http://www.kekaoxing.com

第 28 卷 第 20 期 计算机工程与设计 2007 年 10 月
Vol. 28 No. 20 Computer Engineering and Design Oct. 2007

高可靠软件的可靠性评估技术

杨 彬， 陈丽容
(中国航天科工集团第二研究院 706 所，北京 100854)

摘 要：研究了高可靠软件的可靠性评估技术，给出了失效数据稀少情况下的软件可靠性模型。假设高可靠软件的可靠性测试过程中发生失效是独立同分布的稀有事件，从理论上分析了极值统计理论用于软件可靠性评估的可行性，建立了软件可靠性极值统计模型，讨论了模型的参数估计方法和假设检验方法。

关键词：软件可靠性模型； 高可靠软件； 极值统计； 失效数据
中图法分类号：TP302 **文献标识码**：A **文章编号**：1000-7024 (2007) 20-4839-03

Research on software reliability assessment method for highly reliable software

YANG Bin, CHEN Li-rong
(Institute 706, Second Academy of China Aerospace Science and Industry Corporation, Beijing 100854, China)

Abstract: A software reliability assessment method for highly reliable software systems is investigated. As highly reliable software system seldom fails, the failures as rare events is considered, so extreme value theory (EVT) is applied to model the failures. Analysis of rare event data using EVT does not require a priori assumptions concerning the distributional form of the data, and thus provides an interesting alternative to traditional approaches. The feasibility of using EVT in software reliability assessment is investigated, and the appropriate statistical methods is proposed.

Key words: software reliability; highly reliable software; statistics of extreme; failure data

0 引 言

关于软件可靠性的评估，从 1972 年出现第一个软件可靠性的评估模型以来，人们对软件可靠性评估方法和评估模型的研究就没有间断过。人们不断总结已有模型的优点和缺点，并在已有的评估理论的基础上，针对不同软件的特点和性能，不断地提出新的模型和方法。这些可靠性模型和评估方法都是基于系统运行期间的失效分析，事先假定软件可靠性失效数据服从一定的随机分布，并且要求一定量的软件可靠性失效数据。对大型商业系统来说，这些数据量可能还是足够的，但是对高可靠软件来说就往往不能产生符合要求的数据了。这有 3 方面的原因：①高可靠软件一般功能比较单一，操作较少；②高可靠软件开发过程中引入了大量软件可靠性分析与设计措施，例如容错、冗余和 N-文本编程等；③前期的软件测试比较充分，因此可靠性测试阶段确实很少发生失效。所以对于高可靠软件，难以决定选择哪个模型更合适，甚至就算选定了某一个模型，由于失效数据的稀少，模型也难以适用[1-2]。

本文分析了极值统计理论用于软件可靠性评估的可行性，建立了软件可靠性极值统计模型，讨论了模型的参数估计方法和假设检验方法，最后通过一个实例说明了该模型的应用。

1 极值统计简介

极值统计是专门研究很少发生、一旦发生却有巨大影响的随机变量极端变异性的建模及统计分析方法[9]。它以次序统计量为基础，研究极值事件分布的特性。极值事件常见于某种分布的尾部，用来拟合全部事件的分布有时不能精确地反映其尾部事件，这时母体分布尾部的分布可以由极值分布来精确拟合。设 $X_1, X_2, \cdots, X_n$ 是独立同分布的随机变量序列，如果存在常数列 $\{a_n>0\}$ 和 $\{b_n\}$，使得

$$\lim_{n\to\infty} \Pr\left(\frac{M_n - b_n}{a_n} \le x\right) = H(x), x \in R$$

成立，其中 $H(x)$ 是非退化的分布函数，那么 $H(x)$ 必属于下列 3 种类型之一[9]：

I 型分布：$H_1(x)=\exp\{-e^{-x}\}, -\infty<x<+\infty$。

II 型分布：$H_2(x;\alpha)=\begin{cases}0, x\le 0,\\ \exp\{-x^{-\alpha}\}, x>0,\end{cases}$ $\alpha>0$。

III 型分布：$H_3(x;\alpha)=\begin{cases}\exp\{-(-x)^{\alpha}\}, x\le 0,\\ 1, x>0,\end{cases}$ $\alpha>0$。

其中 I 型分布又称为 Gumbel 分布，II 型分布又称为 Fréchet 分布，III 型分布又称为 Weibull 分布，这 3 种分布统称为极值分布。当 $\alpha=1$ 时，$H_2(x;1)$ 和 $H_3(x;1)$ 分别称标准 Fréchet 分布与标准 Weibull 分布。求得 3 种类型极值分布的密度函数分布为

收稿日期：2007-02-15 E-mail：yangbin62@tom.com
作者简介：杨彬 (1982－)，男，湖南沅江人，硕士，研究方向为软件测试、软件可靠性； 陈丽容 (1962－)，女，湖南邵阳人，研究员，博士生导师，研究方向为计算数学、软件测试。

© 1994-2007 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$h_1(x)=e^{-x}H_1(x),-\infty<x<+\infty$$
$$h_2(x)=\alpha x^{-(1+\alpha)}H_2(x;\alpha),x>0$$
$$h_3(x)=\alpha(-x)^{\alpha-1}H_3(x;\alpha),x\leq 0$$

如果引进位置参数μ和尺度参数σ，则3种类型的极值分布函数为

$$H_1(x;\mu,\sigma)=\exp\{-e^{-\frac{x-\mu}{\sigma}}\},-\infty<x<+\infty$$
$$H_2(x;\mu,\sigma,\alpha)=\begin{cases}0,x\leq\mu,\\ \exp\{-(\frac{x-\mu}{\sigma})^{-\alpha}\},x>\mu,\end{cases}\quad\alpha>0$$
$$H_3(x;\mu,\sigma,\alpha)=\begin{cases}\exp\{-(-\frac{x-\mu}{\sigma})^{\alpha}\},x\leq\mu,\\ 1,x>\mu,\end{cases}\quad\alpha>0$$

2 软件可靠性极值统计模型

由于在高可靠性软件的测试过程中失效很少发生，我们可以认为高可靠性软件测试过程中失效的发生是稀有事件[5]，从而可以用极值统计方法来分析采集到的失效数据。根据这个思想，可以建立软件可靠性极值统计模型(software reliability model based on the statistics of extreme，SE-SRM)，并从假设检验和参数估计这两个方面提高模型的准确度，那么用该模型进行失效数据分析能获得更真实的可靠度。

2.1 基本假设

高可靠软件一旦失效将造成灾难性的后果，因此高可靠软件质量要求较高，在其交付使用之前，一般会进行大量的测试以发现并排除可能存在的缺陷。在此前提下，对高可靠软件进行可靠性测试可能只能收集到有限的失效数据。另一方面，“软件总是有缺陷的”这一观点已经在学术界和实业界得到了共识，测试中发生很少失效并不能保证软件已经达到了预定的可靠度。因此利用这些有限的失效数据对软件可靠性进行评估，是软件发布前进行决策时的一个必须步骤。

假定软件失效服从母体分布$F(x)$，X是收集的失效数据。则对高可靠软件而言，X很可能是位于母体分布$F(x)$的尾部，而不是母体分布的中心。传统的软件可靠性模型提供与母体分布中心有关的信息，它预先假定失效数据服从某一个先验分布(例如G-O模型假定失效数据服从指数分布)。如果选择了一个错误的分布，则预测的结果会不正确。这种先验分布的选取除了需要软件可靠性工作者的经验外，还需要大量失效数据的统计特性支撑。在失效数据很少的情况下，选择一个分布会特别困难，因此用来对整个失效数据范围建模的母体分布函数可能不能正确反映尾部分布。如果单独对尾部分布进行分析，可能会得到比对母体分布进行分析更准确的结果[8-9]。

极值统计关注给定概率分布的尾部特征，提供了一种对尾部分布进行分析的方法。使用极值统计来分布失效数据，不需要对数据假设一个先验分布。为了更加客观地描述高可靠软件可靠性测试阶段可靠性的变化规律，根据高可靠软件测试的特点，可以做如下基本假设：

(1)软件的失效为稀有事件。由于高可靠软件可靠性测试过程中失效次数很少，可以认为其失效为稀有事件；

(2)软件失效是独立同分布的。由于失效为稀有事件，可以认为软件失效是独立的，各失效之间没有关联。又由于软件的可靠性已经很高，失效时缺陷的排除对可靠性的影响不大，因此可以认为软件失效是同分布的。本条件的前提是条件(1)，如果条件(1)不成立，则本条件也值得怀疑。因此，本文的方法主要用于失效极少发生下的软件可靠性评估；

(3)软件的测试操作剖面与实际操作剖面一致。为了更准确地反映实际使用中软件的可靠性，在对软件进行测试时，其测试操作剖面应与实际操作剖面类似。

2.2 基本公式

基于以上假设，可以建立基于极值统计的软件可靠性模型SE-SRM。设$x_1,x_2,\cdots,x_n$为来自测试阶段的n次失效时间，独立同分布，分布函数为$F(x)$。由于高可靠软件的可靠性已经很高，软件所含的缺陷一般处于操作剖面出现概率极小的模块之中，只有软件多次运行时才可能被执行到。记软件失效时间$M_n=\max(x_1,x_2,\cdots,x_n)$，则$M_n$的分布为

$$P_r(M_n\leq t)=P_r(T_1\leq t,T_2\leq t,\cdots,T_n\leq x)=P_r(x_1\leq x)$$
$$P_r(x_2\leq x)\cdots P_r(x_n\leq x)=\{F(x)\}^n \tag{1}$$

由于软件失效时间的分布函数$F(x)$是连续型分布函数，是由操作剖面、执行任务的复杂程度等因素所决定的，在软件测试过程中$F(x)$的确定十分困难。为便于评估高可靠软件的可靠性，可以采用极值分布作为M_n的分布函数。由极值类型定理可知，当$n\to\infty$时，M_n的渐近分布为3种分布形式之一。我们假设这组失效数据符合Gumbel-II型极大值分布

$$H(x)=\exp\{-(\frac{v}{x})^k\},x>0 \tag{2}$$

通过参数估计得到上式中的参数的估计值$\hat{k}$、$\hat{v}$，然后再进行对这组失效数据是否确实符合Gumbel-II型分布进行假设检验。经过参数估计和假设检验以后，可以利用分布函数$H(x)$和估计值$\hat{k}$、$\hat{v}$求解各可靠性指标，从而对高可靠软件的可靠性进行评估。

母体分布为Gumbel-II型极值分布的可靠度函数为(x为失效时间)

$$R=P_r\{X>x\}=1-P_r\{X\leq x\}=1-H(x)=1-\exp[-(\frac{v}{x})^k] \tag{3}$$

母体分布为Gumbel-II型极值分布的失效率函数为

$$\lambda(x)=\frac{H'(x)}{R(x)}=\frac{kv^kx^{-(k+1)}\exp[-(\frac{v}{x})^k]}{1-\exp[-(\frac{v}{x})^k]} \tag{4}$$

2.3 参数估计

在统计学中对参数估计一般采用极大似然估计和最小二乘估计，但是两者的优良性质只有在样本量较大时才能体现。而高可靠软件的可靠性测试失效数据少而且离散性很强，这两种估计方法是否适用还没有定论。针对极值统计推断，本节引入最好线性无偏估计。设分布函数$F(x,\theta)$，其中θ是未知参数。从该总体中抽取一个随机样本$x_1,x_2,\cdots,x_n$，并构造一个统计量$\hat{\theta}$来估计θ，若$\hat{\theta}$满足下列条件：①$\hat{\theta}$是样本$x_1,x_2,\cdots,x_n$的线性函数；②$\hat{\theta}$是θ的无偏估计，即$E(\hat{\theta})=\theta$；③$\hat{\theta}$具有最小方差。

那么，$\hat{\theta}$是θ的最好线性无偏估计(best linear unbiased estimate，BLUE)。由于X服从Gumbel-II型极大值分布，令$Y=\ln X$，则Y服从Gumbel-I型极小值分布，其分布函数为

$$F(y)=1-\exp[-\exp(\frac{y-u}{\sigma})],-\infty<y<+\infty \tag{5}$$

© 1994-2007 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

http://www.kekaoxing.com

其中：$\mu_t=\ln v$，$\sigma=\frac{1}{k}$，于是对v,k的估计可以转换成对μ,σ的估计。

设失效数据$x_1,x_2,\cdots,x_n$为来自分布函数的总体分布的样本，并且是次序统计量。由于$Y=\ln X$是单调函数，因而$y_1,y_2,\cdots,y_n$也是次序统计量。由中国电子技术标准研究所编制的《可靠性试验用表》，当$n\leq 25$时，Gumel-I 型极小值分布中参数μ和σ的 BLUE 是$y_i=\ln x_i, i=1,2,\cdots,n$的线性函数，即

$$\begin{cases}\hat{\sigma}=\sum_{i=1}^{n}C(n,n,i)y_i=\sum_{i=1}^{n}C(n,n,i)\ln x_i\\ \hat{\mu}=\sum_{i=1}^{n}D(n,n,i)y_i=\sum_{i=1}^{n}D(n,n,i)\ln x_i\end{cases}\tag{6}$$

其中：$C(n,n,i)$、$D(n,n,i)$——最好线性无偏估计系数，其值可以查表获得。

根据μ和σ的估计值$\hat{\mu}$和$\hat{\sigma}$，就可以求得参数v,k的估计值

$$\begin{cases}\hat{k}'=\frac{1}{\hat{\sigma}}\\ \hat{v}=e^{\hat{\mu}}\end{cases}\tag{7}$$

虽然$\hat{\sigma}$是σ的无偏估计，但$\hat{k}'$不是k的无偏估计，为了得到k的无偏估计，可以作如下修正，即

$$\hat{k}=g_{n,n}\cdot\hat{k}'=\frac{g_{n,n}}{\hat{\sigma}}\tag{8}$$

其中：$g_{n,n}$——纠偏系数，其值可以查表获得。这样求得 Gumbel-II 型极大值分布中参数的最好线性无偏估计值$\hat{v}$和$\hat{k}$。

用 BLUE 进行参数估计具有相当高的精度，并且当$n<25$时查表方便。由于本文考虑的是失效数较少的情况，因此用 BLUE 来进行可靠性模型参数估计正合适。

2.4 假设检验

为了利用极值分布描述软件可靠性变化规律，我们需要对软件失效数据是否服从极值分布进行检验。设$x_1,x_2,\cdots,x_n$为软件可靠性测试的失效时间数据，需要检验如下假设：H_0：失效时间$x_1,x_2,\cdots,x_n$服从 Gumbel-II 型极大值分布，即$x_1,x_2,\cdots,x_n\sim H(x)$。其中$H(x)=\exp\{-(\frac{v}{x})^k\}$，$x>0$。具体过程如下：

对$H(x)=\exp\{-(\frac{v}{x})^k\}$，$x>0$进行线性化

$$-\ln[\ln\frac{1}{H(x)}]=k\ln x-k\ln v\tag{9}$$

令$s=-\ln[\ln\frac{1}{H(x)}]$，$t=\ln x$，$a=-k\ln v$，则式(9)可化为

$$s=kt+a\tag{10}$$

式中：$H(x)$——未知，但当n较大时，经验分布函数$H_n(i)=\frac{i}{n+1}$几乎处处收敛于$H(x_i)$，因此可以利用经验分布函数$H_n(i)$代替$H(x_i)$。为了检验式(10)是否成立，我们选择相关系数检验，即

$$r=\frac{l_{ts}}{\sqrt{l_{tt}l_{ss}}}\tag{11}$$

其中：$l_{ts}=\sum_{i=1}^{n}[(t_i-\frac{1}{n}\sum_{i=1}^{n}t_i)(S_i-\frac{1}{n}\sum_{i=1}^{n}S_i)]$，$l_{ss}=\sum_{i=1}^{n}(S_i-\frac{1}{n}\sum_{i=1}^{n}S_i)^2$，$l_{tt}=\sum_{i=1}^{n}(t_i-\frac{1}{n}\sum_{i=1}^{n}t_i)^2$。

当测试失效数据$x_1,x_2,\cdots,x_n$是来自于 Gumbel-II 型极值分布的样本时，则相关系数应接近 1。因此给定显著性水平α，$x_1,x_2,\cdots,x_n$来自于母体分布为 Gumbel-II 型的拒绝区域为$W=\{|r|<r_\alpha(n-2)\}$，其中$r_\alpha(n-2)$为临界值，其值可以通过查表获得。对于上述相关系数，其拒绝步骤为：①计算线性相关系数r；②查相关系数临界值表得到$r_\alpha(n-2)$；③看是否接受假设H_0。

3 应用实例

这里给出一个应用基于极值统计的软件可靠性模型的实例，完整给出了参数估计和假设检验的过程，并与其它软件可靠性模型进行了比较。某程序测试总时间为 528 小时，其中共发生了 7 次失效，实际运行时间为 2 148 小时，其中共发生了 6 次失效，各次失效发生的时间如表 1 所示。

表 1 某软件失效时间

失效数	测试失效时间	运行失效时间
1	120	144
2	120	336
3	152	408
4	192	1 128
5	296	1 152
6	368	1 680
7	488	–

(1)针对测试过程中的失效数据，用最好线性无偏估计进行参数估计，根据公式得到参数估计值：$\hat{k}_T=2.3684$，$\hat{v}_T=168.7$。

(2)用相关系数法检验测试过程中的这些数据是否符合极值分布，根据公式得$r_T=0.9798$。给定显著性水平$\alpha=0.05$，查表得$r_\alpha(7-2)=0.755<r_T$。可见，这些失效数据确实是符合极值分布的。

(3)算出软件可靠性测试结束后预计的可靠度函数和失效率函数为

$$\hat{R}_T=P_r\{X\rangle x\}=1-\exp[-(\frac{168.7}{x})^{2.3648}]$$

$$\hat{\lambda}_{ST}(x)=\frac{2.3684*168.7^{2.3684}x^{-3.3684}\exp[-(\frac{168.7}{x})^{2.3684}]}{1-\exp[-(\frac{168.7}{x})^{2.3684}]}$$

软件可靠性测试结束以后预计的软件平均失效时间，根据公式有$MTTF_T=\hat{v}\Gamma(1-\frac{1}{\hat{k}})=260.2166$。实际运行中的平均失效时间$MTTF_R$用各失效时间间隔的均值来代替，有$MTTF_R=280$。

G-O 模型是常用的 NHPP 类软件可靠性模型，它假定检测出的累积错误数是一个独立增量过程，其期望函数服从非齐次 Possion 分布。针对上述测试失效数据，根据 G-O 模型可以算得其可靠度函数为$R_G=\exp(-22.93(1-e^{-6.90\times 10^{-4}x})e^{-6.90\times 10^{-4}x})$，软件平均失效时间$MTTF_G=\int_0^\infty R\mathrm{d}t=98.2230$。

$|MTTF_G-MTTF_R|=181.7770$，而$|MTTF_T-MTTF_R|=19.7934$，这反映了 SE-SRM 模型进行预计的结果要比 G-O 模型更准确。

(4)针对软件实际运行失效数据，用最好线性无偏估计进行参数估计，根据公式得到参数估计值：$\hat{k}_R=1.1978$，$\hat{v}_R=381.08$。根据公式算得相关系数$r_R=0.9407$，给定显著性水平$\alpha=0.05$，查表得$r_\alpha(6-2)=0.811<r_R$。由此可见，实际运行中的失效数据是符合极值分布的。软件实际运行中的可靠度函数和失效率函数为

$$\hat{R}_R=P_r\{X\rangle x\}=1-\exp[-(\frac{381.08}{x})^{1.1978}]$$

$$\hat{\lambda}_{SR}(x)=\frac{1.1978*381.08^{1.1978}x^{-2.1978}\exp[-(\frac{381.08}{x})^{1.1978}]}{1-\exp[-(\frac{381.08}{x})^{1.1978}]}$$

(下转第 4852 页)

© 1994-2007 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

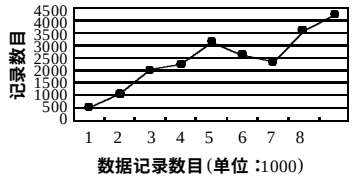

图 2　对不同记录数聚类的可伸缩性

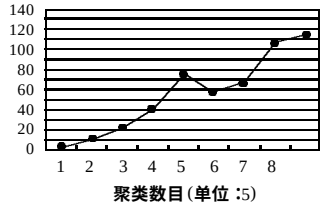

图 3　对不同聚类数目聚类的可伸缩性

256M RAM)机上的测试结果。

图 2 显示在聚类数(即 K)确定的情况下，随着记录数 n)的增加，算法的执行时间呈线形增加，能胜任对大型数据集的聚类。

图3显示在记录数目(n)不变时，随着聚类数目(K)的增加，改进算法的执行时间呈线形增加，表明算法的可伸缩性 较强。

4　结束语

同其它聚类技术一样，对于混合属性数据集的聚类研究仍是一个充满挑战性的研究领域，下一步我们将开展增加若干限制条件后聚类算法性能的提升研究。

参考文献:

[1] Han Jiawei, Michelie Kamber. Data mining concepts and techniques[M].北京:机械工业出版社,2001.

[2] Huang Zhexue. Extensions to the k-means algorithm for clustering large data sets with categorical values[J]. Data Mining and Knowledge Discovery,1998(2):283-304.

[3] Daniel Barbara. Using self-similarity to cluster large data sets[J]. Data Mining and Knowledge Discovery,2003(7):123-152.

[4] Dharmendra S Modha, Scott Spangler W. Feature weighting k-means clustering[J]. Machine Learning, 2003,52(3):217-237.

[5] Sun Y, Zhu Q, Chen Z. An iterative initial-points refinement algorithm for categorical data clustering[J]. Pattern Recognition Letters,2002,23(7):875-884.

[6] Gan G, Yang Z, Wu J. A genetic k-modes algorithm for clustering categorical data[C].Wuhan: Proc of ADMA'05,2005:195-202.

[7] 赵立江. 基于数值型和分类型混合属性数据集的聚类算法研究[D].杭州:浙江大学硕士学位论文,2005.

[8] Blake C, Merz J. UCI repository of machine learning databases [EB/OL]. http://www.ics.uci.edu/~mlearn/MLRepository.html.

(上接第 4841 页)

将软件可靠度函数$\hat{R}_T$、$\hat{R}_R$和失效率函数$\hat{\lambda}_{ST}(x)$、$\hat{\lambda}_{SR}(x)$绘成图，如图 1 所示。

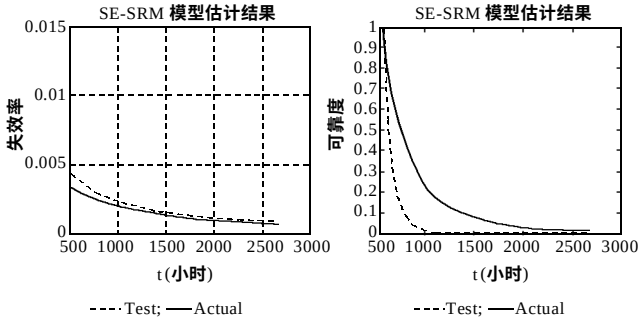

图 1　SE-SRM 模型应用

4　结束语

本章分析了对高可靠软件而言将软件失效视为稀有事件的合理性。在此基础上，利用极值统计理论建立了软件可靠性模型SE-SRM，并提出了适合该模型的参数估计和假设检验方法。根据一个实际的例子，详细阐述了SE-SRM的使用过程。

参考文献:

[1] Lori M Kaufman,Joanne Bechta Dugan,Barry W Johnson.Using statistics of the extremes for software reliability analysis [J]. IEEE Transactions on Reliability,1999,48(3):292-299.

[2] Lori M Kaufman,Ted C Giras.Simulation of rare events in transportation systems[C].Proceedings of the 2001 Winter Simulation Conference,2001.

[3] Uditha Balasooriya, Low C K.Competing causes of failure and reliability tests for weibull lifetimes under type I progressive censoring [J]. IEEE Transactions on Reliability, 2004,53 (1): 29-36.

[4] Toshiyuki Shimokawa.Goodness-of-fit tests for type-I extreme-value and 2-parameter weibull distributions [J]. IEEE Transactions on Reliability,1999,48(1):79-86.

[5] Yangyang Yu,Barry W Johnson.The quantitative safety assessment for safety-critical software[C].Proceedings of the 29th Annual IEEE/NASA Software Engineering Workshop,2005.

[6] Shepperd M,Cartwright M.Predicting with sparse data[J].IEEE Transactions on Software Engineering,2001,27(11):987-998.

[7] Nestoras E Evmorfopoulos.A monte carlo approach for maximum power estimation based on extreme value theory[J].IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems,2002,21(4):415-432.

[8] 陈希孺.高等数理统计学[M].合肥:中国科学技术出版社,1999.

[9] 史道济. 实用极值统计方法 [M]. 天津: 天津科学技术出版社, 2006.

© 1994-2007 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net