第30卷第1期
2001年3月

上海师范大学学报(自然科学版)
J. of Shanghai Teachers Univ. (Natural Sciences)

Vol. 30, No. 1
Mar. 2001

分组数据下G-O模型的参数估计

刘荣官，费鹤良

(上海师范大学 数学科学学院，上海 200234)

摘　要：研究了软件可靠性的重要模型——G-O模型的参数估计，对分组数据情形得出了该模型中参数的最大似然估计存在的充分必要条件.

关键词：非齐次泊松过程；G-O模型；最大似然估计

中图分类号：O213.2　　**文献标识码**：A　　**文章编号**：1000-5137(2001)01-0027-05

0　引　言

在软件可靠性模型中，非齐次泊松过程(NHPP)模型[1,2]在实际中有着广泛的应用，其中较为重要的一类模型是GOEL和OKUMOTO[3]于1979年提出的模型，简称G-O模型，即指数均值函数NHPP模型。关于在故障数据下G-O模型的最大似然估计可参见[4]. 在文献[5]中KNAFL研究了分组数据下G-O模型的最大似然估计存在且唯一的条件，本文用了不同于[5]的方法，给出了G-O模型最大似然估计存在且唯一的充要条件.

1　G-O模型

设 $N(t)$ 为软件在 $(0,t]$ 内发生的累积故障数，$N(t)$ 为非齐次泊松过程(NHPP)，其均值函数为 $\mu(t)$，即 $EN(t)=\mu(t)$. 如果

$$\mu(t)=\alpha(1-e^{-\beta t}),\quad \alpha>0,\beta>0,\tag{1}$$

那么称此模型为G-O模型.

2　参数的最大似然估计

设对软件故障进行观测，观测时间 $t_1<t_2<\cdots<t_n$，设在 $(t_{i-1},t_i]$ 上观测到的故障数为 s_i，$i=1,2,\cdots,n$，令 $t_0=0$，$s=\sum_{i=1}^{n}s_i$ 为总故障数，$f_i=\frac{s_i}{s}$ 为第 i 个子区间 $(t_{i-1},t_i]$ 上出现的故障数占总故障数的比例. 设 $X(i)=N(t_i)-N(t_{i-1})$，它是 $(t_{i-1},t_i]$ 内出现的故障数. 记 $m(i)=\mu(t_i)-\mu(t_{i-1})$，

收稿日期：2000-09-08

基金项目：武汉大学软件工程国家重点实验室资助(AL98002)

作者简介：刘荣官(1958-)，男，上海师范大学数学科学学院讲师；费鹤良(1938-)，男，上海师范大学数学科学学院教授.

易知 $m(i)$ 为 $X(i)$ 的期望. 令 L_i 为 $X(i)$ 取值 s_i 的概率,则

$$L_i = \mathrm{P}[X(i) = s_i] = \frac{e^{-m(i)}[m(i)]^{s_i}}{s_i!} =$$

$$\frac{e^{-\alpha(e^{-\beta t_{i-1}}-e^{-\beta t_i})}[\alpha(e^{-\beta t_{i-1}} - e^{-\beta t_i})]^{s_i}}{s_i!} \tag{2}$$

由于 $X(i)$ $i = 1,2,\cdots,n$ 相互独立,所以 G-O 模型的似然函数为

$$L = \prod_{i=1}^{n} L_i = Ce^{-\alpha(1-e^{-\beta t_n})}\alpha^s \prod_{i=1}^{n}[e^{-\beta t_{i-1}} - e^{-\beta t_i}]^{s_i},$$

其中 C 是不依赖于 α,β 的常数,对 L 取自然对数得

$$\ln L = s\ln\alpha - \alpha(1 - e^{-\beta t_n}) + \sum_{i=1}^{n} s_i \ln(e^{-\beta t_{i-1}} - e^{-\beta t_i}) + \ln C. \tag{3}$$

分别对 $\ln L$ 关于 α,β 求偏导得似然方程

$$\frac{s}{\alpha} - (1 - e^{-\beta t_n}) = 0, \tag{4}$$

$$-\alpha t_n e^{-\beta t_n} + \sum_{i=1}^{n} s_i \frac{-t_{i-1}e^{-\beta t_{i-1}} + t_i e^{-\beta t_i}}{e^{-\beta t_{i-1}} - e^{-\beta t_i}} = 0 \tag{5}$$

从(4)解得

$$\alpha = \frac{s}{1 - e^{-\beta t_n}}. \tag{6}$$

将(6)代入(5)得

$$\sum_{i=1}^{n} f_i \frac{t_i e^{-\beta t_i} - t_{i-1}e^{-\beta t_{i-1}}}{e^{-\beta t_{i-1}} - e^{-\beta t_i}} = \frac{t_n e^{-\beta t_n}}{e^{-\beta t_n} - 1}. \tag{7}$$

若能从(7)解出 β,就可由(6)解得 α, 因此 G-O 模型参数 α,β 的最大似然估计的存在取决于方程(7)的解 β 的存在性与唯一性.

3 β 存在且唯一的充要条件

定理 方程(7)存在唯一解 β 的充分必要条件是

$$\sum_{i=1}^{n} f_i \frac{t_i + t_{i+1}}{2} < \frac{t_n}{2}, \text{且 } f_1 \neq 1.$$

在证明该定理之前先给出以下引理.

引理 若设 $f(x) = \dfrac{e^{\beta x - \frac{\beta t_n}{2}} - e^{-\beta x - \frac{\beta t_n}{2}}}{2x}$, $\beta > 0, t_n > 0$. 则当 $x > 0$ 时, $f(x)$ 为 x 的严格单调递增函数.

证明

$$f'(x) = \frac{e^{-\frac{\beta t_n}{2}}[(\beta e^{\beta x} + \beta e^{-\beta x})x - (e^{\beta x} - e^{-\beta x})]}{2x^2}. \tag{8}$$

从(8)式易知,只要下式成立就可保证 $f'(x) > 0$, 从而证明引理的结论成立.

$$\beta x e^{\beta x} + \beta x e^{-\beta x} > e^{\beta x} - e^{-\beta x}. \tag{9}$$

即

$$e^{-2\beta x} > \frac{1 - \beta x}{1 + \beta x}. \tag{10}$$

(10) 式当 $x \geqslant \dfrac{1}{\beta}$ 是,显然成立,当 $0 < x < \dfrac{1}{\beta}$ 时,对(10)式两边取自然对数得

$$\ln(1 + \beta x) - \ln(1 - \beta x) - 2\beta x > 0. \tag{11}$$

令(11)式左端为 $h(x)$, 并对其求导可得

$$h'(x)=\frac{2\beta}{1-(\beta x)^2}-2\beta. \tag{12}$$

因为 $0<\beta x<1$,所以 $h'(x)>0$. 再由 $h(0)=0$ 可知,不等式(11)成立,从而证明了引理.

在(7)式中,令

$$g_i(\beta)=\frac{t_ie^{-\beta t_i}-t_{i-1}e^{-\beta t_{i-1}}}{e^{-\beta t_i}-e^{-\beta t_{i-1}}},\ i=1,2,\cdots,n, \tag{13}$$

$$g(\beta)=\frac{t_ne^{-\beta t_n}}{e^{-\beta t_n}-1}=\frac{t_ne^{-\beta t_n}-t_0e^{-\beta t_0}}{e^{-\beta t_n}-e^{-\beta t_0}}. \tag{14}$$

为了研究(7)式是否存在解 $\beta\in(0,\infty)$,先考察当 β 趋向于 0^+ 和 $+\infty$ 时,$g_i(\beta)$ 与 $g(\beta)$ 的极限

$$\lim_{\beta\to+\infty}g_i(\beta)=\lim_{\beta\to+\infty}\frac{t_ie^{-\beta t_i}-t_{i-1}e^{-\beta t_{i-1}}}{e^{-\beta t_i}-e^{-\beta t_{i-1}}}=$$

$$\lim_{\beta\to+\infty}\frac{t_ie^{\beta(t_{i-1}-t_i)}-t_{i-1}}{e^{\beta(t_{i-1}-t_i)}-1}=t_{i-1},\quad i=1,2,\cdots,n, \tag{15}$$

$$\lim_{\beta\to+\infty}g(\beta)=t_0=0,$$

$$\lim_{\beta\to0^+}g_i(\beta)=\lim_{\beta\to0^+}\frac{t_ie^{-\beta t_i}-t_{i-1}e^{-\beta t_{i-1}}}{e^{-\beta t_i}-e^{-\beta t_{i-1}}}=$$

$$\lim_{\beta\to0^+}\frac{t_i-t_{i-1}e^{\beta(t_i-t_{i-1})}}{1-e^{\beta(t_i-t_{i-1})}}=-\infty,\quad i=1,2,\cdots,n,$$

同理

$$\lim_{\beta\to0^+}g(\beta)=-\infty.$$

为了比较 $g_i(\beta)$ 与 $g(\beta)$ 当 $x\to0^+$ 时趋向于 $-\infty$ 的快慢有必要对它们进行适当的变形. 令 $u(x)=xe^{-\beta x}$, $v(x)=e^{-\beta x}$, 则

$$g_i(\beta)=\frac{u(t_i)-u(t_{i-1})}{v(t_i)-v(t_{i-1})},\quad i=1,2,\cdots,n, \tag{16}$$

$$g(\beta)=\frac{u(t_n)-u(t_0)}{v(t_n)-v(t_0)}, \tag{17}$$

$$u'(x)=e^{-\beta x}(1-\beta x),$$

$$v'(x)=-\beta e^{-\beta x}\neq0.$$

由柯西中值定理存在 $\xi_i(\beta)\in(t_{i-1},t_i)$,$\xi(\beta)\in(t_0,t_n)$,使得下式成立

$$g_i(\beta)=\frac{u'[\xi_i(\beta)]}{v'[\xi_i(\beta)]}=\xi_i(\beta)-\frac{1}{\beta}, \tag{18}$$

$$g(\beta)=\frac{u'[\xi(\beta)]}{v'[\xi(\beta)]}=\xi(\beta)-\frac{1}{\beta}, \tag{19}$$

易知

$$\lim_{\beta\to+\infty}\xi_i(\beta)=t_{i-1}, \tag{20}$$

$$\lim_{\beta\to+\infty}\xi(\beta)=t_0=0, \tag{21}$$

而且

$$\lim_{\beta\to0^+}\xi_i(\beta)=\lim_{\beta\to0^+}[g_i(\beta)+\frac{1}{\beta}]=\lim_{\beta\to0^+}\frac{\beta(t_ie^{-\beta t_i}-t_{i-1}e^{-\beta t_{i-1}})+e^{-\beta t_i}-e^{-\beta t_{i-1}}}{\beta(e^{-\beta t_i}-e^{-\beta t_{i-1}})}$$

由洛彼塔法则可得

$$\lim_{\beta\to0^+}\xi_i(\beta)=\lim_{\beta\to0^+}\frac{\beta(-t_i^2e^{-\beta t_i}+t_{i-1}^2e^{-\beta t_{i-1}})}{e^{-\beta t_i}-e^{-\beta t_{i-1}}+\beta(-t_ie^{-\beta t_i}+t_{i-1}e^{-\beta t_{i-1}})}=$$

$$\lim_{\beta\to 0^+}\frac{t_{i-1}^2e^{-\beta t_{i-1}}-t_i^2e^{-\beta t_i}+\beta(t_i^3e^{-\beta t_i}-t_{i-1}^3e^{-\beta t_{i-1}})}{2(t_{i-1}e^{-\beta t_{i-1}}-t_ie^{-\beta t_i})+\beta(t_i^2e^{-\beta t_i}-t_{i-1}^2e^{-\beta t_{i-1}})}=\frac{t_i+t_{i-1}}{2}. \tag{22}$$

用同样方法可得

$$\lim_{\beta\to 0^+}\xi(\beta)=\frac{t_n+t_0}{2}=\frac{t_n}{2}. \tag{23}$$

将(18),(19)代入(7)得等价方程

$$\sum_{i=1}^{n}f_i\xi_i(\beta)=\xi(\beta). \tag{24}$$

移项得

$$\sum_{i=1}^{n}f_i[\xi_i(\beta)-\xi(\beta)]=0. \tag{25}$$

并令

$$F(\beta)=\sum_{i=1}^{n}f_i[\xi_i(\beta)-\xi(\beta)],$$

这样求(7)的解的问题就变为求方程 $F(\beta)=0$ 的根的问题. 下面证明 $F(\beta)$ 是 β 的单调递增函数.

$$[\xi_i(\beta)-\xi(\beta)]'=\frac{(t_i-t_{i-1})^2e^{-\beta(t_i+t_{i-1})}}{(e^{-\beta t_i}-e^{-\beta t_{i-1}})^2}-\frac{t_n^2e^{-\beta t_n}}{(e^{-\beta t_n}-1)^2}=$$

$$\frac{(t_i-t_{i-1})^2(e^{-\beta t_n}-1)^2e^{-\beta(t_i+t_{i-1})}-t_n^2(e^{-\beta t_i}-e^{-\beta t_{i-1}})^2e^{-\beta t_n}}{(e^{-\beta t_i}-e^{-\beta t_{i-1}})^2(e^{-\beta t_n}-1)^2}. \tag{26}$$

只要证明上式的分子大于 0,就可推出 $F'(\beta)>0$,为此作函数

$$w(\beta)=\frac{t_n^2(e^{-\beta t_i}-e^{-\beta t_{i-1}})^2e^{-\beta t_n}}{(t_i-t_{i-1})^2(e^{-\beta t_n}-1)^2e^{-\beta(t_i+t_{i-1})}}=$$

$$\left\{\frac{\dfrac{e^{-\beta(\frac{t_n}{2}-\frac{t_i-t_{i-1}}{2})}-e^{-\beta(\frac{t_n}{2}+\frac{t_i-t_{i-1}}{2})}}{(t_i-t_{i-1})}}{\dfrac{e^{-\beta t_n}-1}{t_n}}\right\}^2=\left[\frac{f(\frac{t_i-t_{i-1}}{2})}{f(\frac{t_n}{2})}\right]^2.$$

由引理知 $f(\frac{t_i-t_{i-1}}{2})<f(\frac{t_n}{2})$,所以 $w(\beta)<1$. 从而

$$[\xi_i(\beta)-\xi(\beta)]'>0,\ F'(\beta)>0. \tag{27}$$

由(20),(21),(22),(23)可知

$$\lim_{\beta\to+\infty}F(\beta)=\sum_{i=1}^{n}f_it_{i-1}, \tag{28}$$

$$\lim_{\beta\to 0^+}F(\beta)=\sum_{i=1}^{n}f_i\frac{t_i+t_{i-1}}{2}-\frac{t_n}{2}. \tag{29}$$

综上所述可以给出定理的证明.

证明 充分性. 若 $f_1=1$, $\sum_{i=1}^{n}f_i\frac{t_i+t_{i-1}}{2}<\frac{t_n}{2}$,则 $\lim_{\beta\to 0^+}F(\beta)<0$, $\lim_{\beta\to+\infty}F(\beta)>0$.

又 $\because$ $F(\beta)$ 严格单调递增,$\therefore$ $F(\beta)=0$ 存在唯一解 β,即(7)式存在唯一解 β.

必要性. 若 $f_1=1$,则 $\lim_{\beta\to+\infty}F(\beta)=0$,由 $F(\beta)$ 严格递增可知,$\forall\ \beta\in(0,+\infty)$ 有 $F(\beta)<0$, $\therefore$ $F(\beta)=0$ 无解.

若 $\sum_{i=1}^{n}f_i\frac{t_i+t_{i-1}}{2}>\frac{t_n}{2}$,则 $\lim_{\beta\to 0^+}F(\beta)>0$,因为 $F(\beta)$ 严格递增,所以对 $\forall\ \beta\in(0,+\infty)$,有 $F(\beta)>0$, $\therefore$ $F(\beta)=0$ 无解,即(7)式无解.

参考文献：

[1] 徐仁佐，射旻，郑人杰. 软件可靠性模型及应用[M]. 北京：清华大学出版社，南宁：广西科学技术出版社，1994.

[2] 穆莎 J D，艾里诺 A，奥本 K. 软件可靠性——度量、预计和应用[M]. 北京：机械工业出版社，1992

[3] GOEL A L, OKUMOTO K. Time Dependent Error Detection Rate Model for Software Reliability and Other Performance Measures[J]. IEEE Trans, Reliability, 1979, R-28(3): 206-211.

[4] KNADL G J, MORGAN J. Solvingl ML equations for 2-parameter poisson-process models for ungrouped software-failure data[J]. IEEE Trans, Relibility, 1996, R-45(1): 42-53.

[5] KNAFL G J. Solving maximum likelihood equations for two-parameter software reliability models using grouped data[M]. In third International Symposium on Software Reliability Engineering Proceedings, 1992, 205-213.

Estimation of the Parameters in G-O Model for Grouped Data

LIU Rong-guan, FEI He-liang

(College of Mathematical Sciences, Shanghai Teachers University, Shanghai 200234, China)

Abstract: The G-O model is an important model in software reliability. For grouped data, this paper gives the conditions for the existence and uniqueness of the maximum likelihood estimates of the parameters of the G-O Model.

Key words: nonehomogeneous poisson process; G-O model; maximum likelihood estimate