MAX-MIN PROBABILITIES IN THE VOTING PARADOX[1]

By Zalman Usiskin[2]

University of Illinois

1. Introduction and summary. The voting paradox, that it is possible among three candidates to have A more popular than B, B more popular than C, and C more popular than A (e.g., let $\frac{1}{3}$ of a population prefer A to B, B to C; another $\frac{1}{3}$ prefer B to C, C to A; and the remaining $\frac{1}{3}$ prefer C to A, A to B) naturally raises the question of how much more popular they can be, and what results can be obtained with more than three candidates.

The question corresponds to the mathematical problem of choosing the joint distribution of n real-valued random variables so as to maximize

$$\min \{P(X_1 > X_2), \cdots, P(X_{n-1} > X_n), P(X_n > X_1)\}.$$

The fact that all these probabilities can exceed $\frac{1}{2}$ is well known, (see [1]), but the question of max-min does not appear to have been considered. This note considers this problem (a) with unrestricted $X_1, \cdots, X_n$ and (b) with $X_1, \cdots, X_n$ restricted to be independent.

In Case (a), it is very easy to show that the largest possible minimum is $(n - 1)/n$, easily achievable. In Case (b), which is more interesting, there is also an achievable largest minimum $b(n)$, which can be found by solving a degree $[\frac{1}{2}(n + 1)]$ equation, and has $b(n + 1) > b(n)$, $\lim_{n\to\infty} b(n) = \frac{3}{4}$, and $b(3) = .61803$, $b(4) = \frac{2}{3}, \cdots, b(10) = .73205$.

2. Arbitrary random variables $X_1, \cdots, X_n$. This section considers the problem of choosing the joint distribution of $X_1, \cdots, X_n$ so as to maximize

$$\min \{P(X_1 > X_2), \cdots, P(X_{n-1} > X_n), P(X_n > X_1)\}.$$

One such distribution has probability $1/n$ on each of the points $(n, n - 1, \cdots, 2, 1), \cdots, (2, 1, n, \cdots, 3), (1, n, \cdots, 3, 2)$, where the random variable X_i is a function defined on the ith coordinate of the n-space.

Theorem 1. $\max \min \{P(X_1 > X_2), \cdots, P(X_{n-1} > X_n), P(X_n > X_1)\} = (n - 1)/n$.

Proof. Writing A_i for $(X_i > X_{i+1})$, we have $\sum P(A_i') \geqq P(\sum A_i') = 1$, showing that $P(A_i')$ cannot all be less than $1/n$. Thus $P(A_i)$ cannot all exceed $(n - 1)/n$. It is easy to see (cf. the example above) that this upper limit is attainable. (The author is indebted to Mr. G. Haggstrom for this simple proof.)

Received 28 June 1963; revised 1 December, 1963.

[1] This research was supported in part by the National Science Foundation under grant NSF-G21507.

[2] This work was done while the author was an undergraduate at the University of Illinois. The author is currently doing graduate work at Harvard University.

To achieve this maximum, it is necessary and sufficient that $P(X_i > X_{i+1} > \cdots > X_n > X_1 > \cdots > X_{i-1}) = 1/n$, $i = 1, 2, \cdots, n$, where $X_{n+1} = X_1$.

3. Independent random variables $X_1, \cdots, X_n$. Find n real-valued independent random variables $X_1, X_2, \cdots, X_n$ with distribution functions $F_1, F_2, \cdots, F_n$ so as to maximize

$$\min \{P(X_1 > X_2), \cdots, P(X_{n-1} > X_n), P(X_n > X_1)\}.$$

LEMMA 1. *Attention can be restricted to step functions $F_1, F_2, \cdots, F_n$ with finitely many jumps. In particular, if there is an achievable maximum over all finite discrete F's, this is a maximum over all F's.*

PROOF. The E_i's are approximable by step functions, and $\sup \min \{P(X_i > X_{i+1})\} = \sup \min \{\int_{-\infty}^{\infty} F_{i+1}(x)\, dF_i(x)\}$.

LEMMA 2. *Attention can be further restricted to step functions $F_1, \cdots, F_n$ with finitely many jumps, no two of which have a jump at the same point.*

PROOF. If $X_{i'}$ and $X_{j'}$, $j' \neq i'$, both have positive probability on a point a, then choose ϵ so small that $P(a < X \leqq a + \epsilon) = 0$ for all i.

Shift the probability for X from a to $a + \epsilon$. Then $P(X_{j'} > X_{i'})$ is increased and no others are decreased.

LEMMA 3. *Writing x_i for values of X_i with positive probability, we may restrict attention to distributions where the order of these values from smallest to largest is*

$$x_i, x_{i-1}, \cdots, x_1; x_n, \cdots, x_2, x_1; \cdots; x_n, \cdots, x_2, x_1.$$

PROOF. The order of values as written above can always be achieved if we let some x_i stand for values of X_i with zero probability.

If there are some x_i standing for values of X_i with zero probability, then divide the array into groups, beginning a new group after each x_1, and deleting all x_i with zero probability. If a particular x_i is deleted, then we can shift the probabilities on $x_{i+1}, \cdots, x_n$ upward to the next group if there is one, and the probabilities on $x_1, \cdots, x_{i-1}$ downward to the next group if there is one, without decreasing any $P(X_i > X_{i+1})$. All groups with any deletions can be removed by this procedure. If x_1 is not at the extreme right, then a renaming of the X_i gives the desired array.

LEMMA 4. *Attention can be restricted to distributions in which the ordered possible values are $x_n, \cdots, x_1, x_n, \cdots, x_2$ with respective probabilities $p_n, \cdots, p_2, 1, (1 - p_n), \cdots, (1 - p_2)$.*

PROOF. It is enough to show that in distributions with ordered possible values $\cdots x_n, \cdots, x_1; x_n, \cdots, x_1; x_n, \cdots, x_1$ and respective probabilities $\cdots r_n, \cdots, r_1; p_n, \cdots, p_1; (a_n - p_n), \cdots, (a_1 - p_1)$, where $0 < p_i < a_i \leqq 1$, these probabilities can be changed to $\cdots r_n, \cdots, r_1; q_n, \cdots, q_1; (a_n - q_n), \cdots, (a_1 - q_1)$, where $0 \leqq q_i \leqq a_i$, and some q_i is 0 or a_i, without decreasing any $P(X_i > X_{i+1})$.

In the change from p_i to q_i, $P(X_i > X_{i+1})$ is increased by $p_i(a_{i+1} - p_{i+1}) - q_i(a_{i+1} - q_{i+1})$, $1 \leqq i < n$, and $P(X_n > X_1)$ is increased by $q_1(a_n - q_n) - p_1(a_n - p_n)$.

Multiplying all the X probabilities by $c_i > 0$ multiplies $P(X_i > X_{i+1})$ by $c_i \cdot c_{i+1}$, $c_{n+1} = c_1$, and does not affect the sign of these increases.

We do not have to consider the probabilities $r_n, \cdots, r_1$ since they have no bearing on the change in $P(X_i > X_{i+1})$ caused by changing from p_i to q_i.

Letting $c_i = 1/a_i$ for all i, we need only consider the modification of distributions with values $p_n, \cdots, p_1, (1 - p_n), \cdots, (1 - p_1)$. In this case, $P(X_i > X_{i+1}) = 1 - p_i + p_i p_{i+1}$, for $i = 1, 2, \cdots, n - 1$, and $P(X_n > X_1) = p_1 - p_1 p_n$.

If none of these is to decrease as the p_i's are changed at rates p_i', we must have $-p_i' + p_i p_{i+1}' + p_{i+1} p_i' \geqq 0$, $p_1' - p_n p_1' - p_1 p_n' \geqq 0$; i.e.,

$$(1) \qquad p_{i+1}' \geqq [(1 - p_{i+1})/p_i]p_i', \qquad p_n' \leqq [(1 - p_n)/p_1]p_1'$$

requiring $p_n' \geqq [(1 - p_n)/p_1]\alpha p_1'$, where $\alpha = \prod_{i=2}^{n-1} [(1 - p_i)/p_i]$, $p_n' \leqq [(1 - p_n)/p_1]p_1'$.

If $\alpha \leqq 1$, conditions (1) can be satisfied with $p_i' > 0$ for all i, $p_1' = 1$, in which case $\alpha' < 0$, so α remains $\leqq 1$. The increases in p_i's can be continued until some p_i reaches 1.

If $\alpha \geqq 1$, conditions (1) can be satisfied with $p_i' < 0$ for all i, $p_n' = -1$, in which case $\alpha' > 0$, so α remains $\geqq 1$. The decreases in p_i's can be continued until some p_i reaches 0.

In either case, we have reduced to distributions with probability on one fewer point. This procedure can be similarly applied repeatedly to distributions with probability on $2n$ or more points, modifying to distributions with probability on $2n - 1$ points.

Lemma 5. *Attention can be further restricted to distributions for which the* $P(X_i > X_{i+1})$ *are all equal, since if these are not all equal the probabilities can be changed to increase* $\min_{i=1,\cdots,n} P(X_i > X_{i+1})$.

Proof. For ordered values $x_n, \cdots, x_2, x_1, x_n, \cdots, x_2$ with respective probabilities $p_n, \cdots, p_2, 1, (1 - p_n), \cdots, (1 - p_2)$, we have $P(X_1 > X_2) = p_2$, $P(X_i > X_{i+1}) = 1 - p_i + p_i p_{i+1}$, for $i = 2, \cdots, n - 1$ and $P(X_n > X_1) = 1 - p_n$. As the p_i's are changed at rates p_i', these probabilities change at rates p_2', $p_i p_{i+1}' - (1 - p_{i+1})p_i'$ and $-p_n'$.

If $P(X_i > X_{i+1}) > \min_{j=1,\cdots,n} P(X_j > X_{j+1})$, for some i among $1, 2, \cdots, n$, take $p_2' > 0, p_3' > 0, \cdots, p_i' > 0, p_{i+1}' < 0, \cdots, p_n' < 0$, and with $p_{j+1}' > [(1 - p_{j+1})/p_j]p_j'$. (If $i = 1$, there are no positive p_j''s; if $i = n$, there are no negative p_j''s.) This will decrease $P(X_i > X_{i+1})$, increase all the others, and the minimum will be increased.

Theorem 2. $b(n + 1) > b(n)$, *where* $b(n) = \max\min\{P(X_1 > X_2), \cdots, P(X_{n-1} > X_n), P(X_n > X_1)\}$.

Proof. The existence of an achievable maximum follows from Lemma 4, since the p_i's there form a closed bounded set, the probabilities $P(X_i > X_{i+1})$ are continuous functions of the p_i's, and no increase is possible by taking distributions with more points of positive probability.

Let $p_2, \cdots, p_n$ be values achieving the maximum $b(n)$ in Lemma 4. Add another random variable X_{n+1}, changing the ordered probabilities to

$$p_n, \cdots, p_2, 1, 1, (1 - p_n), \cdots, (1 - p_2).$$

Now $P(X_i > X_{i+1}) = b(n)$ for $i = 1, 2, \cdots, n$, but $P(X_{n+1} > X_1) = 1$.

By Lemma 5, all $P(X_i > X_{i+1}) = P(X_{n+1} > X_1)$ for the values achieving the max-min with $n + 1$ random variables. So the values above do not yield the maximum. By the same Lemma, we can increase $P(X_i > X_{i+1}) = b(n)$ and decrease $P(X_n > X_1) = 1$. This yields distributions of $n + 1$ random variables with $P(X_1 > X_2), P(X_2 > X_3), \cdots, P(X_n > X_{n+1}), P(X_{n+1} > X_1)$ all $> b(n)$. Thus $b(n + 1) > b(n)$.

THEOREM 3. $\lim_{n\to\infty} b(n) = \frac{3}{4}$.

PROOF.

(1) $\lim_{n\to\infty} b(n) \leqq \frac{3}{4}$. Suppose $b(n) > \frac{3}{4}$. Now $p_2 = b(n)$. If $p_i > \frac{1}{2}$ and $b(n) > \frac{3}{4}$, since $p_{i+1} = 1 - \{[1 - b(n)]/p_i\}$, $p_{i+1} \geqq \frac{1}{2}$. So $p_3, p_4, \cdots, p_n$ are all $> \frac{1}{2}$, by induction. But $p_n = 1 - b(n) < \frac{1}{4}$. This contradiction shows that $\lim_{n\to\infty} b(n) \leqq \frac{3}{4}$.

(2) $\lim b(n) \geqq \frac{3}{4}$ by the following example:

Consider n random variables $X_1, \cdots, X_n$, $n > 2$,

$$P(X_1 = 0) = 1$$

with

$$P(X_i = 1 - i) = (n - i + 1)/n,$$

$$P(X_i = n + 1 - i) = (i - 1)/n; \quad i = 2, \cdots, n.$$

For these distributions (not the best possible), $P(X_i > X_{i+1}) = i/n + [(n - i)/n][(n - i + 1)/n]$, $i = 0, 1, \cdots, n - 1$. The minimum $P(X_i > X_{i+1})$ occurs at $i = \frac{1}{2}(n - 1)$ for n odd, and at $i = n/2$ or $\frac{1}{2}(n - 2)$ for n even.

For n odd, $P(X_{\frac{1}{2}(i+1)} > X_{\frac{1}{2}(i+3)}) = (3n^2 - 2n + 1)/4n^2$, the minimum for these distributions.

For n even, $P(X_{\frac{1}{2}(i+2)} > X_{\frac{1}{2}(i+4)}) = (3n^2 - 2n)/4n^2$, the minimum for these distributions.

Thus $\lim \min_{n\to\infty} \{P(X_i > X_{i+1})\} = \frac{3}{4}$ for n odd or even, and this limit is not greater than the limit of the best group of n random variables.

THEOREM 4. *There is a unique set of probabilities $p_1, \cdots, p_n$ achieving $b(n)$, and these have*

(i) $p_{i+2} = 1 - p_{n-i}$ for $i = 0, 1, \cdots, n - 2$.

(ii) $\lim_{n\to\infty} p_i = (1 + i)/2i$ for $i = 2, 3, \cdots$.

The value of $b(n)$ can be found by solving a polynomial equation of degree $[\frac{1}{2}(n + 1)]$.

PROOF. From the recursion relation of the proof of Theorem 3, each p_i is determined uniquely in terms of $b(n)$.

(i) $p_{i+2} = 1 - p_{n-i}$ by induction.

For $i = 0$, $P(X_1 > X_2) = P(X_n > X_1)$ implies $p_2 = 1 - p_n$.

TABLE 1

n	$g_n(b)$	Lower Approx.	b_n	Upper Approx.
3	$b^2 + b - 1$	7/16 = .4375	$(5^{\frac{1}{2}} - 1)/2$ = .61803···	5/8 = .6250
4	$3b^2 - 2b$	5/8 = .6250	2/3 = .66666···	27/40 = .6750
5	$b^3 + 3b^2 - 4b + 1$	95/144 = .6597	.6944···	7/10 = .7000
6	$4b^3 - 2b^2 - 2b + 1$	11/16 = .6875	$2^{\frac{1}{2}}/2$ = .70710···	5/7 = .7143
7	$b^4 + 6b^3 - 9b^2 + 3b$	279/400 = .6975		81/112 = .7234
8	$5b^4 - 9b^2 + 6b - 1$	17/24 = .7083		35/48 = .7292
9	$b^5 + 10b^4 - 15b^3 + 3b^2 + 3b - 1$	559/784 = .7130		11/15 = .7333
10	$6b^5 + 5b^4 - 24b^3 + 18b^2 - 4b$	23/32 = .7188	$3^{\frac{1}{2}} - 1$ = .73205···	81/110 = .7364
20		53/72 = .7361		209/280 = .7467
25		6255/8464 = .7390		243/325 = .7477
30		143/192 = .7448		637/850 = .7494

Suppose true for $i = k$. $p_{k+2} = 1 - p_{n-k}$. From the recursion relation, $p_{k+1+2} = 1 - [1 - b(n)]/p_{k+2} = 1 - [1 - b(n)]/(1 - p_{n-k})$. From the same relation reversed, $p_{n-k-1} = [b(n) - 1]/(p_{n-k} - 1)$. Thus $p_{k+1+2} = 1 - p_{n-k-1}$ and the proposition is true for $i = k + 1$.

(ii) $\lim_{n\to\infty} p_2 = \lim_{n\to\infty} b(n) = \frac{3}{4}$, so $\lim_{n\to\infty} p_i = (1 + i)/2i$ is true for $i = 2$.

Suppose true for $i = k$: i.e., $\lim_{n\to\infty} p_k = (k + 1)/2k$.

If true for $i = k$, $\lim_{n\to\infty} p_{k+1} = \lim \{1 - [1 - b(n)]/p_k\} = (k + 2)/(2k + 2)$, so true for $i = k + 1$. Therefore $\lim_{n\to\infty} p_i = (i + 1)/2i$ by induction.

Determination of $b(n)$. From the recursion relation, $p_n = 1 - [1 - b(n)]/p_{n-1}$. But $p_n = 1 - b(n)$. Thus $b(n) = [1 - b(n)]/p_{n-1}$.

If we let $f_n(b) = b(n) - [1 - b(n)]/p_{n-1}$, and write b for $b(n)$ when convenient, then the solution of $f_n(b) = 0$ between $\frac{1}{2}$ and $\frac{3}{4}$ is the value of $b(n)$.

From the recursive relation for the p_i,

$$f_n(b) = b - \cfrac{1 - b_3}{1 - \cfrac{1 - b_4}{1 - \cfrac{1 - b_5}{1 - \cdot \atop \cdot \atop \cdot \; \cfrac{}{1 - \cfrac{1 - b_n}{b}}}}}$$

where $b_3 = b_4 = \cdots = b_n = b$, indexed to show the number of steps in the continued fraction. Thus $f_n(b) = b - (1 - b)/[1 - b + f_{n-1}(b)]$.

Letting $f_i(b) = g_i(b)/h_i(b)$, we obtain

$$f_n(b) = \frac{bg_{n-1}(b) - (1 - b)^2 h_{n-1}(b)}{g_{n-1}(b) - (1 - b)h_{n-1}(b)} = \frac{g_n(b)}{h_n(b)}.$$

Since the solutions of $f_n(b) = 0$ are identical with those of $g_n(b) = 0$, a recursive relation for $g_n(b)$ will suffice.

From the system of equations, $g_n(b) = bg_{n-1}(b) - (1 - b)^2 h_{n-1}(b)$, and $(1 - b)g_{n-1}(b) = (1 - b)g_{n-2}(b) - (1 - b)^2 h_{n-1}(b)$. So $g_n(b) - g_{n-1}(b) = -g_{n-2}(b) + bg_{n-2}(b)$, i.e., $g_n(b) = g_{n-1}(b) - (1 - b)g_{n-2}(b)$.

Determination of $g_n(b)$ for $n = 3$ and $n = 4$ makes it possible to determine $g_n(b)$ for any n, and gives a means for determining $b(n)$.

$$n = 3: \quad f_3(b) = b - (1 - b)/b = (b^2 + b - 1)/b.$$

$$\text{i. e., } g_3(b) = b^2 + b - 1.$$

$$n = 4: \quad f_4(b) = b - \frac{1 - b}{1 - (1 - b)/b} = (3b^2 - 2b)/(2b - 1).$$

$$\text{i. e., } g_4(b) = 3b^2 - 2b.$$

From the recursive relation, $\deg [g_i(b)] = \deg [g_{i-1}(b)] + 1$ if and only if $\deg [g_{i-1}(b)] = \deg [g_{i-2}(b)]$. The latter is true only when i is odd. Thus the degree of $g_n(b)$ increases by one only when n is odd, otherwise remaining the same. By simple induction we see that $\deg [g_n(b)] = [\frac{1}{2}(n + 1)]$.

4. Approximations to $b(n)$. Rough lower and upper bounds were constructed for $b(n)$ in Theorem 3. They are

$$\text{for } n \text{ odd:} \quad (3n^2 - 2n + 1)/4n^2 \leqq b(n) < \tfrac{3}{4}$$

$$\text{for } n \text{ even:} \quad (3n - 2)/4n \leqq b(n) < \tfrac{3}{4}.$$

A better lower bound has been constructed and a better upper bound is conjectured; namely,

$$\text{for } n \text{ odd:} \quad \tfrac{3}{4} - 1/4(n - 2) - 1/16(n - 2)^2 < b(n) < \tfrac{3}{4}[1 - 2/n(n + 1)]$$

$$\text{for } n \text{ even:} \quad \tfrac{3}{4} - 1/4(n - 2) < b(n) < \tfrac{3}{4}[1 - 2/n(n + 1)].$$

5. Application. Consider n types of steel, numbered 1 to n, the bars of each type having a predetermined distribution of strength along their lengths. Randomly choose a sample bar of each type. Measure relative strength of two bars by pressing one against the other at a random place along the length of each bar—whichever breaks first is weaker. It is quite possible that each randomly chosen bar is stronger than the next and the last is stronger than the first. Let $p_i = P(i\text{th bar is stronger than the } i + 1\text{st})$. Then $b(n) = \max \min_{i=1,\cdots,n} \{p_i\}$.

6. Acknowledgment. My thanks go to Professor Colin Blyth for his suggestion of the problem and for his generous help and guidance throughout its solution.

REFERENCE

[1] BLACK, DUNCAN (1958). *The Theory of Committees and Elections.* Cambridge Univ. Press.