A Recursive Algorithm for Null Distributions for Outliers: I. Gamma Samples

T. Lewis  
Department of Mathematical Statistics  
University of Hull, England

N. R. J. Fieller  
Department of Probability and Statistics  
University of Sheffield, England

A recursive method is described, which may be used to obtain certain distributional results relating to order statistics. The procedure is applied to gamma samples to obtain null distributions of various statistics appropriate for the testing of single and multiple outliers, and some useful inequalities for tail probabilities.

KEY WORDS
Outliers  
Order statistics  
Dixon's statistics  
Recursive algorithm  
Gamma samples

1. INTRODUCTION
Problems of analysis of outliers in samples from exponential parent populations have been studied by several authors, see e.g. Epstein [3], Laurent [6], Likeš [7], and Kabe [5]. As regards testing of outliers, some of this work has centred on statistics similar to Dixon's [2] in the case of testing outliers from a normal parent population. Other work has centred on statistics of the form $T_{n,i} = x_{i}/(x_{1} + \cdots + x_{n})$, where $x_{i}$ denotes the $i$th smallest observation in the random sample $x_{1}, \ldots, x_{n}$. The latter may be viewed as likelihood-ratio criteria, because the statistic is a monotonic function of the increase in the maximized log-likelihood on omission of the queried observation. We consider criteria of both kinds and show how their null distributions can be obtained by means of a recursive algorithm. We start by considering the statistic $T_{n,i}$ for an exponential sample. The result is known, Fisher [4], but this illustrates the procedure and leads naturally to the structure of the distribution noted by Fisher. The method readily extends to the more general case of a gamma population considered by Cochran [1]. We apply the technique to give some new results, and illustrate our procedure by numerical examples.

2. PRELIMINARIES
Let $x_{1}, \ldots, x_{n}$ be a sample from a gamma population

$$g(x, \lambda) = (\lambda^{r}/\Gamma(r)) x^{r-1} \exp(-\lambda x) \quad (x > 0) \quad (1)$$

where $r$ is known but $\lambda$ is unknown. Denote the ordered sample by $x_{(1)}, \ldots, x_{(n)} (x_{(1)} < \cdots < x_{(n)}).$

The criteria developed by Dixon [2] are of the form

$$y = [x_{(p)} - x_{(r)}]/[x_{(q)} - x_{(p)}] \quad (1 \leq p \leq r \leq s \leq q \leq n, q - p > s - r); \quad (2)$$

e.g. for testing the largest observation as an outlier we choose $s = q = n, r = n - 1, p = 1$.

For the likelihood-based criteria, let $S_{n} = \sum_{i=1}^{n} x_{i}$ and $T_{n,j} = x_{j}/S_{n}$. Also write

$$T_{n,i} = x_{i}/S_{n} \quad (3)$$

and in particular

$$T_{n,i} = \max_{j=1}^{n} T_{n,j} = x_{i}/S_{n}; \quad (4)$$

these will be abbreviated to $T_{(i)}$ when the sample size is obviously $n$. For any $j$, $T_{n,j}$ follows a beta distribution with parameters $r$ and $\eta(n-1)$, i.e. its density function is $\beta_{r,\eta(n-1)}(\cdot)$ where

$$\beta_{a,b}(u) = (\Gamma(a + b)/\Gamma(a)\Gamma(b))u^{a-1}(1 - u)^{b-1} \quad (0 < u < 1). \quad (5)$$

(Here, and throughout the paper, a density function whose value is specified only over certain regions is understood to be otherwise zero.)

The density functions and distribution functions of various statistics we now discuss will be denoted by $a_{n}(\cdot), A_{n}(\cdot); b_{n}(\cdot), B_{n}(\cdot); \cdots$ as they arise. The subscript $n$ refers to sample size.

Received August 1975, revised January 1978

371
3. A RECURRENT ALGORITHM FOR $T_n$

Suppose $T_{(n)}$ has density function $a_n(\cdot)$ and distribution function $A_n(\cdot)$. Then

$$a_n(u)du = P(T_{(n)} \leq (u, u + du))$$
$$= \sum_{j=1}^{n} P(T_{n,j} \in (u, u + du), T_{n,\{n\}} = T_{n,j})$$
$$= n P(T_{n,n} \in (u, u + du))$$
$$\cdot \max_{k=1}^{n-1} x_k/(S_n - x_n) < x_n/(S_n - x_n)).$$  (6)

Now

$$T_{n,n} = x_n/S_n = (x_n/(S_n - x_n))$$
$$(1 + x_n/(S_n - x_n)),$$  (7)

and $x_n/(S_n - x_n)$ is independent of $x_k/(S_n - x_n)$ for each $k = 1, \ldots, n-1$, (since the $x_k$ follow a gamma distribution). Hence $T_{n,n}$ and

$$\max_{k=1}^{n-1} x_k/(S_n - x_n)$$

are independent, and so (6) gives

$$a_n(u)du = n P(T_{n,n} \in (u, u + du))$$
$$P(T_{n-1,n-1} < u/(1 - u)).$$  (8)

We thus have the recurrence relation

$$a_n(u) = n A_{n-1}(u) A_n(u/(1 - u))((1/n \leq u \leq 1).$$  (9)

Consider first the particular case of an exponential parent population ($r = 1$). For $1/2 < u < 1$ we have

$$u/(1 - u) \leq 1, \ A_{n-1}(u/(1 - u)) = 1,$$

hence from (9)

$$a_n(u) = n A_{n-1}(u) A_n(u/(1 - u)) = ((1/n \leq u \leq 1).$$  (10)

$$A_n(u) = 1 - (1 - u)^{n-1}.$$  (11)

For $1/3 \leq u \leq 1/2$ we have $1/2 \leq u/(1 - u) \leq 1$; hence from (11) $A_{n-1}(u/(1 - u)) = 1 -(n - 1) ((1 - 2u)/(1 - u))^{n-2}$, and so from (9)

$$a_n(u) = n(n-1)(1 - u)^{n-2} - n(n-1)(1 - 2u)/(1 - u)^{n-2},$$  (12)

$$A_n(u) = 1 - n(1 - u)^{n-1} + \left(\frac{n}{2}\right)(1 - 2u)^{n-1};$$  (13)

and in general for $1/(q + 1) \leq u \leq 1/q, q \in [1, u],$

$$A_n(u) = 1 - n(1 - u)^{n-1} + \left(\frac{n}{2}\right)(1 - 2u)^{n-1}$$
$$- \cdots + (-)^n\left(\frac{n}{q}\right)(1 - qu)^{n-1},$$  (14)

which is the well-known result of Fisher [4].

For general $r$, the same procedure can be applied, using the recursive formula (9) to evaluate the functions $a_n(u), A_n(u)$ in successive intervals $1/(s + 1) \leq u \leq 1/s (s = 1, \ldots, n - 1)$. In particular, since $A_{n-1}(u/(1 - u)) = 1$ when $u \geq 1/2$ and $< 1$ when $u < 1/2$, the upper tail probability $P(T_{(n)} > u)$, which can be used as a significance probability for testing an upper outlier, satisfies

$$P(T_{(n)} > u) = 1 - A_n(u) = n \int_{u}^{1} \beta_{r,r(n-1)}(t)dt$$

$(u \geq 1/2),$  (15a)

$$P(T_{(n)} > u) < n \int_{u}^{1} \beta_{r,r(n-1)}(t)dt$$

$(u < 1/2).$  (15b)

This may conveniently be written (upon making the usual transformation taking the beta density into the $F$-density) as

$$P(T_{(n)} > u) = n P(F_{2r,2r(n-1)} > (n - 1)u/(1 - u))$$

$(u \geq 1/2),$  (16a)

$$P(T_{(n)} > u) < n P(F_{2r,2r(n-1)} > (n - 1)u/(1 - u))$$

$(1/ n < u < 1/2),$  (16b)

where $F_{2r,2r(n-1)}$ denotes the usual variance-ratio statistic on $2r, 2r(n - 1)$ degrees of freedom.

4. FURTHER RESULTS

4.1 Null Distribution of $T_{(1)}$

We first apply the method of Section 3 to find the distribution of $T_{(1)}$, a suitable statistic for testing the smallest observation in a gamma sample of size $n$. If $T_{(1)}$ has density function $a_n(\cdot)$ and distribution function $A_n(\cdot)$, an argument similar to that of Section 3 gives the recursive relation

$$a_n(u) = n A_{n-1}(u) A_n(u/(1 - u))$$

$(0 \leq u \leq 1/n).$  (17)

For $n = 2, B_1(u/(1 - u)) = 0$ or 1 according as $u/(1 - u) < 1$ or $> 1,$ hence

$$b_2(u) = 2\beta_{r,r}(u) = 2(\Gamma(2r)/(\Gamma(r))^2)u^{r-1}(1 - u)^{r-1}$$

$(0 \leq u \leq 1/2).$  (18)

Proceeding recursively we find, in the case of an exponential population, $(r = 1),$

$$b_2(u) = 2(0 \leq u \leq 1/2),$$

$$b_n(u) = n A_{n-1}(1 - nu)^{n-2}$$

$(0 \leq u \leq 1/n).$  (20)

In the case $r = 2,$

$$b_2(u) = 12u(1 - u)$$

$(0 \leq u \leq 1/2),$  (21a)
whence recursively

\[ b_1(u) = 60u(1 - 3u)(1 - 3u^2) \quad (0 \leq u \leq 1/3), \quad (21b) \]
\[ b_2(u) = 168u(1 - 4u)(1 + 3u - 12u^2 - 4u^3) \quad (0 \leq u \leq 1/4), \quad (21c) \]
\[ b_3(u) = 360u(1 - 5u^2)(1 + 8u - 80u^2 - 65u^3) \quad (0 \leq u \leq 1/5) \quad (21d) \]

and so on. As a general inequality, analogous to (7), for the lower tail probability appropriate for testing a lower outlier, we have

\[ P(T_{1,1} < u) \leq n P(F_{2r,2(n-1)} < (n - 1)u/(1 - u)) \quad (u < 1/n). \quad (22) \]

4.2 Null Distribution of \( T_{1,1}, T_{1,n} \)

Suppose \( T_{1,1}, T_{1,n} \) in a gamma sample have joint density function \( c_{n1}(.;.) \) and distribution function \( C_{n1}(.;.) \). \( c_{n1}(u, v) \) is clearly zero outside the region \( R_n \), defined by \( u > 0, (n - 1)u + v < 1, u + (n - 1)v > 1 \). Now the joint density of \( T_{1,1} \) and \( T_{1,n} \) for any \( i \) and \( j \), is

\[ \Gamma(n)/(\Gamma(r))^2 \Gamma((n - 2)) \] 
\[ \cdot u^{r-1}v^{r-1}(1 - u - v)^{r(n - 2) - 1} \quad (23) \]

(see, for example, Wilks [8], section 7.7). Thus, on applying the method of Section 3, we find, for \((u, v) \in R_n \) and \( n \geq 3 \),

\[ c_{n1}(u, v) = n(n - 1)\Gamma(n)/\Gamma(r)^2 \Gamma((n - 2)) \] 
\[ \times u^{r-1}v^{r-1}(1 - u - v)^{r(n - 2) - 1} \Psi \quad (24) \]

where

\[ \Psi = P(u/(1 - u - v) < T_{n-2,11} < T_{n-2,n-2}) \] 
\[ < v/(1 - u - v)). \quad (25) \]

For \( n \geq 5 \), (25) can be written

\[ \Psi = A_{n-2} \{w/(1 - u - v) - C_{n-2} \{u/(1 - u - v), \] 
\[ v/(1 - u - v)\}. \quad (26) \]

When \( n = 3 \), the probability \( \Psi \) in (25) is 1 if \( u/(1 - u - v) < 1 \) and \( v/(1 - u - v) > 1 \), and is 0 otherwise;

\[ c_{11}(u, v) = 3!\{\Gamma(3r)/\Gamma(r)^3\} u^{r-1}v^{r-1}(1 - u - v)^{r-1} \] 
\[ (2u + v < 1, u + 2v > 1, u > 0). \quad (27) \]

When \( n = 4 \), \( \Psi \) is equal to

\[ P(T_{1,1}) > max \{1 - (v/(1 - u - v)), u/(1 - u - v)\}) \] 
\[ = 2\Gamma(2r)/\Gamma(r)^2 \int_0^{1/2} r^{r-1}(1 - t)^{r-1}dt \quad (28) \]

where

\[ w = (1 - u - 2v)/(1 - u - v) \quad (u + 2v < 1, (u, v) \in R_c \] 
\[ u/(1 - u - v) \quad (u + 2v > 1, (u, v) \in R_c \] 

Evaluation of the integral in (28) and application of (24) with \( n = 4 \) gives

\[ c_{41}(u, v) = 4!\{\Gamma(4r)/\Gamma(r)^4\} u^{r-1}v^{r-1} \] 
\[ \times \sum_{j=0}^{r-1} \left( \begin{array}{c} r-1 \\ j \end{array} \right) (1 - u - v)^{r-1-j}z^j \quad (29) \]

where

\[ z = u + 3v - 1 \quad (2u + 2v < 1, \] 
\[ u + 3v > 1, \quad u > 0) \]

and

\[ z = 1 - 3u - v \quad (2u + 2v > 1, \] 
\[ 3u + v > 1, \quad u > 0) \]

Particularizing to the exponential case (\( r = 1 \)), this gives

\[ c_{41}(u, v) = 1.2^r.3 (u + 2v > 1, 2u + v < 1, u > 0), \quad (30a) \]
\[ c_{41}(u, v) = 2.3^r.4 (u + 3v - 1) \] 
\[ (u + 3v > 1, 2u + 2v < 1, u > 0), \quad (30b) \]
\[ c_{41}(u, v) = 3.4^r.5 (u + 4v - 1)^2 \] 
\[ (u + 4v > 1, 2u + 3v < 1, u > 0), \quad (30c) \]

and so on. The general form of the distribution is clear.

Again, particularizing to the case \( r = 2 \) we get

\[ c_{41}(u, v) = \frac{720}{20}w(1 - u - v) \] 
\[ (u + 2v > 1, 2u + v < 1, u > 0), \quad (31a) \]
\[ c_{41}(u, v) = \frac{100800}{20}w(1 - u - v)(1 - u)^2 - 3u^2 \] 
\[ (u + 3v > 1, 2u + 2v < 1, u > 0), \quad (31b) \]
\[ c_{41}(u, v) = \frac{100800}{20}(1 - u - v)(1 - v)^2 - 3u^2 \] 
\[ (2u + 2v > 1, 3u + v < 1, u > 0), \quad (31c) \]

whence \( c_{n1}(u, v) \) can be obtained recursively for further values of \( n \).
4.3 Null Distribution of $T_{(n)} - T_{(1)}$

From the joint density function $e_n(\cdot, \cdot)$ of $T_{(1)}, T_{(n)}$, expressions can be obtained for the density function, $d_n(\cdot)$ say, of $T_{(n)} - T_{(1)}$, a statistic which may be used for testing simultaneously the largest and smallest observations as outlying in a gamma sample of size $n$. Naturally it only makes sense to test for two outliers when $n$ is 5 or more. In the exponential case, $r = 1$, we find successively, for purposes of the recurrence,

\begin{align*}
    d_3(u) &= 4u \quad (0 \leq u \leq 1/2), \\
    4[u-(2u-1)] & \quad (1/2 \leq u \leq 1), \\
    d_4(u) &= 9(2u^2) \quad (0 \leq u \leq 1/3), \\
    9[2u^2-(3u-1)^2] & \quad (1/3 \leq u \leq 1/2), \\
    d_5(u) &= 16(6u^3) \quad (0 \leq u \leq 1/4), \\
    16(6u^3-(4u-1)^3) & \quad (1/4 \leq u \leq 1/3), \\
    d_6(u) &= 16(6u^3-(4u-1)^3+3(3u-1)^3) \quad (1/3 \leq u \leq 1/2), \\
    16(6u^3-(4u-1)^3+3(3u-1)^3-3(2u-1)^3) & \quad (1/2 \leq u \leq 1),
\end{align*}

the form of $d_n(u)$ for general sample size $n$ is clear.

A corresponding calculation in the case $r = 2$ starts with the results

\begin{align*}
    d_3(u) &= (40/9)u(2-5u^2) \quad (0 \leq u \leq 1/2), \\
    (40/9)(1-u)^3(1+4u) & \quad (1/2 < u < 1), \\
    d_4(u) &= (21/32) u^2(90-450u^2+182u^4) \quad (0 \leq u \leq 1/3), \\
    (21/32)u^2(13+101u^2-2(1-2u)^2(13+104u+100u^2)) & \quad (1/3 \leq u \leq 1/2), \\
    (21/32)(1-u)(13+101u) & \quad (1/2 \leq u \leq 1), \\
\end{align*}

further details are not given here.

4.4 Two Upper Outliers

Suppose $j_n(\cdot, \cdot)$ is the joint density function of $T_{(1)}$ and $T_{(n-1)}$ in a gamma sample; $j_n(u, v)$ will be zero outside the region $R_r$ defined by $u + v < 1, u + (n - 1)v > 1, u > v$. For $(u, v) \in R_r$, the procedure of Section 3 gives the recurrence relation

$$
    j_n(u, v) = n(n - 1)(\Gamma(n)/(\Gamma(r))^2\Gamma(r(n - 2))) \times u^{r-1}v^{r-1}(1 - u - v)^{r-2}B_{n-1}(v/(1 - u)).
$$

In the exponential case $(r = 1)$ this becomes, from (34),

$$
    g_n(u, v) = n(n - 1)^2(1 - u - v)^{n-2}B_{n-1}(v/(1 - u)).
$$

A useful statistic in the testing of a pair of upper outliers in a gamma sample is $T_{(n)} + T_{(n-1)}$, say, whose density function $h_n(\cdot)$ can be found in principle from the above joint density. Even without the exact distribution of this statistic, we have by the method of Section 3 the following inequality:

$$
    P(Z_n > u) \leq |n(n - 1)/2|\beta_{2r,\Gamma(r-2)}(u),
$$

whence

$$
    P(Z_n > u) \leq |n(n - 1)/2| P(F_{2r,\Gamma(r-2)} > (n - 2)/2(u/(1 - u))).
$$

The generalization to three or more upper outliers is again clear.

4.5 Two Lower Outliers

Suppose $j_n(\cdot, \cdot)$ is the joint density function of $T_{(1)}$ and $T_{(2)}$; $j_n(u, v)$ is clearly zero outside the region $R_r$ defined by $u + v < 1, u + (n - 1)v > 1, u > v$. For $(u, v) \in R_r$, the procedure of Section 3 gives the recurrence relation

$$
    j_n(u, v) = n(n - 1)(\Gamma(n)/(\Gamma(r))^2\Gamma(r(n - 2))) \times u^{r-1}v^{r-1}(1 - u - v)^{r-2}B_{n-1}(v/(1 - u)).
$$

In the exponential case $(r = 1)$ this becomes, from (38),

$$
    j_n(u, v) = n(n - 1)^2(1 - u - v)^{n-2}B_{n-1}(v/(1 - u)).
$$

In the general case of $m$ lower outliers the same method gives the joint density $j_n(u, v, \cdots, u, z)$ of $T_{(1)}$, 

TECHNOMETRICS®, VOL. 21, NO. 3, AUGUST 1979
A RECURSIVE ALGORITHM FOR NULL DISTRIBUTIONS FOR OUTLIERS

...$T_{(m)}$ for $(u, v, \cdots, z) \in R$, the region defined by $u + v + \cdots + y + (n-m+1)z < 1$.

$$0 < u < v \cdots < z$$

as

$$j_n(u, v, \cdots, y, z) = \frac{(n)!}{(n - k)!} \frac{1}{(r(n - m + 1))} \times b_{n-m+1}(z/(1-u-v-\cdots-y)),$$

where $(n)_k = n!/(n - k)!$, which in the exponential case $(r = 1)$ becomes

$$j_n(u, v, \cdots, y, z) = \frac{(n)!}{(n - k)!} \frac{1}{(r(n - m + 1))} \times b_{n-m+1}(z/(1-u-v-\cdots-y)).$$

An appropriate statistic for testing the significance of $m$ lower outliers in a gamma sample is $T_{(m)}$, say. The density function $k_n(t)$ of $W_n$ can be found in principle by transforming the above density. In the case of two lower outliers in an exponential sample ($r = 1$, $m = 2$) this yields

$$k_n(t) = \frac{n(n-1)}{(n-2)}((1-t)^{n-2}(0 < t < 1/(n-1)),$$

$$1/(n-1) < t < 2/n).$$

Even without the exact distribution of $W_n$ in the general case of $m$ outliers, the method of Section 3 yields the following inequality

$$k_n(u) \leq \left(\frac{n}{m}\right)^{r_m} \beta_{r_m, r(n-m)}(u),$$

whence

$$P(W_n \leq u) \leq \left(\frac{n}{m}\right)^{r_m} P(F_{2p/n-m-1, 2p} > m(1-u)/(u(n-m))).$$

4.6 A Dixon Statistic

We have shown that the recursive methods of Section 3 can be applied to likelihood-ratio statistics to useful effect, producing explicit results in some cases and bounds in others. The methods can also be applied to statistics of the Dixon type, though for the larger sample sizes the calculations become involved. We illustrate the principle by considering the statistic

$$y = (x_{(2)} - x_{(1)})/(x_{(n)} - x_{(1)})$$

which has been discussed in the particular context of an exponential sample by Likeş [7] and Kabe [5].

Consider first the joint distribution of $T_{(1)}$, $T_{(2)}$, $T_{(n)}$ for a gamma sample; write $l_n(t, \cdots)$ for its density function. Clearly $l_n(t, \cdots)$ will be zero outside the region $K_t$ defined by

$$u + (n-2)v + w < 1, \quad u + v + (n-2)w > 1,$$

$$w > v > u > 0.$$ For $(u, v, w) \in R$, and $n \geq 4$, the method of Section 3 gives

$$L_n(u, v, w) = n(1 - u)^{(n-1)} e^{n-1} c_{n-1},$$

where

$$c_{n-1}((1-y)u + yw)/(1-u),$$

for $(u, v, w) \in R$ defined by

$$(1-y)u + yw < 1 - u, 0 < y < 1, u > 0.$$

From this we can find the distribution of $y$ using

$$\int (1-y)u + yw + (n-2)w > 1 - u, 0 < y < 1, u > 0.$$
TABLE 1—Frequencies of excess cycle times in steel manufacture.

<table>
<thead>
<tr>
<th>Excess cycle time X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>18</td>
<td>12</td>
<td>18</td>
<td>16</td>
<td>10</td>
<td>4</td>
<td>9</td>
<td>9</td>
<td>2</td>
<td>7</td>
</tr>
</tbody>
</table>

Integration with respect to u and w gives the density function of y as

\[
\begin{align*}
&3/8 \left( 11(2y + 1)^{-2} + 44(2y + 1)^{-3} \\
&- 69(2y + 1)^{-4} + 24(2y + 1)^{-5} \\
&- 10(2y + 2)^{-2} - 44(2y + 2)^{-3} - 12(2y + 2)^{-4} \\
&+ 192(2y + 2)^{-5} \right) (0 < y < 1).
\end{align*}
\]

5. ILLUSTRATIVE EXAMPLES

5.1 Example 1*

Consider example (6) given by Epstein [3], p. 171. Here the failure times of 10 items are observed, totaling \( \sum x_i = 600 \) units. The failure times of the first two items to fail are the shortest of the 10, and total 24 units, so we can write \( x_{11} + x_{12} = 24 \). Under the assumption that failure times under given conditions are exponentially distributed, Epstein tests whether the first two items to fail can be regarded as having failed abnormally early; making use of a straightforward \( F \)-test, he concludes in favour of this hypothesis. Suppose however that the 10 items were placed on test at different starting times and that the two shortest failure times occurred, not necessarily first, but randomly in chronological sequence, so that there was no a priori reason to consider these two items as different from the rest. How strong would the evidence then be for regarding them as inconsistent with the rest in view of their failure times? In the notation of Section 4.5 we have \( T_{11} + T_{12} = W_{10} \) whose value, with \( m = 2 \), is

\[ w = \frac{24}{600} \text{, lying in the range } 0 < w < \frac{1}{9}. \]

From (42),

\[
P(W_{10} < 0.04) = \int_{0}^{0.04} \frac{810}{8} \left[ (1 - 5u)^6 - (1 - 9u)^6 \right] du
\]

\[ = \frac{90}{8} \left( \frac{(0.64)^6}{9} - \frac{(0.80)^6}{5} - \frac{1}{9} + \frac{1}{5} \right) = 0.721. \]

This is the significance probability attaching to the observed ratio \( w = 0.04 \), i.e. there is no real evidence for regarding the figure of 24 units as abnormally low—a contrary conclusion to Epstein’s on our modified premise.

5.2 Example 2

Table 1 shows a sample of 132 excess cycle times in steel manufacture (source of data: private communication).

The sample of size 130 obtained by omitting the outliers \( x_{131} = 92 \) and \( x_{132} = 97 \) has mean \( \bar{x} = 6.44 \), standard deviation \( s = 6.18 \), and third and fourth moments about the mean \( m_3 = 493 \), \( m_4 = 1.34 \times 10^4 \). Hence \( \bar{x}/s = 1.04 \), \( m_3/s^3 = 2.09 \), \( m_4/s^4 = 9.24 \), suggesting that the \( X \)-distribution may reasonably be assumed exponential (\( \mu/s = 2 \), \( \mu_3/s^3 = 9 \) for an exponential distribution). On this assumption we can test the outliers 92, 97 for consistency with the other 130 values, using the statistic \( Z_0 \) of Section 4.4. Its value for the sample is \( (92 + 97)/1043 = 0.1812 \), so in the null case, from (37),

\[
P(Z_{132} > 0.1812) = 0.1812 \leq \frac{132 \times 131}{2}
\]

\[
P(F_{4,289} > (130/2)(0.1812/0.8188)) \approx 8646 P(x^2 > 57.54) = (8646)(29.77) \exp(-28.77) = 8 \times 10^{-8},
\]

i.e., the evidence for regarding the values 92, 97 as not belonging to the same exponential distribution as the other 130 values is very strong.

6. CONCLUSION

The recursive procedure described in this paper allows ready calculation of the null distributions of various statistics for testing outliers in gamma samples.

Exact significance levels can then be calculated on hand calculators, in many cases avoiding the need for extensive tables of percentage points.

7. ACKNOWLEDGMENTS

We are grateful to the referees for helpful comments.

REFERENCES