THE DISTRIBUTION OF A SUM OF BINOMIAL RANDOM VARIABLES

Ken Butler
Michael Stephens

TECHNICAL REPORT No. 467
APRIL 28, 1993

Prepared Under Contract
N00014-92-J-1264 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065

(12)

THE DISTRIBUTION OF A SUM OF BINOMIAL RANDOM VARIABLES

Ken Butler

Michael Stephens

TECHNICAL REPORT No. 467

APRIL 28, 1993

Prepared Under Contract

N00014-92-J-1264 (NR-042-267)

FOR THE OFFICE OF NAVAL RESEARCH

Professor Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA 94305-4065

The distribution of a sum of binomial random variables

Ken Butler Michael Stephens

Abstract

In this paper we examine the distribution of a sum S of binomial random variables, each with different success probabilities. The distribution arises in reliability analysis and in survival analysis. An algorithm is given to calculate the exact distribution of S, and several approximations are examined. An approximation based on a method of Kolmogorov, and another based on fitting a distribution from the Pearson family, can be recommended.

KEY WORDS: Kolmogorov-type approximation; Pearson distributions; reliability; survival analysis.

1 Introduction

The classical binomial distribution requires a set of independent trials in which the probability of success is constant. However, one may be faced with a situation in which the success probabilities differ. For example, in a common situation in reliability analysis (see Boland and Proschan, 1983) there may be n individual components, each of which has a different probability of functioning correctly, and the probability of observing k or more functioning components must be found in order to determine the overall reliability of the system. Similarly, with survival data, the individual subjects may suffer differing influences on their survival probabilities, but the quantity of interest may be an overall survival rate at a particular time.

Let $X_1, X_2, \ldots, X_r$ be binomially distributed random variables, with X_i having index n_i and probability p_i. In the type of problem discussed above, we are interested in the distribution of the sum

$S = \sum_{i=1}^{r} X_i$.

There exist several approximations for a *single* binomial distribution; these are described in Johnson and Kotz (1969), and some numerical comparisons are provided by Gebhardt (1969). One approximation, due to Bol'shev (1963), exhibits remarkable accuracy, but does not appear to generalize to the distribution of a *sum* S of binomial random variables. For S, Boland and Proschan (1983) give bounds for the cumulative probabilities, in terms of cumulative probabilities of other sums of binomial random variables which have the same mean as S.

In this paper, we provide a method for the exact calculation of the distribution of S, and we examine several approximations to the distribution. These include the well-known normal and Poisson approximations suitable for large n_i and/or small p_i. These have limited accuracy, and two other approximations are considered; one is based on Kolmogorov's refinement of the Poisson approximation to the binomial, while the other involves fitting a suitably chosen distribution from the (continuous) Pearson family.

We compare the above approximations for several sets of n_i and p_i. The Kolmogorov-type approximation performs well in all cases, while the Pearson-family approximation is particularly effective when S can take a large number of values. The normal and Poisson approximations perform poorly in comparison.

2 Exact calculation

With modern computing facilities, it is possible to calculate the exact distribution of S. The calculation depends on the observation that for any two discrete random variables Y and Z taking values $0, 1, 2, \ldots$,

$$P(Y + Z = j) = \sum_{i=0}^{j} P(Y = i) \cdot P(Z = j - i). \tag{1}$$

This suggests that we may calculate the distribution of S by finding the distribution of $X_1 + X_2$, and then adding the remaining X_i one at a time. Corresponding to (??), Y is the sum of those X_i currently added, while Z is the next X_i to be added.

The binomial distribution has a recurrence relation which is useful in the automatic calculation of its probabilities. If X denotes a binomial random variable with index n and probability p, then

$$\left.\begin{aligned} P(X = 0) &= (1 - p)^n \\ P(X = j) &= \{(n - j + 1)/j\} \cdot \{p/(1 - p)\} \cdot P(X = j - 1) \qquad \text{if } j \geq 1 \end{aligned}\right\} \tag{2}$$

Use of (??) requires less computation than would the evaluation of each probability directly.

We are thus led to the following algorithm for the direct calculation of the distribution of S:

1. Calculate the distributions of X_1 and X_2 using (??).

2. Let $Y = X_1$ and $Z = X_2$, and use (??) to find $P(Y + Z = j)$ for all j. Let $U = Y + Z$.

3. For $k = 3, 4, \ldots, r$:

   (a) Calculate the distribution of X_k using (??).

   (b) Let $Y = U$ and $Z = X_k$, and use (??) to find $P(Y + Z = j)$ for all j.

   (c) Let $U = Y + Z$.

4. The distribution of S is then that of the U calculated in step 3(c) when $k = r$.

It is of interest to count the number of arithmetic operations required by this algorithm. Suppose m is the maximum value of S for which $P(S = m)$ is non-trivial (for example, m might be such that $P(S = j) < 10^{-8}$ if $j > m$). Then we assume that each probability distribution is calculated for $j = 0, 1, \ldots, m$.

Formula (??) requires $j + 1$ multiplications and $j + 1$ additions for each j, and hence $\sum_{j=0}^{m}(j + 1) = \frac{1}{2}(m + 1)(m + 2)$ of each operation altogether.

Formula (??) requires an exponentiation to start, then three additions and three multiplications for each use of the recurrence relation, making $3m$ of each altogether.

In the algorithm, (??) is used $r - 1$ times, and (??) is used r times. Consequently, the algorithm requires $\frac{1}{2}(r - 1)(m + 1)(m + 2) + 3rm$ additions and multiplications, and r exponentiations. Since m is likely to be the largest of these quantities, we see that the operation count is of the order of m^2. We also note that storage will be required for three arrays of length m.

3 Kolmogorov-type approximations

Johnson and Kotz (1969) and Dunin-Barkovsky and Smirnov (1955) describe an approximation for a single binomial distribution, due to Kolmogorov, based on probabilities and (backward) differences of probabilities from the Poisson distribution. Multiples of the differences are taken so as to match the moments of the true and approximating distributions. The idea is quite general, allowing any discrete distribution to be approximated by any other more easily calculated distribution. The only requirement is that the moments ν_i of the true distribution can be found, at least up to some order r.

Let $p(i)$ denote the true $P(S = i)$, and let $p_k(i)$ denote the approximation to this probability based on differences up to order k. The probability $p_0(i)$ is then taken to be the approximation to $P(S = i)$ taken directly from the approximating distribution. The k-th backward difference at i is denoted $\nabla^{(k)}p_0(i)$ and is

$$\nabla^{(k)}p_0(i) = \nabla^{(k-1)}p_0(i) - \nabla^{(k-1)}p_0(i-1), \tag{3}$$

where $p_0(i) = 0$ for $i < 0$ and $\nabla^{(0)}p_0(i) = p_0(i)$.

The approximating distribution is improved by adding to it a linear combination of its backward differences, up to some order k:

$$p_k(i) = p_0(i) + \sum_{j=1}^{k} a_j \nabla^{(j)} p_0(i), \tag{4}$$

where the coefficients a_j are chosen to match the first k moments of the true and approximate distributions. The l-th moment of the k-th order approximation, about some value θ, is

$$\mu_{lk} = \mu_{l0} + \sum_{j=1}^{k} a_j e_{lj}, \tag{5}$$

where

$$e_{lj} = \sum_{i=0}^{\infty} (i-\theta)^l \nabla^{(j)} p_0(i).$$

Note that μ_{l0} denotes the l-th moment of the original approximating distribution $p_0(i)$.

After some algebra, we find that

$$e_{lj} = \sum_{i=j}^{l} c_{lji}\mu_{l-i,0}, \tag{6}$$

where $c_{lji} = \binom{l}{i}(-1)^j j!\, \mathcal{S}_i^{(j)}$, and $\mathcal{S}_i^{(j)}$ is the Stirling number of the second kind (see Abramowitz and Stegun, 1970, p. 824). As a result, the c_{lji} can be evaluated by using a table such as that of Abramowitz and Stegun (1970, p. 835). For convenience, we provide a table of the c_{lji} for $l \leq 6$ in Table 1. Note that $\mathcal{S}_i^{(j)} = 0$ for $i < j$, which reduces the range of summation required for e_{lj}, and gives $e_{lj} = 0$ if $l < j$.

Therefore μ_{lk} can be written

$$\mu_{lk} = \mu_{l0} + \sum_{j=1}^{k} a_j \sum_{i=j}^{l} c_{lji}\mu_{l-i,0} \tag{7}$$

In practice, it is convenient to match the moments of the true and approximating distributions sequentially, starting with the mean. Since $e_{lj} = 0$ if $l < j$, any multiple of the k-th differences can be added to $p_{k-1}(i)$ without changing the moments of order $1, 2, \ldots, k-1$. As a result, the coefficients $a_1, a_2, \ldots, a_{k-1}$ do not change; one simply chooses a_k to ensure that $\mu_{kk} = \nu_k$.

The algorithm for computing the Kolmogorov-type approximations follows:

l	j	i	c_{lji}
1	1	1	-1
2	1	1	-2
2	1	2	-1
2	2	2	2
3	1	1	-3
3	1	2	-3
3	1	3	-1
3	2	2	6
3	2	3	6
3	3	3	-6
4	1	1	-4
4	1	2	-6
4	1	3	-4
4	1	4	-1
4	2	2	12
4	2	3	24
4	2	4	14
4	3	3	-24
4	3	4	-36
4	4	4	24

l	j	i	c_{lji}
5	1	1	-5
5	1	2	-10
5	1	3	-10
5	1	4	-5
5	1	5	-1
5	2	2	20
5	2	3	60
5	2	4	70
5	2	5	30
5	3	3	-60
5	3	4	-180
5	3	5	-150
5	4	4	120
5	4	5	240
5	5	5	-120

l	j	i	c_{lji}
6	1	1	-6
6	1	2	-15
6	1	3	-20
6	1	4	-15
6	1	5	-6
6	1	6	-1
6	2	2	30
6	2	3	120
6	2	4	210
6	2	5	180
6	2	6	62
6	3	3	-120
6	3	4	-540
6	3	5	-900
6	3	6	-540
6	4	4	360
6	4	5	1440
6	4	6	1560
6	5	5	-720
6	5	6	-1800
6	6	6	720

Table 1: Coefficients c_{lji}

1. Calculate the probabilities of the initial approximating distribution $p_0(i)$ for all i.

2. For $k = 1, 2, \ldots$:

   (a) Find the k-th moment ν_k of the true distribution about some value θ (which could be, for example, 0 or the true mean ν).

   (b) Find the k-th differences $\nabla^{(k)} p_0(i)$ for all i using (??).

   (c) Calculate $\mu_{k,k-1}$ from (??). This is the k-th moment about θ of the current approximating distribution (which has probabilities $p_{k-1}(i)$).

   (d) Let $a_k = (-1)^k (\nu_k - \mu_{k,k-1})/k!$.

   (e) Then calculate $p_k(i) = p_{k-1}(i) + a_k \nabla^{(k)} p_0(i)$, for all i, which forms the "improved" approximation to the true distribution.

3. Continue until $|p_k(i) - p_{k-1}(i)|$ is sufficiently small for all i, indicating that further differences will not improve the approximation, or until the desired number of moments has been matched.

In implementing this algorithm, there are two sources of numerical instability. One involves the choice of θ; we recommend taking θ equal to the true mean ν in order to keep each moment as small as possible. A more serious problem involves the differences themselves: each difference involves the subtraction of neighbouring lower-order differences, which may well be approximately equal. If this is so, several significant digits will be lost. For this reason, we recommend calculating the differences with extra precision. In fact, even with double precision, this cancellation of significant digits may prevent improvement of accuracy beyond a certain point. The theoretical convergence of the approximating probabilities to the true probabilities may not, for this reason, carry over into computational convergence, although an approximation based on a large number of moments can still be extremely accurate.

4 Approximations using the Pearson family of distributions

An effective technique for approximating sums of continuous random variables is to find the first four moments of the sum, and then to fit a Pearson curve. In some circumstances this has been shown to give good results also even for discrete distributions (see Stephens, 1965). Suppose S^* is a continuous random variable with a distribution in the Pearson family, and suppose S^* and S have the same first four moments or cumulants. Then $P(S \leq s)$, where s is an integer, is approximated by $P(S^* < s + 0.5)$, and the latter can be calculated using programs to fit Pearson curves.

In order to try this approximation, we need the first four cumulants of S. For a binomial random variable with index n and probability p, the first four cumulants are $\kappa_1 = np$, $\kappa_2 = npq$, $\kappa_3 = npq(q-p)$ and $\kappa_4 = npq(1-6pq)$, where $q = 1-p$. The cumulants of S are therefore:

$$\begin{aligned}
\kappa_1 &= \textstyle\sum_{i=1}^{r} n_i p_i \\
\kappa_2 &= \textstyle\sum_{i=1}^{r} n_i p_i (1-p_i) \\
\kappa_3 &= \textstyle\sum_{i=1}^{r} n_i p_i (1-p_i)(1-2p_i) \\
\kappa_4 &= \textstyle\sum_{i=1}^{r} n_i p_i (1-p_i)\{1-6p_i(1-p_i)\}
\end{aligned}$$

¿From the cumulants, the values of the skewness parameter $\sqrt{\beta_1} = \kappa_3/\kappa_2^{3/2}$ and the kurtosis parameter $\beta_2 = \kappa_4/\kappa_2^2$ are found, and these are used to fit a Pearson curve (see, e. g., Solomon and Stephens, 1977, Stephens, 1992) with these parameters, and hence to find $P(S^* < s + 0.5)$ for the desired (integral) values of s.

5 Examples

We present several examples to illustrate the accuracy of the approximations. For comparison, we also compute approximations using the normal distribution (with continuity correction), matching the first two moments, and the Poisson distribution, matching the mean of S. The Kolmogorov-type approximations are started with a binomial with $n = \sum_{i=1}^{r} n_i$ and p chosen so that the means of S and the approximating distribution are equal. From work of Hoeffding (1956), referred to and extended by Barlow and Proschan (1983), this choice of initial approximation will produce cumulative probabilities $F_S(s)$ that are too large when $s < E(S)$ and too small when $s \geq E(S)$. The examples have small p_i, and so the distribution of S will have a long right-hand tail; we have concentrated on this long tail in assessing the approximations.

In the tables, we have shown the results of the Kolmogorov-type approximations obtained using four and six moments, denoted by K (4) and K (6) respectively. K (0) denotes the initial approximation used in each case. The last row of each table shows the maximum absolute error committed by each approximation over the values given in the table.

s	True $P(S \leq s)$	Approximate $P(S \leq s)$					
		K (0)	K (4)	K (6)	Pearson	Poisson	Normal
1	0.551513	0.552660	0.551284	0.551514	0.553052	0.557825	0.500000
2	0.813946	0.812895	0.814174	0.813945	0.808411	0.808847	0.801834
3	0.941627	0.940243	0.941594	0.941628	0.936561	0.934358	0.955093
4	0.985710	0.984951	0.985665	0.985709	0.983959	0.981424	0.994529
5	0.997203	0.996936	0.997203	0.997203	0.997033	0.995544	0.999654
6	0.999554	0.999486	0.999560	0.999554	0.999630	0.999074	0.999989
7	0.999941	0.999928	0.999943	0.999941	0.999973	0.999830	1.000000
	Max. error	0.001384	0.000229	0.000001	0.005535	0.007269	0.051513

Table 2: Exact and approximate probabilities for Example 1.

5.1 Example 1

Here, S is the sum of five binomials with small values of p. The values of n and p for each X_i are:

n_i	p_i
5	0.02
5	0.04
5	0.06
5	0.08
5	0.10

The mean of S is 1.5, and the variance is 1.39. As a result, we would expect the normal distribution to give a poor approximation. The initial approximation is here a binomial with $n = 25$ and $p = 0.06$.

The actual and approximate cumulative probabilities for the upper tail are shown in Table ??.

In this example, the Kolmogorov-type approximations are clearly superior, with the four-moment approximation giving almost four accurate decimals and the six-moment almost six. The Pearson family distribution gives a serviceable two or three accurate decimals, while the Poisson and normal fail to assume the correct form in the extreme tail. The Kolmogorov and Pearson procedures, based on matching at least four moments, capture the tail behaviour of S very well. The Pearson distribution performs very creditably, considering that it is a continuous distribution applied to a discrete distribution with a small number of values.

	True	Approximate $P(S \leq s)$					
s	$P(S \leq s)$	K (0)	K (4)	K (6)	Pearson	Poisson	Normal
275	0.516777	0.516451	0.516712	0.516772	0.516772	0.512027	0.515644
283	0.748050	0.740862	0.748010	0.748048	0.747995	0.695874	0.747548
291	0.901928	0.894073	0.901959	0.901931	0.901869	0.840129	0.902231
296	0.953696	0.947807	0.953738	0.953699	0.953654	0.902598	0.954160
300	0.976850	0.972808	0.976881	0.976851	0.976821	0.937940	0.977271
305	0.991358	0.989215	0.991368	0.991357	0.991343	0.967059	0.991636
311	0.997782	0.996982	0.997776	0.997781	0.997775	0.986133	0.997904
315	0.999197	0.998832	0.999189	0.999196	0.999193	0.992702	0.999256
320	0.999801	0.999682	0.999795	0.999801	0.999798	0.996963	0.999821
326	0.999969	0.999944	0.999966	0.999969	0.999966	0.999050	0.999973
	Max. error	0.007855	0.000065	0.000003	0.000059	0.061799	0.001133

Table 3: Exact and approximate probabilities for Example 2.

5.2 Example 2

Once again, S is the sum of five binomials, but with larger values of n_i and p_i:

n_i	p_i
50	0.1
100	0.2
150	0.3
200	0.4
250	0.5

With these larger values of n and p, we would expect the approximations based on continuous distributions (normal and Pearson family) to perform well. This time, we start the Kolmogorov-type approximations using a binomial with $n = 750$ and $p = 0.3$. The results for selected values of s are shown in Table ??.

For this example, we notice that the six-moment Kolmogorov approximation is the best, but the Pearson family approximation is just as good as the four-moment Kolmogorov approximation. The Poisson approximation is very poor, as we might expect from these larger values of p, while the normal approximation provides almost three-figure accuracy, which is not as good as the other approximations.

s	True $P(S \leq s)$	Approximate $P(S \leq s)$					
		K (0)	K (4)	K (6)	Pearson	Poisson	Normal
10	0.583047	0.583044	0.583047	0.583047	0.582846	0.583040	0.563529
12	0.793728	0.793482	0.793728	0.793728	0.792788	0.791556	0.788033
14	0.918908	0.918643	0.918908	0.918908	0.918030	0.916542	0.924968
15	0.953221	0.953003	0.953221	0.953221	0.952557	0.951260	0.960275
16	0.974420	0.974259	0.974420	0.974420	0.973981	0.972958	0.981192
17	0.986718	0.986609	0.986718	0.986718	0.986462	0.985722	0.991777
19	0.996913	0.996874	0.996913	0.996913	0.996856	0.996546	0.998811
21	0.999405	0.999394	0.999405	0.999405	0.999400	0.999300	0.999883
23	0.999904	0.999901	0.999904	0.999904	0.999904	0.999880	0.999992
25	0.999987	0.999986	0.999987	0.999987	0.999985	0.999982	1.000000
	Max. error	0.000265	$< 10^{-6}$	$< 10^{-6}$	0.000940	0.002360	0.019518

Table 4: Exact and approximate probabilities for Example 3.

5.3 Example 3

In this example, we take larger values of n and smaller values of p:

n_i	p_i
100	0.010
100	0.015
100	0.020
100	0.025
100	0.030

On this occasion, we would expect the simple Poisson approximation to work well, since the p_i are so small and the n_i are quite large. Starting the Kolmogorov approximation with a binomial having $n = 500$, $p = 0.02$, we obtain the results of Table ??.

In this example, the Poisson approximation is, surprisingly, not even as good as the Pearson approximation, while the Kolmogorov approximation using four moments is correct to the six decimals given. The normal approximation is ineffective because of the considerable skewness in the distribution of S.

s	True $P(S \leq s)$	Approximate $P(S \leq s)$ K (0)	K (4)	K (6)	Pearson	Poisson	Normal
5	0.615961	0.615961	0.615961	0.615961	0.615463	0.615961	0.588668
6	0.762519	0.762428	0.762519	0.762519	0.760952	0.762183	0.749322
7	0.867107	0.866977	0.867107	0.867107	0.865215	0.866628	0.868770
8	0.932354	0.932233	0.932354	0.932354	0.930799	0.931906	0.941657
9	0.968503	0.968414	0.968503	0.968503	0.967529	0.968172	0.978156
10	0.986511	0.986456	0.986511	0.986511	0.986034	0.986305	0.993155
11	0.994659	0.994629	0.994659	0.994659	0.994484	0.994547	0.998213
12	0.998036	0.998021	0.998036	0.998036	0.997996	0.997981	0.999613
13	0.999326	0.999319	0.999326	0.999326	0.999330	0.999302	0.999931
14	0.999783	0.999781	0.999783	0.999783	0.999794	0.999774	0.999990
15	0.999935	0.999934	0.999935	0.999935	0.999942	0.999931	0.999999
16	0.999981	0.999981	0.999981	0.999981	0.999986	0.999980	1.000000
	Max. error	0.000130	$< 10^{-6}$	$< 10^{-6}$	0.001892	0.000479	0.027923

Table 5: Exact and approximate probabilities for Example 4.

5.4 Example 4

For our final example, we take still larger values of n_i and still smaller values of p_i:

n_i	p_i
500	0.0020
400	0.0025
300	0.0033
200	0.0050
100	0.0100

We would expect this S to behave like a Poisson random variable with mean 5, since the distribution of each X_i is very close to Poisson with mean 1. We start the Kolmogorov approximations with a binomial distribution which has $n = 1500$ and $p = \frac{1}{300}$ and obtain the results shown in Table ??.

For these n_i and p_i, the Poisson approximation does outperform the Pearson curve, although not by a wide margin. Once again, the Kolmogorov approximations are superior, achieving six-decimal accuracy with only four moments.

6 Conclusions

It is clear that, although the normal approximation is generally accepted for large n_i and the Poisson approximation for small p_i, greater accuracy can be obtained from the Kolmogorov and Pearson curve approximations. Though the two latter techniques performed well throughout this study, the Pearson curve approximation was most effective when the n_i were large, while the Kolmogorov technique was most effective when the p_i were small. The Pearson curve approximation is easier to use, since the Kolmogorov technique requires an iterative procedure with highly accurate intermediate computation, but once suitable programs are available, both methods are very straightforward to apply.

References

Abramowitz, M. and Stegun, I. A. (1970) *Handbook of Mathematical Functions.* Dover, New York.

Boland, P. J. and Proschan, F. (1983) The reliability of k out of n systems. *Ann. Prob.* **11** 760–764.

Bol'shev, L. N. (1963) Asymptotically Pearson transformations. *Theor. Prob. Appl.* **8** 121–146.

Dunin-Barkovsky, I. V. and Smirnov, N. V. (1955) *Theory of Probability and Mathematical Statistics in Engineering.* Nauka, Moscow.

Gebhardt, F. (1969) Some numerical comparisons of several approximations to the binomial distribution. *J. Amer. Statist. Assoc.* **64** 1638–1646.

Hoeffding, W. (1956) On the distribution of the number of successes in independent trials. *Ann. Math. Statist.* **27** 713–721.

Johnson, N. L. and Kotz, S. (1969) *Distributions in Statistics Vol. 1: Discrete Distributions.* Wiley, New York.

Solomon, H. and Stephens, M. A. (1978) The distribution of a sum of weighted chi-square variables. *J. Amer. Statist. Assoc.* **73** 153–160.

Stephens, M. A. (1965) Significance points for the two-sample statistic $U^2_{M,N}$. *Biometrika* **52** 661–663.

Stephens, M. A. (1992) Moments in statistics: approximations to density functions. Proceedings of Conference on Moments in Statistics and Signal Processing, Naval Postgraduate School, Monterey.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
467

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
The Distribution of a Sum of Binomial Random Variables

5. TYPE OF REPORT & PERIOD COVERED
Technical

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Ken Butler and Michael Stephens

8. CONTRACT OR GRANT NUMBER(s)
N0025-92-J-1264

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Department of Statistics
Stanford University
Stanford, CA 94305-4065

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
NR-042-267

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
Statistics & Probability Program
Code 111

12. REPORT DATE
April 28, 1993

13. NUMBER OF PAGES
15

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

KEY WORDS: Kolmogorov-type approximation; Pearson distributions; reliability; survival analysis.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper we examine the distribution of a sum S of binomial random variables, each with different success probabilities. The distribution arises in reliability analysis and in survival analysis. An algorithm is given to calculate the exact distribution of S, and several approximations are examined. An approximation based on a method of Kolmogorov, and another based on fitting a distribution from the Pearson family, can be recommended.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)