指数分布参数具任给可靠度和精度的置信区间*

高　峰
(淮阴工学院 计算科学系,江苏 淮安　223001)

摘　要:利用两阶段抽样,构造了定数截尾试验下指数分布参数的一个置信区间,它同时满足预先任意指定的可靠度和精度。

关键词:置信区间;指数分布;定数截尾试验;两阶段抽样
中图分类号:O211.2　　**文献标识码**:A　　**文章编号**:1671-5322(2004)04-00013-03

区间估计有两个要素:可靠度和精度,可靠度由置信水平表示,设 $X_1,\cdots X_n$ 为来自总体 X 的样本,$\theta\in\Theta$ 为总体的未知参数,若有统计量 $\underline{\theta}=\underline{\theta}(X_1,\cdots X_n)$ 和 $\bar{\theta}=\bar{\theta}(X_1,\cdots X_n)$,使得对于给定的 $\alpha(0<\alpha<1)$,有 $P_\theta(\underline{\theta}\leqslant\theta\leqslant\bar{\theta})\geqslant 1-\alpha$,则称 $[\underline{\theta},\bar{\theta}]$ 为 θ 的置信水平为 $1-\alpha$ 的置信区间。给区间估计规定一个置信水平,也就保证了其可靠度达到一定的要求。置信区间的精度是通过置信区间包含错误值的概率来刻画的,即 $\int_{(\theta'\neq\theta)}P_\theta(\underline{\theta}\leqslant\theta'\leqslant\bar{\theta})\mathrm{d}\theta'$ 越小,置信区间的精度就越高。对于置信区间精度的要求导致两个方面的研究,一个是在具同一可靠度的所有置信区间中,求 $\int_{(\theta'\neq\theta)}P_\theta(\underline{\theta}\leqslant\theta'\leqslant\bar{\theta})\mathrm{d}\theta'$ 最小的置信区间,这个要求通常被更确切地表述为:设 $[\underline{\theta},\bar{\theta}]$ 为 θ 的置信水平为 $1-\alpha$ 的置信区间,若对于任何其它的具有同一置信水平的置信区间 $[\underline{\theta}^*,\bar{\theta}^*]$,以及任何 $\theta'\neq\theta,\theta'\in\Theta$ 有 $P_\theta(\underline{\theta}\leqslant\theta'\leqslant\bar{\theta})\leqslant P_\theta(\underline{\theta}^*\leqslant\theta'\leqslant\bar{\theta}^*)$,则称 $[\underline{\theta},\bar{\theta}]$ 为 θ 的置信水平为 $1-\alpha$ 的一致最精确的置信区间。另外一个颇具实际意义的要求是,在其可靠度达到一定要求下,希望其精度达到预先指定的要求,这就是具任给可靠度和精度的置信区间[1]。

在分析产品的可靠性时,常常需要进行寿命试验,在实际应用中,寿命试验常采用截尾寿命试验;指数分布是可靠性分析中非常重要的寿命分布之一,故基于截尾数据下的指数分布参数的统计推断是有实用价值的。本文将证明:在定数截尾试验下指数分布参数具任给可靠度和精度的置信区间问题。

对于具任给可靠度和精度的置信区间问题的研究,所见文献中主要是关于正态分布总体参数的:对于正态总体均值的估计,Dantzig[1] 指出在一次抽样数据下(也称为固定样本)同时满足可靠度和精度的置信区间是不存在的,Stein[1] 建议用两阶段抽样并得到了结果;王海贤[2] 借助两阶段抽样方法解决了正态总体方差同时满足可靠度和精度的置信区间。而对于截尾寿命试验情形下指数分布参数具任给可靠度和精度的置信区间问题的研究,尚未见之于文献。

1　引理及问题的进一步明确

设总体 X 服从参数为 $\lambda(\lambda>0)$ 的指数分布,其分布函数为

$$F(x,\lambda)=1-e^{-\lambda x},x>0 \tag{1-1}$$

设 $X_1,\cdots X_n$ 为来自总体(1-1)的样本,在定数为 r 的截尾寿命试验下仅能观察到前面 r 个观察值:$0\leqslant t_1\leqslant t_2\leqslant\cdots\leqslant t_r$,则总试验时间为 $T=\sum\limits_{i=1}^{r}t_i+(n-r)t_r$,由文献[3]知,$2\lambda T\sim\chi^2(2r)$,故 λ 的 $1-\alpha$ 的置信区间为:

$$\left[\frac{\chi^2_{1-\alpha/2}(2r)}{2T},\frac{\chi^2_{\alpha/2}(2r)}{2T}\right] \tag{1-2}$$

* 收稿日期:2004-09-11
作者简介:高　峰(1963-),男,江苏东海人,淮阴工学院讲师,主要研究方向为数理统计研究。

这个置信区间可靠度达到了预先指定的要求,但是不能达到预先指定的精度。

关于置信区间的精度,根据文献[1]有如下的引理一。

引理一 设$[\underline{\theta},\bar{\theta}]$为 θ 的置信区间,则有

$$E_{\theta}(\bar{\theta}-\underline{\theta})=\int_{(\theta'\neq\theta)}P_{\theta}(\underline{\theta}\leqslant\theta'\leqslant\bar{\theta})\mathrm{d}\theta'$$

因此,置信区间的精度可以用置信区间的平均宽度来衡量,置信区间的平均宽度越小,则置信区间的精度越高,于是本文的研究问题是:在定数截尾寿命试验模型下求指数分布产品的失效率 λ 的一个置信区间$[\underline{\lambda},\bar{\lambda}]$,使它满足

$$\begin{cases}P_{\lambda}(\underline{\lambda}\leqslant\lambda\leqslant\bar{\lambda})\geqslant 1-\alpha\\ E_{\lambda}(\bar{\lambda}-\underline{\lambda})\leqslant\varepsilon\end{cases}\tag{1-3}$$

其中 $1-\alpha$,ε 为预先指定的置信水平和平均宽度。

引理二 (1)令 $H(r)=\dfrac{\chi^2_{\beta/2}(r)}{r}$,$r$ 为正整数,则 $H(r)$是单调减少函数。其中$\dfrac{\beta}{2}=0.25,0.10,0.05,0.025,0.01,0.005$,$\chi^2_{\beta/2}(r)$表示自由度为 r 的χ^2 分布的上 $\beta/2$ 分位值。

(2)令 $G(r)=\dfrac{\chi^2_{1-\beta/2}(r)}{r}$,$r$ 为正整数,则$G(r)$是单调增加函数。其中$\dfrac{\beta}{2}$的取值如上。

证(1)当 $1\leqslant r\leqslant 45$ 时,运用 χ^2 分布表验证可得,$H(r)$是单调减少函数;

当 $r\geqslant 46$ 时,可以看作大样本来处理。于是 $\chi^2(r)$分布可以用其极限分布 $N(r,2r)$来代替,记$\chi^2(r)$同时表示服从 $\chi^2(r)$分布的随机变量,Z_{β} 表示 $N(0,1)$的上 β 分位值,则由 $P\left\{\dfrac{\chi^2(r)-r}{\sqrt{2r}}>Z_{\beta/2}\right\}=\dfrac{\beta}{2}$ 得:$P\{\chi^2(r)>r+\sqrt{2r}Z_{\beta/2}\}=\dfrac{\beta}{2}$,因此,$\chi^2_{\beta/2}(r)=r+\sqrt{2r}Z_{\beta/2}$,于是 $H(r)=1+\dfrac{\sqrt{2r}Z_{\beta/2}}{\sqrt{r}}$,而当 $\beta/2\leqslant 0.25$ 时,$Z_{\beta/2}>0$,故 $H(r)$ 是单调减少函数。同理可证(2) 成立。

2 主要结果

下文中的 γ、β 的取法规定为:先在 $0.25,0.10,0.05,0.025,0.01,0.005$ 中选取一数作为$\beta/2$值,再由$(1-\gamma)(1-\beta)=1-\alpha$ 定出γ 值(显然 $\beta\leqslant\alpha$,$\gamma\leqslant\alpha$)。

下面运用两阶段抽样解决本问题。

第一步,从总体(1-1)中抽取独立同分布的随机样本 $X_1,\cdots X_{n_1}$ 进行试验,试验的截尾数为 r_1,设试验所得的定数截尾数据为 $t_1,\cdots t_{r_1}$,记 T_1 表示总试验时间,则 λ 的 $1-\gamma$ 置信上限为$\bar{\lambda}_1(X_1,\cdots X_{n_1})=\dfrac{\chi^2_{\beta/2}(2r_1)}{2T_1}$,它的观察值记为$\bar{\lambda}_1(t_1,\cdots t_{r_1})$。

第二步,首先确定第二阶段抽样试验的截尾数 r_2,由引理二知 $H(2r_2)=\dfrac{\chi^2_{\beta/2}(2r_2)}{2r_2}$为单调减少函数,故不等式方程

$$H(2r_2)\leqslant 1+\frac{\varepsilon}{2\bar{\lambda}_1(t_1,\cdots t_{r_1})}\tag{2-1}$$

必有唯一确定的最小整数解 $r_2^{<1>}$,由引理二又知,$G(r)=\dfrac{\chi^2_{1-\beta/2}(r)}{r}$为单调增加函数,故不等式方程

$$G(2r_2)\geqslant 1-\frac{\varepsilon}{2\bar{\lambda}_1(t_1,\cdots t_{r_1})}\tag{2-2}$$

必有唯一确定的最小整数解 $r_2^{<2>}$,取 $r_2=\mathrm{Max}(r_2^{<1>},r_2^{<2>})$。

其次,再从总体(1-1)中抽取独立同分布的随机样本 $X_{n_1+1},\cdots X_{n_1+n_2}$,$(n_2\geqslant r_2)$,作定数为 r_2 的截

尾寿命试验,记 T_2 表示第二阶段抽样试验的总试验时间,令

$$\underline{\lambda}_2 = \frac{\chi^2_{1-\beta/2}(2r_2)}{2T_2},\bar{\lambda}_2 \frac{\chi^2_{\beta/2}(2r_2)}{2T_2}$$

由 $2\lambda T_2 \sim \chi^2(2r_2)$ 知,$E(2\lambda T_2) = 2r_2$。因此

$$E_\lambda[\bar{\lambda}_2 - \lambda] = E_\lambda\left[\frac{\chi^2_{\beta/2}(2r_2)}{2T_2} - \lambda\right] = \lambda\left[\frac{\chi^2_{\beta/2}(2r_2)}{2r_2} - 1\right] \leqslant \bar{\lambda}_1(t_1,\cdots t_{r_1})[H(2r_2) - 1] \leqslant \varepsilon/2$$

$$E_\lambda[\lambda - \underline{\lambda}_2] = E_\lambda\left[\lambda - \frac{\chi^2_{1-\beta/2}(2r_2)}{2T_2}\right] = \lambda\left[1 - \frac{\chi^2_{1-\beta/2}(2r_2)}{2r_2}\right] \leqslant (\bar{\lambda}_1(t_1,\cdots t_{r_1})[1 - G(2\gamma_2)] \leqslant \varepsilon/2$$

于是,$E_\lambda[\bar{\lambda}_2 - \underline{\lambda}_2] = E_\lambda[\bar{\lambda}_2 - \lambda + \lambda - \underline{\lambda}_2] = E_\lambda[\bar{\lambda}_2 - \lambda] + E_\lambda[\lambda - \underline{\lambda}_2] \leqslant \varepsilon$,
故随机区间满足预先指定的精度要求。

又
$$\begin{aligned} P_\lambda(\underline{\lambda}_2 \leqslant \lambda \leqslant \bar{\lambda}_2) = {} & P_\lambda[\underline{\lambda}_2 \leqslant \lambda \leqslant \bar{\lambda}_2 \mid \lambda \leqslant \bar{\lambda}_1(X_1,\cdots X_{n_1})]P_\lambda[\lambda \leqslant \bar{\lambda}_1(X_1,\cdots X_{n_1})] + \\ & P_\lambda[\underline{\lambda}_2 \leqslant \lambda \leqslant \bar{\lambda}_2 \mid \lambda > \bar{\lambda}_1(X_1,\cdots X_{n_1})]P_\lambda[\lambda > \bar{\lambda}_1(X_1,\cdots X_{n_1})] \geqslant \\ & P_\lambda[\underline{\lambda}_2 \leqslant \lambda \leqslant \bar{\lambda}_2 \mid \lambda \leqslant \bar{\lambda}_1(X_1,\cdots X_{n_1})]P_\lambda[\lambda \leqslant \bar{\lambda}_1(X_1,\cdots X_{n_1})] = \\ & P_\lambda(\chi^2_{1-\beta/2}(2r_2) \leqslant 2\lambda T_2 \leqslant \chi^2_\beta(2r_2)(1-\gamma) = (1-\beta)(1-\gamma) = 1-\alpha \end{aligned}$$

故随机区间$[\underline{\lambda}_1,\bar{\lambda}_2]$满足预先指定的可靠度要求。综合得如下结论

定理　设 r_2 为(2-1)和(2-2)组成的不等式方程组的最小整数解,T_2 为定数为 r_2 的截尾试验的总试验时间,则$\left[\frac{\chi^2_{1-\beta/2}(2r_2)}{2T_2},\frac{\chi^2_{\beta/2}(2r_2)}{2T_2}\right]$就是分布(1-1)的参数 λ 的满足条件(1-3)的置信区间。

例　设电池的寿命服从指数分布(1-1),为得到 θ 的置信水平为 $1-\alpha=0.90$,$\varepsilon=0.001$ 的置信区间,进行两阶段抽样截尾试验,先随机地取 20 只电池投入寿命试验,规定试验进行到其中有 6 只失效时停止,测得失效时间(h)为 119,138,147,155,166,172,我们取 $\beta=0.05$,$\gamma=0.053$ 利用单调函数的线性插值,计算得 $\chi^2_{0.053}(12)=20.877$,从而 $\bar{\lambda}_1=0.00316$,不等式方程组(2-1)、(2-2)为 $H(2r_2)\leqslant 1.158$,$G(2r_2)\geqslant 0.8418$,其最小整数解为 $r_3=154$,再随机地取不少于 154 只电池投入寿命试验,规定进行到其中有 154 只失效时结束试验,则 $[\underline{\lambda}_2,\bar{\lambda}_2] = \left[\frac{\chi^2_{0.972}(308)}{2T_2},\frac{\chi^2_{0.025}(308)}{2T_2}\right] = \left[\frac{259.354}{2T_2},\frac{356.646}{2T_2}\right]$ 即为所求。

参考文献:

[1] 陈希孺.数理统计引论[M].北京 科学出版社,1999.
[2] 王海贤,曾建军,陈桂景.Two-Stage Confidence Intervals for the Variance of a Normal Distribution 应用概率统计[J].2003,19(2):118-124.
[3] 茆诗松,王玲玲.可靠性统计[M].上海 华东师范大学出版社,1984.

The Confidence Interval with Prescribed Coverage Probability and Prescribed Width for the Exponential Distribution Parameter

GAO Feng
(Department of Computing, Huaiyin Institute of Technology, Jiangsu Huaian　223001, China)

Abstract: In this paper, we design a confidence interval for the exponential distribution parameter based on Type Ⅱ censored samples by two-stage procedures. This confidence interval has prescribed width and prescribed coverage probability.

Keywords: confidence interval; exponential distribution; Type Ⅱ censored sampling; two-stage procedures