对数正态分布参数的近似极大似然估计

王蓉华，徐晓岭，费鹤良

(上海师范大学 数学科学学院，上海 200234)

摘　要：给出了在定数截尾数据缺失场合下两参数对数正态分布参数的近似极大似然估计.

关键词：定数截尾数据缺失；两参数对数正态分布；近似极大似然估计

中图分类号：O213.2　　**文献标识码**：A　　**文章编号**：1000-5137(2000)01-0039-05

在用统计方法处理实际问题时，常会遇到大量的不完全数据. 最常见的一种不完全数据情形是部分数据丢失的情形. 如在试验设计的统计分析中，由于存在观测手段、试验设备以及其他方面的困难(由于特殊原因使得工作人员暂时离开试验现场等等)，常常会出现某些试验数据未观测到的现象，而只能知道其前后的寿命数据，亦即部分数据丢失的现象. 这种情况在寿命试验中是经常会发生的——我们称其为数据丢失或数据缺失. 因为个别数据丢失就重做试验一般是不值得的，有时甚至是根本不可能的. 这就要求在数据出现丢失的情况下进行统计分析. 关于数据缺失场合下的统计分析已有许多文献作了研究，具体的可参阅文献[1]～[4]. 其中文献[3]对两参数 Weibull 分布(包括极小值分布)给出了参数近似极大似然估计. 本文给出了在定数截尾数据缺失场合下两参数对数正态分布参数的近似极大似然估计.

设产品的寿命服从两参数对数正态分布，其分布密度函数和分布函数分别为：

$$g(t,\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma t}e^{-\frac{(\ln t-\mu)^2}{2\sigma^2}},G(t,\mu,\sigma)=\int_0^t f(x)\mathrm{d}x\ ,t>0.$$

其中：μ、σ^2 分别称为对数均值和对数方差.

令：$X=\ln t$，则 X 服从 $N(\mu,\sigma^2)$. 其分布密度和分布函数分别为：

$$f(x,\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},F(t,\mu,\sigma)=\int_{-\infty}^t f(x)\mathrm{d}x.$$

令：$Z=\dfrac{X-\mu}{\sigma}$，则 Z 服从标准正态分布 $N(0,1)$. 假定有 n 个产品进行寿命试验，到有 r 个产

收稿日期：1997-05-20

基金项目：国家自然科学基金资助项目(69971016)和上海市高校科技发展基金资助项目(98D13)

作者简介：王蓉华(1972-)，女，上海师范大学数学科学学院助教. 费鹤良(1938-)，男，上海师范大学数学科学学院教授.

品失效时停止试验，其次序失效数据为：$t_{(1)}\leqslant t_{(2)}\leqslant\cdots\cdots\leqslant t_{(r)}$. 现考虑如下情形：即上述 r 个失效数据由于某种原因而使得有若干个数据缺失，设剩下 k 个数据，剩下的失效数据为：$0=t_{(r_0)}<t_{(r_1)}\leqslant t_{(r_2)}\leqslant\cdots\cdots\leqslant t_{(r_k)}$，$r_0=0$. 令：$X_{(i)}=\ln t_{(i)}$，$Z_{(i)}=\frac{X_{(i)}-\mu}{\sigma}$，则 $X_{(1)}\leqslant X_{(2)}\leqslant\cdots\cdots\leqslant X_{(r)}$ 为来自 $N(\mu,\sigma^2)$ 的样本容量为 n 的前 r 个次序统计量，而 $Z_{(1)}\leqslant Z_{(2)}\leqslant\cdots\cdots\leqslant Z_{(r)}$ 为来自标准正态分布的样本容量为 n 的前 r 个次序统计量，考虑到数据有缺失场合，剩下的 k 个数据如下：$X_{(r_1)}\leqslant X_{(r_2)}\leqslant\cdots\cdots\leqslant X_{(r_k)}$，及 $Z_{(r_1)}\leqslant Z_{(r_2)}\leqslant\cdots\cdots\leqslant Z_{(r_k)}$.

似然函数 $L(\mu,\sigma)$ 为(C 为正常数)：

$$L(\mu,\sigma)=C\sigma^{-k}[F(Z_{(r_1)})]^{r_1-1}\prod_{i=1}^{k-1}[F(Z_{(r_{i+1})})-F(Z_{(r_i)})]^{r_{i+1}-r_i-1}[1-F(Z_{(r_k)})]^{n-r_k}\prod_{i=1}^{k}f(Z_{(r_i)}). \tag{1}$$

令：$\frac{\partial\ln L(\mu,\sigma)}{\partial\mu}=0$，$\frac{\partial\ln L(\mu,\sigma)}{\partial\sigma}=0$，于是得如下方程：

$$(r_1-1)\frac{f(Z_{(r_1)})}{F(Z_{(r_1)})}+\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)\frac{f(Z_{(r_{i+1})})-f(Z_{(r_i)})}{F(Z_{(r_{i+1})})-F(Z_{(r_i)})}-(n-r_k)\frac{f(Z_{(r_k)})}{1-F(Z_{(r_k)})}-\sum_{i=1}^{k}Z_{(r_i)}=0. \tag{2}$$

$$k+(r_1-1)\frac{f(Z_{(r_1)})}{F(Z_{(r_1)})}Z_{(r_1)}+\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)\frac{Z_{(r_{i+1})}f(Z_{(r_{i+1})})-Z_{(r_i)}f(Z_{(r_i)})}{F(Z_{(r_{i+1})})-F(Z_{(r_i)})}-(n-r_k)\frac{f(Z_{(r_k)})}{1-F(Z_{(r_k)})}Z_{(r_k)}-\sum_{i=1}^{k}Z_{(r_i)}^2=0. \tag{3}$$

令：$p_{r_i}=\frac{r_i}{n+1}$，$q_{r_i}=1-p_{r_i}$，$\xi_{r_i}=F^{-1}(p_{r_i})$.

将函数 $\frac{f(Z_{(r_1)})}{F(Z_{(r_1)})}$ 在点 ξ_{r_1} 处泰勒展开得：

$$\frac{f(Z_{(r_1)})}{F(Z_{(r_1)})}\approx\alpha_0-\beta_0 Z_{(r_1)},$$

其中：

$$\alpha_0=\frac{f(\xi_{r_1})}{p_{r_1}}\left[1+\xi_{r_1}^2+\frac{\xi_{r_1}f(\xi_{r_1})}{p_{r_1}}\right],\beta_0=\frac{f(\xi_{r_1})}{p_{r_1}^2}[f(\xi_{r_1})+p_{r_1}\xi_{r_1}]>0.$$

将函数 $\frac{f(Z_{(r_k)})}{1-F(Z_{(r_k)})}$ 在点 ξ_{r_k} 处泰勒展开得：

$$\frac{f(Z_{(r_k)})}{1-F(Z_{(r_k)})}\approx\alpha_k+\beta_k Z_{(r_k)},$$

其中：

$$\alpha_k=\frac{f(\xi_{r_k})}{q_{r_k}}\left[1+\xi_{r_k}^2-\frac{\xi_{r_k}f(\xi_{r_k})}{q_{r_k}}\right],\beta_k=\frac{f(\xi_{r_k})}{q_{r_k}}[f(\xi_{r_k})-q_{r_k}\xi_{r_k}]>0.$$

将二元函数 $\frac{f(Z_{(r_{i+1})})-f(Z_{(r_i)})}{F(Z_{(r_{i+1})})-F(Z_{(r_i)})}$ 在点 $(\xi_{r_{i+1}},\xi_{r_i})$ 处泰勒展开得：

$$\frac{f(Z_{(r_{i+1})})-f(Z_{(r_i)})}{F(Z_{(r_{i+1})})-F(Z_{(r_i)})}\approx \varepsilon_i+\omega_i Z_{(r_i)}-\delta_i Z_{(r_{i+1})},$$

其中：

$$\varepsilon_i=\frac{f(\xi_{r_{i+1}})-f(\xi_{r_i})}{p_{r_{i+1}}-p_{r_i}}+\xi_{r_{i+1}}f(\xi_{r_{i+1}})\frac{\xi_{r_{i+1}}(p_{r_{i+1}}-p_{r_i})+[f(\xi_{r_{i+1}})-f(\xi_{r_i})]}{[p_{r_{i+1}}-p_{r_i}]^2}-$$
$$\xi_{r_i}f(\xi_{r_i})\frac{\xi_{r_i}(p_{r_{i+1}}-p_{r_i})+[f(\xi_{r_{i+1}})-f(\xi_{r_i})]}{[p_{r_{i+1}}-p_{r_i}]^2}.$$

$$\omega_i=f(\xi_{r_i})\frac{\xi_{r_i}(p_{r_{i+1}}-p_{r_i})+[f(\xi_{r_{i+1}})-f(\xi_{r_i})]}{[p_{r_{i+1}}-p_{r_i}]^2},$$

$$\delta_i=f(\xi_{r_{i+1}})\frac{\xi_{r_{i+1}}(p_{r_{i+1}}-p_{r_i})-[f(\xi_{r_{i+1}})-f(\xi_{r_i})]}{[p_{r_{i+1}}-p_{r_i}]^2}.$$

将二元函数$\frac{Z_{(r_{i+1})}f(Z_{(r_{i+1})})-Z_{(r_i)}f(Z_{(r_i)})}{F(Z_{(r_{i+1})})-F(Z_{(r_i)})}$在点$(\xi_{r_{i+1}},\xi_{r_i})$处泰勒展开得：

$$\frac{Z_{(r_{i+1})}f(Z_{(r_{i+1})})-Z_{(r_i)}f(Z_{(r_i)})}{F(Z_{(r_{i+1})})-F(Z_{(r_i)})}\approx a_i+b_i Z_{(r_i)}-c_i Z_{(r_{i+1})},$$

其中：

$$b_i=f(\xi_{r_i})\frac{[\xi_{r_i}^2-1](p_{r_{i+1}}-p_{r_i})+[\xi_{r_{i+1}}f(\xi_{r_{i+1}})-\xi_{r_i}f(\xi_{r_i})]}{[p_{r_{i+1}}-p_{r_i}]^2},$$

$$c_i=-f(\xi_{r_{i+1}})\frac{[1-\xi_{r_{i+1}}^2](p_{r_{i+1}}-p_{r_i})-[\xi_{r_{i+1}}f(\xi_{r_{i+1}})-\xi_{r_i}f(\xi_{r_i})]}{[p_{r_{i+1}}-p_{r_i}]^2},$$

$$a_i=\frac{\xi_{r_{i+1}}f(\xi_{r_{i+1}})-\xi_{r_i}f(\xi_{r_i})}{p_{r_{i+1}}-p_{r_i}}-$$
$$\xi_{r_{i+1}}f(\xi_{r_{i+1}})\frac{[1-\xi_{r_{i+1}}^2](p_{r_{i+1}}-p_{r_i})-[\xi_{r_{i+1}}f(\xi_{r_{i+1}})-\xi_{r_i}f(\xi_{r_i})]}{[p_{r_{i+1}}-p_{r_i}]^2}-$$
$$\xi_{r_i}f(\xi_{r_i})\frac{[\xi_{r_i}^2-1](p_{r_{i+1}}-p_{r_i})-[\xi_{r_{i+1}}f(\xi_{r_{i+1}})-\xi_{r_i}f(\xi_{r_i})]}{[p_{r_{i+1}}-p_{r_i}]^2}.$$

化简(2)、(3)式求得参数σ,μ的$AMLE$：$\hat{\sigma}=\frac{-D+\sqrt{D^2+4AE}}{2A}$，$\hat{\mu}=B-C\sigma$，其中：

$$M=(r_1-1)\beta_0-\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)(\omega_i-\delta_i)+(n-r_k)\beta_k+k,$$

$$B=\frac{1}{M}\Big[(r_1-1)\beta_0X_{(r_1)}-\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)(\omega_iX_{(r_i)}-\delta_iX_{(r_{i+1})})+$$
$$(n-r_k)\beta_kX_{(r_k)}+\sum_{i=1}^{k}X_{(r_i)}\Big],$$

$$C=\frac{1}{M}\Big[(r_1-1)\alpha_0+\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)\varepsilon_i-(n-r_k)\alpha_k\Big],$$

$$A=k+(r_1-1)C(\alpha_0-\beta_0C)+\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)(a_i+b_i-c_iC)-(n-r_k)C(\alpha_k+\beta_kC)-kC^2,$$

$$D=(r_1-1)(X_{(r_1)}-B)(\alpha_0-2\beta_0C)+\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)(b_iX_{(r_i)}-b_iB-c_iX_{(r_{i+1})}+c_iB)-$$
$$(n-r_k)(X_{(r_k)}-B)(\alpha_k+2\beta_kC)-2C\sum_{i=1}^{k}(X_{(r_i)}-B),$$
$$E=(r_1-1)\beta_0(X_{(r_1)}-B)^2+(n-r_k)\beta_k(X_{(r_k)}-B)^2+\sum_{i=1}^{k}(X_{(r_i)}-B)^2>0.$$

易知:

$$E\left(-\frac{\partial^2\ln L(\mu,\sigma)}{\partial\mu^2}\right)\approx\frac{M}{\sigma^2},\ E\left(-\frac{\partial^2\ln L(\mu,\sigma)}{\partial\mu\partial\sigma}\right)\approx\frac{MV_1}{\sigma^2},\ E\left(-\frac{\partial^2\ln L(\mu,\sigma)}{\partial\sigma^2}\right)\approx\frac{MV_2}{\sigma^2},$$

其中:

$$V_1=\frac{2}{M}\Big[(r_1-1)\beta_0E(Z_{(r_1)})-\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)[\omega_iE(Z_{(r_i)})-\delta_iE(Z_{(r_{i+1})})]+$$
$$(n-r_k)\beta_kE(Z_{(r_k)})+\sum_{i=1}^{k}E(Z_{(r_i)})\Big]-C,$$
$$V_2=\frac{1}{M}\Big\{3\Big[(r_1-1)\beta_0E(Z^2_{(r_1)})+(n-r_k)\beta_kE(Z^2_{(r_k)})+\sum_{i=1}^{k}E(Z^2_{(r_i)})\Big]-$$
$$2\Big[(r_1-1)\alpha_0E(Z_{(r_1)})+\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)[b_iE(Z_{(r_i)})-c_iE(Z_{(r_{i+1})})]-$$
$$(n-r_k)\alpha_kE(Z_{(r_k)})\Big]-k-\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)a_i\Big\}.$$

由此: $Var(\hat\mu)\approx\frac{\sigma^2}{M}\Big[\frac{V_2}{V_2-V_1^2}\Big]$, $Var(\hat\sigma)\approx\frac{\sigma^2}{M}\Big[\frac{1}{V_2-V_1^2}\Big]$, $Cov(\hat\mu,\hat\sigma)\approx-\frac{\sigma^2}{M}\Big[\frac{V_1}{V_2-V_1^2}\Big]$.

例[1]:考察以下由 Gupta(1952)在考察 10 只老鼠样本死亡时间所得到的头 7 个样本数据:1.613,1.644,1.663,1.732,1.740,1.763,1.778. 数据服从正态分布 $N(\mu,\sigma^2)$,如果只知道 $x_{(r_1)}=1.613$,$x_{(r_2)}=1.644$ 及 $x_{(r_3)}=1.778$ 这 3 个数据,也即:$n=10,r_1=1,r_2=2,r_3=7,k=3$,采用本文方法可以得到:

$$p_{r_1}=\frac{1}{11},p_{r_2}=\frac{2}{11},p_{r_3}=\frac{7}{11},\xi_{r_1}=-1.34,\xi_{r_2}=-0.91,\xi_{r_3}=-0.35.$$
$$\frac{f(Z_{(r_1)})}{F(Z_{(r_1)})}\approx\alpha_0-\beta_0Z_{(r_1)},$$
$$\frac{f(Z_{(r_3)})}{1-F(Z_{(r_3)})}\approx\alpha_3+\beta_3Z_{(r_3)},$$
$$\frac{f(Z_{(r_3)})-f(Z_{(r_2)})}{F(Z_{(r_3)})-F(Z_{(r_2)})}\approx\varepsilon_2+\omega_2Z_{(r_2)}-\delta_2Z_{(r_3)},$$
$$\frac{Z_{(r_3)}f(Z_{(r_3)})-Z_{(r_2)}f(Z_{(r_2)})}{F(Z_{(r_3)})-F(Z_{(r_2)})}\approx a_2+b_2Z_{(r_2)}-c_2Z_{(r_3)},$$

其中: $\alpha_0=0.68239$, $\beta_0=0.83724$, $\alpha_3=0.78611$, $\beta_3=0.70532$,
$\varepsilon_2=1.02222$, $\omega_2=-0.38553$, $\delta_2=0.49153$, $a_2=1.13978$,
$b_2=0.37414$, $c_2=-0.05008$, $M=8.62423$, $B=1.71937$,

$$C=0.20066,\quad A=8.41663,\quad D=-0.2397,\quad E=0.02771.$$

于是可得参数 σ 的 AMLE 为：$\hat{\sigma}=\dfrac{-D+\sqrt{D^2+4AE}}{2A}=0.07336$,

进而可得参数 μ 的 AMLE 为：$\hat{\mu}=B-C\hat{\sigma}=1.70465$.

为了作对照下面列出利用前 7 个样本数据所得的参数的 BLUE,GLUE 及 AMLE：

BLUE：　$\hat{\mu}=1.746$,　$\hat{\sigma}=0.091$,

GLUE：　$\hat{\mu}=1.748$,　$\hat{\sigma}=0.094$,

AMLE：　$\hat{\mu}=1.742196$,　$\hat{\sigma}=0.079129$.

参考文献：

[1] BALAKRISHNAN N, CLIFFORD COHEN A. Order Statistic and Inference [M]. Boston: Estimation Methods Academic Press, 1991, 174-175,233-236.

[2] FEI He-liang, KONG Fan-hui. Interval estimations for one — and two — parameter exponential distributions under multiply Type I censoring[J]. Commun Statist-theory Meth. 1994, 23(6):1717-1733.

[3] FEI He-liang,KONG Fan-hui, TANG Ying-cai. Estimations for two-parameter Weibull distributions and extreme-value distributions under multiply TYPE I censoning[J]. Commun Statist-theory Meth. 1995, 24(24): 2087-2104.

[4] KONG Fan-hui, FEI He-liang. Limit Theorems for the maximum likehood estimalors under multiply Type I censoring[J]. Ann Inst Statist Math. 1996, 48(4): 731-755.

Approximate Maximum Likelihood Estimations for Lognormal Distribution

WANG Rong-hua, XU Xiao-ling, FEI He-liang

(College of Mathmatical Science, Shanghai Teachers University, Shanghai, 200234, China)

Abstract: Gives approximate maxmum likelihood estimations of two-parameter lognormal distributions under multiply Type I censoring.

Key words: Type I censoring; two-parameter lognormal distribution; approximate maximum likelihood estimate