SOME ESTIMATES AND TESTS BASED ON THE r SMALLEST VALUES IN A SAMPLE

By John E. Walsh[1]

The Rand Corporation

1. Summary. Let us consider a situation where only the r smallest values of a sample of size n are available. This paper investigates the case where n is large and r is of the form $pn + O(\sqrt{n})$.

Properties of some well known non-parametric point estimates, confidence intervals and significance tests for the $100p\%$ point of the population are investigated. If the sample is from a normal population, these non-parametric estimates and tests have high efficiencies for small values of p (at least 95% if $p \leqq 1/10$).

The other results of the paper are restricted to the special case of a normal population. Asymptotically "best" estimates and tests for the population percentage points are derived for the case in which the population standard deviation is known. For the case in which the population standard deviation is unknown, asymptotically most efficient estimates and tests can be obtained for the smaller population percentage points by suitable choice of p and $O(\sqrt{n})$.

The results derived have application in the field of life testing. There the variable associated with an item is the time to failure and the r smallest sample values can be obtained without the necessity of obtaining the remaining values of the sample. By starting with a larger number of units but stopping the experiment when only a small percentage of the units have "died", it is often possible (using the results of this paper) to obtain the same amount of "information" with a substantial saving in cost and time over that which would be required if a smaller number of units were used and the experiment conducted until all the units have "died". Jacobson called attention to applications of this type in [1].

2. Introduction and statement of results. In life testing, information concerning the smaller population percentage points may be of primary interest. The principal aim of this paper is to investigate the properties of some well known non-parametric estimates and tests of the smaller population percentage points which are based on statistics of the type used for the sign test. These non-parametric results are easy to apply and have several other desirable properties (see Theorem 1 and its discussion). In particular, if the $100p\%$ point is to be investigated, it is only necessary to fail approximately $100p\%$ of the number of starting items to obtain the required statistics (n large). Thus, if the non-parametric results should also happen to be reasonably efficient, they

[1] The author would like to express his appreciation to Max Halperin for calling attention to this problem and for valuable advice and assistance in the preparation of the paper.

would appear to be ideal for a life testing situation where a smaller population percentage point is to be investigated.

Examination shows that life tests of the "wear out" type sometimes yield empirical distributions which are approximately normal. Also in many cases an approximately normal distribution can be obtained by an appropriate monotonic change of variable. Thus the case in which the n observations are a sample from a normal population will receive special consideration in this paper.

Investigation of the efficiency of the non-parametric estimates and tests will be limited to the situation where the n observations are a sample from a normal population. Three cases will be considered:

(A). Asymptotic efficiency of the non-parametric results as compared with the corresponding most efficient results based on the entire sample (population variance unknown).

(B). Asymptotic efficiency of the non-parametric results as compared with the corresponding most efficient results based on the $pn + O(\sqrt{n})$ smallest order statistics for the situation where the variance of the normal population is known.

(C). Asymptotic efficiency of the non-parametric results as compared with the corresponding most efficient results based on the $\beta n + O(\sqrt{n})$ smallest order statistics where β is slightly greater than p (population variance unknown).

The definition of "asymptotic" efficiency together with some of its properties is given in Section 3. Only asymptotic efficiencies will be considered.[2] However, the efficiencies obtained for the asymptotic case would seem to represent lower bounds of the efficiencies for the corresponding non-asymptotic cases since experience indicates that the efficiency of non-parametric results usually decreases as the sample size increases.

First let us consider case (A). From Theorem 3, the asymptotically most efficient results for estimating or testing the $100p\%$ population point on the basis of the entire sample (population variance unknown) are furnished by the non-central t-statistic. An expression for the asymptotic efficiency of the non-parametric results as compared with the corresponding results based on the non-central t-statistic is given in the Corollary to Theorem 3. The reciprocal of this efficiency represents the factor by which the original number of starting items must be multiplied if the non-parametric results are to asymptotically furnish the same "information" as the non-central t-statistic applied to the original number of starting items. Table 1 contains values of this factor. Although a larger number of starting items are used by the "information equivalent" non-parametric results, a noticeably smaller number of items are failed. The factor by which the number of items failed is decreased equals the value of p multiplied by the factor by which the number of starting items was increased for the "equiv-

[2] Some power function comparisons for the non-asymptotic case were given by Paul H. Jacobson in [1].

alent" non-parametric result. Table 2 contains a list of some of the resulting factors.

Next consider case (B). The first step in the analysis for this case consists in obtaining the asymptotically most efficient results. These derivations are contained in Theorems 4 and 5. The Corollary to Theorem 5 contains an expression for the asymptotic efficiency of the non-parametric results for case (B). The factor by which the original number of starting items must be multiplied to obtain "information equivalent" non-parametric results is obtained in the same way as for case (A). Table 1 lists values of this factor. In this case both the number of starting items and the number of items failed are slightly increased by use of the "equivalent" non-parametric results. The factor by which the number of items failed is increased equals the corresponding factor for the increase in number of starting items. For convenience of reference, however, values

TABLE 1

Asymptotic ratio of total numbers of items tested
(Non-parametric test over most efficient test)

Case \ p	01	.02	.05	.10	.20	.30	.40	.50	.70
(A)	377%	270%	190%	160%	150%	153%	155%	157%	
(B)	101%	102%	103%	105%	109%	114%	120%	128%	164%
(C)	111%	114%	118%	122%	129%	140%	148%		

of this factor are also given in Table 2. If the variance of the normal population were unknown, the asymptotic efficiency of the non-parametric results would be at least as great as that obtained for case (B), and likely greater.

Finally consider case (C). Let p be replaced by β in Theorem 5 while the value of β corresponding to a given value of p is defined by the relation in Theorem 6. By suitable choices for the values of β and $O(\sqrt{n})$ in Theorem 5, it is possible to obtain asymptotically most efficient results for the population $100p\%$ point when the population variance is unknown and only the $\beta n + O(\sqrt{n})$ smallest values of the sample are available. These results are presented in Theorem 6. The Corollary to Theorem 6 contains an expression for the asymptotic efficiency of the non-parametric results as compared with the corresponding results of Theorem 6. The factor by which the number of starting items must be increased to obtain "equivalent" non-parametric results is computed as in cases (A) and (B). Table 1 contains values of this factor. The value of β represents the fraction of starting items which are failed if the estimates and tests of Theorem 6 are used. Table 2 contains corresponding values of β for certain values of p. The factor by which the number of items failed is decreased equals p/β times the factor by which the number of starting items was increased to obtain the "equivalent" non-parametric results. Table 2 presents values of this factor.

The results of Theorem 6 furnish an asymptotically efficient method of estimating and testing the smaller population percentage points while only failing a small percentage of the starting items (for the case of normality). Since a larger number of items are failed and much more work is required for computing the necessary statistics, however, this method is not necessarily preferable to the non-parametric method from the viewpoint of "information" per unit cost. In many cases the difference in cost will be slight. Since the non-parametric results are valid under much more general conditions, they would seem to be preferable for these cases.

TABLE 2

Asymptotic ratio of numbers of items failed
(Non-parametric test over most efficient test)

Case \ β	.0113	.0234	.0612	.130	.287	.476	.70	
Case \ p	.01	.02	.05	.10	.20	.30	.40	.50
(A)	3.77%	5.40%	9.50%	16.0%	30.2%	45.9%	62.0%	78.5%
(B)	101%	102%	103%	105%	109%	114%	120%	128%
(C)	99%	98%	96%	94%	90%	88%	85%	

3. Definition of asymptotic efficiency. In this section the n observations are assumed to be a sample from a normal population. Let the $100p\%$ point of the population be denoted by θ_p. Several classes of results for investigating θ_p are considered in this paper. For example, the non-parametric estimates and tests represent one class; the asymptotically most efficient results based on the entire sample (population variance unknown) represent another class; etc. The results considered consist of point estimates of θ_p, confidence intervals for θ_p, and significance tests for θ_p based on these confidence intervals. For a specified class, every point estimate and every endpoint of a confidence interval (a one-sided confidence interval has only one endpoint) consists of some statistic T whose variance is of the form $\sigma_T^2/n + o(1/n)$ for large n. Here σ_T^2 is independent of n and has the same value for all statistics T of the class. Also for every such statistic T the quantity

$$\sqrt{n}(T - \theta_p)/\sigma_T$$

has a distribution which is asymptotically normal with unit variance and some finite mean A which is independent of the unknown parameters of the normal population. By suitable choice of T, the mean A can be made to have any specified value.

Now let us define the asymptotic efficiency of the class of non-parametric results as compared to a class of results of the type defined by (A), (B) or (C). Let the non-parametric results be based on n sample values while the other class of results is based on m sample values. Let the common value of σ_T^2 for the non-parametric results be denoted by σ_1^2 while the common value of this quantity for the other class is denoted by σ_2^2. If $\sigma_1^2/n = \sigma_2^2/m$ when $m = nE$, then the asymptotic efficiency of the non-parametric results (compared to the specified class of results) is defined to be $100E\%$. For the situations considered in this paper, E is independent of n, m and the parameters of the normal population.

Asymptotic efficiency, as defined in the preceding paragraph, has the property that the statistic (or statistics) yielded by a non-parametric result based on n sample values has approximately the same distribution as the corresponding statistic (or statistics) based on m sample values from the specified class if $m = nE$ (n large). For example, consider a non-parametric unbiased estimate T_1 of θ_p based on n sample values and an unbiased estimate T_2 of θ_p from the specified class based on m sample values. Then, if $m = nE$, the distributions of

$$\sqrt{n}(T_1 - \theta_p)/\sigma_1, \qquad \sqrt{n}(T_2 - \theta_p)/\sigma_1$$

are asymptotically identical (note that $\sigma_1^2/n = \sigma_2^2/m$). Similarly for the end-points of confidence intervals. Consequently the power functions of significance tests based on corresponding confidence intervals are asymptotically identical if $m = nE$. It would therefore appear that the definition chosen for asymptotic efficiency is suitable for the situations to which it is applied.

4. Notation. In this paper $t(1), \cdots, t(n)$ will represent the values of the set of all n observations arranged in increasing order of magnitude. Then

$$t(1), \cdots, t(r)$$

are the r smallest values of the set of n observations. The notation $t(r)$ has meaning only if r is an integer such that $1 \leqq r \leqq n$. Often, however, expressions of the form $t[pn + O(\sqrt{n})]$ will be encountered. In what follows, an expression of the form $t(z)$ has the interpretation t (largest integer $\leqq z$). For example,

$$t(487\tfrac{1}{2}) = t(487).$$

Also the $r = pn + O(\sqrt{n})$ smallest observations are frequently referred to; here r is interpreted to be the largest integer contained in $pn + O(\sqrt{n})$; etc.

5. Theorems and derivations. First let us consider some well known estimates and tests of the population percentage points which are based on statistics of the type used for the sign test. These estimates and tests are valid under extremely general conditions. It is not necessary that the observations be drawn from the same population or even that any two observations come from the same population. Population percentage points are not necessarily unique. The strongest continuity restriction imposed is that the population *cdf* be continuous at the percentage point considered. These results follow from

THEOREM 1. *Let $t(1), \cdots, t(n)$ represent the values of n observations arranged in increasing order of magnitude. The n observations are statistically independent and from populations which satisfy the conditions:*

(I). *The populations have at least one $100p\%$ point in common.*

(II). *If the populations have only one common $100p\%$ point, the cdf of each population is continuous at that point.*

Let θ_p denote the value of the common $100p\%$ point if it is unique, or the open interval of common $100p\%$ points otherwise (i.e., the interval of common $100p\%$ points with its endpoints deleted). Then asymptotically ($n \to \infty$)

(i). *$t(pn)$ is a median estimate of θ_p.*

(ii). $\Pr\{t[pn + K_\alpha\sqrt{np(1-p)}] < \theta_p\} = \Pr\{t[pn + K_\alpha\sqrt{np(1-p)}] \leqq \theta_p\} = \alpha,$

where K_α is the standardized normal deviate exceeded with probability α. Relations (i) *and* (ii) *are approximately satisfied if $pn > 5$ and $p \leqq \frac{1}{2}$.*

PROOF. This theorem is a direct application of the binomial theorem. Conditions (I) and (II) assure that the equality between the probabilities in (ii) holds. Relations (i) and (ii) are obtained by using the normal approximation to the binomial theorem; this approximation is reasonably accurate if $pn > 5$ and $p \leqq \frac{1}{2}$ (see [2]).

The non-parametric confidence intervals investigated are of the forms

$$t[pn + B_1\sqrt{n} + o(\sqrt{n})] < \theta_p, \qquad t[pn + B_2\sqrt{n} + o(\sqrt{n})] > \theta_p,$$
$$t[pn + B_1\sqrt{n} + o(\sqrt{n})] < \theta_p < t[pn + B_2\sqrt{n} + o(\sqrt{n})] \qquad (B_1 < B_2),$$

(these intervals have the same confidence coefficient if $<$ is replaced by $\leqq$ and $>$ by $\geqq$). The significance tests considered are those obtained from these confidence intervals while the point estimates of θ_p are based on single order statistics of the form $t[pn + B\sqrt{n} + o(\sqrt{n})]$.

When θ_p is an open interval, (i) and (ii) need interpretation. The meaning of (i) is that the probability of $t(pn)$ exceeding every value of θ_p has the value $\frac{1}{2}$ and that the probability of it being less than all values of θ_p also has the value $\frac{1}{2}$. The inequality $t[pn + K_\alpha\sqrt{np(1-p)}] \leqq \theta_p$ has the interpretaton that every value of θ_p is greater than or equal to $t[pn + K_\alpha\sqrt{np(1-p)}]$. Similarly for $t[pn + K_\alpha\sqrt{np(1-p)}] < \theta_p$.

The purpose in introducing the case where θ_p is an open interval was to point out that situations where population percentage points are not unique cause little difficulty if suitably interpreted.

Non-parametric results of the type considered in Theorem 1 are also available when the sample size is not large. For any sample size n, if the conditions of Theorem 1 are satisfied,

$$\Pr[t(r) < \theta_p] = \Pr[t(r) \leqq \theta_p] = \sum_{s=r}^{n} \frac{n!}{s!(n-s)!} p^s(1-p)^{n-s}.$$

The probability relations in Theorem 1 were obtained by approximating this summation for large n. By suitable choice of r, confidence intervals and significance tests with a wide range of satisfactory confidence coefficients and significance levels can usually be obtained for a given value of n.

The above discussion emphasizes the generality of application of the non-parametric estimates and tests. For most practical situations, however, it is permissible to assume that the observations are a random sample from a population which has a probability density function that is non-zero over the range of definition and differentiable several times. Then asymptotically $t(pn)$ is also a mean estimate of θ_p (which is now necessarily a single point). Moreover, the asymptotic distribution of $t[pn + C\sqrt{n} + o(\sqrt{n})]$ can be found in terms of p, C, θ_p and the value of the probability density function at θ_p. These results are a consequence of

THEOREM 2. *Let the population from which the n sample values were drawn have a pdf $f(t)$ such that $f(t) \neq 0$ over its range of definition and $f'(t)$ exists and is continuous in some neighborhood of $t = \theta_p$. Then the variable*

$$\sqrt{n/p(1-p)}f(\theta_p)\{t[pn + C\sqrt{n} + o(\sqrt{n})] - \theta_p\}$$

has a distribution which approaches the normal distribution with mean

$$C/\sqrt{p(1-p)}$$

and unit variance as $n \to \infty$.

PROOF. If pn is replaced by $pn + C\sqrt{n} + o(\sqrt{n})$, the method used to prove this theorem is completely analogous to the proof presented on pp. 368–69 of [3].

Now let us consider the asymptotically most efficient results for estimating and testing θ_p based on the entire set of observations for the case of a sample from a normal population (population variance unknown).

THEOREM 3. *Let the n observations be a sample from a normal population (unknown variance σ^2). Asymptotically the most efficient point estimates, confidence intervals and significance tests for θ_p using all the observations are those based on the non-central t-statistic. The value of σ_T^2 (see Section 3) for these results based on the non-central t-statistic is $\sigma^2(1 + K_p^2/2)$.*

COROLLARY. *For case (A) the asymptotic efficiency of the non-parametric results equals*

$$100(1 + K_p^2/2)/2\pi p(1-p) \exp(K_p^2) \%.$$

PROOF. The maximum likelihood estimate of θ_p based on all n sample values is

$$\frac{1}{n}\sum_1^n t(i) - K_p\sqrt{\sum_1^n\left[t(i) - \frac{1}{n}\sum_1^n t(j)\right]^2 \Big/ (n-1)}. \tag{1}$$

This quantity is equivalent to the non-central t-statistic, as can be seen by multiplying and dividing $[(1) - \theta_p]$ by

$$\sqrt{\sum_1^n\left[t(i) - \frac{1}{n}\sum_1^n t(j)\right]^2 \Big/ (n-1)}.$$

From maximum likelihood theory, (1) is an efficient estimate of θ_p. Asymptotically ($n \to \infty$) the variance of (1) is of the form

$$\sigma^2(1 + K_p^2/2)/n + o(1/n),$$

and it is easily seen that the variance of an endpoint of a confidence interval for θ_p based on the non-central t-statistic is also of this form. The corollary follows from combining Theorem 2 with Theorem 3.

Next let us investigate the situation where only the $r = pn + O(\sqrt{n})$ smallest values of a sample of size n from a normal population with mean μ and variance σ^2, denoted by $N(\mu, \sigma^2)$, are available. First let us consider the asymptotic distribution of

$$\left[\frac{\sum_1^r t(i) + 2a_p(n-r)t(r)}{r + 2a_p(n-r)} - \mu + \frac{(n-r)(b_p + 2a_pK_p)}{r + 2a_p(n-r)}\sigma\right] \Big/ \frac{\sigma}{\sqrt{r + 2a_p(n-r)}}, \tag{2}$$

where

$$a_p = K_p/2\sqrt{2\pi}\,(1-p)\exp\left(\frac{1}{2}K_p^2\right) + 1/4\pi(1-p)^2\exp(K_p^2),$$

$$b_p = 1/\sqrt{2\pi}(1-p)\exp\left(\frac{1}{2}K_p^2\right).$$

This distribution is given by

THEOREM 4. *Let $t(1), \cdots, t(r)$ be the $r = pn + O(\sqrt{n})$ smallest values (arranged in increasing order of magnitude) of a sample of size n from $N(\mu, \sigma^2)$. Then asymptotically ($n \to \infty$) the distribution of (2) is $N(0, 1)$.*

COROLLARY. *Let $r = pn + C\sqrt{n} + o(\sqrt{n})$. Then as n increases the distribution of*

$$\left[\frac{\sum_1^r t(i) + 2a_p(n-r)t(r)}{r + 2a_p(n-r)} - \mu + \frac{(1-p)(b_p + 2a_pK_p)}{p + 2a_p(1-p)}\sigma\right] \Big/ \frac{\sigma}{\sqrt{r + 2a_p(n-r)}}$$

approaches the normal distribution with unit variance and mean

$$C(b_p + 2a_pK_p)/[p + 2a_p(1-p)]^{3/2}.$$

PROOF. The proof of this theorem is long and will be deferred to section 6 of the paper.

If the value of σ is known, the Corollary to Theorem 4 can be used to obtain point estimates, confidence intervals and significance tests for any population percentage point (including μ). The resulting estimates and tests are asymptotically most efficient. This follows from

THEOREM 5. *Consider the $r = pn + O(\sqrt{n})$ smallest values of a sample of size* *n from $N(\mu, \sigma^2)$ where σ^2 is known. Asymptotically ($n \to \infty$) the variance of every unbiased estimate of μ based on only $t(1), \cdots, t(r)$ and σ^2 is greater than or equal to a quantity of the form*

$$\sigma^2/n[p + 2a_p(1 - p)] + o(1/n).$$

COROLLARY. *For case (B) the asymptotic efficiency of the non-parametric results is*

$$100\left[\frac{\exp(-K_p^2)}{2\pi p(1-p)} \bigg/ \left(p + \frac{K_p \exp\left(-\frac{1}{2}K_p^2\right)}{\sqrt{2\pi}} + \frac{\exp(-K_p^2)}{2\pi(1-p)}\right)\right]\%.$$

PROOF. The proof of this theorem is similar to the proof presented for Theorem 4 and will be given in section 6 following the proof of Theorem 4.

Let p be replaced by β in Theorem 4. Even if σ is unknown asymptotically most efficient estimates and tests can be obtained for the $100p\%$ point of the population if β is defined by

$$K_p = (1 - \beta)(b_\beta + 2a_\beta K_\beta)/[\beta + 2a_\beta(1 - \beta)]. \tag{3}$$

THEOREM 6. *Let p, $(0 < p < \frac{1}{2})$, be given and β defined by (3). Let $t(1), \cdots, t(r)$ be the $r = \beta n + C\sqrt{n} + o(\sqrt{n})$ smallest values of a sample of size n from a normal population. Then asymptotically*

$$\Pr\left\{\left[\sum_1^r t(i) + 2a_\beta(n - r)t(r)\right] \bigg/ [r + 2a_\beta(n - r)] < \theta_p\right\}$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-C(b_\beta + 2a_\beta K_\beta)/[\beta + 2a_\beta(1-\beta)]^{3/2}} e^{-x^2/2}\, dx.$$

COROLLARY. *For case (C) the asymptotic efficiency of the non-parametric results is*

$$100\left[\frac{\exp(-K_p^2)}{2\pi p(1-p)} \bigg/ \left(\beta + \frac{K_\beta \exp\left(-\frac{1}{2}K_\beta^2\right)}{\sqrt{2\pi}} + \frac{\exp(-K_\beta^2)}{2\pi(1-\beta)}\right)\right]\%.$$

PROOF. Theorem 6 is an immediate consequence of relation (3) and the Corollary to Theorem 4. The Corollary to Theorem 6 follows from Theorem 2 and Theorem 6.

6. Long proofs. This section contains the long proof of Theorem 4 and the related proof of Theorem 5.

6.1. Proof of Theorem 4. If $t(r)$ is such that

$$\mu - K_p\sigma - n^{-4/10} \leqq t(r) \leqq \mu - K_p\sigma + n^{-4/10},$$

the ratio of the value of the joint probability density function f of $t(1), \cdots, t(r)$ to the value of the function

$$\begin{aligned}\frac{n!(1-p)^{n-r}}{(n-r)!}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^r \exp\Bigg\{&-\frac{1}{2}\sum_1^r\left[\frac{t(i)-\mu}{\sigma}\right]^2 \\ &- (n-r)a\left[\frac{t(r)-\mu}{\sigma} + K_p\right]^2 - (n-r)b\left[\frac{t(r)-\mu}{\sigma} + K_p\right]\Bigg\}\end{aligned} \tag{4}$$

is of the form $1 + o(1)$. Here (and in the remainder of section 6) $a = a_p$, $b = b_p$. Also, for large n and any positive ϵ, the integral of f over the ranges of the $t(1), \cdots, t(r-1)$ and for $t(r)$ between $\mu - K_p\sigma - n^{-\frac{1}{2}+\epsilon}$ and $\mu - K_p\sigma + n^{-\frac{1}{2}+\epsilon}$ differs from unity by a quantity which is of the order $o(1)$, i.e., a quantity which $\rightarrow 0$ as $n \rightarrow \infty$.

Now consider the moment generating function of (2), i.e., $E[e^{\theta(2)}]$. In evaluating this function of θ, let the range of integration of $t(r)$, (i.e., the range after the other variables have been integrated out), be subdivided into the five intervals

$$-\infty \quad \text{to} \quad \mu - D\sigma, \qquad \mu - D\sigma \quad \text{to} \quad \mu - K_p\sigma - n^{-4/10},$$
$$\mu - K_p\sigma - n^{-4/10} \quad \text{to} \quad \mu - K_p\sigma + n^{-4/10},$$
$$\mu - K_p\sigma + n^{-4/10} \quad \text{to} \quad \mu + D\sigma, \qquad \mu + D\sigma \quad \text{to} \quad \infty.$$

Here D is a positive constant which is independent of n and such that

$$(1/D)^{n-r}(1/p)^{r-1}[1/(1-p)]^{n-r} < \exp\left[-\frac{|\theta|(n-r)(b+2aK_p)}{\sqrt{r+2a(n-r)}}\right]$$

for n sufficiently large and

$$D > |K_p| + n^{-4/10}/\sigma, \qquad 1 - N(D) = N(-D) < e^{-\frac{1}{2}D^2}/D,$$

where

$$N(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{1}{2}y^2}\,dy.$$

First let us consider the interval $\mu - K_p\sigma - n^{-4/10}$ to $\mu - K_p\sigma + n^{-4/10}$. Using (4) in place of f, completing the square in the exponent, making the change of variable

$$x(i) = t(i) - \theta/\sqrt{r + 2a(n-r)} \qquad (i = 1, \cdots, r),$$

integrating $x(1), \cdots, x(r-1)$ over their ranges and then $x(r)$ over the interval

$$\mu - K_p\sigma - n^{-4/10} - \theta/\sqrt{r+2a(n-r)} \text{ to}$$
$$\mu - K_p\sigma + n^{-4/10} - \theta/\sqrt{r+2a(n-r)},$$

an expression of the form

$$\exp(\theta^2/2) + o(1) \tag{5}$$

is obtained. From the above results, this expression differs from the corresponding integration of f by a term of order $o(1)$; hence the contribution to the mgf for the interval considered is of the form (5).

Next consider the interval $\mu - K_p\sigma + n^{-4/10}$ to $\mu + D\sigma$. After $t(1), \cdots, t(r-1)$ have been integrated out, the integrand becomes

$$\begin{aligned} &\frac{n!}{(r-1)!(n-r)!}\left\{N\left[\frac{t(r)-\mu}{\sigma} - \frac{\theta}{\sqrt{r+2a(n-r)}}\right]\right\}^{r-1} \\ &\cdot\left\{1 - N\left[\frac{t(r)-\mu}{\sigma}\right]\right\}^{n-r}\exp\left\{-\frac{1}{2}\left[\frac{t(r)-\mu}{\sigma}\right]^2\right. \\ &\left.+\frac{2\theta a(n-r)}{\sqrt{r+2a(n-r)}}\left[\frac{t(r)-\mu}{\sigma}+K_p\right] + \frac{b\theta(n-r)}{\sqrt{r+2a(n-r)}}\right\}. \end{aligned} \tag{6}$$

By writing $\{N[(t(r) - \mu)/\sigma - \theta/\sqrt{r + 2a(n - r)}]\}^{r-1}$ in the form

$$\left[N\left(\frac{t(r) - \mu}{\sigma}\right)\right]^{r-1} \left\{\left[1 - \theta(1 + o(1))/\sqrt{r + 2a(n - r)} \cdot N\left(\frac{t(r) - \mu}{\sigma}\right) \sqrt{2\pi} \exp\left(\frac{t(r) - \mu}{\sigma}\right)^2\right]^{\sqrt{n}}\right\}^{(r-1)/\sqrt{n}}$$

and maximizing $\exp\{2\theta a(n - r)[t(r) - \mu]/\sigma\sqrt{r + 2a(n - r)}\}$ with respect to $t(r)$ in the specified interval, it is seen that the value of (6) is less than an expression of the form

$$\frac{n! \exp(C_1\theta\sqrt{n})}{(r - 1)!(n - r)!} \left\{N\left[\frac{t(r) - \mu}{\sigma}\right]\right\}^{r-1} \left\{1 - N\left[\frac{t(r) - \mu}{\sigma}\right]\right\}^{n-r} + o(1)$$

for n sufficiently large. Differentiation shows that $\{N[\,]\}^{r-1}\{1 - N[\,]\}^{n-r}$ is a decreasing function of $t(r)$ in the specified interval if n is large enough. Also, if $t(r) = \mu - K_p\sigma + n^{-4/10}$, for large n the value of

$$\left\{N\left[\frac{t(r) - \mu}{\sigma}\right]\right\}^{(r-1)n^{-6/10}} \cdot \left\{1 - N\left[\frac{t(r) - \mu}{\sigma}\right]\right\}^{(n-r)n^{-6/10}} \Big/ p^{(r-1)n^{-6/10}}(1 - p)^{(n-r)^{-6/10}}$$

is less than a constant which is less than unity. Thus the value of (6) is less than a quantity of the form

$$\frac{n!\, p^{r-1}(1 - p)^{n-r}}{(r - 1)!(n - r)!} \exp(-C_3^2 n^{1/10}) + o(1),$$

which in turn is less than an expression of the form

$$C_4\sqrt{n} \exp(-C_3^2 n^{1/10}) + o(1)$$

for n sufficiently large. Thus the integral of (6) over the specified interval is of the order $o(1)$. An analogous proof shows that the contribution to the mgf for the interval $\mu - D\sigma$ to $\mu - K_p\sigma - n^{-4/10}$ is also of order $o(1)$.

Finally consider the interval $\mu + D\sigma$ to ∞. For large n the integral of (6) over this interval is less than an expression of the form

$$\text{(7)} \qquad \frac{n!\, p^{r-1}(1 - p)^{n-r}}{(r - 1)!(n - r)!} \int_{\mu+D\sigma}^{\infty} \exp\left\{-\tfrac{1}{2}(n - r)\left[\frac{t(r) - \mu}{\sigma}\right]^2\right\} dt(r) + o(1);$$

i.e., the contribution to the mgf for this interval is of the order $o(1)$ since the coefficient of the integral is less than an expression of the form $C\sqrt{n}$. The upper limit (7) was obtained by replacing

$N\{[t(r) - \mu]/\sigma - \theta/\sqrt{r + 2a(n - r)}\}$ by 1,

$$1 - N\left[\frac{t(r) - \mu}{\sigma}\right] \text{ by } \frac{1}{D} \exp\left\{-\frac{1}{2}\left[\frac{t(r) - \mu}{\sigma}\right]^2\right\},$$

$$(1/D)^{n-r}\ (1/p)^{r-1}\,[1/(1-p)]^{n-r}\exp\left[-\theta(n-r)\ (b+2aK_p)/\sqrt{r+2a(n-r)}\right]$$

by 1.

A similar type proof shows that the integral of (6) from $-\infty$ to $\mu - D\sigma$ is also of the order $o(1)$.

Thus the mgf of (2) is of the form (5) for large n and Theorem 4 is verified.

6.2. Proof of Theorem 5. Let us consider a single sample value from the multivariate population consisting of the r smallest order statistics of a sample of size n from $N(\mu, \sigma^2)$, where σ^2 is known. Then the variance of every unbiased estimate of μ based on this sample and the value of σ^2 is greater than or equal to the reciprocal of

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{t(2)}\left(\frac{\partial \log f}{\partial\mu}\right)^2 f dt(1)\cdots dt(r) \tag{8}$$
$$= -\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{t(2)}\frac{\partial^2 \log f}{\partial\mu^2} f dt(1)\cdots dt(r),$$

where f is the joint *pdf* of the r smallest order statistics of a sample of size n from $N(\mu, \sigma^2)$. For proof of this statement see pp. 480–81 of [3]. In the lower part of (8) the variables $t(1), \cdots, t(r-1)$ can be integrated out leaving an explicit function of $t(r)$ to be integrated from $-\infty$ to ∞. To evaluate this integral for large n, choose some large but fixed interval $\mu - D\sigma$ to $\mu + D\sigma$ as was done in the proof of Theorem 4. Using a method similar to that presented on pp. 368–69 of [3], the value of the integral for the interval $\mu - D\sigma$ to $\mu + D\sigma$ is found to be of the form

$$n[p + 2a(1-p)]/\sigma^2 + o(n).$$

A procedure analogous to that used in the latter part of section 6.1 shows that integration outside this interval yields an expression of order $o(n)$.

REFERENCES

[1] Paul H. Jacobson, "The relative power of three statistics," *Jour. Amer. Stat. Assn.*, Vol. 42 (1947), pp. 575–584.

[2] Paul G. Hoel, *Introduction to Mathematical Statistics*, John Wiley and Sons, 1947, p. 45.

[3] Harald Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press. 1946.