Application to the Binomial Summation of a Laplacian[1] Method for the Evaluation of Definite Integrals

By E. C. Molina

Introduction

The numerical evaluation of the incomplete Binomial Summation, a problem of major importance for many statistical and engineering applications of the Theory of Probability, is a question for which a satisfactory solution has not as yet been obtained. Several approximation formulas have been presented,[2] each of which gives good results for some limited range of values of the variables involved; but a formula of wide applicability is still a desideratum.

The purpose of this paper is to submit for consideration an approximation formula which seems to meet the situation to a measurable extent. The writer derived it by applying to the equation

$$(1) \qquad \sum_{x=c}^{x=n} \binom{n}{x} p^x(1-p)^{n-x} = \frac{\int_0^p x^{c-1}(1-x)^{n-c}dx}{\int_0^1 x^{c-1}(1-x)^{n-c}dx},$$

a method which is peculiarly efficacious for approximately evaluating definite integrals when the integrands contain factors raised to high powers.

The method used constitutes the subject matter of Chapter I, Part II, Book I of Laplace's "Théorie Analytique des Probabilités." Poisson applied the method to the integrals in the equation

$$(2) \qquad \sum_{x=c}^{x=n} \binom{n}{x} p^x(1-p)^{n-x} = \frac{\int_{(1-p)/p}^{\infty} x^{n-c}/(1+x)^{n+1}dx}{\int_0^{\infty} x^{n-c}/(1+x)^{n+1}dx}$$

and published a first approximation, together with its derivation, in his "Recherches sur la Probabilité des Jugements." Poisson's approximation seems never to have been used and was less fortunate than his famous limit to the binomial expansion which also was lost sight of until it reappeared under the caption "law of small numbers."

[1] Presented before International Congress of Mathematicians at Bologna, Italy in September, 1928.

[2] For an excellent resumé of some well-known formulas, together with a discussion of their limitations, reference may be had to C. Jordan, "Statistique Mathématique," articles 37 and 38.

While the integrals in equations (1) and (2) are well known equivalent forms for the complete and incomplete Beta functions, the equations themselves are not so familiar although one or the other will be found in Laplace, Poisson, Boole (Differential Equations) and at least two other places.

Approximate Formula

The approximate formula derived from equation (1) and submitted herewith for consideration is

$$(3) \qquad \sum_{x=c}^{x=n} \binom{n}{x} p^x(1-p)^{n-x} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{T} e^{-t^2} dt - \frac{S_i e^{-T^2}}{2\sqrt{\pi}},$$

where S_i is the ith approximation to the infinite series

$$(4) \quad S = \frac{\sum_{s=1} R_s T^{s-1}[1 + (s-1)T_1^{-2} + (s-1)(s-3)T_1^{-4} \cdots]}{1 + \sum_{s=1} R_{2s}[1\cdot 3\cdot 5 \cdots (2s-1)]2^{-s}},$$

$T_1 = T\sqrt{2}$,

$$(5) \quad T^2 = (n-1)\log\frac{n}{n-1} + (c-1)\log\frac{c-1}{a} + (n-c)\log\frac{n-c}{n-a},$$

and $a = np$; T to be taken negative when $a < (c-1)n/(n-1)$.

The first, second and third approximations to the infinite series S are

$$S_1 = R_1, \qquad S_2 = \frac{R_1 + R_2 T}{1 + R_2/2}, \qquad S_3 = \frac{R_1 + R_2 T + R_3(1 + T^2)}{1 + R_2/2},$$

where

$$R_1 = 4[(n-c) - (c-1)]/3\sqrt{2(n-1)(n-c)(c-1)},$$
$$R_2 = (1/6)[1/(n-c) + 1/(c-1) - 13/(n-1)],$$
$$R_3 = -(4/15)R_1[R_2 + 6/(n-1)].$$

It will be noted that R_2, $|R_1|$ and $|R_3|$ are symmetric functions of $(n-c)$ and $(c-1)$.

For the limiting case (Poisson's Exponential Binomial Limit) where $n = \infty$, $p = 0$ but $np = a$, we have

$$T^2 = 1 + (c-1)\log(c-1)/a + (a-c),$$
$$R_1 = 4/3\sqrt{2(c-1)},$$
$$R_2 = 1/6(c-1),$$
$$R_3 = -(4/15)R_1R_2.$$

Numerical Results

Since it is easy to compute the binomial summation directly when either c or $n - c$ is small, the practical value of an approximate formula depends on its efficiency for large values of these quantities.

The analysis given below under the heading "Derivation of the Approximate Formula" indicates that the successive R_s's in the series for S decrease when $\sqrt{c-1}$ and $\sqrt{n-c}$ increase. Therefore, when these two quantities are large, a few terms of the approximate formula (3) may be expected to give satisfactory results. As a matter of fact, the formula gives good results when $\sqrt{c-1}$ and $\sqrt{n-c}$ are not large. To confirm this statement the Tables [3] given at the end of this paper are submitted. In the 4th column of each table are given 10^6 times the true values of

$$P = \sum_{x=c}^{x=n} \binom{n}{x} p^x (1-p)^{n-x}.$$

In the columns headed Δ_1, Δ_2 and Δ_3 are given 10^6 times the differences between the true values and those obtained by applying formula (3) with the first, second and third approximations to S respectively. Table I in Czuber's "Wahrscheinlichkeitsrechnung" was used for evaluating the probability integral in equation (3).

The range of values of P covered by the tables is such that at the lower end of each section $P \ngtr .0005$ while at the upper end $P \nless .9995$, except where this latter condition would call for a value of $c < 2$. Of course, a larger or smaller range might have been given. The decision as to this question was based on the fact that several writers on the theory of statistics, when dealing with the normal law of errors, speak of an error exceeding 3 or 4 times the standard deviation as being a very improbable event. In order to keep the number of pages required for the tables within reasonable bounds computations were made only for even values of c.

The values of $a = np$ used are such that each of the values $p = 1/2$, $p = 1/10$ and $p = 1/20$ occurs twice; likewise each of the values $n = 100$, $n = 50$ and $n = 30$ occurs twice.

A greater degree of accuracy than that indicated by the tables can, of course, be obtained by working out and using R_4, $R_5 \cdots$; for this purpose, recourse should be had to equation (12) below and the details immediately following it. The only practical limitation to the use of formula (3) would appear to be the number of places given

[3] I am greatly indebted to Miss Nelliemae Z. Pearson of the Department of Development and Research both for supervising the work of my computers and contributing personally several sections of the tables.

by the existing tables for the probability integral. However, this difficulty is encountered only when P, or $(1 - P)$, is small, in which case T is large and the integral

$$\int_{-\infty}^{T} e^{-t^2}dt$$

may be readily evaluated by computing the first few terms of the series

$$[e^{-T^2}/2T\sqrt{\pi}][1 - T_1^{-2} + (1\cdot 3)T_1^{-4} - (1\cdot 3\cdot 5)T_1^{-6} \cdots],$$

where, as above, $T_1 = T\sqrt{2}$.

When P is very small, the difference $c - a = c - np$ is relatively large compared to a, and for this latter case recourse may be had to the approximate formula published by the writer in the *American Mathematical Monthly* for June, 1913.

Derivation of the Approximate Formula

Following Laplace closely, let us set

$$y(x) = Ye^{-t^2}, \tag{6}$$

where $Y = y(x_0)$ is the maximum value of $y(x)$. Then

$$\int_0^p ydx = Y\int_{-\infty}^{T} e^{-t^2}\left(\frac{dx}{dt}\right)dt, \tag{7}$$

the upper limit T being given by the equation

$$y(p) = y(x_0)e^{-T^2}. \tag{8}$$

Assuming dx/dt expanded in powers of t so that

$$dx/dt = \sum_{s=0} D_{s+1}t^s \tag{9}$$

and setting $R_s = D_{s+1}/D_1$, equation (7) reduces to

$$\int_0^p ydx = YD_1\sum_{s=0} R_s\int_{-\infty}^{T} t^s e^{-t^2}dt.$$

Our fundamental equation (1) may now be written

$$\sum_{x=c}^{x=n}\binom{n}{x}p^x(1 - p)^{n-x} = \frac{\sum_{s=0} R_s\int_{-\infty}^{T} t^s e^{-t^2}dt}{\sum_{s=0} R_s\int_{-\infty}^{\infty} t^s e^{-t^2}dt}. \tag{10}$$

Integrating by parts and separating the terms involving $\int e^{-t^2}dt$ from the terms containing e^{-t^2}, we obtain equations (3) and (4).

To determine $R_s = D_{s+1}/D_1$, note that equation (6) gives $t = (\log Y - \log y)^{1/2}$ and set $v(x) = (x - x_0)/(\log Y - \log y)^{1/2}$ so that x may be written in the form

$$x = x_0 + v(x)t.$$

This form for x gives the expansion (Lagrange's Theorem for the simple case where $f(x) = x$; see "Modern Analysis" by Whittaker and Watson)

$$x = \sum_{s=0} \frac{t^s}{s!}\left(\frac{d^{s-1}v^s}{dx^{s-1}}\right)_{x=x_0}.$$

Comparing this expansion for x with the previous expansion (9) for dx/dt, we obtain

$$D_1 = v(x_0)$$

and

$$\frac{D_{s+1}}{D_1} = R_s = \left(\frac{1}{s!v(x)} \cdot \frac{d^s v^{s+1}}{dx^s}\right)_{x=x_0}.$$

Up to this point no particular form has been attributed to the function $y(x)$. From now on we deal with the function which constitutes the integrand of the integrals in equation (1).

The function $y(x) = x^{c-1}(1 - x)^{n-c}$ gives the expansion $(\log Y - \log y) = (x - x_0)^2[A_0 + A_1(x - x_0) + A_2(x - x_0)^2 \cdots]$, where $x_0 = (c - 1)/(n - 1)$ is the value of x for which $y(x)$ is a maximum and

$$A_s = \frac{1}{(s+2)!}\left[\frac{d^{s+2}(\log Y - \log y)}{dx^{s+2}}\right]_{x=x_0}$$

or

$$(11) \qquad A_s = \frac{(n-1)^{s+2}}{s+2}\left[\left(\frac{1}{n-c}\right)^{s+1} + (-1)^s\left(\frac{1}{c-1}\right)^{s+1}\right].$$

We are now prepared to evaluate R_s. Set

$$g = A_0 + A_1(x - x_0) + A_2(x - x_0)^2 \cdots$$

and

$$g_s = d^s g/dx^s.$$

Then

$$v = g^{-1/2},$$

$$\frac{dv^2}{dx} = -g^{-2}g_1,$$

$$\frac{d^2v^3}{dx^2} = (3/2)g^{-7/2}[(5/2)g_1^2 - g_2 g],$$

$$\frac{d^3v^4}{dx^3} = -2g^{-5}[g_3 g^2 - 9g_2 g_1 g + 12 g_1^3].$$

Therefore, since $g_s = s!A_s$ when $x = x_0$,

$$R_1 = -A_0^{-3/2}A_1,$$
$$R_2 = (3/2)A_0^{-3}[(5/4)A_1^2 - A_0A_2],$$
$$R_3 = -2A_0^{-9/2}[A_3A_0^2 - 3A_2A_1A_0 + 2A_1^3].$$

Substituting for A_0, A_1, A_2 and A_3 the expressions derived by giving s the values 0, 1, 2 and 3 respectively in equation (11), we obtain for R_1, R_2 and R_3 the functions of n and c given on page 2.

For values of s greater than 3 the direct evaluation of d^sv^{s+1}/dx^s by successive differentiation becomes very tedious. It will be found much more practical to use the following procedure,[4] where D is a symbol of operation, $A = A_0$ and $b = A_1$.

$$A_0^{-1/2}R_s = (1/s!)\left(\frac{d^sg^{-(s+1)/2}}{dx^s}\right)$$
$$= \left[\frac{dA^{-(s+1)/2}}{1!dA}\right]D^{s-1}b + \left[\frac{d^2A^{-(s+1)/2}}{2!dA^2}\right]D^{s-2}b^2 + \cdots$$
$$+ \left[\frac{d^{s-1}A^{-(s+1)/2}}{(s-1)!dA^{s-1}}\right]Db^{s-1} + \left[\frac{d^sA^{-(s+1)/2}}{s!dA^s}\right]b^s$$

or

$$R_s = A_0^{1/2}\sum_{m=1}^{m=s}\left[\frac{d^mA^{-(s+1)/2}}{m!dA^m}\right](D^{s-m}b^m). \tag{12}$$

The following equations give the details requisite for the formation of R_s to R_8 inclusive; A_s can be computed from equation (11).

$Db = A_2,\ D^2b = A_3,\ D^3B = A_4,\ D^4b = A_5,$

$D^5b = A_6,\ D^6b = A_7,\ D^7b = A_8,$

$Db^2 = 2A_1A_2,$

$D^2b^2 = 2A_1A_3 + A_2^2,$

$D^3b^2 = 2A_1A_4 + 2A_2A_3,$

$D^4b^2 = 2A_1A_5 + 2A_2A_4 + A_3^2,$

$D^5b^2 = 2A_1A_6 + 2A_2A_5 + 2A_3A_4,$

$D^6b^2 = 2A_1A_7 + 2A_2A_6 + 2A_3A_5 + A_4^2,$

$Db^3 = 3A_1^2A_2,$

$D^2b^3 = 3A_1^2A_3 + 3A_1A_2^2,$

$D^3b^3 = 3A_1^2A_4 + 6A_1A_2A_3 + A_2^3,$

[4] See DeMorgan's "Differential and Integral Calculus," 1842, page 328, art. **214.**

$D^4b^3 = 3A_1^2A_5 + 6A_1A_2A_4 + 3A_1A_3^2 + 3A_2^2A_3,$

$D^5b^3 = 3A_1^2A_6 + 6A_1A_2A_5 + 6A_1A_3A_4 + 3A_2^2A_4 + 3A_2A_3^2,$

$Db^4 = 4A_1^3A_2,$

$D^2b^4 = 4A_1^3A_3 + 6A_1^2A_2^2,$

$D^3b^4 = 4A_1^3A_4 + 12A_1^2A_2A_3 + 4A_1A_2^3,$

$D^4b^4 = 4A_1^3A_5 + 12A_1^2A_2A_4 + 6A_1^2A_3^2 + 12A_1A_2^2A_3 + A_2^4,$

$Db^5 = 5A_1^4A_2,$ $Db^6 = 6A_1^5A_2,$

$D^2b^5 = 5A_1^4A_3 + 10A_1^3A_2^2,$ $D^2b^6 = 6A_1^5A_3 + 15A_1^4A_2^2,$

$D^3b^5 = 5A_1^4A_4 + 20A_1^3A_2A_3 + 10A_1^2A_2^3,$ $Db^7 = 7A_1^6A_2.$

To illustrate the use of the procedure given above, let us evaluate R_4. We have

$$A_0^{-1/2}R_4 = \left(\frac{dA^{-5/2}}{1!dA}\right)D^3b + \left(\frac{d^2A^{-5/2}}{2!dA^2}\right)D^2b^2 + \left(\frac{d^3A^{-5/2}}{3!dA^3}\right)Db^3 + \left(\frac{d^4A^{-5/2}}{4!dA^4}\right)b^4$$
$$= -(5/2)A_0^{-7/2}(A_4) + (1/2)(5/2)(7/2)A_0^{-9/2}(2A_1A_3 + A_2^2)$$
$$-(1/6)(5/2)(7/2)(9/2)A_0^{-11/2}(3A_1^2A_2)$$
$$+(1/24)(5/2)(7/2)(9/2)(11/2)A_0^{-13/2}A_1^4$$

or

$$R_4 = (5/2)A_0^{-6}[-A_0^3A_4 + (7/2)A_0^2(A_1A_3 + A_2^2/2)$$
$$-(1/2)(7/2)(9/2)A_0A_1^2A_2$$
$$+(1/24)(7/2)(9/2)(11/2)A_1^4].$$

Tables Indicating Degree of Accuracy Obtainable by Use of Formula (3) for Evaluating

$$P = \sum_{x=c}^{x=n} \binom{n}{x} p^x(1 - p)^{n-x}.$$

P_1 = 1st approximation, $\Delta_1 = (P - P_1)10^6$,

P_2 = 2d approximation, $\Delta_2 = (P - P_2)10^6$,

P_3 = 3d approximation, $\Delta_3 = (P - P_3)10^6$,

$a = np,$

$$T^2 = (n-1)\log\frac{n}{n-1} + (c-1)\log\frac{c-1}{a} + (n-c)\log\frac{n-c}{n-a},$$

$$I = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{T} e^{-t^2}dt.$$

TABLE I

c	T	$I(10^6)$	$P(10^6)$	Δ_1	Δ_2	Δ_3
			$a = 1.5,\ n = \infty,\ p = 0$			
2	+ .3074653	668154	442174	15991	9518	−1348
4	− .7612106	140849	65643	10816	1989	30
6	−1.5874105	12386	4456	1641	303	8
8	−2.2985028	577	170	103	20	0
			$a = 1.5,\ n = 30,\ p = .05$			
2	+ .2865166	657333	446458	21266	17382	−2083
4	− .8219430	122536	60772	3684	5617	− 97
6	−1.6966449	8211	3282	− 113	914	40
8	−2.4640017	246	85	− 24	50	7

TABLE II

c	T	$I(10^6)$	$P(10^6)$	Δ_1	Δ_2	Δ_3
			$a = 5,\ n = \infty,\ p = 0$			
2	1.5461442	985613	959576	−1681	2590	−798
4	.6837566	833222	734978	−2036	1897	−138
6	.0000000	500000	384044	2986	1036	− 4
8	− .5960752	199621	133376	4219	616	17
10	−1.1358169	54105	31832	2128	286	13
12	−1.6349406	10385	5452	604	123	2
14	−2.1027717	1471	692	107	11	− 5
16	−2.5454242	159	68	14	1	− 1
			$a = 5,\ n = 100,\ p = .05$			
2	1.5596227	986295	962920	− 373	3331	−732*
4	.6839234	833281	742162	639	3248	−141
6	− .0162780	490817	384001	2889	2108	− 38
8	− .6310024	186097	127961	1989	1431	− 8
10	−1.1912234	46029	28188	566	703	1
12	−1.7124507	7723	4274	75	208	2
14	−2.2138799	914	463	2	38	0
16	−2.6715388	79	37	− 1	5	0
			$a = 5,\ n = 50,\ p = .1$			
2	1.5742756	987005	966214	830	3962	−724
4	.6844600	833471	749706	3289	4437	−212
6	− .0335708	481067	383877	2599	3022	− 62
8	− .6689826	172053	122145	− 334	2076	31
10	−1.2524740	38258	24538	− 789	954	51
12	−1.7994619	5467	3220	− 278	244	24
14	−2.3191412	520	285	− 48	36	6
16	−2.8175965	34	17	− 5	3	0

* $|P - P_3| > |P - P_1|$.

TABLE III

c	T	I(10⁶)	P(10⁶)	Δ1	Δ2	Δ3
			$a = 10, n = \infty, p = 0$			
2	2.5879363	999874	999499	− 47	67	−37
4	1.8406742	995381	989662	− 533	275	−53
6	1.2386541	960089	932912	−1532	518	−50
8	.7094191	842135	779778	−1585	547	−27
10	.2274981	626172	542069	79	425	− 8
12	− .2200272	377838	303223	1786	322	1
14	− .6408864	182375	135535	2077	239	4
16	−1.0401811	70640	48740	1374	146	4
18	−1.4215063	22199	14277	628	68	0
20	−1.7875189	5737	3454	216	25	1
22	−2.1402533	1236	699	58	7	0
24	−2.4813121	225	119	12	1	− 1
			$a = 10, n = 100, p = .1$			
2	2.6528972	999912	999679	− 3	71	−24*
4	1.8917619	996268	992164	− 15	432	−39*
6	1.2715533	963931	942424	278	1070	−45
8	.7213308	846163	793949	997	1349	−36
10	.2161911	620099	548710	1213	1179	−19
12	− .2564838	358406	296967	503	982	− 2
14	− .7042404	159638	123877	− 222	771	10
16	−1.1320595	54691	39891	− 376	470	13
18	−1.5434535	14526	10007	− 222	206	9
20	−1.9410214	3025	1979	− 80	66	4
22	−2.3267578	500	312	− 20	15	1
24	−2.7022383	66	40	− 4	2	0

* $|P - P_3| > |P - P_1|$.

TABLE IV

c	T	I(10⁶)	P(10⁶)	Δ1	Δ2	Δ3
			$a = 15, n = \infty, p = 0$			
4	2.6780004	999924	999788	− 18	10	− 4
6	2.1229551	998660	997207	− 141	54	− 9
8	1.6324888	989520	981998	− 525	146	− 14
10	1.1843012	953019	930147	−1066	242	− 15
12	.7670044	860975	815249	−1198	274	− 10
14	.3737500	701445	636783	− 515	245	− 4
16	.0000000	500000	431911	582	203	0
18	− .3574541	306598	251141	1311	168	2
20	− .7009899	160758	124781	1351	132	2
22	−1.0324325	72134	53106	961	89	2
24	−1.3532229	27826	19464	523	49	1
26	−1.6645241	9287	6184	228	22	0
28	−1.9672925	2700	1715	82	8	− 1
30	−2.2623270	689	418	25	2	0

TABLE IV—*Continued*

$a = 15, n = 30, p = .5$

6	2.6019552	999883	999837	52	24	−12
8	2.0184138	997845	997388	559	152	−77
10	1.4626537	980704	978613	2676	411	−239
12	.9237039	904277	899756	5946	488	−343
14	.3942720	711436	707667	5025	227	−205
16	−.1313195	426335	427768	−1916	−73	72
18	−.6581761	175978	180797	−6392	−382	301
20	−1.1915875	45979	49369	−4405	−497	316
22	−1.7378702	6991	8062	−1346	−278	149
24	−2.3057782	555	715	−191	−69	32
26	−2.9097701	19	30	−11	−7	3

TABLE V

c	T	$I(10^6)$	$P(10^6)$	Δ_1	Δ_2	Δ_3
	$a = 25, n = \infty, p = 0$					
10	2.6086661	999888	999778	−11	3	−1
12	2.2291734	999191	998583	−51	11	−2
14	1.8705496	995920	993531	−159	30	−4
16	1.5289263	984700	977705	−364	59	−6
18	1.2015564	955365	939522	−617	90	−7
20	.8863972	894998	866422	−765	109	−7
22	.5818753	794717	752697	−651	110	−7
24	.2867455	657452	606120	−253	101	−6
26	.0000000	500000	447076	268	92	−3
28	−.2791919	346482	299814	678	84	1
30	−.5515253	217703	182105	837	75	2
32	−.8175896	123790	100070	761	62	3
34	−1.0778902	63708	49782	561	45	3
36	−1.3328647	29718	22460	350	29	2
38	−1.5828952	12593	9212	189	17	2
40	−1.8283181	4860	3445	90	9	2
42	−2.0694313	1713	1178	38	4	1
44	−2.3065005	553	370	15	2	1
	$a = 25, n = 50, p = .5$					
14	2.3698187	999598	999531	80	16	−9
16	1.9447371	997023	996699	404	57	−32
18	1.5274793	984621	983580	1329	126	−74
20	1.1159208	942735	940539	2844	170	−113
22	.7083182	841759	838881	3763	140	−109
24	.3031406	665931	664094	2415	60	−56
26	−.1010188	443200	443862	−873	−20	19
28	−.5055162	237333	239944	−3426	−102	87
30	−.9117246	98634	101319	−3499	−166	117
32	−1.3211006	30859	32454	−2056	−157	95
34	−1.7352770	7063	7673	−774	−91	50
36	−2.1561545	1147	1301	−191	−33	17
38	−2.5860897	128	153	−31	−8	4