ISSN 0389-4010
UDC 519.23
519.245
519.65
519.654
519.67

TECHNICAL REPORT OF NATIONAL AEROSPACE LABORATORY

TR-1372T

Goodness-of-Fit Tests for the Type-I Extreme-Value and Two-Parameter Weibull Distributions with Unknown Parameters Estimated by Graphical Plotting Techniques —— Part 2 : Power Study

Min Liao and Toshiyuki Shimokawa

FEBRUARY 1999

NATIONAL AEROSPACE LABORATORY

CHŌFU, TOKYO, JAPAN

Goodness-of-Fit Tests for the Type-I Extreme-Value and Two-Parameter Weibull Distributions with Unknown Parameters Estimated by Graphical Plotting Techniques —— Part 2：Power Study*

Min Liao*1 and Toshiyuki Shimokawa*2

ABSTRACT

The objective of this study was to investigate the power of the Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling statistics for goodness-of-fit tests for the Type-I extreme-value and two-parameter Weibull distributions, when the population parameters were estimated by the combination of three kinds of graphical plotting techniques and the least-squares method and the maximum likelihood estimators. Monte Carlo simulation provided the power results using 10,000 repetitions for each sample size of 5, 10, 25, and 40. Four representative statistical distribution models were selected for alternative distributions in order to conduct the power comparison. The power comparisons indicated that the Anderson-Darling statistic coupled with the symmetrical ranks and the least-squares method is the most powerful statistic for goodness-of-fit tests, and is recommended for practical use.

Keywords: goodness-of-fit tests, power study, Kolmogorov-Smirnov, Cramer-von Mises, Anderson-Darling, Type-I extreme-value distribution, Weibull distribution, graphical plotting techniques, maximum likelihood estimators, Monte Carlo simulation

概　要

本研究の目的は、確率紙上で３種類のプロット法と最小２乗法の組合せおよび最尤法を用いて母数を推定する場合を対象とし、タイプI極値分布と２母数ワイブル分布に対するコルモゴロフ・スミルノフ、クレーマ・フォンミーゼス、およびアンダーソン・ダーリングの適合度検定統計量の検出力を調べることである。モンテカルロ・シミュレーションにより、標本の大きさ5, 10, 25, 40の各々に対して10,000組ずつを発生させ、検出力の結果を与えた。検出力を比較するためには、４種類の代表的統計分布モデルを選択した。計算した検出力を比較することにより、対称ランクと最小２乗法の組合せによるアンダーソン・ダーリングの適合度検定統計量を、適合度検定として最も検出力が高いと判定し実用に推奨した。

1. INTRODUCTION

The first part of this work[1] presented the critical values of the Cramer-von Mises (C-M) and Anderson-Darling (A-D) statistics for goodness-of-fit tests for the Type-I extreme-value and two-parameter Weibull distributions, when the population parameters were estimated by the combination of three kinds of graphical plotting techniques (GPTs) and the least-squares method (LSM). As the second part of this work, the main objective of this paper is to calculate and compare the power of the

---

* Received 22 May 1998

*1 Dr., STA (Science and Technology Agency) Fellow at National Aerospace Laboratory (NAL)(on leave from Northwestern Polytechnical University, Xi'an, P.R. China).

*2 Dr.,Head of Fatigue Strength Laboratory, Structures Division, NAL. Email：shimoka@nal.go.jp.

Kolmogorov-Smirnov (K-S), C-M, and A-D statistics for goodness-of-fit tests by using the three graphical plotting techniques in order to find the best combination of a statistic and one of the graphical plotting techniques. No reports on this subject have been published to date.

Maximum likelihood estimators (MLEs) are widely used to estimate the population parameters in goodness-of-fit tests for the Type-I extreme-value and two-parameter Weibull distributions[2)-5)]. In some handbooks such as MIL-HDBK-5G[6)] and MIL-HDBK-17-1E[7)], MLEs are used as parameter estimators in the goodness-of-fit tests for the Weibull distribution. Therefore, an additional objective of this paper is to compare the power of the K-S, C-M, and A-D statistics for goodness-of-fit tests with that of MLEs.

From an engineering viewpoint, one of the purposes of a goodness-of-fit test is to distinguish the fitness of a set of data to an assumed distribution model from alternative distribution models. Hence, this study compares the Type-I extreme-value or two-parameter Weibull distribution with the following four distribution models, which are frequently used in reliability analysis:
(1) Normal distribution
(2) Log-normal distribution
(3) Two-parameter Weibull distribution
(4) Type-I extreme-value distribution

Meanwhile, in some practical fields of reliability engineering, only a limited range of data scatter is often applied. For example, the coefficient of variation $\eta=3\%$ to 10% covers almost all η values of the strength data of metallic and composite materials. For typical fatigue life data[8)], the shape parameter $\beta=4$[8)] when a two-parameter Weibull distribution is used to fit the test data. The corresponding η of the life data is approximately equal to 28%. Therefore, in this power study, the distribution models in a specified η range also are compared.

Monte Carlo simulation is used to calculate the power of goodness-of-fit tests at seven significance levels, i.e., $\gamma=0.01$, 0.025, 0.05, 0.10, 0.15,0.20 and 0.25, in which 10,000 repetitions are generated for sample sizes $n=5$, 10, 25, and 40.

2. POWER CONCEPT OF A GOODNESS-OF-FIT TEST

Let $x_1<x_2<\ldots<x_n$ be order statistics from a continuous distribution $F(x)$. A goodness-of-fit test for a distribution with unknown parameters is to test the null hypothesis,

$$H_0: F(x)=F_0(x;\theta), \tag{1}$$

where $F_0(x;\theta)$ is a specified family of distribution models and contains a set of parameters θ.

In this paper, the power of a goodness-of-fit test is defined as the probability that the test statistic will reject the null hypothesis, H_0, when it is false, i.e., when a sample is not from the hypothesised population. Let the complement of the null hypothesis be the alternative hypothesis H_a. In the theory of statistical hypothesis tests[9)], there are two types of errors that may be produced when making a decision about the null hypothesis as follows:
(1) A type I error is produced, if H_0 is rejected when H_0 is true. The probability of a type I error is denoted by γ.
(2) A type II error is produced, if H_0 is accepted when H_0 is false. The probability of a type II error is denoted by λ .

The combination of γ and λ provides a measure for the efficiency of a goodness-of-fit test for the null hypothesis. Here, γ is also called the significance level of the tests, which has been considered during calculation of the critical values of a test statistic for the goodness-of-fit test. Thus, the power of the goodness-of-fit test at a given significance level γ is denoted by $(1-\lambda)$, when H_0 is false or H_a is true. An excellent goodness-of-fit test[10)] should reject a sample from the alternative distribution H_a at the highest " power probability "$(1-\lambda)$, and to accept a sample from the true distribution H_0 at the lowest " error probability "γ.

3. GOODNESS-OF-FIT TESTS TO BE INVESTIGATED

3. 1 Goodness-of-Fit Test Statistics

The power of goodness-of-fit tests is investigated for the combination of the statistic and parameter estimators listed in Table 1. The critical values of $W_n^2(1)$, $W_n^2(2)$, $W_n^2(3)$, $A_n^2(1)$, $A_n^2(2)$, and $A_n^2(3)$ in the first part of this work[1)] are used in the

Table 1 Combination of a goodness-of-fit test statstic and parameter estimators.

Statistic \ Estimators	Graphical Plotting Techniques: Median Ranks	Graphical Plotting Techniques: Mean Ranks	Graphical Plotting Techniques: Sym. Ranks	MLEs
K-S Statistic	$\hat{D}_n(1)$	$\hat{D}_n(2)$	$\hat{D}_n(3)$	$\hat{D}_n(4)$
C-M Statistic	$\hat{W}_n^2(1)$	$\hat{W}_n^2(2)$	$\hat{W}_n^2(3)$	$\hat{W}_n^2(4)$
A-D Statistic	$\hat{A}_n^2(1)$	$\hat{A}_n^2(2)$	$\hat{A}_n^2(3)$	$\hat{A}_n^2(4)$

following power calculation.

3. 2 Critical Values of the K-S Statistic Given by the Graphical Plotting Techniques

For the graphical plotting techniques, the critical values of the K-S statistic for goodness-of-fit tests for the Type-I extreme-value and two-parameter Weibull distributions are recalculated in this paper using the same procedures as described in the first part of this study[1)], though these critical values were already reported by one of the authors[11)] using 100,000 sets of samples for each sample size. This Monte Carlo simulation used 1,000,000 sets of samples for each of sample sizes $n=3(1)20, 25(5)50$, and $60(10)100$. The critical values of $\hat{D}_n(1)$, $\hat{D}_n(2)$, and $\hat{D}_n(3)$, at significance level $\gamma=0.05$ are plotted in Figure 1.

3. 3 Critical Values of the K-S, C-M, and A-D Statistics Given by the MLEs

For the MLEs, the critical values of the K-S, C-M, and A-D statistics, i.e., $\hat{D}_n(4)$, $\hat{W}_n^2(4)$, and $\hat{A}_n^2(4)$, for the Type-I extreme-value and two-parameter Weibull distributions are newly calculated by Monte Carlo simulation for 1,000,000 sets of samples for each of sample sizes $n=3(1)20$, 25(5)50, and 60(10)100. The calculation procedures are similar to those used in the first part of this study[1)]. This study used the MLEs subroutine program presented in MIL-HDBK-17-1E[4)].

Figures 1 to 3 indicate the critical values of $\hat{D}_n$, $\hat{W}_n^2$, and $\hat{A}_n^2$, calculated in this study at significance level $\gamma=0.05$. The critical values for the MLEs provided by Littell et al.[4)], Stephens[2),5)], and MIL-HDBK-17-1E[7)] at $\gamma=0.05$ are also presented in these figures. Figures 1 to 3 all clarified the following characteristics:

(1) The critical values of $\hat{D}_n(4)$ calculated are approximately equal to those given by Littell et al.

(2) The critical values of $\hat{W}_n^2(4)$ calculated are approximately equal to those given by Littell et al. and very close to those given by Stephens.

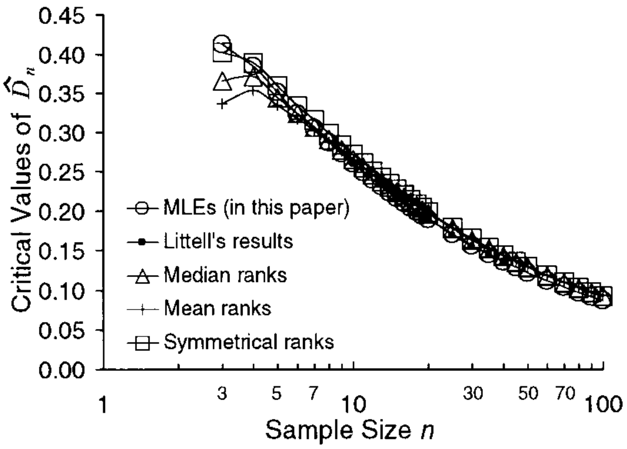

Figure 1 Critical values of the Kolmongorov-Smirnov statistic, $\hat{D}_n$,($\gamma=0.05$).

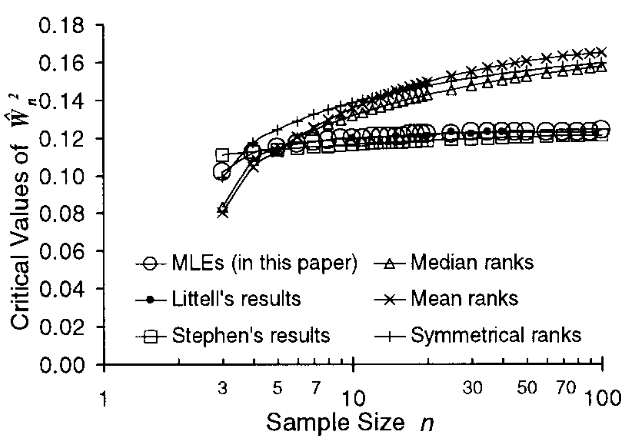

Figure 2 Critical values of the Cramer-von Misese statistic, $\hat{W}_n^2$,($\gamma=0.05$).

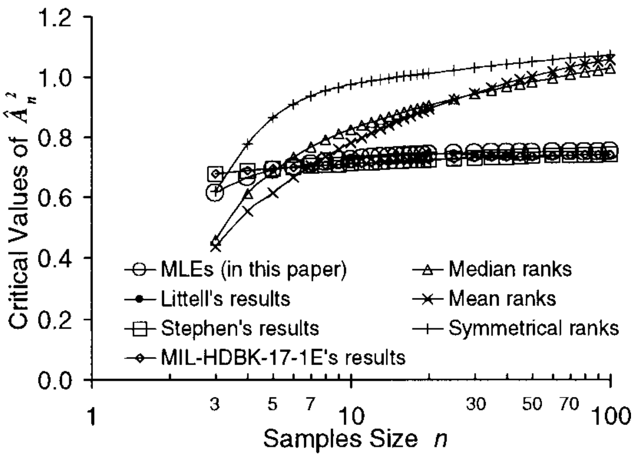

Figure 3 Critical values of the Anderson-Darling statistic, A_n^2,($\gamma = 0.05$).

(3) The critical values of A_n^2(4) calculated are approximately equal to those given by Littell et al. and very close to those given by Stephens and MIL-HDBK-17-1E.

(4) The critical values of the K-S statistic calculated by the graphical plotting techniques are very close to those calculated by the MLEs when sample size n>5. However, the critical values of the C-M and A-D statistics calculated by the graphical plotting techniques are significantly different from those calculated by the MLEs.

4. POWER CALCULATION

4. 1 Alternative Distribution Models and Generation of Samples

This section describes the alternative distribution functions and the detailed procedures of generating pseudo-random samples from each distribution model. In the following calculation, all pseudo-uniform random numbers, $u_i(i=1,2,...,n)$, are generated with the composite generator proposed by Marsaglia and Bray[12].

4. 1. 1 Normal distribution

The cumulative distribution function is,

$$F(x)=\Phi[\frac{x-\mu}{\sigma}], \qquad (-\infty<x<+\infty), \tag{2}$$

where $\Phi[\cdot]$ is the lower probability of a normal distribution with the population parameters, i.e., the mean μ and variance σ^2.

Sample generation

A general pseudo-random normal sample of size n, $x_i(i=1,2,...,n)$, is given by

$$x_i=\mu+\sigma z_i, \tag{3}$$

where z_i is a standard normal variable, which is obtained by transforming a uniform random number $u_i(0<u_i<1)$ using the polar method[13]. In this power study,

(1) Standard normal samples are tested against the Type-I extreme-value distribution without loss of generality. This case is simply expressed as " Std. Normal vs. Type-I " in the following figures. A standard normal sample of size n, $x_i(i=1,2,..., n)$, is generated by using Equation (3) with $\mu=0$ and $\sigma=1$. It should be noted that this test is equivalent to testing a log-normal sample against the two-parameter Weibull distribution[1].

(2) Three sets of samples from the normal distributions with specified $\eta=3\%$, 5%, and 10% are tested against the two-parameter Weibull distribution. This test is equivalent to testing a set of logarithmic values of a normal sample, i.e.,

$$y_i = \ln x_i = \ln\mu + \ln[1 + z_i \cdot \eta], \ (i=1,2,...,n) \tag{4}$$

against the Type-I extreme-value distribution. From Equation (4), this test is independent of $\ln\mu$, but dependent on η as described in the part one of this study[1)]. Therefore, a set of normal samples with a specified η is provided by taking $\ln\mu=1$ as follows:

$$x_i = 1 + z_i \cdot \eta, \ (i=1,2,...,n) \tag{5}$$

It should be noted that there is no negative value in x_i generated from each of the normal distributions with specified $\eta=$ 3%, 5%, and 10% in 10,000 samples.

4. 1. 2 Log-normal distribution

The cumulative distribution function is represented by

$$F(x)=\Phi[\frac{\ln x-\mu_L}{\sigma_L}], \qquad (0<x<+\infty), \tag{6}$$

where $\Phi[\cdot]$ is the lower probability of a normal distribution, and μ_L is the mean and σ_L^2 the variance of $\ln x$.

Sample generation

A general pseudo-random log-normal sample of size n, $x_i(i=1,2,...,n)$, is given by

$$x_i = \exp(\sigma_L z_i) \cdot \exp(\mu_L). \tag{7}$$

where z_i is a standard normal variable. In this study, four sets of samples from the log-normal distributions with specified η = 3%, 5%, 10%, and 28% are tested against the Type-I extreme-value distribution. Equation (7) indicates that this test is independent of μ_L, but dependent on σ_L[1)]. The relation between σ_L and η is given as

$$\sigma_L=\sqrt{\ln(1+\eta^2)}\,. \tag{8}$$

Therefore, a set of log-normal samples with a specified η is provided by taking $\mu_L=0$ as follows :

$$x_i=\exp\left[z_i\sqrt{\ln(1+\eta^2)}\right], \ (i<1,2,....,n). \tag{9}$$

4. 1. 3 Two-parameter Weibull distribution

The cumulative distribution function is,

$$F(x)=1-\exp[-\left(\frac{x}{\alpha}\right)^\beta], \ \alpha>0, \ \beta>0, \ 0<x<+\infty, \tag{10}$$

where α and β are the scale and shape parameters respectively.

Sample generation

A general pseudo-random Weibull sample of size n, $x_i(i=1,2,...,n)$, is given by

$$x_i=\alpha[-\ln(1-u_i)]^{1/\beta}, \tag{11}$$

where u_i is a uniform random number, $0<u_i<1$.

In this paper, four sets of samples from the two-parameter Weibull distributions with specified η = 3%, 5%, 10%, and 28% are tested against the Type-I extreme-value distribution. From Equation (11), these tests are independent of α, but dependent on β[1]. The relation between η and β is deduced as

$$\eta = \sqrt{\frac{\Gamma(1+2/\beta)}{\Gamma^2(1+1/\beta)} - 1}\ . \tag{12}$$

Therefore, a set of Weibull samples with a specified η is given by taking $\alpha = 1$ as follows :

$$x_i = \{-\ln(1-u_i)\}^{1/\beta(\eta)},\ (i=1,2,...,n) \tag{13}$$

where $\beta(\eta)$ is the inverse function of Equation (12).

4. 1. 4 Type-I extreme-value distribution

The cumulative distribution function is,

$$F(x) = 1 - \exp[-\exp(\frac{x-a}{b})],\ (-\infty < x < +\infty), \tag{14}$$

where a and b are the location and scale parameters respectively.

Sample generation

A general pseudo-random Type-I extreme-value sample of size n, $x_i = (i=1,2,...,n)$, is given by

$$x_i = a + b\ln[-\ln(1-u_i)]. \tag{15}$$

In this power study, three sets of samples from the Type-I extreme-value distributions with specified η =3%, 5%, and 10% are tested against the two-parameter Weibull distribution. This test is equivalent to testing a set of logarithmic values of a Type-I extreme-value sample, i.e.,

$$y_i = \ln x_i = \ln b + \ln\left[\ln\{-\ln(1-u_i)\} + \frac{a}{b}\right],\ (i < 1,2,....,n). \tag{16}$$

against the Type-I extreme-value distribution. From Equation (16), this test is independent of b, but dependent on a/b[1]. The relation between a/b and η is deduced as

$$\frac{a}{b} = \frac{\pi}{\sqrt{6}}\frac{1}{\eta} - \nu, \tag{17}$$

where ν = 0.5772 and is known as Euler's constant[2]. Therefore, a set of Type-I extreme-value samples with a specified η is given by taking b=1 as follows:

$$x_i = \ln[-\ln(1-u_i)] + (\frac{\pi}{\sqrt{6}}\frac{1}{\eta} - \nu),\ (i=1,2,....,n). \tag{18}$$

Here again, there is no negative value in x_i generated from each of the Type-I extreme-value distributions with specified η=3%, 5%, and 10% in 10,000 samples.

4. 2 Procedures of Power Calculation

Monte Carlo simulation is used for the power calculation whose procedures are similar to the calculation of the critical values of a statistic for goodness-of-fit tests described in the first part of this work[1]. 10,000 sets of samples for sample sizes n=5, 10, 25, and 40 are generated from each alternative distribution model mentioned above. Figure 4 presents the flow

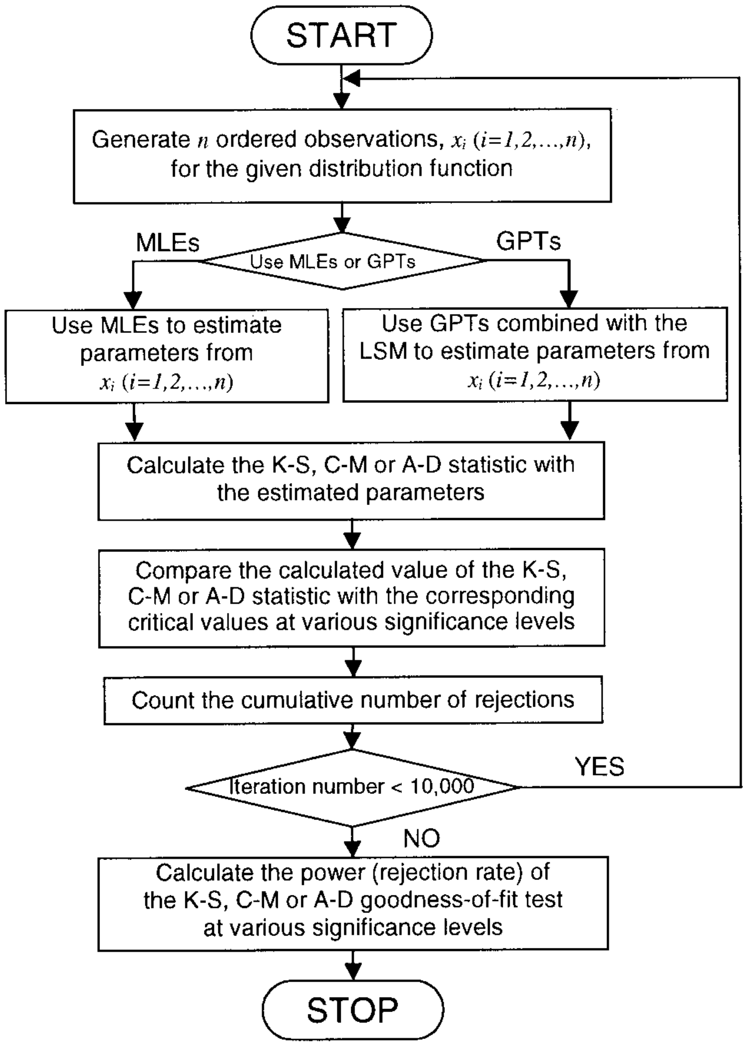

Figure 4 Flow diagram of the power calculation.

diagram of the power calculation.

5. POWER COMPARISONS

The power results of the K-S, C-M, and A-D statistics at the significance level γ=0.05 are presented in Tables 2 to 4. The averaged power for all sample sizes (n=5, 10, 25, and 40) are calculated and listed in each table. The power order is judged by the averaged power.

5. 1 K-S Goodness-of-Fit Tests

Table 2 presents the power results of $D_n(1)$, $D_n(2)$, $D_n(3)$, and $D_n(4)$. The averaged power in decreasing order is: $D_n(3)$, $D_n(4)$, $D_n(1)$, and $D_n(2)$. The comparisons indicate that:

(1) The power given by the symmetrical ranks is more powerful than those given by the median and mean ranks.

(2) The power given by the symmetrical ranks is also more powerful than that given by the MLEs.

5. 2 C-M Goodness-of-Fit Tests

Table 3 shows the power results of $W_n^2(1)$, $W_n^2(2)$, $W_n^2(3)$ and $W_n^2(4)$. The averaged power in decreasing order is: $W_n^2(4)$, $W_n^2(3)$, $W_n^2(1)$ and $W_n^2(2)$. The comparisons indicate that:

(1) The power given by the symmetrical ranks is more powerful than those given by the median and mean ranks.

Tablle 2 Power results, i.e., rejection rate, of the Kolmogorov-Smirnov goodness-of-fit tests.

Alterinative Distribution	Sample Size n	Test Statistics (Significance Level $\gamma = 0.05$) $\hat{D}_n(1)$ (Median Ranks)	$\hat{D}_n(2)$ (Mean Ranks)	$\hat{D}_n(3)$ (Sym. Ranks)	$\hat{D}_n(4)$ (MLEs)	*Littell's Results* (MLEs)
Std. Normal Samples vs. Type-I (Log-normal Samples vs. Weibull) #	5	0.062	0.056	0.070 ##	0.061	/
	10	0.089	0.072	0.101	0.091	0.091
	25	0.181	0.140	0.206	0.192	0.199
	40	0.273	0.220	0.303	0.291	0.315
Log-normal Samples (η=5%) vs. Type-I	5	0.069	0.061	0.077	0.064	/
	10	0.107	0.085	0.119	0.107	/
	25	0.228	0.180	0.255	0.243	/
	40	0.354	0.293	0.389	0.378	/
Log-normal Samples (η=28%) vs. Type-I	5	0.108	0.095	0.119	0.097	/
	10	0.216	0.180	0.237	0.211	/
	25	0.517	0.447	0.553	0.540	/
	40	0.757	0.691	0.787	0.769	/
Weibull Samples (η=5%) vs. Type-I	5	0.047	0.049	0.048	0.046	/
	10	0.049	0.044	0.050	0.050	/
	25	0.046	0.041	0.050	0.049	/
	40	0.047	0.040	0.052	0.050	/
Weibull Samples (η=28%) vs. Type-I	5	0.056	0.052	0.063	0.052	/
	10	0.072	0.058	0.083	0.077	/
	25	0.138	0.101	0.161	0.141	/
	40	0.216	0.163	0.246	0.211	/
Normal Samples (η=5%) vs. Weibull	5	0.058	0.053	0.065	0.057	/
	10	0.077	0.061	0.086	0.079	/
	25	0.138	0.105	0.160	0.150	/
	40	0.203	0.162	0.229	0.218	/
Type-I Samples (η=5%) vs. Weibull	5	0.051	0.055	0.050	0.049	/
	10	0.055	0.057	0.051	0.054	/
	25	0.063	0.070	0.057	0.055	/
	40	0.063	0.069	0.059	0.054	/
Averaged Power for Sample Size n	5	0.065	0.060	0.070	0.061	/
	10	0.095	0.080	0.104	0.096	/
	25	0.187	0.155	0.206	0.196	/
	40	0.273	0.234	0.295	0.282	/
Averaged Power for All Sample Sizes	n=5,10, 25,40	0.155	0.132	0.169	0.158	/
Power Order for All Sample Sizes	n=5,10, 25,40	No.3	No.4	No.1	No.2	/

Note # -- This test is equivalent to test the log-normal samples vs. the two-parameter Weibull distribution.
-- The data with an underline is the maximum value in this row except for *Littell's results.*

(2) The power given by the MLEs is more powerful than those given by the three kinds of graphical plotting techniques.

5. 3 A-D Goodness-of-Fit Tests

Table 4 presents the power results of $A_n^2(1)$, $A_n^2(2)$, $A_n^2(3)$, and $A_n^2(4)$. The averaged power in decreasing order is: $A_n^2(3)$, $A_n^2(1)$, $A_n^2(4)$, and $A_n^2(2)$. The comparisons indicat that:

(1) The power given by the symmetrical ranks is more powerful than those given by the median and mean ranks.

(2) The power given by the symmetrical and median ranks are more powerful than that given by the MLEs.

Tablle 3 Power results, i.e., rejection rate, of the Cramer-von Mises goodness-of-fit tests.

Alterinative Distribution	Sample Size n	Test Statistics (Significance Level $\gamma = 0.05$) $\hat{W}_n^2(1)$ (Median Ranks)	$\hat{W}_n^2(2)$ (Mean Ranks)	$\hat{W}_n^2(3)$ (Sym. Ranks)	$\hat{W}_n^2(4)$ (MLEs)	*Littell's Results* (MLEs)
Std. Normal Samples vs. Type-I (Log-normal Samples vs. Weibull) #	5	0.063	0.051	0.073 ##	0.061	/
	10	0.083	0.055	0.106	0.107	0.098
	25	0.182	0.120	0.218	0.257	0.259
	40	0.302	0.213	0.349	0.402	0.410
Log-normal Samples (η=5%) vs. Type-I	5	0.070	0.056	0.081	0.067	/
	10	0.101	0.067	0.124	0.132	/
	25	0.236	0.165	0.283	0.328	/
	40	0.396	0.300	0.448	0.509	/
Log-normal Samples (η=28%) vs. Type-I	5	0.108	0.088	0.124	0.103	/
	10	0.227	0.175	0.258	0.272	/
	25	0.581	0.484	0.633	0.669	/
	40	0.831	0.756	0.865	0.884	/
Weibull Samples (η=5%) vs. Type-I	5	0.048	0.048	0.049	0.048	/
	10	0.044	0.039	0.048	0.049	/
	25	0.043	0.034	0.050	0.051	/
	40	0.042	0.033	0.049	0.050	/
Weibull Samples (η=28%) vs. Type-I	5	0.057	0.046	0.064	0.056	/
	10	0.065	0.043	0.083	0.087	/
	25	0.135	0.082	0.179	0.183	/
	40	0.230	0.146	0.287	0.281	/
Normal Samples (η=5%) vs. Weibull	5	0.057	0.048	0.066	0.056	/
	10	0.067	0.046	0.088	0.090	/
	25	0.135	0.084	0.170	0.198	/
	40	0.220	0.146	0.263	0.299	/
Type-I Samples (η=5%) vs. Weibull	5	0.054	0.056	0.051	0.052	/
	10	0.058	0.059	0.053	0.051	/
	25	0.070	0.076	0.061	0.056	/
	40	0.073	0.081	0.062	0.057	/
Averaged Power for Sample Size n	5	0.065	0.056	0.073	0.063	/
	10	0.092	0.069	0.108	0.113	/
	25	0.197	0.149	0.228	0.249	/
	40	0.299	0.239	0.332	0.355	/
Averaged Power for All Sample Sizes	n=5,10, 25,40	0.164	0.128	0.185	0.195	/
Power Order for All Sample Sizes	n=5,10, 25,40	No.3	No.4	No.2	No.1	/

Note # -- This test is equivalent to test the log-normal samples vs. the two-parameter Weibull distribution.
-- The data with an underline is the maximum value in this row except for *Littell's results*.

5. 4 Power Comparison Among the K-S, C-M, and A-D Statistics

According to all the power results for the K-S, C-M, and A-D statistics in Tables 2-4, the power order is given as: $A_n^2(3)$, $A_n^2(1)$, $A_n^2(4)$, $W_n^2(4)$, $W_n^2(3)$, $A_n^2(2)$, $D_n(3)$, $W_n^2(1)$, $D_n(4)$, $D_n(1)$, $D_n(2)$, and $W_n^2(2)$,This power order reveals that:

(1) When the MLEs are used to estimate the population parameters, the A-D statistic is more powerful than the C-M statistic, and the C-M statistic is more powerful than the K-S statistic.

(2) The A-D statistic coupled with the symmetrical ranks is the best goodness-of-fit test in most of cases shown in Tables 2 to 4.

Tablle 4 Power results, i.e., rejection rate, of the Anderson-Darling goodness-of-fit tests.

Alternative Distribution	Sample Size n	Test Statistics (Significance Level $\gamma = 0.05$) $\hat{A}_n^2(1)$ (Median Ranks)	$\hat{A}_n^2(2)$ (Mean Ranks)	$\hat{A}_n^2(3)$ (Sym. Ranks)	$\hat{A}_n^2(4)$ (MLEs)	*Littell's Results* (MLEs)
Std. Normal Samples vs. Type-I (Log-normal Samples vs. Weibull) #	5	0.085	0.063	0.098 ##	0.048	/
	10	0.150	0.099	0.181	0.099	0.084
	25	0.315	0.219	0.373	0.277	0.286
	40	0.481	0.353	0.546	0.447	0.462
Log-normal Samples (η=5%) vs. Type-I	5	0.094	0.069	0.110	0.051	/
	10	0.183	0.123	0.213	0.121	/
	25	0.409	0.291	0.464	0.355	/
	40	0.605	0.477	0.670	0.564	/
Log-normal Samples (η=28%) vs. Type-I	5	0.152	0.111	0.177	0.078	/
	10	0.359	0.271	0.405	0.261	/
	25	0.778	0.674	0.825	0.706	/
	40	0.948	0.900	0.963	0.917	/
Weibull Samples (η=5%) vs. Type-I	5	0.052	0.049	0.056	0.049	/
	10	0.050	0.044	0.056	0.046	/
	25	0.049	0.037	0.058	0.052	/
	40	0.048	0.035	0.058	0.051	/
Weibull Samples (η=28%) vs. Type-I	5	0.080	0.056	0.090	0.047	/
	10	0.112	0.076	0.140	0.079	/
	25	0.244	0.149	0.297	0.195	/
	40	0.378	0.250	0.452	0.318	/
Normal Samples (η=5%) vs. Weibull	5	0.077	0.058	0.088	0.045	/
	10	0.123	0.078	0.145	0.080	/
	25	0.239	0.154	0.287	0.211	/
	40	0.359	0.249	0.420	0.338	/
Type-I Samples (η=5%) vs. Weibull	5	0.049	0.056	0.048	0.057	/
	10	0.051	0.056	0.044	0.054	/
	25	0.062	0.076	0.050	0.060	/
	40	0.069	0.084	0.056	0.060	/
Averaged Power for Sample Size n	5	0.084	0.066	0.095	0.054	/
	10	0.147	0.107	0.169	0.106	/
	25	0.299	0.229	0.336	0.265	/
	40	0.413	0.335	0.452	0.385	/
Averaged Power for All Sample Sizes	n=5,10, 25,40	0.236	0.184	0.263	0.202	/
Power Order for All Sample Sizes	n=5,10, 25,40	No.2	No.4	No.1	No.3	/

Note # -- This test is equivalent to testing the log-normal samples vs. the two-parameter Weibull distribution.
-- The data with an underline is the maximum value in this row except for *Littell's results*.

The power comparisons described above indicate that the A-D statistic coupled with the symmetrical ranks is more powerful than the A-D statistic coupled with the MLEs, which is carried in MIL-HDBK-5G[6] and MIL-HDBK-17-1E[7]. However, the power comparisons in Tables 2 to 4 were conducted at $\gamma = 0.05$ only. More detailed power comparisons for the A-D statistic coupled with the symmetrical ranks and the MLEs should be investigated at different significance levels and for specified coefficients of variation of the population. Sample sizes $n = 10$ and 40 representing small and fairly large sample sizes are considered.

5.5 Power Comparison for the A-D Statistic at Different Significance Levels

Figures 5 and 6 present the power results of the A-D statistic coupled with the symmetrical ranks and the MLEs at seven significance levels for sample sizes $n=10$ and 40. These figures show that:

(1) The symmetrical ranks are more powerful than the MLEs in most cases.

(2) The calculated power increases monotonically as γ becomes higher.

(3) High power is obtained between the Normal group (normal and log-normal distributions) and the Weibull group (the two-parameter Weibull and Type-I extreme-value distributions), especially for the large sample size 40 and higher significance levels. Meanwhile, very low power was obtained between the Type-I extreme-value distribution and the two-parameter Weibull distribution for $\eta=5\%$.

5. 6 Power Comparison for the A-D Statistic at Different Coefficients of Variation

Figures 7 and 8 present the power results of the A-D statistic coupled with the symmetrical ranks and the MLEs at three or four values of η ($\eta=3\%$, 5%, 10%, and 28%) for sample sizes $n=10$ and 40. The following conclusions were obtained:

(1) The symmetrical ranks give a higher power than the MLEs in most cases.

(2) Although the power to test the normal samples against the two-parameter Weibull distribution decreases as η becomes larger, the power for other test cases increases as η becomes larger.

(3) A very low power is obtained between the Type-I extreme-value distribution and the two-parameter Weibull distribution for small coefficient of variation, $\eta=3\%\sim10\%$. For large $\eta=28\%$, however, a fairly high power between the two distributions is obtained.

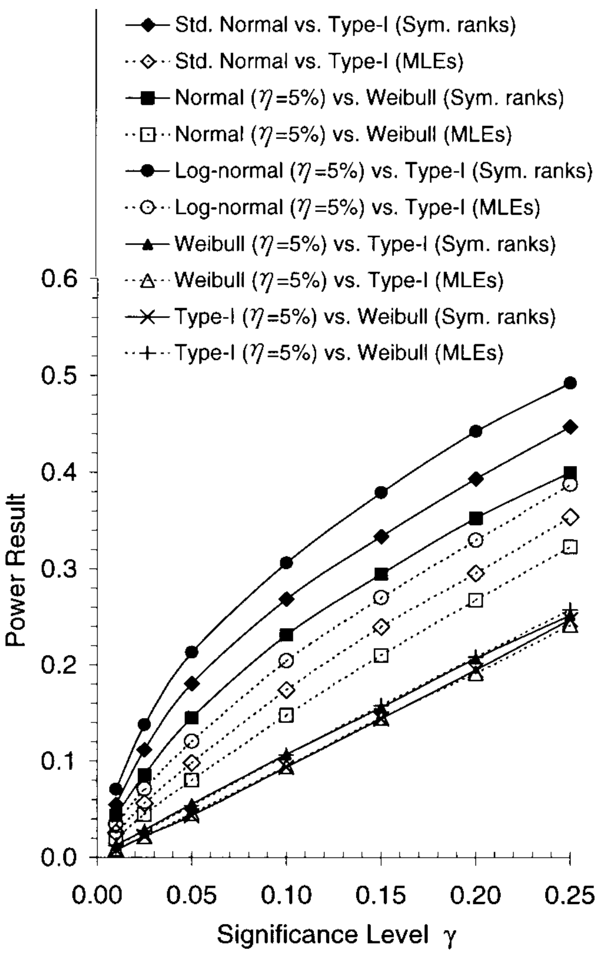

Figure 5 Power results of the Anderson-Darling goodness-of-fit test at different significance levels, (sample size $n=10$).

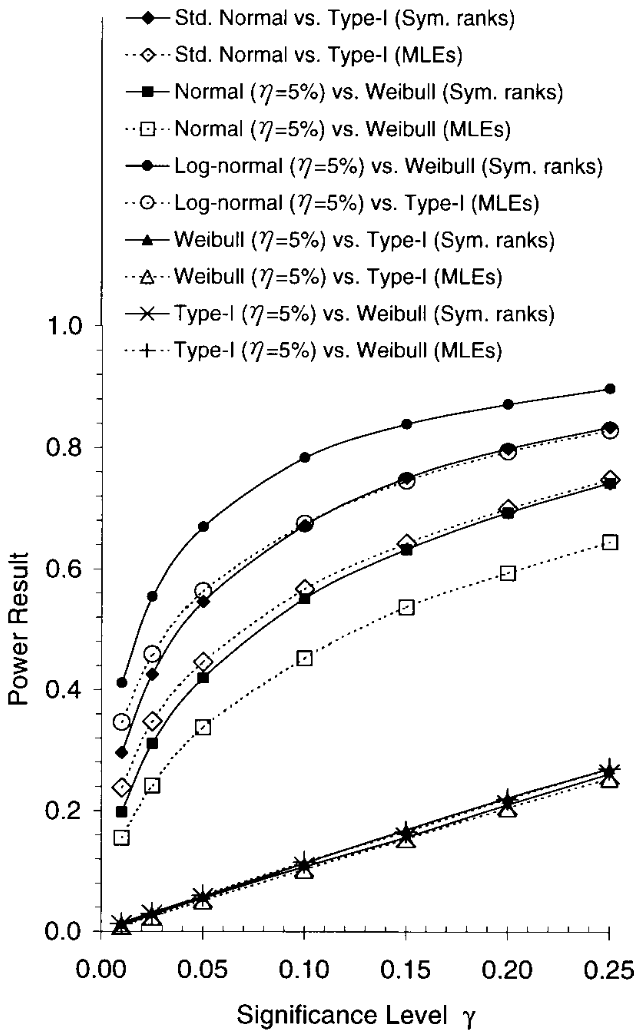

Figure 6 Power results of the Anderson-Darling goodness-of-fit test at different significance levels, (sample size $n=40$).

6. DISCUSSION

6. 1 Validity of the Monte Carlo Simulation and Its Results

The power results calculated by Littell et al. for the K-S, C-M, and A-D statistics coupled with the MLEs are presented in Tables 2, 3, and 4 for sample sizes n=10, 25, and 40. These results are very close to the corresponding results in this study although they were obtained by a different random number generator and ten sets of 5,000 iterations[4]. These facts confirm not only the reproducibility of the power results by the MLEs, but also the validity of the Monte Carlo simulation in this study.

6. 2 Median, Mean, and Symmetrical Ranks in the Graphical Plotting Techniques

For the K-S, C-M, and A-D statistics, the symmetrical ranks provided more powerful results than the median and mean ranks. One explanation for this is that the C-M and A-D statistics[1] include the concept of the symmetrical ranks, $(i-0.5)/n$, which implies that the symmetrical ranks should be used for estimating the unknown population parameters for these statistics. A similar concept is used for the K-S statistic[9].

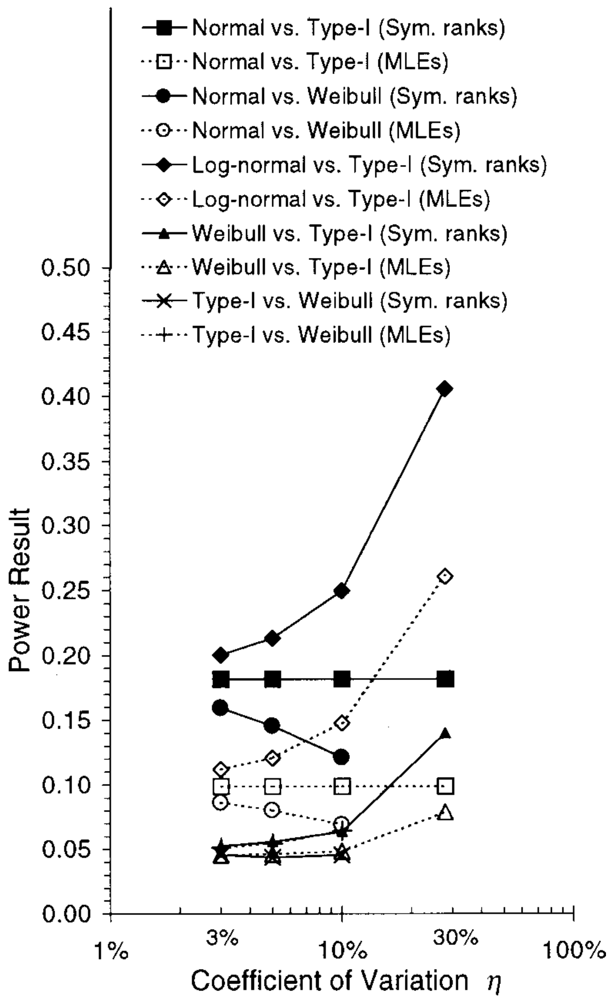

Figure 7 Power results of the Anderson-Darling goodness-of-fit test as a function of the coefficient of variation, (sample size $n=10$, significance level $\gamma=0.05$).

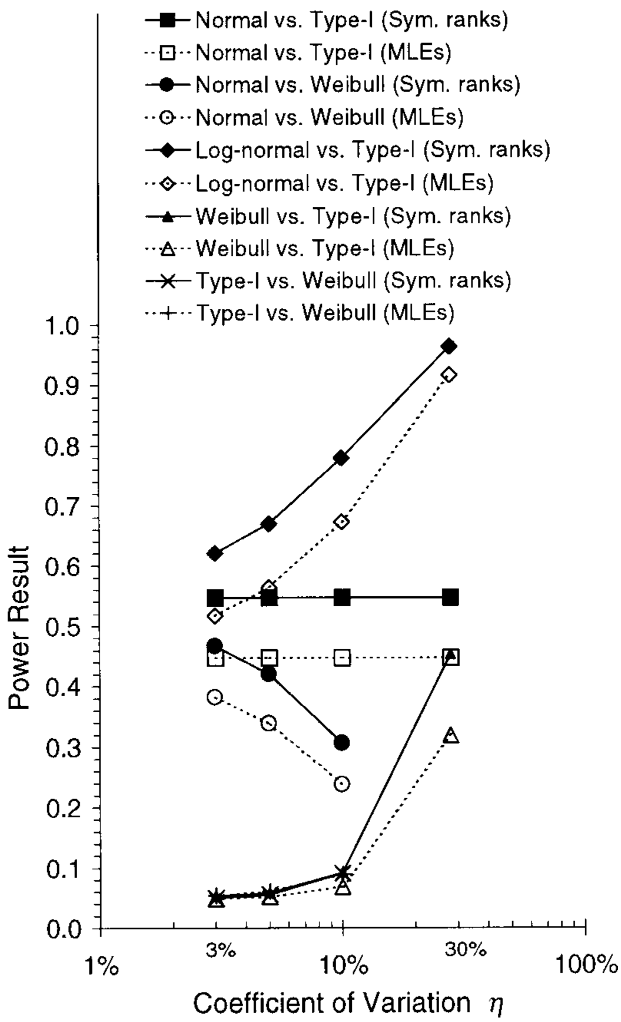

Figure 8 Power results of the Anderson-Darling goodness-of-fit test as a function of the coefficient of variation, (sample size $n=40$, significance level $\gamma=0.05$).

6. 3 Comparison of Concept of the K-S, C-M, and A-D Statistics

The basic concept of the K-S, C-M, and A-D statistics is to test the null hypothesis by measuring the " distance " (discrepancy) between the non-parametric empirical distribution function (EDF) and the assumed cumulative distribution function (CDF) with the estimated parameters. The K-S statistic, D_n, is used to measure the maximum " distance " between EDF and CDF, which uses only the one largest discrepancy calculated from the data. The C-M statistic, W_n^2, is defined by the sum of the square of the " distance " between EDF and CDF on the basis of all the discrepancies over the entire range of the data. Moreover, the A-D statistic, A_n^2, improved W_n^2 by introducing a weight function with heavier weights in both tail regions[2),3),5)]. These weights are effective in producing more powerful results for goodness-of-fit tests.

6. 4 Difference Between the Type-I Extreme-Value Distribution and the Two-Parameter Weibull Distribution

The power comparisons in this paper also revealed the difference between the Type-I extreme-value distribution and the two-parameter Weibull distribution. That is, for small coefficient of variation, $\eta = 3 \sim 10\%$, the difference between the two distributions is rather small; for large $\eta = 28\%$, however, the difference is much larger. Therefore, the following inferences are drawn:

(1) For the typical fatigue life data (with η=28%), there should be a significant difference when either the Type-I extreme-value distribution or the two-parameter Weibull distribution is used for fitting the data.

(2) For the strength data of metallic and composite materials (with $\eta = 3 \sim 10\%$), there should be a small difference when either of the two distributions is used.

7. CONCLUSIONS

The main conclusions of this paper are summarized as follows:

(1) Among the three kinds of graphical plotting techniques, the symmetrical ranks give more powerful results than the median and mean ranks for the K-S, C-M, and A-D statistics.

(2) Among the three graphical plotting techniques and the MLEs, the symmetrical ranks provide more powerful results than the MLEs for the K-S and A-D statistics. However, for the C-M statistic, the MLEs provided more powerful results than the three graphical plotting techniques.

(3) The A-D statistic coupled with the symmetrical ranks and the least-squares method provides the best power results among the competitors in this study, and is recommended for practical use.

(4) It is difficult to separate the Type-I extreme-value distribution and the two-parameter Weibull distribution for small coefficient of variation $\eta = 3 \sim 10\%$, but it is not difficult for large $\eta = 28\%$.

REFERENCES

1) T. Shimokawa and M. Liao; Goodness-of-Fit Tests for the Type-I Extreme-Value and Two-Parameter Weibull Distributions with Unknown Parameters Estimated by Graphical Plotting Techniques — Part 1 : Critical Values, NAL TR-1371T (1999).

2) J.F. Lawless; Statistical Models and Methods for Lifetime Data (1982), John Wiley & Sons.

3) B.W. Woodruff and A.H. Moore; Application of Goodness-of-Fit Tests in Reliability, Handbook of Statistics, Vol. 7 (1988), pp. 113-120, Elsevier Science Publishers B.V.

4) R.C. Littell, J.T. McClave, and W.W. Offen; Goodness-of-Fit Tests for the Two-Parameter Weibull Distribution, Commun. Statist. -Simula. Computa., Vol. 8(B), No. 3 (1979), pp. 257-269.

5) M.A. Stephens; Goodness of Fit for the Extreme Value Distribution, Biometrika, Vol. 64, No. 3 (1977), pp. 583-588.

6) MIL-HDBK-5G, Metallic Materials and Elements for Aerospace Vehicle Structures (1994), Department of Defense, U.S.A.

7) MIL-HDBK-17-1E, Polymer Matrix Composites, Vol. 1 (1997), Department of Defense, U.S.A.

8) I.C. Whittaker and P.M. Besuner; A Reliability Analysis Approach to Fatigue Life Variability of Aircraft Structures, Technical Report, AFML-TR-69-65 (1969), Airforce Material Laboratory.

9) N.R. Mann, R.E. Schafer, and N.D. Singpurwalla; Methods for Statistical Analysis of Reliability and Life Data (1974), John Wiley & Sons.

10) H. Duman and A.H. Moore; A New Goodness-of-Fit Test Based on Spacings for the Gamma Distribution with Known Shape Parameter from Complete and Censored Samples, Air Force Inst. of Tech., Wright-Patterson AFB, U.S., AD-A293875 (1995).

11) T. Shimokawa; Critical Values of Modified Kolmogorov-Smirnov Goodness-of-fit Tests for the Type-I Extreme-Value and Two-Parameter Weibull Distributions, Proceedings of 27th Reliability and Maintainability Symposium, JUSE (1997), pp. 143-148 (in Japanese).

12) G. Marsaglia and T.A. Bray; One-Line Random Number Generators and Their Use in Combinations, Communications of the ACM, Vol. 11, No. 11 (1968), pp. 757-759.

13) W.H. Press, et al.; Numerical Recipes in FORTRAN: the Art of Scientific Computing, 2nd ed. (1992), pp. 280, Cambridge University Press.

TECHNICAL REPORT OF NATIONAL
AEROSPACE LABORATORY
TR-1372T

航空宇宙技術研究所報告1372T号（欧文）

平 成 １１ 年 ２ 月 発 行

発 行 所　科学技術庁航空宇宙技術研究所
東京都調布市深大寺東町７－44－１
電話（0422）47－5911　〒182－8522
印 刷 所　株 式 会 社　実 業 公 報 社
東京都千代田区九段北１－７－８

Published by
NATIONAL AEROSPACE LABORATORY
7－44－1 Jindaijihigashi－machi, Chōfu
Tokyo 182－8522, JAPAN

Printed in Japan