

# **A Tutorial on How to Efficiently Calculate and Format Tables of the Binomial Distribution**

**A. M. Al-Qasimi and A. M. A. Rushdi**

*Department of Electrical and Computer Engineering, King Abdulaziz University, P.O. Box 80204, Jeddah 21589, Saudi Arabia*  
{aqasimi, arushdi}@kau.edu.sa

*Abstract.* Owing to the importance of the binomial distribution, tables for its probability mass function (PMF) and its cumulative distribution function (CDF) have been extensively published. This paper presents a tutorial on a novel technique for constructing and recording these tables. The tables are to be compiled with a high precision and a minimal round-off error through the use of highly efficient iterative algorithms based on binary recursive relations for the PMF and the CDF. The computational cost for each additional table entry is merely a single multiplication plus two additions. This remarkably low cost is due to the fact that each entry is computed in terms of exactly two preceding entries. Certain symmetries and limiting values are noted, thereby allowing a considerable reduction in the number of listed entries. Table entries are given in compact forms of fixed-point or floating-point format, with an objective of minimizing the numerical uncertainty that arises when representing an entry in a limited field width, *i.e.*, when using a limited number of digit or character positions. This kind of novel format is crucial for accurate evaluation of probabilities associated with ultra-high reliability systems or with very rare events.

## **1. Assumptions and Notation**

### ***1.1 Assumptions***

A series of Bernoulli trials are made satisfying the following conditions:

1. The total number of trials  $n$  is fixed.
2. Each trial has exactly two mutually exclusive outcomes, arbitrarily labelled as *success* and *failure*.
3. The outcomes of the various trials are statistically independent.

4. The conditions under which each trial is held are an exact replication of the conditions underlying any of the other trials, implying that the probabilities of successes on the various trials are equal.

### 1.2 Notation

$n$  number of Bernoulli trials = sample size, a non-negative integer.

$p, q$  probabilities of success and failure of each Bernoulli trial;  $p = 1.0 - q$ . Both  $p$  and  $q$  take real values in the closed real interval  $[0.0, 1.0]$ .

$E(k, n, p)$  probability of success in exactly  $k$  out of the  $n$  trials. This can be non-zero only for  $0 \leq k \leq n$ , and represents the probability mass function (PMF) of the total number of successes in  $n$  Bernoulli trials, *i.e.*, it is the PMF of the binomial distribution.

$$E(k, n, p) = c(k, n) p^k (1.0 - p)^{n-k}, \quad 0 \leq k \leq n \quad (1a)$$

$$E(k, n, p) = 0.0, \quad \text{otherwise.} \quad (1b)$$

$c(k, n)$  the binomial (combinatorial) coefficient = the number of ways of choosing  $k$  elements from a set of  $n$  objects, when repetition is not allowed and order does not matter<sup>[1]</sup>. Binomial coefficients satisfy Pascal's identity:

$$c(k, n) = c(k, n-1) + c(k-1, n-1), \quad 0 < k < n, \quad (2)$$

together with the boundary conditions:

$$c(k, n) = 1, \quad (k = 0 \text{ or } k = n) \text{ and } n \geq 0. \quad (3)$$

$\lfloor x \rfloor$  the floor of the real number  $x$  = the greatest integer less than or equal to  $x$ .

$B(t, n, p)$  the cumulative distribution function (CDF) of the binomial distribution (where  $t$  is a real number,  $-\infty < t < \infty$ ).

$$B(t, n, p) = B(\lfloor t \rfloor, n, p) \quad (4)$$

$B(k, n, p)$  a special case of  $B(t, n, p)$  when its first argument  $t$  is integer valued = probability of at most  $k$  successes in the  $n$  trials. It can have non-zero and non-unity values only for  $0 \leq k \leq (n-1)$ .

$$B(k, n, p) = \sum_{i=0}^k E(i, n, p) \quad (5)$$

$$E(k, n, p) = B(k, n, p) - B(k-1, n, p). \quad (6)$$

## 2. Introduction

Whenever a person is away from his office or computer, he can effectively satisfy his computational needs through a mathematical table, a pocket calculator or a combination thereof. This practice is particularly appropriate for a reliability engineer or statistician when he is asked ad hoc questions or while consulting in the field<sup>[2]</sup>. It is also useful for educators and students during class work and in conjunction with typical textbooks.

This paper deals with several issues pertinent to the computation and recording of basic quantities of the binomial distribution. These quantities (either the PMF or the CDF or both functions) have been extensively tabulated. To obtain information on typical tables, the reader may refer, *e.g.*, to Feller<sup>[3]</sup>, Kendall and Stuart<sup>[4]</sup>, Johnson and Leone<sup>[5]</sup>, Trivedi<sup>[6]</sup> or Billinton and Allan<sup>[7]</sup>. Such tables have not been rendered obsolete by the advent of hand electronic calculators, as attested by the inclusion of a table of the binomial CDF in the pocket book by Odeh *et al.*<sup>[2]</sup>. Probably, this is due to the fact (pointed out by Allen<sup>[8]</sup>) that Eq. (1) is a tedious calculation to carry out with a pocket calculator unless the calculator is programmed to do so or equipped by a binomial library program (such as in the case of the HP-21S calculator). Computational aids other than tables also exist in the form of library subprograms to be run on some sort of a computer, including portable PCs. Allen<sup>[8]</sup> offers two APL functions *BINOMIAL* and *BINSUM* for computing the PMF  $E(k,n,p)$  and the CDF  $B(k,n,p)$ , respectively. He also lists similar functions within some statistical packages for personal computers<sup>[8]</sup>. If the value of  $n$  becomes very large, then under appropriate conditions the binomial PMF can be computed via its Laplace's approximation (normal approximation), or via its Poisson approximation<sup>[6]</sup>.

The difficulty in the task of computing the PMF  $E(k,n,p)$  via Eq. (1a) is often aggravated by the usual practice of computing the binomial coefficient  $c(k,n)$  within this equation in terms of factorials, *viz.*,

$$c(k,n) = \text{factorial}(n) / (\text{factorial}(k) * \text{factorial}(n-k)) \quad (7)$$

or, more simply, by multiplying  $m = \min(k, n-k)$  descending consecutive integers that start at  $n$ , and dividing the resulting product by a similar number of ascending consecutive integers that start at 1, namely:

$$c(k, n) = n / 1 * (n-1) / 2 * \dots * (n-m+1) / m \quad (8)$$

Of course, there exists a more efficient technique for computing  $c(k, n)$  through an iterative procedure based on the recursive relations (2) and the boundary conditions (3). Such a procedure is outlined by Rushdi<sup>[9]</sup>, and entails the construction of an array of values  $c(i, j)$ , inclusively bounded in the  $ij$ -plane by the four straight lines  $i = 1$ ,  $i = k$ ,  $i = (j-1)$ ,  $i = (j-n+k)$ . These lines are the edges of a parallelogram that constitutes a sub-area of the well-known Pascal's triangle. However, the underlying principle of this technique is directly applicable to each of the binomial PMF and the binomial CDF, as will be shown in Sections 3 and 4. Based on this observation, efficient iterative procedures for tabulating the binomial distribution are developed. The computational cost for each table entry is merely a single multiplication plus two additions. This remarkably low cost is due to the fact that each entry is computed in terms of exactly two preceding entries. Section 5 outlines the table construction and gives the necessary accompanying description. Section 5 shows also how table entries are written in compact forms of fixed-point or floating-point format, with an objective of minimizing the numerical uncertainty that arises when representing an entry in a limited field width, *i.e.*, when using a limited number of digit or character positions. This kind of novel format avoids the appearance of the approximate values of 0 in the PMF table or of 1 in the CDF table, and hence is crucial for accurate evaluation of probabilities associated with ultra-high reliability systems or with very rare events.

Most modern computational packages offer functions for computing the binomial PMF. For example, MATLAB computes this PMF via its function  $binopdf(x, n, p)$ . This function unnecessarily uses a real argument  $x$ , and states that the PMF is zero unless  $x$  is integer. The binomial PMF is computed in terms of the MATLAB function  $\text{gammaln}(x)$ , which computes the natural logarithm of the gamma function. Obviously, MATLAB does not care to compute the binomial PMF in a specially tailored or a particularly efficient way. It rather computes the binomial PMF as an offshoot or a by-product of its computation of a much more complex function. This serves the purpose of computing a multitude of

useful mathematical functions by as few procedures as possible, and relies heavily on a utilization of the contemporarily abundant and almost free resources of computer time and memory.

Though this paper handles statistical quantities, it should not be considered a paper on statistics as such. Actually, the paper is dealing with computational issues such as analysis and design of algorithms, temporal and spatial complexities, iteration versus recursion, numerical precision, information content and the like. Therefore, this paper is intended to serve as a pedagogical tool or as a tutorial for the above issues. The expected readership of this paper should include the practitioners and students of the disciplines of reliability engineering and risk analysis, statistics and mathematics, and possibly computer engineering and science.

### 3. Recursive Relations

The binomial PMF  $E(k,n,p)$  satisfies the recursive equations<sup>[9]</sup>:

$$E(k,n,p) = q * E(k,n-1,p) + p * E(k-1,n-1,p) \quad (9a)$$

$$= E(k,n-1,p) + p * ( E(k-1,n-1,p) - E(k,n-1,p) ) \quad (9b)$$

in a region of validity  $\{ 0 < k < n \}$  and subject to the set of boundary conditions:

$$E(k,n,p) = q^n, \quad k = 0, n \geq 1 \quad (9c)$$

$$E(k,n,p) = p^n, \quad k = n, n \geq 1 \quad (9d)$$

or in a region of validity  $\{ 0 \leq k \leq n, n \geq 1 \}$  and subject to the alternative set of boundary conditions:

$$E(k,n,p) = 1.0, \quad k = n = 0 \quad (9e)$$

$$E(k,n,p) = 0.0, \quad k = -1, n \geq 0 \quad (9f)$$

$$E(k,n,p) = 0.0, \quad k = (n+1), n \geq 0 \quad (9g)$$

The above recursive relations and boundary conditions are remarkably similar to those of the binomial coefficient  $c(k,n)$  as given by Eqs. (2) and (3).

Similar recursive relations can be derived for the CDF  $B(k,n,p)$  (albeit with a different region of validity and a different set of boundary conditions), namely:

$$B(k,n,p) = q * B(k,n-1,p) + p * B(k-1,n-1,p), \quad 0 \leq k \leq (n-1) \quad (10a)$$

$$= B(k,n-1,p) + p * ( B(k-1,n-1,p) - B(k,n-1,p) ), \quad 0 \leq k \leq (n-1) \quad (10b)$$

$$= B(k-1,n-1,p) + q * ( B(k,n-1,p) - B(k-1,n-1,p) ), \quad 0 \leq k \leq (n-1) \quad (10c)$$

together with:

$$B(k,n,p) = 0.0, \quad k = -1, n \geq 0 \quad (10d)$$

$$B(k,n,p) = 1.0, \quad k = n, n \geq 0 \quad (10e)$$

The boundary conditions above nicely represent the limiting values of the cumulative distribution function  $B(k,n,p)$ .

The quantities  $B(k,n,p)$  and  $E(k,n,p)$  are interrelated by Eqs. (5) and (6). In addition, they also satisfy the following relations<sup>[3]</sup>,

$$B(k,n,p) = B(k,n-1,p) - p * E(k,n-1,p), \quad 0 \leq k \leq (n-1) \quad (11)$$

$$B(k+1,n,p) = B(k,n-1,p) + q * E(k+1,n-1,p), \quad 0 \leq (k+1) \leq (n-1) \quad (12)$$

Equations (11) and (12) can be derived by substitutions of Eq. (6) into Eq. (10b) or Eq. (10c). Their region of validity can be enlarged to all values  $\{ n \geq 1 \}$  by taking into consideration Eq. (9e) together with:

$$B(k,n,p) = E(k,n,p) = 0.0, \quad k \leq -1, n \geq 0 \quad (13a)$$

$$E(k,n,p) = 0.0, \quad 0 \leq n < k \quad (13b)$$

$$B(k,n,p) = 1.0, \quad 0 \leq n < k \quad (13c)$$

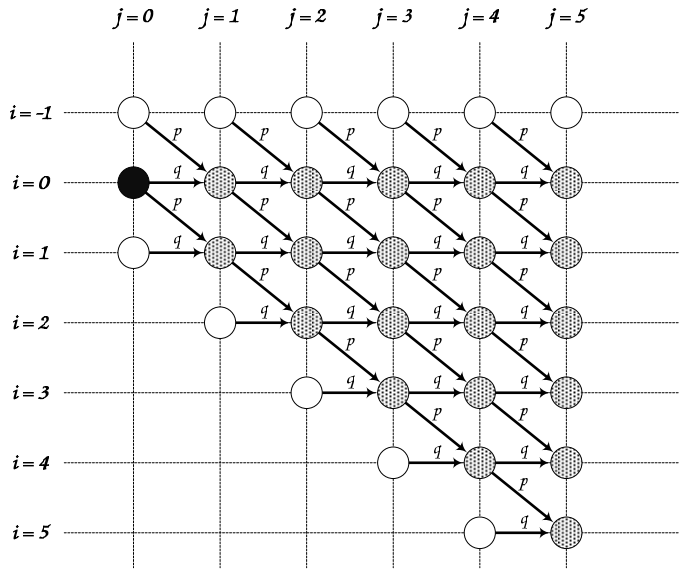
#### 4. Iterative Procedures

This section deals with the construction of purely iterative algorithms that utilize the underlying recursive equations without any resort to recursion. As a step towards the development of these iterative algorithms, we first give an interpretation of the underlying recursive equations in terms of a (Mason) signal flow graph (SFG)<sup>[10]</sup>. Figure 1 gives a very regular SFG that illustrates the computation of the PMF  $E(k,n,p)$  of the binomial distribution via Eqs. (9), while Figure 2 gives a similar SFG that illustrates the computation of the corresponding  $B(k,n,p)$  via Eqs. (10). In these two figures, there are two types of nodes:

(a) Source nodes of known values which are either black or white. A black node has a value of 1.0 while a white node has a value of 0.0.

(b) Non-source nodes drawn as shaded ones. These are of unknown values which are to be computed by the underlying algorithm. These non-source nodes include (at least) one sink node whose value is the final result sought, beside other nodes whose values are intermediate results.

In Fig. 1, a node of a thin circumference at coordinates  $(i,j)$  represents  $E(i,j,p)$ , while in Fig. 2, a node of a thick circumference at coordinates  $(i,j)$  stands for  $B(i,j,p)$ . Of course, white nodes (and the links emanating therefrom) may be discarded, but this would impair the regularity of the SFG.

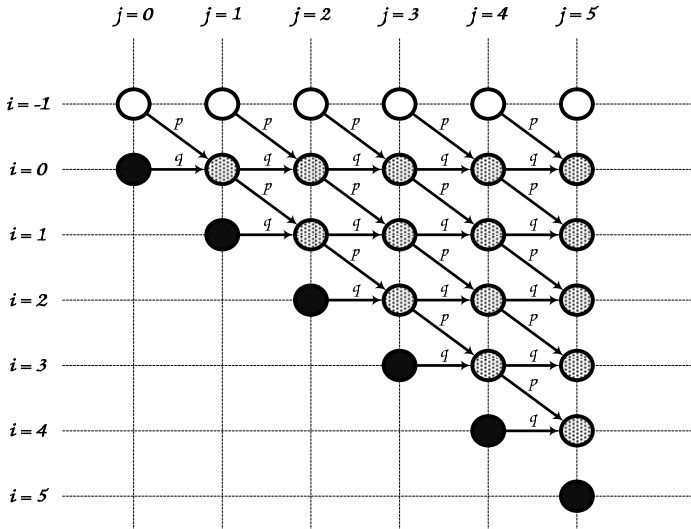


**Fig.1. A Signal flow graph that illustrates the computation of the PMF  $E(i, j, p)$  of the binomial distribution.**

The above SFGs are special cases of those adopted by the AR algorithm of Rushdi<sup>[9]</sup> for computing the k-out-of-n system reliability or unreliability. They also resemble the SFGs pertinent to the algorithms of Rushdi and Al-Thubaity<sup>[11]</sup> for computing the sensitivity of the k-out-of-n system reliability or unreliability.

The pictorial insight provided by the SFGs leads to the development of efficient iterative algorithms for computing  $E(k,n,p)$ ,  $0 \leq k \leq n$ , and  $B(k,n,p)$ ,  $0 \leq k \leq (n-1)$ , as follows. In both algorithms, nodes of the

corresponding SFG are visited columnwise, starting from the leftmost column ( $j = 1$ ) and ending at the rightmost column ( $j = n$ ). Within each column  $j$  the bottom shaded node ( $i = j$  in Fig. 1, or  $i = (j-1)$  in Fig. 2 ) is visited first and then followed by upper nodes till the top shaded node ( $i = 0$ ) is reached. At the end of the traversal of column  $j$ ,  $1 \leq j \leq n$ , its nodes contain the set of values of the PMF that are possibly different from zero, namely  $E(i,j,p)$ ,  $0 \leq i \leq j$ , ( Fig. 1 ) or contain the set of values of the CDF that are possibly different from both zero and unity, namely  $B(i,j,p)$ ,  $0 \leq i \leq (j - 1)$ , ( Fig. 2 ).



**Fig. 2.** A signal flow graph that illustrates the computation of the CDF  $B(i, j, p)$  of the binomial distribution.

### 5. Table Construction and Description

Due to space limitations, the user may have to choose between having a PMF table or having a CDF one. Our advice to him is to have a table of the PMF rather than one of the CDF. We are aware that this advice might be debatable or even controversial, but we base it on the following two observations:

1. A table of the CDF  $B(k,n,p)$  is inferior to one of the PMF  $E(k,n,p)$  in its numerical information content. In fact, entries of unity abound in tables of the CDF. Such entries do not usually represent exact values, but are



worthless approximations representing values different from one that have, nevertheless, been rounded off to one due to the finite precision of a table entry. For example, in a CDF table of Trivedi<sup>[6]</sup>, out of 200 explicit or implied entries for  $B(k,20,p)$ , 70 entries have the value 1.0000, none of which is exact. Table 24 of Odeh *et al*<sup>[2]</sup>  $B(k,30,p)$  has 155 explicit or implied entries out of which 62 entries are given by 1.0000, including 57 entries that are not exact. Of course, some of the unity entries can be omitted, but this can only reduce table size without improving the significance of its content. The use of a higher precision for the table entries will only postpone this problem, and will not eliminate it. The switch from fixed-point format to floating-point format does not solve the problem, either, unless one chooses to replace the unity and near-unity entries by their complements to one, *i.e.*, unless table uniformity is spoiled by reporting  $(1.0 - B(k,n,p))$  instead of  $B(k,n,p)$  for some table entries.

2. A table of the CDF  $B(k,n,p)$  can be used to produce values of the PMF  $E(k,n,p)$  via the differencing (or subtraction) operation of Eq. (6). By contrast, a table of the PMF  $E(k,n,p)$  can produce values of the CDF  $B(k,n,p)$  via the summation (or repeated addition) operation of Eq. (5). Both operations of subtraction and repeated addition can be implemented manually on a hand calculator, though the latter operation is more laborious, and therefore less popular with most users who tend to prefer CDF tables. However, the former operation is more error prone since it involves the subtraction of two numbers that have the same sign and approximately the same magnitude. This may lead to *catastrophic cancellation*, when all or almost all the significant digits of the two numbers are identical. By contrast, the summation operation of Eq. (5) involves the actual accumulation of non-negative numbers, and hence it can be ensured that the level of round-off error is not increased significantly. In the worst case, the summation may lead to a value of  $B(k,n,p)$  that is greater than 1; an error that can be remedied (at least partially) by truncating this value at 1.0.

Tables 1 and 2 present a glimpse of the PMF and the CDF tables for  $n=20$ , reproduced directly from computer printout, thereby avoiding the possibility of typesetting errors. All table entries are less than or equal to 1.0 and are initially intended to be given in fixed-point format. The decimal point is deliberately omitted and is implicitly assumed to be to the left of the leftmost digit. To preserve accuracy, smaller numbers are given in floating-point rather than fixed-point format. A small number in floating-point format

can be recognized by the appearance of a minus sign (  $-$  ) within an entry. The digits to the left of that sign constitute the significant or mantissa of the number with a decimal point implicitly situated to the left of the leftmost digit. The sign itself together with the digit(s) to its right represent the exponent of the number to base 10. The numerical uncertainty of an entry is within  $\pm 0.5$  of the weight of the rightmost digit position in the significand of the entry (A fixed-point number is a floating-point number with a significand equal to the number and a zero exponent). On the other hand, if a number is very near to 1, it is represented by its difference from 1, again in floating-point format but with the character # used in place of the minus sign. Since the field width per entry is constant, the absolute numerical uncertainty of the tables is not uniform, but the relative uncertainty of the tables is more or less uniform. Tables 1 and 2 might look unconventional and unfamiliar, but they are richer in information content than conventional and familiar tables. For comparison purposes, we include the conventional versions of Tables 1 and 2 in Tables 3 and 4, respectively. When dealing with the reliability of an ultra-reliable system, one does not usually want to take it for a perfect 1.0, but wants to know how much it differs from 1.0. Conversely, when handling the probability of a rare event, one does not desire to consider it an exact 0.0, but needs to decide whatever small value it has.

**Table 1. The unconventional table for the PMF of the binomial distribution for  $n=20$ .**

n = 20											
k	p:	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
0		358486	121577	038760	011529	003171	7979-3	1812-3	3656-4	6416-5	9537-6
1		377354	270170	136798	057646	021141	006839	001952	4875-3	1050-3	1907-4
2		188677	285180	229338	136909	066948	027846	009985	003087	8160-3	1812-3
3		059582	190120	242829	205364	133896	071604	032258	012350	004006	001087
4		013328	089779	182122	218199	189685	130421	073821	034991	013930	004621
5		002245	031921	102845	174560	202331	178863	127199	074647	036471	014786
6		2953-3	008867	045373	109100	168609	191639	171230	124412	074600	036964
7		3109-4	001970	016014	054550	112406	164262	184401	165882	122072	073929
8		2659-5	3558-3	004592	022161	060887	114397	161351	179706	162300	120134
9		1866-6	5271-4	001081	007387	027061	065370	115842	159738	177055	160179
10		1080-7	6442-5	2097-3	002031	009922	030817	068614	117142	159349	176197
11		5169-9	6507-6	3365-4	4617-3	003007	012007	033587	070995	118524	160179
12		204-10	5423-7	4454-5	8657-4	7517-3	003859	013564	035497	072731	120134
13		661-12	3708-8	4837-6	1332-4	1542-3	001018	004495	014563	036620	073929
14		174-13	2060-9	4268-7	1665-5	2570-4	2181-3	001210	004854	014981	036964
15		366-15	915-11	3012-8	1665-6	3426-5	3739-4	2606-3	001294	004903	014786
16		602-17	318-12	1661-9	1301-7	3569-6	5008-5	4386-4	2697-3	001254	004621
17		746-19	831-14	690-11	7650-9	2799-7	5050-6	5557-5	4230-4	2413-3	001087
18		654-21	154-15	203-12	319-10	1555-8	3607-7	4987-6	4700-5	3291-4	1812-3
19		362-23	180-17	377-14	839-12	546-10	1627-8	2826-7	3299-6	2834-5	1907-4
20		954-26	100-19	333-16	105-13	909-12	349-10	7610-9	1100-7	1159-6	9537-6

To keep down the sizes of the tables, we have restricted them to the values  $0.05 \leq p \leq 0.50$ , using a step of 0.05 in  $p$ . For the complementary values  $0.50 < p \leq 0.95$ , the PMF table can still be used, but in conjunction with the symmetry relation:

$$E(k,n,p) = E(n - k, n, 1.0 - p) \tag{14}$$

For the limiting values of  $p = 0.0$  and  $p = 1.0$ , we have:

$$E(k,n,0.0) = 1.0, \quad k = 0, n \geq 0 \tag{15a}$$

$$= 0.0, \quad k \neq 0, n \geq 0 \tag{15b}$$

$$E(k,n,1.0) = 1.0, \quad k = n, n \geq 0 \tag{16a}$$

$$= 0.0, \quad k \neq n, n \geq 0 \tag{16b}$$

In view of Eqs. (14)-(16), the PMF table can be considered to cover or list all values of  $p$  in the range  $0.0 \leq p \leq 1.0$ , with a step of 0.05. Similar arguments hold for the CDF table.

**Table 2. The unconventional table for the CDF of the binomial distribution for  $n = 20$ .**

n = 20											
k	p:	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
0		358486	121577	038760	011529	003171	7979-3	1812-3	3656-4	6416-5	9537-6
1		735840	391747	175558	069175	024313	007637	002133	5240-3	1114-3	2003-4
2		924516	676927	404896	206085	091260	035483	012118	003611	9274-3	2012-3
3		984098	867047	647725	411449	225156	107087	044376	015961	004933	001288
4		997426	956826	829847	629648	414842	237508	118197	050952	018863	005909
5		3293#3	988747	932692	804208	617173	416371	245396	125599	055334	020695
6		3395#4	997614	978065	913307	785782	608010	416625	250011	129934	057659
7		2857#5	4156#3	994079	967857	898188	772272	601027	415893	252006	131588
8		1979#6	5986#4	998671	990018	959075	886669	762378	595599	414306	251722
9		1134#7	7151#5	2484#3	997405	986136	952038	878219	755337	591361	411901
10		5380#9	7089#6	3863#4	5634#3	996058	982855	946833	872479	750711	588099
11		211#10	5815#7	4983#5	1017#3	9354#3	994862	980421	943474	869235	748278
12		679#12	3923#8	5295#6	1516#4	1837#3	998721	993985	978971	941966	868412
13		178#13	2155#9	4586#7	1845#5	2951#4	2610#3	998479	993534	978586	942341
14		372#15	948#11	3186#8	1803#6	3813#5	4294#4	3106#3	998388	993566	979305
15		613#17	326#12	1732#9	1380#7	3865#6	5550#5	4994#4	3170#3	998469	994091
16		542#19	847#14	710#11	7978#9	2960#7	5427#6	6084#5	4734#4	2772#3	998712
17		0000#0	156#15	207#12	327#10	1611#8	3773#7	5277#6	5041#5	3586#4	2012#3
18		0000#0	184#17	380#14	849#12	555#10	1662#8	2903#7	3408#6	2950#5	2003#4
19		0000#0	0000#0	332#16	105#13	909#12	349#10	7610#9	1100#7	1159#6	9537#6
20		0000#0	0000#0	0000#0	0000#0	0000#0	0000#0	0000#0	0000#0	0000#0	0000#0

To further reduce table sizes, we suggest also that only entries for odd values of  $n$  be printed. An entry for an even value of  $n$  can be easily computed from the table entries of the immediately preceding odd number,



If the value of  $E(k,n,p)$  is required for an unlisted value of  $p$  with a listed combination of  $k$  and  $n$ , it is a common practice to find the corresponding two values for the surrounding listed values  $p_1$  and  $p_2$  ( where  $p_1 < p < p_2$  and  $p_2 - p_1 = 0.05$  in the present case ), and then use the linear interpolation:

$$E(k,n,p) \approx ((p_2-p) * E(k,n,p_1) + (p-p_1) * E(k,n,p_2)) / (p_2-p_1) \quad (17)$$

Since the second derivative of  $E(k,n,p)$  with respect to  $p$  is a polynomial in  $p$ , this derivative is expected to have a well-behaved change with  $p$ , and hence the maximum truncation error due to the above interpolation in the interval  $(p_1, p_2)$  is more likely to occur at or near the midpoint  $p = 0.5 * (p_1 + p_2)$ . Therefore, we made an empirical study of the total error that is committed when linear interpolation is used to compute values of  $E(k,n,p)$  at midpoints between listed entries. The use of linear interpolation in conjunction with the reported table can be acceptable if the error introduced by this interpolation is of the same order of magnitude as that of the numerical uncertainties in the interpolated entries<sup>[12]</sup>. For most table entries, the above condition is not satisfied. In fact, for many table entries, and in particular those given in floating-point format, the use of linear interpolation leads to relative errors that exceed unity. Therefore, a word of caution must be given against linearly interpolating table entries and assuming the results to retain the original precision. If, however, the following conservative condition

$$ABS( E(k,n,p_2) - E(k,n,p_1) ) \ll \min( E(k,n,p_1) , E(k,n,p_2) ) \quad (18)$$

is satisfied, then precision is preserved when linear interpolation is used. The precision problem can be partially tackled if one constructs larger tables with smaller steps in  $p$ . This observation seems to explain why in the past many large volumes (several hundreds of pages each) have been published for the sole purpose of tabulating the binomial distribution.

## 6. Conclusions

The contemporary progress in computer technology has been, in a sense, a mixed blessing. It has allowed an increase in computational speed, which is so dramatic that there is a current trend among many algorithm designers to belittle and even ignore efficiency issues. This paper is an outcry against this trend. We must learn and practise the

design of efficient algorithms, even when dealing with simple and elementary problems, if we are to be ever successful in handling the more complex and intricate problems of the future.

### References

- [1] **Rushdi, A.M.** and **Al-Qasimi, A.M.**, Efficient computation of the PMF and the CDF of the generalized binomial distribution, *Microelectronics and Reliability*, **34** (9): 1489-1499 (1994).
- [2] **Odeh, R.E., Owen, D.B., Birnbaum, Z. W.** and **Fisher, L.**, *Pocket Book of Statistical Tables*, Marcel Dekker, New York and Basel (1977).
- [3] **Feller, W.**, *An Introduction to Probability Theory and its Applications*, Vol. **I**, Third Edition, John Wiley & Sons, New York (1968).
- [4] **Kendall, M.** and **Stuart, A.**, *The Advanced Theory of Statistics*, Vol. **1**, Fourth Edition. Macmillan, New York (1977).
- [5] **Johnson, N.L.** and **Leone, F.C.**, *Statistics and Experimental Design in Engineering and the Physical Sciences*, Vol. **I**, Second Edition, John Wiley & Sons, New York (1977).
- [6] **Trivedi, K.S.**, *Probability & Statistics with Reliability, Queuing, and Computer Science Applications*, Second Edition, Prentice-Hall, Englewood Cliffs, NJ (2002).
- [7] **Billinton, R.** and **Allan, R.N.**, *Reliability Evaluation of Engineering Systems: Concepts and Techniques*, Second Edition, Springer, New York (2005).
- [8] **Allen, A.O.**, *Probability, Statistics, and Queueing Theory with Computer Science Applications*, Second Edition, Academic Press, San Diego, CA (1990).
- [9] **Rushdi, A.M.**, Reliability of k-out-of-n Systems, Chapter 5, In: Misra, K. B. ( Editor ), *New Trends in System Reliability Evaluation*, Vol. **16**, *Fundamental Studies in Engineering*, Elsevier Science Publishers, Amsterdam, The Netherlands, pp: 185-227 (1993).
- [10] **Kuo, B.C.** and **Golnaraghi, F.**, *Automatic Control Systems*, Eighth Edition, John Wiley & Sons, New York (2002).
- [11] **Rushdi, A.M.** and **Al-Thubaity, A.O.**, Efficient computation of the sensitivity of k-out-of-n system reliability, *Microelectronics and Reliability*, **33** (13): 1963-1979 (1993).
- [12] **Burden, R.L.** and **Faires, J.D.**, *Numerical Analysis*, Eighth Edition, Thomson Brooks/Cole, Boston (2005).

## شرح تعليمي عن كيف تحسب وتشكل بكفاءة جداول التوزيع ذي الحدين

عبد الغني محمد القاسمي، وعلي محمد علي رشدي

قسم الهندسة الكهربائية وهندسة الحاسبات ، جامعة الملك عبد العزيز،  
ص. ب. ٨٠٢٠٤، جدة ٢١٥٨٩، المملكة العربية السعودية

المستخلص. نظراً لأهمية التوزيع الاحتمالي ذي الحدين، فإن كلا من جدولي دالة الكتلة الاحتمالية (د ك ح) ودالة التوزيع التراكمي (د و ر) له قد جرى نشرهما على نطاق واسع. تقدم ورقة البحث شرحاً تعليمياً لأسلوب مبتكر لإنشاء وتسجيل هذين الجدولين. ويمكن تجميع الجدولين بدقة عالية مع تصغير أخطاء القطع إلى أقل ما يمكن، وذلك من خلال استخدام خوارزميتين تكراريتين ذواتي كفاءة عالية تعتمدان على علاقات المعاودة الثنائية لدالتي الكتلة الاحتمالية والتوزيع التراكمي. وتقتصر الكلفة الحسابية لكل مدخل جدولي إضافي على عملية ضرب واحدة، بالإضافة إلى عمليتي جمع فقط. ويرجع الانخفاض الملموس في هذه الكلفة إلى حقيقة أن كل مدخل يتم حسابه بدلالة مدخلين اثنين سابقين له. وقد جرت ملاحظة بعض أنواع التماثل وبعض القيم النهائية، مما أدى إلى تخفيض ملحوظ في عدد المدخلات المدرجة. وقد أعطيت مدخلات الجدولين في صيغ مملومة تستعمل شكل العلامة الثابتة أو شكل العلامة المتحركة، وذلك بهدف التصغير الأعظمي للربية العددية التي تنشأ عند تمثيل المدخل في حيز ذي اتساع محدود، أي عند استعمال عدد محدود من منازل الأرقام العشرية أو المحارف. ويعد هذا النوع من الأشكال المستحدثة ذا أهمية بالغة للتقدير الدقيق للاحتتمالات المستخدمة في النظم ذات المعولية فائقة العلو أو لتلك المستخدمة مع الأحداث شديدة الندرة.