# Comparison of Five Tests of Fit for the Extreme Value Distribution

D. J. Best, *School of Mathematical and Physical Sciences,*
*University of Newcastle, NSW 2308, Australia.*
*Email: Donald.J.Best@newcastle.edu.au*

J. C. W. Rayner, *School of Mathematical and Physical Sciences,*
*University of Newcastle, NSW 2308, Australia.*
*Email: John.Rayner@newcastle.edu.au*

O. Thas, *Department of Applied Mathematics, Biometrics and*
*Process Control, Ghent University, B-9000 Gent, Belgium.*
*Email: olivier.thas@UGent.be*

_____

**Abstract**

Tests for the Extreme Value distribution based on the sample skewness and kurtosis coefficients are shown to be related to components of smooth tests of goodness of fit and are compared with tests due to Anderson-Darling, Shapiro-Brain and Liao-Shimokawa. Two examples are given.

*AMS Subject Classification*: 62F03 and 62G32

*Key Words*: Method of moments estimation; Orthonormal functions; Skewness and kurtosis coefficients; Skewed distributions; Weibull distribution.
_____

## 1. Introduction

The Normal or Gaussian distribution is probably the most used symmetric distribution. Two traditional tests of fit for this distribution are based on the moment statistics for skewness and kurtosis. The Extreme Value distribution is a commonly applied skewed distribution. See, for example, Coles (2004, section 4 of Chapter 3 and the list on page 1). Here we look at the performance of skewness and kurtosis tests for the Extreme Value distribution as well as three other tests of fit, suggested by Stephens (1977), Shapiro-Brain (1987) and Liao-Shimokawa (1999).

Traditional applications of the Extreme Value distribution involve estimating the occurrence of say, the maximum 'one in 100 years' flood based on less than 100 years of data. For example, author Best used annual maximum daily rainfall figures for about 75

years from locations along the Australian 'Sydney to Newcastle freeway' to estimate 'one in 100 years' maximum 24 hour rainfall. This would have helped road engineers plan drainage systems for the freeway during its construction in the 1980s. Similarly the Port Pirie annual maximum sea-levels data given below might help Port authorities decide on the height of a new sea-wall.

A more recent application of the Extreme Value distribution is in molecular biology where DNA protein sequences are aligned with those in a database. Under reasonable assumptions a random score distribution for optimal ungrouped alignments follows the Extreme Value distribution. See Karlin and Altschul (1990).

Besides empirical evidence that the Extreme Value distribution fits various maximum or minimum data, there are theoretical justifications for its use and these are outlined, for example, in Coles (2004, Chapter 3).

The usual Extreme Value distribution has two parameters: a location parameter $\alpha$ and a dispersion parameter $\beta$. If we let $X$ have the Extreme Value distribution then the standardized Extreme Value variable $T = (X - \alpha)/\beta$ has probability density function

$$f_T(t) = \exp\{-t - \exp(-t)\}, \quad -\infty < t < \infty . \tag{1.1}$$

Two potential applications involving the Extreme Value distribution follow.

**Port Pirie data.** The annual maximum sea-levels recorded at Port Pirie (just north of Adelaide), South Australia for 1923-1987 are given in Table 1. These are typical of data that are modelled by the Extreme Value distribution.

**Glass Fibre data**. Sixty-three breaking strengths of glass fibres of length 1.5 cm were reported by Smith and Naylor (1987). No units for the breaking strengths were given. The data are given in Table 2. Are these data well modelled by the Extreme Value distribution?

In section 4 below we will test the fit of the Extreme Value distribution for these small data sets. Section 2 gives some definitions while section 3 gives some critical values and power comparisons.

## 2. Definitions

If $x_1, \ldots, x_n$ are data, and if $\bar{x} = \sum_j x_j / n$ and if $s^2 = \sum_j (x_j - \bar{x})^2 / n$, we define the sample skewness and kurtosis by

$$\sqrt{b_1} = \sum_j \left\{ (x_j - \bar{x})/s \right\}^3 / n \tag{2.1}$$

and

$$b_2 = \sum_j \left\{ (x_j - \bar{x})/s \right\}^4 / n \tag{2.2}$$

respectively. These two moment statistics can be related to smooth tests of fit. Based on a random sample $X_1, \ldots, X_n$, if $\{h_r(x; \beta)\}$ is a set of complete orthonormal functions on the hypothesized distribution, Rayner and Best (1989) test goodness of fit using the

components $V_r = \sum_{j=1}^{n} h_r\left(X_j; \hat{\alpha}, \hat{\beta}\right) / \sqrt{n}$ , where typically, but not necessarily, ($\hat{\alpha}$ , $\hat{\beta}$) is a maximum likelihood estimator of ($\alpha, \beta$).

**Table 1**. Annual maximum sea-levels, recorded in
metres at Port Pirie, South Australia 1923-1987

| | | | | |
|---|---|---|---|---|
| 4.03 | 3.96 | 4.06 | 4.11 | 3.91 |
| 3.83 | 3.85 | 3.71 | 4.24 | 3.72 |
| 3.65 | 3.93 | 3.96 | 3.96 | 4.00 |
| 3.88 | 3.75 | 4.06 | 4.21 | 3.66 |
| 4.01 | 3.63 | 4.55 | 3.74 | 3.62 |
| 4.08 | 3.57 | 3.79 | 3.85 | 4.33 |
| 4.18 | 4.25 | 3.89 | 3.88 | 4.55 |
| 3.80 | 3.97 | 4.11 | 3.66 | 3.75 |
| 4.36 | 4.05 | 3.85 | 4.11 | 4.08 |
| 3.96 | 4.24 | 3.86 | 3.71 | 3.90 |
| 3.98 | 4.22 | 3.86 | 4.18 | 3.88 |
| 4.69 | 3.73 | 4.21 | 3.90 | 3.94 |
| 3.85 | 4.37 | 4.01 | 3.78 | 4.33 |

**Table 2.** Breaking strengths of 63 glass fibres

| | | | | |
|---|---|---|---|---|
| 0.55 | 0.74 | 0.77 | 0.81 | 0.84 |
| 0.93 | 1.04 | 1.11 | 1.13 | 1.24 |
| 1.25 | 1.27 | 1.28 | 1.29 | 1.30 |
| 1.36 | 1.39 | 1.42 | 1.48 | 1.48 |
| 1.49 | 1.49 | 1.50 | 1.50 | 1.51 |
| 1.52 | 1.53 | 1.54 | 1.55 | 1.55 |
| 1.58 | 1.59 | 1.60 | 1.61 | 1.61 |
| 1.61 | 1.61 | 1.62 | 1.62 | 1.63 |
| 1.64 | 1.66 | 1.66 | 1.66 | 1.67 |
| 1.68 | 1.68 | 1.69 | 1.70 | 1.70 |
| 1.73 | 1.76 | 1.76 | 1.77 | 1.78 |
| 1.81 | 1.82 | 1.84 | 1.84 | 1.89 |
| 2.00 | 2.01 | 2.29 | | |

For the normal distribution put $T = (x - \hat{\alpha}) / \hat{\beta}$ in which $\hat{\alpha} = \overline{X}$ and $\hat{\beta} = S$, the sample mean and standard deviation respectively, and then $T$ has the standard normal distribution with orthonormal polynomials the Hermite-Chebycheff polynomials, the first few of which are given by

$h_0(t) = 1$, $h_1(t) = t$, $h_2(t) = (t^2 - 1)/\sqrt{2}$, $h_3(t) = (t^3 - 3t)/\sqrt{6}$, and $h_4(t) = (t^4 - 6t^2 + 3)/\sqrt{24}$.

Thus for the normal $V_1 = \sum_{j=1}^{n} h_1(t_j)/\sqrt{n} \equiv 0$, $V_2 = \sum_{j=1}^{n} h_2(t_j)/\sqrt{n} \equiv 0$, with

$$V_3 = \sum_{j=1}^{n} h_3(t_j)/\sqrt{n} = \sum_{j=1}^{n}(t_j^3 - 3t_j)/\sqrt{6n} = \sum_{j=1}^{n}\{(x_j - \bar{x})/s\}^3/\sqrt{6n} = \sqrt{nb_1/6} \text{ and}$$

$$V_4 = \sum_{j=1}^{n} h_4(t_j)/\sqrt{n} = \sum_{j=1}^{n}(t_j^4 - 6t_j^2 + 3)/\sqrt{24n}$$

$$= \sum_{j=1}^{n}\left[\{(x_j - \bar{x})/s\}^4 - 6\{(x_j - \bar{x})/s\}^2 + 3\right]/\sqrt{24n} = (b_2 - 3)\sqrt{n/24}$$

For the Extreme Value distribution $\sqrt{b_1}$ and $b_2$ can also be simply related to $V_3$ and $V_4$. However the orthonormal polynomials are more complicated. Appendix A gives the formulae for calculating $V_1$, $V_2$, $V_3$ and $V_4$ for any univariate distribution. Using these formulae and the moments also given in Appendix A we find that for the Extreme Value distribution

$h_0(t) = 1$, $h_1(t) = \sqrt{6}(t - \gamma)/\pi$, $h_2(t) = (6/\pi^2)\{(t - \gamma)^2 - 1.462(t - \gamma) - \pi^2/6\}/\sqrt{8.392}$ ,

$h_3(t) = (6\sqrt{6}/\pi^3)\{(t - \gamma)^3 - 4.662(t - \gamma)^2 - 2.069(t - \gamma) + 5.265\}/\sqrt{20.00}$ , and

$h_4(t) = (36/\pi^4)\{(t - \gamma)^4 - 9.693(t - \gamma)^3 + 10.792(t - \gamma)^2 + 31.160(t - \gamma) - 9.060\}/\sqrt{219.72}$ ,

where $\gamma$ is Euler's constant, approximately 0.57722.

For the Extreme Value distribution the MOM (method of moments) estimators $\tilde{\alpha}$ of $\alpha$ and $\tilde{\beta}$ of $\beta$ are given by

$$\tilde{\alpha} = \bar{X} - (\gamma/\pi)\sqrt{6\left\{\sum_j (X_j - \bar{X})^2/n\right\}} \tag{2.3}$$

and

$$\tilde{\beta} = \sqrt{6\left\{\sum_j (X_i - \bar{X})^2/n\right\}}\ /\pi . \tag{2.4}$$

If MOM estimation is used in calculating the $V_r$ for the Extreme Value distribution then $V_1 \equiv 0$, $V_2 \equiv 0$, as for the Normal distribution, and the first two non-trivial smooth test statistics $V_3$ and $V_4$ are given by

$$V_3 = (\sqrt{b_1} - 1.139547)/\sqrt{20/n} \text{ and } V_4 = (b_2 - 7.55\sqrt{b_1} + 3.21)/\sqrt{(219.72/n)} , \tag{2.5}$$

where 1.139547 is the coefficient of skewness for the Extreme Value distribution.

Rayner and Best (1989) observe that for the Normal distribution $V_3$ and $V_4$ have an asymptotic $N(0, 1)$ distribution. In Appendix A we show that for the Extreme Value the $V_r$ variances are no longer unity. Moreover for the normal distribution $V_3$ and $V_4$ are asymptotically independent. Clearly this is not the case for the Extreme Value distribution, as $V_4$ depends on both $\sqrt{b_1}$ and $b_2$. In interpreting $V_4$ we observe that if $V_3$ is not significant and $V_4$ is, then this significance is due to $b_2$. If $V_3$ and $V_4$ are both significant then it is not clear whether the significance of $V_4$ is due to $\sqrt{b_1}$ or $b_2$.

Throughout the remainder of this article we will present results in terms of $V_3$ and $V_4$ instead of $\sqrt{b_1}$ and $b_2$ .

For the Extreme Value distribution if the maximum likelihood estimators of $\alpha$ and $\beta$ are used it is no longer true that $V_1 \equiv 0$ and $V_2 \equiv 0$. Moreover when maximum likelihood estimators are used $V_3$ and $V_4$ no longer have simple links to $\sqrt{b_1}$ and $b_2$ and can no longer be simply interpreted. Note that Rayner et al. (1995) and Henze and Klar (1996) caution about using $\sqrt{b_1}$ and $b_2$ to 'diagnose' the type of alternative. Previous studies on the zero inflated Poisson distribution by Thas and Rayner (2005) and on the logistic by Thas et al. (2007) indicate that tests based on $V_3$ and $V_4$ using MOM estimators have small sample powers very similar to tests based on $V_3$ and $V_4$ using MLE estimators. Although we have not done such a comparison here, we expect a similar result.

An advantage of using the smooth test components $V_3$ and $V_4$ is that their squares are score test statistics for testing for focused alternatives, and hence are weakly optimal. Sums of squares of such components are also score test statistics for testing for more omnibus alternatives, and are again weakly optimal. These comments seem likely to apply when MOM estimators, rather than MLE estimators are used, as $V_3$ and $V_4$ are asymptotically normal. Klar (2000, Theorem 2.1) gives a proof. This asymptotic normality implies a quadratic form statistic similar to a score test statistic can be given.

Further, these components may be used to obtain an easily calculated density estimate for the data. In the Normal case such an estimate is known as a Gram-Charlier Type A series density estimate and is given by $\phi(t)\{1 + V_3 h_3(t)/\sqrt{n} + V_4 h_4(t)/\sqrt{n} + \ldots\}$. An analogous estimate is available for the Extreme Value. Although these density estimates can be negative, this problem can be rectified as, for example, in Gajek (1986).

Stephens (1977) has suggested that tests of fit for the Extreme Value distribution should be based on statistics themselves based on the empirical distribution function (EDF). As with other distributions, the test based on the Anderson-Darling statistic, $A^2$ say, seems to perform well for most alternatives, and so we compare it with the tests based on $V_3^2$ and $V_4^2$. Suppose that $\alpha$ and $\beta$ are estimated by maximum likelihood, with the estimators being denoted by $\hat{\alpha}$ and $\hat{\beta}$ respectively. Further, suppose that $z_{(j)} = \exp\{-\exp(-(x_{(j)} -\hat{\alpha})/\hat{\beta}\}$ where the $x_{(j)}$ are the ordered data and they and the $z_{(j)}$ are in ascending order. Then

$$A^2 = -n - (1/n)\sum_{j=1}^{n}(2j-1)\left\{\ln z_{(j)} + \ln\left(1 - z_{(n+1-j)}\right)\right\} \tag{2.6}$$

Here $\hat{\beta} = \bar{X} - \left\{\sum_j X_j \exp(X_j/\hat{\beta})\right\} / \left\{\sum_j \exp(X_j/\hat{\beta})\right\}$ and this equation needs to be solved iteratively. Then $\hat{\alpha} = -\hat{\beta} \log\left\{\sum_j \exp(-X_j/\hat{\beta})/n\right\}$.

We also consider two more tests of fit that have recently appeared in the literature. Pal et al. (2006, section 9.4) suggest testing for the Extreme Value distribution using the test based on the $W_n$ statistic introduced by Shapiro and Brain (1987). To define $W_n$ we follow Pal et al. (2006, p.244) and first define

$$f_j = \ln\{(n + 1)/(n - j + 1)\}, \text{ for } j = 1, \ldots, (n-1), \text{ with } f_n = n - \sum_{j=1}^{n-1} f_j, \text{ and}$$

$$f_{n+j} = f_j \{1 + \ln(f_j)\} - 1, \text{ for } j = 1, \dots, (n-1), \text{ with } f_{2n} = 0.4228n - \sum_{j=1}^{n-1} f_{n+j} \,. \qquad (2.7)$$

Then write

$$A_1 = \sum_{j=1}^{n-1} f_j X_{(j)} \quad \text{and} \quad A_2 = \sum_{j=1}^{n-1} f_{n+j} X_{(j)} \,, \qquad (2.8)$$

in which the $X_{(j)}$ are the ordered data, and finally

$$W_n = \{0.6079A_2 - 0.2570A_1\}^2 / \{n \sum_{j=1}^{n} (X_j - \bar{X})^2 \}, \qquad (2.9)$$

To carry out a test of fit based on $W_n$ at a given significance level we need upper and lower critical values corresponding to half that significance level. Critical values are given by Pal et al. (2006, p.245); for $n = 40$ the lower 0.025 value is 0.4054 and the upper 0.025 value is 0.90358.

Liao and Shimokawa (1999) suggest testing for the Extreme Value distribution using a statistic $L_n$ given by

$$L_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \max \left\{ \frac{j}{n} - z_{(j)}, z_{(j)} - \frac{j-1}{n} \right\} / \sqrt{z_{(j)} \left(1 - z_{(j)}\right)} \,. \qquad (2.10)$$

This test was suggested in an attempt to combine the desirable properties of the Kolmogorov-Smirnov, Cramer-Von Mises and Anderson-Darling tests. Liao and Shimokawa (1999, Table II) gives the critical values and for $n = 40$ the 0.05 value is 0.992.

In the following section we use one-sided tests for $V_3^2$, $V_4^2$, $A^2$ and $L_n$. We use a two-sided test for $W_n$.

## 3. Critical Values and Powers

In Table 3 we study the approach to the asymptotic critical value of 3.84 for $(20.00/31.68)V_3^2$ and $(219.72/283.86)V_4^2$ as the sample size $n$ increases with test size 0.05. The multipliers $(20/31.68)$ and $(219.72/283.86)$ are discussed in Appendix A. It appears $V_3$ and $V_4$ approach normality as $n$ tends to infinity but that convergence is slow. Hence for finding $p$-values for tests based on $V_3^2$ and $V_4^2$ the parametric bootstrap is recommended. See Gürtler and Henze (2000) for a discussion of the parametric bootstrap in the goodness of fit context.

In the following power study random values from the distributions shown were found, or derived from, IMSL (1995) routines. For $n = 40$ and a test size of 5% a simulation study based on 100,000 Monte Carlo samples found the 5% critical values of $(20/31.68)V_3^2$ and $(219.72/283.86)V_4^2$, to be 1.00 and 2.88 for a null standard Extreme Value distribution. Following Stephens (1977), $A^2(1 + 2/n)$ has 95% critical value of 0.76.

**Table 3**. Critical values of $(20/31.68)\,V_3^2$ and $(219.72/283.86)\,V_4^2$ for test size 0.05, *n* as shown and 10,000 Monte Carlo simulations.

| $n$ | $(20/31.68)\,V_3^2$ | $(219.72/283.86)\,V_4^2$ |
|---|---|---|
| 20 | 1.06 | 2.88 |
| 100 | 2.00 | 2.87 |
| 500 | 3.05 | 2.86 |
| 1,000 | 3.39 | 3.09 |
| 5,000 | 3.68 | 3.30 |
| 20,000 | 3.69 | 3.59 |
| 100,000 | 3.72 | 3.78 |
| $\infty$ | 3.84 | 3.84 |

**Table 4.** Powers of five tests of fit for the Extreme Value distribution for *n* = 40, test size 0.05 and 10,000 Monte Carlo simulations.

| Alternative | $\sqrt{\beta_1}$ | $(20/31.68)\,V_3^2$ | $A^2$ | $W_n$ | $(219.72/283.86)\,V_4^2$ | $L_n$ |
|---|---|---|---|---|---|---|
| Uniform (0, 1) | 0.0 | 0.69 | 0.61 | 0.17 | 0.61 | 0.48 |
| Exponential (1) | 2.0 | 0.16 | 0.71 | 0.61 | 0.20 | 0.72 |
| Weibull (2.5) | 0.4 | 0.23 | 0.15 | 0.18 | 0.31 | 0.10 |
| Weibull (2) | 0.6 | 0.07 | 0.07 | 0.04 | 0.10 | 0.05 |
| Gamma (2) | 1.4 | 0.05 | 0.18 | 0.13 | 0.05 | 0.18 |
| Beta (2,1.5) | − 0.2 | 0.90 | 0.66 | 0.70 | 0.91 | 0.51 |
| Beta (2,2) | 0.0 | 0.63 | 0.37 | 0.38 | 0.67 | 0.25 |
| Beta (2,2.5) | 0.2 | 0.38 | 0.23 | 0.19 | 0.44 | 0.14 |
| Extreme Value | 1.1 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |
| Lognormal (0.5, 1) | 6.2 | 0.52 | 0.90 | 0.90 | 0.65 | 0.92 |
| Normal (0,1) | 0.0 | 0.59 | 0.46 | 0.61 | 0.70 | 0.40 |

Table 4 shows powers for tests based on $V_3^2$, $A^2$, $W_n$, $V_4^2$ and $L_n$ for various alternatives. These alternatives were chosen to give a range of values for $\sqrt{\beta_1}$, the population skewness coefficient, and so that commonly used alternatives to the Extreme Value such as the Weibull, Gamma, Normal and Lognormal were included. We see that the test based on $\sqrt{b_1}$, or, equivalently, $V_3^2$, does well for symmetric alternatives as does that based on $V_4^2$, and the $A^2$ test does well for skewed alternatives. It was expected that the $V_3^2$ test would have poor power for alternatives with skewness coefficient 'near' 1.13955. The $W_n$ test has similar power to that based on $A^2$, except for the uniform and normal alternatives, where the $W_n$ test is worse and better respectively. It appears that $L_n$ fares a little worse than $A^2$, and $V_4^2$ and does a little better than $V_3^2$ for these alternatives.

## 4. Examples

**Port Pirie data**. We find $(\hat{\alpha}, \hat{\beta})$ = (3.87, 0.195) as does Coles (2004, p.63). Further, $V_3^2$ = 0.71, $V_4^2$ = 0.36, $W_n$ = 0.68 and $A^2$ = 0.17 with parametric bootstrap $p$-values, based on 10,000 simulations, of 0.41, 0.55, 0.60 and 0.95 respectively. Appendix B explains how to find parametric bootstrap $p$-values. It appears the Port Pirie data are well fitted by the Extreme Value distribution and this agrees with the graphical assessment of Coles (2004, figure 3.6). For these data the test based on $V_3^2$ gives a smaller $p$-value than that based on $A^2$.

**Glass fibre data**. In his analysis of these data, Coles (2004, p.64) used the negated data. If we do the same we find $(\hat{\alpha}, \hat{\beta})$ = (– 1.66, 0.27), $V_3^2$ = 0.19, $V_4^2$ = 0.03, $W_n$ = 0.73 and $A^2$ = 0.92. The parametric bootstrap $p$-values are 0.60, 0.85, 0.31 and 0.02. This time $A^2$ gives a significantly smaller $p$-value. A quantile-quantile plot as in Coles (2004, Figure 3.7) exhibits some non-linearity in agreement with the $A^2$ $p$-value. We suggest that graphical assessment of fit should usually be complemented by use of $V_3^2$, $V_4^2$, $W_n$ and $A^2$. Given the significant $A^2$ value perhaps a three parameter model is needed for these data and that is the approach of Coles (2004, p.64), although he does not carefully validate his model. For these data with $\sqrt{b_1}$ = 0.90 we would not expect the test based on $V_3^2$ to have large power.

## 5. Conclusion

For the Extreme Value distribution the Anderson-Darling test based on $A^2$ provides a more powerful test than the tests based on the statistics $V_3^2$ and $V_4^2$ when there are skewed alternatives, while the reverse is true for symmetric alternatives. The test based on $W_n$ has powers similar to those based on the Anderson-Darling test for most alternatives. The test based on $V_4^2$ does a little better than that based on $V_3^2$, while the $A^2$ test is probably the best overall. The statistics $V_3^2$ and $V_4^2$ can be used to suggest alternative models and can give density estimates.

We also note that if we wish to test a set of data for agreement with the two parameter Weibull distribution then each data point should be transformed to the negative of its natural logarithm. Then under the null hypothesis the transformed data will follow an Extreme Value distribution.

## References

Coles, S., 2004. *An Introduction to Statistical Modelling of Extreme Values*. Springer, New York.

Gajek, L., 1986. On improving density estimates which are not bone fide functions. *Ann. Statist.* 14, 1612–1618.

Gürtler, N., Henze, N., 2000. Recent and classical goodness-of-fit tests for the Poisson distribution. *Journal of Statistical Planning and Inference*, 90, 207–225.

Henze, N., Klar, B., 1996. Properly rescaled components of smooth tests of fit are diagnostic. *Austral. J. Statist.* 38, 61–74.

IMSL, 1995. *Fortran Numerical Libraries*. Visual Numerics, Houston.

Klar, B., 2000. Diagnostic smooth tests of fit. *Metrika* 52, 237–252.

Lancaster, H.O., 1969. *The Chi-Squared Distribution*. Wiley, New York.

Karlin, S., Altschul, S.F., 1990. Methods for assessing the statistical significance of molecular sequence features using general scoring schemes. *Proc. Math. Acad. Sci. USA*, 87, 2264–2268.

Liao, M., Shimokawa, T., 1999. A new goodness of fit test for Type-I Extreme-Value and Weibull distributions with estimated parameters. *Journal of Statistical Computation and Simulation* 64, 23–48.

Pal, N., Chun, J., Wooi, K.L., 2006. *Handbook of Exponential and Related Distributions for Engineers and Scientists*. Chapman and Hall/CRC, Boca Raton.

Rayner, J.C.W., Best, D.J., 1989. *Smooth Tests of Goodness of Fit*. Oxford University Press, New York.

Rayner, J.C.W., Best, D.J., Mathews, K.L., 1995. Interpreting the skewness coefficient. *Commun. Statist.-Theor. Meth.* 24(3), 593–600.

Shapiro, S.S., Brain, C.W., 1987. *W*-test for the Weibull distribution. *Communications in Statistics – Simulation and Computation* 16, 209–219.

Smith, R.L., Naylor, J.C., 1987. A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Applied Statistics*, 36 358–369.

Stephens, M.A., 1977. Goodness of fit for the extreme value distribution. *Biometrika* 64, 583–588.

Stuart, A. and Ord, K., 1994. *Kendall's Advanced Theory of Statistics*, Volume 1, Sixth edition, Edward Arnold, London.

Thas, O., Rayner, J.C.W., 2005. Smooth tests for the zero inflated Poisson distribution. *Biometrics* 61 (3), 808–815.

Thas, O., Rayner, J.C.W., Best, D.J., 2007. Smooth tests for the logistic distribution. Submitted.

Thas, O., Rayner, J.C.W., De Boeck, B., 2007. Generalised Emerson recurrence relations. In preparation.

## Appendix A: Definition of $V_3$ and $V_4$ for the Extreme Value Distribution

Suppose, as usual, that $\mu$ is the population mean and $\mu_r$, $r = 2, 3, 4, \ldots$ are the central moments. For the standard ($\alpha = 0$, $\beta = 1$) Extreme Value distribution we can derive the following moment values:

$$\mu = 0.57722, \quad \mu_2 = 1.64493, \quad \mu_3 = 2.40411, \quad \mu_4 = 14.61136,$$

$$\mu_5 = 64.43235, \quad \mu_6 = 406.87347, \quad \mu_7 = 2815.13 \text{ and } \mu_8 = 22630.61. \tag{A.1}$$

The constant $C_3$ in the denominator of $V_3$ given subsequently can be derived by substituting these moment values in

$$C_3 = \left\{ \mu_6 - 2a_3\mu_5 + (a_3^2 - 2b_3)\mu_4 + 2(a_3 b_3 - r_3)\mu_3 + (b_3^2 + 2a_3 r_3)\mu_2 + r_3^2 \right\} / \mu_2^3 \tag{A.2}$$

where

$$d_3 = \mu_4 - \mu_3^2 / \mu_2 - \mu_2^2, \quad a_3 = (\mu_5 - \mu_3\mu_4 / \mu_2 - \mu_2\mu_3) / d_3,$$

$$b_3 = (\mu_4^2 / \mu_2 - \mu_2\mu_4 - \mu_3\mu_5 / \mu_2 + \mu_3^2) / d_3 \text{ and } r_3 = (2\mu_3\mu_4 - \mu_3^3 / \mu_2 - \mu_2\mu_5) / d_3. \tag{A.3}$$

For the normal distribution this constant $C_3$ is 6.0 and for the Extreme Value distribution is 20.0. However notice that asymptotically, while var($\sqrt{b_1}$) for a normal distribution is $6/n$, var($\sqrt{b_1}$) $\neq 20/n$ for the Extreme Value distribution. In fact, Stuart and Ord (1994, Exercise 10.26) give, for any distribution, the large sample variance of $\sqrt{b_1}$ as

$$n^{-1} \left\{ \mu_6 / \mu_2^3 - 6\mu_4 / \mu_2^2 + 9 + (\mu_3 / \mu_2^{1.5})(9\mu_4 / \mu_2^2 + 35) - 3\mu_5\mu_3 / \mu_2^4 \right\}. \tag{A.4}$$

Thus for the Extreme Value distribution, in large samples, var($\sqrt{b_1}$) $= 31.68/n$.

The expressions for $V_3$ and $V_4$ quoted in section 2 above can be derived from

$$V_3 = \sum_{j=1}^{n} h_3(t_j)/\sqrt{n}$$

$$= \mu_2^{-3/2} \sum_{j=1}^{n} \left\{ (t_j - \mu)^3 - a_3(t_j - \mu)^2 - b_3(t_j - \mu) - r_3 \right\} / \sqrt{C_3/n} \qquad (A.5)$$

and

$$V_4 = \sum_{j=1}^{n} h_4(t_j)/\sqrt{n}$$

$$= \mu_2^{-2} \sum_{j=1}^{n} \left\{ (t_j - \mu)^4 + a_4(t_j - \mu)^3 + b_4(t_j - \mu)^2 + r_4(t_j - \mu) + d_4 \right\} / \sqrt{C_4/n} \qquad (A.6)$$

Now for $V_4$, let

$$c_1 = \mu_2\mu_4 - \mu_3^2 - \mu_2^3, \quad c_2 = \mu_3\mu_4 - \mu_2^2\mu_3 - \mu_2\mu_5,$$

$$c_3 = \mu_3\mu_5 - \mu_2^2\mu_4 - \mu_4^2 - \mu_2\mu_3^2, \quad \text{and} \quad c_4 = \mu_2^2\mu_5 - 2\mu_2\mu_3\mu_4 + \mu_3^3. \qquad (A.7)$$

Now put $e = -(c_1\mu_6 + c_2\mu_5 + c_3\mu_4 + c_4\mu_3)$ and define $a_4, b_4, c_4$ and $d_4$ by

$$a_4 c_1 e = c_1^2 \mu_7 + c_1 c_2 \mu_6 + c_1 c_3 \mu_5 + c_1 c_4 \mu_4,$$

$$b_4 c_1 e = c_1 c_2 \mu_7 + c_2^2 \mu_6 + c_2 c_3 \mu_5 + c_2 c_4 \mu_4 - e\left(\mu_2\mu_6 - \mu_3\mu_5 - \mu_2^2\mu_4\right),$$

$$r_4 c_1 e = c_1 c_3 \mu_7 + c_2 c_3 \mu_6 + c_3^2 \mu_5 + c_3 c_4 \mu_4 - e\left(\mu_2\mu_3\mu_4 - \mu_3\mu_6 - \mu_2^2\mu_5 + \mu_4\mu_5\right),$$

$$d_4 c_1 e = c_1 c_4 \mu_7 + c_2 c_4 \mu_6 + c_3 c_4 \mu_5 + c_4^2 \mu_4 - e\left(\mu_2\mu_4^2 - \mu_3^2\mu_4 + \mu_2\mu_3\mu_5 - \mu_2^2\mu_6\right), \qquad (A.8)$$

Finally define $C_4$ by

$$\mu_2^4 C_4 = \mu_8 + 2a_4\mu_7 + \left(a_4^2 + 2b_4\right)\mu_6 + 2\left(a_4 b_4 + r_4\right)\mu_5$$

$$+ \left(b_4^2 + 2a_4 r_4 + 2d_4\right)\mu_4 + 2\left(a_4 d_4 + b_4 r_4\right)\mu_3 + \left(r_4^2 + 2b_4 d_4\right)\mu_2 + d_4^2. \qquad (A.9)$$

These formulae for $V_3$ and $V_4$ apply for any distribution and can be derived using the determinant formula in Lancaster (1969, p.49) or the recurrence relation discussed in Thas et al. (2007).

For $V_4$ we find $C_4 = 219.72$ after substituting in $\mu_2^4 C_4$. Although we don't give details here it can be shown that if MOM estimators are used, asymptotically var($V_4$) = 1.2919 = 283.86/219.72.

For completeness note that

$$h_1(t) = (t - \mu)/\sqrt{\mu_2} \text{ and } h_2(t) = \{(t - \mu)^2 - \mu_3(t - \mu)/\mu_2 - \mu_2\}/\sqrt{\left(\mu_4 - \mu_3^2/\mu_2 - \mu_2^2\right)}.$$

## Appendix B: Parametric Bootstrap *p*-values

Gürtler and Henze (2000, p. 223) suggest p-values can be obtained using an analogue of the parametric bootstrap. If $W_n$ denotes a test statistic calculate $w_n := W_n(x_1, x_2, \ldots, x_n)$ where $x_1, x_2, \ldots, x_n$ denotes, as usual, the data. Find estimates $\hat{\alpha}$ and $\hat{\beta}$ (not necessarily the maximum likelihood estimates) from the data and conditional on those estimates, generate $B = 10{,}000$ say pseudo-random samples of size $n$, each having the Extreme Value ($\hat{\alpha}$, $\hat{\beta}$) distribution. For $j = 1, \ldots, B$ compute the value $W^*_{n,j}$ on each random sample. Then the parametric bootstrap p-value is the proportion of $W^*_{n,j}$ that is at least the observed $w_n$, namely $\sum_{j=1}^{B} I\left(W^*_{n,j} \geq w_n\right) / B$.

The above requires random Extreme Value ($\alpha$, $\beta$) values. Assuming a random uniform (0, 1) generator is available, then a random Extreme Value ($\alpha$, $\beta$) value is $\alpha - \beta \ln(-\ln r)$, where $r$ is a random uniform (0, 1) value. To obtain p-values for two-tailed tests proceed as above and find the p-value, $P$ say. Then if $P \leq 0.5$ the two-tailed p-value is $2P$. If $P > 0.5$ the two-tailed p-value is $2(1 - P)$.