Application of general finite mixture models to reliability data using likelihood estimation.

Student Oral Presentation

E. Andries[1], K. Croes[2], L. De Schepper[1], G. Molenberghs[1].

[1]Limburgs University Centre
Institute for Materials Research (IMO), Materials Physics Division
Universitaire Campus,Wetenschapspark 1
B-3590 Diepenbeek (Belgium)
Tel: 32-(0)11-26 88 16
Fax: 32-(0)11-26 88 99
E-mail: ellen.andries@luc.ac.be
[2]XPEQT,Transportstraat 1,B-3980 Tessenderlo (Belgium)

Keywords: general finite mixture model, likelihood estimation, spurious maxima, reliability data.

1. Introduction

A lot of today's reliability data obtained from experiments with micro-electronic components give evidence of bi or even multi modal failure data. This means that a component can fail due to more than one failure mechanism. Although reliability engineers know mostly whether there is more than one mechanism involved, at the end of the experiment it is or too difficult or too expensive (both money and time) to recover the specific failure reason of each device under study. A large part of these multi modal failure data can be modelled by means of a finite mixture model. In particular, mixtures with mixing over all parameters are of interest since, due to the nature of much reliability data, a common shape (or variance) parameter for the component densities cannot a priori be assumed.

A general M-component finite mixture has the following density:

$$f_M(x|\underline{\theta}) = \sum_{m=1}^{M} \pi_m f(x|\mu_m, \sigma_m) \quad (1)$$

with $\sum_{m=1}^{M} \pi_m = 1$, $f(x|\mu_m, \sigma_m)$ the density of a 2-parameter distribution, μ_m a scale and σ_m a shape parameter. The problem, however, with this kind of mixtures is that a maximum likelihood (ML) estimate, defined as the global maximum of the likelihood, does not exist due to singularities in the likelihood. Despite this non-existence, the likelihood does have a local maximum with, very importantly, good statistical properties. Moreover, for normal mixtures this maximum corresponds to the largest local maximum. Nevertheless, not everyone agree on the last fact as they argue that a so-called spurious maximum can be chosen as an estimate.

The aim of this paper is first to discuss shortly this likelihood theory, which is far from new, acknowledged by some authors, but still rarely applied. Second to tackle the problem of spurious maxima and third to demonstrate the method on two sets of reliability data.

2. Likelihood estimation

It is well known that the likelihood function for a mixture with density (1) is unbounded at some points on the edge of the parameter space and therefore has no global maxima [1]. As such, the ML method cannot be applied. Nevertheless, both empirical and theoretical evidence proved for finite normal mixtures the existence of some local maximum of the likelihood function with good statistical properties, i.e. consistent, asymptotically normal and efficient [2,3].

This can be explained through the fact that different conditions determine the existence of a consistent global and a consistent local maximum of the likelihood. Namely, under the conditions of Cramér [4], the likelihood equations (LEQ) have a consistent, asymptotically normal and efficient solution. Further, with probability tending to one as the sample size tends to infinity, this root corresponds to a local maximum and is essentially unique. On the other hand, there are the conditions of Wald [5], which assure the consistency of the classical MLE. These assumptions, however, are totally different and more demanding compared to Cramér's conditions.

Consequently if a density satisfies both conditions the root corresponding to the global maximum is consistent (e.g. the normal or Weibull distribution, a mixture with a common shape parameter). If only Cramér's conditions hold some local maximum is consistent (e.g. a general M-component normal or Weibull mixture). Importantly, whether we either work with a mixture with a common shape parameter or with unequal shape parameters, in essence the same kind of estimate is obtained from the LEQ, in spite of the convention of terminology to only call the first an MLE. The latter will be referred to as a likelihood estimate (LE).

However, the problem is not entirely solved yet since the likelihood function has multiple roots and it is not specified which local maximum is the proper one. It can now be proven that for certain general finite mixtures, such as the (log)normal or weibull mixture, the largest local maximum of the likelihood function corresponds with those well-behaved estimates. This gives a criterion similar to the one for ML estimation. But not everyone agree on this, as McLachlan et al. [7], among others, claim that a so-called spurious maximum could then be chosen as LE.

3. Spurious maxima

What is meant with a spurious maximum? No unambiguous definition exists yet, but most of the time it is characterized by the fact that one of the proportions of the corresponding estimate is small or that one of the component densities has a small standard deviation, …. They cause problems when one of them is the largest local maximum. Some authors suggest to first remove all solutions corresponding to such maxima and then to choose among the remaining roots the solution with the largest likelihood as LE. Although these maxima should be considered with care, this procedure is dangerous, highly subjective and we do not recommend it.

The point is that these spurious maxima are not only related to the largest likelihood criterion and the finite mixture case. They exist as soon as Cramérs conditions hold and as the LEQ have multiple roots; irrespective of the fact whether we search for a local or global maximum. The reason why they sometimes appear as the largest maximum is primarily due to the ambiguity in the statement of a consistent root and related with sample size. Indeed, consistency is a limiting property. Therefore, it is possible to obtain an improper estimate with the likelihood or ML method if the sample size n is too small, although these methods construct a consistent sequence. We define a spurious maximum as each maximum of the likelihood that is not closest to the true values, with *closest* defined by some distance. In this way they can take any form.

How can we then obtain a proper estimate from the LEQ in case of multiple roots? First, use always a consistent procedure, i.e. a method leading to a consistent sequence of estimates (e.g. the largest local or global criterion). Second, choose the sample size large enough. If the latter is not possible, one should take into consideration another method or look whether there is relevant information about the possible true values.

4. Sample size

Since spurious maxima only become a problem when the sample is too small, the question is then how large the sample should be in case of general finite mixtures. This depends clearly on the problem at hand: e.g. for well-separated mixtures this need not be large, but it can be huge for the poor separation case. Based on the likelihood value of distinct spurious maxima, i.e. maxima for which one of the component densities corresponds to no more than a few data values (upon the condition that these data values are not clearly separated from the others), we derived a rule of thumb. Namely, simulations indicated that when n is too small, generally at least one of these maxima has a highest (local) likelihood value. This can be used to recognize too small datasets.

N	$\mu_1 = 0$	$\sigma_1 = 0.5$	$\mu_2 = 3$	$\sigma_2 = 1$	$\pi_1 = 0.2$	Likelihood
50	*-1.029*	*0.00175*	*2.493*	*1.470*	*0.0399*	*-85.122*
	1.387	1.444	3.627	0.553	0.569	-88.398
	-0.249	**0.694**	**2.996**	**0.987**	**0.198**	**-88.571**
120	**-0.255**	**0.627**	**2.895**	**0.984**	**0.128**	**-200.921**
	3.246	0.685	1.693	1.543	0.514	-202.120
	2.874	0.0000541	2.485	1.425	0.0166	-202.628

Table 1: Some important local maxima of two simulated datasets. The estimates in bold correspond to the maximum nearest to the true values.

As an example, Table 1 gives some local maxima of the likelihood of two simulated datasets from a two-component normal mixture: one of size 50 and the other based on the first but with 70 extra values simulated. As seen, for the smallest dataset, the highest likelihood value corresponds to a distinct spurious maximum (the proportion of the first component is less than 2/50), while for the larger dataset the proper maximum has the highest likelihood value and the distinct spurious maxima have a lower value. Note also that for the smallest dataset the proper maximum is not only preceded by distinct spurious maxima but also by a spurious maximum with reasonably looking estimates.

4. Examples

We demonstrate the likelihood estimation method on tow datasets obtained from experiments carried out at IMO. The experiments consisted of accelerating the failure mechanisms of a micro-electronic component by means of increasing certain stress factors (such as temperature, current, …). The failure time of each device under test was measured (i.e. the time-point at which it was not able any more to perform its function). Main interest is in the estimation of the failure-time distribution and more particularly in the prediction of low percentiles. For the components under study it is known that there could be a second failure mechanism, leading to bimodal failure data.

4.1 Example 1

The failure times of 125 commercial metal film resistors, stressed at a temperature of 165°C, were measured. Figure 1 shows a lognormal QQ-plot of the data. Generally, when there is only one failure mode, the failure times for this type of component are lognormally distributed. Since the data suggest two failure modes, a two-component lognormal will be estimated. The local maxima of the likelihood function are searched for and the most important ones are indicated in Table 2 (in decreasing order of likelihood value). Note that the parameters are the means and standard deviations of the corresponding two-component normal mixture of the log failure times. As noticed, the largest local maximum is not a distinct spurious maximum and its value is much larger than the value of the first distinct spurious maximum. So, there is no reason to mistrust the largest local maximum. The fitted distribution is shown in Figure 1. In this case, one can now proceed as in case of ML estimation and carry out tests, construct confidence intervals, … in the usual way.

Example	$\hat{\mu}_1$	$\hat{\sigma}_1$	$\hat{\mu}_2$	$\hat{\sigma}_2$	$\hat{\pi}_1$	Likelihood
1	6.164	0.236	7.022	0.251	0.286	-66.100
	6.745	0.0000355	6.777	0.463	0.0159	-72.545
	6.864	0.0000436	6.775	0.463	0.0158	-72.918
2	5.329	0.0000684	5.009	0.429	0.0292	-40.444
	5.071	0.0000783	5.020	0.432	0.0291	-41.375
	4.086	0.000730	5.041	0.413	0.0286	-43.137
	4.220	0.0518	5.080	0.389	0.105	-44.575
	4.697	0.340	5.374	0.166	0.641	-45.427

Table 2: Some local maxima for the datasets of example 1 and 2. The first line of each example corresponds to the largest local maximum. The estimated parameters are the location and shape/variance parameters of the log failure times.

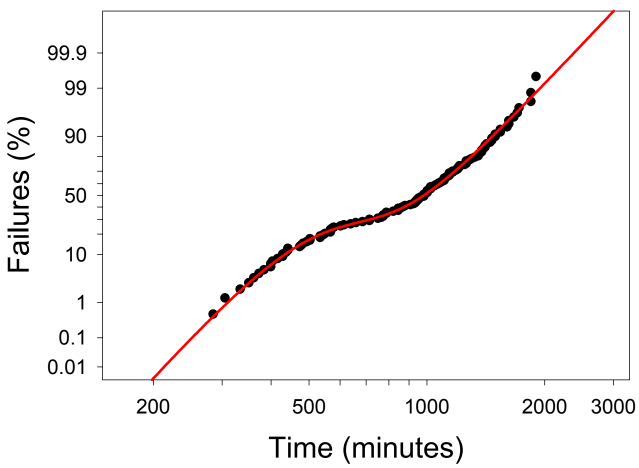

Figure 1: Lognormal probability plot of the failure times of example 1 and estimated mixture distribution.

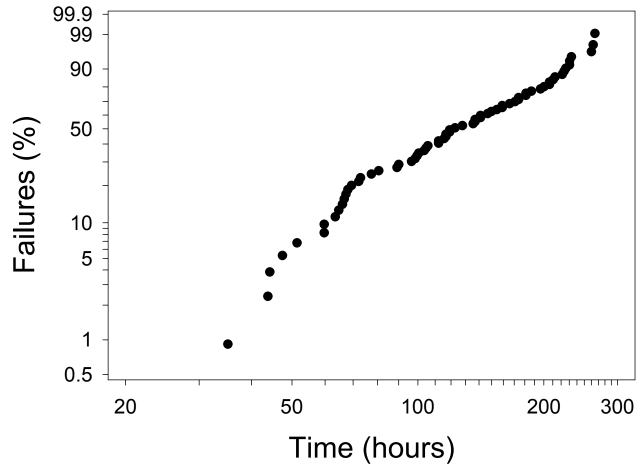

Figure 2: Weibull probability plot of the failure times of example 2.

4.2 Example 2

Interconnects were stressed at 80°C and 0.75 MA/cm^2. All 68 devices under test failed. The Weibull QQ-plot of the failure times is shown in Figure 2. Previous experiments indicated that there could be two failure mechanisms. So, a two-component Weibull mixture is used to fit the data. The likelihood is scanned for local maxima and some of them are tabulated in Table 2. In contrast to the first example, the largest local maximum is now a distinct spurious maximum. Moreover, between the largest local maximum and the first not distinct spurious maximum, several distinct spurious maxima are situated. Although the last two maxima in the table correspond to reasonable estimates, it is dangerous to choose one of these two as the LE. Indeed, depending on the chosen maximum other inference results are obtained, which could lead to wrong reliability predictions and conclusions. If there are truly two failure modes, this is not clearly seen yet. Consequently, more data is needed or other techniques have to be applied.

5. Conclusions

Despite the nonexistence of the MLE for general finite mixtures, there exists a root of the LEQ with good statistical properties. It is the same kind of estimate as the MLE, called the LE and corresponds for a lot of cases to the largest local maximum of the likelihood.

The appearance of spurious maxima is inherently linked to the presence of multiple roots in the LEQ and independent of the fact whether one search for the largest local or global maximum. They only cause problems when the sample size is too small, since one of them then becomes the LE or MLE.

When the likelihood function is dominated by distinct spurious maxima, the sample is most likely too small and none of the roots of the LEQ can be trusted.

References

[1] Kiefer J. and Wolfowitz J., 1956, Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters, *Annals of mathematical statistics*, 27,888-906.

[2] Quandt R.E., 1972, A new approach to estimating switching regressions, *Journal of the American statistical association*, 76, 306-310.

[3] Kiefer N.M., 1978, Discrete parameter variation: efficient estimation of a switching regression model, Econometrica, 46(2), 427-433.

[4] Cramér H., 1946, Mathematical methods of statistics (ch. 33), Princeton University Press.

[5] Wald A., 1949, Note on the consistency of the maximum likelihood estimate, *Annals of mathematical statistics*, 20, 595-601

[6]McLachlan G. and Peel D., 2000, Finite mixture models, New York: John Wiley & Sons.