[1]This research was supported in part by the U. S. Army Research Office, Durham, under Contract No. DAHC04-71-C-0042.

ESTIMATION OF RANK ORDER[1]

by

N. L. Johnson

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series Number 931

June 1974

ESTIMATION OF RANK ORDER

By N. L. Johnson

University of North Carolina at Chapel Hill

1. Introduction

The *rank order* of a member a_i of a set of n numbers $(a_1, a_2, \dots a_n)$ is the number of members of the set less than or equal to a_i (including, of course, a_i itself). If no two a's are equal, the rank of the g-th smallest a_i is g.

If $X_1, X_2, \dots, X_n$ are continuous random variables corresponding to a random sample, then with probability one, no two are equal and so each X_i will have a unique rank order different from that of any other X_j $(j \neq i)$.

In this paper we will be concerned with the estimation of the rank order of a specified X_i when only partial information about the sample values $X_1, X_2, \dots, X_n$ is available. The discussion will be restricted to cases where $X_1, X_2, \dots, X_n$ are independent and identically distributed continuous random variables with common cumulative distribution function $F(x)$ and density function $f(x)$.

Throughout the paper there are two different sampling schemes under consideration. In one an individual value is chosen at random and so its rank order G is a random variable, while in the other it is supposed that the rank order g is fixed, though remaining observed values (if any) are chosen at random.

Tests relating to rank order will be in a later report.

2. Population Distribution Unknown; r Sample Values Known $(2 \le r < n)$

If we are given the values $(X_1', X_2', \ldots X_r')$ of a randomly chosen subset of $r(\ge 2)$ of the n values $X_1, X_2, \ldots, X_n$ we can evaluate the probability that a randomly chosen X' is the g-th smallest among $X_1, X_2, \ldots, X_n$ (i.e. has rank order $G = g$ in this set). Note that G is a random variable.

The conditional probability that $G = g$, given that the rank order, G', among $(X_1', X_2', \ldots, X_r')$ is g' in

$$\Pr[G = g \mid G' = g'] = \frac{\Pr[G=g]\Pr[G'=g' \mid G=g]}{\Pr[G'=g']} .$$

Since $\Pr[G = g] = n^{-1}$, $\Pr[G' = g'] = r^{-1}$ $(g = 1, \ldots, n\ ;\ g' = 1, \ldots, r)$ and

$$\text{(1)} \qquad \Pr[G' = g' \mid G = g] = \binom{g-1}{g'-1}\binom{n-g}{r-g'} \Big/ \binom{n-1}{r-1}$$

$$\left\{\max(1, r + g - n) \le g' \le \min(g, r)\right\}$$

we have

$$\text{(2)} \qquad \Pr[G = g \mid G' = g'] = \binom{g-1}{g'-1}\binom{n-g}{r-g'} \Big/ \binom{n}{r}$$

$$(g = g', g' + 1, \ldots, g' + n - r) .$$

Note that (1) implies that $(G' - 1)$ has a hypergeometric distribution with parameters $(r - 1), (g - 1), (n - 1)$.

Now

$$\frac{\Pr[G=g \mid G=g']}{\Pr[G=g-1 \mid G=g']} = \frac{(g-1)(n-r+1+g'-g)}{(g-g')(n+1-g)}$$

$$(g = g' + 1, \ldots , g' + n - r) .$$

This ratio is greater or less than 1 , according as

$$(g - 1)(n - r + 1) \gtrless n(g' - g) ,$$

i.e. as

$$g \lessgtr \frac{n(g'-1)}{r-1} + 1 .$$

The value of $\Pr[G = g \mid G' = g']$ is therefore minimized by

$$g = \tilde{g} = \left[\frac{n(g'-1)}{r-1} + 1\right] \tag{3}$$

([ ] denotes "integer part of".) . If $n(g' - 1) / (r - 1)$ is an integer (as it is sure to be if $n / (r - 1)$ is an integer) then

$$\Pr[G = \tilde{g} - 1 \mid G' = g'] = \Pr[G = \tilde{g} \mid G' = g']$$

and the common value is $\max_g P[G = g \mid G' = g']$.

One might consider $\tilde{g}$ as a "maximum probability" estimator of g . It does indeed maximize the likelihood (1) (as well as (2)) .

We are thus led to consider the statistic

$$\tilde{G} = \left[\frac{n(G'-1)}{r-1} + 1\right]$$

as an estimator of g .

The properties of $\tilde{G}_a = \frac{n(G'-1)}{r-1} + 1$ are easily established. (Note that if $n/(r - 1)$ is an integer, $\tilde{G}_a = \tilde{G}$) . The distribution of $(G' - 1)$, given $G = g$, is hypergeometric (see (1)) and

$$(4.1) \qquad E[G' \mid g] = 1 + (r - 1)(g - 1) / (n - 1)$$

$$(4.2) \quad \mathrm{var}[G'|g] = [(n - r)\ (n - 2)](r - 1)[(g - 1)/(n - 1)][1 - (g - 1)/(n-1)]$$
$$= (n - r)(r - 1)(g - 1)(n - g)(n - 1)^{-2}(n - 2)^{-1}$$

whence

$$(5.1) \qquad E[\tilde{G}_a \mid g] = [n/(n - 1)]\ (g - 1) + 1$$

$$(5.2) \qquad \mathrm{var}\ [\tilde{G}_a \mid g] = \frac{n^2(n-r)}{(n-1)^2(n-2)} \cdot \frac{(g-1)(n-g)}{r-1}$$

$$(5.3) \qquad \tilde{T} = (r - 1)^{-1}\ (n - 1)\ (G' - 1) + 1$$

is an unbiased estimator of g .

Note that $\mathrm{var}\ [G_a \mid g] = 0$ if either $g = 1$ or $g = n$, as is to be expected.

3. Population Distribution Known: Only One Sample Value Known

The analysis of Section 2 can be applied only if $r \geq 2$. However, if the common density function of $X_1, X_2, \ldots , X_n$ is known, then it is possible to derive a distribution for the rank order G of a randomly chosen X , even if no other X-values are known. The probability that $G = g$, given $X = x$, is the probability that $(g - 1)$ of the remaining X's are less than x , and the other $(n - g)$ are greater than x . This is

(6) $\Pr[G = g \mid X = x] = \binom{n-1}{g-1} \{F(x)\}^{g-1} \{1 - F(x)\}^{n-g} \quad (g = 1, 2, \ldots, n)$.

The value of g which maximizes (6) is

$$\hat{g} = [nF(x) + 1] .$$

If $nF(x)$ is an integer then

$$\Pr[G = nF(x) + 1 \mid X = x] = \Pr[G = nF(x) \mid X = x]$$

and the common value maximizes $\Pr[G = g \mid X = x]$.

We can consider

$$\hat{G} = [nF(X) + 1]$$

at a maximum probability estimator of g .

Considering the related statistic

$$\hat{G}_a = nF(X) + 1$$

we note that $F(X)$ has a beta distribution with parameters $(g, n - g + 1)$ and so

(7.1) $$E[F(X) \mid g] = g/(n + 1)$$

(7.2) $$\operatorname{var}[F(X) \mid g] = g(n - g + 1)(n + 1)^{-2}(n + 2)^{-1}$$

whence

(8.1) $$E[\hat{G}_a \mid g] = gn(n + 1)^{-1} + 1$$

(8.2) $$\operatorname{var}[\hat{G}_a \mid g] = g(n - g + 1)\, n^2 (n + 1)^{-2} (n + 2)^{-1}$$

(8.3) $\hat{T} = (n + 1)\,F(X)$ is an unbiased estimator of g .

4. Population Distribution Known: $r(\geq 2)$ Sample Values Known

In this case we have information of each of the kinds described in Sections 2 and 3. The two sources of information may be combined to give information on the rank G of an individual X among n randomly chosen values, given its rank G' among a randomly chosen subset of $r(\geq 2)$ among the n values and the value of the cumulative distribution function $F(X)$, for the individual concerned. The exact values of X for the other $(r - 1)$ members of the subset contain no further information beyond the value of G' (to which of course they contribute). (In practice, of course, if these values were abnormal they might raise doubts whether the correct function $F(x)$ was being used for the cumulative distribution function.)

The conditional distribution of G is

$$(9) \quad \Pr[G = g \mid (X = x) \cap (G' = g')] = \binom{n-r}{g-g'} \{F(x)\}^{g-g'} \{1 - F(x)\}^{n-r-g+g'}$$

$$(g = g', g' + 1, \ldots , g' + n - r) .$$

This is maximized by

$$g = \hat{g} = [(n - r + 1)\,F(x) + g'] .$$

If $(n - r + 1)\,F(x)$ is an integer, then the values $g = \hat{g}$ and $g = \hat{g}$ give equal values for $\Pr[G = g \mid (X = x) \cap (G' = g')]$ and thus maximize this probability.

Consider now the estimator

$$\hat{\tilde{G}}_a = (n - r + 1)\, F(X) + G' .$$

If $G = g$ then

$$\text{(10.1)} \qquad E[\hat{\tilde{G}}_a \mid g] = (n - r + 1) \frac{g}{n+1} + (r - 1) \frac{g-1}{n-1} + 1$$

$$= \left\{1 + \frac{2r-n-1}{n^2-1}\right\} g + \frac{n-r}{n-1} .$$

Conditional on $G = g$, the variables $F(X)$ and G' are mutually independent, and so

$$\text{(10.2)} \quad \text{var}\, [\hat{\tilde{G}}_a \mid g] = (n - r + 1)^2\, \text{var}\, [F(X) \mid g] + \text{var}\, [G' \mid g]$$

$$= \frac{(n-r+1)^2}{(n+1)^2(n+2)}\, g(n - g + 1) + \frac{(n-r)(r-1)}{(n-1)^2(n-2)} (g - 1)(n - g) .$$

Also

$$\text{(10.3)} \qquad \hat{\tilde{T}} = \left\{G_a - \frac{n-r}{n-1}\right\} \left\{1 + \frac{2r-n-1}{n^2-1}\right\}^{-1}$$ is an unbiased estimator of g .

We have

$$\text{(11.1)} \quad \text{var}\, [\tilde{T} \mid g] = (n - r)(r - 1)^{-1} (n - 2)^{-1} (g - 1)(n - g)$$

$$\text{(11.2)} \quad \text{var}\, (\hat{T} \mid g) = (n + 2)^{-1} g(n - g + 1)$$

$$\text{(11.3)} \quad \text{var}[\hat{\tilde{T}} | g] = \left\{\frac{(n-r+1)^2}{(n+1)^2(n+2)} g(n - g + 1) + \frac{(r-1)(n-r)}{(n-1)^2(n-2)} (g - 1)(n - g)\right\}$$

$$\times \left\{1 + \frac{2r-n-1}{n^2-1}\right\}^{-2}$$

If g, r, and n are large, with $g/n = \gamma$, $r/n = \omega$ then

(12.1) $$n \operatorname{var} [g^{-1}\tilde{T} \mid g] \doteq (\omega^{-1} - 1)(\gamma^{-1} - 1)$$

(12.2) $$n \operatorname{var} [g^{-1}\hat{T} \mid g] \doteq \gamma^{-1} - 1$$

(12.3) $$n \operatorname{var} [g^{-1}\hat{\tilde{T}} \mid g] \doteq (1 - \omega)(\gamma^{-1} - 1) .$$

Hence

$$\operatorname{var} [\tilde{T} \mid g] : \operatorname{var} [\hat{T} \mid g] : \operatorname{var} [\hat{\tilde{T}} \mid g] \doteq \omega^{-1} : (1 - \omega)^{-1} : 1 .$$

These approximate values do not depend on $\gamma (= g/n)$. Since they are based on the assumption that g/n is fixed, and not tending to 0 or 1, they will not apply when g is near 1 or n.

In large samples, we will have

$$\operatorname{var} [\tilde{T} \mid g] \gtrless \operatorname{var} [\hat{T} \mid g]$$

according as $\omega = r/n \lessgtr \frac{1}{2}$. It is to be expected that as r increases the estimator based on rank among a subset of r values will become more accurate. It appears that a "break-even" point (as compared with an estimator based on knowledge of $F(X)$) is attained when about half of the sample values are included in the subset. We note, also, that this estimator can be computed, even when we have no knowledge of $F(X)$, except that it is continuous.

As might be expected, $\hat{\tilde{T}}$ has a smaller variance than either $\tilde{T}$ or $\hat{T}$, since it combines information from both G' and $F(x)$.

Table 1 shows some values of efficiencies (reciprocal ratio of variances) of $\tilde{T}$ and $\hat{T}$ relative to $\hat{\tilde{T}}$ for a few values of n, g and r . The values of eff[T | g] = var [T | g]/var [$\hat{\tilde{T}}$ |g] do not depend on the distribution function F(x) . Even for quite small n , they do not vary much with g either (for n and r fixed), provided g is not near 1 or n .

Table 1: Relative Efficiencies of Some Unbiased Estimators of Rank Order

(For $g > \frac{1}{2}(n + 1)$,use eff [T | n - g + 1] = eff [T | g])

n	r	g	eff[$\tilde{T}$|g]	eff[$\hat{T}$|g]
5	2	2	0.28	0.72
		3	0.24	0.76
	3	2	0.54	0.47
		3	0.49	0.51
	4	2	0.80	0.23
		3	0.75	0.26
10	2	2	0.16	0.85
		4	0.11	0.83
		5	0.11	0.89
	4	2	0.44	0.59
		4	0.34	0.66
		5	0.33	0.67
	5	2	0.57	0.47
		4	0.45	0.55
		5	0.44	0.56
	6	2	0.68	0.36
		4	0.57	0.44
		5	0.56	0.44
	8	2	0.90	0.17
		4	0.79	0.22
		5	0.78	0.22
20	5	4	0.24	0.77
		8	0.21	0.78
		10	0.21	0.79
	10	4	0.51	0.49
		8	0.47	0.52
		10	0.47	0.52
	15	4	0.77	0.24
		8	0.74	0.26
		10	0.74	0.26

Table 1 (cont.)

Relative Efficiencies of Some Unbiased Estimators of Rank Order

(For $g > \frac{1}{2}(n + 1)$, use $\text{eff}\,[T \mid n - g + 1] = \text{eff}[T \mid g]$)

n	r	g	$\text{eff}[\tilde{T} \mid g]$	$\text{eff}[\hat{T} \mid g]$
50	10	10	0.19	0.81
50	10	20	0.18	0.82
50	10	25	0.18	0.82
50	20	10	0.40	0.60
50	20	20	0.39	0.61
50	20	25	0.39	0.61
50	30	10	0.61	0.40
50	30	20	0.60	0.41
50	30	25	0.60	0.41
50	40	10	0.83	0.20
50	40	20	0.82	0.21
50	40	25	0.82	0.21
100	20	10	0.20	0.80
100	20	20	0.20	0.80
100	20	40	0.19	0.81
100	20	50	0.19	0.81
100	40	10	0.41	0.59
100	40	20	0.40	0.60
100	40	40	0.39	0.61
100	40	50	0.39	0.61
100	60	10	0.62	0.39
100	60	20	0.60	0.40
100	60	40	0.60	0.40
100	60	50	0.60	0.40
100	80	10	0.81	0.19
100	80	20	0.80	0.20
100	80	40	0.80	0.20
100	80	50	0.80	0.20

Note that if $\hat{\tilde{T}}$ were the linear function of $\tilde{T}$ and $\hat{T}$ which (a) had expected value g and (b), subject to (a), had minimum variance, then the sum of the two efficiencies would be equal to 1 . The biggest deviation from this situation is when n is small, r/n is near 1 and g/n is near 0 or 1 . (See cases $n = 10$, $g = 2$ in Table 1) .

It should be realized that each of the estimators $\tilde{T}$, $\hat{T}$ and $\hat{\tilde{T}}$ can take functional values. It is not intended that they should replace the integer valued $\tilde{G}$, $\hat{G}$ and $\hat{\tilde{G}}$; but they have been used to facilitate comparison.

4. Population Distribution Unknown: Series of Single Observations of the Same Unknown Rank in Random Samples of Known Size n

If we are given the observed value of a random variable and are told merely that it is selected at random from a set of n independent identically distributed continuous random variables, it is not possible to obtain a useful estimator of the rank of the chosen value among the n original values, without having some further information about the common distribution.

If we have a series of N such values $Y_1, Y_2, \ldots, Y_N$ and know that each takes the same rank (g) among a random sample of size n (but g is not known), we may in some circumstances be able to use the Y's to estimate $F(x)$ and so obtain some information on g. We will suppose the N random samples of size n to be selected independently.

(a) Form of Population Distribution Known

We suppose it is known that $F(x)$ is a known function $F(x \mid \theta_1, \ldots, \theta_s) = F(x \mid \underset{\sim}{\theta})$ of s parameters $\theta_1, \ldots, \theta_s$. The joint function of $Y_1, Y_2, \ldots, Y_N$ is then

$$(13) \quad L_g(\underset{\sim}{\theta}) = \left\{\frac{n!}{(g-1)!(n-g)!}\right\}^N \prod_{j=1}^{N} [\{F(Y_j \mid \underset{\sim}{\theta})\}^{g-1}\{1 - F(Y_j \mid \underset{\sim}{\theta})\}^{n-g} f(Y_j \mid \underset{\sim}{\theta})]$$

where $f(t \mid \underset{\sim}{\theta}) = \frac{d}{dt}\{F(t \mid \underset{\sim}{\theta})\}$.

For a given set of values $\underset{\sim}{\theta}_o$ of the parameters $\underset{\sim}{\theta}$, $L_g(\underset{\sim}{\theta}_o)$ is maximized by taking $g = \hat{g}(\underset{\sim}{\theta}_o)$ where $\hat{g}(\underset{\sim}{\theta}_o)$ is the greatest integer for which

$$L_g(\underset{\sim}{\theta}) > L_{g-1}(\underset{\sim}{\theta})$$

i.e. for which

$$\left(\frac{n-g+1}{g-1}\right)^N \prod_{n=1}^{N} \left\{ \frac{F(Y_j | \underset{\sim}{\theta})}{1-F(Y_j | \underset{\sim}{\theta})} \right\} > 1 .$$

We have

$$(14) \qquad \hat{g}(\underset{\sim}{\theta}_o) = [n\{1 + Q(\underset{\sim}{Y} \mid \underset{\sim}{\theta}_o)\}^{-1} + 1]$$

where

$$Q(\underset{\sim}{Y} \mid \underset{\sim}{\theta}_o) = \left\{ \prod_{j=1}^{N} \frac{F(Y_j | \underset{\sim}{\theta}_o)}{1-F(Y_j | \underset{\sim}{\theta}_o)} \right\}^{1/N} .$$

We then have to seek $\underset{\sim}{\theta}_o$ to maximize $L_{\hat{g}(\underset{\sim}{\theta}_o)}(\underset{\sim}{\theta}_o)$. The corresponding value, $\hat{g}$, of $\hat{g}(\underset{\sim}{\theta}_o)$ is the maximum likelihood estimator of g . The process can be lengthy, but considerable curtailment can be effected by noting that $\hat{g}$ must be an integer, and as $\max L_{\hat{g}(\underset{\sim}{\theta}_o)}(\underset{\sim}{\theta}_o)$ is approached the value of $\hat{g}(\underset{\sim}{\theta}_o)$ does not change over many iterations.

Another method, which is especially useful for small n , is to find the value $\hat{\underset{\sim}{\theta}}_{(g)}$, maximizing $L_g(\underset{\sim}{\theta})$ with respect to $\underset{\sim}{\theta}$, for fixed integer values of g and then choose $\hat{g}$ as the value minimizing $L_g(\hat{\underset{\sim}{\theta}}_{(g)})$. This procedure requires determination of just n values $L_1(\hat{\underset{\sim}{\theta}}_{(1)}) \ldots L_n(\hat{\underset{\sim}{\theta}}_{(n)})$, but the calculations needed to determine each $\hat{\underset{\sim}{\theta}}_{(g)}$ may be heavy. For this reason the procedure based on (14) may be preferable for larger values of n .

As a simple example, let us take

$$(15) \qquad F(x \mid \theta) = 1 - e^{-x/\theta} \quad (x > 0 ; \theta > 0)$$

(corresponding to an exponential distribution for X).

For g fixed,

$$(16)\qquad L_g(\theta) = K_g \prod_{j=1}^{N} \left\{\left(1 - e^{-Y_j/\theta}\right)^{g-1} e^{-(n-g)Y_j/\theta} \theta^{-1} e^{-Y_j/\theta}\right\}$$

$$= K_g \theta^{-N} \exp\left\{-(n - g + 1) \sum_{j=1}^{N} Y_j\right\} \circ \prod_{j=1}^{N} \left(1 - e^{-Y_j/\theta}\right)^{g-1}$$

where $K_g = \left\{\frac{(n-1)!}{(g-1)!(n-g)!}\right\}^N$.

To minimize $L_g(\theta)$ with respect to θ we take θ equal to the root of the equation

$$\frac{2L_g(\theta)}{2\theta} = -\frac{N}{\theta} + \frac{(n-r+1)\sum_{j=1}^{N} Y_j}{\theta^2} - \frac{(g-1)}{\theta^2} \sum_{j=1}^{N} \frac{Y_j e^{-Y_j/\theta}}{1-e^{-Y_j/\theta}} = 0$$

i.e. $(n - g + 1)N^{-1} \sum_{j=1}^{N} \left\{1 - (g - 1)(n - g + 1)^{-1} (e^{Y_j/\theta} - 1)^{-1}\right\}Y_j = \theta$

or

$$(17)\qquad \overline{Y} = \frac{\theta+(g-1)N^{-1} \sum_{j=1}^{N} Y_j (e^{Y_j/\theta} -1)^{-1}}{n-g+1} .$$

Since the right hand side of (17) is a continuous increasing function θ for $0 < \theta$, and tends to $0, \infty$ as $\theta \to 0, \infty$ respectively, the equation has a single solution $\theta = \hat{\theta}_{(g)}$.

In the present case, since

$$(18)\qquad E[Y \mid g] = \theta \sum_{j=0}^{g-1} (-1)^j \binom{g-1}{j} (n - g + 1 + j)^{-2} = \theta a_{g,n} ,$$

say, an initial value of θ could be taken as $\overline{Y}/a_{g,n}$.

(b) <u>Distribution Form Known to Depend only on Scale and Location Parameters</u>

In the special case when $x = 2$ and

$$F(x) = F(x \mid \theta_1, \theta_2) = H\big((x - \theta_1) / \theta_2\big)$$

where $H(\circ)$ is an explicit known function, the following heuristic method of estimating θ_1 and θ_2 which is relatively simple to apply, may give useful results. It is based on the fact that the expected value of the g-th smallest value in a random sample of size n (the g-th order statistic) is

$$\theta_1 + a_{g,n}\,\theta_2 \tag{19.1}$$

where $a_{g,n}$ is a known constant. Similarly the standard deviation of the g-th order statistic is

$$b_{g,n}\,\theta_2 \tag{19.2}$$

where $b_{g,n}$ is a constant.

For a given g, we estimate θ_2 by

$$\tilde{\theta}_2 = S_Y / b_{g,n}$$

where S_Y is the sample standard deviation of $Y_1, Y_2, \ldots, Y_N$; and we estimate θ_1 by

$$\tilde{\theta}_1 = \overline{Y} - a_{g,n}\,\tilde{\theta}_2 \tag{20.2}$$

with $\overline{Y} = N^{-1} \sum_{j=1}^{N} Y_j$.

We then insert these values in the likelihood function giving

$$(21)\quad \tilde{L}_g = L_g(\tilde{\theta}_1,\tilde{\theta}_2) = \tilde{\theta}^{-N}\left\{\frac{(n-1)!}{(g-1)!(n-g)!}\right\}^N \prod_{j=1}^{N}\left[\left\{H\left(\frac{Y_j-\tilde{\theta}_1}{\tilde{\theta}_2}\right)\right\}^{g-1}\left\{1-H_j\left(\frac{Y_j-\tilde{\theta}_1}{\tilde{\theta}_2}\right)\right\}\right] h\left(\frac{Y_j-\tilde{\theta}_1}{\tilde{\theta}_2}\right)$$

where $h(\circ) = H'(\circ)$.

This is repeated for $g = 1, 2, \ldots , n$ and we find the value of g, $\tilde{g}$ say, which maximizes $\tilde{L}_g$.

As a final step we may calculate the values of θ_1 and θ_2 which maximize L , for $g = \tilde{g} - 1$, $\tilde{g}$ and $\tilde{g} + 1$ and take our final estimator as that value of the three for which $L_g\left(\hat{\theta}_{1(g)}, \hat{\theta}_{2(g)}\right)$ is greatest.

In some cases, experience may suggest that this last step might be omitted. Conversely, in situations where we find that $L_g\left(\hat{\theta}_{1(g)}, \hat{\theta}_{2(g)}\right)$ is greater for $g = \tilde{g} + 1$ or $g = \tilde{g} - 1$ than it is for $g = \tilde{g}$, it may be desirable to try also $g = \tilde{g} + 2$ or $g = \tilde{g} - 2$ respectively.

For the case of an exponential distribution we have only one unknown parameter θ , and this would be estimated as $\tilde{\theta} = \bar{Y} / a_{g,n}$ (see text following (18)).

5. Further Possibilities

In the circumstances of the last section, it is possible, in principle, to estimate n , the total sample size (if it is unknown) as well as g , given a sufficient number of values $Y_1, Y_2, \ldots , Y_N$ each of which is the g-th in rank among N random values from the same continuous distributions. Either maximum likelihood or moments can be applied.

For example in the exponential case we can estimate θ , g and n as solutions of the moment equations.

$$(22) \quad m'_r(Y) = \theta^r \Gamma(r+1) \sum_{j=0}^{g-1} (-1)^j \binom{g-1}{j} (n-g+1+j)^{-(r+1)}$$

$$(r = 1, 2, 3)$$

where $m'_r(Y)$ is the r-th sample moment about zero of the N observed values $Y_1, Y_2, \ldots, Y_N$. Alternately we may seek values of n, g and θ maximizing (16) .

Apart from technical difficulties, however, it seems likely that very large values of N would be needed to obtain reasonably accurate estimators of the three parameters n, g and θ .