УДК 519.14

Г. С. АБРАМОВ, И. М. АБРАМОВ
Херсонский Национальный технический университет
Херсонская специализированная школа I-III ступеней №30, МАН Украины

НОРМАЛЬНОЕ ПРИБЛИЖЕНИЕ ДЛЯ РАСПРЕДЕЛЕНИЯ ЧИСЕЛ В СТРОКАХ ТРЕУГОЛЬНИКА ПАСКАЛЯ

Цель статьи состоит в нахождении нормального приближения для чисел в строках треугольника Паскаля. Найдены зависимости от номера строки параметров нормального распределения, аппроксимирующего распределение чисел в строках треугольника Паскаля. Показано, что гауссовское приближение совпадает с факториальным представлением биномиальных коэффициентов, если факториалы вычислять по формуле Стирлинга. Получены уточнения гауссовского приближения, позволяющие с высокой степенью точности вычислять центральные элементы чётных строк треугольника Паскаля.

Ключевые слова: треугольник Паскаля, биномиальные коэффициенты, нормальное (гауссовское) распределение.

G.S. ABRAMOV, I.M. ABRAMOV
Kherson National Technical University
Kherson specialized school I-III stages № 30, MAS Ukraine

NORMAL APPROXIMATION FOR DISTRIBUTION OF NUMBERS IN THE ROW OF PASCAL'S TRIANGLE

Annotation

The purpose of this paper is to find the normal approximation for the number of rows in Pascal's triangle. Found depending on the row of parameters of the normal distribution that approximates the distribution of the numbers in the rows of Pascal's triangle. It is shown that a gaussian approximation coincides with the factorial representation of the binomial coefficients, factorials if calculated by Stirling's formula. Sharpen gaussian approximation, allowing a high degree of accuracy to calculate the central elements of the even rows of Pascal's triangle.

Keywords: Pascal's triangle, binomial coefficients, normal (gaussian) distribution.

Постановка проблемы. В книге Мартина Гарднера «Математические новеллы» [1] можно встретить замечательные слова: “Треугольник Паскаля так прост, что выписать его сможет даже десятилетний ребёнок. В то же время он таит в себе неисчерпаемые сокровища и связывает воедино различные аспекты математики, не имеющие на первый взгляд между собой ничего общего. Столь необычные свойства позволяют считать треугольник Паскаля одной из наиболее изящных схем во всей математике”. По признанию же самого Паскаля, в своём “Трактате” он вынужден был обойти молчанием многие свойства треугольника (сам Паскаль называл его triangle aritmetique – арифметическим треугольником). “Поистине удивительно, – восклицает Паскаль, – сколь неисчерпаемы эти свойства!”. И в самом деле, тот, кто заинтересованно займётся изучением треугольника Паскаля, откроет многие удивительные его свойства.

Поэтому изучению свойств и применению треугольника Паскаля могут быть посвящены последующие работы, так как его числа удовлетворяют, по-видимому, практически бесконечному числу тождеств и свойств [2]. Например, наше внимание обратило на себя то обстоятельство, что распределение чисел в нижних строках треугольника Паскаля (начиная с седьмой и ниже) очень напоминают нормальное (гауссовское) распределение, часто встречающееся в теории вероятностей. Действительно, в литературе можно встретить утверждение о том, что распределение чисел в строках треугольника Паскаля приближается к нормальному. Однако конкретных сведений о виде этого распределения, о его числовых характеристиках (параметрах), об изменении этих параметров в зависимости от номера строки в треугольнике Паскаля, о степени приближения к нормальному распределению найти не удалось.

Анализ публикации. Треугольнику Паскаля посвящена довольно обширная литература [1-8], которая в основном имеет научно-популярную направленность, однако имеются две фундаментальные монографии [2, 3], в которых биномиальным коэффициентам посвящены отдельные главы. В книге Р. Грэхема, Д. Кнута и О. Паташника “Конкретная математика. Основание информатики” [2] в главе “Биномиальные коэффициенты” рассматриваются доказательства многочисленных тождеств и изящные способы вычисления различных частных сумм, а в монографии В. Феллера “Введение в теорию вероятностей и её приложения” [3] рассматривается нормальное приближение для биномиального распределения, однако погрешность полученных результатов (1-2%) нельзя признать удовлетворительной.

Формулировка целей статьи. Целью данной работы было установление связи распределения чисел в строках треугольника Паскаля с нормальным распределением и нахождение высокоточного приближения для центральных элементов чётных строк треугольника Паскаля.

Конкретно же ставились следующие задачи: определение числовых характеристик распределения в строках треугольника Паскаля; выявление зависимости изменения этих параметров от номера строки в треугольнике Паскаля; сравнение полученных распределений с нормальным; исследование асимптотики приближения распределения чисел в строках треугольника Паскаля к нормальному при увеличении номера строки; уточнение нормального приближения для центральных элементов чётных строк треугольника Паскаля.

При решении этих задач использовались возможности Microsoft Office Excel.

Основная часть. В табл. 1 представлен фрагмент треугольника Паскаля (6-16-я строки), иллюстрирующий вычисление числовых характеристик распределения в строке, как дискретной случайной величины. Реально же для вычисления числовых характеристик распределения в строке рассчитывали вплоть до 50-й строки треугольника Паскаля.

Таблица 1

Фрагмент треугольника Паскаля (6-16-я строки), иллюстрирующий вычисление числовых характеристик (*) распределения в строке, как дискретной случайной величины

Номер элемента строки, x_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	…	n+1
Номер строки, n																	
…																	
6	1	6	15	20	15	6	1										
7	1	7	21	35	35	21	7	1									
8	1	8	28	56	70	56	28	8	1								
9	1	9	36	84	126	126	84	36	9	1							
10	1	10	45	120	210	252	210	120	45	10	1						
11	1	11	55	165	330	462	462	330	165	55	11	1					
12	1	12	66	220	495	792	924	792	495	220	66	12	1				
13	1	13	78	286	715	1287	1716	1716	1287	715	286	78	13	1			
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001	364	91	14	1		
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003	1365	455	105	15	1	
16	1	16	120	560	1820	4368	8008	11440	12870	11440	8008	4368	1820	560	120	16	1
… Значение элемента n_i	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}	n_{11}	n_{12}	n_{13}	n_{14}	n_{15}	…	n_{n+1}

Математическое ожидание ($\overline{x}$) и дисперсию (S^2) распределения чисел в строках рассчитывали следующим образом:

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{k} x_i n_i \,, \; S^2 = \frac{1}{N}\sum_{i=1}^{k} x_i^2 n_i - \overline{x}^2 \,, \qquad (*)$$

где x_i – номер позиции (элемента) в строке треугольника Паскаля;
n_i – значение элемента на i-й позиции в строке треугольника Паскаля;
$k = n+1$– число элементов в n-й строке;
N – сумма всех элементов в данной строке треугольника Паскаля: $N=2^n$

Результаты вычислений приведены в таблице 2.

Таблица 2

Математические характеристики распределения чисел в строках треугольника Паскаля

n	$\overline{x}$	S^2	S	n	$\overline{x}$	S^2	S
3	2,5	0,75	0,866	13	7,5	3,25	1,803

4	3	1	1	14	8	3,5	1,871
5	3,5	1,25	1,118	15	8,5	3,75	1,936
6	4	1,5	1,225	16	9	4	2
7	4,5	1,75	1,323	17	9,5	4,25	2,062
8	5	2	1,414	18	10	4,5	2,121
9	5,5	2,25	1,5	19	10,5	4,75	2,179
10	6	2,5	1,581	20	11	5	2,236
11	6,5	2,75	1,658	21	11,5	5,25	2,291
12	7	3	1,732	22	12	5,5	2,345

Результаты вычисления математического ожидания оказались вполне прогнозированными: они равны значению x_i, соответствующего максимальному элементу в строке треугольника Паскаля (если *n* - чётное), или среднему арифметическому между координатами двух максимумов, если номер строки – нечётное число.

Легко установить зависимость матожидания $\bar{x}$, дисперсии S^2 и среднеквадратичного отклонения S от номера строки в треугольнике Паскаля:

$$\bar{x} = \frac{n+2}{2}; \quad S^2 = \frac{n}{4}; \quad S = \frac{\sqrt{n}}{2}.$$

Построим гауссовские распределения для соответствующих значений матожидания и среднеквадратичного отклонения и сравним полученные гауссовские кривые с фактическим частотным распределением в соответствующих строках треугольника Паскаля. Для этого были вычислены относительные частоты распределения чисел в строках треугольника Паскаля, то есть значение данного элемента строки делилось на сумму всех чисел строки.

На рис. 1 приведены соответствующие эмпирические распределения (основанные на вычисленных частотах) в сравнении с теоретическими распределениями (гауссовские кривые) для строк треугольника Паскаля от 8-й до 22-й. Видно, что с увеличением номера строки распределение чисел в строке треугольника Паскаля «нормализуется», то есть всё ближе приближается к кривой нормального распределения. Это даёт основания считать, что нормальное распределение является хорошей аппроксимацией распределения в *n*-й строке треугольника Паскаля (очевидно, тем лучше, чем больше номер строки n): $x_n \in \mathrm{N}(\frac{n+2}{2}; \frac{\sqrt{n}}{2})$.

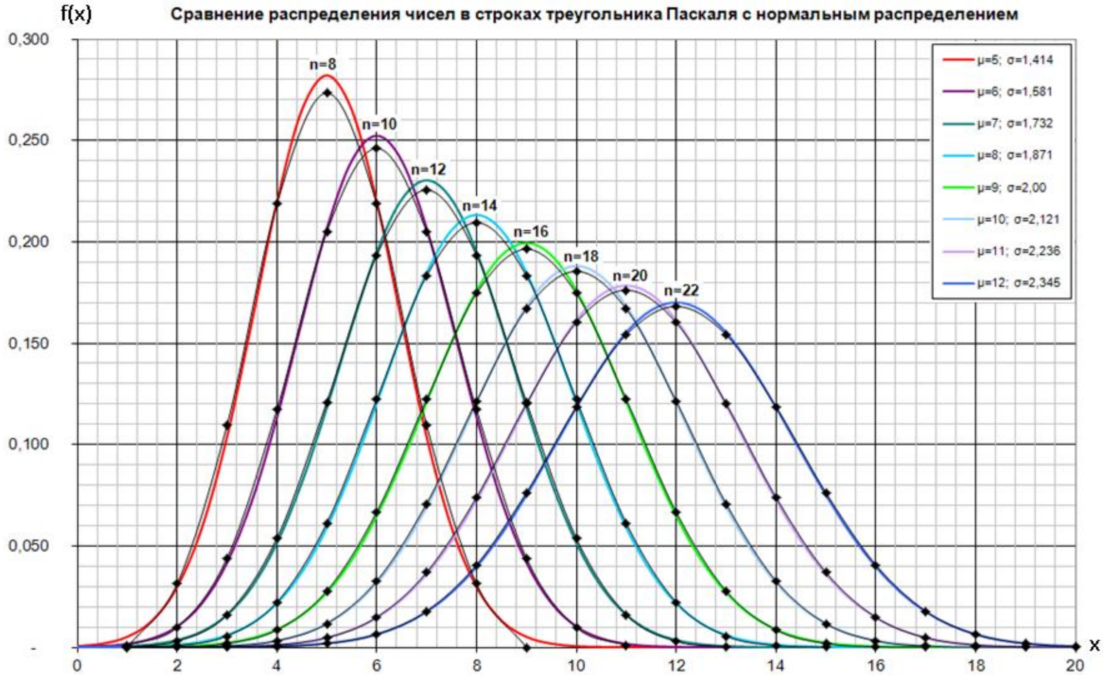

Рис. 1. Сравнение теоретических гауссовских кривых и эмпирических распределений в соответствующих строках треугольника Паскаля.

Из рис. 1 следует, что есть отличия между распределением в строке Паскаля и нормальным распределением, которые тем заметнее, чем меньше номер строки *n*. Эти отличия связаны с «деформацией» кривой по оси ординат. Максимальные значения распределения в строках треугольника Паскаля чуть меньше максимума на соответствующих гауссовских кривых.

Известно, что искажения нормального распределения связаны с моментами третьего и четвёртого порядка, которые называются, соответственно, асимметрией и эксцессом. Асимметрия, описывающая искажения вдоль оси абсцисс, в нашем случае отсутствует, так как распределение в строках треугольника Паскаля абсолютно симметрично. Следовательно, для описания наблюдаемых различий паскалевского распределения от гауссовского, необходимо вычислять значения эксцесса для распределения в каждой строке треугольника Паскаля. Очевидно, что вычисляемые значения должны быть отрицательными (так как "уплощение" гауссовской кривой характеризуется отрицательным значением эксцесса), и с возрастанием номера строки значение эксцесса по абсолютной величине должно уменьшаться.

Эксцесс (E_x) вычисляли с помощью центрального момента 4-го порядка – μ^4:

$$\mu^4 = \frac{\sum_{i=1}^{n+1}(x_i - \bar{x})^4 n_i}{\sum_{i=1}^{n+1} n_i} = \frac{\sum_{i=1}^{n+1}(x_i - \bar{x})^4 n_i}{2^n} \quad ; \qquad E_x = \frac{\mu^4}{\sigma^4} - 3 .$$

Рис. 2. Значение эксцесса для распределения в строках треугольника Паскаля

На рис. 2 приведены значения эксцесса в строках треугольника Паскаля (от 10-й до 50-й включительно). Видно, что, действительно, значения эксцесса отрицательны, и с увеличением номера строки асимптотически приближаются к нулю, то есть для достаточно больших номеров строк можно с большой степенью точности утверждать, что распределение в строке треугольника Паскаля стремится к гауссовскому.

В качестве иллюстрации последнего утверждения на рис. 3 приведены графики гауссовского и паскалевского распределений для 50-й строки треугольника.

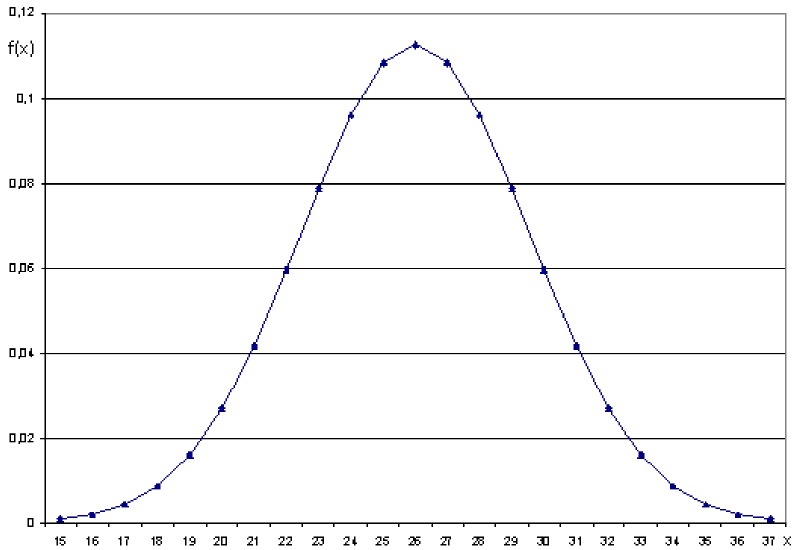

Рис. 3. Сравнение гауссовского распределения (кривая) и паскалевского (точки) для 50-й строки треугольника Паскаля.

Обобщая вышесказанное, можно утверждать, что гауссовское распределение является весьма хорошей аппроксимацией для паскалевского распределения в строках треугольника. Математически плотность распределения в *n*-й строке треугольника Паскаля выглядит следующим образом:

$$f_n(x) \approx \frac{2}{\sqrt{2\pi n}} \exp\left(-\frac{\left(x - \frac{n+2}{2} \right)^2}{\frac{2n}{4}} \right) = \sqrt{\frac{2}{\pi n}} \exp\left(-\frac{2\left(x - \frac{n+2}{2} \right)^2}{n} \right). \tag{1}$$

Таким образом, мы установили значения параметров нормального распределения, которым можно аппроксимировать распределение в строках треугольника Паскаля. С увеличением номера строки точность аппроксимации быстро возрастает: при $n \geq 50$ значения эксцесса практически равно нулю и совпадение с гауссовским распределением становится практически идеальным (различия наблюдаются в 4-ом–5-ом знаке после запятой).

В качестве практического применения полученной формулы (1) для плотности аппроксимирующего гауссовского распределения проведём приближённые вычисления элементов строки треугольника Паскаля (что может быть актуально для больших номеров строк).

Число сочетаний C_n^m в нашем случае является *(m+1)*-м элементом *n*-ой строки треугольника Паскаля (т. к. отсчёт величины *x* мы начали не с нуля, а с единицы).

Вычислим значение аргумента экспоненты в формуле плотности *f(x)* при *x=m+1*:

$$\frac{2\left(x - \frac{n+2}{2} \right)^2}{n} = \frac{(2x-n-2)^2}{2n} \Big|_{x=m+1} = \frac{(2m-n)^2}{2n}. \tag{2}$$

Тогда число сочетаний C_n^m можно вычислить по следующей приближённой формуле:

$$C_n^m \approx 2^n \sqrt{\frac{2}{\pi n}} \exp\left(-\frac{(2m-n)^2}{2n} \right). \tag{3}$$

Наиболее просто с помощью этой формулы вычисляется значение центральных элементов чётных строк треугольника Паскаля, когда *m=n/2*, (или, что то же самое, *n*=2*m*). В этом случае, аргумент экспоненты обнуляется и формула (3) приобретает простой вид:

$$C_n^{m=\frac{n}{2}} \approx \frac{2^n}{\sqrt{\pi m}}. \tag{4}$$

Можно показать, что приближение (4) точно соответствует приближённому вычислению числа сочетаний, если факториалы вычислять с помощью формулы Стирлинга:

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e} \right)^n. \tag{5}$$

Это наиболее известная формула для приближённого вычисления факториала при больших *n*. Факториал n! можно вычислять ещё точнее, т. к. (5) является лишь первым членом ряда Стирлинга:

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e} \right)^n \left(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51840n^3} + \ldots \right). \tag{6}$$

Используя формулу Стирлинга (5) вычислим $C_n^{m=\frac{n}{2}}$:

$$C_n^{m=\frac{n}{2}} = \frac{n!}{(m!)^2} \approx \frac{\sqrt{2\pi n}\left(\frac{n}{e} \right)^{n=2m}}{2\pi m \left(\frac{m}{e} \right)^{2m}} = \begin{bmatrix} \left(\frac{2m}{e} \frac{e}{m} \right)^n = 2^n \\ \sqrt{\frac{4\pi m}{4\pi^2 m^2}} = \frac{1}{\sqrt{\pi m}} \end{bmatrix} = \frac{2^n}{\sqrt{\pi m}}. \tag{7}$$

Как видим, приближение с помощью формулы Стирлинга совпадает с приближением, которое мы получили с помощью гауссовской функции плотности, аппроксимирующей паскалевское распределение в строках треугольника Паскаля.

В работе проведено сравнение точных значений для центральных элементов чётных строк треугольника Паскаля с большими номерами и их приближённых значений, вычисленных с помощью гауссовской аппроксимации (табл. 3).

Таблица 3

Сравнение точных значений и приближённой оценки центральных элементов чётных строк треугольника Паскаля с большими номерами

Номер строки n	Точное значение $C_n^{m=\frac{n}{2}} = \frac{n!}{(m!)^2}$	Приближённое значение $C_n^{m=\frac{n}{2}} \approx \frac{2^n}{\sqrt{\pi m}}$	Относительная ошибка ε, %
30	$1{,}551175 \cdot 10^{8}$	$1{,}564153 \cdot 10^{8}$	0,84
50	$1{,}264106 \cdot 10^{14}$	$1{,}270442 \cdot 10^{14}$	0,50
60	$1{,}182646 \cdot 10^{17}$	$1{,}187583 \cdot 10^{17}$	0,42
100	$1{,}008913 \cdot 10^{29}$	$1{,}011439 \cdot 10^{29}$	0,25
200	$9{,}054851 \cdot 10^{58}$	$9{,}066177 \cdot 10^{58}$	0,125
400	$1{,}029525 \cdot 10^{119}$	$1{,}030169 \cdot 10^{119}$	0,0625
800	$1{,}88042442 \cdot 10^{239}$	$1{,}88101214 \cdot 10^{239}$	0,03125

Здесь относительная ошибка вычислялась как разность между приближённым и точным значением, отнесенная к точному значению.

Из данных табл. 3 следует, что относительная ошибка убывает с возрастанием номера строки по обратно пропорциональной зависимости:

$$\varepsilon = \frac{25}{n}\% . \tag{8}$$

Поскольку относительная ошибка формулы (4) (будем называть её приближением Стирлинга) обнаруживает обратно пропорциональную зависимость от номера строки – это позволяет вывести формулу для следующего, более точного приближения.

Запишем выражение для относительной ошибки:

$$\frac{\frac{2^n}{\sqrt{\pi m}} - C_n^{m=n/2}}{C_n^{m=n/2}} = \frac{0{,}25}{n} . \tag{9}$$

Из (9) следует приближение, более точно оценивающее $C_n^{m=n/2}$:

$$C_n^{m=n/2} \approx \frac{2^n}{\sqrt{\pi m}} \cdot \frac{n}{n+0{,}25} . \tag{10}$$

Таблица 4

Сравнение точных значений и приближённой оценки (10) центральных элементов чётных строк треугольника Паскаля с большими номерами

Номер строки	Точное значение $C_n^{m=\frac{n}{2}} = \frac{n!}{(m!)^2}$	Приближённое значение (10) $C_n^{m=n/2} \approx \frac{2^n}{\sqrt{\pi m}} \cdot \frac{n}{n+0{,}25}$	Относительная ошибка ε_1, %
30	$1{,}551175 \cdot 10^{8}$	$1{,}551226 \cdot 10^{8}$	0,0033
50	$1{,}264106 \cdot 10^{14}$	$1{,}264121 \cdot 10^{14}$	0,0012
60	$1{,}182646 \cdot 10^{17}$	$1{,}182655 \cdot 10^{17}$	0,0008
100	$1{,}0089134 \cdot 10^{29}$	$1{,}0089166 \cdot 10^{29}$	0,0003
200	$9{,}0548515 \cdot 10^{58}$	$9{,}0548578 \cdot 10^{58}$	0,00007
400	$1{,}029525 \cdot 10^{119}$	$1{,}029525 \cdot 10^{119}$	
800	$1{,}88042442 \cdot 10^{239}$	$1{,}88042442 \cdot 10^{239}$	

Видно, что и относительная ошибка этого приближения ε_1 также подчиняется функциональной зависимости, – на этот раз ошибка обратно пропорциональна квадрату номера строки:

$$\varepsilon_1 = \frac{3}{n^2}\% . \tag{11}$$

Это даёт возможность сформулировать ещё более точное приближение:

$$C_n^{m=n/2} \approx \frac{2^n}{\sqrt{\pi m}} \cdot \frac{n}{n+0{,}25} \cdot \frac{n^2}{n^2+0{,}03} . \tag{12}$$

Точность этой формулы настолько высока, что даже для сравнительно небольших n она даёт очень хорошую оценку. Так, например, для n=20 число сочетаний C_{20}^{10}= 184756, а формула (12) даёт результат 184755,5; т. е. относительная ошибка составляет 0,00027%.

Формула (12) может быть записана в следующем виде:

$$C_n^{m=n/2} \approx \frac{2^n}{\sqrt{\pi m}} \cdot \left(\frac{1}{1+\frac{0,25}{n}+\frac{0,03}{n^2}+\frac{0,0075}{n^3}} \right). \tag{13}$$

Поскольку начальное приближение вида (4) (приближение Стирлинга) даёт несколько завышенные значения, то с помощью выражения (13) можно оценить вклад каждого следующего слагаемого в знаменателе скобки в приближение к точному значению.

Имея достаточно точное значение центрального элемента C_{2n}^n чётной строки треугольника Паскаля, можно легко вычислить и остальные элементы строки:

$$C_{2n}^{n\pm k} = C_{2n}^n \cdot \frac{n}{n+1} \cdot \frac{n-1}{n+2} \cdot \ldots \cdot \frac{n+1-k}{n+k} \tag{14}$$

или, в более компактной записи,

$$C_{2n}^{n\pm k} = C_{2n}^n \prod_{i=1}^{k} \frac{n+1-i}{n+i}. \tag{15}$$

Выводы и перспективы дальнейших исследований. В работе найдены значения параметров гауссовского распределения, которое аппроксимирует паскалевское распределение в n-й строке треугольника.

Вычислены значения эксцесса, описывающего деформацию гауссовской кривой вдоль оси ординат, и показано асимптотическое приближение его величины к нулю при увеличении номера строки. Это означает, что при больших значениях номера строки n паскалевское распределение «нормализуется».

Проведены приближённые вычисления элементов строки треугольника Паскаля с помощью функции для плотности аппроксимирующего гауссовского распределения. Наиболее простой вид эта формула имеет для центральных элементов чётных строк треугольника Паскаля. Показано, что этот же результат получается, если факториалы для числа сочетаний вычислять с помощью приближённой формулы Стирлинга.

Проведены оценки относительной погрешности при вычислении по приближённой формуле и получена зависимость относительной погрешности от номера строки, на основе которой была выведена формула для следующего, более точного приближения для центральных элементов чётных строк треугольника Паскаля.

Перспективой дальнейших исследований может стать рассмотрение фрактальных свойств треугольника Паскаля, а также исследование моментов более высоких порядков (5-го, 6-го и следующих) и выяснение их роли и значения в паскалевском распределении. Предложенные методы вычисления биномиальных коэффициентов с высокой точностью могут быть актуальны при вычислении триномиальных (а в общем случае мультиномиальных) коэффициентов, так как они могут быть представлены с помощью произведений биномиальных коэффициентов.

Литература

1. Гарднер М. Математические новеллы / М. Гарднер. – М.: Мир, 1974. – 456 с.
2. Грэхем Р. О. Конкретная математика. Основание информатики: Пер.с англ./Р. Грэхем, Д. Кнут, О. Паташник. – М.: Мир, 1998. – 703 с.
3. Феллер В. Введение в теорию вероятностей и её приложения: Пер. с англ./В. Феллер. – М.: Мир, 1984. – 528 с.
4. Бондаренко Б. А. Обобщённые треугольники и пирамиды Паскаля, их фракталы, графы и приложения/ Б. А. Бондаренко. – Ташкент, 1990. – 192 с.
5. Кузьмин О. В. Некоторые комбинаторные числа в обобщённой пирамиде Паскаля/О. В. Кузьмин //Асимптотические и перечислительные задачи комбинаторного анализа. – Иркутск: Издательство Иркутского университета, 1998. – С. 90-100.
6. Кузьмин О. В. Обобщенные пирамиды Паскаля и их приложения / О. В. Кузьмин. – Новосибирск, 2000 – 64 с.
7. Успенский В. А. Треугольник Паскаля / В. А. Успенский – М.: Наука, 1979 – 48 с.
8. Энзензбергер Х. М. Дух числа / Х. М. Энзензбергер. – Харьков: Книжный клуб «Клуб сімейного дозвілля», 2002. – 272 с.