DOI:10.11721/cqnuj20140421

The Nested EM Algorithm for the Parameters of the Exponential-Geometric Model*

WANG Yanling, WANG Jixia

(College of Mathematics and Information Science, Henan Normal University, Xinxiang Henan 453007, China)

摘要:In this article we introduce the exponential geometric distribution (EG for brevity) as a model, which is obtained by mixing the exponential distribution with one truncated geometric distribution. It has the density of $f(x;\beta,p)=\beta(1-p)\times e^{-2\beta x}(2-pe^{-\beta x})(1-pe^{-\beta x})^{-2}$. By straightforward integration we find that the various moments of the EG is $E(x^r;\beta,p)=p^{-1}(1-p)r!\ \beta^{-r}[p^{-1}L(p,r)-1]$. Firstly, we discuss that the maximum likelihood estimates of β and p can not be get in explicit solution form; it should be solved by numerical algorithm. Then we nest one EM algorithm within the outer EM algorithm to solve the problem. Note that, for the outer EM algorithm the missing data are based on the mixture representation while at the inner EM algorithm, the missing data are the truncated observations. In the end we get the maximum likelihood estimators of parameters.

关键词:EM algorithm; exponential distribution; one truncated geometric distribution; Newton-Raphson algorithm; maximum likelihood estimation

中图分类号:O212.7 **文献标志码**:A **文章编号**:1672-6693(2014)04-0106-04

1 Introduction

The Exponential Poisson(EP) distribution as a model for lifetime data was introduced in the recent paper Kus[1] and Konstantinos Adamidis[2]. Similar models with other mixing distribution have been described in Adamidis and Loukas[3], and Tahmasbi and Rezaei[4] where a geometric and a logarithmic series mixing distributions were used, respectively.

In this article we introduce the exponential geometric distribution(EG for brevity) as a model. The distribution is obtained by mixing the exponential distribution with one truncated geometric distribution. An EM algorithm for estimating the parameters of the EG model has been proposed. This algorithm uses a Newton-Raphson approach at the M-step. In more details, the Newton-Raphson step is replaced by another EM algorithm at alternative approach. This approach has simple explicit solution form by making use of a nested EM algorithm (see, e.g. Van Dyk[5]) where the M-step itself is solved with an EM algorithm. An advantage of this algorithm is that it provides estimates in the admissible range when the initial values are in the admissible range, and avoids overflow problems that may occur during Newton-Raphson iterations. Additionally, it can be faster in some cases. Some simulation evidence on the speed of the new approach compared to the one that uses Newton-Raphson is provided.

* Received:10-20-2013 **Revised**:12-15-2013 **网络出版时间**:2014-7-3 23:03

Foundation:Supported by the Foundation and Frontier Technology Research Projects of Henan Province(122300410396) and the Soft Science Project of Henan Province(No. 122400450180).

First author biography: WANG Yanling, female, lecture, major in statistic inference under restriction.

收稿日期:2013-10-20 **修回日期**:2013-12-15 **网络出版时间**:2014-7-3 23:03

资助项目:河南省基础与前沿技术研究计划项目(No. 122300410396);河南省软科学项目(No. 122400450180)

作者简介:王艳玲,女,讲师,研究方向为数理统计,E-mail: bigduckwyl@163.com

网络出版地址:http://www.cnki.net/kcms/detail/50.1165.N.20140703.2303.021.html

2 The distribution

Let $y_1, y_2, \cdots, y_Z$ be a random sample from the distribution with density of $f(y;\beta)=\beta e^{-\beta y}$, where $\beta, y>0$. Let Z is one truncated geometric variable with probability function

$$P(z;p)=\frac{1}{p}(1-p)p^{z-1}, z=2,3,\cdots, \text{and } p\in(0,1)$$

Define $X=\min(y_1, y_2, \cdots, y_Z)$, then the conditional distribution of X follows an exponential distribution with parameter βz under the condition of $Z=z$, which has the conditional density function

$$f(x\mid z,\beta)=\beta z e^{-\beta z}, x,\beta>0$$

Then, the unconditional distribution of X will be an exponential geometric distribution (EG) with the probability density function

$$f(x;\beta,p)=\beta(1-p)e^{-2\beta x}(2-pe^{-\beta x})(1-pe^{-\beta x})^{-2}, x,\beta>0, p\in(0,1) \tag{1}$$

The latterone defines the distribution that we shall be referring to in the sequel as the EG. In the literature it is customary for such names to be given to distributions arising via the operation of mixing. It is obviously that by assuming that X follows an exponential distribution with parameter βz where Z is a truncated geometric variable with parameter p, the distribution of X is the EG with the density given by (1).

It can be seen that the EG density function is monotone decreasing with model values $\beta(2-p)(1-p)^{-1}$ at $x=0$. For all values of parameters, the density is strictly decreasing in x and tending to zero as $x\to\infty$.

3 Properties and moments of distribution

By straightforward integration we find that the distribution function is given by

$$F(x;\beta,p)=P\{X\leqslant x\}=p^{-1}[(1-pe^{-\beta x})^{-1}-1+p] \tag{2}$$

The raw moments of X may be determined from (1) by direct integration. For $r=1,2,\cdots$, we find that

$$E(x^r;\beta,p)=p^{-1}(1-p)r!\ \beta^{-r}[p^{-1}L(p,r)-1] \tag{3}$$

where $L(p,r)=\sum_{j=1}^{\infty}p^j j^{-r}$ is known as the poly-logarithm function(see [6]). The function is quickly evaluated and readily available in standard software, such as Mathematic. Furthermore, the mean and variance of the EG distribution are easily given by (3).

4 The nested EM algorithm for EG model

4.1 The EM algorithm

Proceeding with the method of likelihood, the log-likelihood function from a sample of n observations, $y_{obs}=(x_1, x_2, \cdots, x_n)$, is given by

$$l(\beta,p;y_{obs})=n\ln[\beta(1-p)]-2\beta\sum_{i=1}^{n}x_i-2\sum_{i=1}^{n}\ln(1-pe^{-\beta x_i})+\sum_{i=1}^{n}\ln(2-pe^{-\beta x_i})$$

and subsequently the associated gradients are found to be

$$\frac{\partial l}{\partial\beta}=n\beta^{-1}-2\sum_{i=1}^{n}x_i-2\sum_{i=1}^{n}[x_i p e^{-\beta x_i}(1-pe^{-\beta x_i})^{-1}]+\sum_{i=1}^{n}[x_i p e^{-\beta x_i}(2-pe^{-\beta x_i})^{-1}]$$

$$\frac{\partial l}{\partial p}=-n(1-p)^{-1}+2\sum_{i=1}^{n}[e^{-\beta x_i}(1-pe^{-\beta x_i})^{-1}]$$

The maximum likelihood estimates (MLE) of β and p must be derived numerically and we propose to use the EM algorithm[7] here. It is a statistically oriented iterative scheme with desirable properties that are fairly well known. To implement the algorithm, we define the hypothetical complete-data distribution with density function

$$f(x,z;\beta,p)=\beta(1-p)zp^{z-2}e^{-\beta zx}, \text{for } x,\beta>0, z=2,3,\cdots, p\in(0,1)$$

Thus, it is straightforward to verify that the E-step of an EM cycle require the computation of the conditional expectation of $(Z\mid X;\beta^{(t)},p^{(t)})$. Using $p(z\mid x;\beta,p)=zp^{z-1}e^{-\beta x(z-1)}(1-pe^{-\beta x})^2(2pe^{-\beta x}-p^2e^{-2\beta x})^{-1}$

We can found that $E(Z\mid X;\beta,p)=2(1-pe^{-\beta x})^{-1}-pe^{-\beta x}(2-pe^{-\beta x})^{-1}$

The cycle is completed with the M-step which is essentially-full data maximum likelihood over (β,p). As a result, an EM iteration, taking $(\beta^{(t)},p^{(t)})$ into $(\beta^{(t+1)},p^{(t+1)})$ is defined by (see [3] and [8])

$$\beta^{(t+1)}=n\left(\sum_{i=1}^{n}x_is_i\right)^{-1},p^{(t+1)}=1-\frac{n}{p^{(t+1)}}\left(\sum_{i=1}^{n}s_i\right)^{-1} \tag{4}$$

where $s_i=E(Z_i\mid x_i,\beta^{(t)},p^{(t)})=2(1-p^{(t)}e^{-\beta^{(t)}x_i})^{-1}-p^{(t)}e^{-\beta^{(t)}x_i}(2-p^{(t)}e^{-\beta^{(t)}x_i})^{-1}$.

The explicit solution can not be get in this equation. Then one needs to use a numerical algorithm, such as NR(see[9]). In general, it is easy to be solved. An alternative algorithm is proposed in the next section.

4.2 **The nested EM algorithm**

Van Dyk[5] showed that nested EM algorithm are useful as they simplify the steps of the algorithm to smaller and easier to handle steps, as well as improving the computational efficiency in certain problems. Van Dyk[5] described how nesting two(or more) EM algorithms can take advantage of conditional expectations in explicit solution form and lead to algorithms which tend to be faster to converge, reducing computational time, are straightforward to implement and program in standard statistical software, and enjoy stable conver-gence (e.g. monotone convergence in likelihood). It is interesting appropriate handling of them. Please refer to Popescu and Wong[10] for other applications of nested EM algorithm.

Truncated data constitute a typical example of data that can be considered as missing data (see, e.g. Mclachlan and Krishnan[11]). The EM algorithm is a standard algorithm to derive MLE for this data with " missing data" representation. For the EM algorithm described above, an important task during the M-step is to derive MLE from a sample of truncated Geometric random variables. This is exactly what the solution of (4) does. Alternatively this equation can be easily solved via an EM algorithm (see e.g. Böhning and Schön[12]).

Consider the casethat MLE for a sample of truncated at one geometric distribution observations is given. The idea is that apart from the observed data, there are some more observations that are one and hence they are missing, since one observation cannot be observed. For the geometric parameter, the sufficient statistic is $\sum Z_i$. The one value doesn't contribute to this sum but it inflates the denominator of the MLE of the parameter which is the sample mean. Hence, a simple EM algorithm could be used to estimate the expected number of one observation, denoted asn_0. Thus, utilizing the observed data $Z_1,Z_2,\cdots,Z_n$ the EM algorithm for a simple geometric distribution, truncated at one is described as:

Use current estimates for

E-step: Obtain $n_0^{\text{new}}=(n+n_0)(1-p)$

In fact this is the number of one observation we would expect if the parameter of the geometric distribution is p(i.e. $(1-p)$estimates the probability of one observation), while $n+n_0$ is the total sample size (i.e. the observed sample size and the one observations as assumed by the current estimates).

M-step: Update geometric parameter $p^{\text{new}}=1-(n+n_0^{\text{new}})\left(\sum_{i=1}^{n}s_i\right)^{-1}$

The algorithm at the E-step estimates how many one observation would expect to observe based on the current estimate, while at the M-step it updates the parameter by considering all the observations, observed or not.

Van Dyk[5] discussed the use of nested EM algorithms where the M-step of an EM algorithm can be another EM algorithm in itself. This is in fact an alternative approach to the use of any numerical maximization algorithm in the M-step, like the NR used in the previous section. Therefore, the idea in the current situation is that while trying to solve the M-step instead of using a NR method, we may nest one EM algorithm within the outer EM algorithm to solve the problem. Note that, for the outer EM algorithm the missing data are based on the mixture representation while at the inner EM algorithm, the missing data are the truncated observations.

From the above discussion a nested EM algorithm is proposed for the EG distribution where the M-step of the algorithm is solved by applying another EM algorithm. Consequently, the nested EM algorithm takes the form:

E-step: With the given values of parameters $\beta^{(t)}$, $p^{(t)}$ calculate E1-step: $s_i = E(Z_i \mid x_i, \beta^{(t)}, p^{(t)}) = 2(1-p^{(t)}e^{-\beta^{(t)}x_i})^{-1} - p^{(t)}e^{-\beta^{(t)}x_i}(2-p^{(t)}e^{-\beta^{(t)}x_i})^{-1}, i=1,2,\cdots,n$; E2-step: $n_0^{new}=(n+n_0)(1-p)$.

M-step: Update theparameters by M1-step: $\beta^{new} = n(\sum_{i=1}^{n} s_i x_i)^{-1}$; M2-step: $p^{new} = 1-(n+n_0)\times(\sum_{i=1}^{n} s_i)^{-1}$.

Note that steps E2 and M2 are the steps based on the second EM algorithm, the inner one. The two steps replace the need for a NR algorithm during the M-step.

An interesting point arises here. There are two strategies on how the EM algorithms will be used. The first one needs to run several steps of the inner EM Algorithm (steps E2 and M2) so as to ensure that the convergence of the estimates for the truncated geometric part. On the contrary, the second strategy, making use of the monotonic property of the algorithm, needs just one iteration of the inner EM at each outer EM(steps E1 and M1) iteration. In both cases the convergence of the algorithm can be based on the results of Van Kyk[5].

References:

[1] Kus C. A new lifetime distribution[J]. Computational Statistics and Data Analysis, 2007, 51: 4497-4509.

[2] Konstantinos A, Theodora D, Sotirios L. On an extension of the exponential-geometric distribution [J]. Statistics & Probability Letters, 2005, 73: 259-269.

[3] Adamidis K, Loukas S. A life time distribution with decreasing failure rate [J]. Statistics and Probability Letters, 1998, 39: 35-42.

[4] Tahmasbi R, Rezaei S. A two-parameter lifetime distribution with decreasing failure rate[J]. Computational Statistics and Data Analysis, 2008, 52: 3889-3901.

[5] Van Dyk D A. Nesting EM algorithms for computational efficiency[J]. Statistica Sinica, 2000, 10: 203-225.

[6] Wang A P, Zhang G Y, Liu F. Research and application of EM algorithm [J]. Computer Technology and Development, 2009, 19(9): 108-110.

[7] Hamid B. The Beta exponential-geometric distribution[J]. Communications in Statistics-Simulation and Computation, 2012, 41(9): 1606-1622.

[8] Theodora D, Konstantinos A, Sotirios L. Bivariate extended exponential- geometric distributions[J]. Communications in Statistics-Theory and Methods, 2012, 41(7): 1129- 1150.

[9] Wang Y L. Applying Grey Markov model to SO_2 emissions from industry[J]. Journal of Chongqing Normal University: Natural Science, 2008, 25(2): 74-77.

[10] Popescu C A, Wong Y S. Nested Monte Carlo EM algorithm for switching state-space models[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17: 1653～1663.

[11] Mclachlan G B, Krishnan N. The EM algorithm and its extensions[M]. Hoboken: Wiley, 1997.

[12] Böhning D, Schön D. Nonparametric maximum likelihood estimation of the population size based upon the counting distribution[J]. Journal of the Royal Statistical Society, Series C, Applied Statistics, 2005, 54: 721-737.

指数几何混合模型参数的嵌套 EM 算法

王艳玲，王继霞

(河南师范大学 数学与信息科学学院，河南 新乡 453007)

Abstract：本文介绍了由指数分布和一个截尾分布混合得到的指数几何混合分布模型，简记为 EG 模型。它的概率密度函数为 $f(x;\beta,p)=\beta(1-p)e^{-2\beta x}(2-pe^{-\beta x})(1-pe^{-\beta x})^{-2}$，通过直接积分得到该分布的矩为 $E(x^r;\beta,p)=p^{-1}(1-p)r!\ \beta^{-r}[p^{-1}L(p,r)-1]$。首先说明了用 EM 算法在 M 步中不能求得参数 β 和 p 的极大似然估计的显式解，需要用数值解法，然后通过嵌套一个 EM 算法在另一个 EM 算法中，外层 EM 算法是基于混合模型的缺失数据讨论，内层 EM 算法是针对截尾观测数据的，得到了参数的极大似然估计量。

Key words：EM 算法；指数分布；截尾几何分布；Newton-Raphson 算法；极大似然估计

(责任编辑 游中胜)