AD641130

DEPARTMENT OF STATISTICS
HARVARD UNIVERSITY

UPPER AND LOWER PROBABILITY INFERENCES BASED ON A SAMPLE FROM A FINITE UNIVARIATE POPULATION

A. P. Dempster

Technical Report No. 14

October 10, 1966

Contract Nonr 1866(37), NR-042-097

This paper is the result of research supported at Harvard University by the Office of Naval Research.
Reproduction in whole or in part is permitted by the U. S. Government. Distribution of this document is unlimited.

CLEARINGHOUSE
FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION

Hardcopy	Microfiche		
$2.00	$0.50	43 pp	ma

1 ARCHIVE COPY

DDC
RECEIVED
NOV 2 1966
C

0. Summary. Following the general approach of Dempster (1966a), this paper provides detailed formulas for upper and lower probability inferences, based on a sample from an unknown finite population with a discrete univariate characteristic, and directed towards parameters of the population or properties of a future sample. The model is set up in Section 2 and the detailed formulas are given in Section 3. Section 4 demonstrates a limited relationship with confidence statements and Section 5 explores some special cases.

1. Introduction. A general framework for reasoning to upper and lower probability inferences was proposed by Dempster (1966a). Two specific ways of implementing the framework were suggested, and the resulting models were called structures of the first and second kinds. The first of these is the simpler of the two, but also the more restrictive since it handles only a univariate characteristic. The infinite population theory of Dempster (1966a) has a natural extension to finite population theory in the case of structures of the first kind, and the purpose of this paper is to derive general formulas for certain of the consequent upper and lower probability inferences.

The particular mathematical theory of upper and lower probabilities used here has been developed in terms unrelated to statistical inference by Dempster (1966b). The basic idea is that a probability measure μ over a sample space X is diffused by a multivalued mapping from X to S so that only bounds may be put on the probabilities of subsets of S. Where appropriate, the notation of Dempster (1966b) will be adopted in the application of this paper.

The ingredients of the situation under study are a finite population of N individuals and two samples of sizes n and m comprising n + m distinct individuals from the population. At the time of inference, only the first sample of size n is supposed to have been observed. Based on these observations, inferences are to be drawn concerning the potentially observable characteristics of the second sample and the population.

To observe an individual means here to determine the category to which the individual belongs among a set of k ordered mutually exclusive categories. The index j running over $j = 1,2,\ldots,k$ will be used to label these categories in their given order. The ordering means that the observable characteristic is essentially univariate, as is required for the specific model structures of the first kind of Dempster (1966a).

The observable information in the population may be characterized by a vector

(1.1) $$\underset{\sim}{N} = (N_1, N_2, \ldots, N_k)$$

where N_j denotes the number of population individuals in category j for $j = 1,2,\ldots,k$. Similarly, the observable information in the first and second samples may be represented by

(1.2) $$\underset{\sim}{n} = (n_1, n_2, \ldots, n_k)$$

and

(1.3) $$\underset{\sim}{m} = (m_1, m_2, \ldots, m_k)$$

where n_j and m_j are the numbers in the first and second samples, respectively, which belongs in category j, for $j = 1,2,\ldots,k$. One could in practice observe the order of sample observations as well as their gross totals, but the inferences to be described come out the same whether the observations are ordered or not, so the grosser but simpler representation (1.2) is adopted. $\underset{\sim}{N}$, $\underset{\sim}{n}$ and $\underset{\sim}{m}$

are allowed to be any vectors consistent with the logic of the situation, i.e., their elements are non-negative integers satisfying

$$n_j + m_j \leq N_j \tag{1.4}$$

for $j = 1,2,\ldots,k$,

$$\sum_{j=1}^{k} N_j = N, \tag{1.5}$$

$$\sum_{j=1}^{k} n_j = n, \tag{1.6}$$

and

$$\sum_{j=1}^{k} m_j = m. \tag{1.7}$$

It is assumed that n, m and N are known from the start.

The probability model set up in Section 2 leads to upper and lower probabilities for any event defined by $\underset{\sim}{m}$ and $\underset{\sim}{N}$ given any data $\underset{\sim}{n}$. Specific formulas are derived in Section 3 for $P^*(T)$ and $P_*(T)$ where

$$T = \left\{ r \leq \sum_{j=a}^{b} m_j \leq t \right\} \tag{1.8}$$

for $0 \leq r \leq t \leq m$ and $1 \leq a \leq b \leq k$.

Having these upper and lower probabilities, one can immediately deduce similar results for the population, since when $m = N - n$ the second sample is the whole of the remainder of the population.

2. The elements of the probability model

The basic elements required for the framework of Dempster (1966b) are a probability measure space X with given measure μ and a multivalued mapping from X to S, where S is the set of outcomes about which probability judgements are desired.

X is taken here to be the set of possible pairs of samples ignoring order within samples. If the population members are labelled by an index I running over $I = 1,2,\ldots,N$, then the first sample is determined by a vector

$$J = (J_1, J_2, \ldots, J_n) \tag{2.1}$$

whose elements are the values of the index I for the individuals in the first sample, and where for uniqueness of the representation one may choose $J_1 < J_2 < \ldots < J_n$. Similarly, the second sample may be represented by a vector

$$K = (K_1, K_2, \ldots, K_m) \tag{2.2}$$

whose elements are the ordered I-index values of the members of the second sample. Finally, a typical member of X is represented by the pair $[J, K]$. Recall that the samples do not overlap, so that J and K together contain $n + m$ distinct integers from $1,2,\ldots,N$.

The probability measure μ is taken to be the familar random sampling law which assigns equal probabilities to each of the $N!/[n!\,m!\,(N-n-m)!]$ elements of X. This simple measure is the only probability measure assumed in this paper, and all of the inferences are consequences of it.

Although the chief interest lies in inferences about $\underset{\sim}{m}$ and $\underset{\sim}{N}$ after $\underset{\sim}{n}$ is fixed by observation, it serves to illustrate the logic of the method to consider the appropriate judgments <u>before</u> $\underset{\sim}{n}$ is fixed as well as after. Consequently, a general member of the space S will be represented by the triple $[\underset{\sim}{n},\underset{\sim}{m},\underset{\sim}{N}]$. Observation of $\underset{\sim}{n}$ requires that the probability judgments established before $\underset{\sim}{n}$ was observed must be appropriately conditioned, in line with the definition of upper and lower conditional probabilities of Dempster (1966b)

A fundamental assumption will now be made which will carry in its train the multivalued mapping from X to S which is a basic feature of the model. As with the structures of the first kind of Dempster (1966a), it is assumed here that the population individuals possess a specific order consistent with the ordering (up to equivalence classes) determined by the observable univariate characteristic. More specifically, any contemplated population distribution $\underset{\sim}{N}$ is constrained to apply to the population individuals $I = 1,2,\ldots,N$ in a unique way, namely, the individual I is assumed to have observable characteristic j if and only if

(2.3) $$N^{j-1} < I \le N^j$$

where $N^0 = 0$ and

(2.4) $$N^j = \sum_{\ell=1}^{j} N_\ell ,$$

for $j = 1,2,\ldots,k$. The assumption (2.3) is illustrated in Table 1.

Individuals	Observable Categories
1	
2	
⋮	1
N_1	
N_1+1	
N_1+2	
⋮	2
N_1+N_2	
N_1+N_2+1	
⋮	
$N_1+N_2+ \quad +N_{k-1}$	
$N_1+N_2+ \quad +N_{k-1}+1$	k
⋮	
N	

Table 1. The population individuals 1,2,...,N. and their observable categories 1,2,...,k for a given $\underset{\sim}{N} = (N_1,N_2,\dots,N_k)$.

Observe that any $\underset{\sim}{N}$ satisfying (1.5) uniquely determines according to (2.3) the observable characteristic of each individual in the population. Thus any $[\underset{\sim}{J},\underset{\sim}{K}] \in X$ is consistent with a unique pair $\underset{\sim}{n}, \underset{\sim}{m}$ for each possible $\underset{\sim}{N}$. Necessary and sufficient conditions for $[\underset{\sim}{n}, \underset{\sim}{m}, \underset{\sim}{N}]$ to be consistent with $[\underset{\sim}{J}, \underset{\sim}{K}]$ are thus easily seen to be

$$J_{n^j} \le N^j < J_{n^j+1}, \text{ and}$$

(2.5)

$$K_{m^j} \le N^j < K_{m^j+1}$$

for $j = 1,2,\ldots,k-1$ where

(2.6) $$m^j = \sum_{\ell=1}^{j} n_\ell \text{ and } m^j = \sum_{\ell=1}^{j} m_\ell .$$

By setting

$$n^0 = m^0 = 0,$$

$$n^k = n,\ m^k = m$$

(2.7)

$$J_0 = K_0 = 0, \text{ and}$$

$$J_{n+1} = K_{m+1} = N + 1,$$

one may extend the range of application of (2.5) from $j = 1,2,\ldots, k-1$ to $j = 0,1,\ldots,k$.

The consistency relations (2.6) may be viewed abstractly as defining a multivalued mapping from X to S, i.e., the range of possible transforms of a given $[\underset{\sim}{J},\underset{\sim}{K}] \in X$ is the set of $[\underset{\sim}{n},\underset{\sim}{m},\underset{\sim}{N}] \in S$ which are consistent with it. Thus (2.6) gives the multivalued mapping essential for the application of Dempster (1966b).

Because there is one and only one possible transform corresponding to each $\underset{\sim}{N}$ satisfying (1.5) it is fairly easy to list the possible transforms in S of any $[\underset{\sim}{J},\underset{\sim}{K}] \in X$. Thus, for example, if $N = 4$, $n = 2$, $m = 1$ and $k = 2$, the element $[\underset{\sim}{J},\underset{\sim}{K}] = [(1,2),(3)]$ of X is consistent with

$$
(2.8) \qquad
\begin{aligned}
[\underset{\sim}{n}, \underset{\sim}{m}, \underset{\sim}{N}] \quad = \quad & [(0,2), \ (0,1), \ (0,4)], \\
& [(1,1), \ (0,1), \ (1,3)], \\
& [(2,0), \ (0,1), \ (2,2)], \\
& [(2,0), \ (1,0), \ (3,1)], \\
\text{or} \quad & [(2,0), \ (1,0), \ (4,0)].
\end{aligned}
$$

The reader may easily supply the analogues of (2.5) for each of the 11 remaining members of X in this example.

Having the elements of the model all specified, the logic of Dempster (1966b) applies directly. Upper and lower probabilities $P^*(T)$ and $P_*(T)$ for any $T \subset S$ are defined to be $\mu(T^*)$ and $\mu(T_*)$, respectively, where $T^* \subset X$ is the set of sample pairs $[\underset{\sim}{J},\underset{\sim}{K}]$ which are consistent with at least one $[\underset{\sim}{n},\underset{\sim}{m},\underset{\sim}{N}]$ in T, and $T_* \subset X$ is the set of $[\underset{\sim}{J},\underset{\sim}{K}]$ which are not consistent with any $[\underset{\sim}{n},\underset{\sim}{m},\underset{\sim}{N}]$ not in T.

The upper and lower probabilities which apply before $\underset{\sim}{n}$ is fixed are not very interesting. Any T defined solely by restrictions on $\underset{\sim}{N}$ has $P^*(T) = 1$ and $P_*(T) = 0$, the reason being that any $[\underset{\sim}{J},\underset{\sim}{K}]$ is consistent with some $[\underset{\sim}{n},\underset{\sim}{m},\underset{\sim}{N}]$ having any given $\underset{\sim}{N}$, so that $T^* = X$ while T_* is empty. A similar result holds if T is defined in terms of $\underset{\sim}{N}$ alone or in terms of m alone. Events defined in terms of m, n and N jointly may have upper and lower probabilities different from unity or zero, but these reflect only the known distribution of m and n once N is fixed.

After n is observed, the inferences must be correspondingly conditioned. This means that S is replaced by a subset R consisting of triples $[n,m,N]$ having the observed n. The multivalued mapping is unchanged in the sense that it is defined by the same consistency relation, but the set of possible transforms of a given $[J,K]$ in X is limited to the subset R' of S. Likewise, the same measure μ is used in the conditional situation to generate conditional upper and lower probabilities. In general models it usually happens that $P^*(R) = \mu(R^*) < 1$, so that some of the probability μ attaches to members of S which are ruled out after restriction to R. This possibility is reflected in the general definition of upper and lower conditional probabilities by a renormalization of the measure μ so that $\mu(R^*)$ becomes unity. The present situation is simpler because any $[J,K]$ in X is consistent with some $[n,m,N]$ for any fixed n. In effect <u>this means that upper and lower conditional probabilities given</u> n <u>are generated by continuing to regard all sample pairs</u> $[J,K]$ <u>to have the same probability</u> $[n!\ m!\ (N - n - m)!]\ /\ N!$ <u>even after</u> n <u>is known</u>.

The computation of upper and lower probabilities is thus reduced to counting $[\underset{\sim}{J},\underset{\sim}{K}]$ pairs. For a given event T and a given observation $\underset{\sim}{n}$, one finds $P^*(T)$ as A/B, where A is the number of $[\underset{\sim}{J},\underset{\sim}{K}]$ pairs which satisfy (2.6) for some triple $[\underset{\sim}{n},\underset{\sim}{m},\underset{\sim}{N}]$ having the given $\underset{\sim}{n}$ and satisfying T, and where B is the total number of $[\underset{\sim}{J},\underset{\sim}{K}]$ pairs, namely $N! / [n!\ m!\ (N-m-n)!]$. $P_*(T)$ is similarly found except that A is replaced by the number of pairs $[\underset{\sim}{J},\underset{\sim}{K}]$ which satisfy (2.6) only for triples $[\underset{\sim}{n},\underset{\sim}{m},\underset{\sim}{N}]$ with the given $\underset{\sim}{n}$ in T.

To provide a concrete illustration, the example $N = 4$, $n = 2$, $m = 1$ and $k = 2$ is analyzed here. Suppose that $\underset{\sim}{n}$ is observed to be (1,1). The restricted multivalued mapping appropriate to this observation is found by picking from (2.8) and its 11 analogues those members of S for which $\underset{\sim}{n} = (1,1)$. This is done in Table 2. Observe that R has only 4 elements.

$$(2.9) \qquad [\underset{\sim}{n},\underset{\sim}{m},\underset{\sim}{N}] = \begin{array}{l} [(1,1),\ (0,1),\ (1,3)] \\ [(1,1),\ (1,0),\ (2,2)] \\ [(1,1),\ (0,1),\ (2,2)] \\ \text{and } [(1,1),\ (1,0),\ (3,1)] \end{array}$$

Upper and lower conditional probabilities given $\underset{\sim}{n} = (1,1)$

Elements [J,K] in X	Consistent elements [m, N] in S with n = (1,1)
[(1,2),(3)]	[(0,1),(1,3)]
[(1,2),(4)]	[(0,1),(1,3)]
[(1,3),(2)]	[(0,1),(1,3)] and [(1,0),(2,2)]
[(1,3),(4)]	[(0,1),(1,3)] and [(1,0),(2,2)]
[(1,4),(2)]	[(0,1),(1,3)] [(1,0),(2,2)] and [(1,0),(3,1)]
[(1,4),(3)]	[(0,1),(1,3)], [(0,1),(2,2)], and [(1,0),(3,1)]
[(2,3),(1)]	[(1,0),(2,2)]
[(2,3),(4)]	[(0,1),(2,2)]
[(2,4),(1)]	[(1,0),(2,2)] and [(1,0),(3,1)]
[(2,4),(3)]	[(0,1),(2,2)] and [(1,0),(3,1)]
[(3,4),(1)]	[(1,0),(3,1)]
[(3,4),(2)]	[(1,0),(3,1)]

Table 2. The multivalued mapping from X to the subset of S with n = (1,1) when N = 4, n = 2, m = 1 and k = 2.

are easily read from Table 2 for any $T \subset S$. For example, if T is the event or hypothesis that $\underset{\sim}{N} = (1,3)$, this $\underset{\sim}{N}$ is observed to appear on six of the twelve lines of Table 2 and appears exclusively on two of these lines so that $P^*(T) = 6/12$ and $P_*(T) = 2/12$. Similarly, if T_1 is the complimentary event that $\underset{\sim}{N} = (2,2)$ or $(3,1)$, then $P^*(T_1) = 10/12$ and $P_*(T_1) = 6/12$, or if T_2 is the event that $\underset{\sim}{m} = (1,0)$, then $P^*(T_2) = 9/12$ and $P_*(T_2) = 3/12$, and so on.

3. Formulas for upper and lower probabilities

Given n and the event T defined by (1.8), what are the appropriate $P^*(T)$ and $P_*(T)$? The first task is to characterize the subsets T^* and T_* of X, where T^* and T_* are assumed here to have been conditioned by the observed n. Thus T^* is the set of [J,K] pairs which are consistent with some [n,m,N] both satisfying T and having the given n, while T_* is the subset of T^* whose [J,K] pairs are consistent with no [n,m,N] outside T.

Theorem 3.1

$$T^* = \{r \le N(O) \le t\} \cup \{N(O) > t,\ N(I) \le t\} \tag{3.1}$$

and

$$T_* = \{N(O) \le t\} \cap \{N(I) \ge r\} \tag{3.2}$$

where O is the subset of I = 0,1,2,...,N + 1 satisfying

$$J_n^{a-1} < I \le J_n^{b}+1, \tag{3.3}$$

where I is the subset of I = 0,1,2,...,N + 1 satisfying

$$\begin{bmatrix} J_n^{a-1}+1 \text{ if } a > 1 \\ 0 \text{ if } a = 1 \end{bmatrix} \le I \le \begin{bmatrix} J_n^{b} \text{ if } b = k \\ n+1 \text{ if } b < k \end{bmatrix} \tag{3.4}$$

and where N(O) and N(I) are respectively the numbers of elements of K common with O and common with I.

Note that O and I may be thought of as intervals on the line, where only the discrete points $0,1,2,\ldots,N+1$ play a role. Since $\mathbf{Q}$ contains $\mathbf{I}$, one may regard $\mathbf{Q}$ as abbreviating *outer interval* and $\mathbf{I}$ as abbreviating *inner interval*.

To prove (3.1) consider the four mutually exclusive and exhaustive events $\{r \le N(\mathbf{Q}) \le t\}$, $\{N(\mathbf{Q}) > t,\ N(\mathbf{I}) \le t\}$, $\{N(\mathbf{Q}) < r\}$ and $\{N(\mathbf{I}) > t\}$. It will be argued directly that any $[\mathbf{J},\mathbf{K}]$ pair in the first or second of these sets is consistent with some $[\mathbf{n},\mathbf{m},\mathbf{N}]$ satisfying T and having the given $\mathbf{n}$, while no pair $[\mathbf{J},\mathbf{K}]$ in the third or fourth set has such a consistency relation. Recall that for a given $[\mathbf{J},\mathbf{K}]$ one need fix only $\mathbf{N}$ to determine a unique consistent $[\mathbf{n},\mathbf{m},\mathbf{N}]$ triple. Given $[\mathbf{J},\mathbf{K}]$ satisfying $\{r \le N(\mathbf{Q}) \le t\}$, choose $\mathbf{N}$ in any way such that N^{a-1} and N^{b} are as far apart as possible consistent with the first line of (2.5). Then the second line of (2.5) together with (2.3) insures that the consistent $\mathbf{m}$ going with that $\mathbf{N}$ and the given $\mathbf{n}$ must satisfy T. Figure 1 should help the reader to visualize this argument. Secondly, given $[\mathbf{J},\mathbf{K}]$ satisfying $\{N(\mathbf{Q}) > t,$ $N(\mathbf{I}) \le t\}$, one may choose $\mathbf{N}$ with N^{a-1} and N^{b} moved in from their most separated positions so that $N(\mathbf{I}) = t$, thus again satisfying T. On the other hand, given $[\mathbf{J},\mathbf{K}]$ satisfying $\{N(\mathbf{Q}) < r\}$, however far apart N^{a-1} and N^{b} are chosen it always happens that $\Sigma_a^b\, m_j < r$.

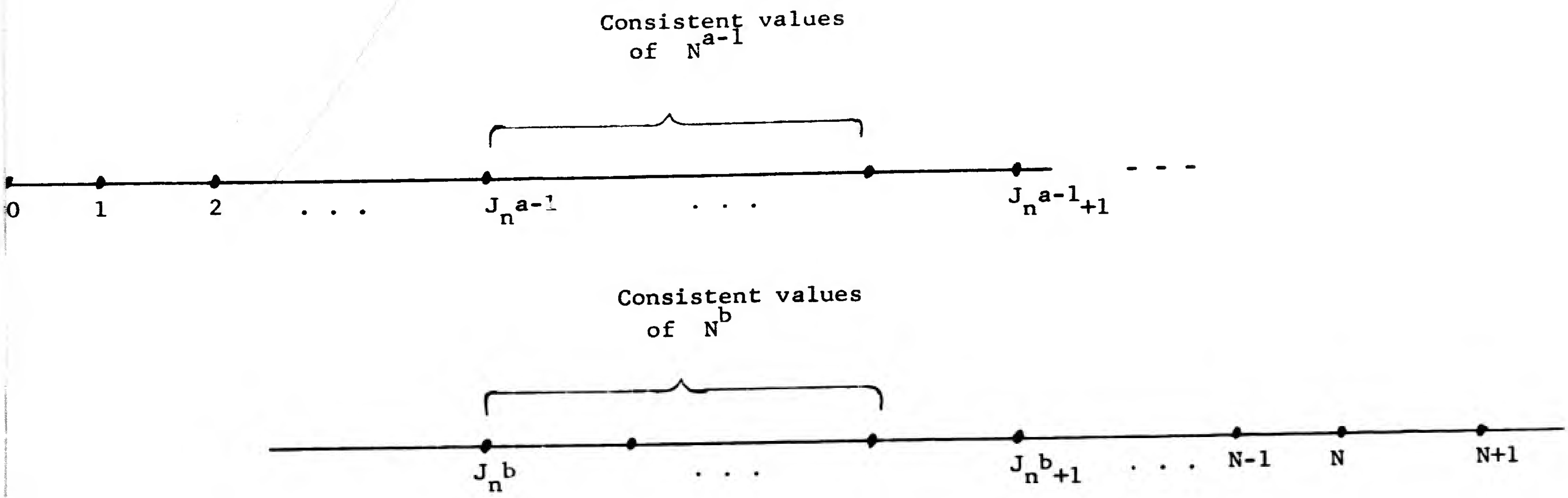

Figure 1 The ranges of N^{a-1} and N^b consistent with $(n_1, n_2, \ldots, n_k)$ according to (2.5). Similar restrictions hold with K in place of J.

Finally, given $[\underset{\sim}{J},\underset{\sim}{K}]$ satisfying $\{N(\underset{\sim}{I}) > t\}$, however close N^{a-1} and N^b are chosen it always happens that $\Sigma_a^b m_j > t$. This completes the proof of (3.1).

By similar reasoning, if the right side of (3.2) fails for some $[\underset{\sim}{J},\underset{\sim}{K}]$ then T need not hold. For, if $\{N(\underset{\sim}{O}) > t\}$, then placing N^{a-1} and N^b as far apart as possible yields $[\underset{\sim}{n},\underset{\sim}{m},\underset{\sim}{N}]$ with $\Sigma_a^b m_j > t$, while if $\{N(\underset{\sim}{I}) < r\}$, then placing N^{a-1} and N^b as close together as possible yields $[\underset{\sim}{n},\underset{\sim}{m},\underset{\sim}{N}]$ with $\Sigma_a^b m_j < r$. On the other hand, if the right side of (3.2) holds for some $[\underset{\sim}{J},\underset{\sim}{K}]$, then, however N^{a-1} and N^b are placed consistent with (2.5), the resulting $[\underset{\sim}{n},\underset{\sim}{m},\underset{\sim}{N}]$ satisfies T. Thus (3.2) is proved.

Since $N(\underset{\sim}{O})$ and $N(\underset{\sim}{I})$ depend only on the relative orders of $J_1,J_2,\ldots,J_n$ and $K_1,K_2,\ldots,K_m$, it follows from (3.1) and (3.2) that $P^*(T)$ and $P_*(T)$ may be computed considering only the sample space of $\binom{n+m}{n}$ equally likely relative orders rather than the full sample space of pairs of samples. This means that $P^*(T)$ and $P_*(T)$ depend on m and n but not on N. The following derivations may be conceptually simplified by assuming $m + n = N$ so that $\underset{\sim}{J}$ and $\underset{\sim}{K}$ together run over the integers $1,2,\ldots,N$.

Some simple notation and lemmas are given in Section 6 for certain discrete distributions which arise frequently in the rest of this section, and the reader who wishes to follow the details must digress at this point to section 6.

Upper and lower probabilities will be derived first for two boundary cases.

Three situations will be distinguished:

Case A : $a = 1$ and $b = k$.

Case B : $n^b - n^{a-1} = 0$ or 1

if $1 < a \leq b \leq k$

(3.5) $n^b - n^{a-1} = 0$

if $a = 1,\ b < k$ or $a > 1,\ b = k$.

Case C : all remaining cases.

Case A may be disposed of immediately, because under it $\sum_1^k m_j$ is necessarily m, so that

$$P^*(T) = P_*(T) = 1 \quad \text{if} \quad t = m$$

$$(3.6) \qquad = 0 \quad \text{if} \quad t < m.$$

Case B has the special feature that $N(I)$ is necessarily zero so that (3.1) and (3.2) reduce to

$$(3.7) \qquad T^* = \{N(Q) > r\}$$

and

$$(3.8) \qquad T_* = \{N(Q) < t\} \quad \text{if } r = 0$$

$$= \phi \qquad \text{if } r > 0$$

$P^*(T)$ may be expressed from Lemma 6.1 as the tail of an FNB distribution and thence from (6.6) as the tail of a hypergeometric distribution. The formulas are not reproduced here since they are covered by (3.14) where the second term on the right side of (3.14) is interpreted as zero. Likewise $P_*(T)$ is zero unless $r = 0$, and, in the case $r = 0$, $P_*(T)$ is given by the right side of (3.15) with the first term unity and the third term zero.

To handle the general case C, consider the general expressions (3.1) and (3.2). Figure 2 shows the regions in the $N(\underset{\sim}{I})$, $N(\underset{\sim}{O})$ plane which determine T^* and T_*. The second picture is included because $N(\underset{\sim}{I})$ and $N(\underset{\sim}{O}) - N(\underset{\sim}{I})$ have the bivariate FNB distribution according to Lemma 6.3.

Theorem 3.2. Under case C

$$P^*(T) = \sum_{u=0}^{t} \sum_{v=\max\{0,r-u\}}^{m-u} c(u,v)$$

(3.9)

and

$$P_*(T) = \sum_{u=r}^{t} \sum_{v=0}^{t-u} c(u,v),$$

(3.10)

where

$$c(u,v) = c(u,v;\ n^b - n^{a-1}-1,\ 2,\ n,\ n+m)$$

(3.11) if $a > 1$ and $b < k$,

$$= c(u,v;\ n^b,\ 1,\ n,\ n+m)$$

if $a = 1$ and $b < k$,

$$= c(u,v;\ n - n^{a-1},\ 1,\ n,\ n+m)$$

if $a > 1$ and $b = k$.

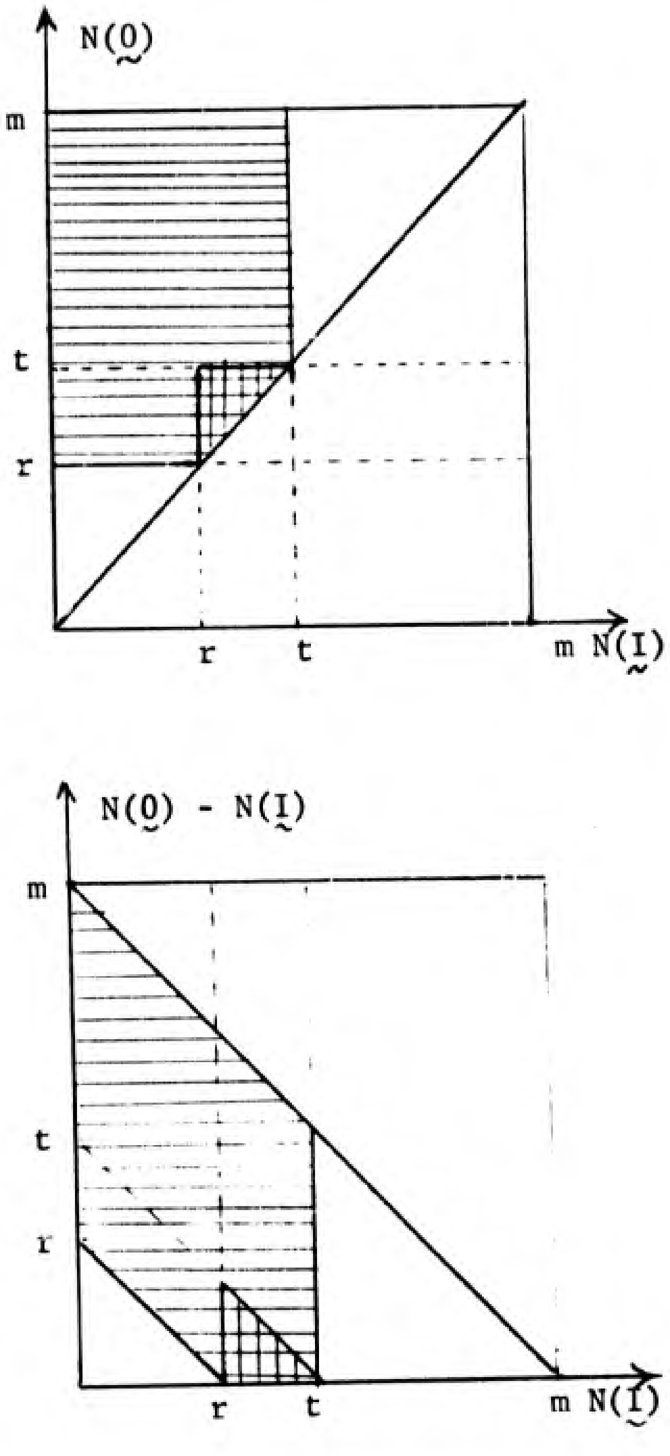

Figure 2. T* shown as horizontal shading and

T_* shown as vertical shading.

The proof of Theorem 3.2 is immediate from (3.1), (3.2) and Lemma 6.3.

It will next be shown that double summation may be eliminated from the expressions for $P^*(T)$ and $P_*(T)$ with the result that such probabilities may be computed with relative ease given tables of the hypergeometric distribution (e.g., Lieberman and Owen (1961)). From Figure 2 it is clear that

$$P^*(T) = \Pr\{N(\underset{\sim}{O}) \geq r\} - \Pr\{N(\underset{\sim}{I}) > t\} \tag{3.12}$$

and

$$\begin{aligned} P_*(T) &= \Pr\{N(\underset{\sim}{I}) \geq r\} - \Pr\{N(\underset{\sim}{O}) > t\} \\ &\quad + \Pr\{N(\underset{\sim}{I}) < r,\ N(\underset{\sim}{O}) > t.\} \end{aligned} \tag{3.13}$$

All of the terms of the right side of (3.12) and (3.13), excepting the last, are expressible from Lemma 6.1 as tail probabilities of FNB distributions and thence from (6.6) as tail probabilities of hypergeometric distributions. The last term of (3.13) is provided by Lemma 6.4. This yields Theorem 3.3.

<u>Theorem 3.3</u> <u>Under Case C</u>

$$\begin{aligned} P^*(T) &= D(n^b - n^{a-1};\ n^b - n^{a-1} + r,\ n,\ n + m) \\ &\quad - D(n^b - n^{a-1} - 2;\ n^b - n^{a-1} + t - 1,\ n,\ n + m) \\ &\qquad \underline{\text{if}}\ \ 1 < a \leq b < k, \\ &= D(n^b;\ n^b + r,\ n,\ n + m) \\ &\quad - D(n^b - 1;\ n^b + t,\ n,\ n + m) \\ &\qquad \underline{\text{if}}\ \ 1 = a \leq b < k \end{aligned} \tag{3.14}$$

$$= D(n - n^{a-1}; n - n^{a-1} + r, n, n + m)$$

$$- D(n - n^{a-1} - 1; n - n^{a-1} + t, n, n + m)$$

$$\text{if } 1 < a \leq b = k,$$

and

$$P_*(T) = D(n^b - n^{a-1} - 2; n^b - n^{a-1} + r - 2, n, n + m)$$

$$- D(n^b - n^{a-1}; n^b - n^{a-1} + t + 1, n, n + m)$$

(3.15) $$+ e(r,t; n^b - n^{a-1} - 1, 2, n, n + m)$$

$$\text{if } 1 < a \leq b < k,$$

$$= D(n^b - 1; n^b + r - 1, n, n + m)$$

$$- D(n^b; n^b + t + 1, n, n + m)$$

$$+ e(r, t; n^b, 1, n, n + m)$$

$$\text{if } 1 = a \leq b < k$$

$$= D(n - n^{a-1} - 1; n - n^{a-1} + r - 1, n, n + m)$$

$$- D(n - n^{a-1}; n - n^{a-1} + t + 1, n, n + m)$$

$$+ e(r, t; n - n^{a-1}, 1, n, n + m)$$

$$\text{if } 1 < a \leq b = k.$$

It may be noted that the first two terms in (3.15) provide a lower bound for $P_*(T)$. When the third term is small, as often is the case, this lower bound is a good approximation to $P_*(T)$. It is also clear from (3.13) that the third term is zero if either $r = 0$ or $t = m$.

Numerical Example Suppose that a sample of size 10 is drawn from a population of size 100 and yields the observation vector $\underset{\sim}{n} = (1, 3, 2, 4, 0)$ where $k = 5$.

Upper and lower probabilities will be computed for the events

$$T_1 = \{m_2 + m_3 + m_4 = 1\} \quad \text{with} \quad m = 1$$

$$T_2 = \{m_2 + m_3 + m_4 \geq 5\} \quad \text{with} \quad m = 10$$

$$T_3 = \{5 \leq m_2 + m_3 + m_4 \leq 9\} \quad \text{with} \quad m = 10$$

$$T_4 = \{N_2 + N_3 + N_4 \geq 50\}$$

$$T_5 = \{50 \leq N_2 + N_3 + N_4 \leq 90.\}$$

From (3.14)

$$\begin{aligned}
P^*(T_1) &= D(9;10,10,11) - D(7;9,10,11) \\
&= .909091 - 0 \\
&= .909091 , \\
P^*(T_2) &= D(9;14,10,20) - D(7;18,10,20) \\
&= .994582 - 0 \\
&= .994582 , \\
P^*(T_3) &= D(9;14,10,20) - D(7;17,10,20) \\
&= .994582 - .105263 \\
&= .889319 , \\
P^*(T_4) &= D(9;50,10,100) - D(7;98,10,100) \\
&= .999406 - 0 \\
&= .999406 , \\
P^*(T_5) &= D(9;50,10,100) - D(7;89,10,100) \\
&= .999406 - .077931 \\
&= .921475
\end{aligned}$$

From (3.15)

$$
\begin{aligned}
P_*(T_1) &= D(7;8,10,11) - D(9;11,10,11) \\
&\quad + e(1,1;8,2,10,11) \\
&= .727273 - 0 + 0 \\
&= .727273, \\
P_*(T_2) &= D(7;12,10,20) - D(9;20,10,20) \\
&\quad + e(5,10;8,2,10,20) \\
&= .915099 - 0 + 0 \\
&= .915099. \\
P_*(T_3) &= D(7;12,10,20) - D(9;19,10,20) \\
&\quad + e(5,9;8,2,10,20) \\
&= .915099 - .500000 + 0 \\
&= .415099 \\
P_*(T_4) &= D(7;48,10,100) - D(9;100,10,100) \\
&\quad + e(41,90;8,2,10,100) \\
&= .965678 - 0 + 0 \\
&= .965678, \\
P_*(T_5) &= D(7;48,10,100) - D(9;91,10,100) \\
&\quad + e(41,81;8,2,10,100) \\
&= .965678 - .628724 + .009897 \\
&= .34685
\end{aligned}
$$

4. Related confidence statements

The assertion $T = \{r \le m^b - m^{a-1} \le t\}$ may be called <u>doubly one-sided</u> if $a = 1$ or $b = k$ <u>and</u> $r = 1$ or $t = m$. For moderate n one could show that there are natural confidence arguments which produce roughly the same answers as the upper and lower probability argument but for the doubly one-sided cases a precise relationship may be exhibited. Specifically, the statement that any doubly one-sided T has lower probability $P_*(T)$ may be re-interpreted after the fact as a confidence statement with confidence coefficient $P_*(T)$.

The four possibilities are essentially the same, so consider $a = 1$ and $t = m$, and ask how one might construct a confidence interval of the form $\{r(\underset{\sim}{n}) \le m^b \le m\}$ with a given confidence coefficient $1 - \alpha$. Such an interval should have the property

$$(4.1) \qquad \Pr\{r(\underset{\sim}{m}) \le m^b < m \mid \underset{\sim}{N}\} \ge 1 - \alpha$$

for all $\underset{\sim}{N}$. Ideally, one would like equality in (4.1), but in the presence of discretness one must settle for inequality, and for some α even strict inequality for all $\underset{\sim}{N}$.

The intervals to be defined will have the property (4.1) conditionally not only given $\underset{\sim}{N}$ but also given $\underset{\sim}{n} + \underset{\sim}{m}$. The derivation parallels and indeed generalizes the familar derivation for binomial p. Consider a horizontal axis for $n^b + m^b$ and a vertical axis for n^b as in Figure 3. Fixing $\underset{\sim}{n} + \underset{\sim}{m}$ also fixes $n^b + m^b$, and along the vertical line with fixed $n^b + m^b$

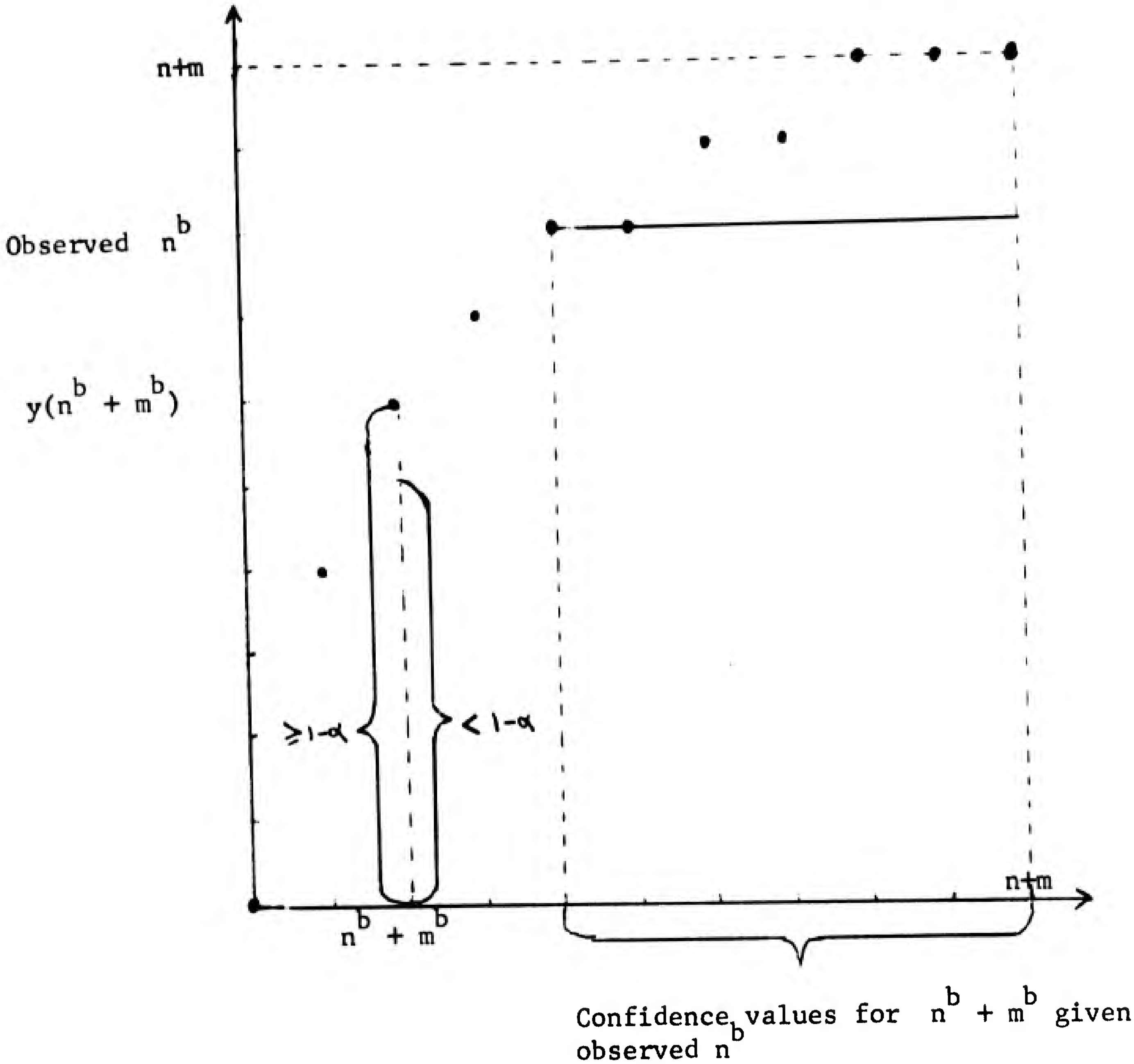

Figure 3. Schematic illustration of the type of confidence statements derived in Section 4.

one may visualize the hypergeometric distribution with density $d(n^b;\ n^b + m^b,\ n,\ n + m)$ which is the conditional density of n^b given $\underset{\sim}{n} + \underset{\sim}{m}$. On each such line one may determine the point whose ordinate $y(n^b + m^b)$ is the smallest integer such that

$$(4.2) \qquad D(y(n^b + m^b);\ n^b + m^b,\ n,\ n + m) \geq 1 - \alpha .$$

The corresponding confidence interval for m^b is of the form $\{r \leq m^b \leq m\}$ where $n^b + r$ is the smallest value of $n^b + m^b$ such that $y(n^b + m^b) \geq m^b$.

The relation of these confidence statements to lower probability statements will be demonstrated by carrying out the following program. First, given any $T = \{r \leq m^b \leq m\}$ and any sample observation $\underset{\sim}{n}$, the lower probability $P_*(T)$ will be computed. Second, the confidence argument defined above with $P_*(T)$ substituted for $1 - \alpha$ will be shown to produce the same interval T.

$P_*(T)$ may be computed as $1 - P^*(\overline{T})$ where $\overline{T}$ is the complementary event $\{0 \leq m^b \leq r - 1\}$. $P^*(\overline{T})$ is given by the second line of (3.9) whose first term in this case in unity. Thus

$$(4.3) \qquad P_*(T) = D(n^b - 1;\ n^b + r - 1,\ n,\ n + m).$$

A comparison of (4.2) with (4.3) shows that, in the construction of the confidence argument with $1 - \alpha = P_*(T)$, one has

(4.4) $$y(n^b + r - 1) = n^b - 1$$

Moreover, equality holds in this application of (4.2). Since replacing $r - 1$, by r increases the right side of (4.3), it follows that

(4.5) $$y(n^b + r) \geq n^b.$$

Relations (4.4) and (4.5) show that the confidence interval has the form $r \leq m^b \leq m$.

It has sometimes seemed puzzling that complementary confidence intervals should in certain discrete situations have confidence coefficients summing to less than unity. Here, the intervals $T = \{r \leq m^b \leq m\}$ and $\bar{T} = \{0 \leq m^b \leq r - 1\}$ may be interpreted as having confidence coefficients whose sum is $P_*(\bar{T}) + P_*(T) = 1 - [P^*(T) - P_*(T)]$ and the shortfall may be explained in terms of a difference between upper and lower probabilities.

Some readers may regard the theory of this section as providing, at least for doubly one-sided intervals, a justification of the upper and lower probability inferences derived in Section 3. The author's attitude is rather the opposite. He regards the confidence property as being very weak, even disturbingly weak in view of the widespread misinterpretation of confidence coefficients as posterior probabilities. The great need is to ascertain when such misinterpretation may be justified.

5. Special Cases.

When $m = N - n$, the statement $T = \{r \le m^b - m^{a-1} \le t\}$ may also be written $T = \{r + n^b - n^{a-1} \le N^b - N^{a-1} \le t + n^b - n^{a-1}\}$, so that formulas may be immediately given for $P^*(Q)$ and $P_*(Q)$ where

(5.1) $$Q = \{r \le \Sigma_a^b N_j \le t\}$$

Proceeding to the limit as $N \to \infty$ does not alter $P^*(T)$ and $P_*(T)$, but does give new forms for $P^*(Q)$ and $P_*(Q)$. If the limits of N_j/N, r/N and t/N are denoted by p_j, ρ and τ then the limiting form of Q may be written $\{\rho \le \Sigma_a^b p_j \le \tau\}$, and $P^*(Q)$ and $P_*(Q)$ provide inference about a set of unknown multinomial probabilities. Formulas (3.9) and (3.10) become

(5.2) $$P^*(Q) = \int_{x=0}^{\tau} \int_{y=\max\{0,\rho-x\}}^{1-x} b(x,y)\,dx\,dy$$

and

(5.3) $$P_*(Q) = \int_{x=\rho}^{\tau} \int_{y=0}^{\tau-x} b(x,y)\,dx\,dy$$

where

(5.4) $$b(x,y) = b(x,y;\, n^b - n^{a-1} - 1,\, 2,\, n - n^b + n^{a-1})$$

if $a > 1$ and $b < k$,

$$= b(x,y;\, n^b,\, 1,\, n - n^b)$$

if $a = 1$ and $b < k$

$$= b(x,y;\, n - n^{a-1},\, 1,\, n^{a-1})$$

if $a > 1$ and $b = k$,

and $b(x,y;\ r,s,t)$ denotes the Dirichlet density

$$(5.5)\quad b(x,y;\ r,s,t) = \frac{(r+s+t-1)!}{(r-1)!\ (s-1)!\ (t-1)!}\ x^{r-1}\ y^{s-1}\ (1-x-y)^{t-1}$$

for $0 \leq x,\ 0 \leq y,\ x + y \leq 1,\ r > 0,\ s > 0,\ t > 0.$

In a similar fashion (3.14) and (3.15) may be extended using tails of beta distribution. Setting

$$(5.6)\quad B(x;\ a,b) = \int_0^x \frac{\Gamma(a+b)}{\Gamma(a)\ \Gamma(b)}\ z^{a-1}\ (1-z)^{b-1} dz$$

one finds from (3.14)

$$
\begin{aligned}
P^*(Q) &= B(\tau;\ n^b - n^{a-1} - 1,\ n - n^b + n^{a-1}) \\
&\quad -B(\rho;\ n^b - n^{a-1} + 1,\ n - n^b + n^{a-1} - 2) \\
&\qquad \text{if } 1 < a \leq b < k \\
&= B(\tau;\ n^b,\ n - n^b - 1) \\
&\quad -B(\rho;\ n^b + 1,\ n - n^b - 2) \\
&\qquad \text{if } 1 = a \leq b < k \\
&= B(\tau;\ n - n^{a-1},\ n^{a-1} - 1) \\
&\quad -B(\rho;\ n - n^{a-1} + 1,\ n^{a-1} - 2) \\
&\qquad \text{if } 1 < a \leq b = k,
\end{aligned}
\tag{5.7}
$$

and from (3.15)

$$P_*(Q) = B(\tau;\ n^b - n^{a-1} + 1,\ n - n^b + n^{a-1} - 2)$$

(5.8) $$-B(\rho;\ n^b - n^{a-1} - 1,\ n - n^b + n^{a-1})$$

$$+\binom{n}{n^b - n^{a-1}}\rho^{n^b - n^{a-1}}(1-\tau)^{n-n^b+n^{a-1}}$$

$$+\frac{n!}{(n^b - n^{a-1} - 1)!\ (n - n^b + n^{a-1})!}\rho^{n^b - n^{a-1} - 1}(\rho-\tau)(1-\tau)^{n-n^b+n^{a-1}}$$

$$+\binom{n}{n^b - n^{a-1} - 1}\rho^{n^b - n^{a-1} - 1}(1-\tau)^{n-n^b+n^{a-1}+1}$$

if $1 < a \leq b < k$,

$$= B(\tau;\ n^b + 1,\ n - n^b - 2)$$

$$-B(\rho;\ n^b,\ n - n^b - 1)$$

$$+\binom{n}{n^b}\rho^{n^b}(1-\tau)^{n-n^b}$$

if $1 = a \leq b < k$,

$$= B(\tau;\ n - n^{a-1} + 1,\ n^{a-1} - 2)$$

$$-B(\rho;\ n - n^{a-1},\ n^{a-1} - 1)$$

$$+\binom{n}{n - n^{a-1}}\rho^{n - n^{a-1}}(1-\tau)^{n^{a-1}}$$

if $1 < a > b = k$.

The special case $k = 2$, $N = \infty$ was worked out in Dempster (1966a) so that (5.8), (5.9), (3.14) and (3.15) generalize formulas given in the previous work.

It is possible to let $k \to \infty$ and treat the case of a continuous observable. This may be done either for finite N or for $N = \infty$. In either case the foregoing formulas apply directly. A similar analysis with $k = \infty$ and $N = \infty$ was carried out in Section 5 of Dempster (1963) with results slightly different from those given here. The difference results from a difference in the hypothesized family of population distributions. The earlier paper restricted the family to continuous distributions, and is therefore able to make certain probability judgements with less uncertainty.

6. Appendix: distributions

Notation and simple results are set up in this appendix for the distributions required in Section 3. The results are neither original nor deep and are collected here chiefly to smooth the reader's path through Section 3. To cover various end cases, it will be assumed that

$$\text{(6.1)} \qquad \binom{x}{y} = \binom{x}{x-y} = 1 \text{ if } y < 0 \text{ and } x = y$$
$$= 0 \text{ if } y < 0 \text{ and } x > y.$$

A discrete distribution over the integers $w = 0, 1, 2, \ldots, Z - Y$ may be defined by the density

$$\text{(6.2)} \qquad c(w;\, X,\, Y,\, Z) = \frac{\binom{X-1+w}{X-1}\binom{Z-X-w}{Y-X}}{\binom{Z}{Y}}$$

for any integers X, Y, Z satisfying $0 \leq X \leq Y \leq Z$. The corresponding cumulative distribution function will be denoted by $C(w;\, X,\, Y,\, Z)$. Note that $X = 0$ provides a degenerate case in the sense that

$$\text{(6.3)} \qquad c(w;\, 0,\, Y,\, Z) = 1 \text{ if } w = 0$$
$$= 0 \text{ if } w > 0.$$

The family of distributions defined by (6.2) is closely related to the family of hypergeometric distributions defined by the density

(6.4) $$d(w;\ W,\ Y,\ Z) = \frac{\binom{W}{w}\binom{Z-W}{Y-w}}{\binom{Z}{Y}}$$

for

(6.5) $$\max\{0,\ W+Y-Z\} \le w \le \min\{W,\ Y\}$$ given

any integers W, Y, Z satisfying $0 \le W \le Z$ and $0 \le Y \le Z$. The corresponding cumulative distribution function will be denoted by $D(w;\ W,\ Y,\ Z)$.

To derive a basic relationship between these distributions, consider an urn containing Z balls of which Y are black. If balls are drawn out sequentially at random, the probability that $X + w$ drawings are required to produce the Xth black ball is clearly $c(w;\ X,\ Y,\ Z)$ On the other hand, a fixed number $X + w$ of drawings produces X black balls with probability $d(X;\ X + w,\ Y,\ Z)$. Moreover, the assertion that the Xth black ball occurs at or before trial $X + w$ is equivalent to the assertion that the first $X + w$ trials yield X or more black balls, so that

(6.6) $$C(w;\ X,\ Y,\ Z) = 1 - D(X - 1;\ X + w,\ Y,\ Z)$$

for $0 \le w \le Z - Y,\ 0 \le X \le Y \le Z$.

The family (6.2) occurs in applications about as widely as the hypergeometric family (6.4) but does not appear to have acquired a widely accepted name. Lieberman and Owen (1961) use the term distribution of the number of exceedances while Raiffa and Schlaifer (1961) prefer beta-binomial distribution in honor of its appearance in the Bayesian theory of conjugate distributions. Both of these sources give (6.6), which, since it is the finite sampling generalization of the well known relation between the tails of the negative binomial and binomial distributions suggests the name finite negative binomial. FNB will be used here as a convenient shorthand.

FNB distributions arise in Section 3 as applications of the following lemma.

Lemma 6.1. *Suppose that* n *objects* (J's) *of one kind and* m *objects* (K's) *of a second kind are randomly arranged on a line with all arrangements equally probable. Suppose that an additional* J *called the 0th* J *is placed at the left of all the objects and another additional* J *called the* (n+1) st J *is placed at the right of all the objects. Define* $N(u,v)$ *to be the number of* K's *between the uth* J *and the vth* J. *Then*

$$\text{(6.7)} \qquad \Pr(N(u,v) = w) = c(w;\, v - u,\, n,\, n + m)$$

for $0 \le u \le v \le n + 1$.

The proof requires only simple combinatorics, treating separately the three cases $1 \le u \le v \le n$, $u = 0$, $v < n$ and $u \ge 1$, $v = n + 1$. The details are omitted.

A bivariate generalization of the FNB distributions also appears in Section 3. The general form of the FNB bivariate density is

$$(6.8) \qquad c(w, w^*; X, X^*, Y, Z)$$

$$= \frac{\binom{X - 1 + w}{X - 1} \binom{X^* - 1 + w^*}{X^* - 1} \binom{Z - X - X^* - w - w^*}{Y - X - X^*}}{\binom{Z}{Y}}$$

where $0 \leq w$, $0 \leq w^*$, $w + w^* \leq Z - Y$,

$0 \leq X$, $0 \leq X^*$, $X + X^* \leq Y \leq Z$.

It is easily checked that, if W and W^* are random variables having the density (6.8), then the marginal distributions of W, W^* and $W + W^*$ are each FNB with the roles of X in (6.2) played by X, X^* and $X + X^*$, respectively. Likewise, the conditional density of W^* given $W = w$ is $c(w^*; X^*, Y - X, Z - X - w)$. If $X = 0$, the distribution (6.8) degenerates to a univariate FNB distribution along the line $w = 0$.

A simple situation in which the distribution (6.8) arises is given by Lemma 6.2 following. A slightly more complicated situation is given by Lemma 6.3. Proofs are omitted.

<u>Lemma 6.2.</u> <u>Under the hypotheses of Lemma 6.1</u>

$$(6.9) \qquad \Pr\{N(0,u) = w, N(u,v) = w^*\}$$

$$= c(w, w^*; u, v - u, n, n + m)$$

<u>for</u> $0 \leq u \leq v \leq m + 1$.

Lemma 6.3. Under the hypotheses of Lemma 6.1.

$$\Pr\left\{N(u,v) = w,\ N(u^*, v^*) - N(u,v) = w^*\right\}$$

(6.10)

$$= c(w, w^*; v - u, v^* - u^* - v + u, n, n + m)$$

for $0 \leq u^* < u \leq v \leq v^* < n + 1$.

On further lemma is used in Section 3.

Lemma 6.4 If

(6.11)
$$e(r,t;\ X, X^*, n, n + m) = \sum_{\substack{w < r \\ w + w^* > t}} \sum c(w,w^*;\ X, X^*, n, n + m),$$

then

(6.12)
$$e(r,t;\ X, 1, n, n + m) = \frac{\binom{X + r - 1}{X} \binom{m + n - X - t - 1}{n - X}}{\binom{m + n}{m}}$$

and

(6.13)
$$e(r, t;\ X, 2, n, n + m) = \left[\left(\frac{X + r}{X + 1}\right) + (t - r) + \left(\frac{m + n - t - X - 1}{n - X}\right)\right] \frac{\binom{X + r - 1}{X} \binom{m + n - t - X - 2}{n - X - 1}}{\binom{m + n}{n}}$$

for $0 \leq r \leq t \leq m$

To prove Lemma 6.4 consider again the formulation of Lemma 6.1 where $J_0, J_1, J_2, \ldots, J_{n+1}$ denotes the J's in order and $K_1, K_2, \ldots, K_m$ denotes the K's. In view of Lemma 6.2, an event whose probability is given by the left side of (6.12) is $\{J_X < K_r < K_{t+1} < J_{X+1}\}$ and this in turn means that the first $X + r - 1$ objects consist of $r - 1$ K's and X J's, the next $t - r + 2$ objects are all K's and the last $m + n - X - t - 1$ objects consist of $m - t - 1$ K's and $n - X$ J's. The numerator of the right side of (6.12) counts the number of arrangements satisfying these restrictions. Similarly, the left side of (6.13) is the probability of the event.

(6.14)
$$\begin{aligned} &\{J_X < K_r < K_{t+1} < J_{X+2}\} \\ &= \{J_{X+1} < K_r < K_{t+1} < J_{X+2}\} \\ &\cup \{J_X < K_r < J_{X+1} < K_{t+1} < J_{X+2}\} \\ &\cup \{J_X < K_r < K_{t+1} < J_{X+1}\} \end{aligned}$$

The three terms on the right side of (6.13) come directly from the three terms in (6.14).

7. Acknowledgement

The author thanks Robert Kleyle for valuable assistance with some of the details.

REFERENCES

DEMPSTER, A. P. (1963) On direct probabilities. J. Roy. Statist. Soc. Ser. B 25 100-110

DEMPSTER, A. P. (1966a) New methods for reasoning towards posterior distributions based on sample data Ann. Math. Statist. 37 355-374

DEMPSTER, A. P. (1966b) Upper and lower probabilities induced by a multivalued mapping.

ONR Technical Report No. 9
Department of Statistics, Harvard University

LIEBERMAN, GERALD J. and OWEN, DONALD B. (1961). Tables of the Hypergeometric Probability Distribution Stanford University Press, Stanford, California.

RAIFFA, HOWARD and SCHLAIFER, ROBERT (1961). Applied Statistical Decision Theory. Division of Research, Harvard Business School, Boston.

unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Harvard University Department of Statistics Cambridge, Massachusetts	2a. REPORT SECURITY CLASSIFICATION 2b. GROUP

3. REPORT TITLE

Upper and lower probability inferences based on a sample from a finite univariate population

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report, October 10, 1966

5. AUTHOR(S) (Last name, first name, initial)

Dempster, A. P.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October 10, 1966	39	5

8a. CONTRACT OR GRANT NO. Nonr 1866(37) b. PROJECT NO. NR 042-097 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Technical report No. 14 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Logistics and Mathematical Statistics Branch Office of Naval Research Department of the Navy Washington, D. C.

13. ABSTRACT

Detailed formulas are provided for upper and lower probability inferences, based on a sample from an unknown finite population with a discrete univariate characteristic, and directed towards parameters of the population or properties of a future sample.

unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Upper and lower probabilities						
Nonparametric inference.						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.