging to $\boldsymbol{S}$, where $\boldsymbol{c}_1 \neq \boldsymbol{c}_2$ and (or) $S_1 \neq S_2$ such that

$$(30) \qquad \int_0^\infty e^{yP(u|\boldsymbol{c}_1)} dS_1(y) = \int_0^\infty e^{yP(u|\boldsymbol{c}_2)} dS_2(y), \quad |u| < 1.$$

Let

$$(31) \qquad \eta_j(z) = \int_0^\infty e^{yP(z|c_j)}dS_j(y), \quad j = 1, 2,$$

and let D_j be the set where the integral given by (31) exists. Without loss of generality we may assume that $c_{1N} = c_{2N} = 1$. Assumption (30) implies that $\eta_1(z)$ and $\eta_2(z)$ coincide on the interval $I = \{z : z = u, |u| < 1\}$. Because the interval I is contained in the interior of $D = D_1 \cap D_2$, it follows from a well-known theorem on analytic functions that for $z \in D$

$$(32) \qquad \eta_1(z) \equiv \eta_2(z).$$

In case N is odd, the proof of Theorem 1 is similar to the proof of Lemma 6 except that an extra care must be taken in selecting z_1 and z_2 so that both these numbers belong to D. Putting $P_j(z) = P(z|c_j), j = 1, 2$, and using the notation introduced in section 2, we can write

$$(33) \qquad A_1(z) - A_2(z) = r^k(c_{1k} - c_{2k})\cos ky + o(r^k),$$

where $c_{1k} \neq c_{2k}$, while $1 \leqslant k \leqslant N-1$.

In view of (33) and of Lemma 5, equation $B_j(z) = 0$, $j = 1, 2$, has solutions $z_j = z_j(r(y_j), y_j)$ with $r_j = r_j(y_j)$ arbitrarily large and such that, for $j = 1$, $A_2(z_1) < A_1(z_1) < 0$, and, for $j = 2$, $A_1(z_2) < A_2(z_2) < 0$.

Consequently, one can select z_1 and z_2 so that

$$(34) \qquad B_1(z_1) = B_2(z_2) = 0$$

and

$$(35) \qquad A_1(z_2) < A_2(z_2) = A_2(z_1) < A_1(z_1).$$

Putting again $a_{jk} = A_j(z_k)$ and $\beta_{jk} = B_j(z_k)$ for $j = 1, 2$, we see that (34) and (35) contradict Lemma 1. Because of this contradiction and because of the assumption $P_1(1) = P_2(1)$, we conclude that for $0 \leqslant k \leqslant N-1$ we have $c_{1k} = c_{2k}$ or, equivalently, that $P_1(z) \equiv P_2(z)$, contrary to the assumption.

Now we proceed to consider the case where N is even. In view of (30), relation $P_1(u_1) = P_1(u_2)$ must imply $P_2(u_1) = P_2(u_2)$ and vice versa, provided that u_1 and u_2 are real numbers belonging to the set D, where both integrals exist. This implies that $P_1(u) \equiv P_2(u)$.

(ii) Theorem 2. If $K(S_E, C_F)$ were non-identifiable, then there would exist two sequences $c_1 = \{c_{1k}\}$ and $c_2 = \{c_{2k}\}$ belonging to C_F and two distributions S_1 and S_2 belonging to S_E, where $c_1 \neq c_2$ and (or) $S_1 \neq S_2$ such that $\{p(k|S_1, c_1)\} = \{p(k|S_2, c_2)\}$. Consequently, for $|u| \leqslant 1$,

$$(36) \qquad \int_0^\infty e^{yP(u|c_1)}dS_1(y) = \int_0^\infty e^{yP(u|c_2)}dS_2(y).$$

Since S_1 and S_2 are entire, relation (21) holds with $P_j(z) = P(u|c_j)$, where $j = 1, 2$, in the whole complex plane. Since $P_1(1) = P_2(1) = 0$. it follows from Lemma 6 that

$$(37) \qquad \sum_k c_{1k}(u^k-1) \equiv a\sum_k c_{2k}(u^k-1) \quad \text{and} \quad S_1(y|a) \equiv S_2(y),$$

where $a > 0$. But $\sum_k c_{1k} = \sum_k c_{2k} = 1$, so that (37) implies $a = 1$. Hence $c_1 = c_2$ and $S_1 = S_2$, contrary to the assumption. This contradiction completes the proof of Theorem 2.

The following example shows that the family $\boldsymbol{K}(\boldsymbol{S}, \boldsymbol{C}_F)$ is non-identifiable, indicating that the assertion of Theorem 2 is not valid without the assumption that the mixing distributions are entire.

Example 2. Let S_1 and S_2 be probability distributions with densities

$$p_1(x) = \frac{a(a^2+b^2)}{b^2}(1-\cos bx)e^{-ax} \quad \text{for } x > 0,$$

while $p_1(x) = 0$ for $x \leqslant 0$, and

$$p_2(x) = e^{-cx}/c \quad \text{for } x > 0,$$

while $p_2(x) = 0$ for $x \leqslant 0$, respectively. Here $a > 0$ and $c > 0$. The characteristic functions of S_1 and S_2 are equal to

$$(38) \qquad F_1(z) = \int_0^\infty e^{yz}dS_1(y) = \frac{1}{1-z/a}\,\frac{1}{1-z/(a+ib)}\,\frac{1}{1-z/(a-ib)}$$

and

$$(39) \qquad F_2(z) = \int_0^\infty e^{yz}dS_2(y) = \frac{1}{1-z/c},$$

respectively.

Clearly, $S_j \in \boldsymbol{S}_A - \boldsymbol{S}_E$, $j = 1, 2$.

Now let

$$P_1(z) = Q(z) - Q(1),$$
$$P_2(z) = Q(z)[Q^2(z)+b^2] - Q(1)[Q^2(1)+b^2],$$

where $Q(z) = (z-d)(z-e)^2$. Note that $P_1(1) = P_2(1) = 0$. Letting d > 1, we put $a = -Q(1)$ and $c = -Q(1)[Q^2(1)+b^2]$. Then, in view of (38) and (39), we obtain $F_1[P_1(z)] \equiv F_2[P_2(z)]$. Hence

$$(40) \qquad \int_0^\infty e^{yP_1(z)}dS_1(y) = \int_0^\infty e^{yP_2(z)}dS_2(y),$$

where both integrals exist for all $z \in \{z\colon \operatorname{Re} Q(z) < 0\}$.

It is clear that (40) implies the non-identifiability of $K(S, C_F)$ provided $\exp\{P_1(z)\}$ and $\exp\{P_2(z)\}$ can be, under the condition $d > 1$, generating functions of composed Poisson distributions. This is shown by proving that there exist real numbers b, e and $d > 1$ such that all coefficients of both derivatives $P_1^{(1)}(z)$ and $P_2^{(1)}(z)$ are non-negative.

First we find

$$\begin{aligned} P_1(0) &= q_0 = -e^2 d, \\ P_1^{(1)}(0) &= q_1 = e^2 + 2ed, \\ P_1^{(2)}(0) &= q_2 = -2(2e+d), \\ P_1^{(3)}(0) &= q_3 = 3! \end{aligned}$$

and

$$\begin{aligned} P_2^{(1)}(0) &= 3q_0^2 q_1 + b^2 q_1, \\ P_2^{(2)}(0) &= 3(2q_0 q_1^2 + q_0^2 q_2) + b^2 q_2, \\ P_2^{(3)}(0) &= 3(2q_1^3 + 6q_0 q_1 q_2 + q_0^2 q_3) + b^2 q_3, \\ P_2^{(4)}(0) &= 6(6q_1^2 q_2 + 4q_0 q_1 q_3 + 3q_0 q_2^2), \\ P_2^{(5)}(0) &= 30(3q_1 q_2^2 + 2q_0 q_2 q_3 + 2q_1^2 q_3), \\ P_2^{(6)}(0) &= 30(3q_2^3 + 12q_1 q_2 q_3 + 2q_0 q_3^2), \\ P_2^{(7)}(0) &= 30(21q_2^2 q_3 + 14q_1 q_3^2), \\ P_2^{(8)}(0) &= 2580 q_2 q_3^2. \end{aligned}$$

Now observe that

$$(41) \qquad \begin{gathered} P_1^{(1)}(0) \geqslant 0, \quad P_1^{(2)}(0) \geqslant 0, \\ P_2^{(1)}(0) \geqslant 0, \quad P_2^{(7)}(0) \geqslant 0, \quad P_2^{(8)}(0) \geqslant 0, \end{gathered}$$

provided that $q_1 \geqslant 0$ and $q_2 \geqslant 0$.

Because of inequalities

$$\lim_{e\to-\infty}\left[-\frac{P_2^{(4)}(0)}{e^5}\right] > 0, \qquad \lim_{e\to-\infty}\frac{P_2^{(5)}(0)}{e^4} > 0, \qquad \lim_{e\to-\infty}\left[-\frac{P_2^{(6)}(0)}{e^3}\right] > 0,$$

we can select e so small that in addition to the inequalities $q_1 \geqslant 0$ and $q_2 \geqslant 0$ the following inequalities hold:

$$(42) \qquad P_2^{(4)}(0) \geqslant 0, \quad P_2^{(5)}(0) \geqslant 0 \quad \text{and} \quad P_2^{(6)}(0) \geqslant 0.$$

Finally, for any fixed $d > 1$ and e we can select b so large that

$$(43) \qquad P_2^{(2)}(0) \geqslant 0 \quad \text{and} \quad P_2^{(3)}(0) \geqslant 0.$$

Combining inequalities (41) to (43), we obtain the desired result.

References

[1] W. Feller, *On a general class of contagious distributions*, Ann. Math. Statist. 14 (1943), p. 389-400.
[2] W. Klonecki, *Mixtures and characteristic functions*, Proc. Nat. Acad. Sci. USA 4 (1970), p. 831-836.
[3] — *Identifiability questions for chance mechanisms underlying stochastic models for carcinogenesis*, Math. Biosciences 7 (1970), p. 365-377.
[4] J. Neyman and E. L. Scott, *Statistical aspects of the problem of carcinogenesis*, Proc. 5-th Berkely Symp. Math. Stat. Prob., vol. 4 (1967), p. 745-775.
[5] Rényi, *On composed Poisson distributions*, II, Acta Math. Acad. Sci. Hung. 2 (1951), p. 83-98.
[6] W. Sierpiński, *Teoria liczb* (*Number theory*), Warszawa 1950.
[7] H. Teicher, *Identifiability of mixtures*, Ann. Math. Statist. 32 (1961), p. 244-248.

Received on 18. 9. 1970

W. KLONECKI (Wrocław)

IDENTYFIKOWALNOŚĆ MIESZANYCH ZŁOŻONYCH ROZKŁADÓW POISSONA

STRESZCZENIE

Niech

$$g_N(u|\mathbf{c}_N, y) = \exp\left\{y \sum_{k=1}^{N} c_k(u^k - 1)\right\},$$

gdzie

$y \geqslant 0$, $\sum_{k=1}^{N} c_k = 1$, $|u| < 1$ oraz $\mathbf{c}_N = \{c_1, \ldots, c_N\}$, przy czym $c_i \geqslant 0$ i $c_N > 0$, będzie funkcją tworzącą nieujemnej zmiennej losowej o wartościach całkowitych. Niech S będzie dystrybuantą określoną na przedziale $[0, \infty)$ i niezdegenerowaną w zerze. Wówczas

$$G_N(u|S, \mathbf{c}_N) = \int_0^\infty g_N dS(y)$$

jest funkcją tworzącą mieszanego złożonego rozkładu Poissona. Jeden z przedstawionych wyników orzeka, że dla każdego ustalonego $N \geqslant 1$ funkcja G_N wyznacza jednoznacznie $\mathbf{c}_N$ oraz S. Wynik ten uogólnia twierdzenie Fellera [1], że rodzina mieszanych rozkładów Poissona jest identyfikowalna. Inny wynik pokazuje, że $G_N \equiv G_M$ nie pociąga za sobą równości $N = M$. Ponadto, w pracy przedstawiony jest kompletny dowód twierdzenia, opublikowanego w [2] i [3], orzekającego, że jeżeli funkcja charakterystyczna dystrybuanty S jest całkowita, to G_N wyznacza też jednoznacznie N.
