RC 14492 (64907) 3/22/89, revised 7/15/96
Mathematics 9 pages

Research Report

Some theoretical results concerning L-moments

J. R. M. Hosking

IBM Research Division
T. J. Watson Research Center
Yorktown Heights, NY 10598

Research Division
Almaden • T.J. Watson • Tokyo • Zurich

Some theoretical results concerning L-moments

J. R. M. Hosking

IBM Research Division
P. O. Box 218
Yorktown Heights, NY 10598

Abstract. L-moments and L-moment ratios are quantities useful in the summarization and estimation of probability distributions. Hosking (*J. R. Statist. Soc. B*, 1990) describes the theory and applications of L-moments. Here we give an expanded discussion of some of the theory in Hosking (1990), including in particular proofs of Theorems 2.1, 2.2 and 2.3 of that paper.

1. L-moments: definitions and basic properties

Let X be a real-valued random variable with cumulative distribution function $F(x)$ and quantile function $x(F)$, and let $X_{1:n} \le X_{2:n} \le \ldots \le X_{n:n}$ be the order statistics of a random sample of size n drawn from the distribution of X. Define the *L-moments* of X to be the quantities

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

The L in "L-moments" emphasizes that λ_r is a *linear* function of the expected order statistics. Furthermore, as noted in Hosking (1990, section 3), 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. The expectation of an order statistic may be written as

$$EX_{j:r} = \frac{r!}{(j-1)!\,(r-j)!} \int x\{F(x)\}^{j-1}\{1-F(x)\}^{r-j}\, dF(x)$$

(David, 1981, p. 33). Substituting this expression in (1), expanding the binomials in $F(x)$ and summing the coefficients of each power of $F(x)$ gives

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

where

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

and

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

$P^*_r(F)$ is the rth shifted Legendre polynomial, related to the usual Legendre polynomials $P_r(u)$ by $P^*_r(u) = P_r(2u-1)$. Shifted Legendre polynomials are orthogonal on the interval $(0,1)$ with constant weight function (Lanczos, 1957, p. 286—though his $P^*_r(\,.\,)$ differs by a factor $(-1)^r$ from ours). The first few L-moments are

$$\begin{aligned}
\lambda_1 &= EX &&= \int x\,.\,dF\,,\\
\lambda_2 &= \tfrac{1}{2}E(X_{2:2} - X_{1:2}) &&= \int x\,.\,(2F-1)\,dF\,,\\
\lambda_3 &= \tfrac{1}{3}E(X_{3:3} - 2X_{2:3} + X_{1:3}) &&= \int x\,.\,(6F^2-6F+1)\,dF\,,\\
\lambda_4 &= \tfrac{1}{4}E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}) &&= \int x\,.\,(20F^3-30F^2+12F-1)\,dF\,.
\end{aligned}$$

The use of L-moments to describe probability distributions is justified by the following theorem.

Theorem 1. (i) The L-moments λ_r, $r=1,2,\ldots$, of a real-valued random variable X exist if and only if X has finite mean.

(ii) A distribution whose mean exists is characterized by its L-moments $\{\lambda_r : r = 1,2,\ldots\}$.

Proof. A finite mean implies finite expectations of all order statistics (David, 1981, p. 33), whence part (i) follows immediately.

For part (ii), we first show that a distribution is characterized by the set $\{EX_{r:r}, r=1,2,\ldots\}$. This has been proved by Chan (1967) and Konheim (1971): the following proof is essentially Konheim's. Let X and Y be random variables with cumulative distribution functions F and G and quantile functions $x(u)$ and $y(u)$ respectively. Let

$$\xi_r^{(X)} \equiv EX_{r:r} = r\int x\{F(x)\}^{r-1} dF(x)\,, \qquad \xi_r^{(Y)} \equiv EY_{r:r} = r\int x\{G(x)\}^{r-1} dG(x)\,.$$

Then

$$\begin{aligned}\xi_{r+2}^{(X)} - \xi_{r+1}^{(X)} &= \int_0^1 \{(r+2)u^{r+1} - (r+1)u^r\}\, x(u)\, du \\ &= \int_0^1 u^r \,.\, u(1-u)\, dx(u) \qquad \text{by parts} \\ &= \int_0^1 u^r \,.\, dz_X(u)\end{aligned}$$

where $z_X(u)$, defined by $dz_X(u) = u(1-u)dx(u)$, is an increasing function on $(0,1)$. If $\xi_r^{(X)} = \xi_r^{(Y)}$, $r=1,2,\ldots$, then

$$\int_0^1 u^r\, dz_X(u) = \int_0^1 u^r\, dz_Y(u)\,, \qquad r = 0,1,\ldots\,.$$

Thus z_X and z_Y are distributions which have the same moments on the finite interval (0,1); consequently (Feller, 1970, pp. 222–224), $z_X = z_Y$. This implies that $x(u) = y(u)$.

We have shown that a distribution with finite mean is characterized by the set $\{\xi_r : r=1,2,\ldots\}$. Using (2), we have

$$\lambda_r = \sum_{k=1}^r p^*_{r-1,k-1} k^{-1} \xi_k\,,$$

whence

$$\xi_r = \sum_{k=1}^r \frac{(2k-1)\, r!\, (r-1)!}{(r-k)!\, (r-1+k)!}\, \lambda_k\,. \tag{3}$$

Thus a given set of λ_r determines a unique set of ξ_r, so the characterization of a distribution in terms of the latter quantities extends to the former.

Thus a distribution may be specified by its L-moments even if some of its conventional moments do not exist. Furthermore, such a specification is always unique: this is of course not true of conventional moments. Indeed, the proof of Theorem 1 shows (in a sense) why L-moments characterize a distribution whereas conventional moments, in general, do not: characterization by L-moments reduces to the classical moment problem on a finite interval (the Hausdorff moment problem)

$$\text{“given } s_r = \int_0^1 u^r \, dz(u), \quad r = 0, 1, \ldots, \text{ find } z(u)\text{”},$$

whereas characterization by conventional moments is the classical moment problem on an infinite interval (the Hamburger moment problem)

$$\text{“given } s_r = \int_{-\infty}^{\infty} u^r \, dz(u), \quad r = 0, 1, \ldots, \text{ find } z(u)\text{”}.$$

Only the Hausdorff problem has a unique solution.

As shown in Hosking (1990), 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. Define, therefore, the *L-moment ratios* of X to be the quantities

$$\tau_r \equiv \lambda_r / \lambda_2, \qquad r = 3, 4, \ldots .$$

It is also possible to define a function of L-moments which is analogous to the coefficient of variation: this is the L-CV, $\tau \equiv \lambda_2/\lambda_1$. Bounds on the numerical values of the L-moment ratios and L-CV are given by the following theorem.

Theorem 2. Let X be a nondegenerate random variable with finite mean. Then the L-moment ratios of X satisfy $|\tau_r| < 1$, $r \geq 3$. If in addition $X \geq 0$ almost surely, then τ, the L-CV of X, satisfies $0 < \tau < 1$.

Proof. Define $Q_r(t)$ by

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

$Q_r(t)$ is the Jacobi polynomial $P_r^{(1,1)}(2t-1)$. From Szegö (1959, chap. 4) it follows that

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

so integrating (2) by parts gives

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

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

$$\lambda_r = \int F(x)\,\{1-F(x)\}\,(r-1)^{-1}Q_{r-2}(F(x))\,dx\,. \tag{4}$$

Since $Q_0(t)=1$ the case $r=2$ gives

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

Now $0\le F(x)\le 1$ 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 it follows that

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

From Szegö (1959, p. 166) it follows that

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

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

If $X\ge 0$ almost surely then $\lambda_1=EX>0$ and $\lambda_2>0$, so $\tau=\lambda_2/\lambda_1>0$; furthermore $EX_{1:2}>0$, so

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

We consider the boundedness of L-moment ratios to be an advantage. Intuitively, it seems easier to interpret a measure such as τ_3, which is constrained to lie within the interval $(-1,1)$, than the conventional skewness, which can take arbitrarily large values.

More stringent bounds on the τ_r can be found. The proof of Theorem 1 implies that a sequence $\xi_1,\xi_2,\ldots$, can be the "$EX_{r:r}$" of some random variable X if and only if there exists an increasing function z such that

$$\xi_{r+2}-\xi_{r+1} = \int_0^1 u^r\,dz(u)\,, \qquad r=0,1,\ldots\,.$$

Establishing conditions for the existence of the function z is part of the classical "moment problem", and was solved by Hausdorff (1923). Akhiezer (1965, p. 74) shows that z exists if and only if the following quadratic forms are nonnegative:

$$\sum_{i,j=0}^{m} (\xi_{i+j+3} - \xi_{i+j+2})\, x_i\, x_j\;, \qquad \sum_{i,j=0}^{m} (-\xi_{i+j+3} + 2\xi_{i+j+2} - \xi_{i+j+1})\, x_i\, x_j\;,$$

for $r = 2m-1$, and

$$\sum_{i,j=0}^{m} (\xi_{i+j+2} - \xi_{i+j+1})\, x_i\, x_j\;, \qquad \sum_{i,j=0}^{m-1} (-\xi_{i+j+4} + 2\xi_{i+j+3} - \xi_{i+j+2})\, x_i\, x_j\;,$$

for $r = 2m-2$. These conditions can be expressed in terms of the nonnegativity of the determinants of certain matrices whose elements are linear combinations of the ξ_r. Mallows (1973) gives an exact statement of the result.

Theorem 3 (Mallows, 1973, Theorem 2(ii)). The sequence $\xi_1, \ldots, \xi_n$ may be regarded as the expectations of the largest order statistics of samples of size $1, \ldots, n$ of a real-valued random variable if and only if either (a), $A_k > 0$, $k = 2, \ldots, n$, and $B_k > 0$, $k = 3, \ldots, n$, or (b), for some even m, $2 \le m < n$, we have $A_k > 0$ and $B_k > 0$ for $k = 2, \ldots, m$, $A_m = 0$, $B_m > 0$ and $A_k = B_k = 0$ for $k = m+1, \ldots, n$. Here A_k and B_k are determinants of matrices, as follows:

$$\begin{aligned}
A_{2k} &= \det\, [\xi_{i+j} - \xi_{i+j-1}]_{i,j=1,\ldots,k}\;, \\
A_{2k+1} &= \det\, [\xi_{i+j+1} - \xi_{i+j}]_{i,j=1,\ldots,k}\;, \\
B_{2k} &= \det\, [-\xi_{i+j} + 2\xi_{i+j-1} - \xi_{i+j-2}]_{i,j=2,\ldots,k}\;, \\
B_{2k+1} &= \det\, [-\xi_{i+j+1} + 2\xi_{i+j} - \xi_{i+j-1}]_{i,j=1,\ldots,k}\;.
\end{aligned}$$

From Theorem 3 and equation (3) we can obtain the possible values of the L-moments of a distribution. In particular, for a nondegenerate distribution the constraints on ξ_1, ξ_2, ξ_3 and ξ_4 are that

$$\xi_2 - \xi_1 > 0, \qquad \xi_3 - \xi_2 > 0, \qquad -\xi_3 + 2\xi_2 - \xi_1 > 0,$$

$$(\xi_4 - \xi_3)(\xi_2 - \xi_1) - (\xi_3 - \xi_2)^2 \ge 0, \qquad -\xi_4 + 2\xi_3 - \xi_2 > 0,$$

and so the constraints on λ_1, λ_2, τ_3 and τ_4 are that

$$0 < \lambda_2, \qquad -1 < \tau_3 < 1, \qquad \tfrac{1}{4}(5\tau_3^2 - 1) \le \tau_4 < 1\,.$$

Here for good measure are the constraints on τ_5 and τ_6:

$$\tfrac{1}{5}\tau_3(7\tau_4-2)-\frac{7(1-\tau_4)(1+4\tau_4-5\tau_3^2)}{5(1+\tau_3)}\;\le\;\tau_5\;\le\;\tfrac{1}{5}\tau_3(7\tau_4-2)+\frac{7(1-\tau_4)(1+4\tau_4-5\tau_3^2)}{5(1+\tau_3)}\,,$$

$$\tfrac{1}{25}(42\tau_4^2-14\tau_4-3)+\frac{6(2\tau_3-7\tau_3\tau_4+5\tau_5)^2}{35(1+4\tau_4-5\tau_3^2)}\;\le\;\tau_6\;\le\;\tfrac{1}{10}(3+7\tau_4)-\frac{15(\tau_3-\tau_5)^2}{14(1-\tau_4)}\,.$$

Equality in these bounds can be attained only by a distribution which can take a finite number (m say) of distinct values. Such a distribution satisfies the lower bound on τ_{2m} and both the lower and upper bounds on τ_r, $r>2m$.

L-moments as measures of distributional shape

Oja (1981), extending work of Bickel and Lehmann (1975, 1976) and van Zwet (1964), has defined intuitively reasonable criteria for one probability distribution on the real line to be located further to the right (more dispersed, more skew, more kurtotic) than another. A real-valued functional of a distribution that preserves the partial ordering of distributions implied by these criteria may then reasonably be called a "measure of location" (dispersion, skewness, kurtosis). The following theorem shows that τ_3 and τ_4 are, by Oja's criteria, measures of skewness and kurtosis respectively.

Theorem 4. Let X and Y be real-valued random variables with cumulative distribution functions F and G respectively, and L-moments $\lambda_r^{(X)}$ and $\lambda_r^{(Y)}$ and L-moment ratios $\tau_r^{(X)}$ and $\tau_r^{(Y)}$ respectively.

(i) If $Y=aX+b$, then $\lambda_1^{(Y)}=a\lambda_1^{(X)}+b$, $\lambda_2^{(Y)}=|a|\lambda_1^{(X)}$, $\tau_3^{(Y)}=\tau_3^{(X)}$ and $\tau_4^{(Y)}=\tau_4^{(X)}$.

(ii) Let $\Delta(x)=G^{-1}(F(x))-x$. If $\Delta(x)\ge 0$ for all x, then $\lambda_1^{(Y)}\ge\lambda_1^{(X)}$. If $\Delta(x)$ is an increasing function of x, then $\lambda_2^{(Y)}\ge\lambda_2^{(X)}$. If $\Delta(x)$ is convex, then $\tau_3^{(Y)}\ge\tau_3^{(X)}$. If X and Y are symmetric and $\Delta(x)$ is convex of order 3 (Oja, 1981, Definition 2.1), then $\tau_4^{(Y)}\ge\tau_4^{(X)}$.

Proof. Part (i) is trivial. Part (ii) was proved by Oja (1981) for λ_1 and λ_2, in Oja's notation $\mu_1(F)$ and $\tfrac{1}{2}\sigma_1(F)$ respectively.

For τ_3, assume that X and Y are continuous, with probability density functions f and g respectively, and let $r(x)=f(x)/g\{G^{-1}(F(x))\}$: because $\Delta(x)$ is convex, $r(x)=d\Delta(x)/dx+1$ is increasing. Now by (4) and the substitution $y=G^{-1}(F(x))$,

$$\lambda_2^{(Y)}=\int G(y)\{1-G(y)\}\,dy=\int F(x)\{1-F(x)\}\,r(x)\,dx\,,$$

and similarly

$$\lambda_3^{(Y)}=\int F(x)\{1-F(x)\}\,\{2F(x)-1\}\,r(x)\,dx\,.$$

Thus $\lambda_2^{(X)}\lambda_2^{(Y)}\{\tau_3^{(Y)} - \tau_3^{(X)}\} = \lambda_3^{(Y)}\lambda_2^{(X)} - \lambda_3^{(X)}\lambda_2^{(Y)}$ may be written as

$$\int F(1-F)(2F-1)r \,.\int F(1-F) - \int F(1-F)(2F-1)\,.\int F(1-F)r\,, \qquad (5)$$

wherein $F(1-F)$ is a positive function of x and $2F-1$ and r are increasing. Chebyshev's inequality for integrals (Mitrinović, 1970, p. 40, Theorem 10) implies that (5) is positive. Because $\lambda_2^{(X)}\lambda_2^{(Y)} > 0$ it follows that $\tau_3^{(Y)} \geq \tau_3^{(X)}$. A discrete random variable can be approximated arbitrarily closely by a continuous random variable, so the result is also valid for discrete random variables.

The proof for τ_4 is similar.

Approximating a quantile function

Sillitto (1969) derived L-moments, without so naming them, as coefficients in the approximation of a quantile function by polynomials. As a matter of taste, we prefer to regard (1) as the fundamental definition; the approximation to the quantile function then becomes an inversion theorem, expressing the quantile function in terms of the L-moments.

Theorem 5 (Sillitto, 1969). Let X be a real-valued continuous random variable with finite variance, quantile function $x(F)$ and L-moments λ_r, $r \geq 1$. Then the representation

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

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

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

Proof. We seek an approximation to the quantile function $x(F)$ of the form

$$x(F) \approx \sum_{r=1}^{s} a_r\,P^*_{r-1}(F)\,, \qquad 0 < F < 1\,. \qquad (6)$$

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 constant weight function. To determine the a_r in (6) we denote the error of the approximation (6) 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

$$\begin{aligned}\int_0^1 \{R_s(F)\}^2 dF &= \int_0^1 \{x(F)\}^2 dF - 2\sum_{r=1}^{s} a_r \int_0^1 x(F) P^*_{r-1}(F) dF \\ &\quad + \sum_{r=1}^{s} \sum_{t=1}^{s} a_r \, a_t \int_0^1 P^*_{r-1}(F) \, P^*_{t-1}(F) \, dF \\ &= \int_0^1 \{x(F)\}^2 dF - 2\sum_{r=1}^{s} a_r \, \lambda_r + \sum_{r=1}^{s} a_r^2/(2r-1)\end{aligned}$$

(where we have used the orthogonality results $\int_0^1 P^*_r(F) P^*_s(F) dF = 0$ if $r \neq s$, $\int_0^1 \{P^*_r(F)\}^2 dF = 1/(2r+1)$), which is minimized by choosing

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

That

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

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).

References

Akhiezer, N. I. (1965). *The classical moment problem.* New York: Hafner.

Bickel, P. J., and Lehmann, E. L. (1975). Descriptive statistics for nonparametric models. I. Introduction. II. Location. *Annals of Statistics*, **3**, 1038–1069.

Bickel, P. J., and Lehmann, E. L. (1976). Descriptive statistics for nonparametric models. III. Dispersion. *Annals of Statistics*, **4**, 1139–1158.

Chan, L. K. (1967). On a characterization of distributions by expected values of extreme order statistics. *American Mathematical Monthly*, **74**, 950–951.

David, H. A. (1981). *Order statistics*, 2nd ed. New York: Wiley.

Feller, W. (1970). *An introduction to probability theory and its applications, vol. 2.* New York: Wiley.

Hausdorff, F. (1923). Momentenprobleme für ein endliches Intervall. *Mathematische Zeitschrift*, **16**, 220–248.

Hosking, J. R. M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B*, **52**, 105–124.

Konheim, A. G. (1971). A note on order statistics. *American Mathematical Monthly*, **78**, 524.

Lanczos, C. (1957). *Applied analysis.* London: Pitman.

Mallows, C. L. (1973). Bounds on distribution functions in terms of expectations of order-statistics. *Annals of Probability*, **1**, 297–303.

Mitrinović, D. S. (1970). *Analytic inequalities.* Berlin: Springer.

Oja, H. (1981). On location, scale, skewness and kurtosis of univariate distributions. *Scandinavian Journal of Statistics*, **8**, 154–168.

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.

Szegö, G. (1959). *Orthogonal polynomials*, 2nd ed. New York: American Mathematical Society.

Titchmarsh, E. C. (1946). *Eigenfunction expansions.* New York: Oxford University Press.

Van Zwet, W. R. (1964). Convex transformations of random variables. *Tract 7*, Mathematisch Centrum, Amsterdam.