ON THE COMPUTATION OF THE MAXIMUM LIKELIHOOD ESTIMATES OF THE PARAMETERS IN A MIXTURE MODEL

G. Nanjundan*

ABSTRACT

A social group may consist of sterile and fertile couples where sterile couples cannot reproduce. When the number of children for a fertile couple is distributed according to a Poisson distribution, the probability distribution of the number of children per couple in the social group is a mixture of a distribution singular at zero and a Poisson distribution. The estimation of the parameters in the mixture distribution is considered in this paper. Since the maximum likelihood (ML) method does not provide estimates in closed forms, it is proposed to obtain the estimates using the EM algorithm. A stepwise procedure for computing the estimates is presented. A numerical study is carried out to compare these estimates with the conditional ML estimates determined using Newton-Raphson iterative procedure.

Key Words: *Poisson distribution, singular distribution, mixture distribution, maximum likelihood (ML) estimation, conditional maximum likelihood estimation, EM algorithm.*

* *Department of Statistics, Bangalore University, Bangalore 560 056, India. e-mail: nanzu@rediffmail.com*

1. Introduction

Suppose that we are interested in the distribution of the number of children (X) per couple in a social group under study. Only fertile couples have children and sterile couples cannot re-produce. Let us not consider the case of a couple with the male partner being sterile and they get an offspring by artificial insemination as well the case where the female partner is sterile and they get a child through a surrogate mother. In other words, we are interested in the distribution of the number of children given birth by couples on their own in the group under the study. If the proportion of the sterile couples in the group is φ and the number of children of a fertile couple has a Poisson distribution with mean θ then the probability mass function of X is given by

$$P(X = x \mid \theta, \varphi) = p(x) = \begin{cases} \varphi + (1-\varphi)e^{-\theta}, & \text{if } x = 0 \\ (1-\varphi)\dfrac{e^{-\theta}\theta^x}{x!}, & \text{if } x = 1, 2, 3, \ldots \end{cases} \tag{1.1}$$

$$= \varphi p_0(x) + (1-\varphi)p_1(x \mid \theta),$$

where $p_0(x) = \begin{cases} 1, & \text{if } x = 0 \\ 0, & \text{if } x \neq 0 \end{cases}$ and $p_1(x \mid \theta) = \dfrac{e^{-\theta}\theta^x}{x!}$, if $x = 0, 1, 2, \ldots$; $\theta > 0$. Thus,

the distribution of X is a mixture of a singular distribution at 0 and a Poisson distribution with mean θ. Mixture distributions of this type have recently gained importance in the context of zero inflated regression models. See Lambert (1992) and Hall (2000).

2. Conditional maximum likelihood estimation

If $\underline{X} = (X_1, X_2, \ldots, X_n)$ is a random sample on X, then the likelihood function is given by

$$L(\theta, \varphi \mid \underline{x}) = \prod_{j=1}^{n} P(X_j = x_j \mid \theta, \varphi).$$

It can be written as

$$L(\theta, \varphi \mid \underline{x}) = \prod_{j=1}^{n} \left\{ \varphi + (1-\varphi)e^{-\theta} \right\}^{1-a_j} \left\{ (1-\varphi)\frac{e^{-\theta}\theta^{x_j}}{x_j!} \right\}^{a_j}, \theta > 0; \, 0 < \varphi < 1 \tag{2.1}$$

where $a_j = \begin{cases} 0, \text{if } x_j = 0 \\ 1, \text{if } x_j \geq 1 \end{cases}$. The above likelihood function does not yield closed form expressions for the maximum likelihood estimates (MLEs) of θ and φ. Hence, a numerical procedure has to be employed to maximize the likelihood function for a given $\underline{x}$. Yip (1988, 1991) and Kale (1998) have discussed the estimation of parameters in such mixture models. Motivated by a criterion of Cox (1958), Yip (1988) has obtained the conditional maximum likelihood estimator (CMLE) of θ as a solution of the equation

$$\tilde{x} = \frac{\theta}{1 - e^{-\theta}}, \tag{2.2}$$

where $\tilde{x} = \dfrac{\sum\limits_{j=1}^{n} a_j x_j}{\sum\limits_{j=1}^{n} a_j}$ and has suggested an estimator of given by the equation

$$(1 - \varphi)(1 - e^{-\tilde{\theta}}) = \frac{a}{n}, \tag{2.3}$$

where $\tilde{\theta}$ is the CMLE of θ and $a = \sum\limits_{j=1}^{n} a_j$. He has also computed the loss of information due to the conditional likelihood estimation for various combinations of θ and φ. Further, Yip (1991) has extended the conditional likelihood approach when p_1 specified in (1.1) is the probability mass function of binomial and negative binomial distributions.

Kale (1998) has discussed, in the spirit of Godambe (1976), the optimal estimating equation for θ ignoring φ when p_1 is the probability mass function of a general power series distribution. When p_1 is the probability mass function of a Poisson distribution as specified in (1.1), the optimal estimating equation for θ obtained by Kale (1998) ignoring φ turns out to be the same as given by (2.2).

It is worth to note that both Yip (1988, 1991) and Kale (1998) have treated φ as a nuisance parameter. But, from a practical point of view, estimating the proportion of sterile couples is also essential and φ cannot be treated as a nuisance parameter. Also, the equation in (2.2) does not yield a closed form expression for the estimate of θ and hence a numerical procedure has to be employed with an appropriate initial value to obtain the estimate of θ. But it is definitely much easier to deal with (2.2) than the likelihood function specified in (2.1).

It can be observed that the likelihood function in (2.1) can also be expressed as,

$$L(\theta, \varphi | \underline{x}) = \{\varphi + (1-\varphi)e^{-\theta}\}^{n_0} (1-\varphi)^{n-n_0} \prod_{j:x_j>0} \{\frac{e^{-\theta}\theta^{x_j}}{x_j!}\}, \quad \theta>0; \ 0<\varphi<1 \qquad (2.4)$$

where n_0 is the number of observations that are equal to 0. Hence, note that

$$a = \sum_{j=1}^{n} a_j = n - n_0 \ .$$

3. Maximum Likelihood Estimation Using the EM Algorithm

When the likelihood functions have complicated structures and their maximi- zation by numerical methods is difficult, the MLEs of the parameters can be computed by the Expectation Maximization (EM) algorithm with ease. It is popular and remarkably simple. It is an iterative procedure and there are two steps in each of the iterations, namely the Expectation Step (E-step) and the Maximization Step (M-step. This algorithm was developed by Dempster, Laird, and Rubin (1977) who synthesized an earlier formulation in many particular cases and gave a general method of finding the MLEs in a variety of problems. Since then the EM algorithm has been applied to a variety of statistical problems such as resolution of mixtures, multi-way contingency tables, variance components estimation, and factor analysis. It has also found applications in specialized areas like genetics, medical imaging, and neural networks. For a detailed discussion, see McLachlan and Krishnan (1997) and Krishnan (2004).

Nanjundan (2006) has obtained the E- and the M – steps by rewriting the likelihood function in (2.1) so as to accommodate missing data.

Let $Z_j = \begin{cases} 1, & \text{if the } j-\text{th sampled couple is fertile} \\ 0, \text{otherwise} \end{cases}$

Then, we have $P(Z_j = 1) = 1 - \varphi = 1 - P(Z_j = 0), \; j = 1, 2, \ldots, n$.

Suppose that $\underline{X} = (X_1, X_2, \ldots, X_n)$ is the observed sample on X. Then, $((X_1, Z_1), (X_2, Z_2), \ldots, (X_n, Z_n))$, becomes the complete sample when $(X_1, X_2, \ldots, X_n)$ is augmented with $(Z_1, Z_2, \ldots, Z_n)$. If $X_j > 0$, then $Z_j = 1$ and if $X_j = 0$, then $Z_j = 0$ or 1. In other words, for $X_j = 0$, we have no information on Z_j. Hence, $\{Z_j : X_j = 0\}$ can be treated as the missing data.

The likelihood function of the complete data is given by

$$L_c(\theta, \varphi | \underline{x}, \underline{u}) = \prod_{j=1}^{n} \varphi^{1-u_j} \left[(1-\varphi) \frac{e^{-\theta} \theta^{x_j}}{x_j!} \right]^{u_j},$$

where $u_j = 1$, if $x_j > 0$ and $u_j = Z_j$ if $x_j = 0$.

In the E – step, the expectation of the likelihood function of the complete data is taken and $E(Z_j)$ is replaced by the conditional expectation $E(Z_j | \theta_0, \varphi_0, X_j)$, where θ_0 and φ_0 are respectively the initial estimates of θ and φ. In the M-step, $E(L_c(\theta, \varphi | \underline{x}, \underline{u}))$ is maximized with respect to θ and φ. If θ_1 and φ_1 are the values of θ and φ which maximize $E(L_c(\theta, \varphi | \underline{x}, \underline{u}))$, then the E-step is repeated using θ_1 and φ_1.

The computational details of these steps can be summarized as follows.

a) Choose the initial estimates θ_0 and φ_0.

b) Compute $w = \dfrac{(1-\varphi_0)e^{-\theta_0}}{\varphi_0 + (1-\varphi_0)e^{-\theta_0}}$.

61

c) Using the realization $(x_1, x_2, \ldots, x_n)$ of the observed sample, compute

$$\theta_1 = \frac{\sum_{j=1}^{n} x_j}{n_g + n_0 w} \quad \text{and} \quad \phi_1 = \frac{n_0 (1-w)}{n},$$

where n_g is the number of observations greater than zero.

d) Repeat the steps b) and c) by fixing $\theta_0 = \theta_1$ and $\phi_0 = \phi_1$ until a required precision is attained.

A reasonable initial estimate of ϕ is n_0 / n and the mean of the observed sample can be taken as the initial estimate of θ. If $\{\theta_n\}_{n=1}^{\infty}$ and $\{\phi_n\}_{n=1}^{\infty}$ are respectively the sequences of EM iterates of the estimates of θ and ϕ and they converge, their limits are the MLEs of θ and ϕ, [for proof see Dempster et al (1977)]. Nanjundan (2006) has observed that the above sequences of EM iterates converge for every sample simulated for various combinations of θ and ϕ.

Since the maximum likelihood estimates of θ and ϕ are those values of θ and ϕ that maximize the likelihood function, $L(\theta, \phi | \underline{x})$ the iterative procedure can be terminated as soon as the absolute difference between the likelihood functions after two successive iterations is less than a pre-assigned threshold value h, say 10^{-5} or 10^{-6}.

4. A Comparison of the Estimation Procedures

Since the conditional maximum likelihood estimate of θ is the solution of the equation in (2.2), usually Newton–Raphson iterative procedure is used to solve such an equation and it is known to converge very fast. After obtaining an estimate of θ with required precision, it is used to compute the estimate of ϕ. That is the values of both θ and ϕ are not simultaneously used to maximize the likelihood function. This is because ϕ is treated as a nuisance parameter in the conditional maximum likelihood procedure. But in the EM algorithm, the likelihood is maximized simultaneously with respect to θ and ϕ. Another drawback of the conditional approach is that the estimate of ϕ suggested by Yip (1988) may turn out to be negative which is not admissible since ϕ is a proportion. From (2.3), it can be

observed that the estimate of φ will be negative when $n_g > n(1-e^{-\bar{\theta}})$. Such a situation is encountered when the sample is

```
0  0  0  0  0  0  0  0  1  0
0  1  1  0  0  1  0  0  0  1
1  0  0  0  1  0  0  2  1  1
0  0  0  1  0  1  1  0  1  0
1  0  0  2  0  0  3  1  0  0
```

For this sample, $n = 50$, $n_0 = 32$, and $n_g = 18$. When the initial estimate of θ is taken to be the sample mean, the following are the Newton-Raphson iterates of the conditional maximum likelihood estimate of θ:

Iteration	CML Estimate of θ
1	0.415807
2	0.415723
3	0.415723

Though the Newton-Raphson procedure for computing the conditional maximum likelihood estimate of θ converges correct to 6 decimal places after the third iteration, the corresponding estimate of φ is -0.058398 which is inadmissible, whereas the EM algorithm gives the estimates of θ and φ to be respectively 0.522885 and 0.158514 after the 25th iteration. The following are the outputs of the first 10 EM iterates of the estimates of θ and φ along with the log-likelihood function for the above sample.

Iteration	MLE of θ	MLE of φ	Log-Likelihood
1	0.829882	0.469804	-41.828278
2	0.770454	0.428908	-41.414387
3	0.728459	0.395985	-41.146732
4	0.696885	0.368618	-40.960281
5	0.672109	0.345345	-40.823498
6	0.652051	0.325206	-40.719246
7	0.635419	0.307543	-40.637405
8	0.621363	0.291879	-40.571636
9	0.609300	0.277860	-40.517757
10	0.598816	0.265216	-40.472912

Note that the log-likelihood function is steadily maximized. Though the rate of convergence of the EM algorithm is slower when compared to the Newton-Raphson method, one need not mind it because 100's of EM iterations can be carried out in no time when high level programming languages like C and C++ are used. But the EM algorithm also may converge very fast depending on the sample. When the given sample is

```
4  3  3  1  5  0  2  2  2  3
0  2  6  0  3  2  2  1  4  0
1  1  2  0  1  0  4  3  3  2
3  1  1  3  0  3  2  4  0  4
5  5  0  2  2  4  1  5  4  1,
```

the following are the outputs of the EM iterates

Iteration	MLE of θ	MLE of φ	Log-Likelihood
1	2.548971	0.121214	-19.959564
2	2.530770	0.114894	-19.940647
3	2.521329	0.111580	-19.935225
4	2.516205	0.109770	-19.933571
5	2.513356	0.108761	-19.933048
6	2.511750	0.108191	-19.932880
7	2.510838	0.107868	-19.932827
8	2.510318	0.107683	-19.932810
9	2.510021	0.107577	-19.932804
10	2.509851	0.107517	-19.932802
11	2.509753	0.107482	-19.932800
12	2.509697	0.107462	-19.932800.

Observe that the EM algorithm converges correct to 6 decimals after the 12th iteration which is very moderate.

5. Discussion

Though the EM algorithm converges slower when compared to the Newton-Raphson procedure used to compute the conditional maximum likelihood estimate of θ, the EM algorithm is preferable because i) it maximizes the likelihood function simultaneously with respect to θ and φ, ii) it does not lead to inadmissible estimate of φ, and iii) when high level programming languages like C and C++ are used the number of iterations is practically immaterial.

It can be realized that treating $\{Z_j: X_j = 0\}$, in terms of the notations of Section 3, as missing data is very appropriate since the couples who do not have any children may not disclose or not even like to know whether they are sterile or not due to personal and social reasons. Hence, missing data are not uncommon in these studies. The computational procedure presented in this paper determines the maximum likelihood estimates using the EM algorithm when there are missing data.

The mixture model discussed here is applicable in other practical situations too. For example, insects live on leaves that are suitable for feeding. If a leaf is unsuitable for feeding, then no insect lives on it. Suppose that the proportion of unsuitable leaves in a tree is φ and the number of insects per suitable leaf has a Poisson distribution with mean θ. Then, the distribution of the number of insects (X) found on a randomly selected leaf is a mixture of a singular distribution and a Poisson distribution.

Acknowledgement

References

1. Cox, D.R. (1958). Some problems connected with statistical inference. *Ann. Math. Statist.*, 29, 357 – 372.

2. Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B*, 39, 1-38.

3. Hall, D.B. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics*, 56, 1030-1039.

4. Godambe, V.P. (1976). Conditional likelihood and unconditional optimum estimating equations. *Biometrika*, 63, 277 – 284.

5. Kale, B.K. (1998). Optimal estimating equations for discrete data with higher frequencies at a point. *J. Ind. Stat. Assoc.*, 36, 125 – 136.

6. Krishnan, T. (2004). The EM Algorithm. In *Statistical Computing*, Eds. D. Kundu and A. Basu, Narosa, New Delhi.

7. Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing, *Technometrics*, 34, No.1, 1 - 14.

8. McLachlan, G.J. and Krishnan, T. (1997). *The EM Algorithm and Extensions*. John Wiley and Sons, New York.

9. Nanjundan, G. (2006). An EM algorithmic approach to maximum likelihood estimation in a mixture model. *Vignana Bharathi*, Vol. 18, 7 – 13.

10. Yip, P. (1988). Inference about the mean of a Poisson distribution in the presence of a nuisance parameter. *Austral. J. Statist.*, 30 (3), 299-306.

11. Yip, P. (1991). Conditional inference on a mixture model for the analysis of count data. *Commn. in Statist. – Theory and Methods*, 20 (7), 2045 – 2057.