ACTA STEREOL 1983; 2/1: 187-191
SHORT NOTE

SOME EMPIRICAL FUNCTIONS FOR USE IN THE PARAMETRIC MODELLING OF STEREOLOGICAL SIZE DISTRIBUTIONS

Laurence E. Scales and Vyvyan Howard(*)

Departments of Computer Science and Anatomy(*)

University of Liverpool, Liverpool, England

ABSTRACT

This paper is concerned with some aspects of the unfolding of sphere size distributions by parametric methods, in particular with a comparison of the performance of three types of empirical distribution upon experimental data obtained from neurological tissue. The results suggest that the log normal function is the most useful in this area. A comparison of the estimated errors involved in the parametric and distribution free unfolding of unimodal systems is attempted, and it is concluded that parametric methods can be superior when the fit of the empirical function is good or the sample size is rather small.

INTRODUCTION

Parametric methods for unfolding sphere size distributions are usually only considered for isolating the components of bimodal or polymodal systems. However, the first part of this paper is concerned with the question: For unimodal systems, can parametric methods in practice ever give lower estimated errors in any quantities of interest than distribution free methods? Simulation work based upon actual experimental data tempts us to answer in the affirmative, but with some reservations to be discussed.

Central to the application of parametric methods is the choice of an empirical analytical distribution whose general shape approximates to the unfolded (underlying) distribution corresponding to the experimental data, but in which there are certain unknown parameters to be determined. The analytical function will normally have to be substantially skewed for cell biological systems at least.

We have considered three skewed distributions, log normal, gamma and Weibull, and have investigated their aptitude for fitting a variety of experimental profile size distributions.

The distribution free analysis used is Cruz-Orive's variation of the traditional Schwartz-Saltykov technique. The parametric approach employs nonlinear least squares fitting of the relevant theoretical profile distribution to the experimental data. In both cases, section thickness effects were included. Details of the theory and further references will appear in a companion paper (Scales et al., 1983).

COMPARATIVE ERROR ESTIMATES FOR UNIMODAL SYSTEMS

The procedure adopted for this part of the work was to take an experimental profile distribution and unfold it by the distribution free method. The resulting distribution was used to generate synthetic profile distributions of various sample sizes by computer simulation. These were then unfolded by parametric and distribution free techniques and the estimated errors compared. The sample sizes taken were 250, 500, 1000 and 2000, chosen to represent what was thought to be the range normally encountered in practice. At each sample size, five synthetic profile distributions were generated. The derived quantity of interest to us is the numerical density N_V, and the average values of the estimated relative standard error in N_V, $\hat{\sigma}(N_V)/N_V$ are considered here.

The whole procedure was repeated for two experimental systems obtained from neurological tissue, one of which (8 degrees of freedom) was thought to be well represented by a log normal model, and the other (4 degrees of freedom) not so well (Scales et al., 1983). For the former, the p-value obtained from the analysis via χ^2 of the goodness of fit of the log normal model to the experimental data was 0.13 (experience has shown that a p-value of this size must be regarded as a reasonably good fit for these experimental systems) while for the latter the p-value was 0.01. For the purposes of our numerical experiments, these distributions were regarded as population distributions, one of which happens to be almost log normal and the other not so. Thus our simulations deal with artificial populations, not the true populations of actual nerve cells. The results are shown in tables 1 and 2. First note that the errors for the parametric approach stay substantially constant because the distribution from which the samples were drawn is not log normal, while those for the distribution free approach decrease as the sample size increases in accordance with the decreasing variance in the histogram class heights. In the 'good fit' case the errors for the

parametric approach are always substantially the smaller especially when the sample size is small. In the 'not so good fit' case the parametric approach fares less well, but still appears superior when the sample size is small. We are therefore tempted to say: Use a parametric method not only when the fit is good but also when the sample size is small.

Table 1. Estimated relative standard errors in N_V as a function of sample size for a 'good fit' case.

sample size	$\hat{\sigma}(N_V)/N_V$	
	parametric	distribution free
250	0.018	0.061
500	0.016	0.047
1000	0.016	0.033
2000	0.017	0.024

Table 2. Estimated relative standard errors in N_V as a function of sample size for a 'not so good fit' case.

sample size	$\hat{\sigma}(N_V)/N_V$	
	parametric	distribution free
250	0.026	0.034
500	0.025	0.023
1000	0.031	0.017
2000	0.029	0.012

EMPIRICAL DISTRIBUTIONS FOR PARAMETRIC UNFOLDING

Three empirical distributions, log normal, gamma and Weibull, were considered for parametric unfolding, each of which has four parameters including a coefficient and a cut-off (Scales et al., 1983). For polymodal systems, each component is modelled by its own

Table 3. Comparative goodness of fit, L = log normal, G = gamma, W = Weibull, d_f = degrees of freedom, p_c = p-value from χ^2 analysis.

data set	model	sample size	d_f	modality	P_c
1	L	1339	8	1	0.126
	G				0.000
	W				0.000
2	L	570	4	1	0.013
	G				0.000
	W				0.000
3	L	2521	7	2	0.026
	G				0.050
4	L	1166	3	2	0.193
	G				0.002
5	L	1108	12	2	0.387
	G				0.226
6	L	1054	10	2	0.110
	G				0.096
7	L	627	7	2	0.282
	G				0.264
8	L	1098	10	2	0.114
	G				0.094
9	L	660	7	2	0.041
	G				0.005

function with its own set of four parameters. The individual coefficients therefore do not restrict normalisation to the original sample size and consequently can to some extent allow for missing profiles. Data in the histogram bins corresponding to smaller profile diameters is usually discarded, being deficient through missing profile effects. The model is freely allowed to predict the profile distribution in this region. Effects of missing profiles on the shape of the profile distribution for the larger diameters can not, however, be catered for in our present approach. The cut-off parameter is thought to be essential for all biological systems and there are biological reasons why this may be so.

These functions were tried out on nine neurological systems, some unimodal, some bimodal, with a variety of sample sizes, degrees of freedom and goodnesses of fit. The results are summarised in table 3. In most cases the log normal function gave the best p-values and in addition had the best convergence properties in the sense that a local minimum of the fitting problem was rapidly reached from a remote starting point. The Weibull function fared worst in this respect, which led us to abandon it after doing the unimodal systems.

REFERENCES

Scales L E, Howard C V, Green J R and Lynch R. (1983) Some observations on the parametric unfolding of size distributions. J Microsc. In press.