Ann. Inst. Statist. Math.
Vol. 41, No. 4, 677–681 (1989)

SYMBOLIC COMPUTING THE EXACT DISTRIBUTIONS OF *L*-STATISTICS FROM A UNIFORM DISTRIBUTION

T. RAMALLINGAM

Division of Statistics, Northern Illinois University, DeKalb, IL 60115, U.S.A.

(Received April 20, 1988; revised March 7, 1989)

Abstract. The exact probability density function of linear combinations of $k = k(n)$ order statistics selected from the whole order statistics (L-statistic) based on a random sample of size n from the uniform distribution on [0, 1] was derived by Matsunawa (1985, *Ann. Inst. Statist. Math.*, **37**, 1–16). As the main expression for the density function given by Matsunawa is not complete for the general situation, we first provide the corrections for this formula. Second, we propose a simple scheme involving symbolic computing for evaluating the corrected version of the density function. The cumulative distribution function and the r-th mean of his L-statistic are also derived.

Key words and phrases: Linear combination, order statistics, uniform distribution, exact distribution, symbolic differentiation.

1. Introduction

The so-called L-statistics, or linear combinations of order statistics, are widely used in estimation and hypothesis testing problems (e.g., Kochar (1984) and D'Agostino and Stephens (1986)). The asymptotic theory of L-statistics has been dealt with adequately in the literature (see Serfling (1980)). However, for small samples, the implementation and evaluation of statistical procedures based on L-statistics requires knowledge of the exact distributions of the statistics.

Let $U_1 < U_2 < \cdots < U_n$ be order-statistics based on a random sample of size n from the uniform distribution on the interval $(0, 1)$. Selecting k statistics $U_{n_1} < U_{n_2} < \cdots < U_{n_k}$, where $k = k(n)$, $1 \le k \le n$ and $0 < n_1 < \cdots < n_k < n + 1$, we denote the L-statistic based on these order statistics by

$$(1.1) \qquad L_n = \sum_{i=1}^{k} a_i U_{n_i} .$$

In (1.1), $a_l = a_l(n)$, $l = 1, 2, \ldots, k$ are real constants with $\sum_{l=1}^{m} a_l^2 \neq 0$ for each m.

Matsunawa (1985) has derived the exact probability density function (p.d.f.) of the general L-statistic (1.1). However, his formula for the p.d.f. is not complete. Furthermore, the calculation of the coefficients in his expression (see his equations (3.7) and (3.8)) are cumbersome. In Section 2, we provide the corrected version of his p.d.f. We then derive formulas for the moments and the cumulative distribution function (c.d.f.) of the L-statistic. Finally, in Section 3, we propose an efficient scheme to compute the coefficients in these formulas.

2. The exact distributional properties of L_n

In order to state the corrected version of the exact p.d.f. of L_n in (1.1), we need these notations:

Let $b_j = b_j(n) = \sum_{l=j}^{k} a_l$, $j = 1, 2, \ldots, k$. Let k^* be the number of distinct non-zero b_j's. Denote such b_j's by $b_1^*, b_2^*, \ldots, b_{k^*}^*$ and the corresponding multiplicities by $\nu_1, \ldots, \nu_{k^*}$. Finally, for a complex variable s, define functions $G(s)$, $G_l(s)$ thus:

$$G(s) = \left[\prod_{j=1}^{k^*} (s + (1/b_j^*))^{\nu_j} \right]^{-1}, \tag{2.1}$$

$$G_l(s) = (s + (1/b_l^*))^{\nu_l} G(s), \qquad l = 1, 2, \ldots, k^*. \tag{2.2}$$

The exact p.d.f. of L_n:

$$f_{L_n}(t) = \sum_{l=1}^{k^*} \sum_{m=1}^{\nu_l} [\operatorname{sgn}(b_l^*) C_{l,m}^{\#} \chi(t/b_l^*) \chi(1 - t/b_l^*) t^{m-1} (1 - t/b_l^*)^{n-m} / B(m, n-m+1)], \tag{2.3}$$

where

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if} \quad x \geq 0, \\ -1 & \text{if} \quad x < 0, \end{cases} \tag{2.4}$$

$$\chi(x) = \begin{cases} 1 & \text{if} \quad x > 0, \\ 0 & \text{if} \quad x \leq 0, \end{cases} \tag{2.5}$$

$$C_{l,m} = (G_l^{(\nu_l - m)}(-1/b_l^*))/(\nu_l - m)!, \tag{2.6}$$

and

$$C_{l,m}^{\#} = \left(\prod_{j=1}^{k^*} (b_j^*)^{-\nu_j} \right) C_{l,m}, \tag{2.7}$$

$G_l^p(s)$ being the p-th derivative of $G_l(s)$ in (2.2).

It should be remarked that (2.3) is the same as equation (3.7) of Matsunawa (1985) except for the factor sgn (b_l^*) defined in (2.4). Without this correction, the function $f(t)$ in (3.7) of Matsunawa (1985) may not always be a bonafide p.d.f. The proof of (2.3) is similar to that of (3.7) of Matsunawa (1985) and is omitted.

When computing the percentage points or the moments of the statistic L_n in (1.1), the following formulas for the cumulative distribution function (c.d.f.), $F_{L_n}(x)$, and the r-th raw moment, $\mu_{r:n}$, are useful:

The c.d.f. of L_n:

$$(2.8) \qquad F_{L_n}(x) = \sum_{l=1}^{k^*} \sum_{m=1}^{\nu_l} C_{l,m}^{\#}(b_l^*)^m [g_1(x,l,m)\chi(b_l^*) + g_2(x,l,m)\chi(-b_l^*)] ,$$

where, denoting the incomplete beta function by

$$IB(y,\alpha,\beta) = \int_0^y t^{\alpha-1}(1-t)^{\beta-1}dt / B(\alpha,\beta), \quad 0<y<1, \quad 0<\alpha,\beta ,$$

and $\chi(x)$ as in (2.5),

$$g_1(x,l,m) = \begin{cases} 0 & \text{if} \quad x<0 , \\ IB(x/b_l^*, m, n-m+1) & \text{if} \quad 0 \le x < b_l^* , \\ 1 & \text{if} \quad x \ge b_l^* , \end{cases}$$

$$g_2(x,l,m) = \begin{cases} 0 & \text{if} \quad x<b_l^* , \\ IB(1-x/b_l^*, n-m+1, m) & \text{if} \quad b_l^* \le x < 0 , \\ 1 & \text{if} \quad x \ge b_l^* . \end{cases}$$

The r-th moment of L_n:

$$(2.9) \qquad \mu_{r:n} = E(L_n^r) = \sum_{l=1}^{k^*} \sum_{m=1}^{\nu_l} C_{l,m}^{\#}(b_l^*)^{m+r}[B(m+r, n-m+1) / B(m, n-m+1)], \qquad r>0 .$$

Equations (2.8) and (2.9) are obtained by appropriate integrations of the p.d.f. in (2.3), and we omit the details.

3. Computing the p.d.f. of L_n

While the formulas (2.6) for the coefficients $C_{l,m}$ are found amidst the proof of Lemma 2.2 in Matsunawa (1985), the expressions for these coefficients via Bell polynomials in his final equation (3.7) are, in general, quite complicated. We therefore suggest that, in using (2.3), (2.8) or (2.9) for computing percentiles, mean, etc., we should first calculate $C_{l,m}$ directly. Indeed, the representation (2.6) of $C_{l,m}$ as functions of derivatives of $G_l(s)$ in (2.2) is ideal for symbolic manipulation using many commercially available softwares such as REDUCE (Seward (1985)). Once $C_{l,m}$ are found by symbolic computing, we can use FORTRAN or other codes to compute the quantities in (2.3), (2.8) or (2.9).

For the benefit of the reader, we now provide a sample REDUCE code to illustrate symbolic computation of $C_{l,m}$ for the following statistic, introduced by Kochar and Ramallingam (1989) in the context of an inference problem for Poisson processes:

$$L_n = \sum_{i=1}^{n} (3i - 2n - 1) U_i . \tag{3.1}$$

Suppose, for example, that $n = 7$. Then, in the notations of Section 2, $k^* = 4$, $b_1^* = -21$, $b_2^* = -9$, $b_3^* = 6$ and $b_4^* = 9$. Also, $\nu_1 = 1$, $\nu_2 = 1$, $\nu_3 = 2$, $\nu_4 = 2$. One may use the following REDUCE program to obtain $C_{l,m}$, $1 \le l \le 4$, $1 \le m \le \nu_l$.

```
COMMENT to find C(L,M) by symbolic computing;
K := 4;
ARRAY LAMDA(K), NU(K), H(K), CNUM(K,K), CDEN(K,K), C(K,K);
LAMDA (1) := - 1/21 $
LAMDA (2) := - 1/9 $
LAMDA (3) := 1/6 $
LAMDA (4) := 1/9 $
NU(1) := 1 $
NU(2) := 1 $
NU(3) := 2 $
NU(4) := 2 $
G := FOR L := 1:K PRODUCT (S + LAMDA(L)**( - NU(L))$
COMMENT use G(s) defined above to find Gl(s) = H(l);
FOR L := 1:k DO
  BEGIN
  H(L) := ((S + LAMDA(L))**NU(L))*G $
  END;
COMMENT Find the numerator, CNUM(L,M), the denominator,
    CDEN(L,M), of the formula in (2.6) and then print C(L,M);
FOR L := 1:K DO
  BEGIN
  FOR M := 1:NU(L) DO
  BEGIN
```

```
CNUM(L,M) := DF(H(L), S, NU(L) - M) $
CDEN(L,M) := FOR I := 1:(NU(L) - M) PRODUCT I $
C(L,M) := CNUM(L,M)/CDEN(L,M) $
WRITE "C("L,M,") = ", SUB(S = - LAMDA(L), C(L,M));
END;
END;
```

The above program yields the following values for $C_{l,m}$ coefficients in (2.6):

$$C_{1,1} = -13613.67, \quad C_{2,1} = 4133.43, \quad C_{3,1} = 240952.32, \quad C_{3,2} = 5443.2\,,$$

$$C_{4,1} = -231472.08, \quad C_{4,2} = 9185.4\,.$$

Acknowledgements

The author is thankful to Dr. Subhash C. Kochar for formulating an inference problem that led to this investigation of the exact distribution of *L*-statistics. Also, without the invaluable help of, and generous discussions with, Dr. R. Mukerjee, this work could not have been completed. Special thanks are due to him for first determining the corrections needed in (3.7) of Matsunawa (1985).

REFERENCES

D'Agostino, R. B. and Stephens, M. A. (1986). *Goodness-of-Fit Techniques*, Marcel Dekker, New York.

Kochar, S. C. (1984). Testing goodness-of-fit in terms of the failure rate function, *IAPQR Trans.*, **9**, 1–6.

Kochar, S. C. and Ramallingam, T. (1989). Testing for superadditivity of the mean value function of a non-homogeneous Poisson process, *Comm. Statist. A—Theory Methods*, **18**, 1549–1562.

Matsunawa, T. (1985). The exact and approximate distributions of linear combinations of selected order statistics from a uniform distribution, *Ann. Inst. Statist. Math.*, **37**, 1–16.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Seward, L. R. (1985). *REDUCE User's Guide for IBM 360 and Derivative Computers*, Version 3.2, The Rand Corporation, Santa Monica, California.