CONFIDENCE SETS FOR MULTIVARIATE MEDIANS[1]

By P. G. Hoel and E. M. Scheuer[2]

University of California, Los Angeles

0. Summary. This paper considers the problem of finding confidence sets of the parallelepiped type based on extreme order statistics for multivariate medians when no parametric assumptions are made. A partial characterization of a multivariate distribution which will minimize the probability of the specified parallelepiped covering the multivariate median is given. This characterization enables one to obtain a sharp lower bound for the probability of coverage, provided the number of medians does not exceed seven and under the assumption that the structure is independent of the sample size.

1. Introduction. There exists a considerable amount of literature on the problem of estimating means of multivariate distributions by means of confidence sets. Most of it is concerned with parametric models as such, or with parametric models that arise from asymptotic considerations. Furthermore, the confidence sets are often ellipsoids, but these are not the most useful kind for applications. Parallelepipeds are considerably more useful in most applications. Since medians are the natural substitutes for means in nonparametric problems, the problem considered here is that of finding confidence parallelepipeds for multivariate medians.

2. Formulation. Let $(x_1, \cdots, x_t)$ be a random variable having the unique median $(\nu_1, \cdots, \nu_t)$. Let $(x_{1j}, \cdots, x_{tj}), j = 1, \cdots, n$, denote the random variables corresponding to a random sample of size n. The ordered values of a sample will be denoted by

$$x_i(1) \leqq x_i(2) \leqq \cdots \leqq x_i(n), \qquad i = 1, \cdots, t.$$

Let the set R be defined by

$$R = \{(x_1, \cdots, x_t) \mid x_i(1) \leqq x_i \leqq x_i(n), i = 1, \cdots, t\}.$$

The problem then is to find a sharp lower bound for

$$\mathcal{P} = P\{(\nu_1, \cdots, \nu_t) \in R\}.$$

The resulting value of $\mathcal{P}$ will be the confidence coefficient that can be guaranteed for the confidence parallelepiped formed by the planes parallel to the coordinate planes which pass through the extreme sample points for each coordinate, regardless of the nature of the distribution of $(x_1, \cdots, x_t)$. A result by Dunn [1]

Received May 25, 1960; revised August 10, 1960.

[1] This research was supported by the Office of Naval Research.

[2] Now with Space Technology Laboratories, Inc.

for $t = 2$ shows that this lower bound is attained when the two variables are independent; however, this property does not hold for higher dimensions.

The method that will be employed here is based on Bonferroni inequalities, and shows that a distribution that minimizes $\mathcal{P}$ must possess a certain structure. In the derivation of these inequalities, the following notation will be needed.

Let E_i be the event that $x_i(1) > \nu_i$ if $i = 1, \cdots, t$, and the event that $x_{i-t}(n) < \nu_{i-t}$ if $i = t + 1, \cdots, 2t$. Then, from the definition of R and $\mathcal{P}$, it follows that

$$1 - \mathcal{P} = P\{(\nu_1, \cdots, \nu_t) \not\in R\} = P\left\{\bigcup_1^{2t} E_i\right\}.$$

The last expression can be written [2], p. 89, in the form

$$P\left\{\bigcup_1^{2t} E_i\right\} = S_1 - S_2 + \cdots - S_{2t}, \tag{1}$$

where

$$S_1 = \sum_{i=1}^{2t} P\{E_i\}, \qquad S_2 = \sum_{i<j} P\{E_iE_j\}, \cdots$$

Because of the nature of the E_i, it follows that $S_{t+1} = \cdots = S_{2t} = 0$ here.

Probabilities such as $P\{E_iE_j\}$ depend only upon the probability mass assigned to each of the orthants determined by a set of coordinate axes through the median point $(\nu_1, \cdots, \nu_t)$. In this connection, let

$$q_{ij\cdots m} = P\{E_iE_j \cdots E_m\}$$

for $n = 1$. This quantity is defined only if all subscripts differ and provided that no two subscripts are equal mod t. The latter restriction is necessary because E_i and E_{i+t}, where the sum $i + t$ is taken mod $2t$, are incompatible events. Thus, $q_{ij\cdots m}$ yields the probability mass for the region determined by the proper positive or negative coordinates for the variables $x_i - \nu_i, x_j - \nu_j, \cdots, x_m - \nu_m$. If a subscript exceeds t, then the corresponding coordinate is negative, otherwise it is positive.

3. The case of $t = 3$. The method of obtaining the desired inequalities for $\mathcal{P}$ is considerably simpler and neater when $t = 3$; therefore this case will be considered first. In view of the definition of E_i and $q_{ij\cdots m}$, it follows that

$$\begin{aligned} P\{E_i\} &= q_i^n = (\tfrac{1}{2})^n, & i &= 1, \cdots, 6 \\ P\{E_iE_j\} &= q_{ij}^n, & i, j &= 1, \cdots, 6 \\ P\{E_iE_jE_k\} &= q_{ijk}^n, & i, j, k &= 1, \cdots, 6. \end{aligned}$$

This notation applied to (1) will yield the expression

$$\mathcal{P} = 1 - S_1 + S_2 - S_3 = 1 - 6(\tfrac{1}{2})^n + \sum_{i<j} q_{ij}^n - \sum_{i<j<k} q_{ijk}^n. \tag{2}$$

Now it follows from the definition of $q_{ij\cdots m}$ that $q_{ij\cdots lm} + q_{ij\cdots l(m+t)} = q_{ij\cdots l}$.

Using this property and the fact that the q's are nonnegative, one can obtain the following inequality for S_3. In this derivation, sums are over all possible permutations of indices for which the q's are defined, unless specified otherwise.

$$S_3 = \tfrac{1}{6} \sum q_{ijk}^n \leqq \tfrac{1}{6} \sum_{k \leqq 3} [q_{ijk} + q_{ij(k+3)}]^n = \tfrac{1}{6} \sum q_{ij}^n = \tfrac{1}{3} \sum_{i<j} q_{ij}^n = \tfrac{1}{3} S_2 .$$

In view of the convexity of v^n for $v \geqq 0$, it follows that

$$v_1^n + v_2^n \geqq 2 \left(\frac{v_1 + v_2}{2} \right)^n .$$

This inequality together with the fact that $q_i = \frac{1}{2}$, may be used to derive the following inequality for S_2.

$$S_2 = \tfrac{1}{2} \sum q_{ij}^n = \tfrac{1}{2} \sum_{j \leqq 3} [q_{ij}^n + q_{i(j+3)}^n] \geqq \tfrac{1}{2} \sum_{j \leqq 3} 2[\tfrac{1}{2} q_j]^n = \tfrac{1}{2} \sum (\tfrac{1}{4})^n = 12(\tfrac{1}{4})^n.$$

If these two inequalities are applied to (2), one will obtain the inequality

$$\mathcal{P} \geqq 1 - 6(\tfrac{1}{2})^n + 8(\tfrac{1}{4})^n. \tag{3}$$

Consider a probability distribution with

$$q_{123} = \tfrac{1}{4}, \quad q_{126} = 0, \quad q_{135} = 0, \quad q_{156} = \tfrac{1}{4},$$
$$q_{345} = \tfrac{1}{4}, \quad q_{456} = 0, \quad q_{234} = 0, \quad q_{246} = \tfrac{1}{4}.$$

These values satisfy the restrictions $q_i = \frac{1}{2}$, $i = 1, \cdots, 6$. Further, it is easily seen that they yield the value $8(\frac{1}{4})^n$ for the sums on the right side of (2); hence the lower bound given by (3) can be attained. This completes the proof of the following theorem.

THEOREM 1. *$P\{x_i(1) < \nu_i < x_i(n), i = 1, 2, 3\} \geqq 1 - 6(\frac{1}{2})^n + 8(\frac{1}{4})^n$, and this lower bound is sharp.*

4. The general case. The method used in the preceding section does not seem to generalize to higher dimensions; therefore a different method of attack is introduced. It will now be necessary to assume that $t < 8$.

Let $q = \max_{i,j} q_{ij}$ and suppose there exists some q_{ijk} possessing the value q. Then, exploiting relations of the type

$$q_{ij} = q_{ijk} + q_{ij(k+t)} \quad \text{and} \quad q_i = q_{ij} + q_{i(j+t)} = \tfrac{1}{2},$$

one can easily prove the following lemma, where, as before, sums such as $k + t$ are taken mod $2t$.

LEMMA 1. *Let $q = \max_{i,j} q_{ij}$ and suppose $q_{abc} = q$, then*

$$q_{ij} = q_{(i+t)(j+t)} = q, \qquad i, j = a, b, c, \ i < j$$
$$q_{i(j+t)} = q_{(i+t)j} = \tfrac{1}{2} - q, \qquad i, j = a, b, c, \ i < j$$
$$q_{ab(c+t)} = q_{a(b+t)c} = q_{(a+t)bc} = 0$$
$$q_{a(b+t)(c+t)} = q_{(a+t)b(c+t)} = q_{(a+t)(b+t)c} = \tfrac{1}{2} - q$$
$$q_{(a+t)(b+t)(c+t)} = 2q - \tfrac{1}{2}.$$

These are the only restrictions on the q's with double and triple subscripts that result directly from the lemma assumptions.

Now consider the problem of counting the number of double and triple subscript q's that assume the maximum value q. The following lemma gives inequalities for those two numbers.

LEMMA 2. *Let there be M quantities q_{abc} that take on the value $q = \max_{i,j} q_{ij}$. Let there be M' quantities $q_{\alpha\beta}$ that arise from these M quantities q_{abc} that have the value q. Then*

(a) *if $q \neq \frac{1}{2}$, $M' > M$ provided $t < 8$*

(b) *if $q = \frac{1}{2}$, $M' > M$ provided $t < 5$.*

PROOF. To each $q_{abc} = q$, Lemma 1 shows that there will correspond six $q_{\alpha\beta}$'s that have the value q; consequently M' will certainly exceed M unless different q_{abc}'s having the value q possess a sufficient number of $q_{\alpha\beta}$'s in common. Since the objective here is to show that $M' > M$, it will suffice to give a proof for the least favorable situation in which M is as large as possible and M' is as small as possible for any fixed number, r, of distinct subscripts.

(a) Suppose $q \neq \frac{1}{2}$. Then one can form at most $\binom{r}{3}$ q_{abc}'s with the value q and, by Lemma 1, there will be $2\binom{r}{2}$ corresponding $q_{\alpha\beta}$'s with the value q. But $2\binom{r}{2} > \binom{r}{3}$ provided $r < 8$.

(b) Suppose $q = \frac{1}{2}$. Then, using the last conclusion of Lemma 1, one can form at most $2\binom{r}{3}$ q_{abc}'s with the value q, while again there will be $2\binom{r}{2}$ corresponding $q_{\alpha\beta}$'s with this value. But $2\binom{r}{2} > 2\binom{r}{3}$ provided $r < 5$.

The preceding two lemmas will be used in the proofs of the following two theorems.

THEOREM 2. *For $2 < t < 8$, a set of orthant probabilities that minimizes $\mathcal{P}$ for all sample sizes must be one for which all $q_{ij} = \frac{1}{4}$.*

PROOF. Consider a set of orthant probabilities for which the q_{ij} are not all equal in value. Let P^* and P denote the values of $\mathcal{P}$ for an orthant probability configuration for which the q_{ij} have the common value $\frac{1}{4}$, and one for which they do not have this common value, respectively. Then, by a Bonferroni inequality [2], p. 100,

$$P^* < 1 - 2t\left(\frac{1}{2}\right)^n + 4\binom{t}{2}\left(\frac{1}{4}\right)^n,$$

$$P > 1 - 2t\left(\frac{1}{2}\right)^n + M' q^n - \sum_{i<j<k} q_{ijk}^n,$$

where M' is the number of q_{ij} assuming the maximum value q. Then

$$P - P^* > M' q^n - \sum_{i<j<k} q_{ijk}^n - 4\binom{t}{2}\left(\frac{1}{4}\right)^n.$$

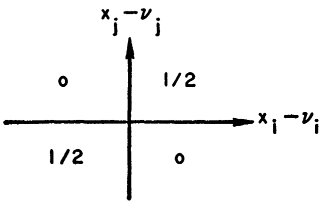

FIG. 1

Under the assumption that the q_{ij} are not all equal in value, it follows that $q > \frac{1}{4}$ here. Now if max $q_{ijk} < q$, the term $M'q^n$ will dominate the right side of this inequality as $n \to \infty$; consequently $P > P^*$ for sufficiently large n. This shows that no probability configuration in which the q_{ij} are not equal can possibly minimize $\mathcal{P}$ for all sample sizes, provided max $q_{ijk} < q$.

If max $q_{ijk} = q \neq \frac{1}{2}$ and $t < 8$, then

$$P - P^* > M' q^n - Mq^n - o(q^n) - 4\binom{t}{2}\left(\frac{1}{4}\right)^n$$
$$= (M' - M)q^n - o(q^n) - 4\binom{t}{2}\left(\frac{1}{4}\right)^n.$$

From Lemma 2, $M' - M > 0$; consequently the term $(M' - M)q^n$ dominates the right side of this inequality, and therefore the same conclusion follows.

This same proof holds for $q = \frac{1}{2}$, provided that $t < 5$. To complete the proof of the theorem it is therefore necessary to show that $\mathcal{P}$ cannot be minimized for all sample sizes when $q = \frac{1}{2}$ and $t = 5, 6$, or 7.

It follows from Lemma 1 that $q_{i+t,j+t} = \frac{1}{2}$ if $q_{ij} = \frac{1}{2}$. In the two dimensional space of the variables x_i and x_j, the probability distribution is therefore as shown in Figure 1. This distribution implies that ν_i will be covered by the confidence parallelepiped if and only if ν_j is covered. If the probability of covering the median point is not zero for a configuration satisfying the preceding restriction shown in Figure 1, then this probability can be decreased by constructing a configuration for which the probability that $x_i - \nu_i$ will assume a given sign is independent of the remaining x's, without changing their probability distribution. This is accomplished by halving all orthant probabilities and shifting one half of each such probability mass to the corresponding orthant with the opposite sign of $x_i - \nu_i$. That the probability of coverage has been decreased follows from the fact that the conditional probability of ν_i being covered, given that the remaining ν's are covered, will no longer be equal to one and that this shift in probabilities does not affect the probability of the remaining ν's being covered. For sufficiently large n, the probability of covering the median point cannot be zero. Thus, $q = \frac{1}{2}$ can be excluded from consideration in characterizing distributions that minimize $\mathcal{P}$[3]. The next theorem places further restrictions on any minimizing configuration.

[3] We are indebted to John W. Pratt for suggesting this method of proof for disposing of $q = \frac{1}{2}$. It is much shorter and neater than our original proof.

THEOREM 3. *For $t < 8$, a set of orthant probabilities that minimizes $\mathcal{P}$ for all sample sizes must be one for which the maximum number of q_{ijk} assume the value $\frac{1}{4}$.*

PROOF. If $q_{ij} = \frac{1}{4}$ for all i, j, then it is easily shown that $q_{ijkl} < \frac{1}{4}$ for all i, j, k, l. Any q with more than four subscripts will of course also satisfy this inequality. From Theorem 2 it follows that for a minimizing set of probabilities,

$$\mathcal{P} = 1 - 2t\left(\frac{1}{2}\right)^n + 4\binom{t}{2}\left(\frac{1}{4}\right)^n - \sum_{i<j<k} q_{ijk}^n + o\left(\frac{1}{4}\right)^n.$$

Let λ denote the number of q_{ijk} having the value $\frac{1}{4}$. Then it is clear that

$$\mathcal{P} = 1 - 2t(\tfrac{1}{2})^n + 4\binom{t}{2}\left(\frac{1}{4}\right)^n - \lambda\left(\frac{1}{4}\right)^n + o\left(\frac{1}{4}\right)^n,$$

and that this quantity can be minimized for all n by a given distribution only if that distribution has maximum possible λ.

5. Numerical lower bounds. The preceding theorems, together with the following lemma, will suffice to yield a theorem on the magnitude of sharp lower bounds for $\mathcal{P}$.

LEMMA 3. *If $q_{ij} = \frac{1}{4}$ for all i, j and $q_{ijk} = \frac{1}{4}$ for some k, then*

$$q_{ijk(l+r)} = q_{i(j+t)(k+t)(l+r)} = q_{(i+t)j(k+t)(l+r)} = q_{(i+t)(j+t)k(l+r)} = \tfrac{1}{8}$$

and

$$q_{ij(k+t)(l+r)} = q_{i(j+t)k(l+r)} = q_{(i+t)jk(l+r)} = q_{(i+t)(j+t)(k+t)(l+r)} = 0$$

for $l \neq i, j, k$ and $r = 0$ and $r = t$.

PROOF. It suffices to consider orthant probabilities in the four dimensional space of the variables x_i, x_j, x_k, x_l as shown in Figure 2. The condition $q_{ijk} = \frac{1}{4}$ implies that $a_1 + a_9 = \frac{1}{4}$. Imposing the condition that each $q_{\alpha\beta} = \frac{1}{4}$ yields $a_1 = a_3 = a_6 = a_8 = a_9 = a_{11} = a_{14} = a_{16} = \frac{1}{8}$ and zero values for the remaining a's. This suffices to prove the lemma.

The desired lower bounds are now given by the following theorem.

THEOREM 4. *Under the assumption that there exists a set of orthant probabilities*

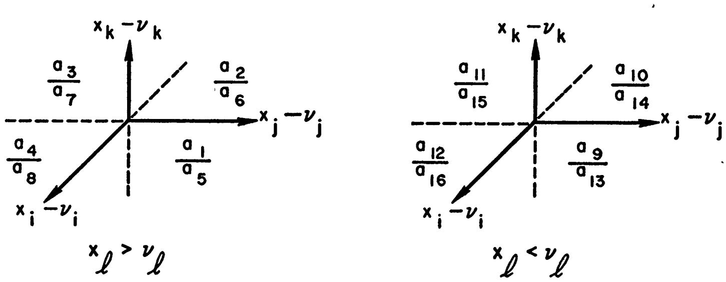

FIG. 2

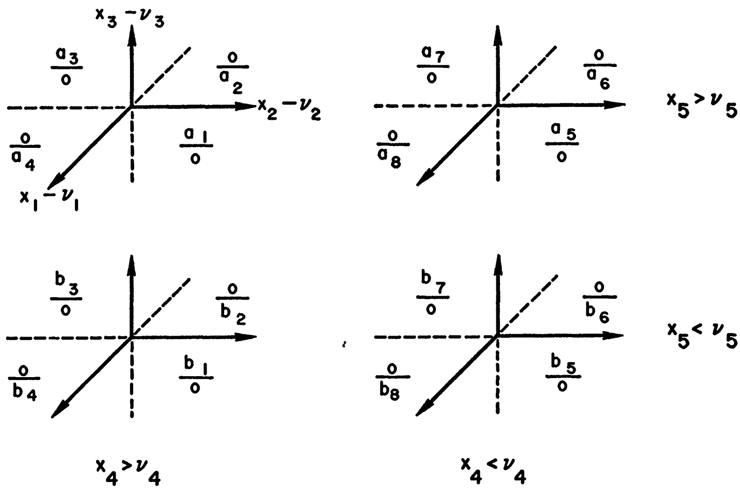

FIG. 3

that minimizes $\mathcal{P}$ *for all sample sizes, sharp lower bounds for* $\mathcal{P}$ *for* $2 < t < 8$ *are given by the formula*

$$\mathcal{P} \geqq 1 - 2t(\tfrac{1}{2})^n + 4(3t - 7)(\tfrac{1}{4})^n - 16(t - 3)(\tfrac{1}{8})^n.$$

PROOF. Theorem 1 has already given these bounds for $t = 3$; therefore consider $t > 3$. The method of proof is essentially the same for all values of t satisfying $3 < t < 8$; consequently only the proof for $t = 5$ will be given to illustrate the nature of the proofs.

Without loss of generality, suppose that q_{123} is a q_{ijk} that assumes the value $\frac{1}{4}$. By Lemma 3, it then follows, for example, that

$$q_{1234} = q_{1239} = q_{1235} = q_{12310} = \tfrac{1}{8}.$$

These values, together with the other values given by this lemma, suffice to yield the orthant configuration shown in Figure 3. The lemma conclusions require, for example, that

$$(4) \quad a_1 + b_1 = \tfrac{1}{8}, \qquad a_1 + a_5 = \tfrac{1}{8}, \qquad a_5 + b_5 = \tfrac{1}{8}, \qquad b_1 + b_5 = \tfrac{1}{8},$$

and therefore that $b_1 = a_5$ and $b_5 = a_1$.

Since a configuration is to be chosen that has the maximum number of q_{ijk} possessing the value $\frac{1}{4}$, suppose that $q_{124} = \frac{1}{4}$. Then $a_1 + b_1 = \frac{1}{4}$, but (4) shows that this choice is not possible. It follows readily that no q_{ijk} with two indices in common with q_{123} can be chosen. Suppose next that $q_{145} = \frac{1}{4}$. Then, by Lemma 3, $q_{1245} = q_{1457} = q_{1345} = q_{1458} = \frac{1}{8}$. The values of the a's and b's in Figure 3 are now completely determined and are given by $a_1 = a_4 = a_6 = a_7 = b_2 = b_3 = b_5 = b_8 = \frac{1}{8}$, and zero values for the remaining symbols.

If instead of assuming that q_{145} had the value $\frac{1}{4}$, one had chosen some other q_{ijk} with one index in common with q_{123}, then the same configuration would have been obtained except for a reflection in an axis. The value of $\mathcal{P}$ would, of course, be unaffected. If one now uses the configuration values just obtained to calculate the values of the respective terms in the Bonferroni expansion of $\mathcal{P}$, he will obtain the value

$$P = 1 - 10(\tfrac{1}{2})^n + 40(\tfrac{1}{4})^n - [8(\tfrac{1}{4})^n + 64(\tfrac{1}{8})^n + 8(0)^n] + [40(\tfrac{1}{8})^n + 40(0)^n] - 8(\tfrac{1}{8})^n.$$

Consequently, collecting terms, it follows that

$$\mathcal{P} \geqq 1 - 10(\tfrac{1}{2})^n + 32(\tfrac{1}{4})^n - 32(\tfrac{1}{8})^n$$

for $t = 5$, and that this result is sharp.

Similar, but considerably more tedious, methods will demonstrate the correctness of the formula given by the theorem for the larger values of t. The demonstration for $t = 4$ is of course the simplest one.

Although the formula of Theorem 4 has been demonstrated only for $2 < t < 8$, it is conjectured that the formula holds in general and that there always exists a configuration of orthant probabilities that minimizes $\mathcal{P}$ for all sample sizes.

6. Other order statistics. As n increases, the confidence coefficient will increase, but so will the size of the confidence parallelepiped. A smaller size parallelepiped, at the expense of a smaller confidence coefficient could be obtained for a symmetric distribution, for example, by taking means of consecutive pairs of samples. The median for the new variables will be the same as for the old variables. This averaging would tend to decrease the size of the parallelepiped as well as the confidence coefficient. If the sample were very large one could use means of more than two consecutive samples. For the general situation, in order to obtain useful confidence regions for various sample sizes, it would be necessary to find corresponding inequalities for other order statistics.

REFERENCES

[1] Olive Jean Dunn, "Estimation of the medians for dependent variables," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 192–197.

[2] William Feller, *An Introduction to Probability Theory and its Applications*, Vol. 1, 2nd Ed., John Wiley and Sons, New York, 1957.