第22卷 第2期 华侨大学学报(自然科学版) Vol.22 No.2
2001年4月 Journal of Huaqiao University (Natural Science) Apr.2001

文章编号 1000-5013(2001)02-128-05

以熵为测度的无失效抽样检验法

吴 绍 敏

(华侨大学经济管理学院, 泉州 362011)

摘要 以信息熵为测度, 建立新的寿命检验方法. 由于应用的范围不爱限制, 因而比较沏底地解决了产品寿命的检验问题. 其所研究的内容, 主要有寿命信息分析与初估公式的建立、信息熵的公式、一次与逐次检验的方法及其试例.

关键词 信息熵, 无失效数据, 抽样检验

中图分类号 O 212.2 文献标识码 A

科技不断进步, 产品寿命愈来愈长, 获得其失效数据的时间长, 花费也大. 许多抽样检验法, 都是建立在有失效数据的基础上制定的, 因而无法适应目前的需要. 所以, 在无失效数据的情况下, 建立抽样检验法就显得十分迫切. 1979年, Martz等人在文[1]中给出一种Bayes无失效检验法, 其应用范围限制在单指数分布范围内. 他们的研究起了个好的开头, 引起许多学者的重视. 本文以信息熵为测度, 建立一种应用范围不受任何限制的检验法. 它对产品寿命检验的问题, 解决的比较好.

1 寿命信息分析

设产品寿命 $T\sim F(t)$, 确定 m 个观测点 $t_1<t_2<\cdots<t_m$, 抽取 n_i 个样品独立寿命试验到 t_i $(i=\overline{1,m})$ 无失效. 停止试验, 称 $(t_i,n_i)_{1\le i\le m}$ 为无失效数据或"样本". 它可提供如下的寿命信息, 记 $R_i=P(T>t_i)$, $S_i=\sum_{k=i}^{m}n_k(i=\overline{1,m})$. (1) (t_i,n_i) 只能提供 $T>t_i$ 的寿命信息, 无法提供 $T>t_{i+1}(i=\overline{1,m-1})$ 的寿命信息; 反之, (t_{i+1},n_{i+1}) 既可提供 $T>t_{i+1}$ 的寿命信息, 又可提供 $T>t_i$ 的寿命信息. (2) 由性质(1), 知有 S_i 个样品的寿命 $T>t_i(i=\overline{1,m})$. (3) 因 $t_i<t_{i+1}$, 故 $R_i=P(T>t_i)>P(T>t_{i+1})=R_{i+1}$. (4) 因 $0<R_i<1$, 可应用Bayes假设, 视 R_i 为 $r.v.$, 其分布密度函数为 $\pi(R_i)=U(0,1)$, $(i=\overline{1,m})$ 且独立.

1.2 可靠度的初估公式

综合"样本"所提供的4个寿命信息, 可知在 $(t_1,t_2,\cdots,t_m)$ 作一次观测, 有 S_1 个样品在 t_1 不失效, 有 S_2 个样品在 t_2 不失效……有 S_m 个样品在 t_m 不失效. 所以, 在点 $(t_1,t_2,\cdots,t_m)$ 不失效

收稿日期 2000-11-09 作者简介 吴绍敏(1959-), 男, 教授

基金项目 福建省自然科学基金资助项目

的概率为 $L(R\ S)=\prod_{i=1}^{m} R_i^{S_i}$, $R_1>R_2>\cdots>R_m$, 其中 $R=(R_1,R_2,\cdots,R_m)$, R_i 是未知参数, S_i 是在 t_i 点不失效的样品数, 非 $r.v.$, 但可视为 $r.v.$ 的特例．因此, 可把 $L(R,S)$ 视为 R 与 $S=(S_1,S_2,\cdots,S_m)$ 的联合似然函数．因 $\pi(R_i)=U(0,1)$, 故 R 的联合密度为 $\pi(R_1,R_2,\cdots,R_m)=\begin{cases}1 & 当\ R_1>R_2>\cdots>R_m\\ 0 & 其它\end{cases}$．这是因 $R_i\sim U(0,1)$, 即 $\pi(R_i)=\begin{cases}1 & 0<R_i<1\\ 0 & 其它\end{cases}$ 且独立$(i=\overline{1,m})$,
$V(D)=m!\int_0^{R_{m-1}}\mathrm{d}R_m\int_0^{R_{m-2}}\mathrm{d}R_{m-1}\cdots\int_0^{R_2}\mathrm{d}R_3\int_0^{R_3}\mathrm{d}R_2\int_0^1\mathrm{d}R_1=m!\dfrac{1}{m!}=1$. 根据 Bayes 定理, 得 R 的后验密度为

$$f(R|S)=\prod_{i=1}^{m}R_i^{S_i}\Big\backslash\int_D\prod_{i=1}^{m}R_i^{S_i}\mathrm{d}R_i\qquad D=\{R:R_1>R_2>\cdots>R_m\}.\tag{1}$$

定理 1　R_1 的后验密度为

$$f(R_1|S)=\left(\sum_{i=1}^{m}S_i+m\right)R_1^{\left(\sum_{i=1}^{m}S_i+m-1\right)}.\tag{2}$$

在二次损失下, R_i 的 Bayes 估计为

$$\hat{R}_1=\left(\sum_{i=1}^{m}S_i+m\right)\Big/\left(\sum_{i=1}^{m}S_i+m+1\right).\tag{3}$$

证明　由式(1) 记 $f(R_1|S)=I_1/W_m$, 其中 $W_m=\int_D\prod_{i=1}^{m}R_i^{S_i}\mathrm{d}R_i$,

$$I_1=R_1^{S_1}\int_0^{R_1}R_2^{S_2}\mathrm{d}R_2\int_0^{R_2}R_3^{S_3}\mathrm{d}R_3\cdots\int_0^{R_{m-1}}R_m^{S_m}\mathrm{d}R_m.$$

经过$(m-1)$次积分, 得

$$I_1=R_1^{\left(\sum_{i=1}^{m}S_i+m-1\right)}\Big/(S_m'+1)(S_{m-1}+S_m+2)\cdots\left(\sum_{i=2}^{m}S_i+m-1\right).$$

因 $\int_0^1 I_1\mathrm{d}R_1=1/(S_m+1)(S_{m-1}+S_m+2)\cdots\left(\sum_{i=1}^{m}S_i+m\right)=W_m$, 故式(2)成立．在二次损失下,

$$\hat{R}_1=\int_0^1 R_1f(R_1|S)\mathrm{d}R_1=\left(\sum_{i=1}^{m}S_i+m\right)\int_0^1 R_1^{\left(\sum_{i=1}^{m}S_i+m\right)}\mathrm{d}R_1=$$

$$\left(\sum_{i=1}^{m}S_i+m\right)\Big/\left(\sum_{i=1}^{m}S_i+m+1\right).$$

因$(t_i,n_i)_{1\le i\le k-1}$无法提供 $T>t_k$ 的寿命信息, 在估计 $R_k(k=\overline{2,m})$ 时不起作用．因此可利用 $f(R(k)|S(k))=\prod_{i=k}^{m}R_i^{S_i}\Big\backslash\int_{D(m-k+1)}\prod_{i=k}^{m}R_i^{S_i}\mathrm{d}R_i$, 其中 $R(k)=(R_k,R_{k+1},\cdots,R_m)$, $S(k)=(S_k,S_{k+1},\cdots,S_m)$, $D(m-k+1)=\{R(k):R_k>R_{k+1}>\cdots>R_m\}$. 可得

推论 1　R_k 的后验密度为

$$f(R_k|S(k))=\left(\sum_{i=k}^{m}S_i+m+1-k\right)R_k^{\left(\sum_{i=k}^{m}S_i+m-k\right)}.\tag{4}$$

在二次损失下

$$\hat{R}_k=\left(\sum_{i=k}^{m}S_i+m+1-k\right)\Big/\left(\sum_{i=k}^{m}S_i+m+2-k\right).\tag{5}$$

1.3　$\hat{R}_k$ 的性质

(1) $\hat{R}_k=1-1/\left(\sum_{i=k}^{m}S_i+m+2-k\right)>1-1/\left(\sum_{i=k+1}^{m}S_i+m+1-k\right)=\hat{R}_{k+1}$. (2) $\hat{R}_k$ 是 n_i 和 m 的

函数, m 与 n_i 不是 $r.v.$. 故 $\hat{R}_k$ 不是统计量. 当 m 或 n_i 增大时, $\hat{R}_k$ 也会增大. (3) $\hat{R}_m=\frac{n_m+1}{n_m+2}$, 当 $n_m=1$ 时, $\hat{R}_m=2/3$, 故 $\hat{R}_k\geq 2/3$. (4) $\hat{R}_k$ 关于分布自由, 应用范围不受限制.

2 信息熵

一只样品在 t_i 点进行一次寿命试验, 其失效数可用 $r.v.X$ 表示. $X=\begin{cases}0 & 表示不失效\\ 1 & 表示失效\end{cases}$, $R_i=P(T>t_i)=P(X=0)(i=\overline{1,m})$. $X\sim\begin{pmatrix}0, & 1\\ R_i, & 1-R_i\end{pmatrix}$, 其信息熵 $H(R_i)=-R_i\ln R_i(e^{-1}\leq R_i\leq 1)$, 它表示事件 $\{X=0\}$ 的不确定性程度的信息量. 当 $R_i=1$, $H(R_i)=0$, 表示 $\{X=0\}$ 的不确定性的信息量为 0; 当 $R_i=e^{-1}$ 时, $H(R_i)$ 达到最大, 此时 $\{X=0\}$ 的不确定性信息量最大. 因此, 在 t_i 点独立寿命试验观测 n_i 次, 其不失效信息量为 $H(R_i)=-n_iR_i\ln R_i$, $(e^{-1}\leq R_i\leq 1,\ i=\overline{1,m})$. 无失效数据组 $(t_i,n_i)_{1\leq i\leq m}$ 的信息量为 $H(R_1,R_2,\cdots,R_m)=-\sum_{i=1}^{m}n_iR_i\ln R_i$. 实际上在 t_i 点观测不止是 n_i 个样品无失效, 而是 S_i 个样品无失效. 因此, 应改用加权信息熵 $H(R_1,R_2,\cdots,R_m)=-\sum_{i=1}^{m}S_iR_i\ln R_i/\sum_{i=1}^{m}S_i$. 比较合理. 因当 $R_i=e^{-1}(i=\overline{1,m})$ 时 $H(R_1,R_2,\cdots,R_m)=e^{-1}$; 当 $R_i=1\ (i=\overline{1,m})$ 时, $H(R_1,R_2,\cdots,R_m)=0$, 即 $0\leq H(R_1,R_2,\cdots,R_m)\leq e^{-1}$. 加以标准化, 得

$$H(m)=H(e_1,e_2,\cdots,e_m)=-e\sum_{i=1}^{m}S_iR_i\ln R_i/\sum_{i=1}^{m}S_i. \tag{6}$$

则有 $0\leq H(m)\leq 1$. 于是, 称式(6)为无失效数据组 $(t_i,n_i)_{1\leq i\leq m}$ 的标准化信息熵. 当 m 越大, 无失效的样品越多, $R_1,R_2,\cdots,R_m$ 就越大, 信息量就越小, 产品的可靠性就越高; 反之也对. 当 m 固定时, 整个寿命试验系统就固定下来. 若变动 m, 则试验系统随之变动. 因此, $H(m)$ 也随之改变. 显然, $H(m+1)<H(m)$.

虽然 $(t_i,n_i)_{1\leq i\leq m}$ 不是 $r.v.$, 但样品是随机抽取的. 从这个角度看, $H(m)$ 带有概率性质, 可以作为抽样检验的一个指标.

3 检验方法的制定

3.1 一次检验法

因 $0<H(m)<1$, 故预先设置的两个界限 $\alpha,\beta(0<\alpha<\beta<1)$, 其大小可由生产方与接收方共同协商确定. 当 m 固定, 可由式(3), (5) 计算 $\hat{R}_k(k=\overline{1,m})$ 及 $\hat{H}(m)$. 若 $\hat{H}(m)<\alpha$ 则接收, 若 $\hat{H}(m)>\beta$ 则拒收; 若 $\alpha<\hat{H}(m)<\beta$, 则追加试验后作决定.

3.2 逐次检验法

当 m 很大时, 要做较多的试验, 时间长化费多. 为节约起见, 可采用逐步检验的方法. 事先, 必须对试验作周密安排; 选择的观测点 $t_i(i=\overline{1,m})$ 要适当地长且均匀, n_i 的个数也均匀一些. 尔后, 进行逐步试验与检验. 若 $\hat{H}(1)<\alpha$ 则接收, 若 $\hat{H}(1)>\alpha$ 则取 $m=2$. 计算 $\hat{R}_1\hat{R}_2$, 若 $\hat{H}(2)<\alpha$ 则接收; 若 $\hat{H}(2)>\alpha$, 取 $m=3$, 等等. 如此继续下去, 若 $H(m)<\alpha$ 则接收; 若 $\hat{H}(m)>\beta$ 则拒收. 若 $\alpha<\hat{H}(m)<\beta$, 则无法判之, 应作追加试验再作判断.

4　试例[2]

从一批某型号的微型发动机中, 随机抽取一只样品．因这种发动机设计复杂, 价格昂贵, 寿命试验是一台一台地进行．假如在给定时刻该发动机没有失效, 应立即停止试验, 并进行维修直到该发动机被认定恢复如新．然后, 再进行上述定时截尾寿命试验．如此寿命试验, 累计 51 次．即相当样品量 $n=51$, 结果无一台失效．其定时截尾时间列于表 1, 其中的数据确实提供了发动机的寿命信息．比如, 各台发动机的寿命, 要大于各自的截尾寿命试验时间.

表 1　试验的截尾时间(s)

250.35	909.77	100.18	1 450.30	110.00	1 451.70	104.91	150.02	110.00	110.00	110.14	110.06	110.18
109.97	110.11	119.94	110.17	110.09	110.13	110.07	110.05	110.14	110.07	110.14	109.93	109.99
109.95	109.97	109.95	849.94	870.03	180.08	180.00	180.09	115.01	180.24	179.94	179.98	179.96
180.00	150.07	150.16	190.36	130.15	850.00	150.00	850.00	783.00	783.00	783.00	783.00	

将表 1 数据作一下处理．按数据从小到大的排序, 相同者归一组, 不同者归不同组, 可得试验"样本"$(t_i, n_i)_{1\le i\le 39}$. 因此, 有(100.18, 1), (104.91, 1), (109.93, 1), (109.95, 2), (109.97, 2), (109.99, 1), (110.00, 3), (110.05, 1), (110.06, 1), (110.07, 2), (110.09, 1), (110.11, 1), (110.13, 1), (110.14, 3), (110.17, 1), (110.18, 1), (115.01, 1), (119.94, 1), (130.50, 1), (150.00, 1), (150.02, 1), (150.07, 1), (150.16, 1), (179.94, 1), (179.96, 1), (179.98, 1), (180.00, 2), (180.08, 1), (180.09, 1), (180.24, 1), (190.36, 1), (250.35, 1), (783.00, 4), (849.94, 1), (850.00, 2), (870.03, 1), (909.71, 1), (1 450.30, 1), (1 451.70, 1).

4.1　一次检验

取 $\alpha=0.05$, $\beta=0.1$, 计算 $\hat{R}_k(k=\overline{1,39})$ 和 $\hat{H}(39)=e\sum_{i=1}^{39}S_i\hat{R}_i\ln\hat{R}_i/\sum_{i=1}^{39}S_i=0.012\,9$. 显然, $\hat{H}(39)<\alpha$, 故可接收．

这个试验并没有预先进行周密的安排, 而是随意地进行试验. 现将 $\hat{H}_k(k=\overline{1,39})$ 全部算出来, 有 $\hat{H}(1)=0.734\,8$, $\hat{H}(2)=0.520\,5$, $\hat{H}(3)=0.389\,0$, $\hat{H}(4)=0.259\,0$, $\hat{H}(5)=0.188\,9$, $\hat{H}(6)=0.161\,9$, $\hat{H}(7)=0.117\,3$, $\hat{H}(8)=0.104\,7$, $\hat{H}(9)=0.095\,1$, $\hat{H}(10)=0.081\,4$, $\hat{H}(11)=0.074\,0$, $\hat{H}(12)=0.068\,1$, $\hat{H}(13)=0.063\,0$, $\hat{H}(14)=0.051\,9$, $\hat{H}(15)=0.048\,1$, $\hat{H}(16)=0.045\,1$, $\hat{H}(17)=0.042\,9$, $\hat{H}(18)=0.039\,9$, $\hat{H}(19)=0.037\,7$, $\hat{H}(20)=0.035\,6$, $\hat{H}(21)=0.033\,6$, $\hat{H}(22)=0.031\,8$, $\hat{H}(23)=0.030\,2$, $\hat{H}(24)=0.028\,7$, $\hat{H}(25)=0.027\,2$, $\hat{H}(26)=0.025\,9$, $\hat{H}(27)=0.023\,9$, $\hat{H}(28)=0.022\,6$, $\hat{H}(29)=0.021\,6$, $\hat{H}(30)=0.020\,6$, $\hat{H}(31)=0.019\,8$, $\hat{H}(32)=0.018\,9$, $\hat{H}(33)=0.016\,4$, $\hat{H}(34)=0.015\,7$, $\hat{H}(35)=0.014\,8$, $\hat{H}(36)=0.014\,2$, $\hat{H}(37)=0.013\,7$, $\hat{H}(38)=0.013\,3$, $\hat{H}(39)=0.012\,9$.
从计算结果发现, $\hat{H}(k)$ 是 k 的单调下降函数, 且 $\hat{H}(15)=0.48\,1<\alpha=0.05$. 因此, 可以接收这批产品．

4.2　逐次检验

若试验是有周密安排, 则只要检验 15 次就可断定．这批发动机是可以接收的, 后面的 24 次试验就可省去.

4.3　分析

(1) 式(4)有个缺点, 就是没跟观测点 t_i($i=\overline{1,39}$) 的大小直接联系, 只与无失效的个数

S_i($i=\overline{1,39}$)联系．显然,它没有反映出产品在试验多长时间无失效的重要信息．人们总希望试验的时间越长,无失效的样品越多．这样的产品才是'绝对'可靠的．(2) 若t_i($i=\overline{1,39}$)选得太小,$\hat{R}(t_i)$虽然大也不能反映产品的可靠性高．为了克服式(4)与试验时间t_i没直接联系的缺点,改用公式

$$\hat{H}(k)=\hat{H}(\hat{R}_1,\hat{R}_2,\cdots,\hat{R}_k)=-\mathrm{e}\sum_{i=1}^{k}t_iS_i\hat{R}_i\ln\hat{R}_i/\sum_{i=1}^{k}t_iS_i \qquad k=\overline{1,m}. \tag{7}$$

若t_i的跳动比较大,S_i增加的个数不多,则无法保证$\hat{H}(k)$具有单调性．计算结果为$\hat{H}(1)=0.7348$,$\hat{H}(2)=0.5236$,$\hat{H}(3)=0.3948$,$\hat{H}(4)=0.2637$,$\hat{H}(5)=0.1934$,$\hat{H}(6)=0.1649$,$\hat{H}(7)=0.1193$,$\hat{H}(8)=0.1064$,$\hat{H}(9)=0.0966$,$\hat{H}(10)=0.0826$,$\hat{H}(11)=0.0750$,$\hat{H}(12)=0.0690$,$\hat{H}(13)=0.0638$,$\hat{H}(14)=0.0526$,$\hat{H}(15)=0.0486$,$\hat{H}(16)=0.0456$,$\hat{H}(17)=0.0430$,$\hat{H}(18)=0.0408$,$\hat{H}(19)=0.0389$,$\hat{H}(20)=0.0378$,$\hat{H}(21)=0.0365$,$\hat{H}(22)=0.0352$,$\hat{H}(23)=0.0338$,$\hat{H}(24)=0.0330$,$\hat{H}(25)=0.0321$,$\hat{H}(26)=0.0312$,$\hat{H}(27)=0.0295$,$\hat{H}(28)=0.0281$,$\hat{H}(29)=0.0270$,$\hat{H}(30)=0.0260$,$\hat{H}(31)=0.0251$,$\hat{H}(32)=0.0249$,$\hat{H}(33)=0.0318$,$\hat{H}(34)=0.0334$,$H(35)=0.0367$,$H(36)=0.0368$,$H(37)=0.0374$,$H(38)=0.0402$,$H(39)=0.0426$.

因t_1至t_{32}时间是渐进地升高,S_1至S_{32}也是逐步增加,故$\hat{H}(1)$至$\hat{H}(32)$仍然保持单调下降．但是,t_{33}至t_{39}跳跃增大,而S_{33}至S_{39}却增大得不多．于是,$H(33)$至$H(39)$不是单调下降反而上升．所以,式(5)不能作为逐步检验的指标,可作为一次检验的指标．由于$H(39)=0.0426<0.05$,它已利用全部的t_i与S_i的信息．因此,仍然判断该批发动机,在水平$\alpha=0.05$下是可以接收的．

正如前面谈及的,如果t_i的长度与n_i的选取是比较均匀的,则仍可保证式(5)具有单调性.这样,它可以作为逐次检验的指标．

参　考　文　献

1 Martz H F, Waller R A. A Bayes zero-failue reliability demonstration resting[J]. Procedure Journal of Quality Technology, 1979, 11(3): 128～137

2 茆诗松,罗朝斌．无失效数据的可靠性分析[J]．数理统计与应用概率,1989,4(4):489～506

3 吴绍敏,彭　沛．零失效数据的可靠度计算与区间估计[J]．华侨大学学报(自然科学版),1999,20(4):334～338

Zero-Faiture Method for Sampling Inspection with Entropy as Measure

Wu Shaomin

(College of Econ. Manag., Huaqiao Univ., 362011, Quanzhou)

Abstract Taking information entropy as measure, the author establishes a new method for sampling inspection of life which has an application range out of limitation and thus has a fairly thorough solution for this problem. The main contents include: information analysis of life and establishment of formula for initial estimation; formula of information entropy; method of single inspection; and method of successive inspection and examples.

Keywords information entropy, zero-failure data, sampling inspection