RC22960 (W0311-006) November 3, 2003
Mathematics

IBM Research Report

Distributions with Maximum Entropy Subject to Constraints on Their *L*-moments or Expected Order Statistics

J. R. M. Hosking
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Distributions with maximum entropy subject to constraints on their *L*-moments or expected order statistics

J. R. M. Hosking

IBM Research Division, T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598

Abstract. We find the distribution that has maximum entropy conditional on having specified values of its first r L-moments. This condition is equivalent to specifying the expected values of the order statistics of a sample of size r. The maximum-entropy distribution has a density-quantile function, the reciprocal of the derivative of the quantile function, that is a polynomial of degree r; the quantile function of the distribution can then be found by integration. This class of maximum-entropy distributions includes the uniform, exponential and logistic, and two new generalizations of the logistic distribution. It provides a new method of nonparametric fitting of a distribution to a data sample. We also derive maximum-entropy distributions subject to constraints on expected values of linear combinations of order statistics.

1. Introduction

The entropy of a continuous probability distribution on the real line with cumulative distribution function $F(x)$ and probability density function $f(x) = dF(x)/dx$ is

$$H = \int_{-\infty}^{\infty} \{-\log f(x)\} f(x) dx \,. \tag{1.1}$$

We restrict attention to distributions whose cumulative distribution functions F are continuous and differentiable, with densities f that are nonzero within the range of the distribution, i.e. $f(x) > 0$ when $0 < F(x) < 1$. Denote this class of distributions by $\mathcal{D}$. A distribution of this type has a quantile function Q, the inverse of the cumulative distribution function, defined by $F(Q(u)) = u$, $0 < u < 1$; the quantile function is continuous and differentiable on $(0,1)$, and $Q'(u) = 1/f(Q(u))$. The function $f(Q(u))$ is known as the density-quantile function (see, e.g., Parzen, 1979). By making the substitution $x = Q(u)$ in (1.1), the entropy can be written in terms of the quantile function as

$$H = \int_0^1 \log Q'(u) du \,. \tag{1.2}$$

L-moments (Hosking, 1990) are measures of location, scale and shape of probability distributions. The rth L-moment of a random variable X with cumulative distribution function F and quantile function Q is

$$\lambda_r = \mathrm{E}[X P^*_{r-1}\{F(X)\}] = \int_0^1 P^*_{r-1}(u) Q(u) du \,, \tag{1.3}$$

where $P^*_r(\,.\,)$ is the rth shifted Legendre polynomial,

$$P^*_r(u) = \sum_{k=0}^{r} (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} u^k \,.$$

In particular, λ_1 is the mean, a location measure, and λ_2 is a scale measure. The dimensionless L-moment ratios $\tau_3 = \lambda_3/\lambda_2$ and $\tau_4 = \lambda_4/\lambda_2$ are measures of skewness and kurtosis, respectively. L-moments can be used as summary statistics for data samples, and to identify probability distributions and fit them to data. A brief description is given in Hosking (1998). In recent work, Karvanen et al. (2002) used L-moments for fitting distributions in independent component analysis in sognal processing, and Jones and Balakrishnan (2002) pointed out some relationships between integrals occurring in the definition of moments and L-moments.

L-moments are related to expected values of order statistics. The order statistic $X_{j:n}$, a random variable distributed as the jth smallest element of a random sample drawn from the distribution of X, has expected value

$$\mathrm{E}\, X_{j:n} = \frac{n!}{(j-1)!(n-j)!} \int_0^1 u^{j-1}(1-u)^{n-j} Q(u) du \,. \tag{1.4}$$

We have (Hosking, 1990)

$$\begin{aligned}
\lambda_1 &= \mathrm{E}(X_{1:1}), \\
\lambda_2 &= \tfrac{1}{2}\, \mathrm{E}(X_{2:2} - X_{1:2}), \\
\lambda_3 &= \tfrac{1}{3}\, \mathrm{E}(X_{3:3} - 2X_{2:3} + X_{1:3}), \\
\lambda_4 &= \tfrac{1}{4}\, \mathrm{E}(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}),
\end{aligned}$$

and in general

$$\lambda_r = r^{-1} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} \mathrm{E}(X_{r-j\,:\,r}) \,.$$

The main problem considered in this paper is the derivation of the distribution that has maximum entropy conditional on having specified values of its first r L-moments. This condition is equivalent to specifying the expected values of the order statistics of a sample of size r. We will show that this maximum-entropy distribution has a density-quantile function that is a polynomial $Z(u)$ of degree r. We call this the PDQ (for "polynomial density-quantile") distribution. The quantile function of the distribution can be found by integrating its derivative $1/Z(v)$.

Some special cases of the PDQ distribution are of interest. Two are well known: on a finite interval, the maximum-entropy distribution is the uniform distribution; on a semi-infinite interval, the maximum-entropy distribution with specified first L-moment (or equivalently, specified mean) is the exponential distribution. We can now add a third: on an infinite interval, the maximum-entropy distribution with specified first two L-moments is the logistic distribution (proved separately by Hosking, 2000). The "maximum entropy Lorenz curves" of Holm (1993) can be interpreted as maximum-entropy probability distributions on a finite or semi-infinite interval, with specified values of the first two L-moments (Holm's conditions (8) and (9)). We shall also describe some other special cases of the PDQ distribution, obtained by deriving maximum-entropy distributions conditional on specifying L-moments of orders $\{1,2,3\}$ and $\{1,2,4\}$; these generate families of distributions that generalize the logistic distribution and may be useful for modelling data.

As well as being of independent mathematical interest, the PDQ distribution can be fitted to a data sample. It provides the maximum-entropy fit among distributions whose first r population L-moments are equal to the sample L-moments. Choice of the parameter r, the degree of the polynomial in the density-quantile function of the PDQ distribution, effectively controls the smoothness of the fitted distribution.

In solving the main problem, we shall consider it in a slightly more general form. The constraints (1.3) and (1.4) have the form

$$\int_0^1 J(u)Q(u)du = g \tag{1.5}$$

where J is a polynomial. We therefore consider the derivation of the distribution that has maximum entropy subject to constraints of the form (1.5). This will enable us to constrain the values of (almost) arbitrary subsets of L-moments, or of (almost) arbitrary sets of linear combinations of expected values of order statistics.

The structure of the paper is as follows. Section 2 derives the quantile function that maximizes (1.2) subject to a set of constraints of the form (1.5). This general solution is applied in Section 3 to constraints on L-moments and in Section 4 to constraints on linear combinations of expected values of order statistics. Section 5 contains some examples of nonparametric density estimation based on maximum entropy and L-moments. Section 6 indicates some further applications and extensions of our results.

2. Derivation of the maximum-entropy distribution

We consider the problem of finding the function Q that maximizes the entropy (1.2) subject to a set of constraints of the form (1.5). It is convenient to write constraints such as (1.5) in the form

$$\int_0^1 K(u)Q'(u)du = h\,. \tag{2.1}$$

Let $K(u) = \int_u^1 J(v)dv$; then integration by parts gives

$$\int_0^1 J(u)Q(u)du = \Big[-K(u)Q(u)\Big]_0^1 + \int_0^1 K(u)Q'(u)du\,. \tag{2.2}$$

If the lower endpoint of the distribution is finite, with $Q(0)=L$, the integrated term is $K(0)L$. If the lower endpoint of the distribution is infinite, we shall require that $K(0)=0$, i.e. that $\int_0^1 J(u)du=0$; since $\int_0^1 J(u)Q(u)du$ is finite, this ensures that $K(u)Q(u)\to 0$ as $u\to 0$, so the integrated term is zero. Thus a constraint of the form (1.5) can be written as (2.1), where $K(u)=\int_u^1 J(v)dv$ and $h=g-K(0)L$ or $h=g$ depending on whether the lower bound of the distribution is finite or infinite.

From constraints in the form (2.1) we can determine $Q'(u)$, as follows.

Theorem 2.1. Consider the problem

$$\text{Maximize} \int_0^1 \log\{Q'(u)\}du \tag{2.3}$$

$$\text{subject to} \int_0^1 K_s(u)Q'(u)du=h_s, \qquad s=1,\dots,S, \tag{2.4}$$

where the K_s are linearly independent polynomials, and the maximization is over functions $Q'(u)$ that are strictly positive on $(0,1)$.

If there exist constants a_s, $s=1,\dots,S$, that satisfy

$$\int_0^1 \frac{K_r(u)du}{\sum_{s=1}^S a_s K_s(u)} = h_r\,, \qquad r=1,\dots,S, \tag{2.5}$$

and

$$\sum_{s=1}^S a_s K_s(u) > 0\,, \qquad 0<u<1, \tag{2.6}$$

then the problem has the solution

$$Q'(u) = Q_0'(u) \equiv 1 \Big/ \sum_{s=1}^S a_s K_s(u)\,. \tag{2.7}$$

The solution is unique up to redefinition of $Q_0'(u)$ on a set of u values that has measure zero.

Proof. Let $\tilde a_1,\dots,\tilde a_S$ be arbitrary constants that satisfy (2.6). We have $\log x\le x-1$ for any x, with equality if and only if $x=1$. Thus, for any $u\in(0,1)$,

$$\log\Big\{\Big(\sum_{s=1}^S \tilde a_s K_s(u)\Big)Q'(u)\Big\} \le \sum_{s=1}^S \tilde a_s K_s(u)Q'(u) - 1\,, \tag{2.8}$$

with equality if and only if

$$\sum_{s=1}^{S} \tilde{a}_s K_s(u) Q'(u) = 1\,. \tag{2.9}$$

Rewriting (2.8), we have

$$\log Q'(u) \le -\log\Big\{\sum_{s=1}^{S} \tilde{a}_s K_s(u)\Big\} + \sum_{s=1}^{S} \tilde{a}_s \{K_s(u)Q'(u) - h_s\} + \sum_{s=1}^{S} \tilde{a}_s h_s - 1\,. \tag{2.10}$$

Integrating over $0 < u < 1$, we have

$$\begin{aligned} \int_0^1 \log Q'(u)du - \sum_{s=1}^{S} \tilde{a}_s \Big(\int_0^1 K_s(u)Q'(u)du - h_s\Big) \\ \le -\int_0^1 \log\Big\{\sum_{s=1}^{S} \tilde{a}_s K_s(u)\Big\}du + \sum_{s=1}^{S} \tilde{a}_s h_s - 1\,. \end{aligned} \tag{2.11}$$

Thus among functions $Q'(u)$ that satisfy the constraints (2.4), we have

$$\int_0^1 \log Q'(u)du \le -\int_0^1 \log\Big\{\sum_{s=1}^{S} \tilde{a}_s K_s(u)\Big\}du + \sum_{s=1}^{S} \tilde{a}_s h_s - 1 \tag{2.12}$$

for any constants $\tilde{a}_1, \ldots, \tilde{a}_S$ that satisfy (2.6). We write this as

$$H[Q'] \le D(\tilde{\mathbf{a}}) \tag{2.13}$$

for any $\tilde{\mathbf{a}} \in \mathcal{A}$, where $\tilde{\mathbf{a}} = (\tilde{a}_1, \ldots, \tilde{a}_S)$ and $\mathcal{A} = \{\tilde{\mathbf{a}} : \sum_{s=1}^{S} \tilde{a}_s K_s(u) > 0 \text{ for all } u \in (0,1)\}$.

Consider a particular $\mathbf{a} = (a_1, \ldots, a_S)$ that satisfies (2.5), and define $Q_0'(u)$ as in (2.7). It is straightforward to show that $Q_0'(u)$ satisfies the constraints (2.4) and attains equality in (2.12), i.e. that $H[Q_0'] = D(\mathbf{a})$. All other functions $Q'(u)$ that satisfy (2.4) have entropy that is bounded according to (2.12), and in particular they satisfy

$$H[Q'] \le D(\mathbf{a})\,. \tag{2.14}$$

Thus Q_0' achieves the maximum possible value of $H[Q']$ for any Q' that satisfies (2.4); Q_0' therefore solves the problem stated in the theorem.

Furthermore, equality in (2.11) is achieved essentially only if $Q'(u) = 1/\sum_s \tilde{a}_s K_s(u)$ for all $u \in (0,1)$. Though $Q'(u)$ may be different for particular values of u, such differences must not alter the values of any of the integrals in (2.3) or (2.4), and can therefore affect $Q'(u)$ only on a set of u values that has measure zero. Similarly, equality in (2.14) is attained only if $Q'(u) = Q_0'(u)$ except on a set of measure zero. Thus, apart from such changes, Q_0' is the unique solution to the problem.

Theorem 2.1 essentially defines $Q'(u)$, and shows that the density-quantile function $f(Q(u)) = 1/Q'(u)$ is a polynomial. One additional constraint is needed to determine $Q(u)$, and how this is done depends on the range of the distribution. Four cases must be considered.

If the range of the distribution is constrained to be the finite interval $[L, U]$, this constraint implies the two conditions $Q(0) = L$ and $Q(1) = U$. We rewrite these as $Q(0) = L$ and $\int_0^1 Q'(u)du = U - L$. The latter constraint is of the form (2.1) and can be added to the original set of constraints; the former constraint determines $Q(u)$ to be $Q(u) = L + \int_0^u Q'(v)dv$.

If the range of the distribution is the semi-infinite interval $[L, \infty)$, this provides the constraint $Q(0) = L$, and determines $Q(u)$ to be $Q(u) = L + \int_0^u Q'(v)dv$.

If the range of the distribution is the entire real line, the range provides no constraint on the location of the distribution. Neither does any constraint that has $\int_0^1 J(u)du = 0$, because $\int_0^1 J(u)Q(u)du$ is then invariant under the location shift $Q(u) \to Q(u) + c$. The entropy (1.2) is also invariant under a location shift. Thus if every constraint satisfies $\int_0^1 J(u)du = 0$, the maximum-entropy distribution can be determined only up to a location shift.

If instead exactly one constraint has $\int_0^1 J(u)du \neq 0$, the other constraints can be used to determine $Q'(u)$, using Theorem 2.1; the constraint $\int_0^1 J(u)Q(u)du = g \neq 0$ then serves to determine Q given Q', provided that this Q' is the derivative of a quantile function Q for which $\int_0^1 J(u)Q(u)du$ is finite. We cannot write $Q(u) = c + \int_0^u Q'(v)dv$ for any c, because the integral is infinite. Instead we write, for any $u \in (0, 1)$,

$$
\begin{aligned}
g - Q(u)\int_0^1 J(t)dt &= \int_0^1 J(t)\{Q(t) - Q(u)\}dt \\
&= \int_0^1 J(t)\int_u^t Q'(v)dv\,dt \\
&= \int_0^u \int_t^u -Q'(v)dv\,J(t)dt + \int_u^1 \int_u^t Q'(v)dv\,J(t)dt \\
&= -\int_0^u \int_0^v J(t)dt\;Q'(v)dv + \int_u^1 \int_v^1 J(t)dt\;Q'(v)dv
\end{aligned}
$$

$$= -\int_0^u \{K(0) - K(v)\} Q'(v)dv + \int_u^1 K(v)Q'(v)dv\,,$$

where, as before, $K(u) = \int_u^1 J(v)dv$; thus

$$Q(u) = \frac{1}{K(0)}\left[g + \int_0^u \{K(0) - K(v)\} Q'(v)dv - \int_u^1 K(v)Q'(v)dv\right]. \qquad (2.15)$$

If more than one constraint has $\int_0^1 J(u)du \neq 0$, the constraints can be redefined by subtraction so that only one of them has $\int_0^1 J(u)du \neq 0$.

The solution $Q(u)$ obtained by integrating $Q'(u)$ is determined up to changing the value of $Q(u)$ on a set of measure zero; given that we are restricting attention to continuous quantile functions, the solution is unique.

Putting the foregoing results together, we obtain a procedure for finding the maximum-entropy distribution subject to a set of constraints of the form (1.5). It has four variants, depending on the range of the distribution. We state the procedure in the form of a theorem.

Theorem 2.2. Consider the problem

$$\text{Maximize} \int_0^1 \log\{Q'(u)\}du \qquad (2.16)$$

$$\text{subject to} \int_0^1 J_r(u)Q(u)du = g_r, \qquad r = 1, \ldots, R, \qquad (2.17)$$

where the J_r are linearly independent polynomials, and the maximization is over quantile functions Q of distributions in the class $\mathcal{D}$ defined in Section 1. Suppose further that one of the following sets of additional constraints is to be satisfied:

(Case 0) $Q(0) = L$ and $Q(1) = U$;

(Case 1) $Q(0) = L$, $Q(1)$ unconstrained;

(Case 2a) no constraints on $Q(0)$ or $Q(1)$, with $\int_0^1 J_r(u)du = 0$ for all r;

(Case 2b) no constraints on $Q(0)$ or $Q(1)$, with $\int_0^1 J_r(u)du \neq 0$ for some r.

The problem is solved by the following procedures, provided that the equations (2.5) referred to below can be solved and, if applicable, that the integrals in (2.19) or (2.23) below are finite. The solution is unique except that in Case 2a the distribution is determined only up to a location shift.

Case 0:

1. Write the constraints in the form (2.4), by setting

$$K_r(u) = \int_u^1 J_r(v)dv, \qquad h_r = g_r - K_r(0)L, \qquad r = 1, \ldots, R. \tag{2.18}$$

2. Add the constraint $\int_0^1 Q'(u)du = U - L$, by defining $K_{R+1}(u) = 1$, $0 \le u \le 1$, and $h_{R+1} = U - L$.
3. Set $S = R + 1$ and solve equations (2.5).
4. The maximum-entropy distribution has Q' given by (2.7) and

$$Q(u) = L + \int_0^u \frac{dv}{\sum_{s=1}^S a_s K_s(v)} \,. \tag{2.19}$$

Case 1:

1. Write the constraints in the form (2.4), via (2.18).
2. Set $S = R$ and solve equations (2.5).
3. Provided that the integral in (2.19) exists, the maximum-entropy distribution has Q' given by (2.7) and Q given by (2.19).

Case 2a:

1. Write the constraints in the form (2.4), via (2.18).
2. Set $S = R$ and solve equations (2.5).
3. The maximum-entropy distribution has Q' given by (2.7). Q is determined only up to an additive constant, by

$$Q(u) = \int^u \frac{dv}{\sum_{s=1}^S a_s K_s(v)} \,. \tag{2.20}$$

Case 2b:

1. Without loss of generality, suppose that $\int_0^1 J_R(u)du \ne 0$.
2. For $r = 1, \ldots, R-1$, set $J_r^*(u) = J_r(u) - \alpha_r J_R(u)$ and $g_r^* = g_r - \alpha_r g_R$, where $\alpha_r = \int_0^1 J_r(u)du / \int_0^1 J_R(u)du$. The first $R-1$ constraints are equivalent to the new constraints

$$\int_0^1 J_r^*(u)Q(u)du = g_r^*, \qquad r = 1, \ldots, R-1, \tag{2.21}$$

for which we have $\int_0^1 J_r^*(u)du = 0$ for all r.

3. Write the new constraints in the form (2.4), by setting

$$K_r(u) = \int_u^1 J_r^*(v)dv, \qquad h_r = g_r^*, \qquad r = 1, \ldots, R-1. \tag{2.22}$$

4. Set $S = R-1$ and solve equations (2.5).
5. Provided that the integrals in (2.23) below exist, the maximum-entropy distribution has Q' given by (2.7) and quantile function given by

$$Q(u) = \frac{1}{K_R(0)} \left[g_R + \int_0^u \{K_R(0) - K_R(v)\} Q'(v)dv - \int_u^1 K_R(v)Q'(v)dv \right], \tag{2.23}$$

where $K_R(u) = \int_u^1 J_R(u)du$.

Remark 2.1. The restriction to class $\mathcal{D}$ is made for mathematical convenience: it ensures that $Q(u)$ is differentiable and enables us, when deriving (2.2), to write $\int K(u)Q'(u)du$ rather than merely $\int K(u)dQ(u)$. Distributions outside this class may have jumps in the quantile function, and, in consequence, constraints of the form (1.5) may not be expressible in the form (2.1).

Remark 2.2. The significance of the condition in Cases 1 and 2b, “provided that the integrals in (2.19) or (2.23) exist”, is illustrated in Examples 3.5 and 4.5 below.

Remark 2.3. The proof of Theorem 2.1 can be related to standard concepts in the theory of optimization. The left side of (2.11) is the Lagrangian for the optimization problem stated in Theorem 2.1, with the $\tilde{a}_s$ as the Lagrange multipliers; the right side of (2.11), and of (2.12), is the criterion function of the dual problem.

The dual problem provides a practical means of solving equations (2.5). The dual problem is to find the minimum of the function $D(\mathbf{a})$ over $\mathbf{a} \in \mathcal{A}$. It is straightforward to show (for details see Appendix, item 1) that the function D is convex and that its local minimum (being a convex function, it can have only one) is the solution of (2.5). Thus solving (2.5) reduces to finding the minimum of a convex function in S-dimensional Euclidean space. This can be achieved by standard iterative numerical methods, provided that a starting value $\mathbf{a} \in \mathcal{A}$ can be found. A simple condition, often satisfied in practice, that ensures the existence of such an $\mathbf{a}$ is that there should be (at least) one constraint that has $K_r(u) > 0$ for all $u \in (0,1)$ and $h_r > 0$. In this case one can take as a starting value the vector $\mathbf{a}$ that has $a_r = 1/h_r$ and $a_s = 0$ for $s \neq r$.

Remark 2.4. Theorems 2.1 and 2.2 provide only sufficient conditions for a solution to exist. This is generally adequate in practice, since the previous remark provides a way of finding the solution when it exists. When no solution can be found, there appear to be three possibilities. The constraints may be mutually inconsistent, i.e., such that no distribution satisfies them all; distributions may exist that satisfy the constraints and have arbitrarily high entropy (as in Example 3.6 below); or a solution for Q' may be found but cannot be translated into a solution for Q because the integrals in (2.19) or (2.23) do not exist (as in Examples 3.5 and 4.5).

Remark 2.5. The problem stated in Theorem 2.2 can also be approached by methods from the calculus of variations. The Euler-Lagrange equations (e.g., Troutman, 1983, Section 6.5) immediately show that the solution is of the form

$$\frac{d}{du}\left(\frac{1}{Q'(u)} - \sum_r \kappa_r J_r(u)\right) = 0\,,$$

where the κ_r are Lagrange multipliers. This implies that $1/Q'(u)$, the density-quantile function, is a polynomial. However, this approach establishes only that the solution gives a stationary value of the entropy. To show that this stationary value is a maximum is not straightforward in general, though the concavity of $\log Q'(u)$ as a function of $Q'(u)$ can be used in some cases. For example, Theorem (3.16) of Troutman (1983, p. 74) covers Case 0, in which $Q(u)$ is bounded. In the other cases, we would require an extension of Troutman's results to the situation in which the function Q has a continuous derivative not on the interval $[0,1]$ but merely on $(0,1)$.

3. Maximum entropy and *L*-moments

We now use Theorem 2.2 to find the distribution that has maximum entropy conditional on having specified values of its first R L-moments. First we write the constraints (1.3) in the form (2.1). Integration by parts gives

$$\lambda_1 - L = \int_0^1 (1-u)Q'(u)du \qquad \text{when the lower bound } L \text{ is finite,} \tag{3.1}$$

$$\lambda_2 = \int_0^1 u(1-u)Q'(u)du\,, \tag{3.2}$$

$$\lambda_3 = \int_0^1 u(1-u)(2u-1)Q'(u)du\,, \tag{3.3}$$

$$\lambda_4 = \int_0^1 u(1-u)(5u^2-5u+1)Q'(u)du\,, \tag{3.4}$$

and in general

$$\lambda_r = \int_0^1 Z_r(u)Q'(u)du\,, \qquad r \ge 2, \tag{3.5}$$

where $Z_r(u) = \int_u^1 P^*_{r-1}(v)dv$ is a polynomial of degree r.

Theorem 3.1. The distribution that has maximum entropy given specified values of its L-moments λ_r, $r = 1, \ldots, R$, is given by the following construction, provided that the equations (3.7) below have a solution. Denote by Cases 0, 1, and 2 the instances in which the range of the distribution is constrained to be the intervals $[L, U]$, $[L, \infty)$, and $(-\infty, \infty)$, respectively. Define

(in Case 0)	$Z_0(u) = 1,$	$k_0 = U - L;$	
(in Cases 0 and 1)	$Z_1(u) = 1 - u,$	$k_1 = \lambda_1 - L;$	
(in all Cases)	$Z_r(u) = \int_u^1 P^*_{r-1}(v)dv\,,$	$k_r = \lambda_r,$	$r \ge 2.$

In Case m ($m = 0$, 1, or 2), the maximum-entropy distribution has quantile function $Q(u)$ with derivative given by

$$Q'(u) = 1 \Big/ \sum_{r=m}^{R} a_r Z_r(u) \tag{3.6}$$

where the a_r satisfy the equations

$$\int_0^1 \frac{Z_r(u)du}{\sum_{s=m}^R a_s Z_s(u)} = k_r\,, \qquad r = m, \ldots, R, \tag{3.7}$$

with

$$\sum_{r=m}^{R} a_r Z_r(u) > 0\,, \qquad 0 < u < 1. \tag{3.8}$$

The quantile function itself is given, in Cases 0 and 1, by

$$Q(u) = L + \int_0^u Q'(v)dv \tag{3.9}$$

or, in Case 2, by

$$Q(u) = \lambda_1 + \int_0^u vQ'(v)dv - \int_u^1 (1-v)Q'(v)dv \tag{3.10}$$

for any $u \in (0,1)$.

Proof. The theorem is a restatement of Theorem 2.2, and follows immediately from it. In Case 1, because λ_1 is constrained, the integral in (3.9) is finite:

$$\begin{aligned}\int_0^u Q'(v)dv &\leq (1-u)^{-1}\int_0^u (1-v)Q'(v)dv \\ &\leq (1-u)^{-1}\int_0^1 (1-v)Q'(v)dv \\ &= (\lambda_1 - L)/(1-u) < \infty\,.\end{aligned}$$

Similarly in Case 2, because λ_2 is constrained the finiteness of the integrals in (3.10) is assured.

Remark 3.1. Since Z_r is a polynomial of degree r, it is clear from (3.6) that the maximum-entropy distribution has a density-quantile function $f(Q(u)) = 1/Q'(u)$ that is a polynomial of degree R. We call such a distribution a PDQ ("polynomial density-quantile") distribution.

Including the coefficients of the polynomial $f(Q(u))$ and the constant of integration that arises when integrating $Q'(u)$ to get $Q(u)$, the quantile function of the PDQ distribution has $R+2$ free parameters. These parameters are determined by the R constraints on the L-moments together with two further conditions that depend on the range of the distribution. If the range of the distribution is a finite interval $[L,U]$, then the quantile function must satisfy $Q(0)=L$ and $Q(1)=U$. If the range of the distribution is a semi-infinite interval, without loss of generality the interval $[L,\infty)$, then the quantile function must satisfy $Q(0)=L$ and the density-quantile function must satisfy $f(Q(1))=0$. If the range of the distribution can be the entire real line, then the density-quantile function must satisfy $f(Q(0))=f(Q(1))=0$.

Remark 3.2. If an endpoint of the distribution is infinite, the corresponding tail of the probability density of the maximum-entropy distribution decays exponentially.

Consider the upper tail, for example. Because $Q'(u)$ is the reciprocal of a polynomial, its behaviour as $u \to 1$ is $Q'(u) \sim c(1-u)^{-m}$ with m an integer, $m \geq 0$ and $c \neq 0$ — in fact $c > 0$, since $Q'(u) > 0$ for all $u \in (0,1)$. When the upper tail is infinite, then so is the integral $\int^1 Q'(u)du$, so we must have $m \geq 1$. The integrals (3.1) (in Case 1) or (3.2) (in Case 2) are finite, so we must have $m \leq 1$. Thus $m=1$, i.e. as $u \to 1$ we have $Q'(u) \sim c(1-u)^{-1}$ for some c, $0 < c < \infty$. Integrating the asymptotic equivalence, we have $Q(u) \sim b - c \log(1-u)$ as $u \to 1$, which upon substituting $u = F(x)$ gives $F(x) \sim 1 - e^{-(x-b)/c}$ as $x \to \infty$.

Remark 3.3. We can also write $Z_r(u)$, $r \geq 2$, as

$$Z_r(u) = \int_u^1 P^*_{r-1}(v) dv = \frac{u(1-u) P^{*\prime}_{r-1}(u)}{r(r-1)} \,. \tag{3.11}$$

The last equality follows from integrating the differential equation satisfied by shifted Legendre polynomials, which can be written as

$$r(r+1) P^*_r(u) + \frac{d}{du}\{u(1-u) P^{*\prime}_r(u)\} = 0 \tag{3.12}$$

(e.g., Sansone, 1959, p. 176, gives the corresponding result for "unshifted" Legendre polynomials).

Remark 3.4. The polynomials $Z_r(u)$, $r \geq 2$, are orthogonal on the interval $(0,1)$ with weight function $\{u(1-u)\}^{-1}$. To see this, take the orthogonality relation of the shifted Legendre polynomials, i.e.

$$\int_0^1 P^*_r(u) P^*_s(u) du = 0 \qquad \text{if } r \neq s, \tag{3.13}$$

and observe that for $r, s \geq 1$

$$\begin{aligned} \int_0^1 P^*_r(u) P^*_s(u) du &= \int_0^1 \Big(\int_u^1 P^*_r(v) dv \Big) P^{*\prime}_s(u) du && \text{by parts} \\ &= \int_0^1 Z_{r+1}(u) \,.\, s(s+1) \frac{Z_{s+1}(u)}{u(1-u)} du && \text{by (3.11)} \\ &= s(s+1) \int_0^1 \{u(1-u)\}^{-1} Z_{r+1}(u) Z_{s+1}(u) du \,. \end{aligned}$$

This gives another relation, in Case 2, between the coefficients a_r in (3.7) and the derivative of the quantile function of the maximum-entropy distribution:

$$\begin{aligned} \int_0^1 \frac{Z_r(u) du}{u(1-u) Q'(u)} &= \int_0^1 \{u(1-u)\}^{-1} \sum_{s=2}^R a_s Z_s(u) Z_r(u) du \\ &= \sum_{s=2}^R a_s \int_0^1 \{u(1-u)\}^{-1} Z_r(u) Z_s(u) du \\ &= a_r \int_0^1 \{u(1-u)\}^{-1} \{Z_r(u)\}^2 du \,. \end{aligned}$$

The final integral can be evaluated explicitly; after a little algebra we obtain

$$a_r = r(r-1)(2r-1)\int_0^1 \frac{Z_r(u)du}{u(1-u)Q'(u)}\,.$$

This orthogonality relation also ensures, in Case 2, that if the constants $a_2, \ldots, a_R$ satisfy (3.8), then $a_2 > 0$. For since $Z_2(u) = u(1-u)$, we have

$$\begin{aligned} 0 < \int_0^1 \sum_{r=2}^R a_r Z_r(u)du &= \int_0^1 \{u(1-u)\}^{-1} Z_2(u) \sum_{r=2}^R a_r Z_r(u)du \\ &= \sum_{r=2}^R a_r \int_0^1 \{u(1-u)\}^{-1} Z_2(u) Z_r(u)du \\ &= a_2 \int_0^1 u(1-u)du \qquad \text{by orthogonality;} \end{aligned}$$

the final integral is positive, whence $a_2 > 0$.

Remark 3.5. A similar result to Theorem 3.1 holds when the L-moments that are constrained are of degrees $r_1, r_2, \ldots, r_R$ rather than $1, 2, \ldots, R$. We require in Case 1 that λ_1 be constrained and in Case 2 that λ_2 be constrained; otherwise no maximum-entropy distribution need exist, as in Examples 3.5 and 3.6 below.

We now give some examples of distributions that have maximum entropy subject to constraints on their L-moments. In each example we give the range of the distribution and the L-moments whose values are constrained.

Example 3.1. Range $[L, U]$; no constraints on L-moments.

Though there are no constraints on L-moments, because the range of the distribution is finite we introduce the constraint $\int_0^1 Q'(u)du = U - L$. Equation (3.6) becomes $Q'(u) = 1/a_0$ and (3.7) is the single equation $1/a_0 = U - L$. Thus $Q'(u) = U - L$ and $Q(u) = L + (U - L)u$, so the maximum-entropy distribution is uniform on the interval $[L, U]$. This is of course a well known result.

Example 3.2. Range $[L, \infty)$; constrain λ_1.

As in the previous example there is one constraint, which is now (3.1), and one equation in the set (3.7). The solution is $Q'(u) = (\lambda_1 - L)/(1-u)$, which can be integrated to give $Q(u) = L - (\lambda_1 - L)\log(1-u)$; thus the maximum-entropy distribution is an exponential distribution with lower bound L and mean λ_1. This too is a well known result.

Example 3.3. Range $(-\infty, \infty)$; constrain λ_1, λ_2.

Again (3.7) consists of a single equation, based on the constraint (3.2). The maximum-entropy solution is $Q'(u) = \lambda_2/\{u(1-u)\}$, and from (3.10) we obtain $Q(u) = \lambda_1 + \lambda_2 \log\{u/(1-u)\}$. This is the quantile function of a logistic distribution: the maximum-entropy distribution is a logistic distribution whose location and scale parameters are chosen to agree with the specified L-moments. This is a limiting case of the results of Holm (1993).

In the corresponding problem when only λ_2 is constrained, the solution is a logistic distribution with undetermined location parameter. This result has also been proved by Hosking (2000).

Example 3.4. Range $[0, \infty)$; constrain λ_1, λ_2.

The solution is, from (3.6),

$$Q'(u) = \frac{1}{a_1(1-u) + a_2 u(1-u)} = \frac{1}{(1-u)(a_1 + a_2 u)}\,, \tag{3.14}$$

with, from (3.7),

$$\lambda_1 = \int_0^1 \frac{du}{(a_1 + a_2 u)} = \frac{1}{a_2} \log\left(\frac{a_1 + a_2}{a_1}\right),$$
$$\lambda_2 = \int_0^1 \frac{u\,du}{(a_1 + a_2 u)} = \frac{1}{a_2} - \frac{a_1}{a_2^2} \log\left(\frac{a_1 + a_2}{a_1}\right).$$

Writing $\beta = a_2/a_1$ we have

$$\lambda_2/\lambda_1 = 1/\log(1+\beta) - 1/\beta \equiv g(\beta)\,, \quad \text{say.} \tag{3.15}$$

The function $g(\beta)$, with $g(0)$ defined to be $\frac{1}{2}$ to maintain continuity at $\beta = 0$, is a continuous monotonic function that decreases from 1 at $\beta = -1$ to 0 as $\beta \to \infty$. Thus provided that $0 < \lambda_2 < \lambda_1$ — precisely the conditions that λ_1 and λ_2 must satisfy to be the first two L-moments of a nondegenerate distribution on $[0, \infty)$ (Hosking, 1990, Theorem 2) — equation (3.15) has a unique solution with $-1 < \beta < \infty$, from which we obtain

$$a_2 = \log(1+\beta)/\lambda_1, \qquad a_1 = \log(1+\beta)/(\beta\lambda_1)\,. \tag{3.16}$$

We can integrate (3.14) to get

$$Q(u) = \frac{1}{(a_1 + a_2)} \log\left(\frac{1 + (a_2/a_1)u}{1 - u}\right). \tag{3.17}$$

For $a_2 > 0$, corresponding to $\lambda_2/\lambda_1 > 1/2$, this is the quantile function of a logistic distribution truncated on the left.

This result has previously been obtained by Holm (1993, p. 388, the case "$r_2 \to 1+$"; Holm's r is our $-a_1/a_2$).

Example 3.5. Range $[0, \infty)$; constrain λ_2.

This problem is excluded from the ambit of Theorem 3.1, because λ_1 is not constrained. In attempting to use Theorem 2.2 we find that equations (2.5)–(2.7) have the solution $Q'(u) = \lambda_2/\{u(1-u)\}$, as in Example 3.3, but now the integral in (2.19) does not exist. Thus no maximum-entropy distribution can be found by the methods of Theorems 2.2 or 3.1.

To understand why no maximum-entropy distribution can be found, consider the previous example, in which λ_1 is also constrained. The maximum value of the entropy is

$$\begin{aligned}
\int_0^1 \log Q'(u)du &= -\int_0^1 \log(1-u)du - \int_0^1 \log(a_1 + a_2 u)du \\
&= 2 + \frac{a_1}{a_2}\log a_1 - \frac{1}{a_2}(a_1 + a_2)\log(a_1 + a_2) \\
&= 2 + \log\lambda_2 + 2\log\beta - \frac{1}{\beta}(1+\beta)\log(1+\beta) - \log\{\beta - \log(1+\beta)\},
\end{aligned}$$

where the last equality follows from expressing a_1 and a_2 in terms of β and λ_2, using (3.16) and (3.15). As $\lambda_1 \to \infty$ with λ_2 fixed, i.e. as $\beta \to \infty$ with λ_2 fixed, the entropy increases monotonically and approaches the limit $2 + \log\lambda_2$ (for details see Appendix, item 2). However, this limit is not attained by any distribution with the specified value of λ_2 and a finite lower bound, so no maximum-entropy distribution exists within this class of distributions.

Example 3.6. Range $[L, \infty)$; constrain $\lambda_3, \lambda_4, \ldots, \lambda_R$.

The same argument as at the end of Remark 3.4 shows that there is no set of constants a_r such that $\sum_{r=3}^{R} a_r Z_r(u) > 0$ for all $u \in (0,1)$. Thus (3.8) cannot be satisfied and no maximum-entropy distribution can be obtained using Theorem 3.1.

The reason that no maximum-entropy distribution can be found is that there exist distributions with lower bound L and the specified values of $\lambda_3, \lambda_4, \ldots, \lambda_R$ but arbitrarily large entropy. For an explicit construction of such a distribution, consider the function $\sum_{r=3}^{R}(2r-1)\lambda_r P^*_{r-1}(u)$. It is a polynomial, so its derivative on $(0,1)$ is bounded below, by M say. Now choose

$$\lambda_2 > |M|/6, \qquad \lambda_1 = L + \sum_{r=2}^{R}(-1)^r(2r-1)\lambda_r\,, \tag{3.18}$$

and define

$$Q_1(u) = \lambda_1 + \sum_{r=2}^{R}(2r-1)\lambda_r P^*_{r-1}(u)\,.$$

Note that, since $P^*_1(u) = 2u-1$,

$$Q_1'(u) = 6\lambda_2 + \sum_{r=3}^{R}(2r-1)\lambda_r P^{*\prime}_{r-1}(u) \geq 6\lambda_2 + M > 0\,;$$

thus the function Q_1 is increasing on $(0,1)$, and is therefore the quantile function of some probability distribution. Because the shifted Legendre polynomials satisfy $P^*_r(0) = (-1)^r$, we have $Q_1(0) = L$, so the distribution has lower bound L. By the orthogonality relation (3.13) and the result

$$\int_0^1 \{P^*_r(u)\}^2 du = 1/(2r+1) \tag{3.19}$$

(Sansone, 1959, p. 245), we see that $\int_0^1 P^*_{r-1}(u)Q_1(u)du = \lambda_r$, i.e. the distribution with quantile function Q_1 has L-moments λ_r, $r = 1, \ldots, R$. The entropy of the distribution is

$$\begin{aligned}\int_0^1 \log Q_1'(u)du &= \int_0^1 \log\Big(6\lambda_2 + \sum_{r=3}^{R}(2r-1)\lambda_r P^{*\prime}_{r-1}(u)\Big)du \\ &\geq \int_0^1 \log(6\lambda_2 + M)du \\ &= \log(6\lambda_2 + M)\end{aligned}$$

and can be made arbitrarily large by letting $\lambda_2 \to \infty$.

When the range of the distribution is $(-\infty, \infty)$ and λ_2 is not constrained, we can also find distributions that satisfy the constraints and have arbitrarily large entropy.

The construction is identical to that given above, except that λ_1 may be specified in the set of constraints rather than being defined as in (3.18).

Example 3.7. Range $(-\infty, \infty)$; constrain λ_1, λ_2, λ_3.

With the given constraints (3.2) and (3.3), the solution (3.6) is

$$Q'(u) = \frac{1}{u(1-u)\{a_2 + a_3(2u-1)\}} \,. \tag{3.20}$$

This can be integrated (for details see Appendix, item 3): writing $\gamma = -a_3/a_2$, we find that the quantile function has the form

$$Q(u) = \xi + \alpha \left\{ \log\left(\frac{u}{1-u}\right) + \gamma \log\left(\frac{\{1-\gamma(2u-1)\}^2}{4u(1-u)}\right) \right\} , \tag{3.21}$$

where ξ, α and γ are constants, related to the L-moments by

$$\tau_3 = \lambda_3/\lambda_2 = \frac{1}{\gamma} - \frac{1}{\operatorname{artanh}\gamma} \,, \tag{3.22}$$

$$\alpha = \frac{\lambda_2(1-\gamma\tau_3)}{(1-\gamma^2)} \,, \tag{3.23}$$

$$\xi = \lambda_1 - \alpha\{(1+\gamma)\log(1+\gamma) - (1-\gamma)\log(1-\gamma) - \gamma \log 4\} \,. \tag{3.24}$$

The L-moment ratio τ_3 is the skewness measure based on L-moments; for nondegenerate distributions its valid values are $-1 < \tau_3 < 1$ and those of λ_2 are $\lambda_2 > 0$ (Hosking, 1990). The right side of (3.22) is a function of γ that increases monotonically from -1 as $\gamma \to -1$ to $+1$ as $\gamma \to +1$. Thus for any valid τ_3 we can find a unique value of γ; furthermore this value satisfies $-1 < \gamma < +1$. Provided that $\lambda_2 > 0$ we also obtain $\alpha > 0$. Thus provided that the specified values of the L-moments are consistent with some nondegenerate probability distribution, the maximum-entropy solution (3.21) will have $\alpha > 0$ and $-1 < \gamma < +1$; this ensures that $Q'(u) > 0$ for all $u \in (0,1)$, i.e. that the solution satisfies (3.8).

The quantile function (3.21) can also be thought of as defining a family of probability distributions with three parameters: ξ is a location parameter, α is a scale parameter, and γ is a shape parameter taking values in $(-1, +1)$. The family generalizes the logistic distribution, which is the special case $\gamma = 0$, and contains both negatively skew ($\gamma < 0$) and positively skew ($\gamma > 0$) distributions. It is, however, not the same as the "generalized logistic" distribution used by Hosking (1996) and Hosking and Wallis (1997). The distributions have infinite range in both directions, with exponentially decreasing tails. The cumulative distribution function $F(x)$ and probability density function $f(x)$ do not have explicit forms, but can be computed numerically: for given x, $F(x)$ is the solution of $Q(F(x))=x$ and $f(x)=1/Q'(F(x))$. Because the distributions have density-quantile functions that are polynomials of degree 3, we call this the PDQ3 family.

We can use (3.4) to compute the fourth-order L-moment λ_4 of the PDQ3 distribution, and the kurtosis measure $\tau_4=\lambda_4/\lambda_2$. We find that $\tau_4=(5\tau_3/\gamma-1)/4$ (for details see Appendix, item 4). The values of τ_3 and τ_4, as γ varies, can be plotted on an L-moment ratio diagram. Figure 1 is such a plot, which compares the τ_3–τ_4 relations for several different families of distributions. It shows that for a given value of τ_3, the PDQ3 distribution has the largest value of τ_4 of any of these distributions.

The PDQ3 family is potentially useful for modelling data that may have been sampled from skew distributions with exponentially decreasing tails. Sample L-moments can be computed from the data (Hosking, 1990, 1996); using these sample L-moments in place of the population L-moments in equations (3.22)–(3.24) provides estimates of the parameters.

As an example of the PDQ3 distribution we consider the case $\lambda_1=0.5772$, $\lambda_2=\log 2=0.6931$, $\lambda_3=\log(9/8)=0.1178$. These are the first three L-moments of a Gumbel (extreme-value type I) distribution. The corresponding parameters of (3.21) are $\gamma=0.4768$, $\alpha=0.8244$, $\xi=0.3680$. The entropies of the two distributions are 1.5772 (Gumbel) and 1.5898 (PDQ3). The two distributions are compared in Figure 2. Both distributions have exponentially decreasing upper tails, and the quantiles of the two distributions diverge only gradually in the upper tail. The lower tail of the PDQ distribution is also exponential, but the lower tail of the Gumbel distribution is much lighter and the quantiles of the two distributions diverge sharply in the lower tail.

Example 3.8. Range $(-\infty,\infty)$; constrain λ_1, λ_2, λ_4.

With the given constraints (3.2) and (3.4), the solution (3.6) is

$$Q'(u)=\frac{1}{u(1-u)\{a_2+a_4(5u^2-5u+1)\}}\,. \tag{3.25}$$

This can be explicitly integrated; after some algebra (for details see Appendix, item 5) we find that the quantile function can be written in one of the forms

$$Q(u) = \xi + \alpha\left[\log\left(\frac{u}{1-u}\right) - 2\delta\,\mathrm{artanh}\{\delta(2u-1)\}\right] \quad \text{with } 0 < \delta < 1 \tag{3.26}$$

or

$$Q(u) = \xi + \alpha\left[\log\left(\frac{u}{1-u}\right) + 2\delta\arctan\{\delta(2u-1)\}\right] \quad \text{with } -\infty < \delta < 0, \tag{3.27}$$

where ξ, α and δ are constants. The value of δ is related to $\tau_4 = \lambda_4/\lambda_2$, the L-moment kurtosis measure, which takes values in $(-\frac{1}{4}, +1)$ for continuous distributions (Hosking, 1990). If $\tau_4 = \frac{1}{6}$ we have $\delta = 0$ and the quantile function (3.26) reduces to that of the logistic distribution. Otherwise δ and τ_4 are related by

$$\tau_4 = \begin{cases} -\dfrac{1}{4} + \dfrac{5}{4\delta}\left(\dfrac{1}{\delta} - \dfrac{1}{\mathrm{artanh}\,\delta}\right) & \text{if } \frac{1}{6} \le \tau_4 < 1, \\[2ex] -\dfrac{1}{4} - \dfrac{5}{4\delta}\left(\dfrac{1}{\delta} - \dfrac{1}{\arctan\delta}\right) & \text{if } -\frac{1}{4} < \tau_4 < \frac{1}{6}. \end{cases} \tag{3.28}$$

The right sides of these equations are monotonically increasing functions of δ, and for given τ_4 we can find a unique solution with $0 < \delta < 1$ when $\frac{1}{6} < \tau_4 < 1$ and $-\infty < \delta < 0$ when $-\frac{1}{4} < \tau_4 < \frac{1}{6}$. Given δ, the other constants are related to the first two L-moments:

$$\alpha = \begin{cases} \dfrac{\lambda_2\delta}{(1-\delta^2)\,\mathrm{artanh}\,\delta} & \text{if } \delta > 0, \\[2ex] \dfrac{\lambda_2\delta}{(1+\delta^2)\arctan\delta} & \text{if } \delta < 0, \end{cases} \tag{3.29}$$

$$\xi = \lambda_1. \tag{3.30}$$

As in the previous example, the quantile functions (3.26)–(3.27) can also be thought of as defining a family of PDQ probability distributions with three parameters: ξ is a location parameter, α is a scale parameter, and δ is a shape parameter taking values in $(-\infty, +1)$. The family generalizes the logistic distribution, which is the special case $\delta = 0$, and contains distributions both lighter-tailed ($\delta < 0$) and heavier-tailed ($\delta > 0$) than the logistic. The distributions are symmetric about ξ and have infinite range in both directions, with exponentially decreasing tails; they are potentially useful for modelling data drawn from symmetric distributions with this tail behaviour.

As an example of the distribution we consider the case $\lambda_1 = 0$, $\lambda_2 = \pi^{-1/2} = 0.5642$, $\tau_4 = 30\pi^{-1} \arctan \sqrt{2} - 9 = 0.1226$. These are the L-moments of a standard normal distribution. The corresponding parameters of (3.27) are $\delta = -0.7029$, $\alpha = 0.4332$, $\xi = 0$. The entropies of the two distributions are 1.4189 (Normal) and 1.4203 (PDQ). The two distributions are compared in Figure 3. Their probability density functions and quantile functions are very similar except in the extreme tails: for exceedance probabilities between 0.011 and 0.989, corresponding to normal variate values between -2.3 and $+2.3$, the quantiles of the two distributions differ by less than 0.03. However, the PDQ distribution has heavier tails and its quantiles ultimately increase much faster than those of the normal distribution.

4. Maximum entropy and order statistics

It is clear from (1.4) that constraints on expected values of order statistics, or of linear combinations thereof, can be put in the form (2.17). Thus Theorem 2.2 can be applied to the problem of finding the distribution that has maximum entropy conditional on having specified values of linear combinations of expected values of its order statistics. We now give some examples of maximum-entropy distributions that can be obtained by this approach.

Example 4.1. Constraints $\mathrm{E}\, X_{j:n} = \xi_j$, $j = 1, \ldots, n$.

Both this problem and that in which the first n L-moments are specified give rise to constraints (2.17) in which the J_r are a set of n linearly independent polynomials of degree at most n. Either set of constraints can be transformed into the other, and the two problems are equivalent. Thus the distribution that has maximum entropy conditional on having specified values of its expected order statistics from a sample of size n is a PDQ distribution in which the density-quantile function is a polynomial of degree n.

Example 4.2. Range $(-\infty, \infty)$; constraint $\mathrm{E}(X_{n:n} - X_{1:n}) = \xi$.

Here we derive the distribution that has maximum entropy conditional on having a specified value of the expected range of a sample of size n. The constraint can be written, using (1.4), as

$$\int_0^1 n\{u^{n-1} - (1-u)^{n-1}\} Q(u) du = \xi \,,$$

or, integrating by parts, as

$$\int_0^1 \{1 - u^n - (1-u)^n\} Q'(u)\,du = \xi\,.$$

Thus the maximum-entropy distribution has

$$Q'(u) = \xi / \{1 - u^n - (1-u)^n\}\,.$$

For $n=2$ and $n=3$ this is the logistic distribution. For general n the quantile function does not have an analytic form, but it and the probability density can be evaluated by numerical integration. The density is illustrated in Figure 4 in the case $\xi=1$ for several values of n. The density has exponentially decreasing tails and, as n increases, an increasingly broad and almost flat peak.

Example 4.3. Range $(-\infty, \infty)$; constraint $\mathrm{E}(X_{3:4} - X_{2:4}) = \xi$.

The constraint can be written, using (1.4), as

$$\int_0^1 12\{u^2(1-u) - u(1-u)^2\} Q(u)\,du = \xi\,, \tag{4.1}$$

or, integrating by parts, as

$$\int_0^1 6u^2(1-u)^2 Q'(u)\,du = \xi\,.$$

Thus the maximum-entropy distribution has

$$Q'(u) = \xi / \{6u^2(1-u)^2\}\,, \tag{4.2}$$

and, by integration,

$$Q(u) = \tfrac{1}{6}\xi \left\{ 2\log\left(\frac{u}{1-u}\right) + \frac{2u-1}{u(1-u)} \right\} + c \tag{4.3}$$

where c is an undetermined constant. This distribution has a probability density $f(x)$ with tails like those of a Cauchy distribution: they decay as $f(x) \sim 1/x^2$ as $x \to \pm\infty$. The mean of the distribution does not exist.

Example 4.4. Range $(-\infty, \infty)$; constraints $\mathrm{E}\, X_{2:4} = \xi_2$, $\mathrm{E}\, X_{3:4} = \xi_3$.

From (1.4) the constraints are

$$\int_0^1 12u(1-u)^2 Q(u)du = \xi_2\,, \qquad \int_0^1 12u^2(1-u)Q(u)du = \xi_3\,.$$

In the notation of Theorem 2.2, both constraints have $\int_0^1 J_r(0) \neq 0$, so we rewrite them as $\mathrm{E}(X_{3:4} - X_{2:4}) = \xi_3 - \xi_2 \equiv \xi$, $\mathrm{E}\, X_{3:4} = \xi_3$. The first constraint is now the same as in Example 4.3, and from it we obtain the same solution, (4.2), for $Q'(u)$. The second constraint, together with (2.23), determines $Q(u)$, which is found to have the form (4.3) with $c = \frac{1}{2}(\xi_2 + \xi_3)$ (for details see Appendix, item 6). Note that, in the notation of Theorem 2.2, we have $R = 2$ and $J_2(u) = 12u^2(1-u)$, whence $K_2(0) - K_2(v) = v^3(4-3v)$ and $K_2(v) = (1-v)^2(1+2v+3v^2)$; these functions have high enough powers of v and $1-v$ respectively to cancel the singularities in $Q'(v)$ as $v \to 0$ or $v \to 1$, so the integrals in (2.23) exist.

Example 4.5. Range $(-\infty, \infty)$; constraints $\mathrm{E}(X_{3:4} - X_{2:4}) = \xi$, $\mathrm{E}\, X = \mu$.

In contrast to the previous example, with the given constraints we can obtain the solution (4.2) for $Q'(u)$, but the integrals in (2.23) do not exist. The situation here is similar to that of Example 3.5: there are distributions that satisfy the constraints and have entropy arbitrarily close to that of the distribution (4.3), but this limit is not attained by any distribution that satisfies the constraints.

An example of a set of distributions that contains members that approach the limit is the set of distributions with $Q'(u) = c/\{u^\alpha(1-u)^\alpha\}$, where $0 < \alpha < 2$ and $c = \frac{1}{6}\xi\Gamma(6-2\alpha)/\{\Gamma(3-\alpha)\}^2$. These distributions have finite mean and satisfy $\mathrm{E}(X_{3:4} - X_{2:4}) = \xi$, and have entropy $\log(\xi/6) + \log\Gamma(6-2\alpha) - 2\log\Gamma(3-\alpha) + 2\alpha$, which increases towards the limit $\log(\xi/6) + 4$ as α approaches 2 from below (for details see Appendix, item 7). However, the limiting case $\alpha = 2$ does not correspond to a distribution with finite mean.

5. Nonparametric estimation of distributions

The first r sample L-moments can be used to summarize a data sample. It is natural to seek an estimate, based on these summary statistics, of the distribution from which the sample was drawn. The maximum-entropy approach provides such an estimate, in the form of a PDQ distribution with a density-quantile function that is a polynomial of degree r and whose first r population L-moments are equal to the sample L-moments.

Choice of r controls the amount of detail in the fitted distribution, and acts similarly to (the reciprocal of) the bandwidth of a kernel density estimator.

Strong philosophical claims have been made for the maximum-entropy approach, for example that it is "maximally noncommittal with regard to missing information" (Jaynes, 1957), but these claims do not apply to applications in exploratory data analysis, and are not necessary to justify the approach. The use of maximum entropy gives a convenient way of summarizing the information in a set of L-moments in the form of a probability distribution, but whether it is preferable to alternatives such as kernel density estimation depends on the practical properties of the final estimates.

We illustrate the maximum-entropy L-moment procedure with two data sets from Silverman (1986). We omit the computational details, which will be described in another paper. Briefly, the PDQ distribution is estimated follwoing the procedure in Theorem 3.1. Equations (3.7) are solved by minimizing the right side of (2.12) — as noted in Remark 2.3, this is a convex minimization problem. Numerical integration is used to evaluate the integral in (2.12), and, after the density-quantile function has been found, to compute the quantiles and the probability density function.

The "geyser" data are 226 waiting times between eruptions of the Old Faithful geyser in Yellowstone National Park. Figure 5 shows a histogram of the data, a kernel density estimate using a Gaussian kernel with bandwidth (standard deviation) 2.56 — this is obtained by the method of Sheather and Jones (1991), as descrtibed in Venables and Ripley (1999, sec. 8.7) — and a PDQ distribution based on 6 L-moments and with range the entire real line. The fits are generally similar, but the maximum-entropy L-moment distribution appears to give a better fit to the peak of the histogram near $x = 50$ and is less influenced by the largest data value at $x = 108$.

A second example is the "suicide" data, a sample of 86 lengths, in days, of psychiatric treatment undergone by patients used as controls in a study of suicide risk. The data values are all positive and some are close to zero. This can be a problem for kernel methods, which in their simplest form will assign substantial probability mass to negative values. Various solutions have been suggested for this problem (see Wand and Jones, 1995, sec. 2.1), and for the suicide data the reflected kernel estimate $\frac{1}{2}\{f(x) + f(-x)\}$ gives reasonable results. For the maximum-entropy L-moment approach however, there is no problem: the constraint that the range of the distribution has a lower bound is dealt with in Theorem 3.1 as "Case 1" and the appropriate PDQ distribution is accordingly obtained. Figure 6 shows a histogram of the data, a kernel density estimate using a Gaussian kernel with bandwidth 50 (subjectively judged to give a more reasonable fit than the Sheather-Jones estimate of 23.6) and a PDQ distribution based on 5 L-moments and with range the positive real line. Again, the fits are generally similar.

Some other authors have used the maximum-entropy priciple to fit distributions based on the L-moments of a sample, though their solutions did not involve the PDQ distribution. Pandey (2000) used an unusual definition of the entropy that is suitable only for distributions that take only positive values. Ulrych et al. (2000) were unable to derive the maximum-entropy solution; instead, they numerically maximized a cost function that combined the entropy and a term that penalized any mismatch between the L-moments of the data sample and the estimated distribution.

6. Further remarks

Given an observed set of n data points sampled from a continuous distribution, we can derive the distribution whose expected order statistics from a sample of size n are equal to the observed data values, and that has maximum entropy subject to this constraint. As in Example 4.1 it is a PDQ distribution whose density-quantile function is a polynomial of degree n. This is arguably the most natural continuous distribution that could be used to describe the data. It is an alternative to the "maximum entropy distribution" defined by Theil and others (e.g. Theil and Laitinen, 1980; Theil and Fiebig, 1984). It has potential applications to bootstrapping, where greater efficiency can sometimes be obtained by using a continuous distribution rather than the empirical distribution in bootstrap sampling (e.g., Silverman and Young, 1987; Hutson and Ernst, 2000).

Alternative entropy measures can be considered in place of H in (1.1). For example one can use Rényi's (1961) generalized entropy

$$H_\alpha = \frac{1}{1-\alpha} \log \int_{-\infty}^{\infty} \{f(x)\}^\alpha \, dx = \frac{1}{1-\alpha} \log \int_0^1 \{Q'(u)\}^{1-\alpha} \, du \,. \tag{6.1}$$

Analogues of the results in sections 2–4 may be found using similar methods of proof. For example, the maximum generalized entropy distribution subject to conditions (2.4) has the form $Q'(u) = \{\sum a_s K_s(u)\}^{1/\alpha}$. When an endpoint of the distribution is infinite and $0 < \alpha < 1$, the tail of the probability density has a power-law form, with $f(x) \sim |x|^{1/(1-\alpha)}$.

Appendix

This Appendix contains mathematical details omitted from the main text.

1. Convexity of D (Remark 2.3)

We have, from (2.12),

$$D(\mathbf{a}) = -\int_0^1 \log\Big\{\sum_{s=1}^{S} a_s K_s(u)\Big\} du + \sum_{s=1}^{S} a_s h_s - 1\,. \tag{A.1}$$

Thus

$$\frac{\partial D}{\partial a_r} = -\int_0^1 \frac{K_r(u)}{\sum_{s=1}^{S} a_s K_s(u)}\, du + h_r \tag{A.2}$$

and

$$\frac{\partial^2 D}{\partial a_r \partial a_s} = \int_0^1 \frac{K_r(u) K_s(u)}{\Big(\sum_{t=1}^{S} a_t K_t(u)\Big)^2}\, du\,. \tag{A.3}$$

For any $x_1, \ldots, x_r$ we have

$$\begin{aligned}
\sum_{r=1}^{S}\sum_{s=1}^{S} x_r x_s \frac{\partial^2 D}{\partial a_r \partial a_s} &= \sum_{r=1}^{S}\sum_{s=1}^{S} \int_0^1 \frac{x_r x_s K_r(u) K_s(u)}{\Big(\sum_{t=1}^{S} a_t K_t(u)\Big)^2}\, du \\
&= \int_0^1 \frac{\Big(\sum_{r=1}^{S} x_r K_r(u)\Big)^2}{\Big(\sum_{t=1}^{S} a_t K_t(u)\Big)^2}\, du \\
&> 0 \qquad \text{unless } x_1 = \ldots = x_r = 0;
\end{aligned}$$

thus the second-derivative matrix of D is positive definite and so D is convex. At a local minimum of D, the right side of (A.2) is zero, and therefore $\mathbf{a} = (a_1, \ldots, a_S)$ satisfies (2.5).

2. Limiting behaviour of the entropy in Example 3.5

The entropy is given by

$$H(\beta, \lambda_2) \equiv 2 + \log\lambda_2 + 2\log\beta - \frac{1}{\beta}(1+\beta)\log(1+\beta) - \log\{\beta - \log(1+\beta)\}\,. \tag{A.4}$$

Differentiating, we obtain

$$\begin{aligned}
\frac{\partial H}{\partial \beta} &= \frac{1}{\beta} + \frac{1}{\beta^2}\log(1+\beta) - \frac{\beta}{(1+\beta)\{\beta - \log(1+\beta)\}} \\
&= \frac{\beta^2 - (1+\beta)\{\log(1+\beta)\}^2}{\beta^2(1+\beta)\{\beta - \log(1+\beta)\}}\,.
\end{aligned} \tag{A.5}$$

As $\beta \to \infty$, $\log(1+\beta) = o(\beta^{1/2})$, so for β sufficiently large the right side of (A.5) is positive and $H(\beta, \lambda_2)$ is an increasing function of β for fixed λ_2. The limiting value of the entropy is obtained by writing (A.4) as

$$H(\beta, \lambda_2) = 2 + \log \lambda_2 + \log\left(\frac{\beta}{1+\beta}\right) - \frac{1}{\beta}\log(1+\beta) - \log\left\{1 - \frac{1}{\beta}\log(1+\beta)\right\};$$

each of the last three terms tends to zero as $\beta \to \infty$, so the limiting value of the entropy is $2 + \log \lambda_2$.

3. Derivation of PDQ3 distribution (Example 3.7)

With $\gamma = -a_3/a_2$, we write (3.20) as

$$\begin{aligned} Q'(u) &= \frac{1}{a_2} \cdot \frac{1}{u(1-u)\{1-\gamma(2u-1)\}} \\ &= \frac{1}{a_2(1-\gamma)^2}\left(\frac{1-\gamma}{u} + \frac{1+\gamma}{1-u} - \frac{4\gamma^2}{\{1-\gamma(2u-1)\}}\right), \end{aligned}$$

which can be integrated to give

$$Q(u) = \xi' + \alpha[(1-\gamma)\log u - (1+\gamma)\log(1-u) + 2\gamma \log\{1-\gamma(2u-1)\}], \tag{A.6}$$

where ξ' and α are constants. It is convenient to set $\xi = \xi' + \alpha\gamma \log 4$ so that the location parameter ξ is the median of the distribution. (A.6) can now be rewritten in the form (3.21), which explicitly represents the quantile function as the logistic quantile function plus an additional term that vanishes if $\gamma = 0$.

The constants ξ, α and γ must be chosen so that the distribution has the specified values λ_1, λ_2 and λ_3 for its first three L-moments. Differentiating (A.6), we have

$$Q'(u) = \frac{\alpha(1-\gamma^2)}{u(1-u)\{1-\gamma(2u-1)\}}, \tag{A.7}$$

so

$$\begin{aligned} \lambda_2 = \int_0^1 u(1-u)Q'(u)du &= \alpha(1-\gamma^2)\int_0^1 \frac{du}{1-\gamma(2u-1)} \\ &= \alpha(1-\gamma^2)\left[\frac{-1}{2\gamma}\log\{1-\gamma(2u-1)\}\right]_0^1 \\ &= \alpha(1-\gamma^2)\{-\log(1-\gamma) + \log(1+\gamma)\}/(2\gamma) \\ &= \alpha(1-\gamma^2)\,\operatorname{artanh}\gamma/\gamma \end{aligned} \tag{A.8}$$

and

$$\begin{aligned}
\lambda_3 &= \int_0^1 u(1-u)(2u-1)Q'(u)du \\
&= \int_0^1 u(1-u)\,\frac{1}{\gamma}[1-\{1-\gamma(2u-1)\}]Q'(u)du \\
&= \frac{1}{\gamma}\int_0^1 u(1-u)Q'(u)du - \frac{1}{\gamma}\int_0^1 u(1-u)\{1-\gamma(2u-1)\}Q'(u)du \\
&= \frac{1}{\gamma}\,\lambda_2 - \frac{1}{\gamma}\,\alpha(1-\gamma^2)\,. \qquad \text{(A.9)}
\end{aligned}$$

Dividing (A.9) by λ_2 and using (A.8), we have

$$\tau_3 = \frac{1}{\gamma} - \frac{1}{\operatorname{artanh}\gamma}\,, \qquad \text{(A.10)}$$

which is (3.22). Eliminating $\operatorname{artanh}\gamma$ from (A.8) and (A.10), we obtain the expression (3.23) for α. Finally, we can integrate (A.6), obtaining

$$\begin{aligned}
\lambda_1 &= \int_0^1 Q(u)du \\
&= \xi' + \alpha\int_0^1 [(1-\gamma)\log u - (1+\gamma)\log(1-u) - 2\gamma\log\{1-\gamma(2u-1)\}]\,du \\
&= \xi' + \alpha\Big[-(1-\gamma)+(1+\gamma)+2\gamma\Big\{1+\frac{1+\gamma}{2\gamma}\log(1+\gamma) - \frac{1-\gamma}{2\gamma}\log(1-\gamma)\Big\}\Big] \\
&= \xi - \alpha\gamma\log 4 + \alpha\{(1+\gamma)\log(1+\gamma) - (1-\gamma)\log(1-\gamma)\}\,,
\end{aligned}$$

which yields (3.24).

4. τ_4 for PDQ3 distribution (Example 3.7)

We have

$$\begin{aligned}
0 &= \alpha(1-\gamma\tau_3)\int_0^1 (2u-1)du \\
&= \int_0^1 (2u-1)\,.\,u(1-u)\{1-\gamma(2u-1)\}Q'(u)du \qquad \text{by (A.7)} \\
&= \int_0^1 u(1-u)\Big[2u-1-\gamma\Big\{\tfrac{1}{5}+\tfrac{4}{5}(5u^2-5u+1)\Big\}\Big]Q'(u)du \\
&= \int_0^1 u(1-u)(2u-1)Q'(u)du - \tfrac{1}{5}\gamma\int_0^1 u(1-u)Q'(u)du \\
&\qquad - \tfrac{4}{5}\gamma\int_0^1 u(1-u)(5u^2-5u+1)Q'(u)du
\end{aligned}$$

$$= \lambda_3 - \tfrac{1}{5}\gamma\lambda_2 - \tfrac{4}{5}\gamma\lambda_4 \qquad \text{by (3.2)–(3.4).}$$

Dividing by λ_2, we have

$$\tau_3 - \tfrac{1}{5}\gamma - \tfrac{4}{5}\gamma\tau_4 = 0\,,$$

i.e. $\tau_4 = (5\tau_3/\gamma - 1)/4$.

5. Derivation of maximum-entropy distribution with specified λ_1, λ_2, λ_4 (Example 3.8)

Let $\omega = a_4/a_2$ and write (3.25) as

$$Q'(u) = \frac{1}{a_2} \cdot \frac{1}{u(1-u)\{1+\omega-5\omega u(1-u)\}} \tag{A.11}$$

$$= \frac{1}{a_2(1+\omega)} \left\{ \frac{1}{u(1-u)} + \frac{5\omega}{1+\omega-5\omega u(1-u)} \right\}, \tag{A.12}$$

Valid values of ω are those for which $Q'(u) > 0$ for all $u \in (0,1)$; from (A.11) this condition is equivalent to

$$T(u) \equiv 1+\omega-5\omega u(1-u) > 0 \qquad \text{for all } u \in (0,1). \tag{A.13}$$

If $\omega \le 0$ the function T is concave and (A.13) is satisfied if and only if $T(0)$ and $T(1)$ are both positive; since $T(0) = T(1) = 1+\omega$ we must have $\omega \ge -1$. If $\omega \ge 0$ the function T is convex and (A.13) is satisfied if and only if the minimum value of $T(u)$ on $u \in (0,1)$ is strictly positive; since the minimum is attained at $u = \frac{1}{2}$ and $T(\frac{1}{2}) = 1 - \frac{1}{4}\omega$, we must have $\omega < 4$. Thus the valid values of ω are $-1 \le \omega < 4$.

If $-1 \le \omega \le 0$ we set $\delta = +\sqrt{-5\omega/(4-\omega)} \in [0,1]$, write (A.12) as

$$Q'(u) = \frac{1}{a_2(1+\omega)} \left\{ \frac{1}{u(1-u)} - \frac{1}{1/(2\delta)^2 - (u-\frac{1}{2})^2} \right\},$$

and integrate it to give

$$Q(u) = \xi + \alpha \left[\log\left(\frac{u}{1-u} \right) - 2\delta \operatorname{artanh}\{2\delta(u-\tfrac{1}{2})\} \right]; \tag{A.14}$$

if $0 < \omega < 4$ we set $\delta = -\sqrt{5\omega/(4-\omega)} \in (-\infty, 0)$, write (A.12) as

$$Q'(u) = \frac{1}{a_2(1+\omega)} \left\{ \frac{1}{u(1-u)} + \frac{1}{1/(2\delta)^2 + (u-\frac{1}{2})^2} \right\},$$

and integrate it to give

$$Q(u) = \xi + \alpha\left[\log\left(\frac{u}{1-u}\right) + 2\delta \arctan\{2\delta(u - \tfrac{1}{2})\}\right];$$

in each case ξ and α are constants. The choice of signs of the square roots in the definition of δ is arbitrary; our choice will ensure that δ is a monotonically increasing function of τ_4, the L-kurtosis of the distribution.

The constants ξ, α and δ must be chosen so that the distribution has the specified values λ_1, λ_2 and λ_4 for its L-moments of orders 1, 2 and 4. We consider only the case $\delta > 0$; the corresponding results in the case $\delta < 0$ can be obtain by a similar procedure. Differentiating (A.14), we have

$$\begin{aligned}
Q'(u) &= \alpha\left\{\frac{1}{u(1-u)} - \frac{1}{1/(2\delta)^2 - (u-\frac{1}{2})^2}\right\} \\
&= \frac{\alpha(1-\delta^2)}{4\delta^2} \cdot \frac{1}{u(1-u)\{1/(2\delta)^2 - (u-\frac{1}{2})^2\}},
\end{aligned}$$

so

$$\begin{aligned}
\lambda_2 = \int_0^1 u(1-u)Q'(u)du &= \frac{\alpha(1-\delta^2)}{4\delta^2}\int_0^1 \frac{du}{1/(2\delta)^2 - (u-\frac{1}{2})^2} \\
&= \frac{\alpha(1-\delta^2)}{4\delta^2}\left[2\delta \operatorname{artanh}\{2\delta(u-\tfrac{1}{2})\}\right]_0^1 \\
&= \frac{\alpha(1-\delta^2)}{4\delta^2}\; 2\delta\{\operatorname{artanh}\delta - \operatorname{artanh}(-\delta)\} \\
&= \alpha(1-\delta^2)\operatorname{artanh}\delta/\delta\,,
\end{aligned}$$

whence α is given by the first equation of (3.29). Further, we have

$$\begin{aligned}
\lambda_4 = \int_0^1 u(1-u)(5u^2-5u+1)Q'(u)du &= \frac{\alpha(1-\delta^2)}{4\delta^2}\int_0^1 \frac{-\frac{1}{4}+5(u-\frac{1}{2})^2}{1/(2\delta)^2-(u-\frac{1}{2})^2}\,du \\
&= \frac{\alpha(1-\delta^2)}{4\delta^2}\int_0^1\left(-5+\frac{(5/\delta^2-1)/4}{1/(2\delta)^2-(u-\frac{1}{2})^2}\right)du \\
&= \frac{\alpha(1-\delta^2)}{4\delta^2}\left\{-5+\tfrac{1}{4}\left(\frac{5}{\delta^2}-1\right)4\delta\operatorname{artanh}\delta\right\};
\end{aligned}$$

thus

$$\begin{aligned}
\tau_4 = \frac{\lambda_4}{\lambda_2} &= \frac{1}{4\delta\operatorname{artanh}\delta}\left\{-5+\left(\frac{5}{\delta^2-1}\right)\right\} \\
&= \frac{-5}{4\delta\operatorname{artanh}\delta} + \frac{5}{4\delta^2} - \frac{1}{4}\,,
\end{aligned}$$

which is the first equation of (3.28). Integrating (A.14) gives

$$\lambda_1 = \int_0^1 Q(u)du = \xi\,,$$

since the factor multiplying α in (A.14) is integrable over $(0,1)$ and symmetric about $u = \frac{1}{2}$; this is (3.30).

6. Derivation of c in Example 4.4

To obtain $Q(u)$ we evaluate (2.23). We have $R=2$, $g_2=\xi_3$, $K_2(v)=\int_v^1 12u^2(1-u)du = (1-v)^2(1+2v+3v^2)$, and $Q'(u)=\frac{1}{6}(\xi_3-\xi_2)/\{u^2(1-u)^2\}$; thus

$$\begin{aligned}
\int_0^u \{K_2(0)-K_2(v)\}Q'(v)dv &= \int_0^u (4v^3-3v^4)\,\frac{\xi_3-\xi_2}{6v^2(1-v)^2}\,dv \\
&= \tfrac{1}{6}(\xi_3-\xi_2)\int_0^u \frac{4v-3v^2}{(1-v)^2}\,dv \\
&= \tfrac{1}{6}(\xi_3-\xi_2)\int_{1-u}^1 \frac{1+2t-3t^2}{t^2}\,dt \qquad (t=1-v) \\
&= \tfrac{1}{6}(\xi_3-\xi_2)\left[-\frac{1}{t}+2\log t-3t\right]_{1-u}^1 \\
&= \tfrac{1}{6}(\xi_3-\xi_2)\left(-1-3u-2\log(1-u)+\frac{1}{1-u}\right)
\end{aligned}$$

and

$$\begin{aligned}
\int_u^1 K_2(v)Q'(v)dv &= \int_u^1 (1-v)^2(1+2v+3v^2)\,\frac{\xi_3-\xi_2}{6v^2(1-v)^2}\,dv \\
&= \tfrac{1}{6}(\xi_3-\xi_2)\int_u^1 \frac{1+2v+3v^2}{v^2}\,dv \\
&= \tfrac{1}{6}(\xi_3-\xi_2)\left[-\frac{1}{v}+2\log v+3v\right]_u^1 \\
&= \tfrac{1}{6}(\xi_3-\xi_2)\left(2-3u-2\log(1-u)+\frac{1}{u}\right),
\end{aligned}$$

so (2.23) gives

$$\begin{aligned}
Q(u) &= \xi_3+\tfrac{1}{6}(\xi_3-\xi_2)\left(-3+2\log u-2\log(1-u)+\frac{1}{1-u}-\frac{1}{u}\right) \\
&= \tfrac{1}{2}(\xi_2+\xi_3)+\tfrac{1}{6}(\xi_3-\xi_2)\left\{2\log\left(\frac{u}{1-u}\right)+\frac{2u-1}{u(1-u)}\right\}.
\end{aligned}$$

7. Limiting behaviour of the entropy in Example 4.5

Given $Q'(u) = c/\{u^\alpha(1-u)^\alpha\}$, to obtain $\mathrm{E}(X_{3:4} - X_{2:4}) = \xi$, i.e. (4.1), we require that

$$\begin{aligned}
\xi = \mathrm{E}(X_{3:4} - X_{2:4}) &= \int_0^1 6u^2(1-u)^2 Q'(u)du \\
&= 6c \int_0^1 u^{2-\alpha}(1-u)^{2-\alpha} du \\
&= 6c\,\{\Gamma(3-\alpha)\}^2/\Gamma(6-2\alpha)\,,
\end{aligned}$$

whence $c = \frac{1}{6}\xi\Gamma(6-2\alpha)/\{\Gamma(3-\alpha)\}^2$. The entropy of the distribution is

$$\begin{aligned}
\bar{H}(\alpha) \equiv \int_0^1 \log Q'(u)du &= \log c - \alpha \int_0^1 \log u\, du - \alpha \int_0^1 \log(1-u)\, du \\
&= \log c + 2\alpha \\
&= \log(\xi/6) + \log\Gamma(6-2\alpha) - 2\log\Gamma(3-\alpha) + 2\alpha\,.
\end{aligned}$$

$\bar{H}(\alpha)$ is a continuous function of $\alpha \in (0, 2]$, and

$$\frac{d}{d\alpha}\bar{H}(\alpha) = -2\psi(6-2\alpha) + 2\psi(3-\alpha) + 2$$

where $\psi(x) = \frac{d}{dx}\log\Gamma(x)$ is Euler's psi function. Now for $x > 1$ we have

$$\begin{aligned}
\psi(2x) - \psi(x) &= \sum_{k=0}^{\infty}\left(\frac{1}{x+k} - \frac{1}{2x+k}\right) \qquad \text{(Gradshteyn and Ryzhik, 1980, eq. 8.363.3)} \\
&= \sum_{k=0}^{\infty} \frac{x}{(x+k)(2x+k)} \\
&< \sum_{k=0}^{\infty} \frac{x}{(x+k)(x+1+k)} \\
&= 1 \qquad \text{(Gradshteyn and Ryzhik, 1980, eq. 0.243.1)},
\end{aligned}$$

so for $\alpha < 2$ we have $\frac{d}{d\alpha}\bar{H}(\alpha) > 0$. Thus as $\alpha \to 2$, $\bar{H}(\alpha)$ increases towards its limiting value $\bar{H}(2) = \log(\xi/6) + 4$.

References

Gradshteyn, I. S., and Ryzhik, I. M. (1980). *Table of integrals, series and products*, 4th edn. Academic Press, New York.

Holm, J. (1993). Maximum entropy Lorenz curves. *Journal of Econometrics*, **59**, 377–389.

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.

Hosking, J. R. M. (1996). Fortran routines for use with the method of L-moments, version 3. *Research Report RC20525*, IBM Research Division, Yorktown Heights, N.Y.

Hosking, J. R. M. (1998). L-moments. In *Encyclopedia of statistical sciences, update vol. 2*, ed. S. Kotz, C. Read and D. L. Banks, 357–362. New York: Wiley.

Hosking, J. R. M. (2000). Maximum-entropy characterization of the logistic distribution using L-moments. *Research Report RC 21691*, IBM Research Division, Yorktown Heights, N.Y.

Hosking, J. R. M., and Wallis, J. R. (1997). *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press, Cambridge, U.K.

Hutson, A. D., and Ernst, M. D. (2000). The exact bootstrap mean and variance of an L-estimator. *Journal of the Royal Statistical Society, Series B*, **62**, 89–94.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, **106**, 620–630.

Jones, M. C., and Balakrishnan, N. (2002). How are moments and moments of spacings related to distribution functions? *Journal of Statistical Planning and Inference*, **103**, 377–390.

Karvanen, J., Eriksson, J., and Koivunen, V. (2002). Adaptive score functions for maximum likelihood ICA. *Journal of VLSI Signal Processing*, **32**, 83–92.

Pandey, M. D. (2000). Direct estimation of quantile functions using the maximum entropy principle. *Structural Safety*, **22**, 61–79.

Parzen, E. (1979). Nonparametric statistical data modeling. *Journal of the American Statistical Association*, **74**, 105–121.

Rényi, A. (1961). On measures of entropy and information. In *Proceedings of the 4th Berkely Symposium on Mathematical Statistics and Probability*, 547–561. University of California Press, Berkeley, Calif.

Sansone, G. (1959). *Orthogonal functions.* Interscience Publishers, New York (reprinted 1991, Dover Publications, New York).

Sheather, S. J., and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Series B*, **53**, 683–690.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* Chapman and Hall, London.

Silverman, B. W., and Young, G. A. (1987). The bootstrap: to smooth or not to smooth? *Biometrika*, **74**, 469–479.

Theil, H., and Fiebig, D. G. (1984). *Exploiting continuity: maximum entropy estimation of continuous distributions.* Ballinger, Cambridge, Mass.

Theil, H., and Laitinen, K. (1980). Singular moment matrices in applied econometrics. In *Multivariate Analysis–V*, ed. P. R. Krishnaiah, pp. 629–649. North-Holland, Amsterdam.

Troutman, J. L. (1983). *Variational calculus with elementary convexity.* Springer, Berlin.

Ulrych, T. J., Velis, D. R., Woodbury, A. D., and Sacchi, M .D. (2000). L-moments and C-moments. *Stochastic Environmental Research and Risk Assessment*, **14**, 50–68.

Venables, W. N., and Ripley, B. D. (1999) *Modern applied statistics with S-PLUS.* Springer-Verlag, New York.

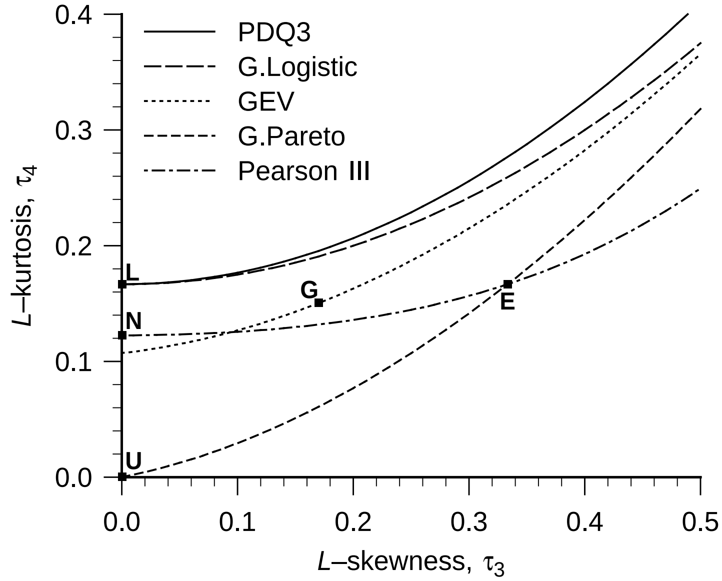

Figure 1. L-moment ratio diagram showing the relation between τ_3 and τ_4 for several three-parameter families of distributions: PDQ3, generalized logistic, generalized extreme-value (GEV), generalized Pareto, and Pearson type III (other than the PDQ3, these are all as defined in Hosking, 1996, or Hosking and Wallis, 1997). Labelled points indicate two-parameter distributions: exponential (E), Gumbel (G), logistic (L), normal (N) and uniform (U).

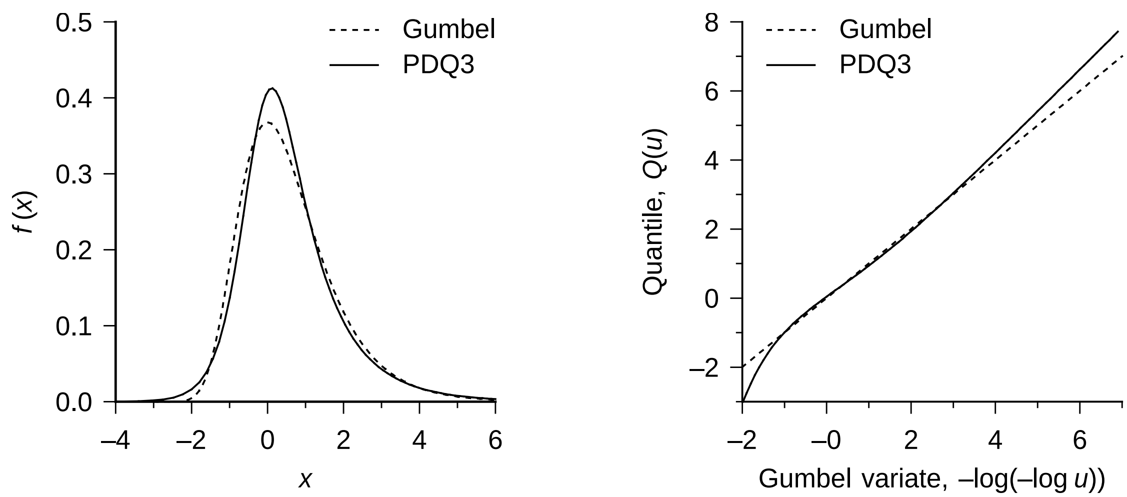

Figure 2. Probability density functions and a quantile-quantile plot for the Gumbel distribution and the maximum-entropy (PDQ3) distribution that has the same first three L-moments.

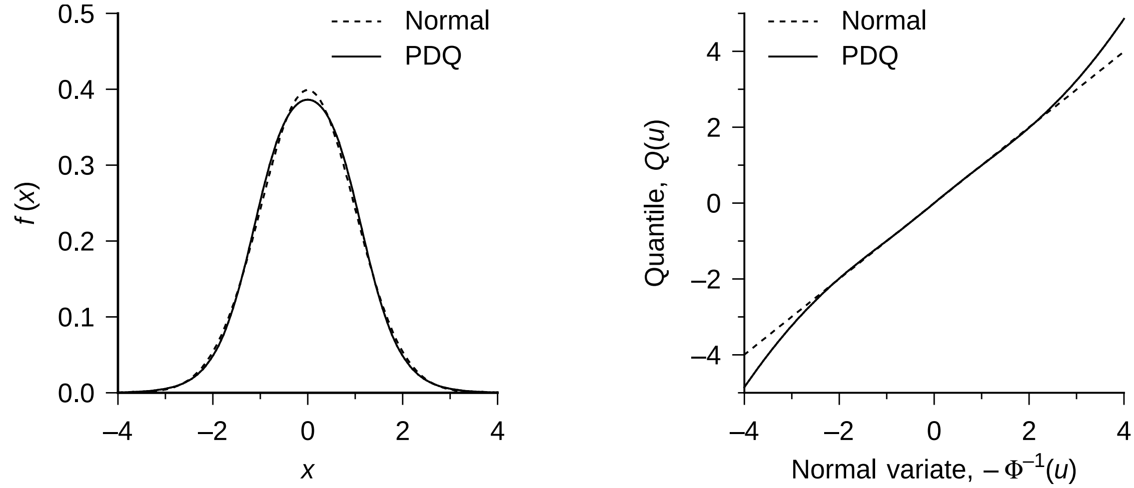

Figure 3. Probability density functions and a quantile-quantile plot for the standard Normal distribution and the maximum-entropy (PDQ) distribution that has the same L-moments λ_1, λ_2, λ_4.

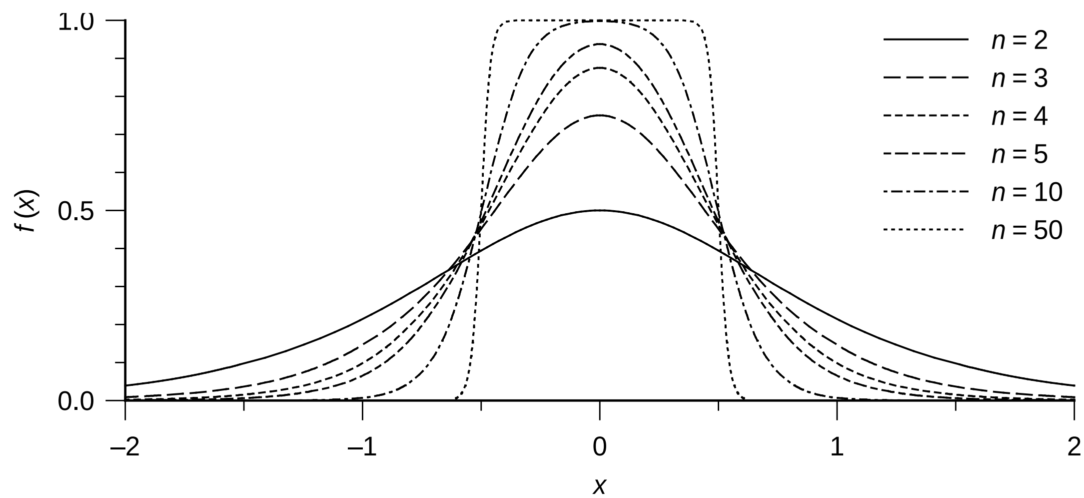

Figure 4. Probability density functions of Example 4.2, the maximum-entropy distribution subject to the constraint that the expected range of a sample of size n is 1.

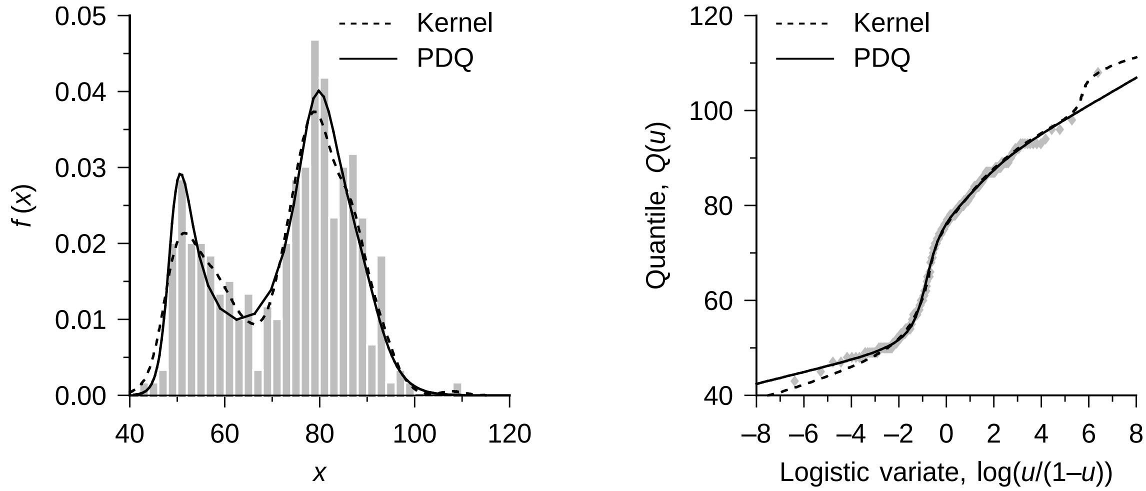

Figure 5. Distributions fitted to the "geyser" data.

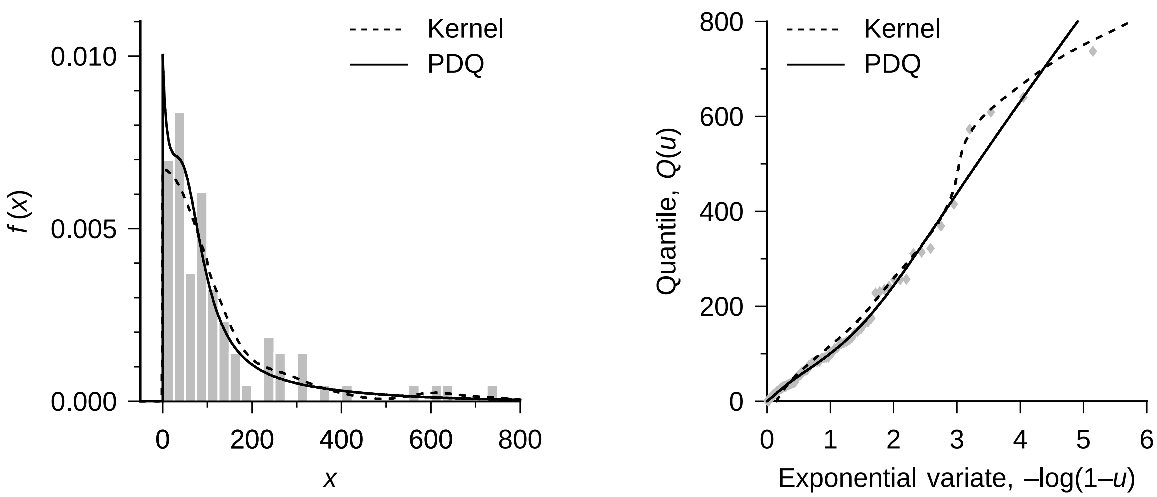

Figure 6. Distributions fitted to the "suicide" data.