Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: http://www.tandfonline.com/loi/lssp20

Maximum Likelihood Estimation of Parameters in a Mixture Model

Satish Bhat, R. Vidya & V. Pandit Parameshwar

To cite this article: Satish Bhat, R. Vidya & V. Pandit Parameshwar (2016) Maximum Likelihood Estimation of Parameters in a Mixture Model, Communications in Statistics - Simulation and Computation, 45:5, 1776-1784, DOI: 10.1080/03610918.2013.879177

To link to this article: http://dx.doi.org/10.1080/03610918.2013.879177

Accepted author version posted online: 19 Jun 2014.
Published online: 19 Jun 2014.

Communications in Statistics—Simulation and Computation®, 45: 1776–1784, 2016

ISSN: 0361-0918 print / 1532-4141 online
DOI: 10.1080/03610918.2013.879177

Maximum Likelihood Estimation of Parameters in a Mixture Model

SATISH BHAT,[1] R. VIDYA,[1]
AND V. PANDIT PARAMESHWAR[2]

[1]Department of Statistics, Yuvaraja's College, Mysore, India
[2]Department of Statistics, Bangalore University, Bangalore, India

The estimation of parameters of the log normal distribution based on complete and censored samples are considered in the literature. In this article, the problem of estimating the parameters of log normal mixture model is considered. The Expectation Maximization algorithm is used to obtain maximum likelihood estimators for the parameters, as the likelihood equation does not yield closed form expression. The standard errors of the estimates are obtained. The methodology developed here is then illustrated through simulation studies. The confidence interval based on large-sample theory is obtained.

Keywords Bootstrapping and Confidence interval; EM algorithm Standard errors; MLE.

Mathematics Subject Classification 62F10

1. Introduction

Let X be the survival time of a component under study. Suppose the component is functioning well even after say warranty period t (> 0), it can be termed as good. If $x_i \le t$, then the component can be called either good or of poor quality. Let φ be the proportion of poor-quality components in a batch. The survival time of number of poor-quality components produced in each batch is distributed over (0, t) with probability function $f_1(x_i)$. Suppose that the survival time of number of good components in a batch is distributed over (0, ∞) with probability function $f_\theta(x_i)$, θ may be real or vector. Thus, survival time of a component selected at random from a lot has the following probability density function $f(x_i)$, defined as:

$$f(x_i) = \begin{cases} \varphi f_1(x_i) + (1-\varphi) f_\theta(x_i), & x_i \le t \\ (1-\varphi) f_\theta(x_i), & x_i > t \end{cases} \tag{1}$$

Equation (1) can be written as:

$$f(x_i) = [\varphi f_1(x_i) + (1-\varphi) f_\theta(x_i)]^{1-\alpha_i} \, [(1-\varphi) f_\theta(x_i)]^{\alpha_i}, \tag{2}$$

Received February 7, 2012; Accepted December 19, 2013
Address correspondence to Dr. Parameshwar, V. Pandit, Department of Statistics, Bangalore University, Bangalore 560056, India; E-mail: panditpv12@gmail.com

where

$$\alpha_i = \begin{cases} 0, & x_i \le t \\ 1, & x_i > t. \end{cases} \tag{3}$$

Suppose $\underline{X} = (X_1, X_2, ..., X_n)$ be a random sample from X. Then the likelihood function of the random sample is given by

$$L(\theta, \varphi | \underline{x}) = \prod_{i=1}^{n} [\varphi f_1(x_i) + (1-\varphi) f_\theta(x_i)]^{1-\alpha_i} [(1-\varphi) f_\theta(x_i)]^{\alpha_i}. \tag{4}$$

It is observed that, in various situations likelihood equation does not yield closed form expression for maximum likelihood estimators (MLEs) of θ and φ. Therefore, in order to maximize the likelihood function for a given $\underline{x}$, one has to apply numerical procedures like Newton–Raphson and Fisher's scoring. But the above iterative procedures could fail due to boundary problem and since the likelihood equation has flat surface, Yip (1988) has observed that in case of power series distributions such as Poisson, binomial, etc., it is difficult to find the MLEs by these numerical procedures. Thus, as justified by Cox (1958), Yip (1988) has applied the conditional likelihood approach to estimate θ and φ. Moreover, Yip (1988) has also computed the loss incurred in estimating θ and φ by the conditional likelihood approach, treating φ as a nuisance parameter.

Here, we do not treat the mixture parameter φ as a nuisance parameter.

In this article, we consider the problem of estimating the parameters of the log normal mixture model with pdf

$$f(x_i) = \begin{cases} \varphi f_1(x_i) + (1-\varphi) f_\theta(x_i), & if \quad x_i \le t \\ (1-\varphi) f_\theta(x_i), & if \quad x_i > t \end{cases}, \tag{5}$$

where

$$f_1(x_i) = \frac{1}{t}, \quad 0 < x_i < t \tag{6}$$

and

$$f_\theta(x_i) = \begin{cases} \dfrac{1}{\sigma\sqrt{2\pi}} \dfrac{1}{x_i} \exp\left\{ -\dfrac{(\log x_i - \mu)^2}{2\sigma^2} \right\}, & x_i > 0 \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

where $\theta = (\mu, \sigma)$. Suppose that $\underline{X} = (X_1, X_2, ..., X_n)$be an observed random sample from log normal mixture model. Then the likelihood function of $\theta$and φ given the sample $\underline{X}$ is given by

$$L(\mu, \sigma, \varphi | \underline{x}) = \prod_{i=1}^{n} \left[\frac{\varphi}{t} + (1-\varphi) f_\theta(x_i) \right]^{1-\alpha_i} [(1-\varphi) f_\theta(x_i)]^{\alpha_i}, \tag{8}$$

The MLE procedure can be applied to estimate θ and φ simultaneously by solving $\frac{\partial \log L}{\partial \theta} = 0$ and $\frac{\partial \log L}{\partial \varphi} = 0$. However, the likelihood equation does not yield closed form expression and usual Newton–Raphson iterative method may fail due to boundary problems (see McLeish and Small (1988) and Sprott (1980). Hence, an alternative procedure is to be used to solve this problem namely, EM algorithm. EM algorithm is applied to a variety of statistical problems such as resolution of mixtures, multiway contingency tables, variance components estimation, and factor analysis. It has also found applications in specialized areas like genetics, medical imaging, and neural networks. For details, see McLachlan and Krishnan (1997) and Krishnan (2004). However, one can refer Daniel and Yip (1993), and McLachlan and Peel (2000) in postulating the complete-data framework for the application of the EM algorithm.

The EM algorithm is a broadly applicable statistical technique for maximizing complex likelihoods and handling the incomplete-data problems. It is a popular and a powerful numerical iterative procedure of computing the MLEs. At each iteration step of the algorithm, two steps are performed: (i) Expectation step (E-Step) and (ii) Maximization step (M-Step). E-Step consisting of projecting an appropriate functional containing the augmented data on the space of the original, incomplete data, and (ii) M-Step consisting of maximizing the functional. The name EM algorithm was coined by Dempster et al. (1977) who synthesized and earlier formulation in many particular cases and gave a general method of finding the MLEs in a variety of problems in their fundamental paper.

Section 2 contains some background on estimation of parameters via the EM algorithm.

In section 3, the tables of comparison of estimators of μ are presented when sample mean, and the sample median as initial estimates forμ. Some observations based on simulation study are made in section 4.

2. Estimation of Parameters via the EM Algorithm

Suppose $\underline{X} = (X_1, X_2, ..., X_n)$ be an observed random sample on X. Then, in the EM framework the data $\underline{X}$ **are** being viewed as an incomplete because, their associated component indicator variable Z_i (say) is not available. That is, to apply EM algorithm first of all we need to accommodate missing data. Thus, we define a random variable Z_i as:

$$Z_i = \begin{cases} 1, & \text{if the } i^{th} \text{component is good} \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

Then, we have $P(Z_i = 1) = 1 - \varphi = 1 - P(Z_i = 0)$, $i = 1,2,\ldots,n$. Note that, if $X_i > t$, then $Z_i = 1$; and if $X_i \le t$, then $Z_i = 0\, or\, 1$. In other words, for $X_i \le t$ we have no information onZ_i. Hence $\{Z_i : X_i \le t\}$ can be treated as the missing data. Thus $((X_1, Z_1), (X_2, Z_2). \ldots (X_n, Z_n))$, becomes the complete sample by augmenting the observed data $\underline{X} = (X_1, X_2, ..., X_n)$ with $\underline{Z} = (Z_1, Z_2, ..., Z_n)$. The likelihood function of the complete data is then given by

$$\begin{aligned} L_c(\mu, \sigma, \varphi | \underline{x}, \underline{z}) &= \prod_{i=1}^{n} f(x_i, z_i) = \prod_{i=1}^{n} f(z_i) f(x_i | z_i) \\ &= \prod_{i=1}^{n} \left\{ [f_\theta(x_i)]^{z_i} \right\}^{1-\alpha_i} [f_\theta(x_i)]^{\alpha_i} (1-\varphi)^{z_i} \varphi^{1-z_i}, \end{aligned} \tag{10}$$

where α_i is as defined in Eq. (3).

Thus, using the Eqs. (3) and (9), the Eq. (10) reduces to

$$\begin{aligned}\log\ L_c(\mu,\sigma,\varphi|,) = {} & n_g\left[\log(1-\varphi)-\log\sigma\right]+n_l\log\varphi-\sum_{i:x_i>t}\log x_i-\frac{1}{2\sigma^2}\sum_{i:x_i>t}(\log x_i-\mu)^2\\ & -\sum_{i:x_i\le t}Z_i\log\varphi+\sum_{i:x_i\le t}Z_i\log(1-\varphi)-\sum_{i:x_i\le t}Z_i\log\sigma-\sum_{i:x_i\le t}Z_i\log x_i\\ & -\frac{1}{2\sigma^2}\sum_{i:x_i\le t}Z_i(\log x_i-\mu)^2,\end{aligned} \tag{11}$$

where n_g and n_l are the number of observations which are, respectively, greater than or equal to t and less than t.

2.1. *The E-step*

By E-step, one can handle the problem of **presence** of the missing data z_i. It takes the conditional expectation of the complete-data log likelihood, $\log L_c(\theta,\varphi)$ given the observed data x_i, using current fit for θ and φ. That is, suppose θ_0 and φ_0 are respectively the initial estimates of θ and φ, then E-step requires the computation of the conditional expectation of $\log L_c(\theta,\varphi)$ for the given x_i, i.e.,

$$Q(\theta:\theta_0;\varphi:\ \varphi_0)=E\{\log L_c(\theta,\varphi|\theta_0,\varphi_0,x_i)\} \tag{12}$$

Since the complete-data log likelihood is linear in z_i, the E-step simply requires the computation of the current conditional expectation of Z_i given the observed data x_i. That is, $E(Z_i|\theta_0,\varphi_0,X_i\le t)$; where θ_0 and φ_0 are respectively the initial estimates of $\theta=(\mu,\sigma)$ and φ. Then by using Eqs. (9) and (12), we have

$$\begin{aligned}E(Z_i|\mu_0,\sigma_0,\varphi_0,X_i\le t) &= 1\times P(Z_i=1|\mu_0,\sigma_0,\varphi_0,X_i\le t)\\ &\quad +0\times P(Z_i=0|\mu_0,\sigma_0,\varphi_0,X_i\le t).\\ &= P(Z_i=1|\mu_0,\sigma_0,\varphi_0,X_i\le t)\end{aligned} \tag{13}$$

Applying Bayes theorem, we have

$$P(Z_i=1|\mu_0,\sigma_0,\varphi_0,x_i)=\frac{P(Z_i=1)\times P(A_i|Z_i=1)}{P(Z_i=0)\times P(A_i|Z_i=0)+P(Z_i=0)\times P(A_i|Z_i=0)}, \tag{14}$$

where $A_i=(X_i=x_i|\mu_0,\sigma_0,\varphi_0)$.

Hence,

$$\begin{aligned}P\left[Z_i=1|\mu_0,\sigma_0,\varphi_0,x_i\right] &= \frac{(1-\varphi_0)\frac{1}{\sigma_0\sqrt{2\pi}}\frac{1}{x_i}\exp\left\{-\frac{(\log x_i-\mu_0)^2}{2\sigma_0^2}\right\}}{\frac{\varphi_0}{t}+(1-\varphi_0)\frac{1}{\sigma_0\sqrt{2\pi}}\frac{1}{x_i}\exp\left\{-\frac{(\log x_i-\mu_0)^2}{2\sigma_0^2}\right\}}\\ &= w_i(say),\ for\ i:x_i\le t.\end{aligned} \tag{15}$$

Using (15), Eq. (11) can be expressed as the conditional expectation of the complete-data log likelihood, i.e.,

$$E[\log L_c(\mu,\sigma,\varphi|\underline{x},\underline{z})]=n_g[\log(1-\varphi)-\log\sigma]+n_l\log\varphi-\sum_{i:x_i>t}\log x_i$$

$$-\frac{1}{2\sigma^2}\sum_{i:x_i>t}(\log x_i-\mu)^2-\sum_{i:x_i\le t}w_i\log\varphi+\sum_{i:x_i\le t}w_i\log(1-\varphi)$$
$$-\sum_{i:x_i\le t}w_i\log\sigma-\sum_{i:x_i\le t}w_i\log x_i-\frac{1}{2\sigma^2}\sum_{i:x_i\le t}w_i(\log x_i-\mu)^2, \tag{16}$$

2.2 *The M-step*

Here we simply maximizing the Eq. (12) for θ and φ. It requires the global maxima of $E\{\log L_c(\mu,\sigma,\varphi|\mu_0,\sigma_0,\varphi_0,\underline{x},\underline{z})\}$ on the (r + 1)$^{\text{th}}$ iteration with respect to θ and φ. The E- and M-steps are replicated until the difference:

$$L(\theta^{(r+1)})-L(\theta^{(r)})$$

is less than a prefixed threshold value h which, is usually taken to be 10^{-4} or 10^{-5} (one can refer Wu (1983) for convergence of EM algorithm). Dempster et al. (1977) showed that for the incomplete-data likelihood function$L(\theta)$; $L(\theta^{(r+1)})\ge L(\theta^{(r)})$,for $r=0,1,2.\ \ldots$ Hence, convergence must be obtained with the sequence of likelihood values which are bounded above. Dempster et al. (1977) also showed that if $Q(\theta:\theta_0;\varphi:\varphi_0)$is continuous weakly in both θ and φ then L^* will be a local maxima of $L(\theta)$ if and only if the sequence is not trapped at some saddle point. And suppose $\{\theta_n\}_{n-1}^{\infty}$ and $\{\varphi_n\}_{n-1}^{\infty}$are respectively the sequences of estimates of θ and φ in the successive iterations and if they converge, then their limits are the MLEs of θ and φ (for the proof see Dempster et al. (1977)).

Since $E\{\log L_c(\mu,\sigma,\varphi|\mu_0,\sigma_0,\varphi_0,\underline{x},\underline{z}\}$ is differentiable with respect to μ, σ and φ, the values ofμ, σ, and φ for which $E\{\log L_c(\mu,\sigma,\varphi|\mu_0,\sigma_0,\varphi_0,\underline{x},\underline{z}\}$ is maximum and can be determined by the method of calculus. Thus, from the Eq. (16) we have

$$\frac{\mathrm{d}E\{\log L_c(\mu,\sigma,\varphi|\mu_0,\sigma_0,\varphi_0,\underline{x},\underline{z}\}}{d\mu}=\frac{1}{\sigma^2}\sum_{i:x_i>t}(\log x_i-\mu)+\frac{1}{\sigma^2}\sum_{i:x_i\le t}w_i(\log x_i-\mu)=0, \tag{17}$$

$$\frac{\mathrm{d}E\{\log L_c(\mu,\sigma,\varphi|\mu_0,\sigma_0,\varphi_0,\underline{x},\underline{z}\}}{d\sigma}=-\frac{n_g}{\sigma}+\frac{1}{\sigma^3}\sum_{i:x_i>t}(\log x_i-\mu)^2-\frac{1}{\sigma}\sum_{i:x_i\le t}w_i+\frac{1}{\sigma^3}\sum_{i:x_i\le t}w_i(\log x_i-\mu)^2=0, \tag{18}$$

and

$$\frac{\mathrm{d}E\{\log L_c(\mu,\ \sigma,\ \varphi|\mu_0,\ \sigma_0,\ \varphi_0,\ \underline{x},\ \underline{z}\}}{d\varphi}=-\frac{n_g}{1-\varphi}+\frac{n_l}{\varphi}-\sum_{i:x_i\le t}\frac{w_i}{\varphi}-\sum_{i:x_i\le t}\frac{w_i}{(1-\varphi)}=0. \tag{19}$$

Table 1
Estimates of μ, σ, and φ for $t = 2.5$, $\mu = 5.10$, $\varphi = 0.05$, and $\sigma = 0.20$

			Estimates		
Sample Size	Initial Estimates	No. of Iterations	μ	φ	σ
50	Mean	03	5.28320	0.06000	0.22190
	Median	03	5.28322	0.06000	0.19858
100	Mean	03	5.25410	0.05000	0.21587
	Median	03	5.25407	0.05000	0.19682
250	Mean	03	5.10460	0.04800	0.21036
	Median	03	5.10457	0.04800	0.19397
500	Mean	03	5.10090	0.04530	0.21000
	Median	03	5.10088	0.04530	0.19266

On simplifying the above simultaneous equations, we get

$$\hat{\mu} = \frac{\sum_{i:x_i>t} \log x_i + \sum_{i:x_i \le t} w_i \log x_i}{n_g + \sum_{i:x_i \le t} w_i} \tag{20}$$

$$\hat{\sigma} = \sqrt{\frac{\sum_{i:x_i>t} (\log x_i - \mu)^2 + \sum_{i:x_i \le t} w_i (\log x_i - \mu)^2}{n_g + \sum_{i:x_i \le t} w_i}} \tag{21}$$

$$\hat{\varphi} = \frac{n_l - \sum_{i:x_i \le t} w_i}{n}, \tag{22}$$

where $n = n_g + n_l$

Steps of the EM algorithm for computing the MLEs of μ, σ and φ:

Table 2
Estimates of μ, σ, and φ for $t = 150.2$ $\mu = 5.10$, $\varphi = 0.05$, and $\sigma = 0.20$

			Estimates		
Sample Size	Initial Estimates	No. of Iterations	μ	φ	σ
50	Mean	03	5.28321	0.06273	0.26940
	Median	52	5.28339	0.06351	0.25263
100	Mean	05	5.18361	0.05978	0.23089
	Median	56	5.18348	0.05931	0.22959
250	Mean	06	5.08555	0.04714	0.21034
	Median	55	5.08554	0.04714	0.21023
500	Mean	23	5.07429	0.03887	0.21033
	Median	67	5.07421	0.03858	0.21007

Table 3
Standard Errors of μ, σ, and φ for $t = 2.5$, $\mu = 5.10$, $\varphi = 0.05$, and $\sigma = 0.20$

Sample Size	Initial Estimates	Standard Errors(SE) from Bootstrapping μ	φ	σ	95% confidence Interval for μ
50	Mean	0.11952	0.03555	0.08955	(4.75712, 5.55935)
	Median	0.02869	0.03555	0.08356	(5.07171, 5.24476)
100	Mean	0.05526	0.01949	0.05544	(4.94423, 5.29001)
	Median	0.01945	0.01949	0.05299	(5.05227, 5.17885)
250	Mean	0.01923	0.01314	0.02453	(5.03624, 5.14428)
	Median	0.01129	0.01314	0.02279	(5.05783, 5.12270)
500	Mean	0.01172	0.01195	0.01547	(5.08374, 5.13958)
	Median	0.00932	0.01195	0.01341	(5.09078, 5.13253)

a. Choose the initial estimates μ_0, σ_0 and φ_0
b. Compute w_i (Ref. Eq. (15))
c. Using the realization $(x_1, x_2, ..., x_n)$of the observed sample, compute

$$\mu_1 = \frac{\sum_{i:x_i>t} \log x_i + \sum_{i:x_i\le t} w_i \log x_i}{n_g + \sum_{i:x_i\le t} w_i}$$

$$\sigma_1 = \sqrt{\frac{\sum_{i:x_i>t} (\log x_i - \mu)^2 + \sum_{i:x_i\le t} w_i(\log x_i - \mu)^2}{n_g + \sum_{i:x_i\le t} w_i}}$$

Table 4
Standard Errors of μ, σ, and φ for $t = 150.2$, $\mu = 5.10$, $\varphi = 0.05$, and $\sigma = 0.20$

Sample Size	Initial Estimates	Standard Errors(SE) from Bootstrapping μ	φ	σ	95% confidence Interval for μ
50	Mean	0.34890	0.05593	0.16783	(4.15237, 6.00875)
	Median	0.03557	0.10475	0.72986	(4.98357, 5.17755)
100	Mean	0.14720	0.04447	0.14420	(4.64913, 5.46422)
	Median	0.02187	0.09939	0.66567	(5.00154, 5.11181)
250	Mean	0.12322	0.03514	0.12762	(4.92599, 5.19865)
	Median	0.01679	0.06228	0.62181	(4.99182, 5.13276)
500	Mean	0.01566	0.02984	0.10459	(5.03304, 5.13608)
	Median	0.01158	0.04495	0.56436	(5.05575, 5.11329)

and

$$\varphi_1 = \frac{n_l - \sum_{i:x_i \le t} w_i}{n}.$$

d. Repeat the steps (b) and (c) by fixing $\mu_0 = \mu_1$, $\sigma_0 = \sigma_1$ and $\varphi_0 = \varphi_1$, until the likelihood function converges.

3. Comparative Study

In this section, a simulation study is conducted by generating random sample of sizes $n = 50$, 100, 250, and 500 from the log normal mixture model with 1000 replications each, and the MLEs are computed for the parameters μ, σ, and φ using EM algorithm which is described in the section 2. The initial estimates, number of iterations required and the estimates of the parameters for different sample sizes are presented in Tables 1 and 2 below. Tables 3 and 4 give the standard error for the estimators and also, the 95% confidence interval for μ, which are obtained through bootstrapping.

4. Discussion

Generally, EM algorithm converges *more slowly* in computing the conditional maximum likelihood estimate of μ, σ, and φ even though, it is preferable because (i) it maximizes the likelihood function simultaneously with respect to parameters and (ii) when the high level programming languages like C++ and softwares like MATLAB, R are used, the number of iterations is practically immaterial. But, here we observe that if the cut-off point "t" is too small as compared to the sample mean, the EM algorithm converges faster, whereas if "t" is around the sample mean, the convergence is at a slower rate. Moreover, when median is an initial estimator of μ, it is observed that the EM algorithm converges *more slowly* as compared to mean as an initial estimator of μ. Also, we observe that the estimates of μ, σ, and φ are nearer to the given values when median is an initial estimator of μ as compared to mean as an initial estimator. The standard errors (SEs) are obtained through Bootstrapping. Here we find that, when median is an initial estimator of μ the standard error (SE) of the estimator of μ is minimum and also, the length of the 95% confidence interval for μ is shorter.

Acknowledgments

The authors would like to thank the Editor-in-chief, Associate Editor, and Referee(s), for their valuable comments which led to significant improvement in the presentation of the work.

References

Cox, D. R. (1958). Two further applications of a model for binary regression. *Biometrika* 45:562–565.

Daniel, Y. T. Fong, Yip, P. (1993). An EM algorithm for a mixture model of count data. *Statistics and Probability Letters* 17:53–60.

Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B* 39:1–38.

Kale, B. K. (1998). Optimal estimating equations for discrete data with higher frequencies at a point. *Journal of the Indian Statistical Association* 36:125–136, University of Pune, Pune.

Krishnan, T. (2004). The EM algorithm. In: Kundu, D., Basu, A., eds. *Statistical Computing*, New Delhi: Narosa.

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

McLachlan, G. J., Peel, D. (2000). *Finite Mixture Models*. New York: Wiley.

McLeish, D. L., Small, C. (1988). *Lecture Notes in Statistics*. 44. Yew York: Springer– Verlag.

Meng, X. L., Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical Association* 86:899–909.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. New Delhi: John Wiley Eastern.

Sprott, D. A. (1980). Maximum Likelihood in small samples: Estimation in the presence of nuisance parameters. *Biometrika* 67:515–523.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *Annals of Statistics* 11:95–103.

Yip, P. (1988). Inference about the mean of a Poisson distribution in the presence of a nuisance parameter. *Australian Journal of Statistics* 30:299–306.

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