RC 12210 (#54860) 10/6/86
Mathematics 160 pages

THE THEORY OF PROBABILITY WEIGHTED MOMENTS

J.R.M. Hosking

IBM T.J. Watson Research Center
Mathematical Sciences Dept.
Yorktown Heights, New York 10598

On leave from:
Institute of Hydrology
Crowmarsh Gifford
Wallingford, Oxon OX10 8BB
England

Typed by: Barbara A. Newman

1 October 1986

Abstract: Probability weighted moments (PWMs) are expectations of certain functions of a random variable. They can be defined for any random variable whose mean exists and were devised by Greenwood *et al.* (1979) primarily as an aid to estimating the parameters of the Wakeby distribution. Greenwood *et al.*'s use of PWMs is, however, but one application of a general theory which is founded on PWMs and covers:

the summarization and description of theoretical probability distributions;
the summarization and description of observed data samples;
nonparametric estimation of the underlying distribution of an observed sample;
estimation of parameters and quantiles of probability distributions;
hypothesis tests for probability distributions.

The theory involves such established and efficient procedures as the use of order statistics and Gini's mean difference statistic, and gives rise to some promising innovations such as the new measures of skewness and kurtosis described in section 3, and new methods of parameter estimation for several distributions. The theory of PWMs parallels the theory of (conventional) moments, as the above list of applications might suggest. The main advantage of PWMs over conventional moments is that PWMs, being linear functions of the data, suffer less from the effects of sampling variability: PWMs are more robust than conventional moments to outliers in the data, enable more secure inferences to be made from small samples about an underlying probability distribution, and frequently yield more efficient parameter estimates than the conventional moment estimates.

CONTENTS

THE THEORY OF PROBABILITY WEIGHTED MOMENTS

1. Introduction

It is standard statistical practice to summarize a probability distribution or an observed data set by its moments or cumulants. It is also common, when fitting a parametric distribution to a data set, to estimate the parameters by equating the sample moments to those of the fitted distribution. Yet moment-based methods, although long-established in statistics, are not always satisfactory. It is sometimes difficult to assess exactly what information about the shape of a distribution is conveyed by its moments of third and higher order; the numerical values of sample moments, particularly when the sample is small, can be very different from those of the probability distribution from which the sample was drawn; and the estimated parameters of distributions fitted by the method of moments are often markedly less accurate than those obtainable by other estimation procedures such as the method of maximum likelihood.

This report follows an alternative approach based on quantities, which we call L-moments, which are analogous to the conventional moments but which can be estimated by linear combinations of an ordered data set, *i.e.* by L-statistics. L-moments have the theoretical advantages over conventional moments of being able to characterize a wider range of distributions and, when estimated from a sample, of being more robust to the presence of outliers in the data. Our experience (not all of it presented here) also shows that, compared with conventional moments, L-moments are less subject to bias in estimation, approximate their asymptotic Normal distribution more closely in finite samples, and frequently yield more accurate estimates of the parameters of a fitted distribution. Indeed, parameter estimates obtained from L-moments are sometimes more accurate in small samples than are the maximum-likelihood estimates.

Many statistical techniques are based on the use of linear combinations of the observed data – see David (1981) for examples – but there has not heretofore been developed a unified theory covering the characterization of probability distributions, the summarization of observed data samples, the fitting of probability distributions to data and the testing of hypotheses about fitted distributions. We present such a theory in sections 2-7 below. We take as our starting point the "probability weighted moments" introduced by Greenwood *et al.* (1979) primarily for estimating the parameters of the Wakeby distribution. The theory also includes and enlarges upon some scattered results and techniques described principally by Gini (1912), Nair (1936), Sillitto (1951, 1969), Downton (1966) and Chan (1967). Results for specific distributions are given in Appendix A. Numbered results are proved in Appendix B.

2. Probability weighted moments

Definitions

Let X be a real-valued random variable with distribution function F. Greenwood *et al.* (1979) defined the *probability weighted moments* of X to be the quantities

$$M_{p,r,s} \equiv E\left[X^p \{F(X)\}^r \{1 - F(X)\}^s \right] = \int x^p \{F(x)\}^r \{1 - F(x)\}^s dF(x) ,$$

where p, r and s are real numbers. Note that

(2.1) $M_{p,r,s}$ exists for all $r, s \geq 0$ if and only if $E|X|^p$ exists.

We are usually interested in cases in which r and s are positive integers: we then have

$$M_{p,r,s} = \frac{r!s!}{(r+s+1)!} \, EX^p_{r+1:r+s+1} \tag{2.2}$$

where $X_{k:n}$ is the kth order statistic of a random sample of size n from the distribution F. The definition of $M_{p,r,s}$ is valid both for continuous and for discrete random variables: in the former case we have

$$M_{p,r,s} = \int x^p \{F(x)\}^r \{1 - F(x)\}^s f(x)\, dx \tag{2.3}$$

where f is the probability density function of X. We also have

$$M_{p,r,s} = \int_0^1 \{x(F)\}^p F^r (1 - F)^s dF \tag{2.4}$$

where $x(F)$ is the quantile function, or inverse cumulative distribution function, of X.

The quantities $M_{p,r,s}$ may be used to describe and characterize probability distributions. One possible approach is to work with $M_{p,0,0}$, $p = 1, 2, \ldots$: these are just the conventional noncentral moments of X. We shall instead work with the moments $M_{1,r,s}$, into which X enters linearly, and in particular with the quantities (which we shall also refer to as PWMs)

$$\alpha_r = M_{1,0,r} = E\left[X\{1 - F(X)\}^r\right], \quad r = 0, 1, \ldots,$$

$$\beta_r = M_{1,r,0} = E\left[X\{F(X)\}^r\right], \quad r = 0, 1, \ldots;$$

note that $r\alpha_{r-1} = EX_{1:r}$ and $r\beta_{r-1} = EX_{r:r}$ are expected values of extreme order statistics.

Characterizing a probability distribution

Either of the sets $\{\alpha_r: r = 0, 1, \ldots\}$ or $\{\beta_r: r = 0, 1, \ldots\}$ is sufficient to characterize a probability distribution, *i.e.*

(2.5) let X and Y be real-valued random variables with distribution functions F and G respectively; suppose that $E|X| < \infty$ and $E|Y| < \infty$; let $\alpha_r^{(X)} = E\left[X\{1 - F(X)\}^r\right]$, $\beta_r^{(X)} = E\left[X\{F(X)\}^r\right]$, and similarly $\alpha_r^{(Y)}$, $\beta_r^{(Y)}$; then F and G are identical if and only if $\alpha_r^{(X)} = \alpha_r^{(Y)}$, $r = 0, 1, \ldots$ (or if and only if $\beta_r^{(X)} = \beta_r^{(Y)}$, $r = 0, 1, \ldots$).

The characterizations of a distribution by the α_r and by the β_r are interchangeable, because the α_r and β_r are functions of each other: we have in general

$$\alpha_r = \sum_{k=0}^{r} (-1)^k \binom{r}{k} \beta_k, \qquad \beta_r = \sum_{k=0}^{r} (-1)^k \binom{r}{k} \alpha_k, \tag{2.6}$$

and in particular

$$
\begin{aligned}
\alpha_0 &= \beta_0, & \beta_0 &= \alpha_0, \\
\alpha_1 &= \beta_0 - \beta_1, & \beta_1 &= \alpha_0 - \alpha_1, \\
\alpha_2 &= \beta_0 - 2\beta_1 + \beta_2, & \beta_2 &= \alpha_0 - 2\alpha_1 + \alpha_2, \\
\alpha_3 &= \beta_0 - 3\beta_1 + 3\beta_2 - \beta_3, & \beta_3 &= \alpha_0 - 3\alpha_1 + 3\alpha_2 - \alpha_3.
\end{aligned}
$$

Because the complete set of α_r (or β_r) characterizes a distribution, it is natural to use the first few α_r (or β_r) to summarize the main features of a distribution: this point is pursued in section 3. Another natural application is to use the first few α_r (or β_r) to estimate the parameters of a distribution: many examples are given in Appendix A.

Since the α_r are sufficient to characterize a distribution, it should be possible to obtain an "inversion theorem" expressing the distribution function of a random variable as a function of the PWMs α_r. This is most easily done in terms of the L-moments of a distribution, defined in section 3, so we postpone consideration of the matter until then.

We can also obtain results which relate the lower (respectively upper) tail of the distribution of a random variable X to the behavior of α_r (respectively β_r) for large r. For example

(2.7) if $E|X|$ exists and X has an exponential upper tail, *i.e.* $1 - F(x) \sim Ae^{-Bx}$ as $x \to \infty$, then $r\beta_{r-1} \sim \gamma \log r$ as $r \to \infty$, with $\gamma = 1/B$;

(2.8) if $E|X|$ exists and X has a power-law upper tail, *i.e.* $1 - F(x) \sim Ax^{-B}$ as $x \to \infty$, then $r\beta_{r-1} \sim \gamma r^{\delta}$ as $r \to \infty$, with $\gamma = A^{1/B}\Gamma(1 - 1/B)$, $\delta = 1/B$.

3. *L*-moments

Definitions and basic properties

Probability weighted moments characterize a distribution, but are not particularly meaningful in themselves. It is therefore useful to define functions of PWMs which give a descriptive summary of the location, scale and shape of a probability distribution. We define the *L-moments* of a real-valued random variable X to be the quantities

$$\lambda_r \equiv r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} EX_{r-k:r}, \quad r = 1, 2, \ldots .$$

The L in "L-moments" emphasizes that λ_r is a *linear* function of the expected order statistics. Furthermore, as we shall see in section 4, the natural estimator of λ_r based on an observed sample of data is a linear combination of the ordered data values, *i.e.* an L-statistic.

In terms of the previously defined quantities α_r and β_r we have

$$\lambda_{r+1} = (-1)^r \sum_{k=0}^{r} p^*_{r,k}\, \alpha_k = \sum_{k=0}^{r} p^*_{r,k} \beta_k, \quad r = 0, 1, \ldots . \tag{3.1}$$

where

$$p^*_{r,k} = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} .$$

The first few L-moments are

$$\begin{aligned}
\lambda_1 &= EX & &= \alpha_0 & &= \beta_0, \\
\lambda_2 &= \tfrac{1}{2} E(X_{2:2} - X_{1:2}) & &= \alpha_0 - 2\alpha_1 & &= 2\beta_1 - \beta_0, \\
\lambda_3 &= \tfrac{1}{3} E(X_{3:3} - 2X_{2:3} + X_{1:3}) & &= \alpha_0 - 6\alpha_1 + 6\alpha_2 & &= 6\beta_2 - 6\beta_1 + \beta_0, \\
\lambda_4 &= \tfrac{1}{4} E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}) & &= \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3 & &= 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0.
\end{aligned}$$

We also have

$$\lambda_r = \int_0^1 x(F)\, P^*_{r-1}(F)\, dF\,, \quad r = 1, 2, \dots\,, \tag{3.2}$$

where

$$P^*_r(F) = \sum_{k=0}^{r} p^*_{r,k} F^k$$

is the rth shifted Legendre polynomial, discussed further in Appendix C.

We note that

(3.3) λ_r exists for $r = 1, 2, \dots$ if and only if $E|X|$ exists.

Thus we can describe a distribution using L-moments even if some of the conventional moments of the distribution do not exist. Furthermore, such a description is meaningful, because

(3.4) a distribution whose mean exists is characterized by its L-moments $\{\lambda_r;\ r = 1, 2, \dots\}$.

As we shall shortly show, λ_2 is a measure of the scale or dispersion of the random variable X. It is often convenient to standardize the higher moments λ_r, $r \geq 3$, so that they are independent of the units of measurement of X. We therefore define, for a nondegenerate random variable X whose mean exists, the *L-moment ratios* of X to be the quantities

$$\tau_r \equiv \lambda_r / \lambda_2, \quad r = 3, 4, \dots\,.$$

Bounds on the numerical values of the τ_r are given by the following theorem.

(3.5) If X is a nondegenerate random variable and $E|X| < \infty$, then the L-moment ratios of X satisfy $|\tau_r| < 1$, $r = 3, 4, \ldots$.

We can also define a function of PWMs which is analogous to the coefficient of variation: this is the *L-CV*, $\tau \equiv \lambda_2 / \lambda_1$. This quantity is particularly relevant for nonnegative random variables, for which we have the following result:

(3.6) If the nondegenerate random variable X satisfies $X \geq 0$ almost surely, and if $EX < \infty$, then τ, the L-CV of X, satisfies $0 < \tau < 1$.

Summarizing a probability distribution

The L-moments $\lambda_1, \ldots, \lambda_r$ and the L-moment ratios $\tau_3, \ldots, \tau_r$ are functions of the first r α's (or β's) and are the most meaningful quantities for summarizing a distribution in terms of a small number of PWMs. The L-moments are in some ways analogous to the (conventional) central moments and the L-moment ratios are analogous to moment ratios. In particular, as we now show, $\lambda_1, \lambda_2, \tau_3$ and τ_4 may be regarded as measures of location, scale, skewness and kurtosis respectively.

Consider the definition of the λ_r as expectations of sums of order statistics. Clearly λ_1, the mean, is a measure of location. To interpret λ_2, consider the typical configuration of a sample of size two: if the two values tend to be close together, as in Figure 1(a), then λ_2 will be smaller than if they are far apart, as in Figure 1(b). Thus λ_2 can be thought of as measuring the scale or dispersion of the distribution. Samples of size three are relevant to λ_3, which is the central second difference of the median of such a sample. Samples like that of Figure 1(c), which yields a positive central second difference, tend to arise from positively skew distributions; Figure 1(d) is more typical of distributions with negative skewness. For a symmetric distribution we have $\lambda_3 = 0$. Thus λ_3 may be thought of as measuring skewness,

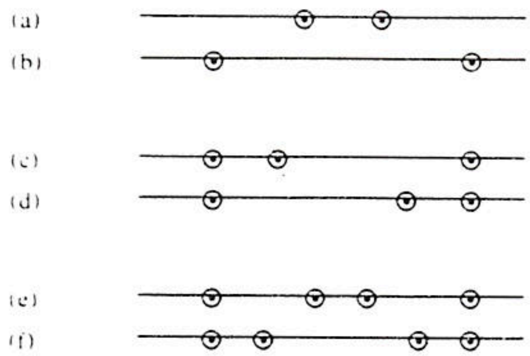

Figure 1. Configurations of samples of size 2, 3 and 4.

although not independently of scale. Similarly, (e) and (f) of Figure 1 illustrate samples of size four. Configuration (e), typical of a heavy-tailed or sharply peaked distribution, has a large positive central third difference, while configuration (f), more typical of a flat or even U-shaped distribution, has a negative central third difference. Thus λ_4, itself the central third difference of the expected order statistics of a sample of size four, measures the same aspects of a distribution as does the fourth central (conventional) moment. The *L*-moment ratios τ_3 and τ_4 are dimensionless analogues of λ_3 and λ_4 respectively and are, therefore, plausible measures of skewness and kurtosis. We conclude that the main features of a probability distribution should be well summarized by the following four measures: the mean or *L-location*, λ_1; the *L-scale*, λ_2; the *L-skewness*, τ_3; and the *L-kurtosis*, τ_4. We now consider these measures in more detail.

The PWM-based measure of location is the mean, λ_1. This is a well-established and familiar quantity which needs no further description or justification here.

The *L*-scale λ_2 is also long-established in statistics, for it is, apart from a scalar multiple, the expectation of Gini's mean difference statistic. This statistic is further discussed in section 4 where we consider the estimation of L-moments. The difference between λ_2 and the more familiar scale measure σ, the standard deviation, may be seen by writing

$$\lambda_2 = \tfrac{1}{2} E(X_{2:2} - X_{1:2}) \qquad \sigma^2 = \tfrac{1}{2} E(X_{2:2} - X_{1:2})^2. \tag{3.7}$$

Thus λ_2 is the half the expected range of sample of size two, whereas σ^2 is half the expected squared range of such a sample. Oja (1981) gives a formal definition of "measure of scale" and shows that λ_2 is indeed a measure of scale according to that definition, *i.e.* that

(a) if X and Y are random variables with L-scale $\lambda_2^{(X)}$ and $\lambda_2^{(Y)}$ respectively, and if $Y = aX + b$, then $\lambda_2^{(Y)} = |a| \lambda_2^{(X)}$;

(b) if X and Y are random variables with L-scale $\lambda_2^{(X)}$ and $\lambda_2^{(Y)}$ respectively, and quantile functions $x(F)$ and $y(F)$ respectively, and if $y(F) - x(F)$ is an increasing function of F, then $\lambda_2^{(Y)} > \lambda_2^{(X)}$

– see Oja's Definition 4.3 and the discussion following it, and Corollary 4.3; note that λ_2 is in Oja's notation $\frac{1}{2}\sigma_2(F)$.

The L-skewness τ_3 is a dimensionless analogue of λ_3. From (3.5) we see that τ_3 takes values between -1 and $+1$; symmetric distributions have $\tau_3 = 0$. Unlike λ_1 and λ_2, τ_3 has not appeared previously in the statistical literature, although Sillitto (1951) defined a scale-dependent skewness measure S_L which is identical to λ_3 (and also a scale-dependent kurtosis measure K_L identical to λ_4). We can, however, write τ_3 as

$$\tau_3 = \frac{EX_{3:3} - 2EX_{2:3} + EX_{1:3}}{EX_{3:3} - EX_{1:3}}. \tag{3.8}$$

which is similar in form to a measure of skewness used by Bowley (1920):

$$B \equiv \frac{(Q_3 - 2Q_2 + Q_1)}{(Q_3 - Q_1)}$$

where $Q_r \equiv x(r/4)$, $r=1, 2, 3$, are the quartiles of X. Skewness measures similar to B but based on quantiles other than the quartiles have been used by Hinkley (1975) and Groeneveld and Meeden (1984). Oja (1981) gives a formal definition of "measure of skewness", but it is not easy to show that τ_3 satisfies it; however, the closely related quantity

$$E\left(\frac{X_{3:3} - 2X_{2:3} + X_{1:3}}{X_{3:3} - X_{1:3}}\right)$$

does satisfy Oja's criteria.

Since $Q_r = F^{-1}[E\{F(X_{r:3})\}]$ there is a kind of conjugacy relationship between λ_3 and B. As a measure of skewness, B could be criticized for being insensitive to the distribution of X any further into the tails than the quartiles. On the other hand, the conventional moment-based measure of skewness,

$$\gamma \equiv E(X - EX)^3/\{E(X - EX)^2\}^{3/2},$$

is so sensitive to the extreme tail of the distribution that it is difficult to estimate accurately in practice when the distribution is markedly skew. We believe that the skewness measure τ_3 steers an advantageous middle course between these extremes.

It is interesting to compare the skewness measures τ_3 and γ for different distributions: the comparison is made graphically in Figure 2. For symmetric distributions both τ_3 and γ are zero, and for many near-symmetric distributions we have $\gamma \approx 6\tau_3$, but in general there is no simple relationship between γ and τ_3. Both γ and τ_3 may yield a large positive skewness either when a distribution has a heavy right tail or when a continuous distribution is reverse J-shaped, *i.e.* has a finite lower bound near which $f(x) \to \infty$. The former case tends to yield particularly high values of γ relative to τ_3, because γ is more sensitive to the extreme tail weight of the distribution. Indeed for some heavy-tailed distributions, γ approaches infinity while τ_3 has still quite a modest value (for example, 0.25 in the case of the generalized logistic distribution).

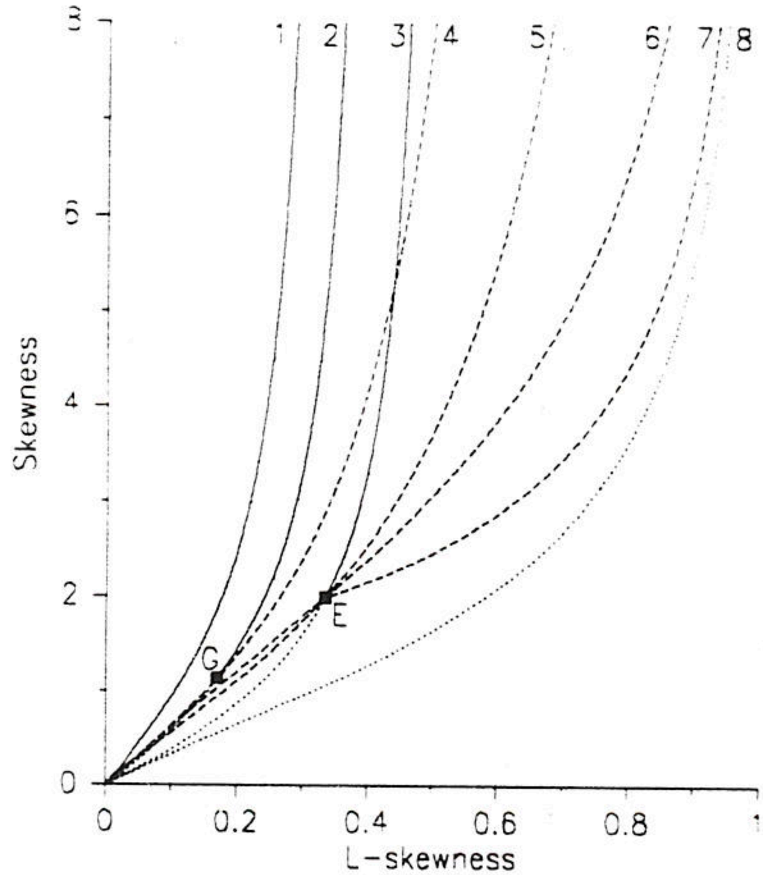

Figure 2. Comparison of skewness and *L*-skewness. Distributions plotted: E = exponential, G = Gumbel, 1 = generalized logistic, 2 = generalized extreme-value, 3 = generalized Pareto, 4 = generalized Normal, 5 = gamma, 6 = Weibull (reverse GEV), 7 = mixture of a standard exponential distribution and a distribution degenerate at zero, 8 = reverse generalized Pareto. Solid, dashed and dotted lines represent distributions whose asymptotic distributions of maxima are of extreme-value types II, I and III respectively, *i.e.*, roughly, distributions with power-law upper tails, exponential upper tails and finite upper limits respectively.

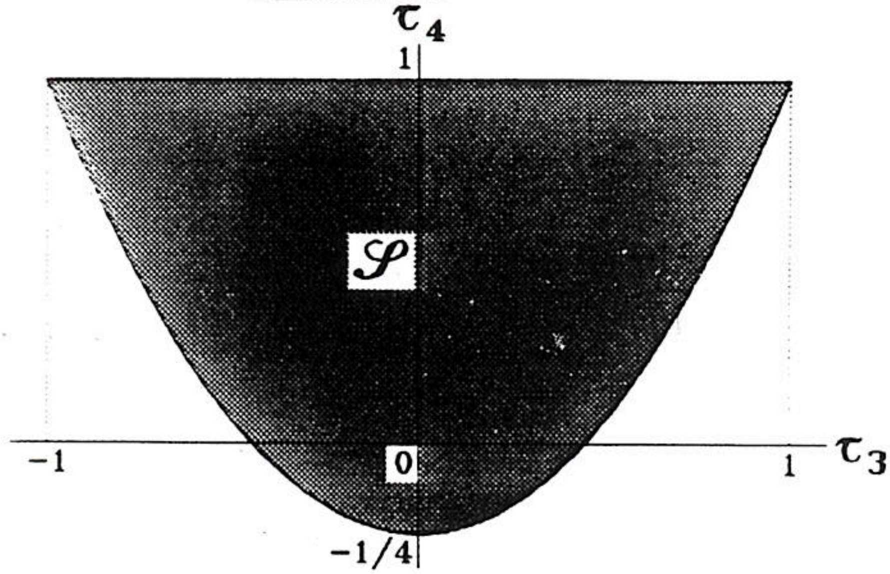

Figure 3. Possible values of τ_3 and τ_4.

Kurtosis, as measured by the moment ratio

$$\kappa \equiv E(X - EX)^4 / \{E(X - EX)^2\}^2,$$

has no unique interpretation. It can be thought of as the "peakedness" of a distribution, or as "tail weight", but only for fairly closely defined families of symmetric unimodal distributions do these interpretations have any demonstrable validity. See Oja (1981) for a discussion. L-kurtosis, τ_4, is equally difficult to interpret uniquely, and is best thought of as a measure similar to κ but giving less weight to the extreme tails of the distribution and being easier to estimate reliably from small samples (as implied by section 4 below). We know from (3.5) that $|\tau_4| < 1$, but a stronger result can be proved:

(3.9) the set of possible values of the L-moment ratios τ_3 and τ_4 of a nondegenerate random variable is

$$\mathscr{S} \equiv \{(\tau_3, \tau_4): (5\tau_3^2 - 1)/4 \le \tau_4 < 1\}$$

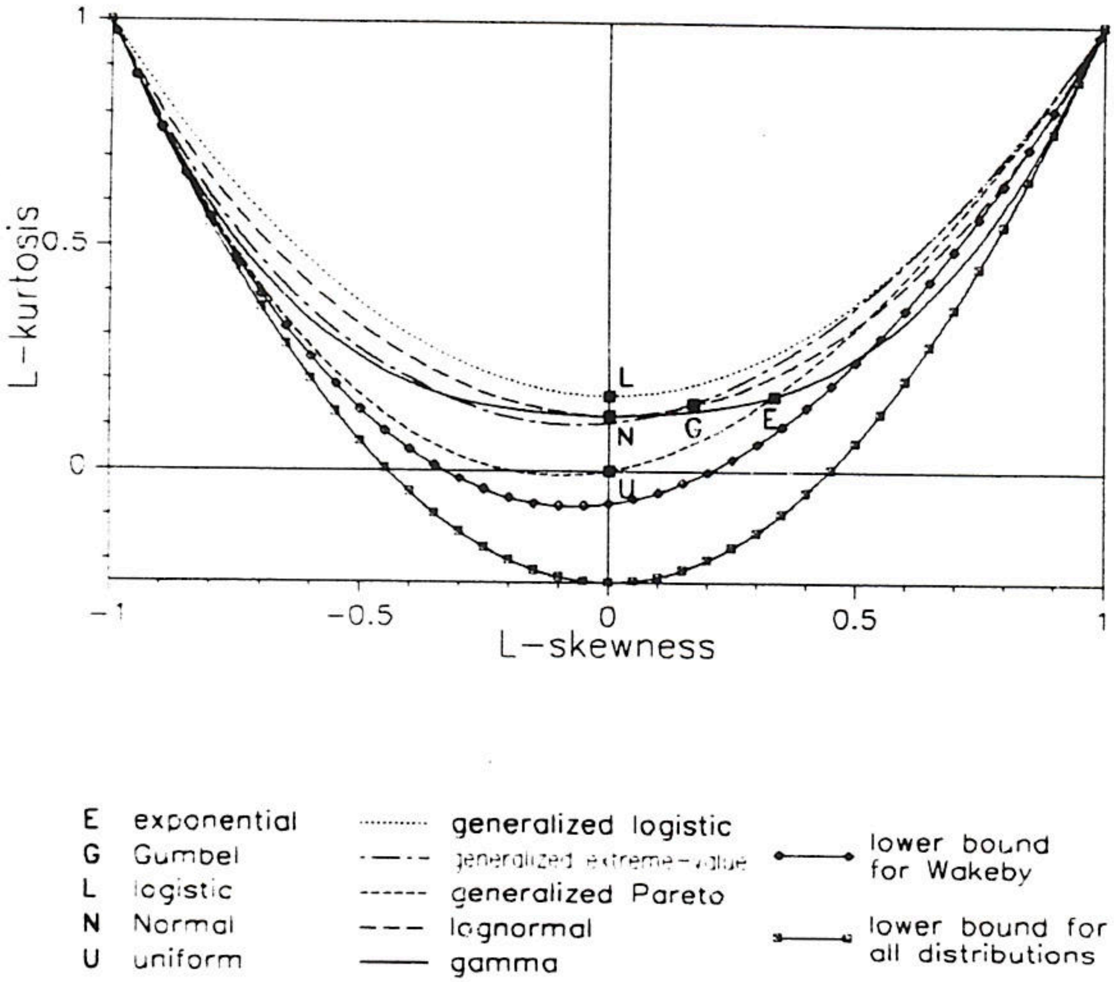

Figure 4 *L*-moment ratios of some common distributions.

The set $\mathscr{P}$ is illustrated in Figure 3. The range of possible values of τ_4 is $-\frac{1}{4} \le \tau_4 < 1$, with the uniform distribution having $\tau_4 = 0$. Indeed the uniform distribution has $\tau_r = 0$ for all $r \ge 3$ and thus plays a central role in L-moment theory akin to that of the Normal distribution in cumulant theory. For the Normal distribution, incidentally, $\tau_4 = 0.1226$. Values of τ_3 and τ_4 for different distributions can be plotted to yield a L-moment ratio diagram, exemplified by Figure 4.

Higher L-moments may be viewed similarly to higher conventional moments. For example τ_5 could be interpreted as a measure of tendency to bimodality, while the τ_r of odd order are generalized skewness measures insofar as symmetric distributions have $\tau_{2r+1} = 0$ for all $r \ge 1$. The use of λ_1, λ_2, τ_3 and τ_4 should, however, be adequate for most applications of L-moment theory.

Approximating a quantile function

An inversion theorem, expressing the quantile function of a random variable in terms of the L-moments, has been given by Sillitto (1969). The theorem is valid for both continuous and discrete random variables, provided that the quantile function is "normalized" (Widder, 1941), *i.e.* that

$$\lim_{\epsilon \to 0} \frac{1}{2} \{x(F + \epsilon) + x(F - \epsilon)\} = x(F) \qquad \text{for all } F \in (0, 1).$$

(3.10) Let X be a real-valued random variable with finite variance, quantile function $x(F)$ and L-moments λ_r, $r \ge 1$. Then the representation

$$x(F) = \sum_{r=1}^{\infty} (2r - 1)\lambda_r P^*_{r-1}(F), \quad 0 < F < 1,$$

is convergent in mean square, *i.e.*

$$R_s(F) \equiv x(F) - \sum_{r=1}^{s} (2r-1)\lambda_r P^*_{r-1}(F),$$

the remainder after stopping the infinite sum after s terms, satisfies

$$\int_0^1 \{R_s(F)\}^2 dF \to 0 \quad \text{as } s \to \infty .$$

A corollary of (3.10) is that when the variance of X exists, it satisfies

$$\operatorname{var} X = \sum_{r=2}^{\infty} (2r-1)\lambda_r^2. \tag{3.11}$$

The representation for $x(F)$ given by the inversion theorem seems to be of limited practical utility. The approximation to $x(F)$ using a finite number of L-moments can be poor in the tails of the distribution, particularly if the distribution has a heavy tail; not uncommonly there are some intervals of F in which the approximation to $x(F)$ is not monotonic increasing. Similar problems arise, of course, with the Cornish-Fisher expansion of $x(F)$ in terms of the cumulants of X. Indeed, just as the Cornish-Fisher expansion is most useful for near-Normal distributions, one would expect the approximation of $x(F)$ by L-moments to be most accurate when the distribution of X is close to uniform. The adequacy of the L-moment approximation to $x(F)$ is further considered in Appendix D.

4. Estimation of PWMs and L-moments

Unbiased estimators

We consider the problem of estimating the PWMs α_r and β_r and the L-moments λ_r of a distribution whose mean exists, given an ordered random sample $x_1 \le \cdots \le x_n$, $n > r$, drawn from the distribution. We define

$$a_r \equiv n^{-1} \sum_{i=1}^{n} \binom{n-i}{r} x_i \Big/ \binom{n-1}{r}, \quad r = 0, 1, \ldots, n-1,$$

$$b_r \equiv n^{-1} \sum_{i=1}^{n} \binom{i-1}{r} x_i \Big/ \binom{n-1}{r}, \quad r = 0, 1, \ldots, n-1,$$

where by convention $\binom{k}{j} = 0$ if $k < j$. Then

(4.1) a_r is an unbiased estimator of α_r, and b_r is an unbiased estimator of β_r.

Special cases of these estimators include the sample mean $\bar{x} \equiv n^{-1} \Sigma x_i = a_0 = b_0$ and the extreme data values $x_1 = n a_{n-1}$ and $x_n = n b_{n-1}$. In general a_r and b_r are linear combinations of the x_i with weights which are polynomials of degree r in i, the first (for b_r) or the last (for a_r) r weights being zero. The a_r and b_r are related in the same way as are the α_r and β_r: we have in general

$$a_r = \sum_{k=0}^{r} (-1)^k \binom{r}{k} b_k, \quad b_r = \sum_{k=0}^{r} (-1)^k \binom{r}{k} a_k \tag{4.2}$$

(*cf.* (2.6)) and in particular

$$
\begin{aligned}
a_0 &= b_0, & b_0 &= a_0, \\
a_1 &= b_0 - b_1, & b_1 &= a_0 - a_1, \\
a_2 &= b_0 - 2b_1 + b_2, & b_2 &= a_0 - 2a_1 + a_2, \\
a_3 &= b_0 - 3b_1 + 3b_2 - b_3, & b_3 &= a_0 - 3a_1 + 3a_2 - a_3.
\end{aligned}
$$

Because the L-moments λ_r are linear combinations of the α_r or β_r we can construct estimators of the λ_r which are the corresponding linear combinations of the a_r or b_r. We define ℓ_r, $r = 1, 2, \ldots, n$, by

$$\ell_{r+1} \equiv (-1)^r \sum_{k=0}^{r} p^*_{r,k}\, a_k = \sum_{k=0}^{r} p^*_{r,k}\, b_k \tag{4.3}$$

where $p^*_{r,k}$ is as defined in (3.1); then

(4.4) ℓ_r is an unbiased estimator of λ_r.

The first few ℓ_r are

$$
\begin{aligned}
\ell_1 &= a_0 & &= b_0, \\
\ell_2 &= a_0 - 2a_1 & &= 2b_1 - b_0, \\
\ell_3 &= a_0 - 6a_1 + 6a_2 & &= 6b_2 - 6b_1 + b_0, \\
\ell_4 &= a_0 - 12a_1 + 30a_2 - 20a_3 & &= 20b_3 - 30b_2 + 12b_1 - b_0.
\end{aligned}
$$

An illuminating representation of the estimators a_r, b_r and ℓ_r is as U-statistics, *i.e.* averages over all subsamples of size $m < n$ contained in a sample of size n. Recalling that $r\alpha_{r-1} = EX_{1:r}$, is the expectation of the smallest order statistic from a sample of size r, we may compare this result with the representation of $r\alpha_{r-1}$ as a U-statistic:

$$r a_{r-1} = \binom{n}{r}^{-1} \sum\sum \cdots \sum_{1 \le i_1 < i_2 < \cdots < i_r \le n} x_{i_1} \, . \tag{4.5}$$

the average of the smallest members of all subsamples of size r. Corresponding to $r\beta_{r-1} = EX_{r:r}$ we have

$$r b_{r-1} = \binom{n}{r}^{-1} \sum\sum \cdots \sum_{i_1 < i_2 < \cdots < i_r} x_{i_r} \, . \tag{4.6}$$

the average of the largest members of all size-r subsamples, while

$$\ell_r = \binom{n}{r}^{-1} \sum\sum \cdots \sum_{i_1 < i_2 < \cdots < i_r} r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} x_{i_{r-k}} \tag{4.7}$$

– *cf.* the definition of λ_r. In particular we have

$$\ell_1 = n^{-1} \sum_i x_i \, ,$$

$$\ell_2 = \tfrac{1}{2} \binom{n}{2}^{-1} \sum_{i > j}\sum (x_i - x_j) \, ,$$

$$\ell_3 = \tfrac{1}{3} \binom{n}{3}^{-1} \sum_{i > j > k}\sum\sum (x_i - 2x_j + x_k) \, ,$$

$$\ell_4 = \tfrac{1}{4} \binom{n}{4}^{-1} \sum_{i > j > k > l}\sum\sum\sum (x_i - 3x_j + 3x_k - x_l) \, .$$

U-statistics were introduced by Hoeffding (1948), and are widely used in nonparametric statistics (see, for example, Fraser, 1957, chap. 4, and Randles and Wolfe, 1979, chap. 3). Their properties of high efficiency, asymptotic normality and resistance to the influence of outliers make them particularly attractive for statistical inference. The above representations of a_r, b_r and ℓ_r as U-statistics bring out particularly clearly the relationships between the statistics and the population quantities which they estimate, and suggest that these statistics are the most natural choice as estimators of PWMs and L-moments. We shall, therefore, refer to a_r and b_r as *sample PWMs* and to the ℓ_r as *sample L-moments.*

Sample L-moments may be used similarly to (conventional) sample moments: they summarize the basic properties – location, scale, skewness, kurtosis – of a data set, they estimate the corresponding properties of the probability distribution from which the data were sampled, and they may be used to estimate the parameters of the underlying distribution (parameter estimation is discussed in section 6). In these applications L-moments are frequently preferable to conventional moments: being linear functions of the data, they are less sensitive than are conventional moments to sampling variability or measurement errors in the extreme data values, and may, therefore, be expected to yield more accurate and robust estimates of the characteristics or parameters of an underlying probability distribution.

Note that under a linear transformation of the data, the sample PWMs and sample L-moments are transformed isomorphically with the corresponding population moments: *i.e.*

(4.8) under the linear transformation of the data $x_i \rightarrow Ax_i + B$, $i = 1, \ldots, n$, the sample PWMs and L-moments are transformed as follows:

$$
\begin{aligned}
ra_{r-1} &\rightarrow Ara_{r-1} + B,\\
rb_{r-1} &\rightarrow Arb_{r-1} + B,\\
\ell_1 &\rightarrow A\ell_1 + B,\\
\ell_r &\rightarrow (\operatorname{sign} A)^r A\ell_r, \quad r \geq 2.
\end{aligned}
$$

Sample L-moments have been used previously in statistics, although not as part of a unified theory. The statistic ℓ_1, the sample mean, has a long history and is the most frequently used location estimator in statistical practice. The sample L-scale, ℓ_2, is a scalar multiple of the statistic

$$G = \binom{n}{2}^{-1} \sum_{i>j}\sum (x_i - x_j)$$

which is known as Gini's mean difference and has been used in statistics since at least as far back as von Andrae (1872) and Gini (1912). Downton (1966a) showed that $\frac{1}{2}\pi^{1/2}G$ is a 98% efficient estimator of the scale parameter of a Normal distribution (see Appendix A.6). G is also related to the "total time on test" statistic for testing exponentiality (see section 7). The statistic ℓ_3 has been used as the basis of a test for Normality by Locke and Spurrier (1976). The statistic ℓ_4 does not appear to have been used previously but Locke and Spurrier's reasoning (also that of section 3 above) suggests that ℓ_4 could be used to test a null hypothesis of Normality, or of uniformity, against symmetric alternatives.

Plotting-position estimators

By a *plotting position* we mean a distribution-free estimator of $F(x_{i:n})$, the nonexcession probability of the ith order statistic from a sample of size n. Reasonable choices for plotting positions include $p_{i:n} \equiv (i+\gamma)/(n+\delta)$ for $\delta > \gamma > -1$, and we restrict attention to plotting positions of this form. The plotting position $p_{i:n}$ is *symmetric*, *i.e.* $p_{i:n} = 1 - p_{n+1-i:n}$, if $\delta = 1 + 2\gamma$. The uses of plotting positions in the graphical display of statistical data are discussed by, among others, Cunnane (1978) and Harter (1984).

Graphical methods are not an essential part of PWM theory, but plotting positions may be used to estimate such quantities as

$$\theta = \int_0^1 x(F)\,\eta(F)\,dF$$

where $x(F)$ is the quantile function of a distribution and η is a function of F alone: given a sample $x_1 \leq \cdots \leq x_n$ we might estimate θ by

$$\tilde{\theta}[\gamma, \delta] \equiv n^{-1} \sum_{i=1}^{n} \eta(p_{i:n})\, x_i ,$$

where $p_{i:n}$ is as defined above. This approach yields in particular the estimators

$$\tilde{\alpha}_r[\gamma, \delta] \equiv n^{-1} \sum_{i=1}^{n} (1 - p_{i:n})^r\, x_i ,$$

$$\tilde{\beta}_r[\gamma, \delta] \equiv n^{-1} \sum_{i=1}^{n} p_{i:n}^r\, x_i ,$$

$$\tilde{\lambda}_r[\gamma, \delta] \equiv n^{-1} \sum_{i=1}^{n} P^*_{r-1}(p_{i:n})\, x_i$$

of α_r, β_r and λ_r respectively. We shall sometimes write $\tilde{\alpha}_r$ instead of $\tilde{\alpha}_r[\gamma, \delta]$, *etc.*, when the precise values of γ and δ are unimportant.

We can show that

(4.9) $\tilde{\alpha}_r[\gamma, \delta]$, $\tilde{\beta}_r[\gamma, \delta]$ and $\tilde{\lambda}_r[\gamma, \delta]$ are consistent estimators of α_r, β_r and λ_r respectively,

but these estimators are in general biased for finite samples. The asymptotic bias of $\tilde{\beta}_r$ is given by

$$E(\tilde{\beta}_r - \beta_r) = [\, -\{\tfrac{1}{2} r(r+1) + r\delta\}\beta_r + \{\tfrac{1}{2} r(r+1) + r\gamma\}\beta_{r-1}\,]\, n^{-1} + O(n^{-2}), \quad (4.10)$$

whence we may find asymptotic biases for the other estimators. For example

$$E(\tilde{\lambda}_1 - \lambda_1) = 0,$$

$$E(\tilde{\lambda}_2 - \lambda_2) \sim \{(1 + 2\gamma - \delta)\lambda_1 - (1 + \delta)\lambda_2\}\, n^{-1} .$$

so $\tilde{\lambda}_2[\gamma, 1+2\gamma]$, calculated using the symmetric plotting position $p_{i:n} = (i+\gamma)/(n+1+2\gamma)$, $\gamma > -1$, should have a small negative bias in large samples.

The estimators $\tilde{\alpha}_r$, $\tilde{\beta}_r$ and $\tilde{\lambda}_r$ satisfy the same linear relationships as do the unbiased estimators, *i.e.*

(4.11) equations (4.2) and (4.3) are still valid if a_r, b_r and l_r are replaced by $\tilde{\alpha}_r$, $\tilde{\beta}_r$ and $\tilde{\lambda}_r$ respectively.

If the data are linearly transformed, plotting-position estimators of PWMs are not transformed isomorphically with the PWMs themselves.

(4.12) Let $y_i = Ax_i + B$, and let $\tilde{\alpha}_r^{[Y]}$, $\tilde{\alpha}_r^{[X]}$ be plotting-position estimators calculated from the y_i and from the x_i respectively:

(a) if $B = 0$ then $\tilde{\alpha}_r^{[Y]} = A\tilde{\alpha}_r^{[X]}$ for all $r < n$;

(b) if $B \neq 0$ then $\tilde{\alpha}_r^{[Y]} = A\tilde{\alpha}_r^{[X]} + B$ if and only if either $r = 0$, or $r = 1$ and $\delta = 1 + 2\gamma$.

Similar results hold for $\tilde{\beta}_r$ and $\tilde{\lambda}_r$. Thus inferences based on plotting-position estimators are equivariant under changes of scale of the data but not in general under changes of location. The lack of equivariance is asymptotically negligible – the difference between $\tilde{\alpha}_r^{[Y]}$ and $A\tilde{\alpha}_r^{[X]} + B$ is of order n^{-1}, whereas the difference between $\tilde{\alpha}_r$ and α_r is of stochastic order $n^{-1/2}$ – but could perhaps be important in some practical situations.

There is no profound theoretical reason for preferring plotting-position estimators of PWMs and L-moments to the unbiased estimators described previously, but practical experience shows that plotting-position estimators sometimes yield better estimates of parameters and quantiles when a distribution is fitted to data. In particular the choice $p_{i:n} = (i - 0.35)/n$ gives good results for the generalized Pareto (Hosking and Wallis, 1986), generalized extreme-value (Hosking *et al.*, 1985) and Wakeby (Landwehr *et al.*, 1979b) distributions.

Estimation of L-moment ratios

The L-moment ratios $\tau_r = \lambda_r/\lambda_2$ are naturally estimated by $t_r \equiv \ell_r/\ell_2$, based on the sample L-moments, or by $\tilde{\tau}_r = \tilde{\lambda}_r/\tilde{\lambda}_2$, based on plotting-position estimators. Analogously to our previous terminology we call t_3 the *sample L-skewness* and t_4 the *sample L-kurtosis*.

Although ℓ_r is an unbiased estimator of λ_r, it is of course not true that t_r is an unbiased estimator of τ_r. However, it is true that

(4.13) t_r and $\tilde{\tau}_r$ are consistent estimators of τ_r,

and we can calculate the asymptotic biases of t_r and $\tilde{\tau}_r$ in many cases of interest. For example:

$$\begin{aligned} E(t_3 - \tau_3) &= o(n^{-1}) && \text{for a symmetric distribution } (\tau_3 = 0), \\ &\sim -0.19n^{-1} && \text{for the Gumbel distribution } (\tau_3 = 0.1699), \\ &\sim -0.22n^{-1} && \text{for the exponential distribution } (\tau_3 = 0.3333); \end{aligned} \tag{4.14}$$

$$\begin{aligned} E(t_4 - \tau_4) &\sim 0.09n^{-1} && \text{for the uniform distribution } (\tau_4 = 0), \\ &\sim 0.03n^{-1} && \text{for the Normal distribution } (\tau_4 = 0.1226), \\ &\sim -0.04n^{-1} && \text{for the logistic distribution } (\tau_4 = 0.1667). \end{aligned} \tag{4.15}$$

Clearly the biases of t_3 and t_4 are negligible for sample sizes $n > 20$ in these cases.

The sample L-moment ratios t_3 and t_4 may be used to measure the skewness and kurtosis of an observed data set, as is commonly done with the conventional sample moment ratios g (skewness) and k (kurtosis). A disadvantage of the conventional moment ratios, noted by Kirby (1974), is that when calculated from finite samples they are bounded and cannot attain the full range of values available to the population skewness and kurtosis. For example the skewness g is bounded by

$$|g| \le (n-2)/(n-1)^{1/2}$$

for a sample of size n, and for many moderately to highly skew distributions it is unusual for g to take a value anywhere near the population skewness γ (Wallis *et al.*, 1974). In contrast,

(4.16) the sample L-moment ratios (t_3, t_4) calculated from a sample of size $n \ge 4$ can take any of the feasible values of the population L-moment ratios (τ_3, τ_4) — these values constitute the set $\mathscr{P}$ defined in (3.9).

Indeed it is possible for the sample (t_3, t_4) values to lie outside the set $\mathscr{P}$! For example, the sample

$x_1 = x_2 = \cdots = x_{n-1} = 0,\ x_n = 1$ has $t_3 = -1,\ t_4 = 1$;

$x_1 = 0,\ x_2 = x_3 = \cdots = x_n = 1$ has $t_3 = 1,\ t_4 = 1$;

$x_1 = \cdots = x_m = 0,\ x_{m+1} = \cdots = x_{2m} = 1,\ n = 2m,$ has $t_3 = 0,\ t_4 = -\frac{1}{4}(n+2)/(n-3)$;

the last of these t_4 values is less than $-\frac{1}{4}$ for all n and is as low as $-\frac{3}{2}$ when $n = 4$. In practice, however, (t_3, t_4) values lying outside the set $\mathscr{P}$ are very rarely encountered when sampling from continuous distributions, and in general t_3 and t_4 seem to be reliable estimators of τ_3 and τ_4.

In summary, then, the statistics ℓ_1, ℓ_2, t_3 and t_4 give a useful and informative description of any random sample of statistical data. They summarize a data set in a manner similar to, but in many ways preferable to, that of the conventional sample moments.

5. Sampling distributions of PWMs and L-moments

Introduction

Exact distributions of sample PWMs and L-moments are difficult to obtain. Joint distributions of order statistics, on which the results depend, are algebraically tractable only for samples from the uniform and exponential distributions, and even these do not yield simple results. Exact variances of the sample PWMs a_r and b_r, and of the L-moments ℓ_r, can be obtained in terms of expectations of quadratic functions of order statistics, and can be estimated in a distribution-free manner. However, the most practically useful results come from asymptotic distribution theory. Asymptotic theory for linear combinations of order statistics, developed by Chernoff *et al.* (1967), Moore (1968) and Stigler (1974) among others, can be immediately applied to estimators of PWMs and L-moments. Provided that the underlying distribution has finite variance we can demonstrate the asymptotic Normality, and calculate the asymptotic bias and variance, of the sample statistics a_r, b_r, ℓ_r and t_r and of the corresponding statistics $\tilde{\alpha}_r$, $\tilde{\beta}_r$, $\tilde{\lambda}_r$ and $\tilde{\tau}_r$ based on plotting-position estimators. The asymptotic theory usually provides a good approximation to the exact distribution for samples of size $n \geq 50$, and is often adequate even for $n = 20$.

Exact results

First we consider the exact means and variances of the sample L-moments ℓ_r calculated from a random sample of size n drawn from the distribution of a random variable X with finite variance. Because ℓ_r is unbiased, as shown in section 4, its mean is given by $E\ell_r = \lambda_r$. The variance of ℓ_1, the sample mean, is of course

$$\operatorname{var} \ell_1 = n^{-1} \operatorname{var} X = \tfrac{1}{2} n^{-1} E(X_{2:2} - X_{1:2})^2 .$$

For ℓ_2 we have

$$\operatorname{var} \ell_2 = \{n(n-1)\}^{-1} \left[\tfrac{2}{3}(n-2)E(X_{3:3} - X_{1:3})^2 - \tfrac{1}{2}(n-3)E(X_{2:2} - X_{1:2})^2 - \tfrac{1}{2}(2n-3)\{E(X_{2:2} - X_{1:2})\}^2 \right]. \tag{5.1}$$

This result was originally derived by Nair (1936), who calculated the variance of Gini's mean difference statistic $2\ell_2$. Some special cases are given in Appendix A, sections A.1, A.2 and A.6.

Expressions similar to (5.1) can be obtained for all the a_r, b_r and ℓ_r. By writing these exact variances as sums of expectations of differences of order statistics $E(X_{s:t} - X_{r:t})^2$ and noting that

(5.2) $E(X_{s:t} - X_{r:t})^2$, $1 \le r < s \le t$, is unbiasedly estimated by

$$\binom{n}{t}^{-1} \sum_{1 \le j < k \le n}\sum \binom{j-1}{r-1} \binom{k-j-1}{s-r-1} \binom{n-k}{t-s} (x_k - x_j)^2 ,$$

we can derive unbiased distribution-free estimators of the exact variances of sample PWMs and L-moments. Sillitto (1969) followed this approach. He found that it gave as an estimator of $\operatorname{var} \ell_1$ the familiar statistic $n^{-1}s^2$ where s^2 is the sample variance, and he gave (rather complicated) explicit expressions for estimators of $\operatorname{var} \ell_2$ and $\operatorname{var} \ell_3$. Sillitto suggested that ℓ_3 divided by its estimated standard error could be used as a test for symmetry of the distribution underlying an observed data set. On the whole, however, we would prefer to base such a test on the statistic t_3 (see section 7).

Asymptotic results

The main results are (5.3) - (5.5) below.

(5.3) Let X be a real-valued random variable with cumulative distribution function F, PWMs $\alpha_r = E[X\{1 - F(X)\}^r]$ and $\beta_r = E[X\{F(X)\}^r]$, L-moments λ_r and finite variance. Let a_r, b_r, ℓ_{r+1}, $r = 0, 1, \ldots, m$, be sample PWMs and L-moments calculated from a random sample of size n drawn from the distribution of X. Then as $n \to \infty$:

(a) $n^{1/2}(a_r - \alpha_r)$, $r = 0, 1, \ldots, m-1$, converge in distribution to the multivariate Normal distribution $N(0, A)$, where the elements A_{rs} $(r, s = 0, 1, \ldots, m-1)$ of A are given by

$$A_{rs} = I_{rs} + I_{sr},$$
$$I_{rs} = \iint_{x<y} \{1 - F(x)\}^r \{1 - F(y)\}^s \, F(x)\{1 - F(y)\} \, dx \, dy ;$$

(b) $n^{1/2}(b_r - \beta_r)$, $r = 0, 1, \ldots, m-1$, converge in distribution to the multivariate Normal distribution $N(0, B)$, where the elements B_{rs} $(r, s = 0, 1, \ldots, m-1)$ of B are given by

$$B_{rs} = J_{rs} + J_{sr},$$
$$J_{rs} = \iint_{x<y} \{F(x)\}^r \{F(y)\}^s \, F(x)\{1 - F(y)\} \, dx \, dy ;$$

(c) $n^{1/2}(\ell_r - \lambda_r)$, $r = 1, 2, \ldots, m$, converge in distribution to the multivariate Normal distribution $N(0, \Lambda)$, where the elements Λ_{rs} $(r, s = 1, 2, \ldots, m)$ of Λ are given by

$$\Lambda_{rs} = \iint_{x<y} \{P^*_{r-1}(F(x))P^*_{s-1}(F(y)) + P^*_{s-1}(F(x))P^*_{r-1}(F(y))\} \, F(x)\{1 - F(y)\} \, dx \, dy ,$$

$P^*_r(x)$ being the rth shifted Legendre polynomial as defined in Appendix C.

(5.4) Under the assumptions of (5.3), let $\tau_r = \lambda_r / \lambda_2$ and $t_r = \ell_r / \ell_2$, $r = 3, 4, \ldots, m$. Then as $n \to \infty$ the vector

$$n^{1/2}\left[\,(\ell_1-\lambda_1)\quad(\ell_2-\lambda_2)\quad(t_3-\tau_3)\quad(t_4-\tau_4)\ \ldots\ (t_m-\tau_m)\,\right]^{\mathrm{T}}$$

converges in distribution to the multivariate Normal distribution $N(0,T)$ where the elements T_{rs} $(r,s=1,2,\ldots,m)$ of T are given by

$$T_{rs}=\begin{cases}\Lambda_{rs} & \text{if } r\le 2,\ s\le 2,\\ (\Lambda_{rs}-\tau_r\Lambda_{2s})/\lambda_2 & \text{if } r\ge 3,\ s\le 2,\\ (\Lambda_{rs}-\tau_r\Lambda_{2s}-\tau_s\Lambda_{2r}+\tau_r\tau_s\Lambda_{22})/\lambda_2^2 & \text{if } r\ge 3,\ s\ge 3.\end{cases}$$

(5.5) Results (5.3) and (5.4) remain valid if a_r, b_r, ℓ_r and t_r are replaced by $\tilde{\alpha}_r[\gamma,\delta]$, $\tilde{\beta}_r[\gamma,\delta]$, $\tilde{\lambda}_r[\gamma,\delta]$ and $\tilde{\tau}_r[\gamma,\delta]$ respectively, with $\delta>\gamma>-1$.

If X has a symmetric distribution, some of the foregoing results can be simplified.

(5.6) If the assumptions of (5.3) and (5.4) hold, and if the random variable X has a symmetric distribution, then when $r+s$ is odd we have $\Lambda_{rs}=0$ and $T_{rs}=0$.

Alternative expressions can be found for the asymptotic covariances in (5.3). When the random variable X is continuous, with a differentiable quantile function $x(F)$, we have

$$I_{rs}=\iint_{0<u<v<1}(1-u)^r(1-v)^s\,u(1-v)\,x'(u)\,x'(v)\,du\,dv \tag{5.7}$$

and corresponding expressions for J_{rs} and Λ_{rs}: these are often the most convenient forms for evaluating these quantities for specific distributions. The asymptotic covariances can also be expressed in terms of squared differences of order statistics, using the result (David, 1981, p. 50)

$$\iint_{x<y} \{F(x)\}^r \{1 - F(y)\}^s dx\, dy = \frac{r!s!}{(r+s)!} E(X_{r+1:r+s} - X_{r:r+s})^2 .$$

Expressions such as (5.7) can be used to derive distribution-free estimators of asymptotic covariances of sample L-moments. For example, we can show that

$$\Lambda_{22} = \iint_{0<u<v<1} (1 - 4u)(3 - 4v)\{x(v) - x(u)\}^2 du\, dv . \tag{5.8}$$

which suggests that var ℓ_2 and var $\tilde{\lambda}_2$ could be estimated by

$$n^{-1} \binom{n}{2}^{-1} \sum\sum_{1 \le i < j \le n} (1 - 4p_{i:n})(3 - 4p_{j:n})(x_j - x_i)^2 .$$

Asymptotic distribution theory for linear combinations of order statistics can be extended to higher orders of approximation. Helmers (1980) derives an Edgeworth expansion for the distribution function, and Easton and Ronchetti (1986) obtain a saddle-point approximation to the density function. The procedures described in these papers can be applied immediately to (either unbiased or plotting-position) estimators of PWMs and L-moments, although a lot of algebra is needed to obtain explicit results. Both papers give as an example the third-order approximation to the distribution of an estimator of the mean of the logistic distribution – in our notation this estimator is $\tilde{\lambda}_1[0.1] - \tilde{\lambda}_3[0.1]$.

The asymptotic biases of L-moment statistics are in general of order n^{-1}. This is a result whose formal proof requires second-order asymptotic theory, but asymptotic biases can often be obtained in particular instances by cruder methods: (4.14) and (4.15) are cases in point.

The asymptotic results presented in this section have many practical applications. They make possible the construction of confidence limits for sample L-moments and of critical values for tests based on them. First-order Taylor series expansion, as used in the proof of (5.4), enables the asymptotic theory to be applied to functions of sample L-moments and in particular to PWM-based estimators of parameters and quantiles of probability distributions. In these applications we typically find that first-order asymptotic theory gives a reliable approximation to finite-sample distributions for samples of size $n \geq 50$ (see, for example, Hosking *et al.*, 1985; Hosking and Wallis, 1986).

Of course the true devotee of PWMs would prefer to describe the limiting Normal distributions of sample PWMs and L-moments in terms of their L-moments rather than their means and variances. Since a random variable distributed as $N(\mu, \sigma^2)$ has L-moments $\lambda_1 = \mu$ and $\lambda_2 = \pi^{-1/2}\sigma$ (Appendix A.6) this is easily done. A PWM-based analogue of covariance is, however, not so easy to define.

6. Parameter estimation using PWMs

The method of probability weighted moments

A very common problem in statistics is the estimation, from a random sample of size n, of a probability distribution whose specification involves a finite number, p, of unknown parameters. In its basic form, the *method of probability weighted moments* solves this problem by equating the first p sample PWMs (or, equivalently, the first p L-moments) to the corresponding population quantities. Many examples of parameter estimators derived using this method can be found in Appendix A.

Two refinements of the method of PWMs are sometimes useful. When the distribution being estimated has an endpoint which is a function of the parameters, it is sometimes efficient to estimate the endpoint by one of the extreme values $x_{1:n}$ or $x_{n:n}$ of the sample. Smith (1985) discusses this phenomenon in the context of maximum-likelihood estimation. Using a PWM-based approach we can equate $x_{1:n} = na_{n-1}$ and $x_{n:n} = nb_{n-1}$ to $n\alpha_{n-1}$ and $n\beta_{n-1}$ respectively and use one or both of these equations, together with those obtained from equating low-order sample and population PWMs, to estimate the parameters. Distributions for which this approach is useful include the uniform, exponential and generalized Pareto – see Appendix A, sections A.1, A.2 and A.8. Another refinement is to use more than p PWMs in the estimation procedure: for example, to estimate the mean μ of a symmetric distribution not by ℓ_1 but by the linear combination

$$\hat{\mu} \equiv \ell_1 + A_1\ell_3 + A_2\ell_5 + \cdots + A_m\ell_{2m+1},$$

the A_i being chosen to minimize $\operatorname{var}\hat{\mu}$. This approach car be used for, among others, the logistic and Normal distributions – see Appendix A, sections A.5 and A.6.

When estimating a distribution which has only location and scale parameters, the use of the unbiased estimators a_r (or b_r or ℓ_r) in the method of PWMs yields estimators of parameters and quantiles which are also unbiased. Otherwise there is no general reason to estimate PWMs by any particular unbiased or plotting-position estimator. The asymptotic standard errors of parameter estimators will be the same whether λ_r is estimated by ℓ_r or by $\tilde{\lambda}_r$, so the choice of whether to use ℓ_r or $\tilde{\lambda}_r$ can be made on grounds of computational convenience or the accuracy in finite samples of the ultimate estimators of parameters or quantiles. The use of plotting-position estimators with $p_{i:n} = (i - 0.35)/n$ seems to give particularly good results in many cases, including the Wakeby, generalized extreme-value and generalized Pareto distributions, particularly when they are positively skewed (Landwehr *et al.*, 1979b; Hosking *et al.*, 1985; Hosking and Wallis, 1986).

Exact distributions of parameter estimators obtained by the method of PWMs are in general difficult to derive but asymptotic distributions can usually be found using the results of section 5. For example

(6.1) if the parameters $\theta_1, \ldots, \theta_p$ and L-moments $\lambda_1, \ldots, \lambda_p$ of a distribution with finite variance are related by

$$\theta_r = g_r(\lambda_1, \ldots, \lambda_p), \quad r = 1, \ldots, p,$$

and if the functions g_r are continuous in a neighborhood of the true values of $\lambda_1, \ldots, \lambda_p$, then the method of probability weighted moments yields the estimators

$$\hat{\theta}_r = g_r(\ell_1, \ldots, \ell_p), \quad r = 1, \ldots, p,$$

and $n^{1/2}(\hat{\theta}_r - \theta_r)$, $r = 1, \ldots, p$, converge in distribution as $n \to \infty$ to the multivariate Normal distribution $N(0, \Theta)$, where the elements Θ_{rs} $(r, s = 1, \ldots, p)$ of Θ are given by

$$\Theta_{rs} = \sum_{t=1}^{p} \sum_{u=1}^{p} G_{rt} \Lambda_{tu} G_{su},$$

where $G_{rs} = \partial\theta_r / \partial\lambda_s$.

Similar results hold when the parameters are expressed as functions of the α_r or β_r. For most standard distributions, (6.1) can be used to show that PWM estimators of parameters and quantiles are asymptotically Normally distributed and to find standard errors and confidence intervals. In many cases the resulting algebraic expressions are analytically intractable and numerical calculations must be used.

Comparisons with other estimation methods

The method of maximum likelihood yields estimators which are consistent and asymptotically efficient, and is for this reason favoured by statisticians: our first comparison is therefore between the method of PWMs and the method of maximum likelihood. The method of PWMs is usually computationally more tractable than the method of maximum likelihood, and needs less frequent recourse to iterative procedures. The asymptotic standard errors of PWM estimators, when compared with those of the maximum-likelihood estimators, usually show the method of PWMs to be reasonably efficient. Asymptotic efficiencies of parameters estimated by the method of PWMs are in cases of practical interest typically greater than 90% for location parameters and about 70-80% (but sometimes higher) for scale or shape parameters. In practice, of course, only a finite sample is available, and asymptotic theory is not always a reliable guide to finite-sample performance. Maximum-likelihood estimation often gives worse results than asymptotic theory would suggest and in some cases yields parameter and quantile estimators which are less efficient than those obtained by the method of PWMs. For example, PWM estimators of extreme upper quantiles $x(F)$, $F \geq 0.9$, are preferable to the maximum-likelihood estimators for samples of size up to 100 drawn from the generalized extreme-value or generalized Pareto distributions (Hosking *et al.*, 1985; Hosking and Wallis, 1986).

The method of PWMs clearly parallels the method of (conventional) moments, so a comparison of the two methods is of interest. Neither method has a strong theoretical advantage over the other. As remarked in section 4, however, because sample PWMs are linear functions of the data while conventional moments involve higher powers of the data, we might expect inferences based on PWMs to be more robust and less affected by sampling variability or the presence of outliers than inferences based on conventional moments. Sample PWMs, L-moments and L-moment ratios are much less biased than are conventional sample moments and cumulants, so equating sample moments and population moments yields more secure inferences for PWMs than for conventional moments. Comparing the practicality of the two methods for estimating specific distributions, we find some distributions (*e.g.* gamma) where moment estimators can be explicitly written down while PWM estimators cannot, and some (*e.g.* generalized logistic) where the reverse is true. It should be noted, however, that for several distributions for which explicit expressions for the parameters in terms of the PWMs cannot be found, either simple approximations exist which are adequate for all practical purposes (*e.g.* generalized extreme-value, generalized Normal), or PWM parameter estimators can be expressed as solutions of simple equations (*e.g.* Wakeby, symmetric lambda, exponential mixture). The two estimation methods can also be compared in terms of the efficiency, asymptotic or finite-sample, of the estimators. When this comparison is made the advantage generally seems to lie with the method of PWMs, especially when the distribution being estimated has three or more parameters.

The method of PWMs may also be compared with other methods which use linear combinations of order statistics to estimate the parameters of probability distributions. These come in many varieties: best linear estimates, best linear unbiased estimates, nearly best nearly unbiased linear estimates, to name but a few. See David (1981, chapter 6) for further definitions and discussion. One of these varieties of estimates, Downton's (1966a) "linear estimates with polynomial coefficients", is a special case of the method of PWMs. In general, one can say that of these linear estimation methods, the method of PWMs yields estimators which are easy to compute and have tractable asymptotic distributions and reasonably high to very high efficiency.

7. Hypothesis testing using PWMs

Introduction

The use of PWMs and L-moments for describing the main characteristics and estimating the parameters of probability distributions extends naturally to testing hypotheses about distributional form. Several categories of testing can be distinguished, according to the degree of precision with which the hypotheses are formulated.

First we can consider very general hypotheses such as "the distribution has zero mean" or "the distribution is symmetric". Tests of these hypotheses may naturally be based on the location and skewness measures ℓ_1 and ℓ_3.

Tests of specific distributional form, or goodness-of-fit tests, are more precisely formulated: the null hypothesis is that the observed sample is derived from a probability distribution whose specification involves at most a finite number, p say, of unknown parameters. The alternative hypothesis is not defined but is presumed to be of a vey general and vague nature. When inference is based on PWMs the natural approach is to estimate the p unknown parameters of the distribution by the method of PWMs, using the sample L-moments $\ell_1, \ldots, \ell_p$, and to test the goodness of fit of the distribution by comparing the next few sample L-moments $\ell_{p+1}, \ldots, \ell_{p+r}$ (or $t_{p+1}, \ldots, t_{p+r}$) to their expected values under the null hypothesis.

The chi-squared test of Pearson (1900) is a well-known test which can be used to test the goodness of fit of any probability distribution. A test of similar scope can be devised which is particularly suitable for use with PWMs. It is based on Neyman's (1937) "smooth test", which has some theoretical and practical advantages over the chi-squared test.

Finally we can consider tests in which both the null and the alternative hypotheses are specified apart from a finite number of parameters, and the null hypothesis is that the parameters satisfy certain restrictions. Tests of these parametric hypotheses can be based on estimated parameters obtained by the method of PWMs. The asymptotic distribution theory of section 5 enables us to derive critical values for the test statistics.

General tests

First we consider the hypothesis "the distribution has zero mean". The familiar Student's t test may be regarded as being based on PWMs, for the test statistic $t = n^{1/2}\bar{x}/s$ is equal to the sample L-moment ℓ_1 divided by an estimate of the standard error of ℓ_1. When the underlying distribution is known to be close to Normal, the estimator s of the standard deviation could be replaced by the PWM-based estimator derived for the Normal distribution, *viz.* $\pi^{1/2}\ell_2$ (Barnett *et al.*, 1967).

A test for zero skewness analogous to the t test could be based on the statistic ℓ_3 divided by a distribution-free estimate of its standard error. The variance of ℓ_3 is given by an expression similar to (5.1), and (5.2) can be used to obtain an estimator of this variance and thence of the standard error of ℓ_3. The variance estimator is given explicitly by Sillitto (1969, equation (35)). Alternatively, the asymptotic variance

$$\lim_{n\to\infty} n \operatorname{var} \ell_3 = \Lambda_{33}$$
$$= \tfrac{1}{2} \iint_{u<v} (1 - 12u + 18u^2)(7 - 24v + 18v^2)\{x(v) - x(u)\}^2 \, du \, dv \,. \qquad (7.1)$$

can be estimated by the plotting-position estimator

$$\tfrac{1}{2} \binom{n}{2}^{-1} \sum\sum_{i<j} (1 - 12p_{i:n} + 18p_{i:n}^2)(7 - 24p_{j:n} + 18p_{j:n}^2)(x_j - x_i)^2 \,.$$

Although no specific distributional form has been assumed in the construction of the statistic, one would expect the exact distribution of the statistic and the critical values of the resulting test to depend on the distribution from which the sample is drawn (as is the case for Student's t test). One might reasonably hope, however, that the statistic would have an asymptotic standard Normal distribution which would provide an adequate approximation to the exact distribution in moderate or large samples.

Goodness-of-fit tests for specific distributions

Tests of specific distributional form may be based on the characterization of distributions by their PWMs or L-moments. The uniform distribution, characterized by its L-moment ratios τ_r, $r \geq 3$, being zero, provides the canonical example of this approach. Using the asymptotic Normal distribution of sample L-moment ratios (see (5.4) above and Appendix A.1) we can construct the test statistics $U_3 \equiv (35n/6)^{1/2} t_3$, which has an asymptotic $N(0,1)$ distribution if the null hypothesis of uniformity is true, and $U_4 \equiv 35(t_3^2/6 + t_4^2/2)$, whose asymptotic distribution under the uniformity hypothesis is χ_2^2, or other statistics involving higher-order L-moments. If the range of the data is known *a priori*, as the interval (0,1) say, we may also take into account the deviations of the L-moments ℓ_1 and ℓ_2 from their respective expectations $\frac{1}{2}$ and $\frac{1}{6}$.

Tests of distributional form for other distributions may be similarly derived. For example, the statistics $2.315n^{1/2}t_3$ and $3.375n^{1/2}t_4$ test the null hypothesis of Normality against asymmetric and symmetric alternatives respectively; both statistics have asymptotic $N(0,1)$ distributions when the null hypothesis is true. The test based on t_3 is similar in spirit to a test of Normality proposed by Locke and Spurrier (1976), which can be regarded as being based the statistic ℓ_3/s where s is the sample standard deviation. However, when working with L-moments it seems more natural to measure the scale of the distribution by ℓ_2 rather than s. The statistic t_3 has an asymptotic variance which is fairly insensitive to the exact specification of the underlying distribution provided that this is symmetric and not too heavy-tailed (*e.g.* $n \operatorname{var} t_3 \sim 0.1714$ for the uniform distribution, 0.1866 for the Normal), and, therefore, yields a fairly robust test of the general hypothesis that the underlying distribution is symmetric. The sample L-CV, ℓ_2/ℓ_1, can be used for testing a hypothesis of exponentiality. It is a linear transformation of the "total time on test" statistic, used for example by Barlow *et al.* (1972, section 6.2), which has been shown by Gail and Gastwirth (1978) to be a particularly accurate and powerful test of goodness of fit for the exponential distribution.

A general goodness-of-fit test

An important special case of testing for uniformity is when a continuous distribution $F(x;\theta)$ dependent upon a vector parameter θ has been fitted to a data set $x_1 \le \cdots \le x_n$, yielding an estimate $\hat{\theta}$ of θ, and we calculate the statistics $\hat{u}_i \equiv F(x_i;\hat{\theta})$. Since the quantities $u_i \equiv F(x_i;\theta)$ constitute an ordered random sample from a uniform distribution if the data were indeed drawn from the distribution F, a test of the uniformity of the $\hat{u}_i$ yields a test of goodness of fit valid for any continuous distribution F, provided that the effect of estimation of θ can be allowed for.

Neyman (1937) based his "smooth test" of goodness of fit on the conventional moments of the u_i, with θ assumed known. The test statistic is constructed using the asymptotic Normal distribution of the sample moments, and is a quadratic form in the first m moments with an asymptotic χ^2_m distribution if the true distribution is indeed F. Neyman suggests using $m = 3$ or $m = 4$. An equivalent test could instead be based on the PWMs of the u_i, because

(7.2) if $u_1 \le \cdots \le u_n$ is an ordered random sample from the uniform distribution on $[0, 1]$, then the rth conventional noncentral moment $m_r \equiv n^{-1}\Sigma u_i^r$ and the scaled rth probability weighted moment rb_{r-1} (or $r\tilde{\beta}_{r-1}$) calculated from the u_i are, apart from a constant independent of n, asymptotically equivalent as $n \to \infty$.

Thomas and Pierce (1979) have extended Neyman's test to the case in which θ is estimated by the method of maximum likelihood. They also give some theoretical reasons why the smooth test should be preferable to alternative testing procedures based on variants of Pearson's chi-squared statistic, and provide some simulation results which support their theoretical arguments. Their approach can also be used when θ is estimated by the method of PWMs, but the test procedure is rather ungainly – after calculating $\hat{\theta}$ the method in effect uses one step of the likelihood-based method of scoring to obtain an estimator $\bar{\theta}$ which is asymptotically equivalent to the maximum-likelihood estimator, and the test is then based on $\bar{u}_i \equiv F(x_i; \bar{\theta})$.

We can instead construct a test which makes direct use of the $\hat{u}_i$ calculated from the estimator $\hat{\theta}$ obtained by the method of PWMs.

(7.3) Let X be a random variable with distribution function $F(x; \theta)$ and quantile function $G(u; \theta)$, where $\theta \equiv (\theta_1 \ldots \theta_p)^T$ is a vector of real parameters with true (but unknown) value θ_0. Let $\kappa_r(u)$, $r = 1, \ldots, p$, be a set of linearly independent polynomials. Let

$$\gamma_r(\theta) \equiv \int_0^1 \kappa_r(u)\, G(u; \theta)\, d\theta, \qquad r = 1, \ldots, p,$$

be probability weighted moments (or linear combinations thereof) and let their estimators be

$$g_r \equiv n^{-1} \sum_{j=1}^{n} \kappa_{rj}^{(n)} X_{j:n}, \qquad r = 1, \ldots, p,$$

where

$$\kappa_{rj}^{(n)} = \kappa_r\left(\frac{j}{n+1}\right) + O(n^{-1})$$

and $X_{j:n}$, $j = 1, \ldots, n$, are the order statistics of a random sample of size n from the distribution F. Let $\hat{\theta}$ be the method-of-PWMs estimator of θ, defined by $g_r = \gamma_r(\hat{\theta})$, $r = 1, \ldots, p$. Let $\delta:[0, 1] \to \mathbb{R}^m$ be a continuous bounded function and let

$$d = n^{-1} \sum_{j=1}^{n} \delta_j^{(n)} \hat{U}_{j:n}$$

where

$$\delta_j^{(n)} = \delta\left(\frac{j}{n+1}\right) + O(n^{-1})$$

and $\hat{U}_{j:n} \equiv F(X_{j:n}; \hat{\theta})$. Then under suitable regularity conditions

$$n^{1/2}(d - \mu) \xrightarrow{\mathcal{D}} N(0, \Sigma) \quad \text{as } n \to \infty,$$

where

$$\mu = \int_0^1 \delta(u)\, u\, du$$

and

$$\Sigma = 2\iint_{u<v} \eta(u)\{\eta(v)\}^{\mathrm{T}} u(1 - v)\, du\, dv$$

where

$$\eta(u) = \delta(u) - \sum_{r=1}^{p} \sum_{s=1}^{p} \left\{\int_0^1 \delta(u) \frac{\partial G(u; \theta_0)}{\partial \theta_r} \frac{du}{G'(u; \theta_0)}\right\} H^{rs}(\theta_0)\, \kappa_s(u)\, G'(u; \theta_0) ;$$

here $H^{rs}(\theta)$ is the (r, s) element of the $m \times m$ matrix $H(\theta)$ whose (r, s) element is $H_{rs}(\theta) \equiv \partial\gamma_r(\theta)/\partial\theta_s$.

(7.4) In the notation of (7.3), a goodness of fit test for the distribution F may be based on the statistic

$$Z \equiv n(d - \mu)^{\mathrm{T}} \hat{\Sigma}^{-1} (d - \mu)$$

where $\hat{\Sigma}$ is a consistent estimator of Σ. If F is the true distribution from which the sample was drawn then Z has asymptotically as $n \to \infty$ a χ^2 distribution with $\text{rank}(\Sigma)$ degrees of freedom.

The test statistic in (7.4) does not depend on the particular choice of $\kappa_r(u)$, provided that none of the κ_r is of degree greater than p. In practice one would usually choose $\kappa_r(u)$ equal to $(1-u)^r$ or u^r or $P^*_{r-1}(u)$, making $\gamma_r(\theta)$ equal to α_r or β_r or λ_r respectively, depending on whichever variety of PWMs could be expressed most conveniently in terms of the parameters. Similarly $\delta(u)$ can be chosen to make d a vector of α's, β's or λ's of the $\hat{u}_i$, whichever is most convenient. The rank of Σ in (7.4) will generally be m. Exceptional cases can occur in which linear dependencies between elements of d arise from the equations defining $\hat{\theta}$ – for example if the location parameter of the logistic distribution is estimated by $\ell_1 - \ell_3$, then $\Sigma \hat{u}_i = 0$ – but such cases are rare.

This PWM-based variant of Neyman's smooth test cannot yet be recommended for general use. It is inconvenient to use, because the calculation of the elemnts of Σ usually requires numerical iteration. Furthermore, some small-scale simulations from the Normal and Gumbel distributions indicate that fairly large sample sizes, 100 or more, are needed before the asymptotic significance levels are attained in finite samples. The main problem seems to be slow convergence of the mean of $d - \mu$ to zero, so a bias correction for the elements of $d - \mu$ should lead to an improved test.

Tests of parametric hypotheses

Tests concerning the parameters of specific distributions can be based on the estimators of those parameters obtained by the method of PWMs. For example, a test of the hypotheses $H_0 : \theta = \theta_0$ *vs.* $H_1 : \theta \neq \theta_0$ concerning a scalar parameter θ may be based on the statistic $(\hat{\theta} - \theta_0) / \hat{\sigma}(\hat{\theta})$ where $\hat{\theta}$ is the PWM estimator of θ and $\hat{\sigma}(\hat{\theta})$ is a consistent estimator of the standard error of $\hat{\theta}$. Under very general conditions the test statistic has an asymptotic

$N(0,1)$ distribution on H_0. This method of constructing tests parallels Wald's (1943) method for deriving test statistics from maximum-likelihood estimators. To some extent, of course, this approach overlaps with derivation of tests of specific distributional form considered above. For example, testing whether the shape parameter k of the generalized extreme-value distribution is zero may be regarded as testing whether the distribution is Gumbel, the Gumbel distribution being the special case $k = 0$ of the GEV distribution. Hosking *et al.* (1985) derive a test of this hypothesis based on the PWM estimator of k, which is a function of the plotting-position L-moment estimator $\tilde{\lambda}_3[-0.35, 0]$. Monte Carlo simulations found the performance in small samples of the PWM-based test to be almost equal to that of the modified likelihood-ratio test recommended by Hosking (1984); the PWM-based test involves much less computation.

APPENDIX A. RESULTS FOR SPECIFIC DISTRIBUTIONS

Introduction

Sections A.1 *et seq.* contain results for the following distributions.

2-parameter (location-scale) distributions

1. Uniform
2. Exponential
3. Gumbel (extreme-value type I)
4. Laplace (double exponential)
5. Logistic
6. Normal
7. Rayleigh

3-parameter (location-scale-shape) distributions

8. Generalized Pareto
9. Generalized extreme-value (GEV; includes Weibull)
10. Generalized logistic
11. Generalized Normal (lognormal)
12. Gamma (Pearson type III)

Distributions with more than 3 parameters

13. Generalized Lambda
14. Wakeby

Other distributions

15. Bernoulli
16. Exponential mixture

For each distribution we give expressions, in terms of the parameters of the distribution, for as many as we have derived of the quantities defined in the "Notation" section (page 151) — *viz.* the α_r, β_r, λ_r, τ_r, A_{rs}, B_{rs}, Λ_{rs} and T_{rs}, and the matrices A, B, Λ and T. Where possible we also give expressions for the parameters in terms of the PWMs (α_r, β_r, λ_r or τ_r, whichever is most convenient) — these define, by substituting the estimators a_r or $\tilde{\alpha}_r$ for α_r, *etc.*, the PWM estimators of the parameters of the distribution. We also give, if they can be simply expressed, the asymptotic variances of the PWM estimators of parameters and quantiles and their asymptotic efficiencies relative to the maximum-likelihood estimators. When the efficiency of a vector of parameters is given, it is defined to be the ratio of the covariance determinants of the maximum-likelihood and PWM estimators.

Proofs of numbered results are given in Appendix B. Other results follow directly from one or other of the definitions in the "Notation" section.

A.1. Uniform

Definition

$$f(x) = 1/(\beta - \alpha), \quad \alpha \le x \le \beta$$

$$F(x) = (x - \alpha)/(\beta - \alpha)$$

$$x(F) = \alpha + (\beta - \alpha)\, F$$

We begin with the uniform distribution because all its L-moment ratios are zero: *i.e.*, as stated below, $\tau_r = 0$ for all $r \ge 3$. The distribution, therefore, plays a central role in PWM theory, rather as the Normal distribution does in cumulant theory.

PWMs, etc.

$$r\alpha_{r-1} = \frac{r\alpha + \beta}{r + 1}$$

$$r\beta_{r-1} = \frac{\alpha + r\beta}{r + 1}$$

$$\lambda_1 = \frac{1}{2}(\alpha + \beta)$$

$$\lambda_2 = \frac{1}{6}(\beta - \alpha)$$

$$\lambda_3 = 0$$

$$\lambda_4 = 0$$

$$\lambda_r = 0 \quad \text{for all } r \ge 3$$

$$\tau_r = 0 \quad \text{for all } r \ge 3 \qquad \text{(A.1.1)}$$

Asymptotic sampling variances

$$A_{rs} = B_{rs} = \frac{(\beta - \alpha)^2}{(r + 2)(s + 2)(r + s + 3)} \qquad \text{(A.1.2)}$$

$$\Lambda_{rr} = \frac{(\beta - \alpha)^2}{2(2r-3)(2r-1)(2r+1)} - \lambda_r\lambda_s$$

$$\Lambda_{r,r\pm 2} = \frac{-(\beta - \alpha)^2}{4(2r-1)(2r+1)(2r+3)} \tag{A.1.3}$$

$$\Lambda_{rs} = 0 \quad \text{otherwise}$$

$$A = B = (\beta - \alpha)^2 \begin{bmatrix} 1/12 & & & \\ 1/24 & 1/45 & & \\ 1/40 & 1/72 & 1/112 & \\ 1/60 & 1/105 & 1/160 & 1/225 \end{bmatrix}$$

$$\Lambda = (\beta - \alpha)^2 \begin{bmatrix} 1/12 & & & \\ 0 & 1/180 & & \\ -1/60 & 0 & 1/210 & \\ 0 & -1/420 & 0 & 1/630 \end{bmatrix}$$

$$T = \begin{bmatrix} (\beta - \alpha)^2/12 & & & \\ 0 & (\beta - \alpha)^2/180 & & \\ -(\beta - \alpha)/10 & 0 & 6/35 & \\ 0 & -3(\beta - \alpha)/35 & 0 & 2/35 \end{bmatrix}$$

The exact variance of ℓ_2 is (Nair, 1936)

$$\text{var}\,\ell_2 = \frac{(\beta - \alpha)^2(n+3)}{180n(n-1)} .$$

Parameter estimation

PWM estimators based on λ_1 and λ_2 are inferior to the maximum-likelihood estimators $\hat{\alpha}_{ML} = x_{1:n}$, $\hat{\beta}_{ML} = x_{n:n}$; indeed the PWM estimators have asymptotic efficiency zero. However the bias-corrected maximum-likelihood estimators

$$\hat{\alpha} = x_{1:n} - \frac{x_{n:n} - x_{1:n}}{n-1}, \quad \hat{\beta} = x_{n:n} + \frac{x_{n:n} - x_{1:n}}{n-1} .$$

which are also the "maximum product of spacings" estimators of Cheng and Amin (1983), may be regarded as PWM estimators derived from the PWMs

$$n\alpha_{n-1} = (n\alpha + \beta)/(n+1), \quad n\beta_{n-1} = (\alpha + n\beta)/(n+1),$$

estimated respectively by $na_{n-1} = x_{1:n}$, $nb_{n-1} = x_{n:n}$.

A.2. Exponential

Definition

$$f(x) = \alpha^{-1} e^{-(x-\xi)/\alpha}, \quad \xi \le x < \infty$$

$$F(x) = 1 - e^{-(x-\xi)/\alpha}$$

$$x(F) = \xi - \alpha \log(1 - F)$$

PWMs, etc.

$$r\alpha_{r-1} = \xi + \alpha/r$$

$$r\beta_{r-1} = \xi + \alpha\{\psi(r+1) - \psi(1)\}$$

$$= \xi + \alpha\left(1 + \tfrac{1}{2} + \cdots + \frac{1}{r}\right) \qquad \text{(A.2.1)}$$

$$\lambda_1 = \xi + \alpha$$

$$\lambda_2 = \alpha/2$$

$$\lambda_3 = \alpha/6$$

$$\lambda_4 = \alpha/12$$

$$\lambda_r = \alpha/\{r(r-1)\} \qquad \text{(A.2.2)}$$

$$\tau_3 = \tfrac{1}{3}$$

$$\tau_4 = \tfrac{1}{6}$$

$$\tau_r = 2/\{r(r-1)\}$$

Asymptotic sampling variances

$$A_{r,s} = \frac{\alpha^2}{(r+1)(s+1)(r+s+2)} \qquad \text{(A.2.3)}$$

$$A = \alpha^2 \begin{bmatrix} 1/2 & & & \\ 1/6 & 1/12 & & \\ 1/12 & 1/24 & 1/45 & \\ 1/20 & 1/40 & 1/72 & 1/112 \end{bmatrix}$$

$$\Lambda = \alpha^2 \begin{bmatrix} 1 & & & \\ 1/2 & 1/3 & & \\ 1/6 & 1/6 & 1/3 & \\ 1/12 & 1/12 & 3/4 & 19/6 \end{bmatrix}$$

$$T = \begin{bmatrix} \alpha^2 & & & \\ \alpha^2/2 & \alpha^2/3 & & \\ 0 & \alpha/9 & 28/27 & \\ 0 & \alpha/18 & 77/27 & 340/27 \end{bmatrix}$$

Note that ℓ_1 and t_r, $r \geq 3$ are independent, by Basu's theorem (Basu, 1955). The exact variance of ℓ_2 is (Nair, 1936)

$$\operatorname{var} \ell_2 = \frac{\alpha^2(2n-1)}{6n(n-1)}.$$

Parameter estimation

When ξ is known we have $\alpha = \lambda_1 - \xi$ and the PWM and maximum-likelihood estimators are identical. When ξ is unknown the situation is similar to that arising with the uniform distribution: PWM estimators based on λ_1 and λ_2 are inefficient, but estimators based on λ_1 (estimated by the sample mean $\bar{x}$) and $n\alpha_{n-1}$ (estimated by $x_{1:n}$) yield the bias-corrected maximum-likelihood estimators

$$\hat{\xi} = (nx_{1:n} - \bar{x})/(n-1), \quad \hat{\alpha} = n(\bar{x} - x_{1:n})/(n-1).$$

A.3. Gumbel

Definition

$$f(x) = \alpha^{-1} \exp\{-(x-\xi)/\alpha\} \exp[-\exp\{-(x-\xi)/\alpha\}], \qquad -\infty < x < \infty$$
$$F(x) = \exp[-\exp\{-(x-\xi)/\alpha\}]$$
$$x(F) = \xi - \alpha \log(-\log F)$$

PWMs, etc.

$$r\beta_{r-1} = \xi + \alpha(\gamma + \log r) \qquad \text{(A.3.1)}$$

where $\gamma = 0.5772...$ is Euler's constant.

$$\lambda_1 = \xi + \alpha\gamma$$
$$\lambda_2 = \alpha \log 2$$
$$\lambda_3 = \alpha(2 \log 3 - 3 \log 2)$$
$$\lambda_4 = \alpha(5 \log 4 - 10 \log 3 + 6 \log 2)$$

$$\tau_3 = \log(9/8)/\log 2 \qquad = 0.1699$$
$$\tau_4 = (16 \log 2 - 10 \log 3)/\log 2 \qquad = 0.1504$$

Asymptotic sampling variances

$$B_{r,r} = \alpha^2 \left\{ \frac{\pi^2}{6} - 2g\left(\frac{r}{r+1}\right) \right\}$$

$$B_{r,r+1} = \frac{\alpha^2}{2} \left[\frac{\pi^2}{6} - 2g\left(\frac{r}{r+2}\right) + \left\{ \log\left(\frac{r+2}{r+1}\right) \right\}^2 \right]$$

$$B_{r,r+s} = \frac{\alpha^2}{2} \left[2g\left(\frac{r+1}{r+2}\right) - 2g\left(\frac{r}{r+s+1}\right) + \left\{ \log\left(\frac{r+s+1}{r+1}\right) \right\}^2 - \left\{ \log\left(\frac{r+s}{r+1}\right) \right\}^2 \right] \qquad \text{(A.3.2)}$$

where

$$g(x) = x\,{}_3F_2(1, 1, 1; 2, 2; -x) = x - \frac{x^2}{2^2} + \frac{x^3}{3^2} - \frac{x^4}{4^2} + \cdots$$

$$B = \alpha^2 \begin{bmatrix} 1.6449 & & & \\ 1.0627 & 0.7481 & & \\ 0.8117 & 0.5956 & 0.4854 & \\ 0.6665 & 0.5019 & 0.4154 & 0.3594 \end{bmatrix}$$

$$\Lambda = \alpha^2 \begin{bmatrix} 1.6449 & & & \\ 0.4805 & 0.3866 & & \\ 0.1387 & 0.1570 & 0.1540 & \\ 0.0868 & 0.0794 & 0.0792 & 0.0813 \end{bmatrix}$$

$$T = \begin{bmatrix} 1.6449\alpha^2 & & & \\ 0.4805\alpha^2 & 0.3866\alpha^2 & & \\ 0.0824\alpha & 0.1317\alpha & 0.2327 & \\ 0.0210\alpha & 0.0307\alpha & 0.1082 & 0.1376 \end{bmatrix}$$

Parameter estimation

$$\alpha = \lambda_2 / \log 2, \quad \xi = \lambda_1 - \gamma\alpha$$

These estimators were derived independently by Downton (1966b) and Greenwood *et al.* (1979). Efficiencies with respect to the maximum-likelihood estimators were calculated asymptotically by Downton and evaluated in finite samples using Monte Carlo simulation by Landwehr *et al.* (1979a). Asymptotic results:

$$n \operatorname{var} \begin{bmatrix} \hat{\xi} \\ \hat{\alpha} \end{bmatrix} \sim \alpha^2 \begin{bmatrix} 1.1128 & 0.2287 \\ 0.2287 & 0.8046 \end{bmatrix}.$$

cf.

$$n \operatorname{var} \begin{bmatrix} \hat{\xi}_{ML} \\ \hat{\alpha}_{ML} \end{bmatrix} \sim \alpha^2 \begin{bmatrix} 1.1087 & 0.2570 \\ 0.2570 & 0.6079 \end{bmatrix};$$

thus

$$\operatorname{eff}(\hat{\xi}) \sim 0.9963, \quad \operatorname{eff}(\hat{\alpha}) \sim 0.7555, \quad \operatorname{eff} \begin{bmatrix} \hat{\xi} \\ \hat{\alpha} \end{bmatrix} \sim 0.6832.$$

For comparison, estimators based on the conventional moments have asymptotic efficiencies 0.95 for the estimator of ξ and 0.55 for the estimator of α (Tiago de Oliveira, 1963), and are thus less efficient than the PWM estimators. Landwehr *et al.*'s (1979a) simulation results indicate that PWM estimators based on ℓ_r also outperform the conventional moment estimators in finite samples; the method of PWMs is less efficient than the method of maximum likelihood, but is also less biased (being, indeed, unbiased for all sample sizes) and requires much less computation. Downton (1966b) also considers estimators based on λ_1, λ_2 and λ_3: these estimators are more efficient than those given above but have not so simple a form.

A.4. Laplace

Definition

$$f(x) = \tfrac{1}{2}\alpha^{-1} e^{-|x-\xi|/\alpha}, \quad -\infty < x < \infty$$

$$F(x) = \begin{cases} \tfrac{1}{2} e^{(x-\xi)/\alpha}, & x \le \xi \\ 1 - \tfrac{1}{2} e^{-(x-\xi)/\alpha}, & x \ge \xi \end{cases}$$

$$x(F) = \begin{cases} \xi + \alpha \log(2F), & F \le \tfrac{1}{2} \\ \xi - \alpha \log\{2(1-F)\}, & F \ge \tfrac{1}{2} \end{cases}$$

PWMs, etc.

$$\lambda_1 = \xi$$
$$\lambda_2 = 3\alpha/4$$
$$\lambda_3 = 0$$
$$\lambda_4 = 17\alpha/96$$

$$\lambda_r = \begin{cases} 0, & r \text{ odd}, \ge 3, \\ \dfrac{2\alpha}{r(r-1)}\{1 - {}_2F_1(-r, r-1; 1; \tfrac{1}{2})\}, & r \text{ even}, \ge 4. \end{cases} \tag{A.4.1}$$

$$\tau_3 = 0$$
$$\tau_4 = 17/72 = 0.2361$$

$$\tau_r = \begin{cases} 0, & r \text{ odd}, \ge 3, \\ \dfrac{8\alpha}{3r(r-1)}\{1 - {}_2F_1(-r, r-1; 1; \tfrac{1}{2})\}, & r \text{ even}, \ge 4. \end{cases}$$

Asymptotic sampling variances

$$A_{rs} = B_{rs} = J_{rs} + J_{sr}$$

where

$$\begin{aligned} J_{rs} = {} & \frac{2^{-(r+s+2)}(r+s+3)}{(r+1)(r+s+1)(r+s+2)} + \frac{2^{-(r+1)}(1-2^{-(s+1)})}{(r+1)(s+1)} \\ & + \frac{1}{s+1}\left\{\frac{1-2^{-(r+2)}}{r+2} + \frac{1-2^{-(r+3)}}{r+3} + \cdots + \frac{1-2^{-(r+s+2)}}{r+s+2}\right\} \end{aligned} \tag{A.4.2}$$

$$A = B = \alpha^2 \begin{bmatrix} 2 & & & \\ 1 & 0.6458 & & \\ 0.7361 & 0.5139 & 0.4215 & 0 \\ 0.6042 & 0.4378 & 0.3652 & 0.3197 \end{bmatrix}$$

$$\Lambda = \alpha^2 \begin{bmatrix} 2 & & & \\ 0 & 7/12 & & \\ 5/12 & 0 & 31/120 & \\ 0 & 17/96 & & 187/1344 \end{bmatrix}$$

$$= \alpha^2 \begin{bmatrix} 2 & & & \\ 0 & 0.5833 & & \\ 0.4167 & 0 & 0.2583 & 0 \\ 0 & 0.1771 & 0 & 0.1390 \end{bmatrix}$$

$$T = \begin{bmatrix} 2\alpha^2 & & & \\ 0 & 0.5833\alpha^2 & & \\ 0.5556\alpha & 0 & 0.4593 & \\ 0 & 0.0525\alpha & 0 & 0.1565 \end{bmatrix}$$

Parameter estimation

$$\xi = \lambda_1, \quad \alpha = \tfrac{4}{3}\lambda_2$$

Thus the PWM estimator of ξ is the sample mean ℓ_1. This is an inefficient estimator: the maximum-likelihood estimator of ξ is the sample median. We can obtained improved PWM estimators by estimating the successive median-like quantities $EX_{r:2r+1}$, $r = 1, 2, \ldots$: for example

$EX_{1:1}$ is estimated by ℓ_1 with asymptotic variance $2\alpha^2 n^{-1}$;

$EX_{2:3}$ is estimated by $\ell_1 - \ell_3$ with asymptotic variance $1.4250\alpha^2 n^{-1}$;

$EX_{3:5}$ is estimated by $\ell_1 - \frac{10}{7}\ell_3 + \frac{3}{7}\ell_5$ with asymptotic variance $1.2974\alpha^2 n^{-1}$;

$EX_{4:7}$ is estimated by $\ell_1 - \frac{5}{3}\ell_3 + \frac{9}{11}\ell_5 - \frac{5}{33}\ell_7$ with asymptotic variance $1.2380\alpha^2 n^{-1}$.

Alternatively we can construct minimum-variance unbiased linear combinations of the ℓ_{2r+1} for estimating ξ: for example the best combination of ℓ_1 and ℓ_3 is

$$\ell_1 - \frac{50}{31}\ell_3,$$

with asymptotic variance $1.3280\alpha^2 n^{-1}$.

The estimator of α can be similarly refined: we have

$$\operatorname{var}\left(\tfrac{4}{3}\ell_2\right) \sim \frac{28}{27}\alpha^2 n^{-1} \sim 1.0370\alpha^2 n^{-1},$$

whereas the best linear combination of ℓ_2 and ℓ_4 has

$$\operatorname{var}(1.4741\ell_2 - 0.5960\ell_4) \sim 1.0058\alpha^2 n^{-1},$$

this variance being within 1% of $\alpha^2 n^{-1}$, the asymptotic variance of the maximum-likelihood estimator of α.

A.5. Logistic

Definition

$$f(x) = e^{-(x-\xi)/\alpha}/\{1 + e^{-(x-\xi)/\alpha}\}^2, \quad -\infty < x < \infty$$

$$F(x) = 1/\{1 + e^{-(x-\xi)/\alpha}\}$$

$$x(F) = \xi + \alpha \log\{F/(1-F)\}$$

PWMs, etc.

$$r\alpha_{r-1} = \xi - \alpha\{\psi(r) - \psi(1)\} = \xi - \alpha \sum_{s=1}^{r-1} s^{-1}$$

$$r\beta_{r-1} = \xi + \alpha\{\psi(r) - \psi(1)\} = \xi + \alpha \sum_{s=1}^{r-1} s^{-1} \qquad \text{(A.5.1)}$$

$$\lambda_1 = \xi$$
$$\lambda_2 = \alpha$$
$$\lambda_3 = 0$$
$$\lambda_4 = \alpha/6$$

$$\lambda_r = \begin{cases} 0, & r \text{ odd}, \geq 3, \\ 2\alpha/\{r(r-1)\}, & r \text{ even}, \geq 4. \end{cases} \qquad \text{(A.5.2)}$$

$$\tau_3 = 0$$

$$\tau_4 = \tfrac{1}{6}$$

$$\tau_r = \begin{cases} 0, & r \text{ odd}, \geq 3, \\ 2/\{r(r-1)\}, & r \text{ even}, \geq 4. \end{cases}$$

Asymptotic sampling variances

$$
\begin{aligned}
A_{00} &= B_{00} = 2\psi'(1) \\
A_{0s} &= B_{0s} = \psi'(s+1) + s^{-1}\{\psi(s+1) - \psi(1)\} \\
A_{rs} &= B_{rs} = \frac{\alpha^2}{rs}\{(r+s)\psi(r+s+1) - r\psi(r+1) - s\psi(s+1)\}
\end{aligned}
\qquad \text{(A.5.3)}
$$

$$
A = B = \alpha^2 \begin{bmatrix} \pi^2/3 & & & \\ \pi^2/6 & 1 & & \\ \pi^2/6 - 1/2 & 3/4 & 7/12 & \\ \pi^2/6 - 3/4 & 11/18 & 35/72 & 37/90 \end{bmatrix}
$$

$$
\Lambda = \alpha^2 \begin{bmatrix} \pi^2/3 & & & \\ 0 & 4 - \pi^2/3 & & \\ \pi^2/3 - 3 & 0 & \pi^2/3 - 3 & \\ 0 & 31/9 - \pi^2/3 & & 31/9 - \pi^2/3 \end{bmatrix}
$$

$$
= \alpha^2 \begin{bmatrix} 3.2899 & & & \\ 0 & 0.7101 & & \\ 0.2899 & 0 & 0.2899 & 0 \\ 0 & 0.1546 & 0 & 0.1546 \end{bmatrix}
$$

$$
T = \begin{bmatrix} \alpha^2\pi^2/3 & & & \\ 0 & \alpha^2(4 - \pi^2/3) & & \\ \alpha(\pi^2/3 - 3) & 0 & \pi^2/3 - 3 & \\ 0 & \alpha(25/9 - 5\pi^2/18) & 0 & 65/27 - 25\pi^2/108 \end{bmatrix}
$$

$$
= \begin{bmatrix} 3.2899\alpha^2 & & & \\ 0 & 0.7101\alpha^2 & & \\ 0.2899\alpha & 0 & 0.2899 & \\ 0 & 0.0362\alpha & 0 & 0.1228 \end{bmatrix}
$$

$$\xi = \lambda_1, \quad \alpha = \lambda_2$$

The resulting estimators have $n \operatorname{var} \hat{\xi} \sim 3.2899$, $n \operatorname{var} \hat{\alpha} \sim 0.7101$. However, we also have $\xi = \lambda_1 - \lambda_3$: this gives an estimator $\hat{\xi} = \ell_1 - \ell_3$ with $n \operatorname{var} \hat{\xi} \sim 3$. We therefore define the PWM estimators to be

$$\hat{\xi} = \ell_1 - \ell_3, \quad \hat{\alpha} = \ell_2;$$

we have

$$n \operatorname{var} \begin{bmatrix} \hat{\xi} \\ \hat{\alpha} \end{bmatrix} \sim \alpha^2 \begin{bmatrix} 3 & 0 \\ 0 & 0.7101 \end{bmatrix},$$

cf.

$$n \operatorname{var} \begin{bmatrix} \hat{\xi}_{ML} \\ \hat{\alpha}_{ML} \end{bmatrix} \sim \alpha^2 \begin{bmatrix} 3 & 0 \\ 0 & 9/(3 + \pi^2) \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 0.6993 \end{bmatrix}$$

(see, for example, Johnson and Kotz, 1970, vol. 2, p. 7); thus

$$\operatorname{eff}(\hat{\xi}) = 1, \quad \operatorname{eff}(\hat{\alpha}) = 0.9848, \quad \operatorname{eff} \begin{bmatrix} \hat{\xi} \\ \hat{\alpha} \end{bmatrix} = 0.9848.$$

Note that $\lambda_1 - \lambda_3 = EX_{2:3}$, the expected value of the median of a sample of size 3. The finite-sample properties of these estimators are not known.

In general we could define estimators

$$\tilde{\xi} = \sum_{j=1}^{\infty} P_j \ell_{2j-1}, \quad \tilde{\alpha} = \sum_{j=1}^{\infty} Q_j \ell_{2j}$$

with the P_j and Q_j chosen to minimize the asymptotic variance of the estimators. This would yield, asymptotically, the "optimal linear estimators" of Chernoff *et al.* (1967). In fact, $\tilde{\xi} = \hat{\xi}$ (one cannot find an estimator with asymptotic variance smaller than that of $\hat{\xi}_{ML}$) while $\tilde{\alpha}$ is a rather complicated linear combination of the $x_{k:n}$ and in practice the 98½% efficient estimator $\hat{\alpha}$ should be adequate.

A.6. Normal

Definition

$$f(x) = \phi\left(\frac{x-\mu}{\sigma}\right), \quad -\infty < x < \infty$$

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$x(F)$ has no explicit analytical form

Here

$$\phi(x) = (2\pi)^{-1/2} e^{-x^2/2}, \quad \Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} \phi(t)\, dt .$$

PWMs, etc.

$$\alpha_0 = \beta_0 = \mu$$

$$\alpha_0 - 2\alpha_1 = 2\beta_1 - \beta_0 = \pi^{-1/2}\sigma = 0.5642\sigma \qquad (A.6.1)$$

$$\alpha_0 - 3\alpha_2 = 3\beta_2 - \beta_0 = \tfrac{3}{2}\pi^{-1/2}\sigma = 0.8463\sigma \qquad (A.6.2)$$

$$\alpha_0 - 4\alpha_3 = 4\beta_3 - \beta_0 = 6\pi^{-3/2}\arctan(\sqrt{2}\,)\,\sigma = 1.0294\sigma \qquad (A.6.3)$$

PWMs of higher order do not in general have analytical expressions, but can be found using tables of expected values of Normal order statistics.

$$\lambda_1 = \mu$$

$$\lambda_2 = \pi^{-1/2}\sigma = 0.5642\sigma$$

$$\lambda_3 = 0$$

$$\lambda_4 = (30\pi^{-3/2}\arctan\sqrt{2} - 9\pi^{-1/2})\,\sigma = 0.0692\sigma$$

$$\tau_3 = 0$$

$$\tau_4 = 30\pi^{-1}\arctan\sqrt{2} - 9 = 0.1226$$

Asymptotic sampling variances

$$A = B = \sigma^2 \begin{bmatrix} 1 & & & \\ 0.5000 & 0.2907 & & \\ 0.3333 & 0.2074 & 0.1535 & \\ 0.2500 & 0.1619 & 0.1227 & 0.0997 \end{bmatrix} \quad \text{(A.6.4)}$$

$$\Lambda = \sigma^2 \begin{bmatrix} 1 & & & \\ 0 & 0.1627 & & \\ 0 & 0 & 0.0594 & \\ 0 & 0.0107 & 0 & 0.0281 \end{bmatrix}$$

$$T = \begin{bmatrix} \sigma^2 & & & \\ 0 & 0.1627\sigma^2 & & \\ 0 & 0 & 0.1866 & \\ 0 & -0.0164\sigma & 0 & 0.0878 \end{bmatrix}$$

By Basu's theorem (Basu, 1955), ℓ_1 is independent of ℓ_r for all $r \geq 2$ and for all sample sizes. The exact variance of ℓ_2 is (Nair, 1936)

$$\text{var}\, \ell_2 = \frac{\sigma^2}{n(n-1)} \left\{ 1 + (n-2)\left(\frac{1}{3} + \frac{2\sqrt{3}}{\pi} \right) - \frac{2(2n-3)}{\pi} \right\}.$$

Parameter estimation

$$\mu = \lambda_1, \quad \sigma = \pi^{1/2}\lambda_2$$

The resulting estimators, particularly $\hat{\sigma} = \pi^{1/2}\ell_2$, have been discussed by Downton (1966a), who shows that

$$n\, \text{var} \begin{bmatrix} \hat{\mu} \\ \hat{\sigma} \end{bmatrix} \sim \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 0.5113 \end{bmatrix};$$

cf.

$$n \operatorname{var}\begin{bmatrix}\hat{\mu}_{ML}\\ \hat{\sigma}_{ML}\end{bmatrix} \sim \sigma^2\begin{bmatrix}1 & 0\\ 0 & 0.5\end{bmatrix}.$$

Thus

$$\text{eff}\,(\hat{\mu}) \sim 1, \quad \text{eff}\,(\hat{\sigma}) \sim 0.9779.$$

More efficient estimators of σ could be defined if required (*cf.* the logistic distribution, section A.5). For example, the best linear combination of ℓ_2 and ℓ_4 is

$$1.7004\,\ell_2 + 0.5875\,\ell_4$$

with mean σ, asymptotic variance $0.5014n^{-1}$ and asymptotic efficiency 99.9%.

A.7. Rayleigh

Definition

$$f(x) = \frac{(x-\xi)}{\sigma^2} \exp\{-\tfrac{1}{2}(x-\xi)^2/\sigma^2\}, \quad \xi \le x < \infty$$

$$F(x) = 1 - \exp\{-\tfrac{1}{2}(x-\xi)^2/\sigma^2\}$$

$$x(F) = \xi + \{-2\sigma^2 \log(1-F)\}^{1/2}$$

It is often assumed (for example by Johnson and Kotz, 1970, vol. 1, p. 197) that $\xi = 0$.

PWMs, etc.

$$r\alpha_{r-1} = \xi + \sigma\sqrt{\pi/(2r)} \qquad \text{(A.7.1)}$$

$$\lambda_1 = \xi + \sigma\sqrt{\pi/2}$$

$$\lambda_2 = \tfrac{1}{2}\sigma\sqrt{\pi}(\sqrt{2} - 1)$$

$$\lambda_3 = (1 - 3/\sqrt{2} + 2/\sqrt{3})\,\sigma\sqrt{\pi/2}$$

$$\lambda_4 = (1 - 6/\sqrt{2} + 10/\sqrt{3} - 5/\sqrt{5})\,\sigma\sqrt{\pi/2}$$

$$\tau_3 = (1 - 3/\sqrt{2} + 2/\sqrt{3})/(1 - 1/\sqrt{2}) = 0.1140$$

$$\tau_4(1 - 6/\sqrt{2} + 10/\sqrt{3} - 5/\sqrt{4})/(1 - 1/\sqrt{2}) = 0.1054$$

Asymptotic sampling variances

$$A_{r,s} = \sigma^2 \left\{ \frac{\arctan\sqrt{r/(s+1)}}{\sqrt{r(s+1)}} + \frac{\arctan\sqrt{s/(r+1)}}{\sqrt{(r+1)s}} - \frac{\pi/2}{\sqrt{(r+1)(s+1)}} \right\} \qquad \text{(A.7.2)}$$

The elements of Λ and T can be calculated from (A.7.2).

Parameter estimation

$$\sigma = 2\pi^{1/2}\lambda_2/(\sqrt{2} - 1), \quad \xi = \lambda_1 - \sigma\sqrt{\pi/2}$$

Properties of the resulting estimators have not been investigated.

A.8. Generalized Pareto

Definition

$$f(x) = \alpha^{-1} e^{-(1-k)y} \quad \text{where} \quad y = \begin{cases} -k^{-1} \log\{1 - k(x-\xi)/\alpha\}, & k \neq 0, \\ (x-\xi)/\alpha, & k = 0. \end{cases}$$

$0 \le x \le \alpha/k$ if $k > 0$, $0 \le x < \infty$ if $k \le 0$.

$$F(x) = 1 - e^{-y}$$

$$x(F) = \begin{cases} \xi + \alpha\{1 - (1-F)^k\}/k, & k \neq 0, \\ \xi - \alpha \log(1-F), & k = 0. \end{cases}$$

Special cases: $k = 0$ is the exponential distribution, $k = 1$ is the uniform distribution on the interval $0 \le x \le \alpha$.

PWMs, etc.

$$r\alpha_{r-1} = \xi + \alpha/(r+k), \qquad k > -1 \qquad \text{(A.8.1)}$$

$$r\beta_{r-1} = \xi + \frac{\alpha}{k}\left\{1 - \frac{\Gamma(1+k)\Gamma(1+r)}{\Gamma(1+k+r)}\right\}, \quad k > -1 \qquad \text{(A.8.2)}$$

$$\lambda_1 = \xi + \alpha/(1+k)$$
$$\lambda_2 = \alpha/\{(1+k)(2+k)\}$$
$$\lambda_3 = \alpha(1-k)/\{(1+k)(2+k)(3+k)\}$$
$$\lambda_4 = \alpha(1-k)(2-k)/\{(1+k)(2+k)(3+k)(4+k)\}$$

$$\lambda_r = \alpha \frac{\Gamma(1+k)\Gamma(r-1-k)}{\Gamma(1-k)\Gamma(r+1+k)} \qquad \text{(A.8.3)}$$

$$\tau_3 = (1-k)/(3+k)$$
$$\tau_4 = (1-k)(2-k)/\{(3+k)(4+k)\}$$

$$\tau_r = \frac{\Gamma(3+k)\Gamma(r-1-k)}{\Gamma(1-k)\Gamma(r+1+k)}$$

Asymptotic sampling variances

$$A_{rs} = \frac{\alpha^2}{(r+1+k)(s+1+k)(r+s+1+2k)} \qquad \text{(A.8.4)}$$

Λ_{rs}, *etc.*, can be calculated as required but have no simple general form.

Parameter estimation

Case ξ known: this is the more realistic case, for X has a lower bound at which $f(x) > 0$, so the location parameter can be estimated (by $x_{1:n}$, for example) with variance $O(n^{-2})$. Assuming ξ known and, without loss of generality, $\xi = 0$, we have

$$k = \lambda_1/\lambda_2 - 2, \quad \alpha = (1+k)\lambda_1.$$

The resulting estimators have been investigated by Hosking and Wallis (1986). They recommend the use of plotting-position estimators $\tilde{\lambda}_r[-0.35, 0]$, and show that

$$n\,\mathrm{var}\begin{bmatrix}\hat{\alpha}\\ \hat{k}\end{bmatrix} \sim \frac{1}{(1+2k)(3+2k)} \times \begin{bmatrix}\alpha^2(7+18k+11k^2+2k^3) & \\ \alpha(2+k)(2+6k+7k^2+2k^3) & (1+k)(2+k)^2(1+k+2k^2)\end{bmatrix}, \quad k > -\tfrac{1}{2}.$$

while

$$n\,\mathrm{var}\begin{bmatrix}\hat{\alpha}_{ML}\\ \hat{k}_{ML}\end{bmatrix} \sim \begin{bmatrix}2\alpha^2(1-k) & \\ \alpha(1-k) & (1-k)^2\end{bmatrix}, \quad k < \tfrac{1}{2}.$$

PWM estimators are most efficient for $k \approx -0.2$, with eff $(\hat{k}) = 0.92$ for $k = -0.2$. For $k > -0.1$, conventional moment estimators are asymptotically more efficient than PWM estimators, being, indeed, 100% efficient at $k = 0$. Hosking and Wallis's (1986) simulation results for finite samples show that PWM estimation is most useful when $k \leq -0.2$, because PWM parameter and quantile estimators are then less biased than moment or maximum-likelihood estimators, or when the asymptotic sampling variances are to be used for constructing finite-sample confidence intervals.

Case ξ unknown: analogy with the uniform and exponential distributions (sections A.1, A.2) suggests that we base PWM estimators on λ_1, λ_2 and $n\alpha_{n-1}$, the last-named quantity being estimated by $x_{1:n}$. We obtain

$$\hat{k} = \frac{n(\ell_1 - x_{1:n}) - 2(n-1)\ell_2}{(n-1)\ell_2 - (\ell_1 - x_{1:n})}, \quad \hat{\alpha} = (1 + \hat{k})(2 + \hat{k})\ell_2, \quad \hat{\xi} = x_{1:n} - \hat{\alpha}/(n + \hat{k}).$$

This should ensure that var $\hat{\xi} = O(n^{-2})$ and that the asymptotic variances of $\hat{\alpha}$ and $\hat{k}$ are the same as in the case of ξ known.

A.9. Generalized extreme-value

Defintion

$$f(x) = \alpha^{-1} e^{-(1-k)y} e^{-e^{-y}} \quad \text{where} \quad y = \begin{cases} -k^{-1} \log\{1 - k(x - \xi)/\alpha\}, & k \neq 0, \\ (x - \xi)/\alpha, & k = 0, \end{cases}$$

$$\xi + \alpha/k \le x < \infty \text{ if } k < 0, \quad -\infty < x < \infty \text{ if } x = 0, \quad -\infty < x \le \xi + \alpha/k \text{ if } k > 0$$

$$F(x) = e^{-e^{-y}}$$

$$x(f) = \begin{cases} \xi + \alpha\{1 - (-\log F)^k\}/k, & k \neq 0, \\ \xi - \alpha \log(-\log F), & k = 0. \end{cases}$$

Special cases: $k = 0$ is the Gumbel distribution; $k = \frac{1}{2}$ is a reverse Rayleigh distribution; $k = 1$ is a reverse exponential distribution. The Weibull distribution defined by

$$F(x) = 1 - \exp[-\{(x - \lambda)/\beta\}^{\gamma}], \quad \lambda \le x < \infty,$$

is a reverse GEV distribution with parameters

$$k = 1/\gamma, \quad \alpha = \beta/\gamma, \quad \xi = \lambda + \beta,$$

and results for it may be derived from those below.

PWMs, etc.

$$r\beta_{r-1} = \xi + \alpha\{1 - r^{-k}\Gamma(1 + k)\}/k, \quad k > -1 \qquad (A.9.1)$$

$$\lambda_1 = \xi + \alpha\{1 - \Gamma(1 + k)\}/k$$

$$\lambda_2 = \alpha(1 - 2^{-k})\Gamma(1 + k)/k$$

$$\lambda_3 = \alpha(1 - 3.2^{-k} + 2.3^{-k})\Gamma(1 + k)/k$$

$$\lambda_4 = \alpha(1 - 6.2^{-k} + 10.3^{-k} - 5.4^{-k})\Gamma(1 + k)/k$$

$$\tau_3 = 2(1 - 3^{-k})/(1 - 2^{-k}) - 3$$

$$\tau_4 = \{5(1 - 4^{-k}) - 10(1 - 3^{-k}) + 6(1 - 2^{-k})\}/(1 - 2^{-k})$$

Asymptotic sampling variances

$$\begin{aligned}
B_{r,r} &= \alpha^2 k^{-2}(r+1)^{-2k}[\Gamma(1+2k)G\{r/(r+1)\} - \{\Gamma(1+k)\}^2] \\
B_{r,r+1} &= \tfrac{1}{2}\alpha^2 k^{-2}[(r+2)^{-2k}\Gamma(1+2k)G\{r/(r+2)\} \\
&\quad + (r+1)^{-k}\{(r+1)^{-k} - 2(r+2)^{-k}\}\{\Gamma(1+k)\}^2] \\
B_{r,r+s} &= \tfrac{1}{2}\alpha^2 k^{-2}[(r+s+1)^{-2k}\Gamma(1+2k)G\{r/(r+s+1)\} \\
&\quad - (r+s)^{-2k}\Gamma(1+2k)G\{(r+1)/(r+s)\} \\
&\quad + 2(r+1)^{-k}\{(r+s)^{-k} - (r+s+1)^{-k}\}\{\Gamma(1+k)\}^2], \quad s \geq 2.
\end{aligned} \tag{A.9.2}$$

where

$$G(x) = {}_2F_1(k, 2k; 1+k; -x).$$

$\Lambda_{r,s}$, B, Λ, T can be calculated numerically as required.

Parameter estimation

The shape parameter k is the solution of the equation

$$\frac{(3\beta_2 - \beta_0)}{(2\beta_1 - \beta_0)} = \frac{(1 - 3^{-k})}{(1 - 2^{-k})}.$$

Hosking *et al.* (1985) suggest the approximate solution

$$k = 7.8590c + 2.9554c^2, \quad c = \frac{(2\beta_1 - \beta_0)}{(3\beta_2 - \beta_0)} - \frac{\log 2}{\log 3}.$$

The other parameters are then given by

$$\alpha = \frac{\lambda_2 k}{(1 - 2^{-k})\Gamma(1 + k)}, \quad \xi = \lambda_1 - \alpha\{1 - \Gamma(1 + k)\}/k.$$

Asymptotic variances of parameter and quantile estimators can be calculated from the $B_{r,s}$. Hosking *et al.* (1985) report some such results, and also assess the performance of the PWM estimators in finite samples. They recommend the use of plotting-position estimators $\tilde{\beta}_r[-0.35, 0]$. PWM estimation seems preferable to maximum-likelihood estimation, at least for samples of size 100 or less.

Other expressions for k, and thence estimators of k, can be found, for example

$$k = -\log_2\{(4\beta_3 - 2\beta_1)/(2\beta_1 - \beta_0)\},$$

obtained by Greenwood *et al.* (1979) for the GEV distribution in its Weibull parametrization. Our unpublished research indicates that the estimator based on $(3\beta_2 - \beta_0)/(2\beta_1 - \beta_0)$ is preferable because it has greater asymptotic efficiency over a range of especially plausible k values, *viz.* $-0.2 \le k \le 0$.

PWM estimates of the GEV distribution can be infeasible in the sense that the fitted distribution has an upper or lower bound and one of the data values lies outside this bound. For example, the data set $\{-4, 0, 1, 1, 2\}$ yields, using unbiased PWM estimators, a fitted GEV distribution with upper bound 1.92. In such cases it may be preferable to modify the parameter estimates so that all the data lie within the range of the fitted distribution: specifically, that either the largest or the smallest data value lies exactly on the bound of the distribution. To do this we equate $x \equiv x_{1:n}$ or $x \equiv x_{n:n}$ to the bound of the distribution obtaining

$$x = \xi + \alpha/k = \lambda_0 + \lambda_1/(1 - 2^{-k})$$

and

$$k = -\log\left(\frac{2\beta_1 - x}{\beta_0 - x}\right) / \log 2 .$$

The parameters α and ξ can then be estimated as before. For typical values of k, $-\frac{1}{2} < k < \frac{1}{2}$ say, infeasibility of PWM parameter estimates occurs very rarely even in small samples and should have a negligible effect on the asymptotic properties of the estimators. Only when $k > 1$ should we expect otherwise, for only then do we have $f(x) > 0$ at the bound. In this case we could consider (*cf.* the uniform and exponential distributions, and the generalized Pareto distribution with ξ unknown) basing parameter estimation on λ_0, λ_1 and $n\beta_{n-1}$, the last being estimated by $x_{n:n}$, but the resulting equation for k is

$$\frac{1 - n^{-k}}{1 - 2^{-k}} = \frac{x_{n:n} - \lambda_0}{\lambda_1}$$

and does not have for general n a neat approximate solution like that found by Hosking *et al.* (1985) for $n = 3$. Note that the $k > 1$ case could be important in practice since it corresponds to a Weibull distribution with shape parameter $\gamma < 1$.

A.10. Generalized logistic

Definition

$$f(x) = \alpha^{-1} e^{-(1-k)y}/(1+e^{-y})^2 \quad \text{where} \quad y = \begin{cases} -k^{-1}\log\{1-k(x-\xi)/\alpha\}, & k \neq 0, \\ (x-\xi)/\alpha, & k = 0, \end{cases}$$

$$\xi + \alpha/k \le x < \infty \text{ if } k < 0, \quad -\infty < x < \infty \text{ if } k = 0, \quad -\infty < x \le \xi + \alpha/k \text{ if } k > 0$$

$$F(x) = 1/(1+e^{-y})$$

$$x(F) = \begin{cases} \xi + \alpha[1 - \{(1-F)/F\}^k]/k, & k \neq 0, \\ \xi - \alpha\log\{(1-F)/F\}, & k = 0. \end{cases}$$

Special case: $k = 0$ is the logistic distribution. This generalization of the logistic distribution differs from others which have been defined in the literature (*e.g.* Gumbel, 1944; Dubey, 1969); it has not previously been used in statistics. The name is chosen to reflect the distribution's similarity to the generalized Pareto and generalized extreme-value distributions.

PWMs, etc.

$$\begin{aligned} r\alpha_{r-1} &= \xi + \frac{\alpha}{k}\left\{1 - \frac{\Gamma(1-k)\Gamma(r+k)}{\Gamma(r)}\right\}, \quad |k| < 1 \\ r\beta_{r-1} &= \xi + \frac{\alpha}{k}\left\{1 - \frac{\Gamma(1+k)\Gamma(r-k)}{\Gamma(r)}\right\}, \quad |k| < 1 \end{aligned} \qquad \text{(A.10.1)}$$

$$\lambda_1 = \xi + \alpha\{1 - \Gamma(1-k)\Gamma(1+k)\}/k$$
$$\lambda_2 = \alpha\Gamma(1-k)\Gamma(1+k)$$
$$\lambda_3 = -k\alpha\Gamma(1-k)\Gamma(1+k)$$
$$\lambda_4 = \tfrac{1}{6}(1+5k^2)\alpha\Gamma(1-k)\Gamma(1+k)$$

$$\tau_3 = -k$$
$$\tau_4 = (1+5k^2)/6$$

There seems to be no simple general expression for λ_r or τ_r, though we can work out individual values from the definitions, *e.g.*

$$\tau_5 = -k(5 + 7k^2)/12, \quad \tau_6 = (4 + 35k^2 + 21k^4)/60.$$

Asymptotic sampling variances

$$B_{rs} = J_{rs} + J_{sr}$$

where

$$J_{rs} = \frac{\alpha^2}{1+k} \cdot \frac{\Gamma(1+2k)\Gamma(r+s+1-2k)}{\Gamma(r+s+2)} \, {}_3F_2\left[\begin{matrix} 1, s+1, 1+2k \\ r+s+2, 2+k \end{matrix}\right], \quad |k| < \tfrac{1}{2} \; ; \qquad \text{(A.10.2)}$$

$$A_{rs} = I_{rs} + I_{sr}$$

where I_{rs} is J_{rs} with the substitution $k \to -k$. One special case gives a simple result:

$$A_{00} = B_{00} = \alpha^2 \Gamma(1-2k)\Gamma(1+2k)\{\psi(1+k) - \psi(1-k)\}/k.$$

Otherwise A_{rs} and B_{rs}, and thence Λ_{rs} and T_{rs}, can be obtained numerically from (A.10.2).

Parameter estimation

The shape parameter is of course estimated using $k = -\tau_3$: indeed this simple relationship suggests the generalized logistic distribution as a natural choice for modelling data with L-moments. The full set of parameter/L-moment relationships is

$$k = -\tau_3, \quad \alpha = \lambda_2/\{\Gamma(1-k)\Gamma(1+k)\}, \quad \xi = \lambda_1 - \alpha\{1 - \Gamma(1-k)\Gamma(1+k)\}/k.$$

Properties of the resulting estimators have not been investigated.

As with the GEV distribution (section A.9) parameter estimates for the generalized logistic distribution could on rare occasions be infeasible. To overcome this the largest or smallest data value, x say, could be equated to the bound of the distribution, *viz.* $\xi + \alpha/k = \lambda_1 + \lambda_2/k$, to yield $k = \lambda_2/(x - \lambda_1)$ and thence estimates of α and ξ as above.

Conventional moments

This generalized logistic distribution has not featured previously in the literature, so for completeness we give here its conventional moments. We have

$$E\{1 - k(X-\xi)/\alpha\}^r = \int_0^1 \{(1-F)/F\}^r dF = \Gamma(1-rk)\Gamma(1+rk), \quad |k| < 1/r,$$

so

$$\begin{aligned}
EX &= \xi + \alpha(1-g_1)/k & &\sim \xi & &\text{if } k = 0,\\
\operatorname{var} X &= \alpha^2(g_2 - g_1^2)/k^2 & &\sim \alpha^2\pi^2/3 & &\text{if } k = 0,\\
\text{skewness }(X) &= (-\operatorname{sign} k)(g_3 - 3g_2g_1 + 2g_1^3)/(g_2 - g_1^2)^{3/2} & &\sim 0 & &\text{if } k = 0,\\
\text{kurtosis }(X) &= (g_4 - 4g_3g_1 + 6g_2g_1^2 - 3g_1^4)/(g_2 - g_1^2)^2 & &\sim 21/5 & &\text{if } k = 0,
\end{aligned}$$

where

$$g_r = \Gamma(1-rk)\Gamma(1+rk).$$

A.11. Generalized Normal

Definition

$$f(x) = (2\pi)^{-1/2}\alpha^{-1}e^{ky-y^2/2} \quad \text{where} \quad y = \begin{cases} -k^{-1}\log\{1 - k(x-\xi)/\alpha\}, & k \neq 0, \\ (x-\xi)/\alpha, & k = 0, \end{cases}$$

$$\xi + \alpha/k \le x < \infty \text{ if } k < 0, \quad -\infty < x < \infty \text{ if } x = 0, \quad -\infty < x \le \xi + \alpha/k \text{ if } k > 0$$

$F(x) = \Phi(y)$ where Φ is as defined in section A.6

$x(F)$ has no explicit analytical form

Special case: $k = 0$ is the Normal distribution. The lognormal distribution defined by

$$F(x) = \Phi[\{\log(x-\gamma) - \mu\}/\sigma], \quad \gamma \le x < \infty,$$

is a generalized Normal distribution with parameters

$$k = -\sigma, \quad \alpha = \sigma e^{\mu}, \quad \xi = \gamma + e^{\mu},$$

and results for it may be derived from those below.

The reparametrization of the lognormal distribution in terms of ξ, α and k is a small modification of the parametrization of Munro and Wixley (1970). It has several advantages over the usual parametrization using μ, σ and γ:

(a) within a single distribution it includes both lognormal distributions with positive skewness and a lower bound ($k < 0$), and lognormal distributions with negative skewness and an upper bound ($k > 0$);

(b) it includes the Normal distribution as a special case ($k = 0$) rather than as an unattainable limit;

(c) it exhibits the similarity in structure of the lognormal distribution to the generalized Pareto and generalized extreme-value distributions;

(d) its parameters are more meaningful and more stable to estimate than are those of the lognormal distribution, particularly when the skewness is close to zero.

PWMs, etc.

$$\beta_0 = \xi + \frac{\alpha}{k}(1 - e^{k^2/2}) \tag{A.11.1}$$

$$2\beta_1 - \beta_0 = \frac{\alpha}{k} e^{k^2/2} \{1 - 2\Phi(-k/\sqrt{2})\} \tag{A.11.2}$$

$$3\beta_2 - \beta_0 = \frac{\alpha}{k} e^{k^2/2} \left\{1 - 6\int_{-\infty}^{-k/\sqrt{2}} \Phi(t/\sqrt{3})\phi(t)dt\right\} \tag{A.11.3}$$

$$4\beta_3 - \beta_0 = -24\frac{\alpha}{k} e^{k^2/2} \int_0^{-k/\sqrt{2}} \left\{ \int_0^{t/\sqrt{3}} \Phi(u/\sqrt{2})\phi(u)du \right\} \phi(t)dt \tag{A.11.4}$$

where Φ and ϕ are as defined in section A.6.

$$\begin{aligned}
\lambda_1 &= \xi + \alpha k^{-1}(1 - e^{k^2/2}) \\
\lambda_2 &= \alpha k^{-1} e^{k^2/2}\{1 - 2\Phi(-k/\sqrt{2})\} \\
\lambda_3 &= 6\alpha k^{-1} e^{k^2/2}\{\Phi(-k/\sqrt{2}) - 2\Psi(-k/\sqrt{2})\} \\
\lambda_4 &= 12\alpha k^{-1} e^{k^2/2}[(1 + 5\pi^{-1} \arctan\sqrt{2})\{2\Phi(-k/\sqrt{2}) - 1\} \\
&\quad - 5\Psi(-k/\sqrt{2}) + 10\Xi(-k/\sqrt{6})]
\end{aligned} \tag{A.11.5}$$

where

$$\Psi(x) = \int_{-\infty}^{x} \Phi(t/\sqrt{3})\phi(t)dt, \quad \Xi(x) = \int_0^x \{\Phi(x\sqrt{3}) - \Phi(t\sqrt{3})\}\Phi(t/\sqrt{2})\phi(t)dt.$$

$$\tau_3 = \frac{12\Psi(-k/\sqrt{2}) - 6\Phi(-k/\sqrt{2})}{2\Phi(-k/\sqrt{2}) - 1}$$

$$\tau_4 = \frac{120\Xi(-k/\sqrt{6}) - 60\Psi(-k/\sqrt{2})}{2\Phi(-k/\sqrt{2}) - 1} + 12(1 + 5\pi^{-1} \arctan \sqrt{2})$$

No general expressions for α_r, β_r, λ_r or τ_r are known.

Asymptotic sampling variances

None of these is known except

$$A_{00} = B_{00} = \operatorname{var} X = \alpha^2 e^{k^2} (e^{k^2} - 1)/k^2$$

Parameter estimation

To estimate k we must solve the equation

$$\tau_3 = \frac{12\Psi(-k/\sqrt{2}) - 6\Phi(-k/\sqrt{2})}{2\Phi(-k/\sqrt{2}) - 1}$$

for k. No explicit solution is possible, but the approximate solution

$$k \approx s(0.999281 - 0.006118 s^2 + 0.000127 s^4), \quad s = -\sqrt{8/3}\,\Phi^{-1}\left(\frac{1 + \tau_3}{2}\right), \qquad \text{(A.11.6)}$$

seems adequate for almost all practical purposes. The other parameters are then given by

$$\alpha = \lambda_2 k e^{-k^2/2} / \{1 - 2\Phi(-k/\sqrt{2})\}, \quad \xi = \lambda_1 - \alpha k^{-1}(1 - e^{k^2/2}).$$

Properties of the resulting estimators have not been investigated.

A.12. Gamma

Definition

$$f(x) = \frac{(x-\xi)^{\alpha-1} e^{-(x-\xi)/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, \quad \xi \le x < \infty$$

$$F(x) = \gamma\left(\alpha, \frac{x-\xi}{\beta}\right) / \Gamma(\alpha)$$

$x(F)$ has no explicit analytical form

Here

$$\gamma(\alpha, x) = \int_0^x t^{\alpha-1} e^{-t} dt$$

is the incomplete gamma function.

Special case: $\alpha = 1$ is the exponential distribution.

PWMs, etc.

$$\beta_0 = \xi + \alpha\beta$$

$$2\beta_1 - \beta_0 = \pi^{-1/2}\beta\,\Gamma(\alpha + 1/2)/\Gamma(\alpha) \qquad \text{(A.12.1)}$$

$$3\beta_2 - \beta_0 = 3\pi^{-1/2}\beta\, I_{1/3}(\alpha, 2\alpha)\,\Gamma(\alpha + 1/2)/\Gamma(\alpha) \qquad \text{(A.12.2)}$$

$$4\beta_3 - \beta_0 = 6\pi^{-1/2}\beta\,\Delta\,\Gamma(\alpha + 1/2)/\Gamma(\alpha) \qquad \text{(A.12.3)}$$

where

$$I_x(p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^x t^{p-1}(1-t)^{q-1} dt$$

is the incomplete β-function ratio and

$$\Delta = \frac{\Gamma(3\alpha)}{\Gamma(\alpha)\Gamma(2\alpha)} \int_0^{1/3} t^{\alpha-1}(1-t)^{2\alpha-1} I_{(1-t)/(3-t)}(\alpha, 3\alpha)\, dt.$$

$$\lambda_1 = \xi + \alpha\beta$$

$$\lambda_2 = \pi^{-1/2} \beta\, \Gamma(\alpha + 1/2)/\Gamma(\alpha)$$

$$\lambda_3 = \{6\, I_{1/3}(\alpha, 2\alpha) - 3\}\pi^{-1/2}\beta\, \Gamma(\alpha + 1/2)/\Gamma(\alpha)$$

$$\lambda_4 = 6\{1 - 5I_{1/3}(\alpha, 2\alpha) + 5\Delta\}\pi^{-1/2}\beta\, \Gamma(\alpha + 1/2)/\Gamma(\alpha)$$

$$\tau_3 = 6\, I_{1/3}(\alpha, 2\alpha) - 3$$

$$\tau_4 = 6\{1 - 5I_{1/3}(\alpha, 2\alpha) + 5\Delta\}$$

No general expressions for α_r, β_r, λ_r or τ_r are known.

Asymptotic sampling variances

None of these is known except

$$A_{00} = B_{00} = \operatorname{var} X = \alpha\beta^2.$$

Parameter estimation

Case ξ known and, without loss of generality, $\xi = 0$: we must solve the equation

$$\tau \equiv \lambda_2/\lambda_1 = \pi^{-1/2}\Gamma(\alpha + ½)/\Gamma(\alpha + 1)$$

for α. Iterative methods seem necessary: we suggest the iteration

$$\alpha^{(k+1)} = \frac{\pi^{-1/2}\Gamma(\alpha^{(k)} + 3/2)}{\tau\, \Gamma(\alpha^{(k)} + 1)} - \frac{1}{2}$$

with starting value

$$\alpha^{(0)} = \begin{cases} 1/(\pi\tau^2) & \text{if } \tau < 0.6, \\ -\log\tau/\log 4 & \text{if } \tau \geq 0.6 \end{cases}$$

(these expressions are derived from the asymptotic forms of the τ-α equation for large α and small α respectively). Relative error less than 0.1% should be achieved in 6-8 iterations. We can then estimate β from

$$\beta = \lambda_1/\alpha.$$

Case ξ unknown: we estimate α by solving

$$\tau_3 = 6\, I_{1/3}(\alpha, 2\alpha) - 3$$

for α. Again, iterative methods seem to be required, because the relationship between τ_3 and α is not easily inverted (even approximately). Given α, the other parameters may be found from

$$\beta = \pi^{1/2}\lambda_2\Gamma(\alpha)/\Gamma(\alpha + \tfrac{1}{2}), \quad \xi = \lambda_1 - \alpha\beta.$$

Properties of the PWM estimators, whether ξ is known or not, have not been investigated.

A.13. Generalized lambda

Definition

$$f(x),\ F(x) \text{ not explicitly defined}$$
$$x(F) = m + aF^{b} - c(1-F)^{d}$$

The 5-parameter form of the distribution, considered here, is mentioned by Joiner and Rosenblatt (1971). When the distribution has bene used in practice (which has been mainly in computer simulation studies, as an alternative to the Normal distribution), constraints have been placed on the *a*, *b*, *c* and *d* parameters, *e.g.*

$a=c,\ b=d$	(Tukey, 1960; Ramberg and Schmeiser, 1972);
$b=d$	(Shapiro and Wilk, 1965);
$a=c$	(Ramberg and Schmeiser, 1974; Ozturk and Dale, 1985).

PWMs, etc.

$$\begin{aligned} r\alpha_{r-1} &= m + ra\Gamma(b)/\Gamma(r+b) - rc/(r+d), \quad b>-1,\ d<-1 \\ r\beta_{r-1} &= m + rb/(r+b) - rc\Gamma(d)/\Gamma(r+d), \quad b>-1,\ d>-1 \end{aligned} \tag{A.13.1}$$

$$\lambda_1 = m + \frac{a}{(1+b)} - \frac{c}{(1+d)}$$
$$\lambda_2 = \frac{ab}{(1+b)(2+b)} + \frac{cd}{(1+d)(2+d)}$$
$$\lambda_3 = \frac{ab(b-1)}{(1+b)(2+b)(3+b)} - \frac{cd(d-1)}{(1+d)(2+d)(3+d)}$$
$$\lambda_4 = \frac{ab(b-1)(b-2)}{(1+b)(2+b)(3+b)(4+b)} + \frac{cd(d-1)(d-2)}{(1+d)(2+d)(3+d)(4+d)}$$

$$\lambda_r = \frac{a\{\Gamma(1+b)\}^2}{\Gamma(1+r+b)\Gamma(2-r+b)} + \frac{(-1)^r c\{\Gamma(1+d)\}^2}{\Gamma(1+r+d)\Gamma(2-r+d)} \tag{A.13.2}$$

There is no simple expression for τ_r.

Asymptotic sampling variances

$$B_{rs} = J_{rs} + J_{sr},$$

where

$$\begin{aligned}
J_{rs} = {} & \frac{a^2 b^2}{(r+1+b)(r+s+1+2b)(r+s+2+2b)} \\
& + \frac{abcd\Gamma(d)\Gamma(2+r)}{(s+b)(s+1+b)\Gamma(2+r+d)} - \frac{abcd\Gamma(d)\Gamma(r+s+2+b)}{(s+b)\Gamma(r+s+2+b+d)} \\
& + \frac{abcd\Gamma(d)\Gamma(r+s+3+b)}{(s+1+b)\Gamma(r+s+3+b+d)} + \frac{abcd\Gamma(1+d)\Gamma(r+s+2+b)}{(r+1+b)\Gamma(r+s+2+b+d)} \\
& + \frac{c^2 d^2 \Gamma(1+2d)\Gamma(r+2)}{(1+d)\Gamma(r+3+2d)}\, {}_3F_2\left[\begin{matrix} -2,\ 1+d,\ 1+2d \\ 2+d,\ r+3+2d \end{matrix}\right], \qquad b > -\tfrac{1}{2},\ d > -\tfrac{1}{2}
\end{aligned} \tag{A.13.3}$$

A_{rs} is obtained by the substitution $a \leftrightarrow c$, $b \leftrightarrow d$ in B_{rs}. A, B, Λ and T can be calculated from (A.13.3).

Parameter estimation

Case $a = c$, $b = d$ (symmetric distribution): Greenwood *et al.* (1979) give the following algorithm:

$$b = \{R - 7J_2 \pm (J_2^2 - 10J_2R + R^2)^{1/2}\}/(2J_2), \quad a = -R(b+1)(b+2)/b, \quad m = \alpha_0,$$

where

$$R = \alpha_1 - \alpha_0/2, \quad J_2 = \alpha_2 - \alpha_3 - \alpha_0/12.$$

There are in general two possible solutions for b, which can lead to very different fitted distributions. In practice one would choose the fitted distribution which more closely resembled the observed sample; thus one might choose the fitted distribution whose t_5 (or t_6) was closer to t_0 (or τ_0) calculated from the sample. Properties of these estimators have not been investigated.

In other cases ($a = c$, or $b = d$, or no parameter constraints at all) we have not been able to derive explicit representations for the parameters. The phenomenon of multiple solutions for the PWM estimators seems likely to recur, however. Appendix E provides some corroboration of this conjecture for the unrealistic case m known, a, b, c, d unknown and unconstrained.

A.14. Wakeby

Definition

$f(x)$, $F(x)$ not explicitly defined

$x(F) = \xi + \alpha\{1 - (1 - F)^{\beta}\}/\beta - \gamma\{1 - (1 - F)^{-\delta}\}/\delta$

Note that our parametization is different from the one used previously in the literature by Houghton (1978), Landwehr *et al.*, (1978, 1979b) and others. Hosking (1986) gives reasons for preferring the present parametrization. We shall assume that $\beta + \delta \geq 0$ (Hosking, 1986).

PWMs, etc.

$$r\alpha_{r-1} = \xi + \alpha/(r + \beta) + \gamma/(r - \delta), \qquad \delta < 1 \tag{A.14.1}$$

$$\lambda_1 = \xi + \frac{\alpha}{(1 + \beta)} + \frac{\gamma}{(1 - \delta)}$$

$$\lambda_2 = \frac{\alpha}{(1 + \beta)(2 + \beta)} + \frac{\gamma}{(1 - \delta)(2 - \delta)}$$

$$\lambda_3 = \frac{\alpha(1 - \beta)}{(1 + \beta)(2 + \beta)(3 + \beta)} + \frac{\gamma(1 + \delta)}{(1 - \delta)(2 - \delta)(3 - \delta)}$$

$$\lambda_4 = \frac{\alpha(1 - \beta)(2 - \beta)}{(1 + \beta)(2 + \beta)(3 + \beta)(4 + \beta)} + \frac{\gamma(1 + \delta)(2 + \delta)}{(1 - \delta)(2 - \delta)(3 - \delta)(4 - \delta)}$$

$$\lambda_r = \frac{\alpha\,\Gamma(1 + \beta)\Gamma(r - 1 - \beta)}{\Gamma(1 - \beta)\Gamma(r + 1 + \beta)} + \frac{\gamma\,\Gamma(1 - \delta)\Gamma(r - 1 + \delta)}{\Gamma(1 + \delta)\Gamma(r + 1 - \delta)} \tag{A.14.2}$$

There is no simple expression for τ_r.

Asymptotic sampling variances

$$
\begin{aligned}
A_{r,s} = {} & \frac{\alpha^2}{(r+1+\beta)(s+1+\beta)(r+s+1+2\beta)} \\
& + \frac{2\alpha\gamma}{(r+1+\beta)(s+1+\beta)(r+1-\delta)(s+1-\delta)(r+s+1+\beta-\delta)} \qquad \text{(A.14.3)} \\
& + \frac{\gamma^2}{(r+1-\delta)(s+1-\delta)(r+s+1-2\delta)}, \qquad \delta < \tfrac{1}{2}
\end{aligned}
$$

Elements of B, Λ and T can be calculated from (A.14.3).

Parameter estimation

The parameters of the Wakeby distribution may be expressed in terms of the PWMs α_r, as follows.

(i) If ξ is unknown, let

$$
\begin{aligned}
N_1 &= \alpha_0 - 24\alpha_1 + 81\alpha_2 - 64\alpha_3, & C_1 &= 8\alpha_1 - 81\alpha_2 + 192\alpha_3 - 125\alpha_4, \\
N_2 &= \alpha_0 - 12\alpha_1 + 27\alpha_2 - 16\alpha_3, & C_2 &= 4\alpha_1 - 27\alpha_2 + 48\alpha_3 - 25\alpha_4, \\
N_3 &= \alpha_0 - 6\alpha_1 + 9\alpha_2 - 4\alpha_3, & C_3 &= 2\alpha_1 - 9\alpha_2 + 12\alpha_3 - 5\alpha_4.
\end{aligned}
$$

Then β and $-\delta$ are the roots of the quadratic equation

$$(N_2C_3 - N_3C_2)z^2 + (N_1C_3 - N_3C_1)z + (N_1C_2 - N_2C_1) = 0. \qquad (*)$$

β being the larger of the two roots, and the other parameters are then given by

$$
\begin{aligned}
\gamma &= -(1-\delta)(2-\delta)(3-\delta)\{(4N_2 - N_1) + (4N_3 - N_2)\beta\}/\{6(\beta+\delta)\}, \\
\alpha &= (1+\beta)(2+\beta)(3+\beta)\{(4N_2 - N_1) - (4N_3 - N_2)\delta\}/\{6(\beta+\delta)\}, \\
\xi &= \alpha_0 - \alpha/(1+\beta) - \gamma/(1-\delta).
\end{aligned}
$$

(ii) If ξ is known, assume without loss of generality that $\xi = 0$ and let

$$
\begin{aligned}
N_1 &= \alpha_0 - 16\alpha_1 + 27\alpha_2, & C_1 &= 4\alpha_1 - 27\alpha_2 + 32\alpha_3, \\
N_2 &= \alpha_0 - 8\alpha_1 + 9\alpha_2, & C_2 &= 2\alpha_1 - 9\alpha_2 + 8\alpha_3, \\
N_3 &= \alpha_0 - 4\alpha_1 + 3\alpha_2, & C_3 &= \alpha_1 - 3\alpha_2 + 2\alpha_3.
\end{aligned}
$$

Then β and $-\delta$ are the roots of (*), β being the larger of the two roots, and the other parameters are given by

$$\gamma = -(1-\delta)(2-\delta)\{(3N_2 - N_1) + (3N_3 - N_2)\beta\}/\{2(\beta+\delta)\},$$
$$\alpha = (1+\beta)(2+\beta)\{(3N_2 - N_1) - (3N_3 - N_2)\delta\}/\{2(\beta+\delta)\}.$$

These algorithms were devised by Landwehr *et al.* (1979b), who also investigated the finite-sample properties of the PWM parameter and quantile estimators. Landwehr *et al.* recommend the use of the biased estimators $\tilde{\alpha}_r[-0.35, 0]$. Asymptotic variances of the estimators can in principle be calculated from the A_{rs} given above, but the calculations are rather complicated.

A.15. Bernoulli

Definition

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } q = 1 - p \end{cases}$$

$$F(x) = \begin{cases} q, & 0 \le x < 1 \\ 1, & x = 1 \end{cases}$$

$$x(F) = \begin{cases} 0, & 0 \le F < q \\ \frac{1}{2}, & F = q \\ 1, & q < F \le 1 \end{cases}$$

PWMs, etc.

$$r\alpha_{r-1} = p^r$$
$$r\beta_{r-1} = 1 - q^r$$

$$\lambda_1 = p$$
$$\lambda_2 = p(1 - p)$$
$$\lambda_3 = p(1 - p)(1 - 2p)$$
$$\lambda_4 = p(1 - p)(1 - 5p + 5p^2)$$

$$\tau_3 = 1 - 2p$$
$$\tau_4 = 1 - 5p + 5p^2$$

We have $t_4 = (5\tau_3^2 - 1)/4$, so the Bernoulli distribution attains the lower bound for values of τ_4 given τ_3. This parallels the distribution's property of attaining the lower bound for values of kurtosis (κ) given skewness (γ), *i.e.* $\kappa = \gamma^2 + 1$.

Sampling variances and parameter estimation are of little interest for this distribution and are omitted.

A.16. Exponential mixture

Definition

$$X \sim \begin{cases} \mathcal{E}(\alpha) \text{ with probability } p \\ \mathcal{E}(\beta) \text{ with probability } q = 1 - p \end{cases}$$

where $\mathcal{E}(\mu)$ denotes the exponential distribution with mean μ and lower bound zero.

$$f(x) = p\alpha^{-1} e^{-x/\alpha} + q\beta^{-1} e^{-x/\beta}, \quad 0 \le x < \infty$$

$$F(x) = 1 - pe^{-x/\alpha} - qe^{-x/\beta}$$

$x(F)$ has no explicit analytical form

PWMs, etc.

$$\alpha_0 = p\alpha + q\beta$$

$$\alpha_1 = \tfrac{1}{2} p^2 \alpha + 2pq\alpha\beta/(\alpha + \beta) + \tfrac{1}{2} q^2 \beta$$

$$\alpha_2 = \tfrac{1}{3} p^3 \alpha + 3p^2 q\alpha\beta/(\alpha + 2\beta) + 3pq^2 \alpha\beta/(2\alpha + \beta) + \tfrac{1}{3} q^3 \beta$$

$$\alpha_3 = \tfrac{1}{4} p^4 \alpha + 4p^3 q\alpha\beta/(\alpha + 3\beta) + 3p^2 q^2 \alpha\beta/(\alpha + \beta) + 4pq^3 \alpha\beta/(3\alpha + \beta) + \tfrac{1}{4} q^4 \beta$$

$$r\alpha_{r-1} = \sum_{k=0}^{r} \binom{r}{k} \frac{p^{r-k} q^k \alpha\beta}{k\alpha + (r - k)\beta} \qquad \text{(A.16.1)}$$

λ_r and τ_r can be calculated from the α_r but have no simple general form.

Asymptotic sampling variances

$$A_{rs} = K_{r,s+1} + K_{s,r+1} - K_{r+1,s+1} - K_{s+1,r+1}$$

where

$$K_{r,s} = \alpha^2\beta^2 \sum_{j=0}^{r} \sum_{k=0}^{s} \binom{r}{k} \binom{s}{k} \frac{p^{j+k} q^{r+s-j-k}}{\{s\alpha + k(\beta - \alpha)\}\{(r+s)\alpha + (j+k)(\beta - \alpha)\}} \qquad \text{(A.16.2)}$$

It is possible that this expression may simplify in general, or in particular cases. B, Λ and T can be calculated from the $A_{r,s}$.

Parameter estimation

The parameters of the distribution may be expressed in terms of the PWMs α_r as follows.

(A.16.3) Let $S = \alpha + \beta$, $P = \alpha\beta$. Then S satisfies the quadratic equation

$$(18\alpha_2 - 8\alpha_1)S^2 + (2\alpha_0^2 - 12\alpha_0\alpha_1 + 36\alpha_1\alpha_2)S + (\alpha_0^3 - 9\alpha_0^2\alpha_2) = 0$$

and P is given by

$$P = 4\alpha_1 S - \alpha_0^2 .$$

The parameters α and β are the pair of values

$$\tfrac{1}{2}(S \pm \sqrt{S^2 - 4P}\,).$$

and p and q are given by

$$p = \frac{(\alpha_0 - \beta)}{(\alpha - \beta)}, \quad q = \frac{(\alpha - \alpha_0)}{(\alpha - \beta)}$$

The set of parameters thus obtained is feasible provided that p, q, α and β are all real and positive. The parameter pairs (p, α) and (q, β) are interchangeable and so cannot be individually identified. It is conceptually possible that the roots of the quadratic equation for S may yield two, one or no feasible parameter sets. The algorithm's properties have not yet been investigated so we cannot say how frequently each possibility arises, nor how well this estimation algorithm performs relative to others which have been proposed.

A special case

The case $\alpha = 0$ is of some special interest since it yields distributions with arbitrarily high skewness and L-skewness arbitrarily close to 1, but only an exponential tail, *i.e.* $1 - F(x) = O(e^{-\lambda x})$ as $n \to \infty$, for some $\lambda > 0$. The PWMs, *etc.*, are then given by

$$\alpha_0 = 1 - p,$$
$$2\alpha_1 = \tfrac{1}{2}(1-p)^2,$$
$$3\alpha_2 = \tfrac{1}{3}(1-p)^3,$$
$$4\alpha_3 = \tfrac{1}{4}(1-p)^4,$$

$$\lambda_1 = 1 - p,$$
$$\lambda_2 = \tfrac{1}{2}(1-p^2),$$
$$\lambda_3 = \tfrac{1}{6}(1-p)(1+p+4p^2),$$
$$\lambda_4 = \tfrac{1}{12}(1-p)(1+p-5p^2+15p^3),$$

$$\tau_3 = \frac{1+p+4p^2}{3(1+p)} \quad (\to 1 \text{ as } p \to 1),$$
$$\tau_4 = \frac{1+p-5p^2+15p^3}{6(1+p)}.$$

For comparison, the first three conventional moments are

$$EX = 1 - p,$$
$$\operatorname{var} X = 1 - p^2,$$
$$\text{skewness }(X) = 2(1+p+p^2)(1-p)^{-1/2}(1+p)^{-3/2}.$$

APPENDIX B. PROOFS OF NUMBERED RESULTS

(2.1) If $E|X|^p < \infty$ then since $|\{F(x)\}^r\{1 - F(x)\}^s| \le 1$ for all x and for all $r, s \ge 0$, the existence and finiteness of $M_{p,r,s}$ follows. The converse is trivial since $EX^p = M_{1,0,0}$. Note, however, that $M_{p,r,s}$ can exist for some $r, s \ge 0$ with $r + s > 0$ even if $E|X|^p$ does not exist: for example the Cauchy distribution, with density $f(x) = \{\pi(1 + x^2)\}^{-1}$, has $M_{1,1,1} = \frac{1}{6} EX_{2:3} = 0$.

(2.2) This is a well-known result in the theory of order statistics (see, for example, David, 1981, section 3.1).

(2.3),(2.4) These follow trivially from the definition of $M_{p,r,s}$.

(2.5) We reproduce, with minimal modifications, Chan's (1967) result that a distribution may be characterized by either of the sets $\{EX_{1:r}; r = 1, 2, \dots\}$ or $\{EX_{r:r}; r = 1, 2, \dots\}$.

That "$F = G$" implies "$\beta_r^{[X]} = \beta_r^{[Y]}$, $r = 0, 1, \dots$" is trivial. To prove the converse, suppose that $\beta_r^{[X]} = \beta_r^{[Y]}$, $r = 0, 1, \dots$, and let $x(u)$ and $y(u)$, $0 \le u \le 1$, be the quantile functions of X and Y respectively. Since for every real t and every positive integer s

$$\left| x(u) \sum_{r=0}^{s} \frac{(iut)^r}{r!} \right| \le \sum_{r=0}^{\infty} \left| x(u) \frac{(ut)^r}{r!} \right| \le e^{|t|} |x(u)|,$$

and

$$\int_0^1 e^{|t|} |x(u)| \, du = e^{|t|} E|X| < \infty,$$

by the Lebesgue dominated convergence theorem we have

$$\sum_{r=0}^{\infty} \frac{(it)^r}{r!} \int_0^1 x(u)\, u^r\, du = \int_0^1 x(u)\, e^{iut}\, du.$$

Similarly

$$\sum_{r=0}^{\infty} \frac{(it)^r}{r!} \int_0^1 y(u)\, u^r\, du = \int_0^1 y(u)\, e^{iut}\, du.$$

Then by assumption for every real number t

$$\int_0^1 x(u)\, e^{iut}\, du = \int_0^1 y(u)\, e^{iut}\, du.$$

By the uniqueness of the Fourier transform this implies that $x(u) = y(u)$ almost everywhere. The fact that both F and G are nondecreasing and continuous from the right implies that both $x(u)$ and $y(u)$ are nondecreasing and continuous from the left; hence $x(u) = y(u)$ for every $u \in [0, 1]$. Therefore we conclude that F and G are identical. The case where $\alpha_r^{[X]} = \alpha_r^{[Y]}$, $r = 0, 1, \ldots$, can be similarly proved.

(2.6) By expanding $F^r(1 - F)^s$ in (2.4) in powers of F or of $(1 - F)$ by the binomial theorem, $M_{p,r,s}$ may be expressed as a sum of the quantities $M_{p,k,0}$ or $M_{p,0,k}$, $k = 1, 2, \ldots, r + s$; (2.6) gives two special cases of this result.

(2.7) Because $1 - F(x) \sim Ae^{-Bx}$ as $x \to \infty$ we know that given any $\epsilon > 0$

$$\exists\, x_0 \text{ such that } x \geq x_0 \Rightarrow (1 - \epsilon)Ae^{-Bx} \leq 1 - F(x) \leq (1 + \epsilon)Ae^{-Bx}.$$

whence

$$\exists F_0,\ 0 \le F_0 < 1, \text{ such that } F \ge F_0 \Rightarrow \frac{1}{B}\log(1-\varepsilon) \le x(F) - y(F) \le \frac{1}{B}\log(1+\varepsilon)$$

where

$$y(F) = -B^{-1}\log\{(1-F)/A\};$$

indeed, $F(x_0)$ is such an F_0. Thus taking $\varepsilon = \frac{1}{2}$ for example we have that

$$\exists F_0,\ 0 \le F_0 < 1, \text{ such that } F \ge F_0 \Rightarrow |x(F) - y(F)| \le B^{-1}\log 2.$$

Now choose $r_0 > 1$ such that $r_0 F_0^{r_0 - 1} \le 1$: this can always be done, since $0 \le F_0 < 1$. For $r \ge r_0$ we have

$$\begin{aligned}
\left| r\beta_{r-1} - \int_0^1 y(F) rF^{r-1}\,dF \right| &= \left| \int_0^1 x(F) rF^{r-1}\,dF - \int_0^1 y(F) rF^{r-1}\,dF \right| \\
&\le \left| \int_0^{F_0} x(F) rF^{r-1}\,dF \right| + \left| \int_0^{F_0} y(F) rF^{r-1}\,dF \right| \\
&\quad + \int_{F_0}^1 |\{x(F) - y(F)\}|\, rF^{r-1}\,dF \\
&\le \int_0^{F_0} |x(F)|\,dF + \int_0^{F_0} |y(F)|\,dF + B^{-1}\log 2 \int_{F_0}^1 rF^{r-1}\,dF \\
&\le E|X| + E|Y| + B^{-1}\log 2
\end{aligned}$$

where Y is a random variable with quantile function $y(F)$, *i.e.* Y has an exponential distribution with lower bound $B^{-1}\log A$ and mean B^{-1}. Now $\int y(F) rF^{r-1}\,dF$ is the "$r\beta_{r-1}$" of the random variable Y, and this is (from Appendix A.2)

$$B^{-1}\left(\log A + 1 + \tfrac{1}{2} + \tfrac{1}{3} + \cdots + \tfrac{1}{r}\right) \sim B^{-1}\log r \quad \text{as } r \to \infty.$$

Thus $r\beta_{r-1}$, whose difference from $\int y(F) r F^{r-1} dF$ is for $r \geq r_0$ bounded by a constant independent of r, itself satisfies $r\beta_{r-1} \sim B^{-1} \log r$. This is the required result.

Note that the foregoing method of proof remains valid if the assumption $1 - F(x) \sim Ae^{-Bx}$ is weakened to "for all sufficiently large x, $\{1 - F(x)\}e^{Bx}$ is bounded away from 0 and $+\infty$". Other weakenings of the assumptions might be possible. It might also be possible to prove the converse (or a partial converse) of (2.7).

(2.8) Proved similarly to (2.7); the approximating $y(F)$ is a generalized Pareto rather than an exponential quantile function.

(3.1) We have

$$\lambda_{r+1} = (r+1)^{-1} \sum_{j=0}^{r} (-1)^j \binom{r}{j} EX_{r+1-j:r+1} .$$

Now by (2.2)

$$EX_{r+1-j:r+1} = \frac{(r+1)!}{j!(r-j)!} \cdot M_{1,r-j,j} ,$$

and

$$\begin{aligned} M_{1,r-j,j} &= E[X\{F(X)\}^{r-j}\{1 - F(x)\}^j] \\ &= \sum_{i=0}^{j} (-1)^i \binom{j}{i} \beta_{r-j+i} \end{aligned}$$

by binomial expansion of $\{1 - F(x)\}^j$; thus

$$\lambda_{r+1} = \sum_{j=0}^{r} (-1)^j \binom{r}{j}^2 \sum_{i=0}^{j} (-1)^i \binom{j}{i} \beta_{r-j+i},$$

in which the coefficient of β_k is

$$\begin{aligned}
&\sum_{j=0}^{r} (-1)^j \binom{r}{j}^2 (-1)^{j+k-r} \binom{j}{j+k-r} \\
&= (-1)^{r-k} \binom{r}{k} \sum_{j=0}^{r} \binom{r}{j} \binom{k}{j} \\
&= (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} \\
&= p^*_{r,k}.
\end{aligned}$$

The expression for λ_{r+1} as a sum of the α_k may be proved similarly.

(3.2) Substitute $\beta_k = \int x(F) F^k dF$ from (2.4) into (3.1).

(3.3) Proof as for (2.1).

(3.4) The relationship (3.1), which gives λ_{r+1} in terms of β_k, $k = 0, 1, \ldots, r$, can clearly be inverted to give β_r in terms of λ_k, $k = 1, 2, \ldots, r+1$. Thus the sets $\{\beta_r: r = 0, 1, \ldots\}$ and $\{\lambda_r: r = 1, 2, \ldots\}$ are in 1-1 correspondence and the characterization theorem (2.5) for the former set extends to the latter.

(3.5) The Jacobi polynomials $P_r^{(\alpha,\beta)}(t)$ are defined by

$$(1-x)^\alpha (1+x)^\beta P_r^{(\alpha,\beta)}(t) = \frac{(-1)^r}{2^r r!} \frac{d^r}{dt^r} \{(1-x)^{r+\alpha}(1+x)^{r+\beta}\}, \quad r = 0, 1, \ldots.$$

(Szego, 1959, chap. 4). Let $Q_r(t) = P_r^{(1,1)}(2t-1)$. Then

$$t(1-t)Q_r(t) = \frac{(-1)^r}{r!}\frac{d^r}{dt^r}\{t(1-t)\}^{r+1}$$

and so, using (C.2) from Appendix C,

$$\frac{d}{dt}\{t(1-t)Q_r(t)\} = -(r+1)P^*_{r+1}(t).$$

From (3.2) we have

$$\lambda_r = \int x\, P^*_{r-1}(F(x))\, dF(x),$$

and integrating by parts we obtain

$$\lambda_r = \Big[-xF(x)\{1-F(x)\}(r-1)^{-1}Q_{r-2}(F(x)) \Big] + \int F(x)\{1-F(x)\}(r-1)^{-1}Q_{r-2}(F(x))\, dx\,.$$

The integrated term vanishes, for the finiteness of $E|X|$ ensures that $xF(x)\{1-F(x)\} \to 0$ as x approaches the endpoints of the range of integration: thus

$$\lambda_r = \int F(x)\,\{1-F(x)\}\,(r-1)^{-1}Q_{r-2}(F(x))\, dx\,.$$

Since $Q_0(t) = 1$ we have in particular

$$\lambda_2 = \int F(x)\{1-F(x)\}\, dx\,.$$

We have $0 \le F(x) \le 1$ for all x, and because X is nondegenerate there exists a set of nonzero measure on which $0 < F(x) < 1$: thus $\lambda_2 > 0$. Since $F(x)\{1 - F(x)\} \ge 0$ for all x we also have

$$|\lambda_r| \le (r-1)^{-1} \sup_{0 \le t \le 1} |Q_{r-2}(t)| \int F(x)\{1 - F(x)\} dx$$
$$= (r-1)^{-1} \sup_{0 \le t \le 1} |Q_{r-2}(t)| \ \lambda_2 .$$

From Szego (1959, p. 166) we have

$$\sup_{0 \le t \le 1} |Q_r(t)| = r + 1$$

with the supremum being attained only at $t = 0$ or $t = 1$. Thus $|\lambda_4| \le \lambda_2$, with equality only if $F(x)$ can take only the values 0 and 1, *i.e.* only if X is degenerate. Thus for a nondegenerate distribution we have $|\lambda_4| < \lambda_2$, which together with $\lambda_2 > 0$ implies $|\tau_r| < 1$.

Note that since for r even we have

$$\inf_{0 \le t \le 1} (r+1)^{-1} Q_r(t) > -1,$$

the inequality $\tau_r > -1$ can be strengthened when r is even: (3.9) is a case in point.

(3.6) Because $X \ge 0$ almost surely we have $\lambda_1 = EX > 0$ as well as $\lambda_2 > 0$, so $\tau = \lambda_2 / \lambda_1 > 0$; furthermore $EX_{1:2} > 0$, so

$$\tau - 1 = (\lambda_2 - \lambda_1)/\lambda_1 = -2\alpha_1/\lambda_1 = -EX_{1:2}/\lambda_1 < 0 .$$

(3.7) The first equation is just the definition of λ_2. For the second equation we have

$$\tfrac{1}{2} E(X_{2:2} - X_{1:2})^2 = \iint_{x<y} (y-x)^2 dF(x)dF(y)$$
$$= \tfrac{1}{2} \iint_{x\ y} (y-x)^2 dF(x)dF(y)$$
$$= \int y^2 dF(y). \int dF(x) - \int ydF(y). \int xdF(x)$$
$$= EX^2 - (EX)^2$$
$$= \operatorname{var} X.$$

(3.8) From (B.3) below we obtain

$$EX_{3:3} - EX_{1:3} = \tfrac{3}{2}(EX_{2:2} - EX_{1:2}),$$

which together with the definition of τ_3 yields (3.8).

(3.9) Let $\mu_{r:n} = EX_{r:n}$. We use the result

$$\mu_{r+1:n} - \mu_{r:n} = \binom{n}{r} \int_{-\infty}^{\infty} \{F(x)\}^r \{1 - F(x)\}^{n-r} dx \tag{B.1}$$

(David, 1981, p. 50). By Cauchy's inequality we have

$$\left[\int_{-\infty}^{\infty} \{F(x)\}^2 \{1 - F(x)\}^2 dx \right]^2$$
$$\le \int_{-\infty}^{\infty} \{F(x)\}^3 \{1 - F(x)\} dx \, . \int_{-\infty}^{\infty} F(x)\{1 - F(x)\}^3 dx \tag{B.2}$$

which, using (B.1), we may write as

$$(\mu_{3:4} - \mu_{2:4})^2 \le 9(\mu_{4:4} - \mu_{3:4})(\mu_{2:4} - \mu_{1:4}).$$

Substituting

$$\mu_{4:4} - \mu_{3:4} = \tfrac{6}{5}\,\lambda_2 + 2\lambda_3 + \tfrac{4}{5}\,\lambda_4,$$

$$\mu_{3:4} - \mu_{2:4} = \tfrac{6}{5}\,(\lambda_2 - \lambda_4),$$

$$\mu_{2:4} - \mu_{1:4} = \tfrac{6}{5}\,\lambda_2 - 2\lambda_3 + \tfrac{4}{5}\,\lambda_4,$$

which may easily be verified from the result

$$n\mu_{r:n-1} = (n - r)\mu_{r:n} + r\mu_{r+1:n} \tag{B.3}$$

(David, 1981, p. 46) and the definition of the λ_r, we obtain

$$4\lambda_4\lambda_2 - 5\lambda_3^2 + \lambda_2^2 \geq 0. \tag{B.4}$$

Dividing by λ_2^2, which as in the proof of (3.5) we can show to be strictly positive, we obtain

$$(5\tau_3^2 - 1)/4 \leq \tau_4.$$

Since $\tau_4 < 1$ by (3.5), we have shown that all possible (τ_3, τ_4) pairs must lie in $\mathscr{P}$.

To show that each point in $\mathscr{P}$ is the (τ_3, τ_4) pair of some distribution, we need consider only distributions whose probability mass is concentrated at two or three points. The L-moment ratios of the Bernoulli random variables X_p, $0 < p < 1$, defined by

$$P(X_p = 0) = p, \quad P(X_p = 1) = 1 - p,$$

occupy the lower boundary of $\mathscr{P}$ (see Appendix A.15), and the distribution of a random variable Y defined by

$$P(Y = 0) = p_1, \quad P(Y = x_0) = p_2 - p_1, \quad P(Y = 1) = 1 - p_2,$$
$$0 < x_0 < 1, \quad 0 < p_1 < p_2 < 1.$$

accounts, as x_0 varies with p_1 and p_2 fixed, for all (τ_3, τ_4) values on the line segment in (τ_3, τ_4)-space which joins the (τ_3, τ_4) values of the random variables X_{p_1} and X_{p_2}. All points in the set $\mathscr{P}$ are thereby attainable.

Note that the Bernoulli random variables X_p, $0 \le p \le 1$, and linear transformations thereof, are the only ones which attain the bound (B.4). This follows from the criterion for equality in (B.2), *viz.* that there exists a constant k for which

$$k\,\{F(x)\}^{3/2}\,\{1 - F(x)\}^{1/2} = \{\dot{F}(x)\}^{1/2}\,\{1 - F(x)\}^{3/2} \quad \text{for all } x\,;$$

this implies that $F(x)$ can take only the values 0, $(1 + k)^{-1}$ or 1 and thus that X can take at most two distinct values.

(3.10) The following proof is a more rigorous version of Sillitto's, and is a standard method of approximating a function as a weighted sum of orthogonal functions. We seek an approximation to the quantile function $x(F)$ of the form

$$x(F) \approx \sum_{r=1}^{s} a_r P^{*}_{r-1}(F)\,, \quad 0 < F < 1\,. \tag{B.5}$$

The shifted Legendre polynomials $P^{*}_{r-1}(F)$ are a natural choice as the basis of the approximation because they are orthogonal on $0 < F < 1$ with unit weight function – see Appendix C. To determine the a_r in (B.5) we denote the error of the approximation (B.5) by

$$R_s(F) = x(F) - \sum_{r=1}^{s} a_r P^{*}_{r-1}(F)$$

and seek to minimize the mean square error $\int_0^1 \{R_s(F)\}^2\,dF$. The condition that X has finite variance ensures that the mean square error is finite. We have

$$\int_0^1 \{R_s(F)\}^2 dF = \int_0^1 \{x(F)\}^2 dF - 2\sum_{r=1}^{s} a_r \int_0^1 x(F) P^*_{r-1}(F) dF$$

$$+ \sum_{r=1}^{s} \sum_{t=1}^{s} a_r a_t \int_0^1 P^*_{r-1}(F)\, P^*_{t-1}(F)\, dF$$

$$= \int_0^1 \{x(F)\}^2 dF - 2\sum_{r=1}^{s} a_r \lambda_r + \sum_{r=1}^{s} a_r^2/(2r-1)$$

(where we have used (C.1) and (C.4) from Appendix C), which is minimized by choosing

$$a_r = (2r-1)\lambda_r.$$

That

$$\int_0^1 \{R_s(F)\}^2 dF \to 0 \quad \text{as } s \to \infty,$$

i.e. that the set of orthogonal functions $P^*_{r-1}(F)$ is complete, is a standard result in Sturm-Liouville theory. A proof is given by Titchmarsh (1946, chap. 4).

Note: our use of the symbols λ_r to denote the L-moments of a distribution follows Sillitto (1969), who defined λ_r in the course of proving (3.10).

(3.11) In the notation of the proof of (3.10), we have

$$0 = \lim_{s\to\infty} \int_0^1 \{R_s(F)\}^2 dF = \int_0^1 \{x(F)\}^2 dF - \sum_{r=1}^{\infty} (2r-1)\lambda_r^2;$$

since

$$\operatorname{var} X = EX^2 - (EX)^2 = \int_0^1 \{x(F)\}^2 dF - \lambda_1^2$$

the result follows.

(4.1) Unbiasedness of a_r was proved by Landwehr *et al.* (1979a). A quicker proof uses the result

$$EX_{1:m} = \binom{n}{m}^{-1} \sum_{i=0}^{n-1} \binom{n-1-i}{m-1} EX_{i+1:n}$$

(Sillitto, 1964), whence

$$\begin{aligned}\alpha_r &= (r+1)^{-1} EX_{1:r+1} \\ &= (r+1)^{-1} \binom{n}{r+1}^{-1} \sum_{i=1}^{n} \binom{n-i}{r} EX_{i:n} \\ &= n^{-1} \binom{n-1}{r}^{-1} \sum_{i=1}^{n} \binom{n-i}{r} EX_{i:n}\end{aligned}$$

which is clearly unbiasedly estimated by a_r. Unbiasedness of b_r follows from a similar expression for $EX_{m:m}$ in terms of $EX_{i:n}$, $1 \le i \le n$.

(4.2) From the definition of b_r we have

$$\sum_{k=0}^{r} (-1)^k \binom{r}{k} b_k = \sum_{k=0}^{r} (-1)^k \binom{r}{k} n^{-1} \sum_{i=1}^{n} \binom{i-1}{k} x_i / \binom{n-1}{k},$$

in which the coefficient of x_i is

$$n^{-1}\sum_{k=0}^{r}(-1)^k\binom{r}{k}\frac{(i-1)(i-2)\ldots(i-k)}{(n-1)(n-2)\ldots(n-k)}$$
$$= n^{-1}\,{}_2F_1(-r, 1-i; 1-n; 1)$$
$$= n^{-1}\frac{(i-n)(i-n+1)\ldots(i-n+r)}{(1-n)(2-n)\ldots(r-n)}$$

which is also the coefficient of x_i in the definition of a_r. The second equality of (4.2) is proved similarly.

(4.3) To show that the two expressions for ℓ_{r+1} are equivalent we write

$$(-1)^r\sum_{k=0}^{r}\dot{p}_{r,k}\,a_k = (-1)^r\sum_{k=0}^{r}\dot{p}_{r,k}\sum_{j=0}^{k}(-1)^k\binom{k}{j}b_j$$

by (4.2), in which the coefficient of b_j is

$$\sum_{k=j}^{r}(-1)^{r+j}\dot{p}_{r,k}\binom{k}{j} = \sum_{k=j}^{r}(-1)^{j-k}\binom{k}{j}\binom{r}{k}\binom{r+k}{k}$$
$$= \sum_{k=0}^{r-j}(-1)^k\binom{k+j}{j}\binom{r}{k+j}\binom{r+k+j}{k+j}$$
$$= \frac{(r+j)!}{(j!)^2(r-j)!}\sum_{k=0}^{r-j}(-1)^k\binom{r-j}{k}\frac{(r+k+j)!}{(r+j)!}\frac{j!}{(j+k)!}$$
$$= \frac{(r+j)!}{(j!)^2(r-j)!}\,{}_2F_1(j-r, r+j+1; j+1; 1)$$
$$= \frac{(-1)^{r-j}(r+j)!}{(j!)^2(r-j)!}$$
$$= \dot{p}_{r,j}.$$

(4.4) Follows immediately from (3.1), (4.1) and (4.3).

(4.5) When the right side of (4.5) is written out in full, the number of occurrences of x_i is the number of size-r subsamples of which x_i is the smallest member, *viz.* $\binom{n-i}{r-1}$. Thus when the right side of (4.5) is written as a weighted sum of the x_i, the coefficient of x_i is

$$\binom{n-i}{r-1} \Big/ \binom{n}{r} = \frac{r}{n} \binom{n-i}{r-i} \Big/ \binom{n-1}{r-1}.$$

which from the definition of a_r is also the coefficient of x_i in ra_{r-1}. This proves the result.

(4.6),(4.7) Proved similarly to (4.5). Note that many other similar results may be derived: for example

$$(3b_2 - b_0) - (2b_1 - b_0) = \frac{1}{3} \binom{n}{3}^{-1} \sum_{i > j > k} \sum \sum (x_j - x_k).$$

whence

$$3b_2 - b_0 > 2b_1 - b_0$$

and so

$$(3b_2 - b_0)/(2b_1 - b_0) > 1.$$

with consequences for PWM-based estimators of parameters of the generalized extreme-value distribution (Hosking *et al.*, 1985, Appendix B).

(4.8) Proof straightforward.

(4.9) When $EX^2 < \infty$, (4.9) follows from the asymptotic distribution theory presented in section 5. A proof of consistency should also be possible when $E|X|$ exists but EX^2 does not.

(4.10) We have

$$\tilde{\beta}_r = n^{-1}\sum_{i=1}^{n}\left(\frac{i+\gamma}{n+\delta}\right)^r x_i .$$

Write $(i+\gamma)^r$ as a sum of factorial powers of $(i-1)$, say

$$(i+\gamma)^r = \sum_{k=0}^{r} g_{rk}(\gamma)(i-1)(i-2)\ldots(i-r+k)\,;$$

then

$$\tilde{\beta}_r = \sum_{k=0}^{r} g_{rk}(\gamma)(n-1)(n-2)\ldots(n-r+k)b_{r-k}/(n+\delta)^r; \qquad \text{(B.6)}$$

for example

$$\tilde{\beta}_0 = b_0\,,$$

$$\tilde{\beta}_1 = \frac{(n-1)}{(n+\delta)}b_1 + \frac{(1+\gamma)}{(n+\delta)}b_0\,,$$

$$\tilde{\beta}_2 = \frac{(n-1)(n-2)}{(n+\delta)^2}b_2 + \frac{(3+2\gamma)(n-1)}{(n+\delta)^2}b_1 + \frac{(1+\gamma)^2}{(n+\delta)^2}b_0\,.$$

Thus

$$E\tilde{\beta}_r = \sum_{k=0}^{r} g_{rk}(\gamma)(n-1)(n-2)\ldots(n-r+k)\beta_{r-k}/(n+\delta)^r\,.$$

whence the bias of $\tilde{\beta}_r$ may be found. As $n \to \infty$ the asymptotic bias can be found by noting that

$$\frac{(n-1)(n-2)\ldots(n-r+k)}{(n+\delta)^r} = \begin{cases} 1 - \{\frac{1}{2}r(r+1) + r\delta\}n^{-1} + O(n^{-2}) & \text{if } k = 0, \\ n^{-1} + O(n^{-2}) & \text{if } k = 1, \\ O(n^{-k}) & \text{if } k \geq 2, \end{cases}$$

and (proof by induction) that

$$g_{r0}(\gamma) = 1, \quad g_{r1}(\gamma) = \tfrac{1}{2}r(r+1) + r\gamma;$$

(4.10) follows.

(4.11) Proof straightforward: expand $(1 - p_{i:n})^r$ by binomial theorem, *etc.*

(4.12) Proof straightforward.

(4.13) Follows from the consistency of ℓ_r and $\tilde{\lambda}_r$ as estimators of λ_r, and the continuity of $\tau_r = \lambda_r/\lambda_2$ as a function of λ_2 and λ_r (note that $\lambda_2 > 0$).

(4.14),(4.15) The asymptotic distribution theory of section 5 gives, when $EX^2 < \infty$,

$$\ell_r = \lambda_r + O_p(n^{-1/2}).$$

Making a Taylor-series expansion of t_r about τ_r (as in Rao, 1973, p. 388) we obtain

$$E(t_r - \tau_r) \sim \lambda_2^{-2}\{\tau_r \operatorname{var} \ell_2 - \operatorname{cov}(\ell_2, \ell_r)\}.$$

The results then follow by substituting the relevant values for each distribution from Appendix A, sections A.1 through A.6.

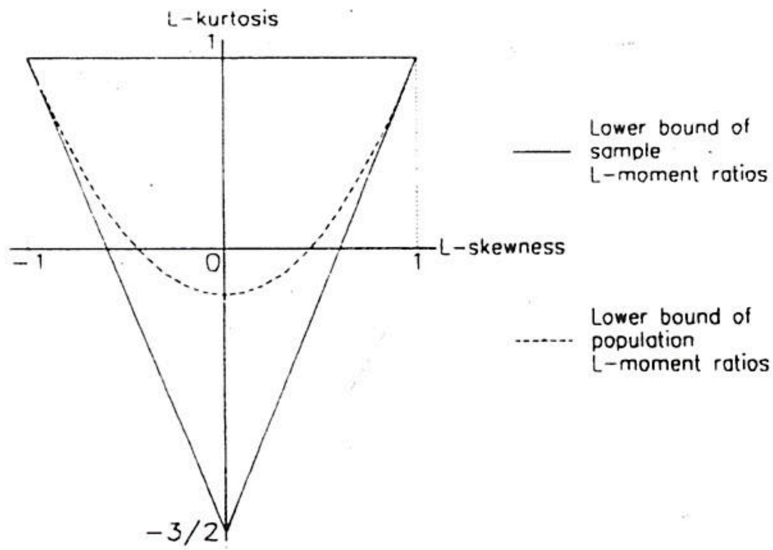

Figure B.1. *L*-moment ratios from a sample of size four.

(4.16) We need only consider samples whose numbers take one of the four values 0, α, $1-\beta$ and 1, where $0 \le \alpha + \beta \le 1$. For example, for a sample of size 4 with $x_1 = 0$, $x_2 = \alpha$, $x_3 = 1 - \beta$, $x_4 = 1$, we have

$$\ell_2 = (4 - \alpha - \beta)/12, \quad \ell_3 = (\beta - \alpha)/4, \quad \ell_4 = (3\alpha + 3\beta - 2)/4,$$

and it is straightforward to show that t_3 and t_4 can take any values such that

$$|t_3| \le 1, \quad t_4 \ge (5|t_3| - 3)/2;$$

the set of possible (t_3, t_4) values includes the set $\mathscr{P}$ of possible (τ_3, τ_4) values – see Figure B.1. Proofs for other sample sizes are similar.

(5.1) From the *U*-statistic representation of ℓ_2 we have

$$\ell_2^2 = \frac{1}{n^2(n-1)^2} \sum_{i>j} \sum_{k>l} (x_i - x_j)(x_k - x_l).$$

By breaking up the summations in this equation according to the ordering of i, j, k, ℓ, and taking expectations, Kendall and Stuart (1979, section 10.14), following Lomnicki (1952), show that

$$E\ell_2^2 = \{n(n-1)\}^{-1}\{\mathrm{var}\, X + (n-2)J + (n-2)(n-3)\lambda_2^2\}$$

and thence that

$$\mathrm{var}\, \ell_2 = \{n(n-1)\}^{-1}\{\mathrm{var}\, X + (n-2)J - 2(2n-3)\lambda_2^2\}$$

where

$$\begin{aligned} J = \tfrac{1}{3} E\{&(X_{3:3} - X_{2:3})(X_{3:3} - X_{1:3}) + (X_{3:3} - X_{2:3})(X_{2:3} - X_{1:3}) \\ &+ (X_{3:3} - X_{1:3})(X_{2:3} - X_{1:3})\}. \end{aligned}$$

We can rewrite this last equation as

$$\begin{aligned} 3J &= \tfrac{1}{2} E(X_{1:3} + X_{2:3} + X_{3:3})^2 - \tfrac{3}{2} E(X_{1:3}^2 + X_{2:3}^2 + X_{3:3}^2) + 2E(X_{3:3} - X_{1:3})^2 \\ &= \tfrac{3}{2} EX^2 + 3(EX)^2 - \tfrac{9}{2} EX^2 + 2E(X_{3:3} - X_{1:3})^2 \\ &= 2E(X_{3:3} - X_{1:3})^2 - 3\mathrm{var}\, X, \end{aligned}$$

which together with (3.7) yields the required result.

(5.2) $E(X_{s:t} - X_{r:t})^2$ is unbiasedly estimated by the U-statistic

$$\binom{n}{t}^{-1} \sum_{1 \le i_1 < i_2 <} \sum \cdots \sum_{< i_t \le n} (x_{i_s} - x_{i_r})^2,$$

and by counting the occurrences of each squared difference $(x_k - x_j)^2$, $1 \le j < k \le n$, similarly to the proof of (4.5), we obtain

$$\binom{n}{t}^{-1} \sum_{i_1 < i_2 < \cdots < i_t} \sum \cdots \sum (x_{i_s} - x_{i_r})^2$$
$$= \binom{n}{t}^{-1} \sum_{j < k} \sum \binom{j-1}{r-1} \binom{k-j-1}{s-r-1} \binom{n-k}{t-s} (x_k - x_j)^2 .$$

(5.3) We use Stigler's (1974) form of the theorem giving the asymptotic distribution of a linear combination of order statistics, which is as follows. Let

$$S_n = n^{-1} \sum_{i=1}^{n} w\left(\frac{i}{n+1} \right) x_i$$

where $x_1 \leq \cdots \leq x_n$ is an ordered random sample from the distribution of a random variable X with cumulative distribution function $F(x)$ and quantile function $x(F)$, and w is a weight function defined on the interval $(0, 1)$. Suppose that $EX^2 < \infty$ and that w is bounded and continuous almost everywhere. Then as $n \to \infty$,

(a) $n^{1/2}(ES_n - \mu) \to 0$,

(b) $n \operatorname{var} S_n \to \sigma^2$,

(c) $(S_n - ES_n)/(\operatorname{var} S_n)^{1/2} \xrightarrow{\mathcal{D}} N(0,1)$,

where

$$\mu = \int_0^1 w(u)x(u)du ,$$
$$\sigma^2 = 2 \iint_{x<y} w(F(x))w(F(y)) F(x)\{1 - F(y)\} \, dx \, dy$$

(Stigler, 1974, Theorem 6).

This theorem immediately yields the univariate asymptotic distributions of the plotting-position estimators $\tilde{\alpha}_r[0,1]$, $\tilde{\beta}_r[0,1]$ and $\tilde{\lambda}_r[0,1]$; take $w(u)$ equal to $(1-u)^r$, u^r and $P^*_{r-1}(u)$ respectively. The theorem also applies to any finite linear combination of $\tilde{\alpha}_r[0,1]$, $\tilde{\beta}_r[0,1]$ and $\tilde{\lambda}_r[0,1]$, showing that the asymptotic joint distribution of any finite set of $\tilde{\alpha}_r[0,1]$, $\tilde{\beta}_r[0,1]$ and $\tilde{\lambda}_r[0,1]$, is multivariate Normal. The covariances of the limiting multivariate Normal distribution, calculated from

$$\operatorname{cov}(\tilde{\alpha}_r, \tilde{\alpha}_s) = \tfrac{1}{2}\left\{ \operatorname{var}(\tilde{\alpha}_r + \tilde{\alpha}_s) - \operatorname{var}\tilde{\alpha}_r - \operatorname{var}\tilde{\alpha}_s \right\},$$

etc., yield A_{rs}, B_{rs} and Λ_{rs} as defined in (5.3). By expressing a_r as a sum of the $\tilde{\alpha}_s[0,1]$, $0 \le s \le r$, similarly to (B.6), it is straightforward to show that

$$a_r - \tilde{\alpha}_r[0,1] = O_p(n^{-1}),$$

etc., and, therefore, that the asymptotic distributions of the a_r and the $\tilde{\alpha}_r$, of the b_r and the $\tilde{\beta}_r$, and of the l_r and the $\tilde{\lambda}_r$, are identical.

Note that the asymptotic distributions of the a_r, b_r, and l_r can also be obtained by considering these statistics as U-statistics: asymptotic Normality is then an immediate consequence of Theorem 7.1 of Hoeffding (1948). This approach, however, yields the variance of the limiting Normal distribution in a less tractable form than that given above.

(5.4) We use the result that, for $\hat{\theta}$ an estimator of the vector parameter θ and f a vector function differentiable in a neighborhood of θ, if

$$n^{1/2}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N(0, V)$$

then

$$n^{1/2}\{f(\hat{\theta}) - f(\theta)\} \xrightarrow{\mathcal{D}} N(0, GVG^{\mathrm{T}})$$

where $G = \partial f/\partial\theta^{\mathrm{T}}$ (Rao, 1973, p. 388). Taking $\hat{\theta} = (\ell_1\ \ell_2 \ldots \ell_m)^{\mathrm{T}}$ and $\theta = (\lambda_1\ \lambda_2 \ldots \lambda_m)^{\mathrm{T}}$, and treating the τ_r, $r = 3, \ldots, m$, as functions of θ, the result follows.

(5.5) Proof as for (5.3).

(5.6) Assuming without loss of generality that the center of symmetry is at zero, so that

$$F(-x) = 1 - F(x),$$

we have

$$\begin{aligned}
&\iint_{x<y} P^*_{r-1}(F(x))\, P^*_{s-1}(F(y))\, F(x)\{1 - F(y)\}\, dx\, dy \\
&= \iint_{x<y} P^*_{r-1}(F(-y))\, P^*_{s-1}(F(-x))\, F(-y)\{1 - F(-x)\}\, dx\, dy \\
&= \iint_{x<y} P^*_{s-1}(1 - F(x))\, P^*_{r-1}(1 - F(y))\, F(x)\{1 - F(y)\}\, dx\, dy \\
&= (-1)^{r+s-2} \iint_{x<y} P^*_{s-1}(F(x)) P^*_{r-1}(F(y)) F(x)\{1 - F(y)\}\, dx\, dy
\end{aligned}$$

by (C.3). Thus $\Lambda_{rs} = 0$ if $r + s$ is odd. Since for a symmetric distribution we also have $\tau_r = 0$ if r is odd, it follows from the definition of T_{rs} that $T_{rs} = 0$ if $r + s$ is odd.

(5.7) Make the change of variables $u = F(x)$, $v = F(y)$ in the definition of Λ_{rs} in (5.3).

(5.8) Corresponding to (5.7) we have

$$\Lambda_{22} = 2\iint\limits_{u<v} (2u-1)(2v-1)u(1-v)x'(u)x'(v)\,du\,dv$$
$$= \iint\limits_{u<v} 2(1-v)(1-2v)(1-4u)\{x(v)-x(u)\}x'(v)\,du\,dv$$
$$= \iint\limits_{u<v} (1-4u)(3-4v)\{x(v)-x(u)\}^2\,du\,dv,$$

integrating by parts with respect to u and v successively.

(6.1) This is a straightforward consequence of (5.3) and the result of Rao (1973) used in the proof of (5.4).

(7.1) Proved similarly to (5.8).

(7.2) Taylor expansion of u_j about its mean $\zeta_j \equiv j/(n+1)$ gives

$$u_j^r = \zeta_j^r + r\zeta_j^{r-1}(u_j-\zeta_j) + \tfrac{1}{2}r(r-1)\zeta_j^{r-2}(u_j-\zeta_j)^2 + O_p(n^{-3/2}).$$

Averaging this expression over $j = 1, \ldots, n$ gives

$$m_r = r\,n^{-1}\sum_{j=1}^{n}\zeta_j^{r-1}u_j - (r-1)n^{-1}\sum_{j=1}^{n}\zeta_j^r + \tfrac{1}{2}r(r-1)n^{-1}\sum_{j=1}^{n}\zeta_j^{r-2}(u_j-\zeta_j)^2 + O_p(n^{-1/2}).$$

The first term on the right side of this equation is $r\tilde{\beta}_{r-1}[0, 1]$, a plotting-position estimator of $r\beta_{r-1}$. The second term is, as $n \to \infty$, within $O(n^{-1})$ of

$$(r-1)\int_0^1 z^r\,dz = \frac{r-1}{r+1}.$$

The third term is bounded, by

$$\tfrac{1}{2} r(r-1)n^{-1}\sum \zeta_j^{r-2} \to \tfrac{1}{2} r(r-1)\int_0^1 z^{r-2}\,dz = \tfrac{1}{2} r,$$

and has mean

$$\tfrac{1}{2} r(r-1)n^{-1}\sum_{j=1}^{n} \zeta_j^{r-2} \frac{\zeta_j(1-\zeta_j)}{n+2} \sim \frac{r(r-1)}{2(n+2)}\int_0^1 z^{r-1}(1-z)\,dz$$

which is of order n^{-1}; thus this term is of stochastic order $n^{-1/2}$ and we have

$$m_r = r\tilde{\beta}_{r-1}[0,1] + (r-1)/(r+1) + O_p(n^{-1/2}).$$

This expression remains valid if $r\tilde{\beta}_{r-1}[0,1]$ is replaced by rb_{r-1} or any other $r\tilde{\beta}_{r-1}[\gamma,\delta]$, for the differences between these quantities are all of stochastic order $n^{-1/2}$.

(7.3) Let $U_{j:n} = F(X_{j:n};\theta_0)$: the $U_{j:n}$ are order statistics of a sample of size n from the uniform distribution on [0, 1]. Suppressing for convenience the dependence on n of various subscripts, we have

$$d = n^{-1}\sum_j \delta_j \hat{U}_j = n^{-1}\sum_j \delta_j F(G(U_j;\theta_0);\hat{\theta}),$$

which we shall expand in a first-order Taylor series about $U_j = \zeta_j$, where

$$\zeta_j \equiv E\,U_j = j/(n+1).$$

The validity of such an expansion is subject to regularity conditions on the function F and its derivatives, but we shall assume that these are satisfied by the standard families of distributions used in statistics. We denote by $\tilde{\theta}$ the estimator of θ obtained when $U_j = \zeta_j$; *i.e.* $\tilde{\theta}$ is defined by $\tilde{g}_r = \gamma_r(\tilde{\theta})$, $r = 1, \ldots, p$, where

$$\tilde{g}_r = n^{-1}\sum_{j=1}^{n}\kappa_{rj}\, G(\zeta_j;\, \theta_0).$$

We then have, to first order in the $(U_j - \zeta_j)$,

$$d \sim n^{-1}\sum_j \delta_j \Bigg\{ F(G(\zeta_j;\, \theta_0);\, \tilde{\theta}) + F'(G(\zeta_j;\, \theta_0);\, \tilde{\theta})\, G'(\zeta_j;\, \theta_0)\, (U_j - \zeta_j)$$
$$+ \sum_r \sum_s \frac{\partial F}{\partial \theta_r}(G(\zeta_j;\, \theta_0);\, \tilde{\theta})\, H^{rs}(\tilde{\theta})\, n^{-1}\sum_k \kappa_{sk} G'(\zeta_k;\, \theta_0)\, (U_k - \zeta_k) \Bigg\} .$$

We may replace $\tilde{\theta}$ by θ_0 in this expression, for $\tilde{\theta} - \theta_0$ is deterministic and of order n^{-1}. Noting that

$$F'(G(x;\, \theta);\, \theta) G'(x;\, \theta) = 1$$

and

$$\frac{\partial F}{\partial \theta}(G(x;\, \theta)) = \frac{-1}{G'(x,\, \theta)}\, \frac{\partial G}{\partial \theta}(x;\, \theta) \;.$$

because F and G are inverse functions, we have

$$d \sim n^{-1}\sum_j \delta_j \zeta_j + n^{-1}\sum_j \delta_j (U_j - \zeta_j)$$
$$- \sum_r \sum_s \left\{ n^{-1}\sum_j \delta_j \frac{\partial G(\zeta_j,\, \theta_0)}{\partial \theta_r}\, \frac{1}{G'(\zeta_j,\, \theta_0)} \right\} \times$$
$$H^{rs}(\theta_0) \left\{ n^{-1}\sum_k \kappa_{sk} G'(\zeta_k,\, \theta_0)(U_k - \zeta_k) \right\}$$

The last two terms in this expression have mean zero, so to first order

$$E\,d = n^{-1}\sum_j \delta_j \zeta_j \sim \int_0^1 \delta(z)\, z\, dz\,;$$

and we also have

$$d \sim K + n^{-1}\sum_j \eta_j U_j$$

where K is a constant and $\eta_j = \eta(\zeta_j)$ where $\eta(u)$ is as defined in (7.3). Thus d is asymptotically equivalent, apart from a constant, to a linear combination of order statistics from a uniform distribution, and the asymptotic distribution of d now follows from the theorem of Stigler (1974) given in the proof of (5.3) above.

(7.4) This result is a standard consequence (see, for example, Rao, 1973, sections 3b.4, 6a.2) of the asymptotic distribution of d derived in (7.3).

(A.1.1) From (3.2) we have

$$\lambda_r = \int_0^1 x(F)\, P^*_{r-1}(F)\, dF\,.$$

The shifted Legendre polynomials P^*_{r-1} are orthogonal on [0, 1] with unit weight function, so $x(F)$, which is a polynomial of degree 1, is orthogonal to P^*_{r-1} if $r-1 \geq 2$, *i.e.* if $r \geq 3$. Thus $\lambda_r = 0$ if $r \geq 3$.

(A.1.2) Set $k = 1$ in the corresponding result (A.8.4) for the generalized Pareto distribution.

(A.1.3) If X is a real-valued random variable with distribution function F and T is a differentiable function such that $E\{T(X)\}^2 < \infty$, then

$$\operatorname{var} T(X) = 2\iint_{x<y} T'(x)T'(y)\, F(x)\{1 - F(y)\}\, dx\, dy$$

(proof: integrate the right side by parts successively). Thus if X has a uniform distribution with, without loss of generality, $\alpha = 0$ and $\beta = 1$, we have

$$\Lambda_{rr} = \operatorname{var}\{\dot{Q}_{r-1}(X)\},$$

and similarly

$$\Lambda_{rs} = \operatorname{cov}\{\dot{Q}_{r-1}(X), \dot{Q}_{s-1}(X)\},$$

where

$$\dot{Q}_r(x) \equiv \int_0^x \dot{P}_r(u)\, du.$$

It is easy to show that

$$\dot{Q}_0(x) = \tfrac{1}{2}\{\dot{P}_1(x) - \dot{P}_0(x)\},$$
$$\dot{Q}_r(x) = (4r+2)^{-1}\{\dot{P}_{r+1}(x) - \dot{P}_{r-1}(x)\}, \qquad r \geq 1,$$

and that

$$E\,\dot{Q}_{r-1}(X) = \int_0^1 \dot{Q}_{r-1}(u)\, du = -\lambda_r,$$

whence

$$\lambda_{rs} = \int_0^1 Q^{\bullet}_{r-1}(u)\, Q^{\bullet}_{s-1}(u)\,du - \lambda_r \lambda_s,$$

which when expressed as a sum of integrals of products of the $P^{\bullet}_r$ and evaluated using (C.3) and (C.4) yields (A.1.3).

(A.2.1) Using (2.6) we have

$$\begin{aligned} r\beta_{r-1} &= r \sum_{s=0}^{r-1} (-1)^s \binom{r-1}{s} \frac{1}{(s+1)^2} \\ &= r\, {}_3F_2\left[\begin{matrix} 1, 1, 1-r \\ 2, 2 \end{matrix}\right] \\ &= \psi(r+1) - \psi(1) \end{aligned}$$

(Exton, 1978, eqn. 2.1.15; Gradshteyn and Ryzhik, 1980, p. 538).

(A.2.2) Set $k = 0$ in the corresponding result (A.8.3) for the generalized Pareto distribution.

(A.2.3) Set $k = 0$ in the corresponding result (A.8.4) for the generalized Pareto distribution.

(A.3.1) Assuming, without loss of generality, that $\xi = 0$ and $\alpha = 1$, we have

$$\begin{aligned} r\beta_{r-1} &= r \int_0^1 \{-\log(-\log F)\}\, F^{r-1}\, dF \\ &= \int_0^{\infty} \log(r/u) e^{-u} du \\ &= \log r + \gamma \end{aligned}$$

(A.3.2) Let $k \to 0$ in the corresponding result (A.9.2) for the generalized extreme-value distribution.

(A.4.1) We assume, without loss of generality, that $\xi = 0$ and $\alpha = 1$. The Laplace distribution is symmetric about its mean, so if r is odd then $\lambda_r = 0$, while if r is even then

$$\lambda_r = 2\int_0^{1/2} P^*_{r-1}(F)\, x(F)\, dF$$
$$= \int_0^1 P^*_{r-1}(\tfrac{1}{2}F)\, \log(F)\, dF\,.$$

Writing $P^*_{r-1}(\frac{1}{2}F)$ as a sum of powers of F and integrating term by term we eventually obtain

$$\lambda_r = \sum_{j=1}^{r} \frac{(-1)^{r-j+1}(r-2+j)!}{2^{j-1}(j!)^2(r-j)!}$$
$$= \frac{-2}{r(r-1)}\{{}_2F_1(-r, r-1; 1; \tfrac{1}{2}) - 1\}\,.$$

(A.4.2) Assuming without loss of generality that $\xi = 0$ and $\alpha = 1$, we have

$$J_{rs} = \iint_{u<v} u^r v^s u(1-v)\, x'(u) x'(v)\, du\, dv\,,$$

with

$$x'(u) = \begin{cases} 1/u, & u < \frac{1}{2}\,, \\ 1/(1-u), & u > \frac{1}{2}\,. \end{cases}$$

Now J_{rs} is the sum of integrals over the regions $\{0 < u < v < \frac{1}{2}\}$, $\{0 < u < \frac{1}{2} < v < 1\}$ and $\{\frac{1}{2} < u < v < 1\}$, and it is straightforward to show that these three integrals yield the three terms on the right side of (A.4.2).

(A.5.1) Let $k \to 0$ in the corresponding result (A.10.1) for the generalized logistic distribution.

(A.5.2) Assuming without loss of generality that $\xi = 0$ and $\alpha = 1$, and using (3.2), we have

$$\lambda_r = \int_0^1 x(F) P^*_{r-1}(F)\,dF$$
$$= \int_0^1 P^*_{r-1}(F)\{\log F - \log(1 - F)\}\,dF.$$

Now

$$P^*_{r-1}(F) = (-1)^{r-1} P^*_{r-1}(1 - F),$$

so if r is odd then $\lambda_r = 0$, while if r is even and $r \geq 4$ then

$$
\begin{aligned}
\lambda_r &= 2\int_0^1 P^*_{r-1}(F)\log F\,dF \\
&= 2\sum_{k=0}^{r-1} p^*_{r-1,k}\int_0^1 F^k \log F\,dF \\
&= 2\sum_{k=0}^{r-1}(-1)^{r-1-k}\binom{r-1}{k}\binom{r-1+k}{k}\cdot\frac{(-1)}{(k+1)^2} \\
&= \frac{-2}{r(r-1)}\sum_{k=0}^{r-1}(-1)^{r-1-k}\binom{r}{k+1}\binom{r-1+k}{k+1} \\
&= \frac{-2}{r(r-1)}\sum_{k=1}^{r}(-1)^{r-k}\binom{r}{k}\binom{r-2+k}{k} \\
&= \frac{2}{r(r-1)}\left\{1-\sum_{k=0}^{r}(-1)^{r-k}\binom{r}{r-k}\binom{r-2+k}{k}\right\}
\end{aligned}
$$

The sum in the last expression is, by binomial expansion, the coefficient of x^r in $(1-x)^r(1-x)^{2-r}$ and is zero, so we have $\lambda_r = 2/\{r(r-1)\}$. This completes the proof.

(A.5.3) By letting $k \to 0$ in the corresponding result (A.10.2) for the generalized logistic distribution, we have

$$A_{rs} = B_{rs} = J_{rs} + J_{sr},$$

where

$$
\begin{aligned}
J_{rs} &= \frac{\alpha^2}{r+s+1}\,{}_3F_2\left[\begin{matrix}1,1,s+1\\ 2,r+s+2\end{matrix}\right] \\
&= \alpha^2 s^{-1}\{\psi(r+s+1)-\psi(r+1)\}
\end{aligned}
$$

(Exton, 1978, eqn. 2.1.15; Gradshteyn and Ryzhik, 1980, p. 538). The result follows.

(A.6.1) Assume without loss of generality that $\mu = 0$ and $\sigma = 1$; we have $\beta_0 = 0$ and

$$
\begin{aligned}
2\beta_1 = EX_{2:2} &= 2\int x\Phi(x)\phi(x)dx \\
&= -2\int \Phi(x)\phi'(x)dx \qquad \text{since } \phi'(x) = -x\phi(x) \\
&= 2\int \{\phi(x)\}^2 dx \qquad \text{by parts} \\
&= \pi^{-1}\int_{-\infty}^{\infty} e^{-x^2} dx \\
&= \pi^{-1/2} \qquad \text{(Gradshteyn and Ryzhik, 1980, p. 307).}
\end{aligned}
$$

(A.6.2) We have

$$3\beta_2 = EX_{3:3} = 3\int x\Phi^2\phi = -3\int \Phi^2\phi' = 6\int \Phi\phi^2 = 3\int \phi^2$$

since $(\Phi - ½)$ is an odd function and ϕ is an even function; the result follows, as in the proof of (A.6.1).

(A.6.3) We have

$$4\beta_3 = EX_{4:4} = 4\int x\Phi^3\phi = -4\int \Phi^3\phi' = 12\int \Phi^2\phi^2.$$

Now let

$$I(a) = \int_{-\infty}^{\infty} \{\Phi(ax)\}^2\{\phi(x)\}^2 dx;$$

then

$$dI/da = \int_{-\infty}^{\infty} x\,\Phi(ax)\,\phi(ax)\,\{\phi(x)\}^2$$

$$= (2\pi)^{-3/2}\int x\Phi(ax)e^{-(a^2+2)x^2/2}dx$$

$$= \tfrac{1}{2}\,\pi^{-3/2}a(1+a^2)^{-1/2}(2+a^2)^{-1}$$

(Gradshteyn and Ryzhik, 1980, p. 649), while

$$I(0) = \tfrac{1}{4}\int\phi^2 = \tfrac{1}{8}\,\pi^{-1/2};$$

thus

$$4\beta_3 = 12\left(\tfrac{1}{8}\,\pi^{-1/2} + \int_0^1 \frac{dI}{da}\,da\right);$$

hence result.

Note: the expressions for expected Normal order statistics used in (A.6.1-3) were proved by Jones (1948) and Godwin (1949).

(A.6.4) Use the result

$$\iint_{x<y}\{F(x)\}^r\{1-F(y)\}^s dx\,dy = \frac{r!s!}{(r+s)!}E(X_{r+1:r+s} - X_{r:r+s})^2$$

and tables of means and covariances of Normal order statistics.

(A.7.1) Assuming, without loss of generality, that $\xi = 0$ and $\sigma = 1$, we have

$$
\begin{aligned}
r\alpha_{r-1} &= r\int_0^\infty x\{1 - F(x)\}^{r-1} f(x)\,dx \\
&= \frac{r}{\sigma^2}\int_0^\infty x^2 e^{-rx^2/2\sigma^2}\,dx \\
&= \sigma\sqrt{\pi/(2r)}
\end{aligned}
$$

(Gradshteyn and Ryzhik, 1980, p. 337).

(A.7.2) Assuming, without loss of generality, that $\xi = 0$ and $\sigma = 1$, we have

$$
\begin{aligned}
K_{rs} &\equiv \iint_{x<y} \{1 - F(x)\}^r \{1 - F(y)\}^s\,dx\,dy \\
&= \iint_{x<y} e^{-rx^2/2} e^{-sy^2/2}\,dx\,dy \\
&= \int_0^\infty e^{-rx^2/2} (2\pi/s)^{1/2} \{1 - \Phi(x\sqrt{s})\}\,dx \\
&= 2\pi(rs)^{-1/2} \int_0^\infty \phi(t)\,\{1 - \Phi(t\sqrt{s/r})\}\,dt \\
&= (rs)^{-1/2} \arctan\,(r/s)^{1/2}
\end{aligned}
$$

(Gradshteyn and Ryzhik, 1980, p. 649). Since

$$
A_{rs} = K_{r,s+1} + K_{s,r+1} - K_{r+1,s+1} - K_{s+1,r+1}
$$

the result follows.

(A.8.1) Let $\gamma, \delta \to 0$ in the corresponding result (A.14.1) for the Wakeby distribution.

(A.8.2) Assuming, without loss of generality, that $\xi = 0$ and $\alpha = 1$, and using (2.6), we have

$$
\begin{aligned}
r\beta_{r-1} &= r\sum_{s=0}^{r-1}(-1)^r \binom{r-1}{s} \alpha_s \\
&= r\sum_{s=0}^{r-1} \frac{(-1)^s (r-1)!}{(r-1-s)!s!} \cdot \frac{1}{(s+1)(s+1+k)} \\
&= \sum_{s=0}^{r-1} \frac{(-1)^s r!}{(r-1-s)!(s+1)!} \cdot \frac{1}{(s+1+k)} \\
&= -\sum_{s=1}^{r} \frac{(-1)^s r!}{(r-s)!s!} \cdot \frac{1}{(s+k)} \\
&= k^{-1}\{1 - {}_2F_1(-r, k; 1+k; 1)\} \\
&= k^{-1}\{1 - \Gamma(1+k)\Gamma(1+r)/\Gamma(1+k+r)\}
\end{aligned}
$$

(Gradshteyn and Ryzhik, 1980, p. 1042), which is the required result.

(A.8.3) Assuming, without loss of generality, that $\xi = 0$ and $\alpha = 1$, we have

$$
\begin{aligned}
\lambda_r &= (-1)^{r-1}\sum_{s=0}^{r-1} p^*_{r-1,s}\,\alpha_s \\
&= \sum_{s=0}^{r-1} \frac{(-1)^s (r-1+s)!}{(s!)^2 (r-1-s)!} \cdot \frac{1}{(s+1)(s+1+k)} \\
&= \sum_{s=0}^{r-1}(-1)^s \binom{r-1}{s} \frac{(r-1+s)!}{(r-1)!} \frac{1}{(s+1)!\,(s+1+k)} \\
&= (1+k)^{-1}\,{}_3F_2\left[\begin{matrix} 1-r, r, 1+k \\ 2, 2+k \end{matrix}\right] \\
&= (1+k)^{-1}\frac{\Gamma(2+k)\Gamma(r-1-k)\Gamma(2)}{\Gamma(r+1+k)\Gamma(1-k)}
\end{aligned}
$$

(Mathai and Saxena, 1973, eqn. 4.1.29), which is the required result.

(A.8.4) Set $\gamma = 0$, $\delta = 0$ in the corresponding result (A.14.3) for the Wakeby distribution.

(A.9.1-2) Proved by Hosking *et al.* (1985). Note that Hosking *et al.*'s expression for $B_{r,r+s}$, $s \geq 2$, contains a typographical error.

(A.10.1) We have

$$\begin{aligned} r\alpha_{r-1} &= r\int_0^1 x(F)(1-F)^{r-1}\,dF \\ &= \xi + \alpha\left\{\frac{1}{k} - \frac{r}{k}\int_0^1 F^{-k}(1-F)^{r+k-1}\,dF\right\} \\ &= \xi + \frac{\alpha}{k}\left\{1 - r\frac{\Gamma(1-k)\Gamma(r+k)}{\Gamma(r+1)}\right\}, \end{aligned}$$

as required; the proof for $r\beta_{r-1}$ is similar.

(A.10.2) We have

$$x'(u) = \alpha u^{-k-1}(1-u)^{k-1},$$

so

$$\begin{aligned} J_{rs} &= \iint_{0<u<v<1} u^{r+1} v^{s}(1-v)\, x'(u)x'(v)\,du\,dv, \\ &= \alpha^2 \iint_{u<v} u^{r-k}(1-u)^{k-1} v^{s-k-1}(1-v)^{k}\,du\,dv. \end{aligned}$$

Now

$$\begin{aligned} \int_u^1 v^{s-k-1}(1-v)^k\,dv &= \int_0^{1-u} v^k(1-v)^{s-k-1}\,dv \\ &= \frac{(1-u)^{1+k}}{1+k}\,{}_2F_1(1+k-s, 1+k; 2+k; 1-u) \\ &= \frac{u^{s-k}(1-u)^{1+k}}{1+k}\,{}_2F_1(1, s+1; 2+k; 1-u) \end{aligned}$$

(Gradshteyn and Ryzhik, 1980, pp. 1040, 1043), so

$$J_{rs} = \frac{\alpha^2}{1+k}\int_0^1 u^{r+s-2k}(1-u)^{2k}\ {}_2F_1(1, s+1; 2+k; 1-u)\,du$$

$$= \frac{\alpha^2}{1+k}\int_0^1 u^{2k}(1-u)^{r+s-2k}\ {}_2F_1(1, s+1; 2+k; u)\,du$$

$$= \frac{\alpha^2}{1+k}\cdot\frac{\Gamma(1+2k)\Gamma(r+s+1-2k)}{\Gamma(r+s+2)}\ {}_3F_2\begin{bmatrix}1, s+1, 1+2k \\ r+s+2, 2+k\end{bmatrix}$$

(Exton, 1978, eqn. 2.1.1.5).

(A.11.1) In proving (A.11.1)-(A.11.5) we shall for simplicity derive the PWMs of the random variable $X = e^{cZ}$ where Z is a random variable with a standard Normal distribution: results for the general case are then given by a straightforward linear transformation. We have

$$\beta_r = \int_{-\infty}^{\infty} e^{cz}\{\Phi(z)\}^r \phi(z)dz$$

where

$$\phi(z) = (2\pi)^{-1/2}e^{-z^2/2}, \quad \Phi(z) = \int_{-\infty}^{z}\phi(t)dt,$$

and we note that

$$\int_{-\infty}^{\infty}\exp(-px^2 \pm qx)\,dx = \exp\left(\frac{q^2}{4p}\right)\sqrt{\frac{\pi}{p}} \qquad \text{(B.7)}$$

(Gradshteyn and Ryzhik, 1980, p. 307). Thus

$$\beta_0 = \int_{-\infty}^{\infty} e^{cz}\phi(z)dz = (2\pi)^{-1/2}\int_{-\infty}^{\infty} e^{-z^2/2+cz}dz$$

$$= e^{c^2/2}$$

by (B.7).

(A.11.2) Let

$$I(c) \equiv \int_{-\infty}^{\infty} e^{cz}\Phi(z)\phi(z)dz$$

be the integral defining β_1. We have

$$\begin{aligned}
\frac{d}{dc}I(c) &= \int_{-\infty}^{\infty} ze^{cz}\Phi(z)\phi(z)dz \\
&= -\int_{-\infty}^{\infty} e^{cz}\Phi(z)\phi'(z)dz \\
&= \int_{-\infty}^{\infty} ce^{cz}\Phi(z)\phi(z)dz + \int_{-\infty}^{\infty} e^{cz}\{\phi(z)\}^2dz \qquad \text{by parts} \\
&= cI(c) + (2\pi)^{-1}\int_{-\infty}^{\infty} e^{-z^2+cz}dz \\
&= cI(c) + \tfrac{1}{2}\pi^{-1/2}e^{c^2/4}
\end{aligned}$$

by (B.7). Thus $I(c)$ satisfies a linear first-order differential equation. We can write down the solution:

$$I(c) = e^{c^2/2}(2\pi)^{-1/2}\int^{c/\sqrt{2}} e^{-t^2/2}dt\ .$$

Now

$$I(0) = \int_{-\infty}^{\infty} \Phi(z)\phi(z)dz = \left[\tfrac{1}{2} \{\Phi(z)\}^2 \right]_{-\infty}^{\infty} = \tfrac{1}{2} = (2\pi)^{-1/2} \int_{-\infty}^{0} e^{-t^2/2} dt,$$

so

$$\beta_1 = I(c) = e^{c^2/2} (2\pi)^{-1/2} \int_{-\infty}^{c/\sqrt{2}} e^{-t^2/2} dt = e^{c^2/2} \Phi(c/\sqrt{2}).$$

(A.11.3) Let

$$J(c) = \int_{-\infty}^{\infty} e^{cz} \{\Phi(z)\}^2 \phi(z) dz$$

be the integral defining β_2. Integrating by parts as in the proof of (A.11.2) we obtain

$$\frac{d}{dc} J(c) = cJ(c) + 2K(c) \tag{B.8}$$

where

$$K(c) = \int_{-\infty}^{\infty} e^{cz} \Phi(z) \{\phi(z)\}^2 dz.$$

Similarly we find that

$$\frac{d}{dc} K(c) = \tfrac{1}{2} cK(c) + \frac{e^{c^2/6}}{4\pi\sqrt{3}}.$$

The solution of this equation is

$$K(c) = \frac{e^{c^2/4}}{2\sqrt{2}\,\pi}\int^{c/\sqrt{6}} e^{-t^2/2}\,dt$$

which, since

$$K(0) = \int_{-\infty}^{\infty} \Phi(z)\{\phi(z)\}^2\,dz = \tfrac{1}{4}\,\pi^{-1/2},$$

as in the proof of (A.6.2), gives

$$\begin{aligned} K(c) &= \tfrac{1}{2}\,\pi^{-1/2}e^{c^2/4}\left\{\tfrac{1}{2} + (2\pi)^{-1/2}\int_0^{c/\sqrt{6}} e^{-t^2/2}\,dt\right\} \\ &= \tfrac{1}{2}\,\pi^{-1/2}e^{c^2/4}\Phi(c/\sqrt{6}\,). \end{aligned}$$

We can now solve (B.8) to get

$$J(c) = e^{c^2/2}(2/\pi)^{1/2}\int^{c/\sqrt{2}} e^{-t^2/2}\Phi(t/\sqrt{3}\,)\,dt$$

which, since

$$J(0) = \int_{-\infty}^{\infty}\{\Phi(z)\}^2\phi(z)\,dz = \left[\tfrac{1}{3}\{\Phi(z)\}^3\right]_{-\infty}^{\infty} = \tfrac{1}{3}$$

gives

$$J(c) = e^{c^2/2}\left\{\tfrac{1}{3} + 2\int_0^{c/\sqrt{2}} \Phi(t/\sqrt{3}\,)\phi(t)\,dt\right\}.$$

Now

$$L(a) \equiv \int_{-\infty}^{\infty} \Phi(az)\phi(z)dz = \operatorname{arccot}(a)/2\pi$$

(proof: differentiate both sides with respect to a and use (B.7)), so $L(1/\sqrt{3}) = \frac{1}{6}$ and we obtain, finally,

$$\beta_2 = J(c) = 2e^{c^2/2}\int_{-\infty}^{c/\sqrt{2}} \Phi(t/\sqrt{3})\phi(t)dt .$$

(A.11.4-5) To find β_3 we follow our previous procedure, solving three first-order differential equations along the way, to get

$$\beta_3 = e^{c^2/2}\left\{ \tfrac{1}{4} + 6\int_0^{c/\sqrt{2}} \Omega(t/\sqrt{3})\phi(t)dt \right\}$$

where

$$\Omega(x) = \int_{-\infty}^{x} \Phi(t/\sqrt{2})\phi(t)dt$$

or, by changing the order of integration,

$$\beta_3 = e^{c^2/2}\left[\tfrac{1}{4} + \frac{6\{\Phi(c/\sqrt{2}) - 1/2\} \arctan\sqrt{2}}{2\pi} + 6\int_0^{c/\sqrt{6}} \{\Phi(c/\sqrt{2}) - \Phi(u\sqrt{3})\}\Phi(u/\sqrt{2})\phi(u)du \right].$$

from which (A.11.4) and (A.11.5) follow straightforwardly.

(A.11.6) The expression

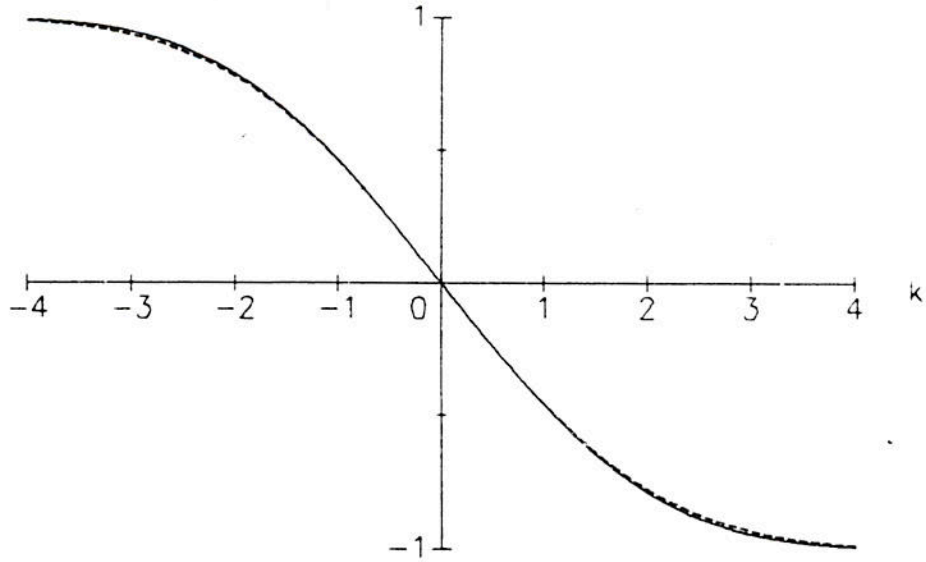

Figure B.2. Comparison of the function (B.9) (solid line) and the function $2\Phi(-k\sqrt{3/8}) - 1$ (dotted line).

$$\tau_3 = \frac{12\Psi(-k/\sqrt{2}) - 6\Phi(-k/\sqrt{2})}{2\Phi(-k/\sqrt{2}) - 1}. \tag{B.9}$$

considered as a function of k, closely resembles the function $2\Phi(-k\sqrt{3/8}) - 1$: see Figure B.2. (The factor $\sqrt{3/8}$ is obtained by equating the derivatives with respect to k of (B.9) and $2\Phi(-Ak) - 1$ at $k = 0$.) Thus as a first approximation

$$k \approx s \equiv -\sqrt{8/3}\ \Phi^{-1}\left(\frac{1+\tau_3}{2}\right). \tag{B.10}$$

We considered refinements of this approximation, of the form

$$k \approx s \sum_{j=0}^{n} a_j s^{2j};$$

the most suitable seems to be the $n = 2$ approximation which yields k with a relative error of less than 0.075% provided that $|k| \leq 4$.

Note that the quantity $(1 + \tau_3)/2$ used in (B.10) is itself a simple function of PWMs:

$$\frac{(1 + \tau_3)}{2} = \frac{3\beta_2 - 2\beta_1}{2\beta_1 - \beta_0} = \frac{EX_{3:3} - EX_{2:2}}{EX_{2:2} - EX_{1:1}} .$$

(A.12.1) Assuming, without loss of generality, that $\xi = 0$ and $\beta = 1$, we note that

$$f'(x) = (\alpha - 1)\, x^{-1} f(x) - f(x);$$

thus

$$\begin{aligned}
\beta_1 &= \int_0^\infty xF(x)f(x)dx \\
&= (\alpha - 1)\int_0^\infty F(x)f(x)dx - \int_0^\infty xF(x)f'(x)dx \\
&= (\alpha - 1)\int_0^\infty F(x)f(x)dx - [xF(x)f(x)]_0^\infty + \int_0^\infty F(x)f(x)dx + \int_0^\infty x\{f(x)\}^2 dx \\
&= \alpha\int_0^\infty F(x)f(x)dx + \int_0^\infty x^{2\alpha-1}e^{-2x}dx/\{\Gamma(\alpha)\}^2 \\
&= \tfrac{1}{2}\alpha + 2^{-2\alpha}\Gamma(2\alpha)/\{\Gamma(\alpha)\}^2 ,
\end{aligned}$$

whence the duplication formula for the gamma function yields the required result.

(A.12.2) Assuming that $\xi = 0$ and $\beta = 1$, proceed as in the proof of (A.12.1):

$$\beta_2 = \int xF^2 f = (\alpha - 1)\int F^2 f - \int xF^2 f'$$
$$= (\alpha - 1)\int F^2 f - [xF^2 f] + \int F^2 f + 2\int xf^2$$
$$= \alpha \int F^2 f + 2J(1)$$

where

$$J(t) = \int_0^\infty xF(tx)\{f(x)\}^2 dx$$

Now

$$\frac{dJ}{dt} = \int_0^\infty x^2 f(tx)\{f(x)\}^2 dx$$
$$= \frac{t^{\alpha-1}}{\{\Gamma(\alpha)\}^3}\int_0^\infty x^{3\alpha-1}e^{-(2+t)x}dx$$
$$= \frac{\Gamma(3\alpha)}{\{\Gamma(\alpha)\}^3}\cdot\frac{t^{\alpha-1}}{(2+t)^{3\alpha}}$$

and $J(0) = 0$, so

$$J(1) = \int_0^1 \frac{d}{dt}J(t)\,.\,dt$$
$$= \Gamma(3\alpha)\{\Gamma(\alpha)\}^{-3}\int_0^1 t^{\alpha-1}(2+t)^{-3\alpha}dt$$
$$= \Gamma(3\alpha)\{\Gamma(\alpha)\}^{-3}2^{-2\alpha}\int_0^{1/3} u^{\alpha-1}(1-u)^{2\alpha-1}du \qquad (u = t/(2+t))$$
$$= 2^{-2\alpha}\Gamma(2\alpha)I_{1/3}(\alpha, 2\alpha)/\{\Gamma(\alpha)\}^2.$$

Since $\int F^2 f = \frac{1}{3}$, the result follows.

(A.12.3) Assuming that $\xi = 0$ and $\beta = 1$, and proceeding as in the proof of (A.12.2), we obtain

$$\beta_3 = \tfrac{1}{4}\,\alpha + 3K(1,1)$$

where

$$K(s,t) = \int_0^\infty xF(sx)F(tx)\{f(x)\}^2\,dx\,.$$

We have $K(s,0) = K(0,t) = 0$ and

$$\partial^2 K/\partial s \partial t = \Gamma(4\alpha)\{\Gamma(\alpha)\}^{-4}(st)^{\alpha-1}(2+s+t)^{-4\alpha},$$

and the result follows, after a little algebra, from

$$K(1,1) = \int_0^1\int_0^1 \frac{\partial^2 K}{\partial s \partial t}\,ds\,dt\,.$$

(A.13.1-2) These results may be obtained by a straightforward extension of the methods used to obtain the corresponding results (A.8.2) and (A.8.3) for the generalized Pareto distribution.

(A.13.3) We have $B_{rs} = J_{rs} + J_{sr}$ with

$$\begin{aligned} J_{rs} &= \iint\limits_{0<u<v<1} u^r v^s\, u(1-v)\, x'(u)x'(v)\,du\,dv \\ &= \iint\limits_{0<u<v<1} u^{r+1} v^s (1-v)\{abu^{b-1} + cd(1-u)^{d-1}\}\{abv^{b-1} + cd(1-v)^{d-1}\}\,du\,dv \end{aligned}$$

Multiplying out the {} brackets we obtain four integrals. Three of these quickly reduce to simple sums of beta functions and yield the first five terms on the right side of (A.13.3): since

$$\int_u^1 v^s(1-v)^d dv = \int_0^{1-u} v^d(1-v)^s dv$$
$$= \sum_{k=0}^{s} (-1)^k \binom{s}{k} \int_0^{1-u} v^{k+d} dv$$
$$= \sum_{k=0}^{s} (-1)^k \binom{s}{k} \frac{(1-u)^{k+1+d}}{k+1+d}.$$

the fourth term is

$$\iint_{u<v} u^{r+1} v^s (1-v)\, cd(1-u)^{d-1} cd(1-v)^{d-1} du\, dv$$
$$= c^2 d^2 \sum_{k=0}^{s} (-1)^k \binom{s}{k} \frac{1}{k+1+d} \int_0^1 u^{r+1}(1-u)^{k+2d} du$$
$$= c^2 d^2 \sum_{k=0}^{s} (-1)^k \binom{s}{k} \frac{1}{k+1+d} \cdot \frac{\Gamma(r+2)\Gamma(1+k+2d)}{\Gamma(r+3+2d+k)}$$
$$= \frac{c^2 d^2 \Gamma(r+2)\Gamma(1+2d)}{(1+d)\Gamma(r+3+2d)}\, {}_3F_2\left[\begin{matrix} -s, 1+d, 1+2d \\ 2+d, r+3+2d \end{matrix}\right].$$

(A.14.1) From (2.4) we have

$$r\alpha_{r-1} = r\int_0^1 x(F)\,(1-F)^{r-1}\, dF$$
$$= \xi + \frac{\alpha}{\beta}\left(1 - \frac{r}{r+\beta}\right) - \frac{\gamma}{\delta}\left(1 - \frac{r}{r-\delta}\right)$$
$$= \xi + \frac{\alpha}{r+\beta} + \frac{\gamma}{r-\delta}.$$

(A.14.2) Proof as for the corresponding result (A.8.3) for the generalized Pareto distribution.

(A.14.3) We have $A_{rs} = I_{rs} + I_{sr}$, where

$$I_{rs} = \iint_{0<u<v<1} u(1-u)^r (1-v)^{s+1} x'(u)x'(v)\,du\,dv$$

and

$$x'(F) = \alpha(1-F)^{\beta-1} + \gamma(1-F)^{-\delta-1}.$$

Now

$$\begin{aligned}\int_u^1 (1-v)^{s+1} x'(v)\,dv &= \int_0^{1-u} v^{s+1}(\alpha v^{\beta-1} + \gamma v^{-\delta-1})dv \\ &= \frac{\alpha(1-u)^{s+1+\beta}}{s+1+\beta} + \frac{\gamma(1-u)^{s+1-\delta}}{s+1-\delta},\end{aligned}$$

so

$$\begin{aligned}I_{rs} &= \int_0^1 (1-u)u^r(\alpha u^{\beta-1} + \gamma u^{-\delta-1})\left(\frac{\alpha u^{s+1+\beta}}{s+1+\beta} + \frac{\gamma u^{s+1-\delta}}{s+1-\delta}\right)du \\ &= \frac{\alpha^2}{(r+1+\beta)(r+s+1+2\beta)(r+s+2+2\beta)} \\ &\quad + \frac{\alpha\gamma(2s+2+\beta-\delta)}{(s+1+\beta)(s+1-\delta)(r+s+1+\beta-\delta)(r+s+2+\beta-\delta)} \\ &\quad + \frac{\gamma^2}{(s+1-\delta)(r+s+1-2\delta)(r+s+2-2\delta)};\end{aligned}$$

the result follows.

(A.16.1) Write

$$r\alpha_{r-1} = r\int x\{1 - F(x)\}^{r-1} f(x)dx$$

$$= r\int_0^\infty x(pe^{-x/\alpha} + qe^{-x/\beta})^{r-1}(p\alpha^{-1}e^{-x/\alpha} + q\beta^{-1}e^{-x/\beta})dx\,;$$

multiply out the bracketed terms and integrate them individually.

(A.16.2) Write $A_{rs} = I_{rs} + I_{sr}$, where

$$I_{rs} = \iint_{x<y} F(x)\{1 - F(x)\}^r\{1 - F(y)\}^{s+1}\,dx\,dy$$

$$= K_{r,s+1} - K_{r+1,s+1}$$

where

$$K_{rs} = \iint_{x<y} \{1 - F(x)\}^r\{1 - F(y)\}^s\,dx\,dy\,;$$

proceed as in the proof of (A.16.1).

(A.16.3) Eliminating p and q from the equations for $2\alpha_1$ and $3\alpha_2$, we get

$$2\alpha_1 = \frac{\alpha_0^2 + \alpha\beta}{2(\alpha + \beta)}, \quad 3\alpha_2 = \frac{2\alpha_0^3 + 3\alpha\beta\alpha_0 + 2\alpha\beta(\alpha + \beta)}{3(\alpha + 2\beta)(2\alpha + \beta)}. \tag{B.11}$$

Noting that

$$(\alpha + 2\beta)(2\alpha + \beta) = 2(\alpha + \beta)^2 + \alpha\beta\,,$$

we can rewrite (B.11) as

$$4\alpha_1 S = \alpha_0^2 + P, \quad 18\alpha_2 S^2 + 9\alpha_2 = 2\alpha_0^3 + 3\alpha_0 P + 2PS. \tag{B.12}$$

where

$$S = \alpha + \beta, \quad P = \alpha\beta. \tag{B.13}$$

Eliminating P from (B.12) yields the quadratic equation in S given in the text. Given S, we obtain P from the first equation of (B.12), α and β from (B.13) as the roots of

$$z^2 - zS + P = 0,$$

and p and q from

$$\alpha_0 = p\alpha + q\beta, \quad p + q = 1.$$

APPENDIX C. SHIFTED LEGENDRE POLYNOMIALS

The shifted Legendre polynomials $P_n^*(x)$, $n = 0, 1, 2, \ldots$, are defined to be the orthogonal polynomials on the interval $[0, 1]$ with unit weight function and $P_n^*(1) = 1$, *i.e.*

$$
\begin{aligned}
&\text{(a)} \quad P_n^* \text{ is a polynomial of degree } n; \\
&\text{(b)} \quad P_n^*(1) = 1; \qquad\qquad (C.1) \\
&\text{(c)} \quad \int_0^1 P_m^*(x) P_n^*(x)\,dx = 0 \text{ if } m \neq n.
\end{aligned}
$$

Shifted Legendre polynomials are derived from the Legendre polynomials $P_n(x)$ which are orthogonal on the interval $[-1, 1]$, also with unit weight function and $P_n(1) = 1$. We therefore have

$$P_n(x) = P_n^*(2x - 1).$$

From Rodrigues' formula

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n$$

we then obtain

$$
\begin{aligned}
P_n^*(y) &= \frac{1}{2^n n!} \left. \frac{d^n}{dx^n} (x^2 - 1)^n \right|_{x=2y-1} \\
&= \frac{(-1)^n}{n!} \frac{d^n}{dy^n} \{y(1 - y)\}^n . \qquad (C.2)
\end{aligned}
$$

which yields the coefficients of

$$P_n^*(x) = \sum_{k=0}^{n} p_{n,k}^* x^k$$

as

$$P^{*}_{n,k} = \frac{(-1)^{n-k}(n+k)!}{(k!)^2(n-k)!} = (-1)^{n-k}\binom{n}{k}\binom{n+k}{k}.$$

The first few shifted Legendre polynomials are as follows:

$$P^{*}_0(x) = 1,$$
$$P^{*}_1(x) = 2x - 1,$$
$$P^{*}_2(x) = 6x^2 - 6x + 1,$$
$$P^{*}_3(x) = 20x^3 - 30x^2 + 12x - 1,$$
$$P^{*}_4(x) = 70x^4 - 140x^3 + 90x^2 - 20x + 1,$$
$$P^{*}_5(x) = 252x^5 - 630x^4 + 560x^3 - 210x^2 + 30x - 1,$$
$$P^{*}_6(x) = 924x^6 - 2772x^5 + 3150x^4 - 1680x^3 + 420x^2 - 42x + 1,$$
$$P^{*}_7(x) = 3432x^7 - 12012x^6 + 16632x^5 - 11550x^4 + 4200x^3 - 756x^2 + 56x - 1.$$

The following results are easily proved:

$$P^{*}_n(0) = (-1)^n.$$

$$P^{*}_n(x) = (-1)^n P^{*}_n(1-x). \tag{C.3}$$

$$\int_0^1 \{P^{*}_n(x)\}^2 dx = 1/(2n+1). \tag{C.4}$$

Further results concerning shifted Legendre polynomials are given by Lanczos (1957, chap. 4) – though he writes $(-1)^n P^{*}_n(x)$ for our $P^{*}_n(x)$.

APPENDIX D. POLYNOMIAL APPROXIMATION OF THE INVERSE DISTRIBUTION FUNCTION

The inversion theorem (3.10) leaves unanswered many questions about the utility of the approximation of $x(F)$ by the finite series

$$x_s(F) \equiv \sum_{r=1}^{s} (2r-1)\lambda_r P^*_{r-1}(F) \tag{D.1}$$

or by the corresponding series of sample L-moments

$$\hat{x}_s(F) \equiv \sum_{r=1}^{s} (2r-1)\, \ell_r P^*_{r-1}(F)\,. \tag{D.2}$$

For example ...

1. Is mean-square convergence of $x_s(F)$ to $x(F)$ adequate, or do we require pointwise convergence (for estimating a fixed quantile) or even uniform convergence (for simultaneously estimating an arbitrary number of quantiles)?
2. What is the rate of convergence of $R_s(F)$ to zero (in any of the senses mentioned in question 1)?
3. How many terms of (D.1) are needed to get a usable approximation to $x(F)$?
4. Is $x_s(F)$ monotonic increasing? Does it matter if it isn't?
5. In (D.2), what is the optimal choice $s^*(n)$ of s as a function of n?
6. Do we have even mean-square convergence of $\hat{x}_{s^*(n)}(F)$ to $x(F)$?

7.-10. The same questions as 1-4 but expressed in terms of $\hat{x}_{s^*(n)}(F)$ rather than $x_s(F)$.

Initial thoughts on these questions are not encouraging. For example, for question 1 uniform convergence of $R_s(F)$ to zero is impossible if X is unbounded (though it might be possible for $R_s(F)/x(F)$ if $x(F)$ can be bounded away from zero). However, Weierstrass's approximation theorem implies that there exist $b_{s,r}$, $r = 1, \ldots, s$, $s = 1, 2, \ldots$, such that

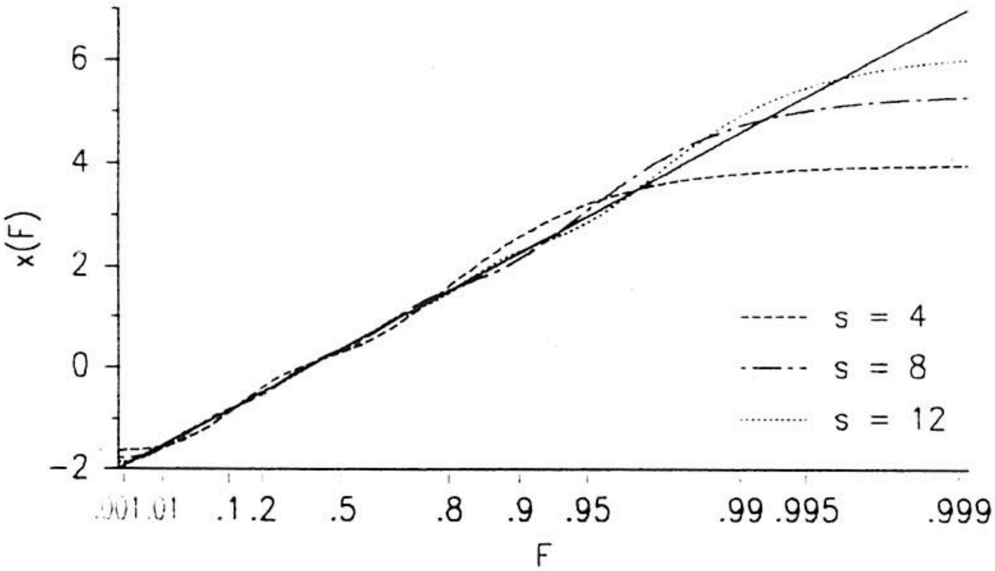

Figure D.1. Approximations to the quantile function of the Gumbel distribution.

$$x(F) - \sum_{r=1}^{s} b_{s,r} P^*_{r-1}(F) \to 0 \quad \text{as } s \to \infty$$

uniformly in $F \in [\delta, 1-\delta]$, $\delta > 0$, and presumably $b_{s,r}$ and a_r must be close to each other when r is large. For question 3, Figures D.1 and D.2 give approximations to Gumbel and generalized extreme-value distributions for various values of s: it looks as though many terms will be needed to attain usable accuracy in the tails of markedly skew distributions. For question 4, it is not difficult to find non-monotonic x_s's: *e.g.* $x_3(F)$ is not monotonic if $|\lambda_3/\lambda_2| > 0.2$, as occurs, for example, with the exponential distribution; monotonicity usually seems easier to attain if s is even but even here we have $x_4(F)$ decreasing at $F = \frac{1}{2}$ for generalized Pareto distributions with $k < -0.275$. For questions 6-10 note that we must surely have $s^*(n) \to \infty$ as $n \to \infty$, so standard asymptotic theory for the ℓ_r will not suffice.

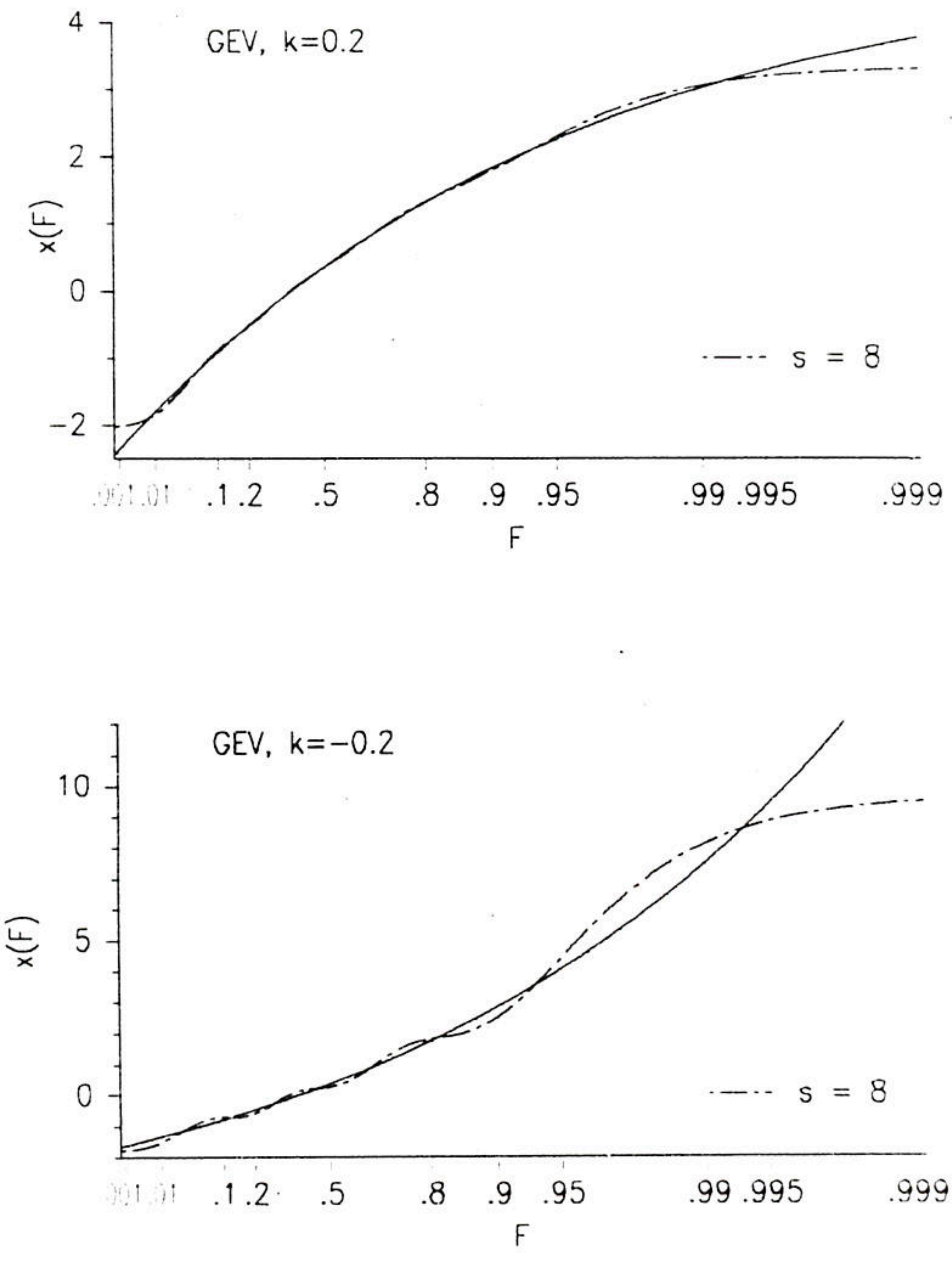

Figure D.2. Approximations to the quantile function of the generalized extreme-value distribution.

APPENDIX E. PWM ESTIMATION FOR A SPECIAL CASE OF THE GENERALIZED LAMBDA DISTRIBUTION

Assuming m known and, without loss of generality, $m = 0$, we have

$$x(F) = aF^b + c(1 - F)^d. \tag{E.1}$$

From $\alpha_r = E\{X(1 - F)^r\}$ we obtain

$$\begin{aligned}
\alpha_0 &= \frac{a}{B} - \frac{c}{D}, \\
\alpha_1 &= \frac{a}{B(1+B)} - \frac{c}{1+D}, \\
\alpha_2 &= \frac{2a}{B(1+B)(2+B)} - \frac{c}{2+D}, \\
\alpha_3 &= \frac{6a}{B(1+B)(2+B)(3+B)} - \frac{c}{3+D},
\end{aligned} \tag{E.2}$$

where $B = 1 + b$, $D = 1 + d$. Rewrite equations (E.2) as follows:

$$\begin{aligned}
B\alpha_0 D &= aD - cB, \\
(1+B)B\alpha_1 + (1+B)B\alpha_1 D &= a + aD - cB(1+B), \\
2(2+B)(1+B)B\alpha_2 + (2+B)(1+B)B\alpha_2 D &= 4a + 2aD - cB(1+B)(2+B), \\
3(3+B)(2+B)(1+B)B\alpha_3 + (3+B)(2+B)(1+B)B\alpha_3 D &= 18a + 6aD \\
&\quad - cB(1+B)(2+B)(3+B).
\end{aligned}$$

We can eliminate from these equations the terms involving a, aD and c; we thereby obtain

$$(D_0 + D_1 B + D_2 B^2) - (E_0 + E_1 B + E_2 B^2)D = 0, \tag{E.3}$$

where

$$\begin{aligned}
D_0 &= 10\alpha_1 - 28\alpha_2 + 18\alpha_3, & E_0 &= 2\alpha_0 - 10\alpha_1 + 14\alpha_2 - 6\alpha_3, \\
D_1 &= 4\alpha_1 - 18\alpha_2 + 15\alpha_3, & E_1 &= -4\alpha_1 + 9\alpha_2 - 5\alpha_3, \\
D_2 &= -2\alpha_2 + 3\alpha_3, & E_2 &= \alpha_2 - \alpha_3.
\end{aligned}$$

The PWMs $\beta_r = E(XF^r)$ are given by (E.2) with the substitutions $\alpha_r \to \beta_r$, $a \leftrightarrow -c$, $b \leftrightarrow -d$. Following the above procedure we therefore obtain

$$(B_0 + B_1 D + B_2 D^2) - (C_0 + C_1 D + C_2 D^2)B = 0, \qquad \text{(E.4)}$$

where

$$\begin{aligned} B_0 &= 10\beta_1 - 28\beta_2 + 18\beta_3, & C_0 &= 2\beta_0 - 10\beta_1 + 14\beta_2 - 6\beta_3, \\ B_1 &= 4\beta_1 - 18\beta_2 + 15\beta_3, & C_1 &= -4\beta_1 + 9\beta_2 - 5\beta_3, \\ B_2 &= -2\beta_2 + 3\beta_3, & C_2 &= \beta_2 - \beta_3. \end{aligned}$$

Using (E.3) to substitute for D in (E.4) we obtain a quintic polynomial of which B is a root. Given B we can find D from (E.3) and the parameters of (E.1) from

$$b = B - 1, \quad d = D - 1, \quad c = \frac{D(1 + D)}{1 - BD}\{(1 + B)\alpha_1 - \alpha_0\}, \quad a = b\{\alpha_0 + c/D\}. \qquad \text{(E.5)}$$

But which root of the quintic corresponds to B? It must be a real root, and the parameters (E.5) must yield an $x(F)$ which is a valid quantile function, *i.e.* $x'(F) \geq 0$ for all $F \in (0, 1)$, and for which EX exists. Now "$B > 0$, $D > 0$" is a necessary and sufficient condition for EX to exist given that $x(F)$ is valid, while "$ab > 0$, $cd > 0$" is a sufficient condition for the validity of $x(F)$ (necessary conditions are more complicated). It is possible that these conditions are satisfied by more than one solution of the quintic, and hence that the generalized lambda distribution is not uniquely determined by its first four PWMs.

As an example suppose that $a = -10$, $b = -0.1$, $c = -15$ and $d = -0.15$. The L-moment ratios of the distribution are $\lambda_1 = 6.5359$, $\lambda_2 = 2.0156$, $\tau_3 = 0.1764$, $\tau_4 = 0.2192$. The quintic equation for B is

$$-5.5834 + 14.5712B - 13.5832B^2 + 5.5063B^3 - 0.8147B^4 - 0.01331B^5 = 0,$$

with roots -67.5554, 0.9000, 1.6804, $1.8851 \pm 0.7429i$. The second and third of these roots yield valid parameters. The parameters and quantiles of the corresponding $x(F)$'s are tabulated below.

Root	Parameters *a*	*b*	*c*	*d*
0.9000	-10.00	-0.1000	-15.00	-0.1500
1.6804	9.32	0.6804	-0.2359	-0.7613

Root	Quantiles $x(F)$ F: 0.001	0.01	0.1	0.2	0.5	0.8	0.9	0.99	0.999
0.9000	-4.95	-0.83	2.65	3.76	5.93	8.87	11.08	19.92	32.27
1.6804	0.32	0.64	2.20	3.40	6.22	8.81	10.04	17.12	54.67

NOTATION

The symbol "≡" indicates that the quantity on the left side is defined as being equal to the expression on the right side; the symbol "=" indicates alternative definitions (which may not always be meaningful, *e.g.* if $f(x)$ does not exist). Definitions of PWMs and related quantities are valid for any random variable X whose mean exists; definitions of asymptotic sampling variances are valid only if X has finite variance. Proofs of the equivalence of the alternative definitions, if not immediate, are provided in the main text.

Preliminaries

$X_{k:n}$, $k = 1, \dots, n$: order statistics of a real-valued random variable X

$f(x)$: probability density function of X (when it exists)

$F(x)$: cumulative distribution function of X

$x(F)$: quantile function (inverse cumulative distributive function) of X

$x_{1:n} \le \cdots \le x_{n:n}$: ordered random sample of size n from the distribution of X

Probability weighted moments and related quantities

Probability weighted moments (PWMs):

$$M_{p,r,s} \equiv \frac{r!s!}{(r+s+1)!} EX^p_{r+1:r+s+1} = E[\,X^p\{F(X)\}^r\{1-F(X)\}^s\,],$$

$$\alpha_r \equiv M_{1,0,r} = E[\,X\{1-F(X)\}^r\,] = \int_0^1 x(F)(1-F)^r dF = \int x\{1-F(x)\}^r f(x)dx,$$

$$\beta_r \equiv M_{1,r,0} = E[\,X\{F(X)\}^r\,] \quad = \int_0^1 x(F)F^r dF \quad = \int x\{F(x)\}^r f(x)dx.$$

L-moments:

$$\lambda_1 \equiv EX \qquad = \alpha_0 \qquad = \beta_0,$$

$$\lambda_2 \equiv \tfrac{1}{2} E(X_{2:2} - X_{1:2}) \qquad = \alpha_0 - 2\alpha_1 \qquad = 2\beta_1 - \beta_0,$$

$$\lambda_3 \equiv \tfrac{1}{3} E(X_{3:3} - 2X_{2:3} + X_{1:3}) \qquad = \alpha_0 - 6\alpha_1 + 6\alpha_2 \qquad = 6\beta_2 - 6\beta_1 + \beta_0,$$

$$\lambda_4 \equiv \tfrac{1}{4} E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}) = \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0,$$

and in general

$$\lambda_r \equiv r^{-1} \sum_{k=1}^{r-1} (-1)^k \binom{r-1}{k} EX_{r-k:r} = (-1)^{r-1} \sum_{k=0}^{r-1} p^*_{r-1,k}\,\alpha_k = \sum_{k=0}^{r-1} p^*_{r-1,k}\,\beta_k$$

where

$$p^*_{r,k} \equiv \frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!}.$$

L-moment ratios:

$$L\text{-skewness} \qquad \tau_3 \equiv \lambda_3/\lambda_2,$$

$$L\text{-kurtosis} \qquad \tau_4 \equiv \lambda_4/\lambda_2,$$

and in general

$$\tau_r \equiv \lambda_r/\lambda_2.$$

Estimators of PWMs, etc.

Unbiased estimators of α_r:

$$a_r \equiv n^{-1} \sum_{k=1}^{n} \frac{(n-k)(n-k-1)\ldots(n-k-r+1)}{(n-1)(n-2)\ldots(n-r)}\, x_{k:n}.$$

and of β_r:

$$b_r \equiv n^{-1} \sum_{k=1}^{n} \frac{(k-1)(k-2)\dots(k-r)}{(n-1)(n-2)\dots(n-r)} x_{k:n} .$$

Plotting-position (biased) estimators of α_r:

$$\tilde{\alpha}_r \equiv \tilde{\alpha}_r[\gamma, \delta] \equiv n^{-1} \sum_{k=1}^{n} (1 - p_{k:n})^r x_{k:n} ,$$

and of β_r:

$$\tilde{\beta}_r \equiv \tilde{\beta}_r[\gamma, \delta] \equiv n^{-1} \sum_{k=1}^{n} p_{k:n}^r \, x_{k:n} ,$$

where $p_{k:n} \equiv (k + \gamma)/(n + \delta)$, $\delta > \gamma > -1$, is the plotting position of $x_{k:n}$.

Unbiased (respectively plotting-position) estimators ℓ_r, t_r (respectively $\tilde{\lambda}_r$, $\tilde{\tau}_r$) are the same functions of a_r, b_r (respectively $\tilde{\alpha}_r$, $\tilde{\beta}_r$) as λ_r, τ_r are of α_r, β_r.

Asymptotic sampling variances of sample PWMs, etc.

$$A_{rs} \equiv \lim_{n \to \infty} n \operatorname{cov}(a_r, a_s) = I_{rs} + I_{sr}$$

where

$$\begin{aligned} I_{rs} &\equiv \iint_{0<u<v<1} (1-u)^r (1-v)^s . \, u(1-v) \, x'(u) x'(v) \, du \, dv \\ &= \iint_{x<y} \{1 - F(x)\}^r \{1 - F(y)\}^s . \, F(x)\{1 - F(y)\} \, dx \, dy ; \end{aligned}$$

$$B_{rs} \equiv \lim_{n \to \infty} n \operatorname{cov}(a_r, a_s) = J_{rs} + J_{sr}$$

where

$$J_{rs} \equiv \iint_{0<u<v<1} u^r v^s . u(1-v)\, x'(u)x'(v)\, du\, dv$$
$$= \iint_{x<y} \{F(x)\}^r \{F(y)\}^s . F(x)\{1-F(y)\}\, dx\, dy\,;$$

$$\Lambda_{rs} \equiv \lim_{n\to\infty} n \operatorname{cov}(\ell_r, \ell_s) = \iint_{0<u<v<1} \{P^*_{r-1}(u)P^*_{s-1}(v) + P^*_{s-1}(u)P^*_{r-1}(v)\}.\, u(1-v)\, x'(u)x'(v)\, du\, dv$$

where

$$P^*_r(x) \equiv \sum_{k=0}^{r} p^*_{r,k}\, x^k.$$

$$A \equiv \lim_{n\to\infty} n \operatorname{var}(a_0\ a_1\ a_2\ a_3)^{\mathrm{T}} = \begin{bmatrix} A_{00} & A_{01} & A_{02} & A_{03} \\ A_{10} & A_{11} & A_{12} & A_{13} \\ A_{20} & A_{21} & A_{22} & A_{23} \\ A_{30} & A_{31} & A_{32} & A_{33} \end{bmatrix};$$

$$B \equiv \lim_{n\to\infty} n \operatorname{var}(b_0\ b_1\ b_2\ b_3)^{\mathrm{T}} = \begin{bmatrix} B_{00} & B_{01} & B_{02} & B_{03} \\ B_{10} & B_{11} & B_{12} & B_{13} \\ B_{20} & B_{21} & B_{22} & B_{23} \\ B_{30} & B_{31} & B_{32} & B_{33} \end{bmatrix};$$

$$\Lambda \equiv \lim_{n\to\infty} n \operatorname{var}(\ell_1\ \ell_2\ \ell_3\ \ell_4)^{\mathrm{T}} = MAM^{\mathrm{T}} = NBN^{\mathrm{T}}$$

where

$$M \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & -2 & 0 & 0 \\ 1 & -6 & 6 & 0 \\ 1 & -12 & 30 & -20 \end{bmatrix}, \quad N \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 2 & 0 & 0 \\ 1 & -6 & 6 & 0 \\ -1 & 12 & -30 & 20 \end{bmatrix};$$

$$T \equiv \lim_{n\to\infty} n \operatorname{var}(\ell_1\ \ell_2\ t_3\ t_4)^{\mathrm{T}} = Q\Lambda Q^{\mathrm{T}}$$

where

$$Q \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -\tau_3/\lambda_2 & -\tau_4/\lambda_2 \\ 0 & 0 & 1/\lambda_2 & 0 \\ 0 & 0 & 0 & 1/\lambda_2 \end{bmatrix}.$$

We can also write T as

$$\begin{bmatrix} \sigma_{11} & \cdot & \cdot & \cdot \\ \sigma_{12} & \sigma_{22} & \cdot & \cdot \\ (\sigma_{13} - \tau_3\sigma_{12})/\lambda_2 & (\sigma_{23} - \tau_3\sigma_{22})/\lambda_2 & (\sigma_{33} - 2\tau_3\sigma_{23} + \tau_3^2\sigma_{22})/\lambda_2^2 & \cdot \\ (\sigma_{14} - \tau_4\sigma_{12})/\lambda_2 & (\sigma_{24} - \tau_4\sigma_{22})/\lambda_2 & \dfrac{\sigma_{34} - \tau_3\sigma_{24} - \tau_4\sigma_{23} + \tau_3\tau_4\sigma_{22}}{\lambda_2^2} & \dfrac{\sigma_{44} - 2\tau_4\sigma_{24} + \tau_4^2\sigma_{22}}{\lambda_2^2} \end{bmatrix}$$

where $\sigma_{ij} \equiv (\Lambda)_{ij}$.

Miscellaneous mathematical notation

$\Gamma(x)$: gamma function

$\psi(x)$: digamma function (Euler's psi function)

${}_2F_1(a, b; c; x)$: hypergeometric function

${}_3F_2\begin{bmatrix} a, b, c \\ d, e \end{bmatrix}$: generalized hypergeometric function of unit argument

REFERENCES

BARLOW, R.E., BARTHOLOMEW, D.J., BREMNER, J.M. and BRUNK, H.D. (1972). *Statistical inference under order restrictions.* New York: Wiley.

BARNETT, F.C., MULLEN, K. and SAW, J.G. (1967). Linear estimates of a population scale parameter. *Biometrika*, **54**, 551-554.

BASU, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhya*, **15**, 377-380.

BOWLEY, A.L. (1920). *Elements of statistics*, 4th ed. New York: Scribner's.

CHAN, L.K. (1967). On a characterization of distributions by expected values of extreme order statistics. *Amer. Math. Mon.*, **74**, 950-951.

CHENG, R.C.H. and AMIN, N.A.K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *J.R. Statist. Soc. B*, **45**, 394-403.

CHERNOFF, H., GASTWIRTH, J.L. and JOHNS, M.V. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. *Ann. Math. Statist.*, **38**, 52-72.

CUNNANE, C. (1978). Unbiased plotting positions - a review. *J. Hydrol.*, **37**, 205-222.

DAVID, H.A. (1981). *Order statistics*, 2nd ed. New York: Wiley.

DOWNTON, F. (1966a). Linear estimates with polynomial coefficients. *Biometrika*, **53**, 129-141.

DOWNTON, F. (1966b). Linear estimates of parameters in the extreme value distribution. *Technometrics*, **8**, 3-17.

DUBEY, S.D. (1969). A new derivation of the logistic distribution. *Naval Res. Logist. Quart.*, **16**, 37-40.

EASTON, G.S. and RONCHETTI, E. (1986). General saddlepoint approximations with applications to *L* statistics. *J. Amer. Statist. Assoc.*, **81**, 420-430.

EXTON, H. (1978). *Handbook of hypergeometric integrals.* Chichester: Ellis Horwood.

FRASER, D.A.S. (1957). *Nonparametric methods of statistics.* New York: Wiley.

GAIL, M.H. and GASTWIRTH, J.L. (1978). A scale-free goodness-of-fit test for the exponential distribution based on the Gini statistic. *J. R. Statist. Soc. B*, **40**, 350-357.

GINI, C. (1912). Variabilità e mutabilità, contributo allo studio delle distribuzioni e delle relazioni statistiche. *Stud. Econ.-Giuridici R. Univ. Cagliari*, **3**, 3-159.

GODWIN, H.J. (1949). Some low moments of order statistics. *Ann. Math. Statist.*, **20** 279-285.

GRADSHTEYN, I.S. and RYZHIK, M. (1980). *Table of integrals, series and products*. 2nd ed. New York: Academic Press.

GREENWOOD, J.A., LANDWEHR, J.M., MATALAS, N.C. and WALLIS, J.R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. *Water Resour. Res.*, **15**, 1049-1054.

GROENEVELD, R.A. and MEEDEN, G. (1984). Measuring skewness and kurtosis. *Statistician*, **33**, 391-399.

GUMBEL, E.J. (1944). Ranges and midranges. *Ann. Math. Statist.*, **15**, 414-422.

HARTER, H.L. (1984). Another look at plotting positions. *Commun. Statist. Theor. Meth.*, **13**, 1613-1633.

HELMERS, R. (1980). Edgeworth expansions for linear combinations of order statistics with smooth weight functions. *Ann. Statist.*, **8**, 1361-1374.

HINKLEY, D.V. (1975). On power transformations to symmetry. *Biometrika*, **62**, 101-111.

HOEFFDING, W. (1948). A class of statistics with asymptotically Normal distribution. *Ann. Math. Statist.*, **19**, 293-325.

HOSKING, J.R.M. (1984). Testing whether the shape parameter is zero in the generalized extreme-value distribution. *Biometrika*, **71**, 367-374.

HOSKING, J.R.M. (1986). The Wakeby distribution. *IBM Res. Rep.*, in preparation.

HOSKING, J.R.M. and WALLIS, J.R. (1986). Estimation and hypothesis testing for the generalized Pareto distribution. Submitted for publication.

HOSKING, J.R.M., WALLIS, J.R. and WOOD, E.F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, **27**, 251-261.

HOUGHTON, J.C. (1978). Birth of a parent: the Wakeby distribution for modeling flood flows. *Water Resour. Res.*, **14**, 1105-1109.

JOHNSON, N.L. and KOTZ, S. (1970). *Distributions in statistics: continuous univariate distributions*, 2 vols. New York: Wiley.

JOINER, B.L. and ROSENBLATT, J.R. (1971). Some properties of the range in samples from Tukey's symmetric lambda distributions. *J. Amer. Statist. Assoc.*, **66**, 394-399.

JONES, H.L. (1948). Exact lower moments of order statistics in small samples from a normal distribution. *Ann. Math. Statist.*, **19**, 270-273.

KENDALL, M.G. and STUART, A. (1979). *The advanced theory of statistics, vol. 1*, 4th ed. London: Griffin.

KIRBY, W. (1974). Algebraic boundedness of sample statistics. *Water Resour. Res.*, **10**, 220-222.

LANCZOS, C. (1957). *Applied analysis*. London: Pitman.

LANDWEHR, J.M., MATALAS, N.C. and WALLIS, J.R. (1978). Some comparisons of flood statistics in real and log space. *Water Resour. Res.*, **14**, 902-920.

LANDWEHR, J.M., MATALAS, N.C. and WALLIS, J.R. (1979a). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resour. Res.*, **15**, 1055-1064.

LANDWEHR, J.M., MATALAS, N.C. and WALLIS, J.R. (1979b). Estimation of parameters and quantiles of Wakeby distributions. *Water Resour. Res.*, **15**, 1361-1379.

LOCKE, C. and SPURRIER, J.D. (1976). The use of *U*-statistics for testing Normality against non-symmetric alternatives. *Biometrika*, **63**, 143-147.

LOMNICKI, Z.A. (1952). The standard error of Gini's mean difference. *Ann. Math. Statist.* **23**, 635-637.

MATHAI, A.M. and SAXENA, R.K. (1973). *Generalized hypergeometric functions with applications in statistics and physical sciences.* Berlin: Springer.

MOORE, D.S. (1968). An elementary proof of asymptotic normality of linear functions of order statistics. *Ann. Math. Statist.*, **39**, 263-265.

MUNRO, A.H. and WIXLEY, R.A.J. (1970). Estimators based on order statistics of small samples from a three-parameter lognormal distribution. *J. Amer. Statist. Assoc.* **65**, 212-225.

NAIR, U.S. (1936). The standard error of Gini's mean difference. *Biometrika*, **28**, 428-436.

NEYMAN, J. (1937). "Smooth test" for goodness of fit. *Skand. Aktuar.*, **20**, 150-199.

OJA, H. (1981). On location, scale, skewness and kurtosis of univariate distributions. *Scand. J. Statist.*, **8**, 154-168.

OZTURK, A. and DALE, R.F. (1985). Least squares estimation of the parameters of the generalized lambda distribution. *Technometrics*, **27**, 81-84.

PEARSON, K. (1900). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Phil. Mag., 5th Ser.*, **50**, 157-175.

RAMBERG, J.S. and SCHMEISER, B.W. (1972). An approximate method for generating symmetric random variables. *Commun. Assoc. Comput. Mach.*, **15**, 987-990.

RAMBERG, J.S. and SCHMEISER, B.W. (1974). An approximate method for generating asymmetric random variables. *Commun. Assoc. Comput. Mach.*, **17**, 78-82.

RANDLES, R.H. and WOLFE, D.A. (1979). *Introduction to the theory of nonparametric statistics.* New York: Wiley

RAO, C.R. (1973). *Linear statistical inference and its applications*, 2nd ed. New York: Wiley.

SHAPIRO, S.S. and WILK, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, **52**, 591-611.

SILLITTO, G.P. (1951). Interrelations between certain linear systematic statistics of samples from any continuous population. *Biometrika*, **38**, 377-382.

SILLITTO, G.P. (1964). Some relations between expectations of order statistics in samples of different sizes. *Biometrika*, **51**, 259-262.

SILLITTO, G.P. (1969). Derivation of approximants to the inverse distribution function of a continuous univariate population from the order statistics of a sample. *Biometrika*, **56**, 641-650.

SMITH, R.L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, **72**, 67-90.

STIGLER, S.M. (1974). Linear functions of order statistics with smooth weight functions. *Ann. Statist.*, **2**, 676-693. Correction: *Ann. Statist.*, **7** (1979), 466.

SZEGÖ, G. (1959). *Orthogonal polynomials*, 2nd ed. New York: American Mathematical Society.

THOMAS, D.R. and PIERCE, D.A. (1979). Neyman's smooth goodness-of-fit test when the hypothesis is composite. *J. Amer. Statist. Assoc.*, **74**, 441-445.

TIAGO de OLIVEIRA, J. (1963). Decision results for the parameters of the extreme value (Gumbel) distribution based on the mean and the standard deviation. *Trabajos de Estadistica*, **14**, 61-81.

TITCHMARSH, E.C. (1946). *Eigenfunction expansions*. New York: Oxford University Press.

TUKEY, J.W. (1960). The practical relationship between the common transformations of percentages of counts and of amounts. *Tech. Rep. 36*, Statistical Techniques Research Group, Princeton Univ.

von ANDRAE, C.C.G. (1872). Über die Bestimmung das wahrscheinlichen Fehlers durch die gegebenen Differenzen von ein gleich genauen Beobachtungen einer Unbekannten. *Astron. Nach.*, **79**, 257-272.

WALD, A. (1943). Tests of hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.*, **52**, 426-482.

WALLIS, J.R., MATALAS, N.C. and SLACK, J.R. (1974). Just a moment! *Water Resour. Res.*, **10**, 211-219.

WIDDER, D. (1941). *The Laplace transform*. Princeton: Princeton University Press.