NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.
USEFUL PROPERTIES OF THE
MAJORITY BINOMIAL PROBABILITY FUNCTION

Bobby L. Buchanan

November 1961
Requests for additional copies by Agencies of the Department of Defense, their contractors, and other government agencies should be directed to the:

Armed Services Technical Information Agency
Arlington Hall Station
Arlington 12, Virginia

Department of Defense contractors must be established for ASTIA services, or have their 'need-to-know' certified by the cognizant military agency of their project or contract.

All other persons and organizations should apply to the:

U. S. DEPARTMENT OF COMMERCE
OFFICE OF TECHNICAL SERVICES,
WASHINGTON 25, D. C.
USEFUL PROPERTIES OF THE
MAJORITY BINOMIAL PROBABILITY FUNCTION

Bobby L. Buchanan

Project 4608

November 1961

ELECTRONIC MATERIAL SCIENCES LABORATORY
ELECTRONICS RESEARCH DIRECTORATE
AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
BEDFORD MASSACHUSETTS
ABSTRACT

Some useful properties of the majority binomial probability function

\[ f(p,n) = \sum_{x = \frac{n+1}{2}}^{n} \binom{n}{x} p^x (1-p)^{n-x} \]

are derived. An important result is the determination of a function which bounds \( f(p,n) \) and yet is easy to evaluate numerically. Thus, a means is provided of determining the bounding values of the majority binomial probability function for probability values much smaller than those listed in available tables.
CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Definitions and Assumptions</td>
<td>2</td>
</tr>
<tr>
<td>Theorems</td>
<td>3</td>
</tr>
<tr>
<td>Graphical Evaluation of ( n_0 )</td>
<td>11</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>13</td>
</tr>
</tbody>
</table>
USEFUL PROPERTIES OF THE
MAJORITY BINOMIAL PROBABILITY FUNCTION

INTRODUCTION

During the writer's research on the synthesis of reliable automata from unreliable constituents, the majority binomial probability function, 

\[ f(p,n) = \sum_{x = \frac{n+1}{2}}^{n} \binom{n}{x} p^x (1-p)^{n-x}, \quad n = 1, 3, 5, \ldots, \quad 0 < p < 1, \]  

arose. This function is a generalization of the probability function corresponding to \( n = 3 \), utilized by Von Neumann and others in conjunction with majority organs. The function, \( f(p,n) \), is a special case of the well known cumulative binomial probability distribution function for which at least two different volumes of tables have been published. In these tables the smallest entry value of the function is \( 10^{-7} \) and the smallest entry value of \( p \) is \( 10^{-2} \). The tables are not adequate since values of \( p \) and \( f(p,n) \) much smaller than \( 10^{-2} \) and \( 10^{-7} \) are of interest in automata theory and in certain other applications. Thus a problem exists of evaluating the function without a great amount of calculative labor. Approximating \( f(p,n) \) with a continuous function (such as the error function) does not relieve the problem, since, even if the approximating conditions are satisfied, the tabulated values in the tables of the approximating continuous functions do not cover the range of interest.
An explicit solution of (1) for \( n \) in terms of \( f \) and \( p \) is desirable, since for most problems involving the majority binomial, the value of \( p \) is given and the problem is to determine the value of \( n \) required to obtain a certain value of \( f \). However, even if \( n \) were a continuous variable, it is evident that it would not be possible to obtain an exact explicit solution for \( n \).

The purposes of this paper are to derive for ready reference useful properties of the majority binomial function and to find functions which bound \( f(p,n) \) that are easy to evaluate numerically and from which a useful approximation to the explicit solution for \( n \) can be obtained.

**DEFINITIONS AND ASSUMPTIONS**

The well known binomial probability distribution is defined as

\[
E(n,r,p) = \sum_{x=r}^{n} \binom{n}{x} p^x (1-p)^{n-x}, \quad 0 \leq p \leq 1, \quad n = 1,2,3,... \quad (2)
\]

where

\[
\binom{n}{x} = \frac{n!}{x!(n-x)!}
\]

This function gives the probability of \( r \) or more occurrences of an event in \( n \) independent trials where \( p \) is the probability of occurrence of the event in one trial.

The following relationships are not difficult to probe, but are assumed to be true since they appear in the literature.\(^1,2\)

\[
E(n+1,r,p) = pE(n,r-1,p) + (1-p)E(n,r,p) \quad (3)
\]

\[
E(n,r,p) = 1 - E(n,n-r+1,1-p) \quad (4)
\]

\[
\frac{dE}{dp} = r \binom{n}{r} p^{r-1}(1-p)^{n-r}
\]
The majority binomial probability function is defined by (1) as

\[ f(p,n) = \sum_{x=\frac{n+1}{2}}^{n} \binom{n}{x} p^x (1-p)^{n-x} \quad 0 \leq p \leq 1, \quad n = 1, 3, 5, \ldots \]

which is just \( E(n,r,p) \) with the constraints that \( r = \frac{n+1}{2} \) and \( n \) is odd.

It follows directly from the definitions and assumptions that

\[ f(p,n) = 1 - f(1-p,n), \quad f(p,n) = \frac{1}{2} \quad (6) \]

and that \( f(p,n) \) increases if \( p \) increases.

**Theorems**

**Theorem I:** \( f(p,n+2) = f(p,n) \)

\[ \sum_{r=1}^{n} \left[ \binom{n}{r} \frac{E(n,r)}{p^r} \right] \quad (3) \]

**Proof:** By (3)

\[ E(n+1,r,p) = pE(n,r-1,p) + (1-p)E(n,r,p) \]

From this recurrence relation it follows that

\[ E(n+1,r+1,p) = pE(n,r,p) + (1-p)E(n,r+1,p) \quad (7) \]

and

\[ E(n+2,r+1,p) = pE(n+1,r,p) + (1-p)E(n+1,r+1,p) \quad (8) \]

Substituting (7) and (8) into (8) we have after some manipulation

\[ E(n+2,r+1,p) = p^2E(n,r-1,p) + 2p(1-p)E(n,r,p) + (1-p)^2E(n,r+1,p) \quad (9) \]

Note that

\[ E(n,r-1,p) = (\frac{n}{r-1})p^{r-1}(1-p)^{n-(r-1)} + E(n,r,p) \]

and that

\[ E(n,r+1,p) = E(n,r,p) - \binom{n}{r}p^r(1-p)^{n-r} \]

Substituting into (9) and simplifying, we have

\[ E(n+2,r+1,p) = E(n,r,p) + \binom{n}{r-1}p^{r-1}(1-p)^{n-r+1} - (1-p)^2\binom{n}{r}p^r(1-p)^{n-r} \]

Note that \( \binom{n}{r-1} = \frac{r}{n-r+1}\binom{n}{r} \). Substituting again, we have
\[ E(n+2,r+1,p) = E(n,r,p) + \binom{n}{p} p^{r-p} (1-p)^{n-r} \left[ \frac{rp(1-p)}{n-r+1} - (1-p)^2 \right] \]

when \( r = r_1 = \frac{n+1}{2} \) and \( n \) is odd, \( \frac{r}{n-r+1} = 1 \) and by definition

\[ E(n+2,r_1+1,p) = f(p,n+2) = f(p,n) + \left( \frac{n+1}{2} \right) [p(1-p)]^2 \left[ 2p-1 \right]. \]

Theorem II: \( f(p,n) \) increases if \( n \) increases for any given value of \( p \) in the interval \( \frac{1}{2} < p < 1 \), and \( f(p,n) \) decreases if \( n \) increases for any given value of \( p \) in the interval \( 0 < p < \frac{1}{2} \).

Proof: Since \( n \) is an odd integer by definition, the smallest increment in any value of \( n \) is two. From theorem I we have

\[ f(p,n+2) = f(p,n) + \left( \frac{n+1}{2} \right) [p(1-p)]^2 \left[ 2p-1 \right] \quad n = 1,3,5,\ldots \]

therefore

\[ f(p,n+2) - f(p,n) = \left( \frac{n+1}{2} \right) [p(1-p)]^2 \left[ 2p-1 \right] \]

The sign of this difference is independent of \( n \). This difference is positive when \( (2p-1) > 0 \), which requires \( p > \frac{1}{2} \), and it is negative when \( (2p-1) < 0 \), which requires \( p < \frac{1}{2} \). It follows then, that

\[ f(p,n+2) > f(p,n) \text{ if } \frac{1}{2} < p < 1 \]
\[ f(p,n+2) < f(p,n) \text{ if } 0 < p < \frac{1}{2} \]

Since this is true for any odd \( n \), the theorem is proved.

Theorem III: \( f(p,n) > p \) if \( \frac{1}{2} < p < 1 \) and
\[ f(p,n) < p \text{ if } 0 < p < \frac{1}{2} \text{ for } n = 3,5,7,\ldots \]

Proof: By theorem II \( f(p,n) \) is an increasing function of \( n \) if
\( \frac{1}{2} < p < 1 \) and \( f(p,n) \) is a decreasing function of \( n \) if \( 0 < p < \frac{1}{2} \) and \( f(p,\frac{1}{2}) = \frac{1}{2} \). Therefore, the theorem is proved if \( f(p,3) > p \) for 
\( \frac{1}{2} < p < 1 \) and \( f(p,3) < p \) for \( 0 < p < \frac{1}{2} \).

By definition

\[
f(p,3) = \sum_{x=2}^{3} \left( \frac{3}{2} \right) p^x (1-p)^{3-x} = 3p^2 - 2p^3
\]

If \( f(p,3) = p \), then

\[
3p^2 - 2p^3 - p = 0
\]

which has solutions, \( p = 0 \), \( p = \frac{1}{3} \), \( p = 1 \).

\[
\frac{df(p,3)}{dp} = 6p(1-p), \text{ which is } > 1 \text{ at } p = \frac{1}{2}
\]

Since \( f(p,3) = p \) in the interval \( 0 < p < 1 \) only at \( p = \frac{1}{2} \) and the slope is > 1 at \( p = \frac{1}{2} \), \( f(p,3) > p \) for \( \frac{1}{2} < p < 1 \) and \( f(p,3) < p \) for \( 0 < p < \frac{1}{2} \), which proves the theorem.

Theorem IV: \( 3p - 2p^2 \leq R_n < 4p - 4p^2 \),

where \( R_n = \frac{f(p,n+2)}{f(p,n)} \)

Proof: By (10)

\[
R_n = 1 - \left( \frac{\pi}{p} \right)^r (1-p)^{n-r} (1-p)(1-2p)
\]

Define

\[
L(p,n) = \frac{(\pi)^r (1-p)^{n-r} (1-p)(1-2p)}{f(p,n)}
\]

when \( 0 < p < \frac{1}{2} \), \( 1-2p > 0 \)

Therefore \( L(p,n) > 0 \)

Dividing numerator and denominator of \( L(p,n) \) by \( \frac{\pi}{p} \) we have -
\[ L = \frac{(1-p)(1-2p)}{n} \sum_{x=r}^{n} a_x y^{x-r} \]

where \( a_x = \frac{x!(n-x)!}{x!n!} \)

and \( y = \frac{p}{1-p} \)

Define

\[ s = \sum_{x=r}^{n} a_x y^{x-r}, \text{ let } k = x - r, \text{ then} \]

\[ s = \sum_{k=0}^{n-r} a_r + ky^k \]

Note that \( a_r = 1 \) and \( \frac{a_r + k + 1}{a_r + k} = \frac{n - r - k}{r + k + 1} \)

By definition \( r = \frac{n+1}{2} \) therefore

\[ \frac{a_r + k + 1}{a_r + k} = \frac{n - 1 - 2k}{n + 2k + 3} < 1 \quad \text{if } k \geq 0 \]

Hence \( a_{r+k} \leq 1 \) if \( k = 0, 1, 2, \ldots \)

It is well-known that \( \frac{1}{1-y} = \sum_{k=0}^{\infty} y^k, \quad y^2 < 1 \)

Since \( a_{r+k} \leq 1 \) and \( \frac{1}{1-p} < 1 \) if \( 0 < p < \frac{1}{2} \), it is evident that

\[ s < \frac{1}{1-y} = \frac{1-p}{1-2p}. \text{ Therefore } L > \frac{(1-p)(1-2p)}{1-p} = (1-2p) \]

and \( R_n = 1 - L < 4p - 4p^2 \)

The minimum value of \( n \), by definition, is \( n = 1 \). If \( n = 1 \), then

\[ s = 1, \text{ and } s > 1 \text{ if } n > 1, \text{ therefore } s \geq 1. \]
Hence

\[ L = \frac{(1-p)(1-2p)}{8} \leq (1-p)(1-2p) \quad \text{and} \quad R_n = 1 - L \geq 3p - 2p^2 \]

which completes the proof of the theorem.

Theorem V:

\[ \frac{n-1}{2} \leq f(p,n) < \frac{n-1}{2} \quad 0 < p < \frac{1}{2} \]

Proof: By definition

\[ R_n = \frac{f(p,n+2)}{f(p,n)} \quad \text{and} \quad f(p,1) = p, \]

from which it follows that

\[ f(p,3) = R_1 f(p,1) = R_1 p \]
\[ f(p,5) = R_3 f(p,3) = R_3 R_1 p \]

\[ \ldots \ldots \]

\[ f(p,n) = \prod_{k=1}^{n-1} R_{n-2k} p, \quad n = 3,5,7,\ldots \]

By theorem IV

\[ 3p - 2p^2 \leq R_n < 4p - 4p^2, \quad n = 1,3,5,\ldots \]

\[ p < \frac{1}{2} \]

therefore

\[ 3p - 2p^2 \leq R_{n-2k} < 4p - 4p^2, \quad k = 1,2,\ldots,\frac{n-1}{2} \]

\[ n = 3,5,7,\ldots \]

and

\[ \prod_{k=1}^{n-1} R_{n-2k} < (4p - 4p^2)^{n-1} \]

\[ (3p - 2p^2)^{\frac{1}{2}} \leq \prod_{k=1}^{n-1} R_{n-2k} < (4p - 4p^2)^{\frac{n-1}{2}} \]
Hence
\[ p(3p-2p^2)^2 \leq f(p,n) < (4p-4p^2)^2. \]

Theorem VI: \((1-p)^2 \leq f(p,n) \leq \left( \frac{n+1}{2} \right)^{\frac{n+1}{2}}, p < \frac{1}{2}\)

Proof: Let \(r = \frac{n+1}{2}\), then

\[ f(p,n) = \sum_{x=r}^{n} \binom{n}{x} (1-p)^{n-x} = (1-p)^n \sum_{x=r}^{n} \binom{n}{x} \left( \frac{p}{1-p} \right)^x \]

Let \(k = x-r\), then

\[ f(p,n) = (1-p)^n \sum_{k=0}^{n-r} \binom{n}{n+k} \left( \frac{p}{1-p} \right)^{n+k} = (1-p)^n p^r \sum_{k=0}^{n-r} \binom{n}{n+k} \left( \frac{p}{1-p} \right)^k \]

where \(a_{r+k} = \frac{\binom{n}{r+k}}{\binom{n}{r}}\). It was shown in theorem IV that

\[ \sum_{k=0}^{n-r} a_{r+k} \left( \frac{p}{1-p} \right)^k < \frac{1-p}{1-2p}, \text{ therefore} \]

\[ f(p,n) < (1-p)^n p^r \left( \frac{1-p}{1-2p} \right) = \left( \frac{n+1}{2} \right)^{\frac{n+1}{2}} \frac{\left[ p(1-p) \right]^2}{1-2p} \]

Hence
\[ f(p,n) < \left( \frac{n+1}{2} \right)^{\frac{n+1}{2}} \frac{\left[ p(1-p) \right]^2}{1-2p}. \]

Since all the terms in the \(f(p,n)\) series are positive
\[ f(p,n) > \left( \frac{n}{n+1} \right)^{n+1} \left( \frac{2}{1-p} \right)^{n-1} \]

which completes the proof of the theorem.

Theorem VII: If \( f_u = \left( \frac{n}{n+1} \right) \frac{f_{p(1-p)}}{1-2p} \) is taken to be the value of \( f(p,n) \), the error, \( e = f_u - f \), is less than \( \frac{pf_u}{1-p} \) and the relative error, \( e_r = \frac{e}{f_u} \) is less than \( \frac{p}{1-p} \).

Proof: By theorem VI \( f_L \leq f \leq f_u \) where

\[ f_L = (1-p)^2 \left( \frac{n}{n+1} \right)^n \]

therefore

\[ e = f_u - f < f_u - f_L \]

and

\[ \frac{f_u}{f_u} = \frac{1-2p}{1-p} \quad \text{or} \quad f_u = \frac{1-2p}{1-p} f_L \]

Hence

\[ f_u - f_L = \frac{pf_u}{1-p} > e. \]

It follows directly that

\[ e_r = \frac{e}{f_u} < \frac{p}{1-p} \].

Theorem VIII: \( n_L \leq n_0 \leq n_u \), where

(1) \( n_0 \) is the minimum value of \( n = 1, 3, 5, \ldots \) for which \( f(p,n) \leq f_0 \) and \( f_0 \) is any prescribed probability, \( 0 < f_0 < p < \frac{1}{2} \).
(2) \( n_L \) is the smallest odd integer which exceeds or is equal to

\[
\frac{f_0}{2 \log \frac{p}{p}} + 1
\]

\( n_L = \frac{2 \log \frac{f_0}{p}}{\log(3p-2p^2)} + 1 \)  \hspace{1cm} (11)

(3) \( n_u \) is the smallest odd integer which exceeds or is equal to

\[
\frac{f_0}{2 \log \frac{p}{p}} + 1
\]

\( n_u = \frac{2 \log \frac{f_0}{p}}{\log(4p-4p^2)} + 1 \)  \hspace{1cm} (12)

Proof: By theorem V \( f(p,n) < p(4p-4p^2)^{n-1} \)

Let \( f_0 = p(4p-4p^2)^{n-1} \), and solve for \( n'_u \). Then

\[
\frac{f_0}{2 \log \frac{p}{p}} + 1
\]

\( n'_u = \frac{2 \log \frac{f_0}{p}}{\log(4p-4p^2)} + 1 \)

\( n_u \geq n'_u \) by definition, and \( 4p - 4p^2 < 1 \) if \( 0 < p < \frac{1}{2} \);

therefore \( f_0 \geq p(4p-4p^2)^{n-1} \) and \( f(p,n'_u) < p(4p-4p^2)^{n-1} \leq f_0 \), hence \( f(p,n'_u) < f_0 \).

This proves that \( n = n_u \) is sufficient to guarantee that \( f(p,n) \leq f_0 \), therefore \( n_0 \leq n_u \).

By theorem V

\[
f(p,n) \geq p(3p-2p^2)^{n-1}
\]
Let \( f_0 = p(3p-2p^2)^2 \) and solve for \( n'_L \) and we have

\[
2 \log \frac{f_0}{p} = \frac{n'_L - 1}{\log(3p-2p^2) + 1}
\]

It follows from the definitions of \( n_L \) and \( n'_L \) that \( n_L - 2 < n'_L \); therefore since \( 3p - 2p^2 < 1 \),

\[
f_0 < p(3p-2p^2)^2
\]

By theorem V

\[
f(p, n_{L-2}) \geq p(3p-2p^2)^2, \text{ hence } f(p, n_{L-2}) > f_0.
\]

By theorem II \( f(p, n) \) decreases as \( n \) increases; therefore if \( f(p, n_0) \leq f_0 \), \( n_0 \) must be greater than \( n_{L-2} \). However, the smallest increment in \( n_0 \) is 2; therefore a necessary condition for \( f(p, n_0) \leq f_0 \) is \( n_0 \geq n_L \), which completes the proof of the theorem.

**GRAPHICAL EVALUATION OF \( n_0 \)**

In order to simplify calculations, let \( f_0 = 10^{-B} \), or equivalently, \( B = -\log f_0 \). Substituting into (11) and (12) we have

\[
n'_L = \frac{-2B - 2 \log p + 1}{\log(3p-2p^2)} \tag{13}
\]

and

\[
n'_u = \frac{-2B - 2 \log p + 1}{\log(4p-4p^2)} \tag{14}
\]
where \( \log 2 < -\log p < B < \infty \).

For any given value of \( p \), \( n'_L \) and \( n'_u \) are functions of \( B \) only, in which case it is evident that (13) and (14) are equations of straight lines. The plot of (13) or (14) is then completely determined by evaluation at any two values of \( B \).

An efficient procedure for graphical evaluation of \( n_0 \), assuming a fixed value of \( p \), \( 0 < p < \frac{1}{2} \), is as follows:

Let \( n' \) be the axis of ordinates and \( B \) the axis of abscissas. Evaluate \( n'_u \) at any two convenient values of \( B \) and pass a straight line through the resulting two points, remembering that points which do not satisfy the condition, \( \log 2 < -\log p < B < \infty \), have no practical significance. Project values of the ordinates of the straight line, which are not odd integers up to the first value of the ordinate which is an odd integer. The resulting step function gives an upper bound of \( n_0 \) for any value of \( B \). Apply the same procedure to \( n'_L \) to obtain a graph of lower bounds. This procedure is illustrated in Fig. 1. It should be noted that for \( p \leq 10^{-3} \), the upper and lower bounds on \( n_0 \) coincide for most values of \( B < 25 \) and never differ by more than two. Therefore if \( n_u \) is taken to be the value of \( n \) and \( p \leq 10^{-3} \), \( f_0 \geq 10^{-25} \), then \( f(p,n) < f_0 \) and \( n_u \) is at most two greater than \( n_0 \). If greater economy in the determination of \( n_0 \) is desired, then for the values of \( B \) of interest for which \( n_u \) and \( n_L \) do not coincide, theorem VII can be used in most cases to determine if \( n_L \) would suffice.
REFERENCES


Fig. 1. Graphical evaluation of $n_0$
4. Probability, mathematical analysis

2. Probability, applications

I. Buchanan, B. L.

Some useful properties of the majority binomial probability function

\[ f(p,n) = \sum_{x = \text{int}(\frac{n}{2})}^{n} \binom{n}{x} p^x (1-p)^{n-x} \quad n = 1, 3, 5, \ldots, \]

are derived. An important result is the determination of a function which bounds \( f(p,n) \) and yet is easy to evaluate numerically. Thus, a means is provided of determining the bounding values of the majority binomial probability function for probability values much smaller than those listed in available tables.

---

UNCLASSIFIED

1. Probability, mathematical analysis

2. Probability, applications

I. Buchanan, B. L.