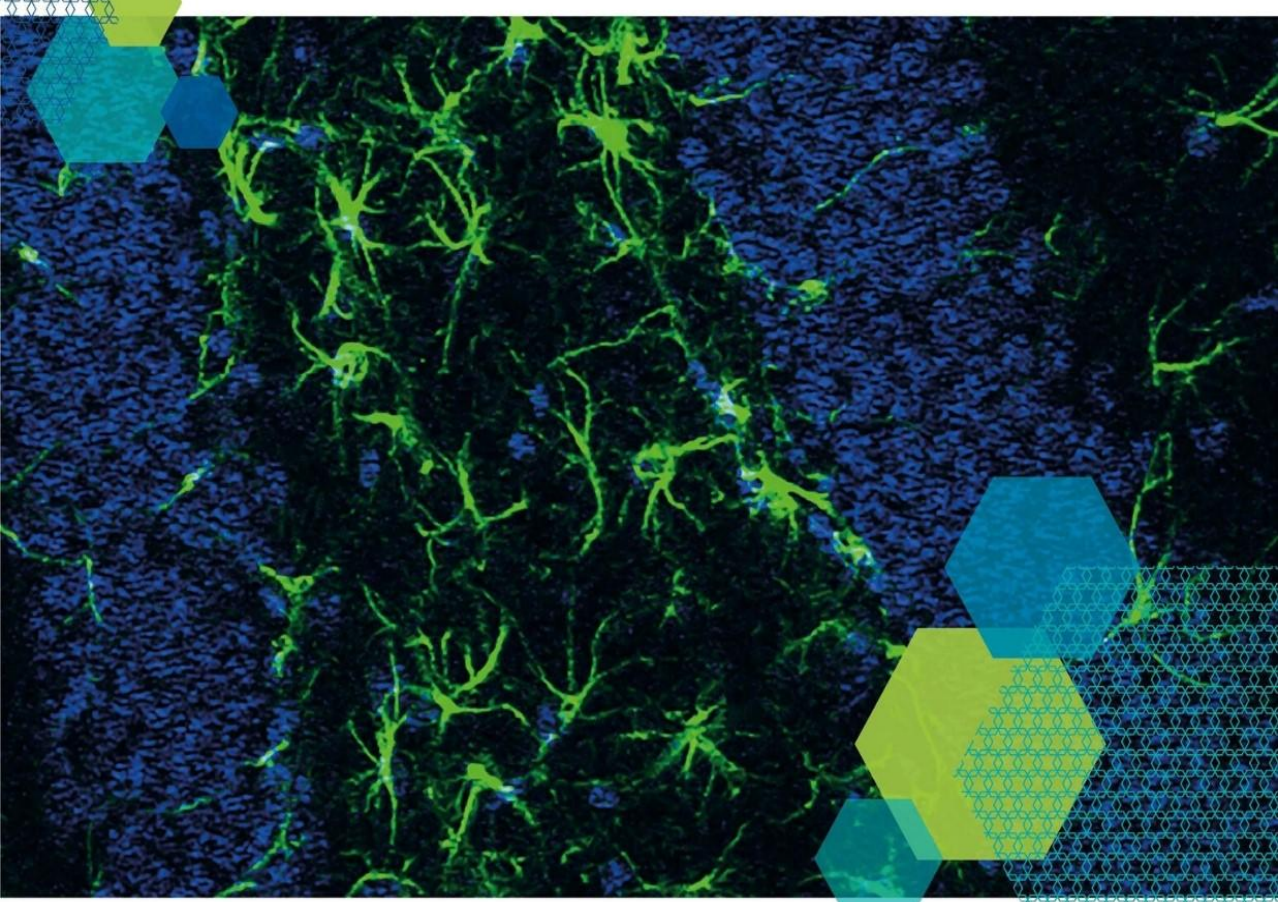
BRITISH PHARMACOLOGICAL SOCIETY

TODAY'S SCIENCE TOMORROW'S MEDICINES

Submit your abstract and register now:

CELL SIGNALLING 2020

20 – 21 April 2020
Stamford Court, University of Leicester, Leicester, UK



Registration details:

	Early bird	Standard
Member registration	£250	£285
Student registration	£150	£185
Non-member registration	£280	£305
Non-member student registration	£170	£195
Printed programme booklet	£10	£10

Abstract submission deadline:	Friday 31 January 2020
Member bursary deadline:	Friday 31 January 2020
Early bird registration deadline:	Friday 6 March 2020

For more information about attending or presenting please contact: meetings@bps.ac.uk or visit: www.bps.ac.uk/cell-signalling

Br. J. clin. Pharmac. (1989), **28**, 655–662

Testing for bimodality in frequency distributions of data suggesting polymorphisms of drug metabolism—hypothesis testing

P. R. JACKSON, G. T. TUCKER & H. F. WOODS
University Department of Pharmacology and Therapeutics, Royal Hallamshire Hospital, Sheffield S10 2JF

1 The theory of methods of hypothesis testing in relation to the detection of bimodality in density distributions is discussed.
2 Practical problems arising from these methods are outlined.
3 The power of three methods of hypothesis testing was compared using simulated data from bimodal distributions with varying separation between components. None of the methods could determine bimodality until the separation between components was 2 standard deviation units and could only do so reliably (> 90%) when the separation was as great as 4–6 standard deviation units.
4 The robustness of a parametric and a non-parametric method of hypothesis testing was compared using simulated unimodal distributions known to deviate markedly from normality. Both methods had a high frequency of falsely indicating bimodality with distributions where the components had markedly differing variances.
5 A further test of robustness using power transformation of data from a normal distribution showed that the algorithms could accurately determine unimodality only when the skew of the distribution was in the range 0–1.45.

Keywords polymorphic drug metabolism bimodality testing

Introduction

We have previously considered the use of graphical methods to detect polymorphisms of drug metabolism from *in vivo* data (Jackson *et al.*, 1989). Although these techniques can be used as an initial screen to identify the presence of bimodality in frequency distributions they cannot be used to test the strength of the hypothesis that two or more populations are present. Hypothesis testing requires a statistical test of the goodness of fit of the data to a theoretical distribution. Such testing generally assumes that the data are normally distributed.

Applying statistical tests to graphical displays

The application of the chi-square test to data in the form of histograms is the simplest method of testing for goodness-of-fit to a distribution. The mean and variance of the data are calculated and the number of subjects to be found in each cell is determined from the theoretical shape of the distribution. The chi-square test is then used to show whether the differences between the observed and predicted numbers in each cell could have arisen by chance. However, as suggested previously (Jackson *et al.*, 1989), the form of the histogram may be sensitive to the

Correspondence: Dr P. R. Jackson, University Department of Pharmacology and Therapeutics, Royal Hallamshire Hospital, Sheffield S10 2JF

positioning of the data cells and the distribution is invariably assumed to be normal or log-normal. The first of these problems is overcome by using the Kolmogoroff-Smirnov test (Kolmogoroff, 1941), which checks goodness-of-fit of the cumulative distribution to that of the predetermined distribution over a given range. A practical method of implementing both the chi-square and the Kolmogoroff-Smirnov tests is to increase the number of components of the distributions until the null hypothesis of an acceptable fit is not rejected. Fitting a multicomponent distribution to the data requires a method of estimating the means, variances and relative contributions of the component distributions. This may be done, for example, by applying the expectation maximization (EM) algorithm (Dempster *et al.*, 1977) as outlined by Everitt (1981).

Although the fit of the data may deviate insignificantly from a mixture of two normal distributions whilst being significantly different from the single normal distribution, this may follow from the use of more parameters alone and does not necessarily imply a significantly improved fit to the bimodal distribution.

Direct application of hypothesis tests

Maximum likelihood methods A method of checking statistically whether the data are best described by two distributions rather than one is to use a maximum likelihood fitting technique. The likelihood function of the whole data set is calculated as the multiplicand of the likelihoods of each individual datum point predicted by the distribution. Standard Newton-Raphson (Fletcher, 1980) or Nelder Mead simplex (Nelder & Mead, 1965) algorithms can be used to alter the parameters of the distributions to maximise the likelihood of the data according to the particular model. To determine whether the maximum likelihood achieved by a model with a mixture of two or more distributions is significantly superior to that achieved with a single distribution the value of twice the log of the ratio of maximum likelihoods is estimated. This is claimed to be distributed as chi-square at a number of degrees of freedom equal to the difference in the number of parameters between the two models (Wilks, 1938). However, as both Wolfe (1971) and Binder (1978) have pointed out, according to the null hypothesis the continuity assumptions as laid down by Wilks (1938) are violated. Binder (1978) suggested an alternative index for testing the probability of one distribution *vs* two with common variances, the suitability of which was subsequently confirmed by Everitt (1981) using Monte Carlo simulation.

Practical problems of model fitting When determining the maximum likelihood for the data according to a model comprising a mixture of two distributions occasional data sets cause the likelihood to increase monotonically towards infinity. This failure occurs when the value of one observation is selected by the algorithm as the mean of one component distribution. The likelihood of this data point, and hence the total data set, increases towards infinity as the variance of the distribution centred at that point falls (Murphy & Bolling, 1967). To prevent this happening an artificial constraint on the variance parameters is needed. The simplest solution is to make the variance of both component distributions equal. However, this may make the test less conservative by reducing the maximum likelihood when the data are fitted by the sum of two distributions. Evans *et al.* (1983) used an alternative approach, making an initial estimate of the parameters of the distribution using the graphical method of Bhattacharya (1967), and then performing a local search in this constrained area to find the point of maximum likelihood.

In an attempt to overcome the limitations of using only normal distributions (Jackson *et al.*, 1989) Maclean (1976) performed a power transformation on the index prior to fitting the data with normal distributions. This has the effect of providing good fits of normal distributions to unimodal but skewed data sets. The use of such transformations does have limitations, demanding that the degree of skew of each individual component is identical, and distorting the variances of the component distributions to different extents depending upon their mean values. Nevertheless this modification does prevent the selection of a model with two distributions to fit the data simply to accommodate skew within a data set derived from a unimodal distribution. On the other hand, it cannot be guaranteed that the modified likelihood ratio suggested by Everitt (1981) will be as useful in distinguishing between models using the maximum likelihood method with power transform (mlpt). Another disadvantage of using a transformed index is that the two-step method of Evans *et al.* (1983) used to prevent the variance parameters approaching zero cannot be applied and an alternative constraint must be introduced to prevent one of the variances approaching zero.

Kao (1959) suggested an alternative theoretical distribution which was subsequently fitted to pharmacogenetic data by Jack (1983). This is the flexible distribution of Rosin, Rammler, Sperling and Weibull (RRSW) (equation 1).

$$f(x) = \frac{\beta(x - \Omega)^{\beta - 1}}{\alpha} e^{-\theta} \qquad \text{equation 1}$$

where $\theta = \frac{(x - \Omega)^{\beta}}{\alpha}$

and Ω is the position parameter
α is the scale parameter
β is the shape parameter.

This distribution accommodates skew in the data of either a positive or a more limited negative extent. If the data are fitted using more than one distribution each component may be skewed to a different degree. The possibility of singularities arising as the scale parameter (α) approaches zero is reduced as the terms inside and outside the exponentiation (equation 1) move in opposite directions as the value of the scale parameter falls. Some constraint may be required on the value of the location parameter (Ω) to maintain the term within the central brackets ($x - \Omega$) above zero.

Non-parametric methods of hypothesis testing Another method of hypothesis testing for bimodality which does not make assumptions concerning the underlying distribution of the data was suggested by Silverman (1981). Each datum point is represented by a kernel, in this case a small normal distribution. The kernels are then summed to describe the likely population distribution from which the data was originally sampled. The parameter, *Hcrit*, which determines the spread of the individual kernels, is adjusted to the minimum value sufficient to produce a unimodal population distribution. To test the hypothesis that the sampling distribution is unimodal a bootstrap technique is used. Samples, of the same size as the data set, are taken at random from the observations with replacement (i.e. some points may be sampled more than once). An error term is added to each sample point, the kernels of this new set of points are summed, and the modality of this distribution is determined. If the original sample has arisen from a bimodal parent distribution the value of *Hcrit* would have to be large to produce unimodality. Thus, when this same value of *Hcrit* is applied to the simulated bootstrap samples from the smooth unimodal distribution a higher than expected proportion will be unimodal. Arbitrarily 95% of samples are expected to be unimodal and when the proportion falls below this the null hypothesis, under which the parent distribution is unimodal, is accepted. This method may be extended to examine the data for any number of modes by determining a new value of *Hcrit* for that number of modes and testing the proportion of bootstrap samples having $\leq$ n modes. In practice the null hypothesis is tested by increasing the modality until a number of modes is discovered at which the hypothesis is not rejected.

Scaling A problem in detecting bimodality in a density distribution is the scaling of the observations. Intuitively it would appear that the graphical and maximum likelihood methods, based on fitting normal distribution, are most likely to be sensitive to changes of scale. Limitations of the power transform and the shapes of the Weibull distributions may make the other maximum likelihood techniques scale sensitive. Experimental data from a study (McGourty *et al.*, 1985) of the population distribution of the urinary drug/metabolite ratio for debrisoquine and metoprolol of man indicate that even the kernel density method of Silverman (1981) may be sensitive to transformation of the data. Thus, when the urinary drug/metabolite ratio data were presented to the algorithm the null hypothesis of unimodality was not rejected, yet when the data were log transformed bimodality was indicated (Jackson, 1988).

The effects of scaling are not only due to deficiencies of the algorithms, since transformations of scale can alter the modality of a distribution. Analytical proof of this is difficult but using simulation techniques the extent of alterations may be illustrated (Jackson, 1988). Figure 1a, b shows the same simulated distribution plotted on linear and log scale respectively. Whereas the linear plot is unimodal, it is clear that log transformation produces marked bimodality. This is because transformations of probability density functions requires multiplication by the first derivative of the transforming function (Lindgren, 1968) unlike changes of scale with simple graphs.

We have used a computer simulation to investigate the power of the various methods of hypothesis testing for the modality of unknown distributions from which samples are available. In addition the effects of deviations from normality on the results obtained with the different methods were examined.

Methods

Evaluation of statistical power

Pseudo-random samples, each of 100 points, were generated from a mixture of two normal distributions. The algorithm NAG G05DDF was used to produce random numbers from a normal distribution and NAG G05CAF was used to allocate each number at random to one of

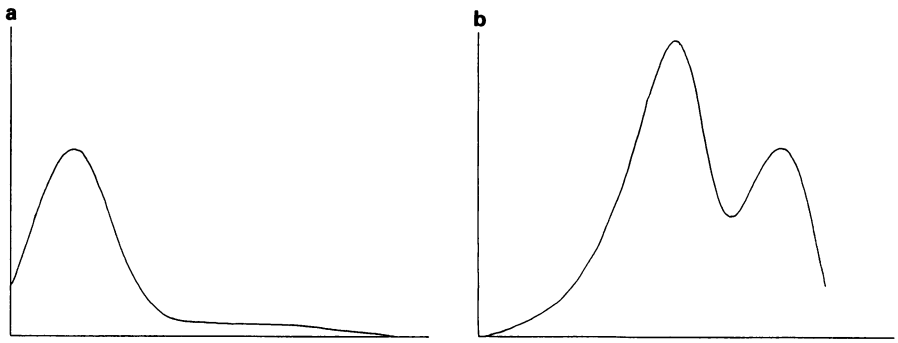


Figure 1 (a) A skewed distribution (generated from the sum of two normal distributions) and (b) its log transformation.

the components according to the mixing proportion. The separation between the means of the distributions, expressed as the common variance, was varied from 2–6 standard deviation units. The samples were then fitted by parametric models (Murphy & Bolling, 1967) of single or double normal distributions transformed by a power factor (Maclean *et al.*, 1976). The values for the means, common variance, mixing proportion and power transform factor were determined using the Newton-Raphson algorithm NAG E04CCF to produce a model with maximum likelihood for the data. The fitting procedure was repeated with models of one and two component distributions. The mixing proportions were constrained to allow an area under the total curve of unity and the power transform factor was constrained to have a value of between −10 and +10. The test statistic described by Everitt (1981), based on the Naperian log of the ratio of the likelihoods, was then compared with a chi-square distribution. The 5% level was taken as being significant. The same samples were then subjected to the Silverman algorithm and again the 5% level was taken as significant. Each procedure was repeated 100 times with different generated samples. In a separate series of simulations the power of the maximum likelihood Weibull method was tested in the same way using the same NAG algorithms to generate the data sets but using a computer with a different word length. The mixing proportions were again constrained to keep the total probability equal to unity and the shape, scale and position parameters in the range 0–100.

Sensitivity to deviations from normality

To determine the sensitivity of the transformed maximum likelihood and kernel density methods to deviations from normality simulations were performed using a series of samples from a number of unimodal distributions constructed from the sum of two normal distributions. To ensure unimodality the parameters of the mixture distributions and their mixing proportions were constrained within the limits described by Behboodian (1970). These samples were then tested for their modality using the kernel density and maximum likelihood with power transform methods as described previously. In a more rigorous investigation of the sensitivity to non-normality of these two methods, samples from distributions with a wide range of skews were tested and the effect of prior log transformation of the data was studied. Samples of 100 data points were generated from random normal distributions using the subtractive algorithm of Knuth (1981) followed by the Box-Mueller algorithm (Press *et al.*, 1986). These variates were then transformed using the Maclean power transform to produce samples from distributions with skew values between −3 and +10. The samples were examined using the kernel density and mlpt algorithms with and without prior log transformation.

Results

Two power curves generated by the kernel density and mlpt methods were parallel between a separation of 2 and 4 standard deviations, the kernel method being slightly less powerful (Figure 2). When the Weibull maximum likelihood method was tested in a similar fashion this appeared to be more conservative than both the mlpt and kernel density methods (Figure 3). This became more apparent as the separation between the components of the sampled distribution increased.

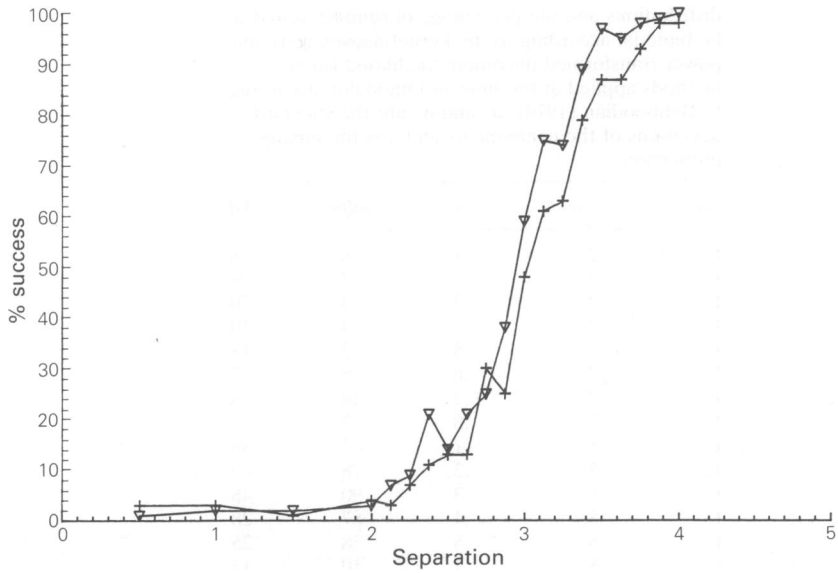


Figure 2 The percentage of samples from a mixture distribution found to be bimodal by the mlpt (▽) and kernel density (+) methods in relation to the separation (expressed as standard deviation units) between the means of the components of the mixture distribution from which the samples were taken.

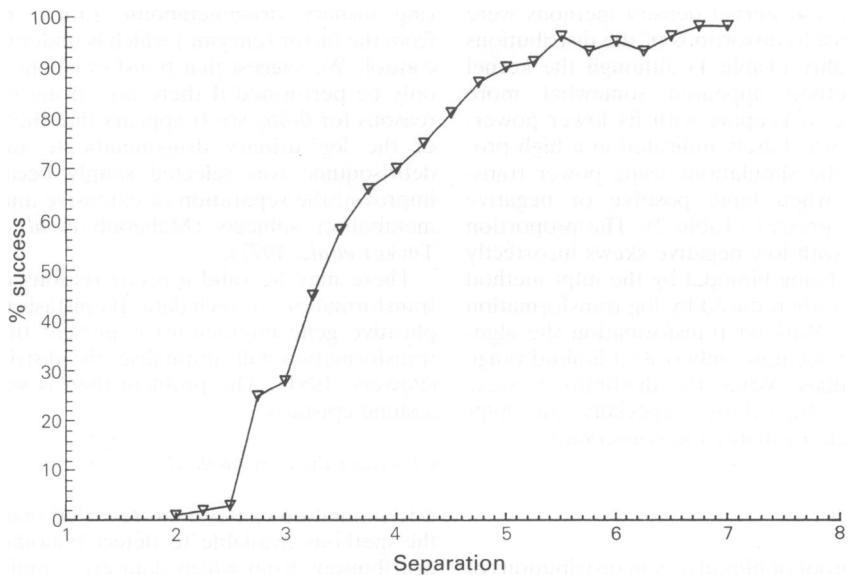


Figure 3 The percentage of samples from a mixture distribution found to be bimodal by the maximum likelihood method using a Weibull distribution in relation to the separation (expressed as standard deviation units) between the means of the components of the mixture distribution from which the samples were taken.

Table 1 Parameter values for theoretical mixture distributions and the percentage of samples found to be bimodal according to the kernel density (kd) and power transformed maximum likelihood (mlpt) methods applied at the limit of bimodality according to Behboodian (1970). σ_1 and σ_2 are the standard deviations of the components and γ is the mixing proportion.

σ_1	σ_2	γ	*mlpt*	*kd*
1.	2.	.1	8	38
1.	2.	.2	7	23
1.	2.	.3	5	20
1.	2.	.4	4	10
1.	2.	.5	2	13
1.	2.	.6	5	9
1.	2.	.7	10	8
1.	2.	.8	7	5
1.	5.	.1	72	88
1.	5.	.2	78	69
1.	5.	.3	90	48
1.	5.	.4	87	28
1.	5.	.5	58	25
1.	5.	.7	10	17
1.	5.	.8	5	11
1.	5.	.9	1	9
.691	.691	.1	5	20
.7685	.7685	.2	2	18
.838	.838	.3	4	23
.9119	.9119	.4	6	17
1.	1.	.5	9	4

The mlpt and kernel density methods were both sensitive to distortions of the distributions from normality (Table 1) although the kernel density method appeared somewhat more conservative in keeping with its lower power. Bimodality was falsely indicated in a high proportion of the simulations using power transformations when large positive or negative skews were present (Table 2). The proportion of samples with low negative skews incorrectly assigned as being bimodal by the mlpt method was significantly reduced by log transformation of the data. Without transformation the algorithms were accurate only over a limited range of skew values. When the distributions were normal the algorithms, especially the mlpt method, were probably too conservative.

Discussion

Statistical proof of bimodality in distributions of known form is difficult. The problem is compounded further when the genetic basis of the distribution of characteristics within a population is unknown. Additional difficulties are introduced by the use of experimental indices (log urinary drug/metabolite ratios) remote from the factor (enzyme) which is under genetic control. We suggest that transformations should only be performed if there are strong *a priori* reasons for doing so. It appears that the choice of the log urinary drug/metabolite ratio of debrisoquine was selected simply because it improved the separation of extensive and poor metaboliser subjects (Mahgoub *et al.*, 1977; Tucker *et al.*, 1977).

There may be valid *a priori* reasons for log transformations of such data. If epistasis (multiplicative gene interaction) is present then log transformation will 'normalise' the distribution (Powers, 1951). The problem then is when to assume epistasis.

Choosing the best method

The simulation exercise has revealed that all of the methods available to detect bimodality in distributions from which data are sampled are not particularly powerful and are sensitive to skew within the distribution. Although the non-parametric kernel density method makes very few assumptions about the form of the distribution it is nevertheless sensitive to certain

Table 2 The percentage of simulations found to be bimodal by the kernel density [kd] and mlpt methods when samples were drawn from a single normal distribution and then skewed using the power transform of Maclean *et al.* (1976) with different values of the power factor TR. The results are also given for the effect of log transforming the data prior to submission to the algorithms (Lmlpt and Lkd).

		% simulations found to be bimodal			
TR	*skew*	*mlpt*	*kd*	*Lmlpt*	*Lkd*
-2.5	-31.5	98	97	90	100
-2.0	-31.5	99	95	82	100
-1.5	-20.0	97	94	78	100
-1.0	-2.28	100	93	81	100
-0.5	-2.01	93	79	67	100
0.0	-0.99	84	72	17	100
0.5	-0.46	38	35	11	100
1.0	-0.01	6	2	7	100
1.5	0.39	0	3	1	100
2.0	0.65	0	16	9	99
2.5	1.16	0	26	37	95
3.0	1.45	16	39	70	84
3.5	1.90	60	51	82	66
4.0	2.21	83	71	75	59
4.5	2.65	93	72	78	46
5.0	3.18	97	78	72	39
5.5	4.36	99	81	63	39
6.0	4.19	99	83	73	28
6.5	5.70	100	94	67	28
7.0	4.90	100	90	68	35
7.5	4.90	100	92	63	28
8.0	8.37	100	95	61	29
8.5	10.80	100	93	54	43
9.0	8.40	99	95	54	43
9.5	20.10	100	96	49	50
10.0	30.6	100	97	41	37

distribution profiles. It is also scale sensitive and relatively expensive in computer time. The mlpt method appears to be as robust and slightly more sensitive. With certain data sets the results of the two methods may diverge because they are testing slightly different hypotheses. The kernel density method tests a hypothesis based on the number of modes in the parent distribution whereas the mlpt method tests whether the sample is best fitted by one or more distributions independent of the modality of the distribution from which the sample is drawn. We believe that because the genetic mechanisms by which variability of enzyme activity is produced are unknown, assumptions as to the form of the distribution in the population are unwarranted and likely to lead to the spurious over-diagnosis of polymorphism. Furthermore as a general point we suggest that the diagnosis of a polymorphism should only be made when a distribution is shown to be bi- or multimodal. Diagnosis of the phenomenon from *in vivo* data is virtually impossible if these data are merely non-normally distributed.

To prevent skew within the data giving rise to a false indication of bimodality it might be sensible to check the skew of each data set prior to presentation to the algorithms.

If frequency histograms of experimental data or transformations of such data indicate that the components are normal the mlpt method is preferred for establishing bimodality since it requires less computer time. However, when the form of the underlying distributions is not known the non-parametric kernel density method may be superior. Both approaches may fail if the component distributions have markedly different variances or the summed distribution is highly skewed, unless a variance stabilising transformation is applied.

Computer Hardware for this study was kindly supported by donations from the University of Sheffield Research Fund and the Hallamshire Therapeutics Trust.

References

Behboodian, J. (1970). On the modes of a mixture of two normal distributions. *Technometrics*, **12**, 131–139.

Bhattacharya, C. G. (1967). A simple method of resolution of a distribution into Gaussian components. *Biometrics*, **23**, 115–135.

Binder, D. A. (1978). Bayesian cluster analysis. *Biometrika*, **65**, 31–38.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc.*, Series B, **39**, 1–38.

Evans, D. P., Harmer, D., Downham, D. Y., Whibley, E. J., Idle, J. R., Ritchie, J. & Smith, R. L. (1983). The genetic control of sparteine and debrisoquine metabolism in man with new methods of analysing bimodal distributions. *J. med. Genet.*, **20**, 321–329.

Everitt, B. S. (1981). A Monte Carlo investigation of the likelihood ratio test for the number of components in a mixture of normal distributions. *Multivar. Behav. Res.*, **16**, 171–180.

Fletcher, R. (1980). *Practical methods of optimization*. New York: Wiley.

Jack, D. B. (1983). Statistical analysis of polymorphic drug metabolism data using the Rosin Rammler Sperling Weibull distribution. *Eur. J. clin. Pharmac.*, **25**, 443–448.

Jackson, P. R. (1988). *Theoretical and experimental studies of variability in drug disposition*. Ph.D. Thesis, University of Sheffield.

Jackson, P. R., Tucker, G. T. & Woods, H. F. (1989). Testing for bimodality in frequency distributions of data suggesting polymorphisms of drug metabolism—histograms and probit plots. *Br. J. clin. Pharmac.*, **28**, 647–653.

Kao, J. H. K. (1959). A graphical estimation of mixed Weibull parameters in life-testing of electron tubes. *Technometrics*, **1**, 389–407.

Knuth, D. E. (1981). *The art of computer programming*. Vol. 2 Seminumerical algorithms. 2nd ed. Reading, Mass: Addison-Wesley.

Kolmogoroff, A. (1941). Confidence limits for an unknown distribution function. *Ann. Math. Stat.*, **12**, 461–463.

Lindgren, B. W. (1968). *Statistical theory*. London: Collier Macmillan.

Maclean, C. J., Morton, N. E., Elston, R. C. & Yee, S. (1976). Skewness in commingled distributions. *Biometrics*, **32**, 695–699.

Mahgoub, A., Idle, J. R., Dring, L. G., Lancaster, R. & Smith, R. L. (1977). Polymorphic hydroxylation of debrisoquine in man. *Lancet*, **ii**, 584–586.

McGourty, J. C., Silas, J. H., Lennard, M. S., Tucker, G. T. & Woods, H. F. (1985). Metoprolol metabolism and debrisoquine oxidation polymorphism—population and family studies. *Br. J. clin. Pharmac.*, **20**, 555–566.

Murphy, E. A. & Bolling, D. R. (1967). Testing of single locus hypotheses where there is incomplete separation of the phenotypes. *Am. J. Hum. Gen.*, **19**, 322–334.

Numerical Algorithm Group (1984). Fortran Routine Library. Oxford: N.A.G.

Nelder, J. A. & Mead, R. (1965). A simplex method for function minimisation. *Computer J.*, **7**, 308–313.

Powers, L. (1951). Gene analysis by the partitioning method when interactions of genes are involved. *Botanical Gazette*, **113**, 1–23.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1986). *Numerical recipes. The art of scientific computing*. Cambridge: Cambridge University Press.

Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *J. Roy. stat. Soc.*, **43**, 97–99.

Tucker, G. T., Silas, J. H., Iyun, A. O., Lennard, M. S. & Smith, A. J. (1977). Polymorphic hydroxylation of debrisoquine. *Lancet*, **ii**, 718.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stats.*, **9**, 60–62.

Wolfe, J. H. (1971). *A Monte Carlo study of the sampling distribution of the likelihood ratio for mixtures of multinomial distributions*. Technical Bulletin STB 72-2. Naval Personnel and Training Research Lab., San Diego, California.

(Received 16 March 1989, accepted 7 August 1989)