DIVERSITY SENSITIVITY AND MULTIMODAL BAYESIAN STATISTICAL ANALYSIS BY RELATIVE ENTROPY

ROY B. LEIPNIK[1] and C. E. M. PEARCE[2]

Abstract

A list of recognised social diversities is assembled, including those used in social action programmes in the USA. Responses to diversity are discussed and diversity sensitivity defined as the derivative of response with respect to a defining parameter of a diversity distribution. Rewards (or penalties) for diversity are listed also; sensitivities to the responses to the rewards for diversity are called diversity sensitivities of the second kind. The statistics of bimodal and multimodal distributions are discussed, including the parametric estimation of such distributions by mixtures of multivariate normal distributions. An example is considered in detail.

1. Introduction

1.1. Diversity   In recent years diversity has received attention in such human social contexts such as social welfare, empowerment and affirmative action. From an administrative viewpoint, quantitative aspects and legal complications inevitably arise. Standard statistical approaches are compromised by the multimodal nature of many types of diversity. However modern statistical estimation has something to contribute when the various concepts involved are properly delineated and sufficiently circumscribed. We may begin by defining diversity as manifest difference as opposed to latent difference.

An overwhelming majority of studies address biological diversity, at the genus, species, organism, organ or cell level; of 1058 diversity references in a recent Internet search, 994 were strictly biological and 64 social or sociobiological. A list of 32 types of social diversity was assembled by collating several references on human diversity.

[1]Mathematics Department, University of California, Santa Barbara, CA 93106–3080, USA; e-mail: leipnik@math.ucsb.edu.
[2]School of Mathematical Sciences, The University of Adelaide, Adelaide SA 5005, Australia; e-mail: charles.pearce@adelaide.edu.au.

Eight or nine are employed explicitly in social programmes and others used tacitly. The 32 diversities are: ability, age, class, clothing, conduct, disability, education, ethnicity, family type, foods, health, hobbies, housing, ideology, income, intelligence, language, "looks", marital status, nationality, occupation, personality, politics, race, religion, self esteem, sex/gender, sexual orientation, somatotype, sports, transport and values.

An early reference to human diversity (of tastes) is from J. S. Mill (1859). "Looks" have several components — hair colour, eye colour, nose shape, posture, *etc*. "Language" includes accent. The term "cultural" is applied to many types of diversity, but there is significant disagreement as to which are cultural or acquired and which primarily or partly genetic. The distinction is important legally and psychologically but pragmatically unhelpful. Most types of diversity can be modified by individual, family or social action. Some diversities are absent in some societies or communities. Others occur primarily in clusters, such as when clothing indicates gender, religion, age and occupation. Further diversities may be hybrid; thus ethnicity, for example, may be Polish-Belgian. Compound diversities, such as Afro-American, are also often encountered.

1.2. Hands   We term the set of diversities exhibited by an individual a *hand*. Responses to a diversity-hand set A may be individualised or statistically predictable within individual responders having in common a diversity-hand set B. This is complicated by the fact that, within a particular diversity, the statistical distribution of any element of that diversity is likely to be bimodal or multimodal rather than following a Gaussian or other unimodal model. Bimodal distributions of response to prescription medicine are well known and documented. For social diversities, data are partly anecdotal, though bimodal or trimodal raw data are quite well documented in musical and mathematical test scores. Such results are often attributed to significant mini-diversities already within the test populations before the tests are taken and diversity induced by the test experience neglected. Whatever the reason, multimodality occurs sufficiently often and strongly to invalidate statistical reasoning based on unimodal distributions, normal or otherwise. The central limit theorem, which supports normality in large samples, has conditions and limitations often violated under diversity conditions. In any case, an estimation procedure that provides for multimodality will relapse to unimodality when all but one of the estimated weights is numerically negligible. The importance to policy making of a substantial bimodality and the capability of modern computing to handle the increased complexity of bimodal estimation justify the moderate increase in conceptual difficulty. Certainly in factor analysis an insistence on a single-factor model of intelligence on the grounds of simplicity and efficiency would not be taken seriously now.

1.3. Sensitivity   One way of describing response is in terms of the intensity of response to a particular level of stimulus. The cumulative probability distribution function is less indicative than the probability density function. Thus we refer to a function that characterises the state of a system as the *sensitivity* of that state. We note that high negative and positive rates to a condition both indicate high sensitivity. Insensitivity to a particular condition indicates a lack of change of response as the condition changes.

Responses can be prioritised for a diversity pair, say ethnicity e and income i. Thus we may, for example, have response $r(e, i) = 0$ for $e \leq e_1$, regardless of i, while for $e > e_1, r(e, i)$ is weakly dependent on e but strongly on i. Many other peculiarities of response to diversity A by B depend on the relative hands involved, with symmetry or trade-offs between e and i absent as in the example above. Another phenomenon is the *venal* response in which for any r_1, e_1 there exists an i_1 such that $r(e_1, i) > r_1$ whenever $i > i_1$, that is, money talks.

We define *diversity of the first type* as sensitivity of a response by B to the diversity revealed by A. A third observer C may respond to B's sensitivity to A, if that is also revealed. That response of C has also a diversity, which we term a second-order sensitivity of the first type. Since C may quite easily share a good deal of the diversity of A, the second-order sensitivity can easily amplify the first-order sensitivity. If $r(B|A)$ is the response of B to A and $s(B|A)$ its sensitivity, then $r(C|r(B|A))$ is a second-order response and $s(C|s(B|A))$ a second-order sensitivity. We have also $r(C|s(B|A))$ and $s(C|r(B|A))$. Responses will in general depend on parameters and derivatives may be with respect to any of these, so sensitivities will be vectors or tensors if responses are scalars or vectors respectively.

1.4. Rewards   A distinction of importance in the context of social action relates to rewards or penalties for diversity. Some institutionally oriented (encouraged, tolerated or mandated) rewards are bonus points towards admission, retention or promotion, higher marks, quotas, timetables, accelerations, opportunities connected with present or future spouses or relatives, pay differentials, extra praise or less criticism, lower expectations, access to networks, grants and loans. Some of these intended rewards may be perceived as insulting. Penalties are the opposites of intended rewards, plus overt insults, snubs, slander, intimidation, threats, dismissal and violence. Longer lists of rewards can be found in various types of affirmative action plans; lists of punishments are often not overtly promulgated.

The effective total reward is seldom a linear functional of the separate items and similarly for penalties or a mixture of the two. Some rewards are "remedies", intended to correct previous penalties to individuals, their relatives or others sharing or considered to be sharing the same or some of the same diversities. History and perception are therefore important.

Reward systems are often "logistic" in that reward increments are larger away from the ends of the diversity scale. Thus with skin colour in the USA, responses to different light colours are often similar as are those to different dark colours. The sensitivity is probably greatest between light and medium brown. The tone of greatest sensitivity will define a de facto colour line. This varies regionally, by period and with the particular diversity which is responding. Similarly in political and legal questions there is a tendency to find a "bright line" of distinction by exaggerating certain small differences near the centre while playing down much larger differences near the extremes. While nonlinearities of reward are perhaps less pronounced than those of response, when these are compounded the results can be effectively discontinuous or categorical and often disastrous.

1.5. Quantitative formulation   We may treat a particular hand as a point α in diversity space, yielding a reward $\rho(\alpha)$. There is an associated reward response function $r_\rho^b(\alpha, \rho(\alpha))$ depending on the responder B, the diversity hand α and the kind and amount of the reward $\rho(\alpha)$. Suppose the vector v of parameters characterises a class of distributions and $\psi(\alpha, v)$ a class of distributions of individuals over the diversity α. This leads to a density $\hat{\psi}(\alpha)$, obtained by picking from $\psi(\alpha, v)$ by a choice $\hat{v}(f)$ a distribution closest in a suitable sense to an empirical distribution $f(\alpha)$.

The total linear response of B to the reward system ρ over the whole population is then

$$\hat{r}_\rho^B = \int r_\rho^B(\alpha, \rho(\alpha)) \hat{\psi}(\alpha) \, d\alpha.$$

If A is a particular set of individuals eligible for the reward ρ, then

$$^A\hat{r}_\rho^B = \int_A r_\rho^B(\alpha, \rho(\alpha)) \hat{\psi}(\alpha) \, d\alpha$$

is the total response of B to the reward system ρ as applied to A. If α is a particular point in A, then the derivative of $^A\hat{r}_\rho^B$ with respect to A, evaluated at α, is $r_\rho^B(\alpha, \rho(\alpha))$. This definition takes account of the "spectre" test, taking account of the frequency of similar diversities. This is a sensitivity of the second type, the sensitivity of the total response of B to the reward obtained by the individual α in A on the reward system ρ. Similarly if $r^B(\alpha)$ is the response of B to α in A, then the total response to A by B of the first type is

$$^A\hat{r}^B = \int_A r^B(\alpha) \hat{\psi}(\alpha) \, d\alpha$$

and the sensitivity of that response to α is $r^B(\alpha)\hat{\psi}(\alpha)$. Likewise the parametric

sensitivities are

$$\frac{\partial}{\partial v} \int_A r^B(\alpha)\psi(\alpha, v)\, d\alpha \bigg|_{v=\hat{v}(f)} \quad \text{and} \quad \frac{\partial}{\partial v} \int_A r^B_\rho(\alpha, \rho(\alpha))\psi(\alpha, v)\, d\alpha \bigg|_{v=\hat{v}(f)}.$$

The total sensitivity to individuals in a small group can be estimated by the response of one of the individuals times the density, evaluated at that individual, of the probability $\hat{\psi}(\alpha) = \psi(\alpha, \hat{v}(f))$. Choice of $\hat{v}(f)$ from least squares is relatively easy but known to be of poor quality in multimodal, multivariate cases [3, page 35]. Choice by maximum likelihood usually requires iterative methods. This was used in unimodal linear models by Koopmans, Rubin and Leipnik [8] and for multimodal density estimation by Hasselblad [7] and Wolfe [14], based on the EM method. See Dempster *et al.* [2]. These iterations may or may not converge to the global maximum [6]. A Bayesian method intermediate in quality and accuracy is maximum relative entropy. A non-iterative version of this, applied to an educational problem discussed in Section 3, is outlined in Section 4.

2. Statistics of diversity

Three overlapping statistical literatures of interest here are physical statistics, psychometrics or educational statistics, and general mathematical statistics. The contributions of the first include three estimation methods: least squares (Gauss-Legendre), maximum entropy (Gibbs) and minimum entropic distance (Boltzmann). These can be used to choose distributional forms and parameters within a form.

The second includes three related techniques: multivariate normal distributions of manifest variables, linear structures and the analysis of latent factors linearly underlying a larger set of measured variables. This factor analysis of Thurstone suggests equations for factor loading and covariance, later derived by Lawley from maximum likelihood (Fisher's general estimation method), in the unimodal multivariate normal case called homogeneous.

Muthén and his associates have extended in two stages conventional homogeneous factor analysis to a limited collection of heterogeneous variables. In the first stage the variables come from differently labelled distributions of Fisher's first class, that is, with identical covariance matrices. In the second the variables are generated by a set of auxiliary random variables rather than fixed labels, but again the covariance matrices are identical. This does not handle realistic heterogeneity, in which the covariances are non-identical across the variables (Fisher's second class).

In general statistics, the Fisher-Behrendt problem of choosing between two proposed distributions or members of a parametric family is still controversial. The original solution of Pearson was based on chi-squared tests for normal distributions.

This was generalised by von Mises to omega-squared tests and converted by Smirnov and Kolmogorov to a nonparametric form, further developed by Grenander and Geman and Geman.

Fisher's own solution is linear for Fisher's first class and quadratic for his second. Another nonparametric method of Fix and Hodges [4], Rosenblatt [11] and Parzen [9], termed the kernel method, is discussed by P. Rao *et al.* [10]. When the empirical distribution is discrete, the kernel method as modified by Nadaraya performs a smoothing before final analysis (see Hand [5]). Cox [1] and Silverman [12, 13] improved and reoriented these theories, often using mixtures of normals in the univariate case. It is clear from their results that combining multimodal and multivariate aspects, as in diversity problems, leads to massive data storage and retrieval problems.

We have noted above the EM method. Just when this exhibits convergence and when divergence is not well understood, but increase from an initial value has been demonstrated. The Newton-Kantorovitch method applied to maximum likelihood problems need not behave well from a general initial value, but converges quadratically from appropriate initial values. The convergence rate of the EM method is not well established from any initial value.

As noted, least-squares methods are poor for this class of problems, though far simpler. Intermediate in quality and difficulty are the entropy methods, maximum relative entropy and minimum entropy distance. These permit direct non-iterative calculations provided certain logarithmic integrals are evaluated.

3. An education diversity problem

A salient diversity problem in the USA relates to employment and unemployment. Closely related is training at various levels of sophistication, for which a useful preliminary is secondary or tertiary education. Under current US plans, voluntary national examinations are to be instituted at various stages, such as 4th, 6th, 8th, 10th and 12th grades. The first stage has partially materialised as part of the "leave no child behind" legislation. Presumably the ETS organisation, arguably one of the most advanced psychometric firms in the USA, will be active here. One of its better-known operations is the SAT tests, consisting of verbal and mathematical components and given several times a year to secondary-school students nearing graduation.

This involves six diversities, four from the original list and two, the two SAT scores, peculiar to the SAT. The former four are ethnicity and income as status diversities and verbal and mathematical ability as ability variables. Class, gender and personality could be helpful but are dropped. The score diversities are given quantitatively. Income can be given also, subject to appreciable noise from family size, spending habits, *etc.* Ethnicity is difficult to score on a linear scale. Even colours are usually described by

three numbers relating to hue, intensity and saturation. But suppose "colour ethnicity" is made linear by a photographer's gray scale. Since we allow six-dimensional multimodality, the location of the mode centres does not imply simple ordering of quality, although five of the six dimensions do have quality interpretations. We do not assume any particular relation between these four performance and two status diversities, though it is plausible that mathematical ability as estimated by teachers, grades, *etc.* will be positively correlated with SAT mathematics scores and verbal ability with SAT verbal scores. As for the other correlations, we need no assumptions. Moreover we need not assume any particular number of distinct modes in this seven-dimensional space of six diversities and one probability dimension. For definiteness, assume four modes associated with four six-dimensional ellipsoids, each with a centre and nonnegative weight. This results in a space with $3 + 4(6 + 6 \cdot 7/2) = 111$ parameters to determine the best approximate distribution. Assuming a conservative 50 data points per degree of freedom, we need only 5550 data points, which could easily be generated by one year's college hopefuls in a couple of congressional districts. The complete distribution fitting is unmanageable, though.

With some loss of generality, remove verbal ability and score and assume a bimodal distribution, but retain the two status variables. There are now $1 + 2(4 + 4 \cdot 5/2) = 29$ parameters and only 1450 data points are needed. To do the necessary Newton-Kantorovitch calculations, many 29×29 matrices would be inverted, each fitting on a PC. Of course this gives us only the $\hat{\psi}(e, i, a, s)$ density, where e, i are the individual ethnicity and income numbers, a mathematical ability and s the score. We need also an estimated response $r(e, i, a, s)$ of some group or groups to the information that an individual of diversity hand (e, i, a, s) exists or that an individual of pre-test hand (e, i, a) has received a score s. This will produce a sensitivity of the first kind to that post-test hand when multiplied by $\hat{\psi}$. If a reward function $\rho(e, i, a, s)$ is projected, responses $r_\rho(e, i, a, s, \rho(e, i, a, s))$ to that combination of hand and reward would be required. Assuming ten levels of each diversity variable, $10,000$ reward values would be inserted and, of 100 respondents, each would have to be asked $1,000$ quite similar questions, making $100,000$ person-questions. Or 1450 people could be asked 70 questions each, scattered on a grid, once for each proposed reward system and once to determine the basic first-type diversity sensitivity for this particular diversity. The political feasibility of various proposed reward systems could then be assessed.

An entirely different calculation would be needed to estimate the long-term individual advantage or social advantage in providing these rewards related to SAT scores and the diversities. After receiving a reward an individual may be found in another income level, social and occupational diversity, *etc.*, depending on conduct, self-esteem and the like. Anticipation of such changes would affect the responses to the projected rewards.

The procedures outlined above seem cumbersome, especially in the light of current

diversity approaches. However before discarding the present disorganised system of diversity reward allocation as unsuitable, it is interesting to see whether and how more rationality can be infused by straightforward statistical and survey methods based on commonly accepted definitions of diversity and sensitivity.

Larger problems are tackled regularly in modern scientific and technological work. If the field of diversity sensitivity is to become more credible, improvement in its statistical methodology would be appropriate.

4. The four-diversities case

In the reduced example with four diversities and two modes, the assumed bimodal-type normal distribution has the form

$$\psi(x, v) = (2\pi)^{-2} \sum_{j=1}^{2} g_j (\det B_j)^{1/2} \exp(-z_j B_j z_j^T),$$

where $B_j^{-1} = A_j$ is a modal covariance matrix, μ_j the mean of the j-th modal distribution and $z_j = x - \mu_j$. If the scripts e, i, a, s refer to ethnicity, income, (pre-test) ability and the test score, then

$$\mu_j = (e_j, i_j, a_j, s_j)$$

give the components of the modal means and

$$A_j = \begin{bmatrix} \sigma_{j,ee} & \sigma_{j,ei} & \sigma_{j,ea} & \sigma_{j,es} \\ \sigma_{j,ie} & \sigma_{j,ii} & \sigma_{j,ia} & \sigma_{j,is} \\ \sigma_{j,ae} & \sigma_{j,ai} & \sigma_{j,aa} & \sigma_{j,as} \\ \sigma_{j,se} & \sigma_{j,si} & \sigma_{j,sa} & \sigma_{j,ss} \end{bmatrix}$$

the covariances. For $j = 1, 2$ there are 28 parameters plus one free weight g_1, since $g_1 + g_2 = 1$, giving 29 parameters to be estimated. Let

$$H = -\int^{(4)} \psi(x, v) \ln(\psi(x, v)/f(x)) \, d^{(4)}x,$$

where $\int^{(4)}$ is a quadruple integral over all values $x = (e, i, a, s)$ and $f(x)$ is the empirical distribution to be fitted. One quantity to be calculated is

$$H_2 = \int^{(4)} \psi(x, v) \ln f(x) \, d^{(4)}x$$

$$= \sum_{j=1}^{2} g_j (\det B_j)^{1/2} \int^{(4)} \exp(-z_j B_j z_j^T/2) \, d^{(4)}x,$$

which could be estimated from the presumed 29×50 data points using (say) a Monte Carlo approach for selecting one percent of the 5^{14} parameter points obtained by choosing five points for each of the 14 parameters entering each integral. The other quantity,

$$
\begin{aligned}
H_1 &= -\int^{(4)} \psi(x, v) \ln \psi(x, v) \, d^{(4)} x \\
&= -\sum_{j=1}^{2} g_j (\det B_j)^{1/2} \int^{(4)} \exp(-z_j B_j z_j^T /2) \ln \psi \, d^{(4)}x,
\end{aligned}
$$

can be calculated in closed form as a function of v. Then $H = H_1 + H_2$.

What is required is a formula for calculating

$$
\ln \psi = \ln \left(\sum_{j=1}^{2} a_j \exp u_j \right)
$$

as a function of u_1 and u_2, where $a_j = g_j (\det B_j)^{1/2}$ and $u_j = -z_j B_j z_j^T /2$. The quantities

$$
\cdot q_j = \frac{a_j \exp u_j}{\sum_{k=1}^{2} a_k \exp u_k} = \frac{\partial \ln \psi}{\partial u_j}
$$

are desirable auxiliary variables. A Taylor series for $\ln \psi$ is

$$
\ln \psi(v^{(0)}) + \sum_{j=1}^{2} (u_j - u_j^{(0)}) q_j (u^{(0)}) + \sum_{m} P_m (u^{(0)}) \prod_{j=1}^{2} \frac{(u_j - u_j^{(0)})^{m_j}}{m_j!},
$$

where m is summed over all pairs on nonnegative integers (m_1, m_2) with $m_1 + m_2 = |m| \geq 2$ and

$$
P_m (y) = \left(\frac{\partial}{\partial y_1} \right)^{m_1} \left(\frac{\partial}{\partial y_2} \right)^{m_2} \ln \psi (y, v).
$$

The perhaps surprising fact is that $P_m(y)$ is expressible as a homogeneous polynomial of degree $|m|$ in q_1 and q_2 with coefficients satisfying a simple one-step recursion with variable coefficients. For any particular choice of $u^{(0)}$, determining $H_2(v)$ is reduced to calculating "two-centred" integrals of the form

$$
\int^{(4)} \exp(u_j) (u_1 - u_1^{(0)})^{m_1} (u_2 - u_2^{(0)})^{m_2} \, d^{(4)}x.
$$

These are of known type, being expressible by differentiation of

$$\int^{(4)} \exp(a_1 u_1 + a_2 u_2)\, d^{(4)}x$$

m_1 times with respect to a_1 and m_2 times with respect to a_2. But $a_1 u_1 + a_2 u_2$ is expressible as a single quadratic form in x, so the integral is known.

The expression for $P_m(y)$ does not seem to appear in the usual sources for double series or in the works of Hand, Grenander, Nadaraya or Geman in which multimodal theory is discussed extensively. We have

$$P_{2,0}(y) = \frac{\partial}{\partial y_1}\left(\frac{a_1 e^{y_1}}{\psi}\right) = \frac{a_1 e^{y_1}\psi - (a_1 e^{y_1})^2}{\psi^2}$$

$$= \frac{a_1 a_2 e^{y_1 + y_2}}{\psi^2} = q_1 q_2 = P_{0,2}(y)$$

and $P_{1,1}(y) = -q_1 q_2$. Likewise

$$P_{m_1,m_2}(y) = (-1)^{m_2} P_{m_1+m_2,0}(y).$$

If $P_{r,0}(y) = \sum_k b_{k,r} q_1^k q_2^{r-k}$, then

$$P_{r+1,0}(y) = \sum_{k=0}^{r} b_{k,r}\left[k q_1^{k-1} q_1 q_2 q_2^{r-k} - (r-k) q_1^k q_1 q_2 q_2^{r-k-1}\right]$$

$$= \sum_{k=0}^{r+1} b_{k,r+1} q_1^k q_2^{r+1-k}.$$

Hence the relevant recurrence is $b_{k,r+1} = k b_{k,r} - (r-k+1) b_{k,r-1}$ with $b_{0,2} = b_{2,2} = 0$ and $b_{1,2} = 1$.

Thus

$$P_{m-1,m_2}(y) = (-1)^{m_2} \sum_{k=0}^{m_1+m_2} b_{k,m-1+m_2} q_1^k q_2^{m_1+m_2-k}$$

and

$$P_m(u^{(0)}) = (-1)^{m_2} \frac{\sum_{k=0}^{|m|} b_{k,|m|}(a_1 \exp(u_1^{(0)}))^k (a_2 \exp(u_2^{(0)}))^{|m|-k}}{\sum_{k=0}^{|m|} \binom{|m|}{k}(a_1 \exp(u_1^{(0)}))^k (a_2 \exp(u_2^{(0)}))^{|m|-k}}.$$

We may calculate $H_2(v)$ to any desired accuracy by taking $|m|$ large enough, while $H_1(v)$ can be approximated by Monte Carlo methods. Comparing the values of $H(v) = H_1(v) + H_2(v)$, a maximum (or competing maxima) may be estimated.

Another approach in such an optimisation problem is the cooling or annealing method. A third approach is to calculate the various integrals that arise in finding

$$h_k = \frac{\partial H(v)}{\partial v_k} = \frac{\partial H_1(v)}{\partial v_k} + \frac{\partial H_2(v)}{\partial v_k}$$

and solve for the vectors $\hat{v}(f)$ that satisfy

$$h_k(\hat{v}) = 0 \quad \text{for} \quad k = 1, 2, \ldots, 29.$$

If this is done by Newton's method, the quantities $h_{k,\ell} = \partial h_k / \partial v_\ell$ must be determined too.

References

[1] D. R. Cox, "Notes on the analysis of mixed frequency distributions", *Brit. J. Math. Statist. Psych.* **19** (1966) 39–47.
[2] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", *J. Roy. Statist. Soc. Ser. B* **39** (1977) 1–38.
[3] R. S. Everitt and D. J. Hand, *Finite Mixture Distributions* (Chapman and Hall, London, 1981).
[4] E. Fix and J. Hodges, "Discriminatory analysis", Report 4 Project 21-49-004, USAF School of Aviation Medicine, (1951).
[5] D. J. Hand, *Discrimination and Classification* (J. Wiley and Sons, Chichester, 1981).
[6] D. J. Hand, *Kernel Discriminant Analysis* (Research Studies Press/Wiley, Chichester, 1982).
[7] V. Hasselblad, "Estimation of parameters on a mixture of normal distributions", *Technometrics* **8** (1966) 431–444.
[8] T. C. Koopmans, H. Rubin and R. B. Leipnik, "Measuring the equation systems of dynamic economics", in *Statistical Inference in Dynamic Economic Models* (ed. T. J. Koopmans), (Wiley and Sons, New York, 1950) 53–237.
[9] E. Parzen, "On the estimation of a probability function", *Ann. Math. Statist.* **33** (1962) 1067–1076.
[10] P. Rao, S. Menjoge and P. S. R. S. Rao, "Improved estimators with the weighted and compounded loss functions", *Commun. Statist. A – Theory, Methods* **12** (1983) 131–139.
[11] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function", *Ann. Math. Statist.* **27** (1956) 832–837.
[12] B. W. Silverman, "Using kernel density estimates to investigate multimodality", *J. Roy. Statist. Soc. Ser. B* **43** (1981) 97–99.
[13] B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, London, 1986).
[14] J. H. Wolfe, "Pattern clustering by multivariate mixture analysis", *Multivar. Behavioral Res.* **5** (1970) 329–350.