JOURNAL OF MULTIVARIATE ANALYSIS 10, 351–362 (1980)

Extreme Sample Censoring Problems with Multivariate Data: Indirect Censoring and the Farlie–Gumbel–Morgenstern Distribution*

N. L. JOHNSON*

University of North Carolina, Chapel Hill, North Carolina 27514

Communicated by P. R. Krishnaiah

Indirect censoring is defined as the effect on observed variables of censoring on unobserved variables. Methods of testing for indirect censoring are discussed, and exemplified, using a bivariate Farlie–Gumbel–Morgenstern distribution.

1. INTRODUCTION

Johnson [4] has given a survey of various problems which can arise in testing for censoring of extreme values from univariate data. When data are multivariate, there is a much richer variety of possible problems; some possibilities are described in Johnson [3, 5]. The present paper discusses the detection of *indirect censoring*, and investigates lines of attack for Farlie–Gumbel–Morgenstern bivariate distributions.

We suppose that observed values on m characters $X_1, X_2, \ldots, X_m$ are available for each of r individuals. We wish to investigate whether these represent a complete random sample, or are the remainder of such a sample (original size $n > r$) after some form of censoring of extreme values has been applied.

As in Johnson [4], we will restrict attention to random sampling from large populations in which the joint distribution of $X_1, \ldots, X_m$ is absolutely continuous, with joint probability density function (PDF)

$$f_{X_1,\ldots,X_m}(x_1, \ldots, x_m) = f_{\mathbf{X}}(\mathbf{x}) = f_{12\cdots m}(x_1, \ldots, x_m). \tag{1}$$

Received July 7, 1979; revised February 1, 1980.

AMS 1970 subject classification: 62H15, 62G30.

Key words: Indirect censoring, Farlie–Gumbel–Morgenstern distribution, likelihood ratio tests, order statistics, moments.

* Research sponsored by the Army Research Office under Grant DAAG29-77-C-0035.

0047-259X/80/030351–12$02.00/0

We will denote the (unordered) observations on the ith available individual by

$$\mathbf{X}^* = (X_{1i}^*, \ldots, X_{mi}^*) \qquad (i = 1, \ldots, r).$$

We also use the notation:

(i) $\Pr[\bigcap_{j=1}^{m} (X_{ji}^* \leqslant x_j)] = F_{12\cdots m}(x_1, \ldots, x_m)$ (in particular $\Pr[X_{ji}^* \leqslant x] = F_j(x)$),

(ii) for the conditional PDF of $X_{a_1}^*, \ldots, X_{a_s}^*$ given $X_{b_1}^*, \ldots, X_{b_t}^*$ $g_{a_1\cdots a_s;b_1\cdots b_t}(x_{a_1}, \ldots, x_{a_s} \mid x_{b_1}, \ldots, x_{b_t})$ (in particular $g_{12}(x_1 \mid x_2)$, $g_{21}(x_2 \mid x_1)$), (2)

and

(iii) for the order statistics corresponding to $X_{j1}^*, \ldots, X_{jr}^*$

$$X_{j1} \leqslant X_{j2} \leqslant \cdots \leqslant X_{jr}.$$

We also will focus on the forms of censoring accorded special attention in Johnson [4]:

(i) from above (exclusion of s_r greatest values) or below (exclusion of s_0 least values),

and

(ii) symmetrical (exclusion of equal number of greatest and least values ($s_0 = s_r$)).

We denote the hypothesis that the s_0 least and s_r greatest values of an original complete random sample of size n ($= r + s_0 + s_r$) have been excluded by H_{s_0, s_r}, so that

(i) corresponds to H_{0,s_r} or $H_{s_0,0}$ ($s_0, s_r > 0$),

(ii) corresponds to $H_{s,s}$ ($s > 0$).

To indicate that the censoring is applied to the variable X_j we use the symbol $H_{s_0,s_r}^{(j)}$.

2. Indirect Censoring

Suppose that there may be censoring on values of one variable—X_1, say—but values of this variables are not observed. How should the (observed) values of $(X_{2i}, \ldots, X_{mi})$ $(i = 1, \ldots, r)$ be used to detect if there has been censoring on X_1? For simplicity, we consider the bivariate case ($m = 2$), but extension of the theory to general m is straightforward. We first derive a likelihood ratio test. As we shall see, there appear to be considerable technical difficulties in applying this test in many natural situations. Therefore, we also suggest some other procedures which may sometimes be applied more easily.

The available data consist of the r observed values of X_2, denoted by $X_{21}^*, \ldots, X_{2r}^*$. Their joint PDF, if $H_{s_0,s_r}^{(1)}$ is valid, is

$$f_{\mathbf{X}_2^*}(\mathbf{x}_2 \mid H_{s_0,s_r}^{(1)}) = \frac{(r+s_0+s_r)!}{r!\, s_0!\, s_r!} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{F_1(l)\}^{s_0} \{1-F_1(u)\}^{s_r}$$

$$\times \prod_{i=1}^{r} [f_1(x_{1i})\, g_{21}(x_{2i} \mid x_{1i})]\, dx_{11} \cdots dx_{1r}, \tag{3}$$

where $l \equiv \min(x_{11}, \ldots, x_{1r})$; $u \equiv \max(x_{11}, \ldots, x_{1r})$.

Since

$$f_1(x_{1i})\, g_{21}(x_{2i} \mid x_{1i}) = f_{12}(x_{1i}, x_{2i}) = f_2(x_{2i})\, g_{12}(x_{1i} \mid x_{2i})$$

we also have

$$f_{\mathbf{X}_2^*}(\mathbf{x}_2 \mid H_{s_0,s_r}^{(1)}) = \frac{(r+s_0+s_r)!}{r!\, s_0!\, s_r!} \left\{ \prod_{i=1}^{r} f_2(x_{2i}) \right\} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \{F_1(l)\}^{s_0} \{1-F_1(u)\}^{s_r}$$

$$\times \prod_{i=1}^{r} g_{12}(x_{1i} \mid x_{2i})\, dx_{11} \cdots dx_{1r}. \tag{3'}$$

In particular

$$f_{\mathbf{X}_2^*}(\mathbf{x}_2^* \mid H_{0,0}^{(1)}) = \prod_{i=1}^{r} f_2(x_{2i}^*).$$

It follows that the likelihood ratio is

$$L = \frac{f_{\mathbf{X}_2^*}(\mathbf{X}_2^* \mid H_{s_0,s_r}^{(1)})}{f_{\mathbf{X}_2^*}(\mathbf{X}_2^* \mid H_{0,0}^{(1)})} = \frac{(r+s_0+s_r)!}{r!\, s_0!\, s_r!} \int_{\infty}^{\infty} \cdots \int_{\infty}^{\infty} \{F_1(l)\}^{s_0} \{1-F_1(u)\}^{s_r}$$

$$\times \prod_{i=1}^{r} g_{12}(x_{1i} \mid X_{2i}^*)\, dx_{11} \cdots dx_{1r}$$

$$= \frac{(r+s_0+s_r)!}{r!\, s_0!\, s_r!} E[\{F_1(X'_{11})\}^{s_0} \{1-F_1(X'_{1r})\}^{s_r} \mid \mathbf{X}_2^*], \tag{4}$$

where $X'_{11} \leqslant X'_{12} \leqslant \cdots \leqslant X'_{1r}$ are the order statistics of r *independent* random variables with densities $g_{12}(x_1 \mid X_{2j}^*)$ $(j = 1, \ldots, r)$. (Note that $X'_{11}, \ldots, X'_{1r}$ do *not*, in general, have the same joint distribution (given $\mathbf{X}_2^*$) as $X_{11}, \ldots, X_{1r}$, unless $s_0 = s_r = 0$.)

Calculation of L from the observed values $\mathbf{X}_2^*$ is usually quite difficult. When this is done, determination of the distribution of L (even when the null hypothesis, $H_{0,0}^{(1)}$, is valid) is likely to be even more difficult. In Section 3 we use a Farlie–Gumbel–Morgenstern (FGM) (see, e.g., Johnson and Kotz [7]) joint distribution for illustrative purposes. Calculation of L is not very difficult in this special case, but even here, the distribution of L is not easily derived. The conditional joint distribution of X'_{11} and X'_{1r} can be derived from

$$\Pr[l \leqslant X'_{11} \leqslant X'_{1r} \leqslant u \mid \mathbf{X}_2^*] = \prod_{i=1}^{r} \left[\int_l^u g_{12}(x_1 \mid X_{2i}^*)\, dx_1 \right] \tag{5}$$

but this expression is usually quite complicated.

We note that the value of L (and so its distribution) is unchanged by any monotonic increasing transformations of X_1^* and X_2^*. This means that we can take, without loss of generality, each of the variables to have a standard uniform distribution ($f_i(x) = 1$ for $0 \leqslant x \leqslant 1$, $i = 1, 2$). However, the joint PDF would then have to be that resulting from application of the appropriate transformations to the original joint PDF. The bivariate FGM distribution discussed in Section 3 does have standard uniform marginal distributions.

A simpler criterion, suggested by the above analysis, is

$$L_1 = \{F_1(\min_i E[X_1 \mid X_{2i}^*])\}^{s_0}\{1 - F_1(\max_i E[X_1 \mid X_{2i}^*])\}^{s_r}. \tag{6}$$

If $E[X_1 \mid X_2]$ is a monotonic increasing function of X_2 then

$$L_1 = \{F_1(E[X_1 | \min(X_{21}^*, ..., X_{2r}^*)])\}^{s_0}\{1 - F_1(E[X_1 | \max(X_{21}^*, ..., X_{2r}^*)])\}^{s_r}. \tag{7}$$

If $E[X_1 \mid X_2]$ is a monotonic decreasing function of X_2, then “min” and “max” in (7) are interchanged.

Another related criterion, generally more difficult to compute, is

$$L_1' = E[\{F_1(X_{1(l)})\}^{s_0}\{1 - F_1(X_{1(u)})\}^{s_r}], \tag{8}$$

where $X_{1(l)}^*, X_{1(u)}^*$ are independent with PDF’s $g_{12}(x_1 \mid X_{2(l)}^*)$, $g_{12}(x_1 \mid X_{2(u)}^*)$, respectively, and $i = (l)$, (u) respectively minimize and maximize $E[X_1 \mid X_{2i}^*]$ with respect to i.

If $E[X_1 \mid X_2]$ is a monotonic increasing (decreasing) function of X_2, then

$$X_{2(l)}^* = \min(\max)(X_{21}^*, ..., X_{2r}^*),$$

$$X_{2(u)}^* = \max(\min)(X_{21}^*, ..., X_{2r}^*).$$

3. Detection of Indirect Censoring in Farlie–Gumbel–Morgenstern (FGM) Distributions

3.1. *Relevant Properties of* FGM *Distributions*

We consider the bivariate FGM joint distribution

$$\Pr[(X_1^* \leqslant x_1) \cap (X_2^* \leqslant x_2)] = x_1 x_2\{1 + \theta(1-x_1)(1-x_2)\}$$
$$(0 \leqslant x_j \leqslant 1; j = 1, 2; |\theta| < 1). \quad (9)$$

Each X_j^* has a marginal standard uniform distribution ($F_j(x_j) = x_j$ $(0 \leqslant x_j \leqslant 1)$). This distribution has been chosen for analytical convenience. It is not claimed that the results will apply for other joint distributions, even after transformation to make the marginals be standard uniform. However, there are some speculative analogies which might be drawn.

From (9) it follows that the joint PDF is

$$f(x_1, x_2) = 1 + \theta(1-2x_1)(1-2x_2) \qquad (0 \leqslant x_j \leqslant 1; j = 1, 2) \quad (10)$$

and the conditional PDF's are

$$\begin{aligned} g_{12}(x_1 \mid x_2) &= 1 + \theta(1-2x_2)\cdot(1-2x_1) \qquad (0 \leqslant x_1 \leqslant 1), \\ g_{21}(x_2 \mid x_1) &= 1 + \theta(1-2x_1)\cdot(1-2x_2) \qquad (0 \leqslant x_2 \leqslant 1). \end{aligned} \quad (11)$$

Hence

$$\Pr[l \leqslant X_1^* \leqslant u \mid x_2] = (u-l)[1 + \theta(1-2x_2)(1-u-l)] \qquad (0 \leqslant l \leqslant u \leqslant 1)$$

and

$$\Pr[l \leqslant X'_{11} \leqslant X'_{1r} \leqslant u \mid \mathbf{x}_2^*] = (u-l)^r \prod_{j=1}^{r} [1 + \theta(1-2x_{2j}^*)(1-u-l)].$$

3.2. *Derivation of L*

The conditional joint PDF of X'_{11} and X'_{1r} is therefore

$$\frac{-\partial^2 \Pr[l \leqslant X'_{11} \leqslant X'_{1r} \leqslant u \mid \mathbf{x}_{(2)}^*]}{\partial l\, \partial u}$$
$$= r(r-1)(u-l)^{r-2} \prod_{j=1}^{r} \{1 + \theta_{2j}^*(1-u-l)\}$$
$$- 2\alpha^2(u-l)^r \sum_{j<j'}\sum z_{2j}^* z_{2j'}^* \prod_{h \neq j,j'} \{1 + \theta z_{2h}^*(1-u-l)\}, \quad (12)$$

where $z_{2j}^* = 1 - 2x_{2j}^*$.

From (4),

$$L = \frac{(r+s_0+s_r)!}{r!\,s_0!\,s_r!} E[X_{11}^{\prime s_0}(1-X'_{1r})^{s_r} \mid \mathbf{X}^*_{(2)}]$$

$$= \frac{(r+s_0+s_r)!}{r!\,s_0!\,s_r!} \iint_{0 \leqslant l \leqslant u \leqslant 1} l^{s_0}(1-u)^{s_r}$$

$$\times \left[\frac{-\partial^2 \Pr[l \leqslant X'_{11} \leqslant X'_{1r} \leqslant u \mid \mathbf{X}^*_{(2)}]}{\partial l\,\partial u}\right] dl\,du$$

$$= \frac{(r+s_0+s_r)!}{r!\,s_0!\,s_r!} \left\{ r(r-1) \sum_{h=0}^{r} J(r-2, s_0, s_r; h)\,\theta^h Y_h \right.$$

$$\left. - 2\alpha^2 \sum_{h=0}^{r} J(r, s_0, s_r; h) \binom{h+2}{2} \theta^h Y_{h+2} \right\}, \tag{13}$$

where $Y_0 = 1$; $Y_h = \sum \cdots \sum_{j_1 < \cdots < j_h} \prod_{i=1}^{h} Z^*_{2j_i}$; $Z^*_{2j} = 1 - 2X^*_{2j}$ $(h, j = 1,..., r)$ and (with β, γ, δ, ε positive integers)

$$J(\beta, \gamma, \delta; \varepsilon) = \int_0^1 \int_0^u (u-l)^\beta\, l^\gamma (1-u)^\delta (1-u-l)^\epsilon \, dl\, du$$

$$= \sum_{i=0}^{\epsilon} (-1)^i \binom{\varepsilon}{i} \frac{\beta!\,(\gamma+i)!\,(\delta+\varepsilon-i)!}{(\beta+\gamma+\delta+\varepsilon+2)!} = \frac{\beta!\,\gamma!\,\delta!}{(\beta+\gamma+\delta+\varepsilon+2)!} G(\gamma, \delta; \varepsilon), \tag{14}$$

where

$$G(\gamma, \delta; \varepsilon) = \sum_{i=0}^{\epsilon} (-1)^\epsilon \binom{\varepsilon}{i} (\gamma+1)^{[i]} (\delta+1)^{[\epsilon-i]} \tag{15}$$

and $a^{[b]} = a(a+1) \cdots (a+b-1)$ is the bth ascending factorial of a.

Formula (13) can be written

$$L = \sum_{h=0}^{r} K(r, s_0, s_r; h)\theta^h Y_h \tag{16}$$

with

$$K(r, s_0, s_r; h) = \{G(s_0, s_r; h) - h(h-1)\, G(s_0, s_r; h-2)\}/(r+s_0+s_r+1)^{[h]}. \tag{17}$$

(If $\varepsilon < 0$, $G(s_0, s_r; \varepsilon)$ can be defined arbitrarily.)

We note that

$$K(r, s_0, s_r; 0) = G(s_0, s_r; 0) = 1. \tag{18}$$

For censoring from below ($s_r = 0$)

$$(r + s_0 + 1)^{[h]} K(r, s_0, 0; h) = (-1)^h s_0^{[h]} \tag{19}$$

and for censoring from above ($s_0 = 0$)

$$(r + s_r + 1)^{[h]} K(r, 0, s_r; h) = s_r^{[h]}. \tag{20}$$

For $s_0, s_r \geqslant 1$ we have the simple formula

$$(r + s_0 + s_r + 1)^{[h]} K(r, s_0, s_r; h) = G(s_0 - 1, s_r - 1; h). \tag{21}$$

This can be established by noting that $G(s_0, s_r; h) = h! \times$ (coefficient of x^h in expansion of $(1+x)^{-s_0-1}(1-x)^{-s_r-1}$), as also can the formula for symmetrical censoring,

$$\begin{aligned}(r + 2s + 1)^{[h]} K(r, s, s; h) &= 0 && \text{if } h \text{ is odd} \\ &= (k+1)^{[k]} s^{[k]} && \text{if } h = 2k.\end{aligned} \tag{22}$$

Summarizing, we have the following expressions for the likelihood ratios:

For detecting censoring from below:

$$L = 1 + \sum_{h=1}^{r} (-1)^r \frac{s_0^{[h]}}{(r + s_0 + 1)^{[h]}} \theta^h Y_h. \tag{23}$$

For detecting censoring from above:

$$L = 1 + \sum_{h=1}^{r} \frac{s_r^{[h]}}{(r + s_r + 1)^{[h]}} \theta^h Y_h. \tag{24}$$

For detecting symmetrical censoring:

$$L = 1 + \sum_{k \leqslant r/2} \frac{(k+1)^{[k]} s^{[k]}}{(r + 2s + 1)^{[2k]}} \theta^{2k} Y_{2k}. \tag{25}$$

In each case large values of the statistic are to be regarded as significant of censoring of the relevant type. Some numerical values for calculating the coefficients of $\theta^h Y_h$ in (23)–(25) are shown in Table I.

3.3. *Moments of L*

From the general theory of testing hypotheses, we have

$$E[L \mid H_{0,0}^{(1)}] = 1. \tag{26}$$

TABLE I

Values of $(r + s_0 + s_r + 1)^{[h]} K(r, s_0, s_r; h)$

		h					
s_0	s_r	1	2	3	4	5	6
0	1	1	2	6	24	120	720
0	2	2	6	24	120	720	5040
1	1	0	2	0	24	0	720
0	3	3	12	60	360	2520	20160
1	2	1	4	12	72	360	2880
0	4	4	20	120	840	6720	60480
1	3	2	8	36	216	1440	11520
2	2	0	4	0	72	0	2880

Note. (i) To obtain the coefficient of $\theta^h Y_h$ in the formula for L (see (16)) these numbers must be divided by $(r + s_0 + s_r + 1)^{[h]}$. Thus for $r = 5$, $s_0 = 1$, $s_5 = 2$ we have

$$L = 1 + \frac{1}{9}\theta Y_1 + \frac{4}{90}\theta^2 Y_2 + \frac{12}{990}\theta^3 Y_3 + \frac{72}{1180}\theta^4 Y_4 + \frac{360}{154,440}\theta^5 Y_5$$

$$= 1 + 0.111\theta Y_1 + 0.0444\theta^2 Y_2 + 0.0121\theta^3 Y_3 + 0.00606\theta^4 Y_4 + 0.00233\theta^5 Y_5.$$

(ii) Values of s_0 and s_r can be interchanged by multiplying entries by $(-1)^h$.

Under $H_{0,0}^{(1)}$, the Z_2^*'s are mutually independent and each is distributed uniformly over the interval $(-1, 1)$ so, for all θ,

$$\begin{aligned} E[(Z_2^*)^q \mid H_{0,0}^{(1)}] &= 0 && \text{if } q \text{ is odd} \\ &= (q+1)^{-1} && \text{if } q \text{ is even.} \end{aligned} \tag{27}$$

If follows that for any h, h' $(\neq h)$

$$E[Y_h \mid H_{0,0}^{(1)}] = 0 = E[Y_h Y_{h'} \mid H_{0,0}^{(1)}]$$

and (28)

$$\operatorname{var}(Y_h \mid H_{0,0}^{(1)}) = E[Y_h^2 \mid H_{0,0}^{(1)}] = \binom{r}{h} (\tfrac{1}{3})^h.$$

Hence when using the statistics (23), (24) testing for censoring from below or above ($s_0 = s$, $s_r = 0$ or $s_0 = 0$, $s_r = s$)

$$\operatorname{var}(L \mid H_{0,0}^{(1)}) = \sum_{h=1}^{r} \left\{ \frac{s^{[h]}}{(r+s+1)^{[h]}} \right\}^2 \binom{r}{h} \left(\frac{1}{3}\theta^2 \right)^h, \tag{29}$$

while when testing for symmetrical censoring ($s_0 = s_r = s$)

$$\operatorname{var}(L \mid H_{0,0}^{(1)}) = \sum_{k \leqslant r/r} \left\{ \frac{(k+1)^{[k]} s^{[k]}}{(r+2s+1)^{[2k]}} \right\}^2 \binom{r}{2k} \left(\frac{1}{3} \theta^2 \right)^{2k}. \tag{30}$$

Approximate significance limits for L may be obtained by supposing the distribution under $H_{0,0}^{(1)}$ to be approximately normal.

3.4. *Alternative Tests*

Since

$$\begin{aligned} E[Z_1^* \mid Z_2^*] = E[1 - 2X_1^* \mid Z_2^*] &= \int_0^1 (1 - 2x_1)\{1 + \theta Z_2^*(1 - 2x_1)\}\, dx_1 \\ &= \tfrac{1}{3}\theta Z_2^* \end{aligned} \tag{31}$$

it follows that

$$\begin{aligned} E[X_1^* \mid X_2^*] &= \tfrac{1}{2}[1 - \tfrac{1}{3}\theta(1 - 2X_2^*)] \\ &= \tfrac{1}{2} - \tfrac{1}{6}\theta(1 - 2X_2^*). \end{aligned} \tag{32}$$

Hence, the simplified test statistic L_1, defined in (6), is, for our FGM distribution and with $\theta > 0$,

$$L_1 = \{\tfrac{1}{2} + \tfrac{1}{6}\theta(2X_{21} - 1)\}^{s_0}\{\tfrac{1}{2} - \tfrac{1}{6}(2X_{2r} - 1)\}^{s_r}. \tag{33}$$

We note that if $s_0(s_r) = 0$ (and $\theta > 0$), the critical region becomes simply $X_{2r} < (X_{21} >)K$, with an appropriate value for the constant K. This would, of course, be the appropriate likelihood ratio test of the null hypothesis, with the alternative that X_2 itself has been subjected to censoring from above (below).

In the case of symmetric censoring ($s_0 = s_r = s$), the critical region (for all $s > 0$) is of the form

$$\{\tfrac{1}{2} + \tfrac{1}{6}\theta(2X_{21} - 1)\}\{\tfrac{1}{2} - \tfrac{1}{6}\theta(2X_{2r} - 1)\} > K \tag{34}$$

or equivalently

$$\tfrac{1}{9}\theta^2 X_{21}(1 - X_{2r}) + \tfrac{1}{3}\theta(\tfrac{1}{2} - \tfrac{1}{6}\theta)(X_{21} + \overline{1 - X_{2r}}) > K'. \tag{34$'$}$$

This can be compared with the critical regions

$X_{21}(1 - X_{2r})$ (for symmetrical censoring),

$X_{21} + (1 - X_{2r})$ (for general censoring—see Johnson and Kotz [6])

for likelihood ratio tests of $H_{0,0}^{(1)}$.

The values in Table I suggest that useful tests might be constructed by taking as test statistics the first terms only in the summations in (23)–(25). This would lead to critical regions (which do not depend on θ).

$$\text{For censoring from below:} \quad Y_1 < C. \tag{35}$$

$$\text{For censoring from above:} \quad Y_1 > C. \tag{36}$$

$$\text{For symmetrical censoring:} \quad Y_2 > C. \tag{37}$$

Since $Y_1 = \sum_{j=1}^{r} Z_{2j}^* = \sum_{j=1}^{r} (1 - 2X_{2j}^*)$, (35) and (36) are equivalent to

$$\sum_{j=1}^{r} X_{2j}^* > C', \tag{35$'$}$$

$$\sum_{j=1}^{r} X_{2j}^* < C', \tag{36$'$}$$

respectively. (The signs of the inequalities would be reversed if $\theta < 0$.)

On the null hypothesis $H_{0,0}^{(1)}$ (no censoring) the X_{2j}^*'s are mutually independent standard uniform variables. Therefore, even for r as small as 5, the distribution of their sum is closely approximated by a normal distribution with expected value $\frac{1}{2}r$ and variance $\frac{1}{12}r$ (e.g., Johnson and Kotz [6, p. 64]). So we obtain an approximate significance level α by taking

$$C' = \frac{1}{2}r + \lambda_\alpha \left(\frac{r}{12}\right)^{1/2} \qquad \text{in (35)}',$$

$$C' = \frac{1}{2}r - \lambda_\alpha \left(\frac{r}{12}\right)^{1/2} \qquad \text{in (36)}',$$

where $\Phi(\lambda_\alpha) = 1 - \alpha$.

From (A20)

$$E[Y_2 \mid H_{0,0}^{(1)}] = 0, \qquad \operatorname{var}(Y_2 \mid H_{0,0}^{(1)}) = (1/18)r(r-1). \tag{38}$$

Assuming Y_2 has an approximately normal distribution under $H_{0,0}^{(1)}$, we obtain an approximate significance level α for the test for symmetrical censoring (37) by taking

$$C = \lambda_\alpha \left(\frac{r(r-1)}{18}\right)^{1/2}. \tag{39}$$

The moments of the Y_h's under $H_{s_0',s_r'}^{(1)}$ may be evaluated by the following steps:

(i) find the conditional expected value, given $\mathbf{X}_1^*$, and

(ii) find the expected value of (i) when the joint distribution of the $\mathbf{X}_1^*$'s is that of the $(s_0'+1)$th, $(s_0'+2)$th,..., $(s_0'+r)$th order statistics among $(r+s_0'+s_r')$ variables, each with PDF $f_1(x)$.

Technical details are given in [5]. The results can be expressed in terms of quantities

$$(T_p(r, s_0', s_r')=)T_p = \sum_{j_1<\cdots<j_p}^{r-p+1} \cdots \sum^{r} E\left[\prod_{i=1}^{p} Z_{1j_i}^* \mid H_{s_0',s_r'}^{(1)}\right]$$

$$= \sum_{u=0}^{p} (-2)^u\{(r+s_0'+s_r'+1)^{[u]}\}^{-1} \sum_{h_1}^{r-p+1} \cdots \sum_{h_p}^{r}$$

$$\times \sum_{i_1<\cdots<i_u}^{p-u+1} \cdots \sum^{p} \sum_{\alpha=1}^{u} (s_0'+h_{i_\alpha}+\alpha-1). \tag{40}$$

After some rather heavy algebra we obtain (using $a^{(b)}=a(a-1)\cdots(a-b+1)$ to denote the bth descending factorial of a):

$$T_1 = \frac{r}{r+s_0'+s_r'+1}(s_r'-s_0'),$$

$$T_2 = \frac{r^{(2)}}{2\cdot(r+s_0'+s_r'+1)^{[2]}}\{(s_r'-s_0')^2+s_0'+s_r'\},$$

$$T_3 = \frac{r^{(3)}}{6\cdot(r+s_0'+s_r'+1)^{[3]}}\{(s_r'-s_0')\{(s_r'-s_0')^2+3(s_0'+s_r')+2\},$$

$$T_4 = \frac{r^{(4)}}{24\cdot(r+s_0'+s_r'+1)^{[4]}}\{(s_r'-s_0')^4+6(s_r'-s_0')^2(s_0'+s_r')$$

$$+11(s'_r-s_0')^2+6(s_0'+s_r')\}.$$

These values suggest the conjectural formula

$$T_p = \frac{r_{(p)}}{p!\,(r+s_0'+s_r'+1)^{[p]}} \sum_{j=0}^{p-1} |S_{p-1,j}|\,(s_r'-s_0')^{j+1}\left(\frac{s_0'+s_r'}{s_r'-s_r'}\right)^{\delta(p+j)}, \tag{41}$$

where

$$\delta(i) = 1 \qquad \text{if } i \text{ is even}$$
$$= 0 \qquad \text{if } i \text{ is odd}$$

and $|S_{p-1,j}|$ is the coefficient of a^j in the expansion of $a^{[p-1]}$ (Stirling number of the first kind).

In particular

$$E[Y_1 \mid H^{(1)}_{s_0', s_r'}] = \frac{r(s_r' - s_0')}{3(r + s_0' + s_r' + 1}\theta \tag{42}$$

and as special cases

$$E[Y_1 \mid H^{(1)}_{s,0}] = -\frac{rs}{3(r + s + 1)}\theta = -E[Y_1 \mid H^{(1)}_{0,s}]. \tag{43}$$

Also

$$\operatorname{var}(Y_2 \mid H^{(1)}_{s_0', s_r'}) = \frac{1}{3}r + \frac{r\{(r-1)(r + s_0' + s_r' + 1) - (2r + s_0' + s_r' + 1)(s_0' - s_r')^2\}}{9(r + s_0' + s_r' + 1)^2(r + s_0' + s_r' + 1)}. \tag{44}$$

Further useful formulae include

$$E[Y_1 Y_2 \mid H^{(1)}_{s_0', s_r'}] = \tfrac{1}{9}\theta(r-1)T_1 + \tfrac{1}{9}\theta^3 T_3, \tag{45}$$

$$E[Y_2^2 \mid H^{(1)}_{s_0', s_r'}] = \frac{r(r-1)}{18} + \frac{2\theta^2}{27}(r-2)T_2 + \frac{2\theta^4}{27}T_4. \tag{46}$$

Acknowledgments

I am grateful to a referee for comments leading to clarification of the explanation following Eq. (4).

References

[1] David, F. N. and Johnson, N. L. (1954). Statistical treatment of censored data. Part I: Fundamental formulae. *Biometrika* **41** 228–240.

[2] Johnson, N. L. (1971). Comparison of some tests of sample censoring of extreme values. *Austral. J. Statist.* **13** 1–16.

[3] Johnson, N. L. (1975). *Extreme Sample Censoring Problems with Multivariate Data—I.* Institute of Statistics, University of North Carolina Mimeo Series No. 1010.

[4] Johnson, N. L. (1978). Tests for censoring of extreme values (expecially) when population distributions are incompletely defined. In *Proceedings, Army Research Office Workshop on Robustness.*

[5] Johnson, N. L. (1979). Extreme sample censoring problems with multivariate data—II. Institute of Statistics, University of North Carolina, Mimeo Series No. 1237.

[6] Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics. Continuous Univariate Distributions—2.* Wiley, New York.

[7] Johnson, N. L. and Kotz, S. (1975). On some generalized Farlie–Gumbel–Morgenstern distributions. *Comm. Statist.* **4** 415–427.