Turkish Journal of Analysis and Number Theory, 2020, Vol. 8, No. 5, 80-90
Available online at http://pubs.sciepub.com/tjant/8/5/1
Published by Science and Education Publishing
DOI:10.12691/tjant-8-5-1

Introduction of *p*-nomial Distribution as a Generalization of Binomial Distribution

Aziz ATTA*

Mathematics and Structural Analysis, Atta Engineering Design Office, El Jadida, Morocco
*Corresponding author: azizatta20@gmail.com

Received August 16, 2020; Revised September 17, 2020; Accepted September 24, 2020

Abstract The theory of probability and statistics, thanks to its continuous modernization, has become more and more important in our life given its presence in several fields such as economics and prevision [8]. The binomial distribution is among the oldest probability distributions introduced by Bernoulli [1]. In the same context, we thought of generalizing this probability distribution under the name *p-nomial distribution* using *p*-nomial coefficients *p*-nomial theorem [7]. In this article, we are going to be interested in the introduction of this new probability distribution as well as an establishment of its various standard characteristics. the purpose of this article is therefore summarized in the tracing of the theoretical framework with some examples of application of the said *p*-nomial distribution.

Keywords: *p-nomial coefficients, p-nomial identity, p-nomial distribution, probability tree, trinomial distribution*

Cite This Article: Aziz ATTA, "Introduction of *p*-nomial Distribution as a Generalization of Binomial Distribution." *Turkish Journal of Analysis and Number Theory*, vol. 8, no. 5 (2020): 80-90. doi: 10.12691/tjant-8-5-1.

1. Introduction

The binomial distribution is one of the oldest probability laws studied [1]. It was introduced by Jacques Bernoulli who referred to it in 1713 in his work *Ars Conjectandi*. Between 1708 and 1718, the multinomial distribution (multidimensional generalization of the binomial distribution), the negative binomial distribution as well as the approximation of the binomial distribution by the Poisson's distribution, the law of large numbers for the binomial distribution and an approximation of the tail of the binomial distribution are discovered [2].

Thanks to the expression of its mass function, the binomial distribution has been used by several scientists to perform calculations in concrete situations. This is the case of Abraham de Moivre who succeeds in finding an approximation of the binomial distribution by the normal distribution. In 1812, Pierre-Simon de Laplace resumed this work. Francis Galton creates the Galton plate which allows a physical representation of this convergence [3]. In 1909, Émile Borel states and proves, in the case of binomial law, the first version of the strong law of large numbers [4].

Binomial law appeared in many applications in the 20th century [5]: in genetics, animal biology, plant ecology, for statistical tests, in different physical models such as telephone networks or the Ehrenfest's urn model, etc.

The name "binomial" of this law comes from [6] the writing of its mass function (see below) which contains a binomial coefficient resulting from the development of the binomial $(p+q)^n$. Indeed, if X is a random variable following a binomial law $B(n,p)$ of characteristics n and p, then: $\forall k \in [\![0,n]\!]$,

$$\mathrm{P}(X=k) = C_n^k p^k q^{n-k} = C_n^k p^k (1-p)^{n-k}$$

In the same sense, I thought of generalizing the binomial coefficients by calling them *p*-nomial coefficients; this is based on a generalization of Pascal's formula. After this definition, I introduced the notion of the *p*-nomial law as a generalization of the binomial distribution in probability.

In this article, we'll recall the *p*-nomial coefficients that we have already defined [7]. We'll also present the properties and characteristics of this distribution as well as some examples of application. Finally, we'll be interested in the particular case of trinomial distribution.

Note that this article aims to introduce this new distribution and show its importance through a few examples. This introduction will contain as expected an establishment of the characteristics of this distribution such as its expectation, its variance and its moments.

2. *p*-nomial Coefficients

In this part, we will quickly recall the definition of *p*-nomial coefficients, their expression as well as the *p*-nomial theorem. For that, the reader interested in more details on these concepts can consult the article cited in the reference [7].

2.1. Definition

Let p be a non-zero natural integer. We define p-nomial coefficient and we note $\pi^k_{n|p-1\rangle}$ the k^{th} p-nomial coefficient among $(p-1)n$ by the following recurrent relation:

$$\forall n\ge 0\ \ \forall k\in[\![0,(p-1)(n+1)]\!]\ \begin{cases}\pi^0_{0|p-1\rangle}=\pi^0_{1|p-1\rangle}=..=\pi^{p-1}_{1|p-1\rangle}=1\\ \pi^k_{(n+1)|p-1\rangle}=\pi^k_{n|p-1\rangle}+..+\pi^{k-p+2}_{n|p-1\rangle}+\pi^{k-p+1}_{n|p-1\rangle}\end{cases}$$

Such that $\pi^q_{n|p\rangle}=0$ if $q<0$ or $q>np$.

2.2. Expressions

Using the fundamental theorem [7], we can establish an expression for the p-nomial coefficients:

$$\pi^q_{q_0|0\rangle}=1\ ;\ \pi^q_{q_0|1\rangle}=C^q_{q_0};\quad \forall p\ge 1:$$

$$\pi^q_{q_0|p+1\rangle}=\sum_{q_1=\left[\frac{q+p}{p+1}\right]}^{\min(q_0,q)}\ \sum_{q_2=\left[\frac{q-q_1+p-1}{p}\right]}^{\min(q_1,q-q_1)}..\sum_{q_p=\left[\frac{q-\sum_{i=1}^{p-1}q_i+1}{2}\right]}^{\min\left(q_{p-1},\,q-\sum_{i=1}^{p-1}q_i\right)} C_{q_p}^{q-\sum_{i=1}^{p} q_i}\left(\prod_{i=1}^{p} C^{q_i}_{q_{i-1}}\right)$$

To simplify the writing, we introduce the symbol $\coprod$ to group the summations. For $p\ge 1$:

$$\pi^q_{q_0|p+1\rangle}=\coprod_{j=0}^{p-1}\ \sum_{q_{j+1}=\left[\frac{q-\sum_{i=1}^{j} q_i+p-j}{p-j+1}\right]}^{\min\left(q_j,\,q-\sum_{i=1}^{j} q_i\right)}\left(\prod_{i=1}^{p} C^{q_i}_{q_{i-1}}\right)C_{q_p}^{q-\sum_{i=1}^{p} q_i}$$

$$=\coprod_{j=0}^{p-1}\ \sum_{q_{j+1}=\left[\frac{q-\sum_{i=1}^{j} q_i+p-j}{p-j+1}\right]}^{\min\left(q_j,\,q-\sum_{i=1}^{j} q_i\right)}\binom{q_0}{q_0-q_1,..,q_{p-1}-q_p,\,q-\sum_{i=1}^{p}q_i,\,q_p+\sum_{i=1}^{p}q_i-q}$$

So, the p-nomial coefficient $\pi^q_{q_0|p-1\rangle}$ is a combination of multinomial coefficients [13].

2.2. *p*-nomial Theorem

Let p a non-zero natural integer. We have following equality [7]: $\forall p\in\mathbb{N}^*\quad \forall n\in\mathbb{N}\quad \forall(x_1,x_2)\in\mathbb{C}^2$

$$\left(x_1^p+x_1^{p-1}x_2+..+x_1x_2^{p-1}+x_2^p\right)^n$$
$$=\sum_{k=0}^{np}\pi^k_{n|p\rangle}x_1^k x_2^{np-k}$$

3. *p*-nomial Distribution

3.1. Definition

Let n and p two non-zero natural integers. A random variable X follow a $(p+1)-$nomial distribution $B_p(n,q)$ where $0<q<1$ if:

$$\forall k\in[\![0,np]\!]\quad \mathrm{P}(X=k)=\pi^k_{n|p\rangle}\frac{q^k(1-q)^{np-k}}{\left(\sum_{j=0}^{p}q^j(1-q)^{p-j}\right)^n}$$

Indeed, this distribution is well defined:

$$\sum_{j=0}^{p} q^j (1-q)^{p-j} = \begin{cases} \dfrac{p+1}{2^p} & \text{if } q = \dfrac{1}{2} \\ \dfrac{q^{p+1} - (1-q)^{p+1}}{2q-1} & \text{if } q \neq \dfrac{1}{2} \end{cases}$$

Which shows that: $\forall q \in]0,1[\quad \sum_{j=0}^{p} q^j (1-q)^{p-j} \neq 0$.

Moreover, using p-nomial theorem:

$$\forall n \in \mathbb{N} \quad \sum_{k=0}^{np} \mathrm{P}(X = k)$$

$$= \sum_{k=0}^{np} \pi_{n|p\rangle}^{k} \frac{q^k (1-q)^{np-k}}{\left(\sum_{j=0}^{p} q^j (1-q)^{p-j} \right)^n}$$

$$= \frac{\sum_{k=0}^{np} \pi_{n|p\rangle}^{k} q^k (1-q)^{np-k}}{\left(\sum_{j=0}^{p} q^j (1-q)^{p-j} \right)^n} = 1$$

3.2. Probability Tree of the (p+1)-nomial Distribution

We consider p identical coins noted $c_1,..,c_p$. We will launch the p coins at the same time for n independent trials. Each outcome has a fixed probability, the same from trial to trial. Also, the probability to have a head for each coin is P_H and to have a tail is $P_T = 1 - P_H$.

For each trial, we count the number of heads and tails. To simplify, if we get k heads $(k \in [\![0,p]\!])$, we will write the result as follows $(kH,(p-k)T)$ where H means heads and T means tails.

We note T_k the k$^{\text{th}}$ trial. The possible results for each trial are $(0H, pT)$, $(1H,(p-1)T),..,(pH,0T)$. We represent below the probability tree:

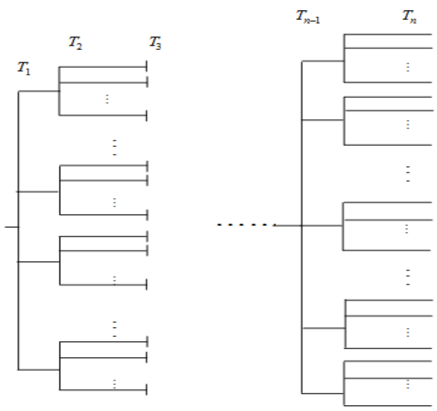

Figure 1. Probability tree of the (p + 1)-nomial distribution

We count, after n trials, the total number of heads obtained. If we denote by X this random variable then the possible values of X are $0,1,..,np$. In addition, if we note $q\left(q \in [0,1]\right)$ the probability of having a head $(P_H = q)$, then the random variable X follows a $(p+1)-$nomial distribution noted $B_p(n,q)$ such that :

$$\forall k \in [\![0,np]\!]$$

$$\mathrm{P}(X = k) = \pi_{n|p\rangle}^{k} \frac{q^k (1-q)^{np-k}}{\left(\sum_{j=0}^{p} q^j (1-q)^{p-j} \right)^n}$$

We note: $X \sim B_p(n,q)$.

3.3. Probability Measure

Since the $(p+1)-$nomial $B_p(n,q)$ is a discrete distribution, it's possible to define it using its probability measure:

$$\mathrm{P} = \sum_{k=0}^{np} \pi_{n|p\rangle}^{k} \frac{q^k (1-q)^{np-k}}{\left(\sum_{j=0}^{p} q^j (1-q)^{p-j} \right)^n} \delta_k$$

where δ_k is the measure of Dirac at point k [12].

3.4. Expectation

We denote:

$$\begin{cases} Q_p(q,j) = q^p + q^{p-1}(1-j) + .. + q(1-q)^{p-1} + (1-j)^p \\ Q_p = Q_p(q,q) \end{cases}$$

The Esperance of the (p+1)-nomial distribution is:

$$E_p(X) = nq\left(p + (1-q)\frac{1}{Q_p}\frac{dQ_p}{dq} \right)$$

$$= E_1(X) \frac{\sum_{j=1}^{p} j q^{j-1} (1-q)^{p-j}}{\sum_{j=0}^{p} q^j (1-q)^{p-j}}$$

Proof. We proceed as follows:

$$Q_p{}^n E_p(X) = \sum_{k=0}^{np} k \pi_{n|p\rangle}^{k} q^k (1-q)^{np-k}$$

$$= q\left[\frac{d\left(Q_p{}^n - (1-q)^{np}\right)}{dq} + \sum_{k=1}^{np-1} (np-k) \pi_{n|p\rangle}^{k} q^k (1-q)^{np-k-1} \right]$$

$$\frac{Q_p{}^n}{1-q}E_p(X) = nq\left[\begin{array}{c}\dfrac{d\left(Q_p{}^n-(1-q)^{np}\right)}{dq}\\ +\dfrac{p}{1-q}\left(Q_p{}^n-(1-q)^{np}\right)\end{array}\right]$$

After simplifications:

$$Q_p{}^n E_p(X) = nq\left[(1-q)Q_p{}^{n-1}\frac{dQ_p}{dq}+pQ_p{}^n\right]$$

$$E_p(X) = nq\left(p+(1-q)\frac{1}{Q_p}\frac{dQ_p}{dq}\right)$$

We can easily deduce that:

$$Q_p{}^n E_p(X) = q\left(\frac{\partial}{\partial q}\left(Q_p(q,j)^n\right)\right)_{j=q}$$

$$= nqQ_p{}^{n-1}\left(\sum_{k=1}^{p}kq^{k-1}(1-q)^{p-k}\right)$$

$$E_p(X) = nq\frac{\sum_{k=1}^{p}kq^{k-1}(1-q)^{p-k}}{Q_p(q)}$$

$$= E_1(X)\frac{\sum_{k=1}^{p}kq^{k-1}(1-q)^{p-k}}{\sum_{k=0}^{p}q^k(1-j)^{p-k}}$$

For $p=1$, we have $Q_1 = q+1-q = 1$ and $\frac{dQ_p}{dq} = 0$. We find $E_1(X) = nq$: Esperance of binomial distribution.

For $p=2$, $Q_2 = q^2-q+1$ and $\frac{dQ_p}{dq} = 2q-1$. We find $E_2(X) = \frac{nq(q+1)}{q^2-q+1}$: Esperance of trinomial distribution.

3.5. Variance

Theorem. For $p \in \mathbb{N}^*$, the variance of the $(p+1)-$nomial distribution is given by the formula below:

$$V_p(X) = V_1(X)\frac{P_{2p-2}(q)}{Q_p(q)^2}$$

Where is a P_{2p-2} polynomial of degree $2p-2$.

Proof. First, calculate the Esperance of X^2 :

$$Q_p(q)^n E_p\left(X^2\right) = \sum_{k=0}^{np}k^2\pi_{n|p\rangle}^{k}q^k(1-q)^{np-k}$$

$$= q^2\frac{\partial^2}{\partial q^2}\left(\sum_{k=0}^{np}\pi_{n|p\rangle}^{k}q^k(1-j)^{np-k}\right)\Bigg|_{j=q}$$

$$+q\frac{\partial}{\partial q}\left(\sum_{k=0}^{np}\pi_{n|p\rangle}^{k}q^k(1-j)^{np-k}\right)\Bigg|_{j=q}$$

$$= q\left(q\frac{\partial^2}{\partial q^2}\left(Q_p(q,j)^n\right)\Big|_{j=q}+\frac{\partial}{\partial q}\left(Q_p(q,j)^n\right)\Big|_{j=q}\right)$$

$$= q\frac{\partial}{\partial q}\left(q\frac{\partial}{\partial q}\left(Q_p(q,j)^n\right)\right)\Bigg|_{j=q}$$

$$= nq\frac{\partial}{\partial q}\left(Q_p(q,j)^{n-1}\sum_{k=1}^{p}kq^k(1-j)^{p-k}\right)\Bigg|_{j=q}$$

$$= nQ_p(q)^{n-2}\left(\begin{array}{c}(n-1)\left(\sum_{k=1}^{p}kq^k(1-q)^{p-k}\right)^2\\ +Q_p(q)\sum_{k=1}^{p}k^2q^k(1-q)^{p-k}\end{array}\right)$$

$$= nQ_p(q)^{n-2}\left(\begin{array}{c}(n-1)\left(\dfrac{Q_p(q)}{n}E_p(X)\right)^2\\ +Q_p(q)\sum_{k=1}^{p}k^2q^k(1-q)^{p-k}\end{array}\right)$$

$$= Q_p(q)^{n-1}\left(\begin{array}{c}\left(1-\dfrac{1}{n}\right)Q_p(q)\left(E_p(X)\right)^2\\ +n\sum_{k=1}^{p}k^2q^k(1-q)^{p-k}\end{array}\right)$$

By applying the variance definition formula:

$$V_p(X) = E_p\left(X^2\right)-\left(E_p(X)\right)^2$$

$$= \left(1-\frac{1}{n}\right)\left(E_p(X)\right)^2+n\frac{\sum_{k=1}^{p}k^2q^k(1-q)^{p-k}}{Q_p(q)}-\left(E_p(X)\right)^2$$

$$= n\frac{\sum_{k=1}^{p}k^2q^k(1-q)^{p-k}}{Q_p(q)}-\frac{\left(E_p(X)\right)^2}{n}$$

$$= n\left(\frac{\sum_{k=1}^{p}k^2q^k(1-q)^{p-k}}{Q_p(q)}-\left(\frac{\sum_{k=1}^{p}kq^k(1-q)^{p-k}}{Q_p(q)}\right)^2\right)$$

$$= \frac{nq}{Q_p(q)^2}\left[\begin{array}{c}Q_p(q)\sum_{k=1}^{p}k^2q^{k-1}(1-q)^{p-k}\\ -q\left(\sum_{k=1}^{p}kq^{k-1}(1-q)^{p-k}\right)^2\end{array}\right]$$

Let consider the following polynomial:

$$R_{2p-1}(q)=Q_p(q)\sum_{k=1}^{p}k^2q^{k-1}(1-q)^{p-k}$$
$$-q\left(\sum_{k=1}^{p}kq^{k-1}(1-q)^{p-k}\right)^2.$$

We have:

$$\begin{cases}R_{2p-1}(0)=Q_p(0)=1\\ R_{2p-1}(1)=Q_p(1)p^2-p^2=0\end{cases}$$

This means that R_{2p-1} is divisible by $(1-q)$ and not divisible by q. We therefore, let factor the polynomial R_{2p-1} as follows:

$$R_{2p-1}(q)=(1-q)P_{2p-2}(q)$$

Thus, the variance of the $(p+1)-$nomial distribution becomes:

$$V_p(X)=nq(1-q)\frac{P_{2p-2}(q)}{Q_p(q)^2}=V_1(X)\frac{P_{2p-2}(q)}{Q_p(q)^2}$$

We find the expression of the polynomial P_{2p-2} using the following relation linking it to the polynomial Q_p:

$$P_{2p-2}(q)$$
$$=Q_p(q)\left[\begin{array}{l}q(1-q)Q_p(q)''+\\(1-2q)Q_p'(q)+pQ_p(q)\end{array}\right]$$
$$-q(1-q)\left(Q_p'(q)\right)^2$$

3.6. Some Properties of Q_p and P_{2p-2}

The purpose of this part is to simplify the determination of the polynomials Q_p and P_{2p} so that the establishment of the characteristics of the *p*-nomial law are simple as much as possible. For this, we will proceed to factorization of the polynomial Q_p.

3.6.1. First Values of *p*

$$Q_1(q)=1$$
$$Q_2(q)=1-q+q^2$$
$$Q_3(q)=1-2q+2q^2$$
$$Q_4(q)=1-3q+4q^2-2q^3+q^4$$
$$Q_5(q)=1-4q+7q^2-6q^3+3q^4$$
$$P_0(q)=1$$
$$P_2(q)=1+2q-2q^2$$
$$P_4(q)=1+4q^2-8q^3+4q^4$$
$$P_6(q)=1-2q+5q^2-15q^4+18q^5-6q^6$$
$$P_8(q)=1-4q+10q^2-12q^3+15q^4-36q^5$$
$$+54q^6-36q^7+9q^8$$

3.6.2. Degree of Polynomial Q_p

It's easy to notice that $\deg(Q_p)\le p$. If we note $Q_p(q)=\sum_{k=0}^{p}a_{p,k}q^k$, then:

$$a_{p,p}=\sum_{k=0}^{p}(-1)^{p-k}=\sum_{k=0}^{p}(-1)^k=\begin{cases}1 & \text{if } p \text{ even}\\0 & \text{if } p \text{ odd}\end{cases}$$

So, we can deduce that:

$$\deg(Q_{2p})=\deg(Q_{2p+1})$$
$$=2p\quad\left(\deg(Q_p)=2\left[\frac{p}{2}\right]\right)\quad(a)$$

3.6.3. Factorization of the Polynomial Q_p

We know, using the definition of the $(p+1)-$nomial distribution, that the polynomial Q_p has no roots on $\mathbb{R}$. So, let's look for the complex roots of Q_p $(p\ge 2)$:

$$Q_p(q)=\sum_{k=0}^{p}q^k(1-q)^{p-k}=0$$
$$\Rightarrow\begin{cases}q^{p+1}-(1-q)^{p+1}=0\\ q\ne\frac{1}{2}\end{cases}$$
$$\Rightarrow\left(\frac{q}{1-q}\right)^{p+1}=1 \text{ and } q\ne\frac{1}{2}$$
$$\Rightarrow q_{p,k}=\frac{e^{i\frac{k\pi}{p+1}}}{2\cos\left(\frac{k\pi}{p+1}\right)}$$
$$=\frac{1}{2}\left(1+i\tan\left(\frac{k\pi}{p+1}\right)\right)\quad k\in[\![1,p]\!]\setminus\left\{\frac{p+1}{2}\right\}\quad(b)$$

In addition, we have:

$$q_{p,p+1-k}=\frac{1}{2}\left(1-i\tan\left(\frac{k\pi}{p+1}\right)\right)=\bar{q}_{p,k}$$

By discussing the cases, we'll have (using $(a),(b)$),

$$Q_{2p}(q)=a_{2p,2p}\prod_{k=1}^{2p}\left(q-q_{2p,k}\right)$$
$$=a_{2p,2p}\prod_{k=1}^{p}\left(q-q_{2p,k}\right)\left(q-\bar{q}_{2p,k}\right)$$
$$=a_{2p,2p}\prod_{k=1}^{p}\left(q^2-q+\frac{1}{4\cos\left(\frac{k\pi}{2p+1}\right)^2}\right)$$

$$Q_{2p+1}(q) = a_{2p+1,2p} \prod_{\substack{k=1\\k\neq p+1}}^{2p+1} \left(q - q_{2p+1,k}\right)$$

$$= a_{2p+1,2p} \prod_{k=1}^{p} \left(q - q_{2p+1,k}\right)\left(q - \overline{q}_{2p+1,k}\right)$$

$$= a_{2p+1,2p} \prod_{k=1}^{p} \left(q^2 - q + \frac{1}{4\cos\left(\frac{k\pi}{2(p+1)}\right)^2} \right)$$

we can summarize the two cases (for p odd or even) by the following formula:

$$\begin{cases} Q_1(q) = 1 \\ Q_p(q) = a_{p,2\left[\frac{p}{2}\right]} \prod_{k=1}^{\left[\frac{p}{2}\right]} \left(q^2 - q + \frac{1}{4\cos\left(\frac{k\pi}{p+1}\right)^2} \right) & (p \geq 2) \end{cases}$$

We recall that: $a_{p,0} = Q_p(0) = 1$. Now let's look for the dominant coefficient of the polynomial Q_p. Using the relationship between the coefficients of a polynomial and its roots, we find $(p \geq 1)$:

$$a_{2p,2p} = (-1)^{2p} \frac{a_0}{\prod_{k=1}^{2p} q_{2p,k}}$$

$$= \frac{1}{\prod_{k=1}^{p} q_{2p,k}\overline{q}_{2p,k}} = \frac{1}{\prod_{k=1}^{p} \left\| q_{2p,k} \right\|^2}$$

$$= 2^{2p} \left(\prod_{k=1}^{p} \cos\left(\frac{k\pi}{2p+1}\right) \right)^2$$

$$a_{2p+1,2p} = \frac{a_0}{\prod_{\substack{k=1\\k\neq p+1}}^{2p+1} q_{2p+1,k}}$$

$$= \frac{1}{\prod_{k=1}^{p} q_{2p+1,k}\overline{q}_{2p+1,k}} = \frac{1}{\prod_{k=1}^{p} \left\| q_{2p+1,k} \right\|^2}$$

$$= 2^{2p} \left(\prod_{k=1}^{p} \cos\left(\frac{k\pi}{2(p+1)}\right) \right)^2$$

We use these two results in this article proved in [10]:

$$\prod_{k=1}^{p} \cos\left(\frac{k\pi}{2p+1}\right) = \frac{1}{2^p}$$

and $$\prod_{k=1}^{p} \cos\left(\frac{k\pi}{2(p+1)}\right) = \frac{\sqrt{p+1}}{2^p}$$

We get:

$$a_{2p,2p} = 2^{2p} \frac{1}{2^{2p}} = 1 \text{ and } a_{2p+1,2p} = 2^{2p} \frac{p+1}{2^{2p}} = p+1$$

if we define *the staircase function* of $\mathbb{Z}$ by:

$$\forall x \in \mathbb{R} \quad \{x\} = \begin{cases} 0 & \text{if } x \in \mathbb{Z} \\ [x] & \text{if } x \notin \mathbb{Z} \end{cases}$$

We get: $a_{p,2\left[\frac{p}{2}\right]} = \left\{\frac{p}{2}\right\} + 1 \quad (p \geq 2)$.

We can thus write the factorization of the polynomial Q_p:

$$Q_p(q)$$

$$= \left(\left\{\frac{p}{2}\right\} + 1 \right) \prod_{k=1}^{\left[\frac{p}{2}\right]} \left(q^2 - q + \frac{1}{4\cos\left(\frac{k\pi}{p+1}\right)^2} \right) \quad (p \geq 2)$$

3.6.4. Rational Fraction of Variance F_p

We define the rational fraction appearing in the expression of the variance of the $(p+1)-$nomial distribution, and we call it rational fraction of variance of order p, by:

$$F_p(q) = \frac{P_{2p-2}(q)}{Q_p(q)^2} \quad (p \geq 1)$$

We represent in the following figure the rational fractions of variance of order 1 to 5 for $q \in [0,1]$:

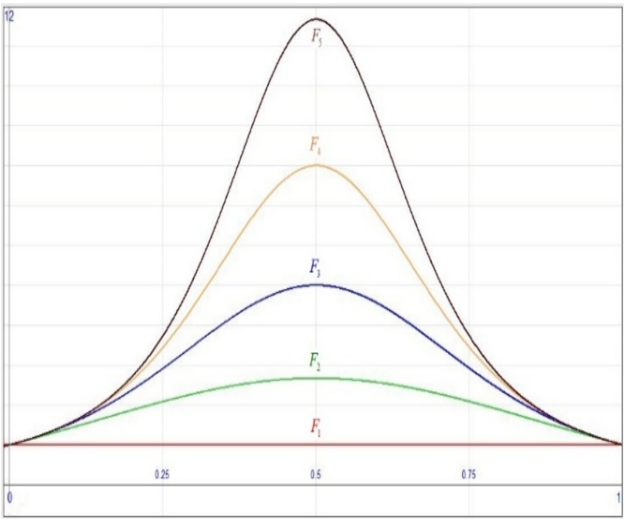

Figure 2. Curves of rational fractions of variance for p = 1 to 5

We notice that: $F_5 \geq F_4 \geq F_3 \geq F_2 \geq F_1$.

Property. We have the following property:

$$\forall p \in \mathbb{N}^* \quad \int_0^1 F_p(q)\,dq = p$$

Proof. Let p be a non-zero natural integer.

By integration by part, we find that:

$$\int_0^1 \frac{q(1-q)\left(Q_p'(q)\right)^2}{Q_p(q)^2}dq = \left[-\frac{q(1-q)Q_p'(q)}{Q_p(q)}\right]_0^1 + \int_0^1 \frac{q(1-q)Q_p(q)''+(1-2q)Q_p'(q)}{Q_p(q)}dq = \int_0^1 \frac{q(1-q)Q_p(q)''+(1-2q)Q_p'(q)}{Q_p(q)}dq$$

Which proves that:

$$\int_0^1 \frac{Q_p(q)\begin{pmatrix} q(1-q)Q_p(q)'' \\ +(1-2q)Q_p'(q)\end{pmatrix} - q(1-q)\left(Q_p'(q)\right)^2}{Q_p(q)^2}dq = 0$$

Or by adding p on both sides:

$$\int_0^1 \frac{\begin{bmatrix} Q_p(q)\begin{pmatrix} q(1-q)Q_p(q)'' \\ +(1-2q)Q_p'(q)+pQ_p(q)\end{pmatrix} \\ -q(1-q)\left(Q_p'(q)\right)^2\end{bmatrix}}{Q_p(q)^2}dq = p$$

So, we get:

$$\forall p \in \mathbb{N}^* \qquad \int_0^1 F_p(q)\,dq = p.$$

3.7. Ordinary Moments

The ordinary moments [9] of the $(p+1)-$nomial distribution are obtained by the recurrence relation:

$$\mu_{r+1} = nq\left\{\begin{array}{l}(1-q)\dfrac{\mu_r'}{n} \\ +\left(p+(1-q)\dfrac{Q_p'(q)}{Q_p(q)}\right)\mu_r\end{array}\right\}.$$

Proof. Starting from the ordinary moment of order $r + 1$:

$$Q_p(q)^n E_p\left(X^{r+1}\right) = \sum_{k=0}^{np} k^{r+1}\pi_{n|p\rangle}^k q^k (1-q)^{np-k}$$
$$= q\frac{d}{dq}\left(\sum_{k=0}^{np} k^r \pi_{n|p\rangle}^k q^k (1-q)^{np-k}\right) + \frac{q}{1-q}\sum_{k=0}^{np} k^r (np-k)\pi_{n|p\rangle}^k q^k (1-q)^{np-k}$$
$$= q\frac{d}{dq}\left(Q_p(q)^n E_p\left(X^r\right)\right) + \frac{q}{1-q}\begin{pmatrix} npQ_p(q)^n E_p\left(X^r\right) \\ -Q_p(q)^n E_p\left(X^{r+1}\right)\end{pmatrix}$$

After simplifications, we get the desired formula.

3.8. Distribution Function

The distribution function of random variable X following a $(p+1)-$nomial distribution $B_p(n,q)$ is given by:

$$f_p(x) = \begin{cases} 1 & \text{if} \quad x \geq np \\ \dfrac{\sum_{k=0}^{[x]} \pi_{n|p\rangle}^k q^k (1-q)^{np-k}}{\left(\sum_{k=0}^{p} q^k (1-q)^{p-k}\right)^n} & \text{if} \quad 0 \leq x < np \\ 0 & \text{if} \quad x < 0 \end{cases}$$

3.9. Characteristic Function

The characteristic function of a random variable X following a $(p+1)-$nomial distribution $B_p(n,q)$ is given by:

$$\forall t \in \mathbb{R} \qquad \phi_p(t) = \mathrm{E}\left(e^{itX}\right) = \frac{\sum_{k=0}^{np} \pi_{n|p\rangle}^k \left(qe^{it}\right)^k (1-q)^{np-k}}{\left(\sum_{j=0}^{p} q^j (1-q)^{p-j}\right)^n} = \left(\frac{\sum_{j=0}^{p}\left(qe^{it}\right)^j (1-q)^{p-j}}{\sum_{j=0}^{p} q^j (1-q)^{p-j}}\right)^n$$

3.10. Moments Generating Function

The moments generating function [9] of a random variable X following a $(p+1)-$nomial distribution $B_p(n,q)$ is given by:

$$\forall t \in \mathbb{R} \qquad G_p(t) = \mathrm{E}\left(e^{tX}\right) = \frac{\sum_{k=0}^{np} \pi_{n|p\rangle}^k \left(qe^{t}\right)^k (1-q)^{np-k}}{\left(\sum_{j=0}^{p} q^j (1-q)^{p-j}\right)^n} = \left(\frac{\sum_{j=0}^{p}\left(qe^{t}\right)^j (1-q)^{p-j}}{\sum_{j=0}^{p} q^j (1-q)^{p-j}}\right)^n$$

3.11. Cumulants Generating Function

We directly deduce the generating function of cumulants:

$$\forall t \in \mathbb{R} \quad \mathrm{C}_p(t) = \ln\left(G(t)\right) = n\ln\left(\frac{\sum_{j=0}^{p}\left(qe^{t}\right)^j (1-q)^{p-j}}{\sum_{j=0}^{p} q^j (1-q)^{p-j}}\right)$$

The generating function of factorial cumulants $(\forall t \in \mathbb{R})$:

$$C_{f,p}(t) = \ln\left[E\left((1+t)^X\right)\right]$$

$$= \ln\left(\frac{\sum_{k=0}^{np} \pi_{n|p\rangle}^k (q(1+t))^k (1-q)^{np-k}}{\left(\sum_{j=0}^{p} q^j (1-q)^{p-j}\right)^n}\right)$$

$$= n\ln\left(\frac{\sum_{j=0}^{p} (q(1+t))^j (1-q)^{p-j}}{\sum_{j=0}^{p} q^j (1-q)^{p-j}}\right)$$

3.12. Beta-(p+1)-nomial Distribution

Similarly, we can define the Beta $-(p+1)-$ nomial distribution $\beta_p(n,a,b)$:

$$\forall k \in [\![0, np]\!]$$

$$P(X=k) = \frac{\pi_{n|p\rangle}^k}{B(a,b)} \int_0^1 \frac{q^{k+a-1}(1-q)^{np-k+b-1}}{\left(\sum_{j=0}^{p} q^j (1-q)^{p-j}\right)^n} dq$$

Where B is the Euler's beta function.

3.13. Markov's Inequality

The Markov's inequality applied for a random variable X following a $(p+1)-$nomial distribution $B_p(n,q)$ give :

$$\forall x > 0 \quad P\left(\frac{X}{n} > x\right) \le \frac{q}{x} \frac{\sum_{k=1}^{p} kq^{k-1}(1-q)^{p-k}}{\sum_{k=0}^{p} q^k (1-j)^{p-k}}$$

3.14. Bienaymé-Tchebychev's Inequality

The Bienaymé-Tchebychev's inequality [8] for a random variable X following a $(p+1)-$ nomial distribution $B_p(n,q)$ is obtained thanks to the moments:

$$\forall x > 0$$

$$P\left(\left|\frac{X}{n} - q\frac{\sum_{k=1}^{p} kq^{k-1}(1-q)^{np-k}}{Q_p(q)}\right| > x\right) \le \frac{q(1-q)}{nx^2} F_p(q).$$

3.15. Central Limit Theorem

Let $(X_m)_{m\in\mathbb{N}^*}$ a series of independent random variables of the same distribution $B_p(n,q)$ and $\bar{X}_m = \frac{X_1 + .. + X_m}{m}$. The application of the central limit theorem [8,9] gives, $\forall x \in \mathbb{R}$:

$$\lim_{m\to+\infty} P\left(\sqrt{\frac{m}{n}} \frac{\bar{X}_m - nq\frac{\sum_{k=1}^{p} kq^{k-1}(1-q)^{np-k}}{Q_p(q)}}{\sqrt{q(1-q)P_{2p-2}(q)}} < \frac{x}{Q_p(q)}\right)$$

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

4. Application

In this example, we are interested in the case of a pandemic. It is assumed that a population P of $m \times p$ individuals is affected by a contaminating pandemic. We subdivide this population into p groups of the same number of individuals n . We will therefore realize n $(n \le m)$ trials so that for each trail we will realize p screening tests by taking an individual from each group, and this in order to isolate the infected individuals.

If the result of the test is that the individual is infected, we say that it is positive and we write P, if not, we say that it is negative and we write N.

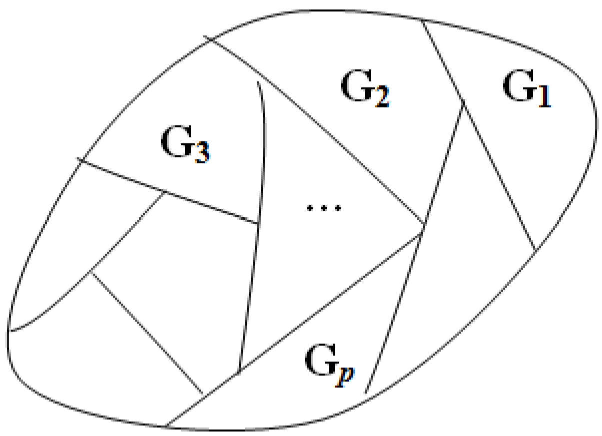

Figure 3. Subdivision of the population into p groups of n individuals

We assume that the rate of infected in the p groups is the same and it's equal to q .

Detection of infected follow $(p+1)-$nomial distribution:

$$D_I \sim B_p(n,q)$$

We consider a simple example. Let a population of 16 individuals. We are dealing with three cases:

1- We realize 16 tests on the entire population without subdividing it into groups. This is the case with binomial distribution.

2- W subdivide this population into two groups and we'll realize 8 tests. This is the case of trinomial distribution.

3- We subdivide this population into four groups and we'll realize 4 tests. This is the case of pentanomial distribution.

We will study the three cases for the infection rate $q = 0,10;\ q = 0,30$ and $q = 0,50$.

a) First case: binomial distribution

The distribution of detection of infected D_I is:

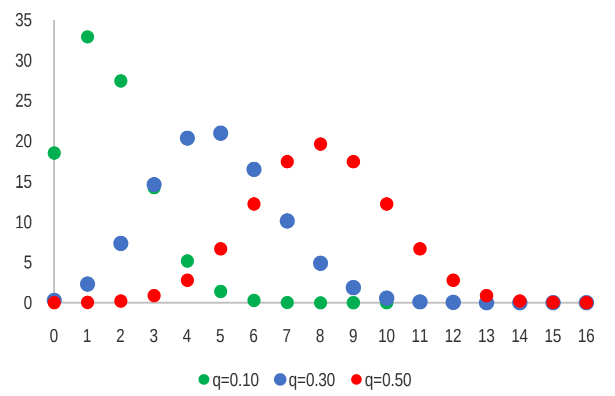

Figure 4. Curves of detection infected in case of binomial distribution

b) Second case: trinomial distribution

The staircase of the trinomial coefficients for $n = 0$ to $n = 8$ is as follows:

								1								
							1	1	1							
						1	2	3	2	1						
					1	3	6	7	6	3	1					
				1	4	10	16	19	16	10	4	1				
			1	5	15	30	45	51	45	30	15	5	1			
		1	6	21	50	90	126	141	126	90	50	21	6	1		
	1	7	28	77	161	266	357	393	357	266	161	77	28	7	1	
1	8	36	112	266	504	784	1016	1107	1016	784	504	266	112	36	8	1

The distribution of detection of infected D_I is:

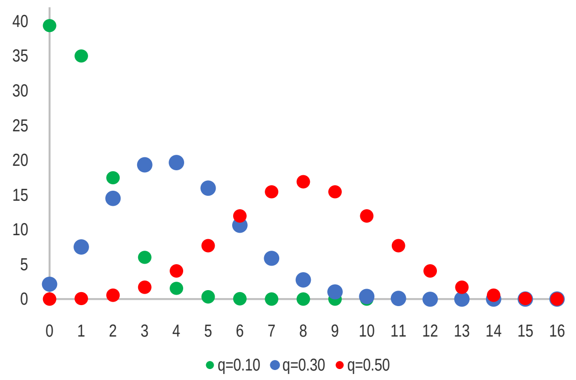

Figure 5. Curves of detection infected in case of trinomial distribution

c) Third case: pentanomial distribution

The staircase of the pentanomial coefficients for $n = 0$ to $n = 4$ is as follows:

								1								
						1	1	1	1	1						
				1	2	3	4	5	4	3	2	1				
		1	3	6	10	15	18	19	18	15	10	6	3	1		
1	4	10	20	35	52	68	80	85	80	68	52	35	20	10	4	1

The distribution of detection of infected D_I is:

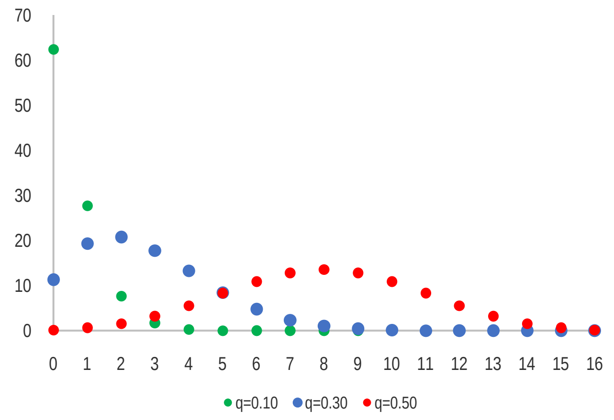

Figure 6. Curves of detection infected in case of pentanomial distribution

Note. We observe that as long as we use a *p*-nomial distribution of parameter *p* large as long as the distribution of probabilities flattens out. In other words, the peak of curves softens. This shows the usefulness of dividing the population into several groups in order to reduce the spread and infection by the pandemic.

5. Trinomial Distribution

5.1. Definition

We consider two sets E_1 and E_2 containing the same elements denoted $e_1,..,e_n$ with cardinal n:

$$\begin{cases} E_1 = E_2 = \{e_1,..,e_n\} \\ \text{card}(E_1) = \text{card}(E_2) = n \end{cases}$$

We are interested in the number of possible choices of k elements among the $2n$ elements that we have $(\text{card}(E_1) + \text{card}(E_2) = 2n)$ without repetition or also in the number of sets A_i that we can construct from sets E_1 and E_2 such as $\text{card}(A_i) = k$ and $A_i \neq A_j$ for $i \neq j$.

5.2. Theorem

This number of partitions is exactly the trinomial coefficient $\pi^k_{n|2\rangle}$.

Proof. The idea of demonstration of this theorem consists in considering a set of cardinal q such that

$q \geq \frac{k}{2}$ of the set E_1 then completing the $k-q$ elements which remain of the same set of the set E_2. By gradually varying q, we will get all the possible sets of cardinal k:

$$E_1 = \left\{ \boxed{e_1, e_2, .., e_{\left[\frac{k}{2}\right]}, .., e_q}, .., e_n \right\}$$

$$E_2 = \left\{ \boxed{\boxed{e_1, e_2, .., e_{k-q}}, .., e_q}, .., e_n \right\}$$

It is quite clear from the above diagrams that:

$$\frac{k}{2} \leq q \leq k \text{ and } q \leq m.$$

This means that: $\frac{k}{2} \leq q \leq \min(n,k)$. Furthermore:

$$\frac{k}{2} \leq q \Rightarrow \left[\frac{k+1}{2}\right] \leq q$$

Now the choice of q elements among n is C_n^q and the choice of $k-q$ elements among q is C_q^{k-q}. Thus, by following this reasoning approach, the number of construction of sets of cardinal k for a given $q\left(\left[\frac{k+1}{2}\right] \leq q \leq \min(n,k)\right)$ is $C_n^q C_q^{k-q}$.

Finally, the total number of construction of sets of cardinal $k(0 \leq k \leq 2n)$ of the two sets E_1 and E_2 is:

$$\pi_{n|2\rangle}^{k} = \sum_{q=\left[\frac{k+1}{2}\right]}^{\min(n,k)} C_n^q C_q^{k-q}$$

Note. In the same way, by considering p identical sets of cardinal n each, we can demonstrate that the number of choices of k element among the np elements that we have or even the number of construction of sets of cardinal k is exactly the $(p+1)-$ nomial coefficient $\pi_{n|p\rangle}^{k}$.

5.3. Probability Tree of the Trinomial Distribution

We consider two identical coins noted c_1 and c_2. We will launch these two coins at the same time for n independent trials. Each outcome has a fixed probability, the same from trial to trial.

If we get a head, we note a success S and if we get a tail, we note a failure F. The possible outcomes are (S,S), (S,F) or (F,F). We note T_k the k[th] trial. We represent below the probability tree:

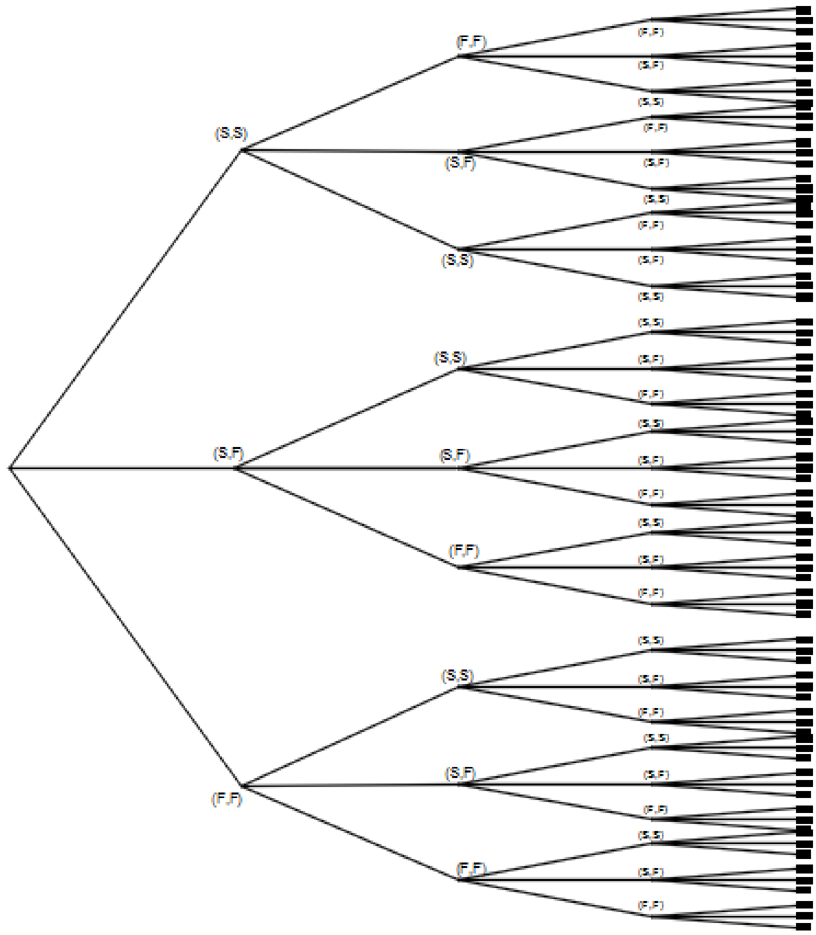

Figure 7. Probability tree of the trinomial distribution

We count, after n tests, the total number of success obtained. If we denote by X this random variable then the possible values of X are $0,1,..,2n$. In addition, if we note $q\left(q \in [0,1]\right)$ the probability of having a success, then the random variable X follows a trinomial distribution noted $B_2(n,q)$ such that :

$$\forall k \in [\![0,2n]\!] \quad \mathrm{P}(X=k) = \pi_{n|2\rangle}^{k} \frac{q^k (1-q)^{2n-k}}{\left(q^2-q+1\right)^n}$$

We note: $X \sim B_2(n,q)$.

5.4. Examples

5.4.1. Case where *n* = 1

In this case, one trial is realized. The possible outcomes to obtain are: (S,S), (S,F) or (F,F). If we denote by $N_1(k)$ the number of obtaining k successes, then:

$$N_1(0)=1=\pi_{1|2\rangle}^{0};\ N_1(1)=1=\pi_{1|2\rangle}^{1};\ N_1(2)=1=\pi_{1|2\rangle}^{1}$$

The probabilities are thus:

$$\begin{cases} \mathrm{P}(X=0)=\dfrac{(1-q)^2}{q^2-q+1};\ \mathrm{P}(X=1)=\dfrac{q(1-q)}{q^2-q+1} \\ \mathrm{P}(X=2)=\dfrac{q^2}{q^2-q+1} \end{cases}$$

For $q=\frac{1}{2}$: $\mathrm{P}(X=0)=\mathrm{P}(X=1)=\mathrm{P}(X=2)=\frac{1}{3}$.

5.4.2. Case where n = 2

In this case, two trials are realized. The possible outcomes to obtain are: (S,S,S,S), (S,S,S,F), (S,S,F,F), (S,F,F,F) or (F,F,F,F). If we denote by $N_2(k)$ the number of obtaining k successes, then:

$$N_2(0)=1=\pi_{2|2\rangle}^{0};\ N_2(1)=2=\pi_{2|2\rangle}^{1};\ N_2(2)=3=\pi_{2|2\rangle}^{1}$$

$$N_2(3)=2=\pi_{2|2\rangle}^{3};\ N_2(4)=1=\pi_{2|2\rangle}^{4}$$

The probabilities are thus:

$$\begin{cases} \mathrm{P}(X=0)=\dfrac{(1-q)^4}{\left(q^2-q+1\right)^2};\quad \mathrm{P}(X=1)=2\dfrac{q(1-q)^3}{\left(q^2-q+1\right)^2} \\ \mathrm{P}(X=2)=3\dfrac{q^2(1-q)^2}{\left(q^2-q+1\right)^2};\ \mathrm{P}(X=3)=2\dfrac{q^3(1-q)}{\left(q^2-q+1\right)^2} \\ \mathrm{P}(X=4)=\dfrac{q^4}{\left(q^2-q+1\right)^2} \end{cases}$$

For $q=\frac{1}{2}$, we get: $\mathrm{P}(X=0)=\mathrm{P}(X=4)=\frac{1}{9}$

$\mathrm{P}(X=2)=\frac{1}{3}$ and $\mathrm{P}(X=1)=\mathrm{P}(X=3)=\frac{2}{9}$.

6. Conclusion

This article is an introduction for constructing of a new probability distribution called *p-nomial distribution* as a generalization of binomial distribution. The examples of application of this new probability law introduced in this research give us some ideas about its frequent meeting in practice.

References

[1] Yadolah Dodge, Statistique, disctionnaire encyclopédique, Paris 2007.
[2] Anders Hald, A History of Probability and Statistics and Their Applications before 1750, John Wiley & Sons, 2005.
[3] Eric Gossett, Discrete Mathematics with Proof, John Wiley & Sons, 2009.
[4] Michiel Hazewinkel, Encyclopaedia of Mathematics, vol. 2, Springer Science+Business Media, 1994.
[5] Norman Johnson, Adrienne Kemp and Samuel Kotz, Univariate Discrete Distributions, John Wiley & Sons, 2005
[6] Alan Ruegg, Probabilités et statistique, vol. 3, PPUR, 1994.
[7] Aziz ATTA, "p-nomial Coefficients and p-nomial Theorem." Turkish Journal of Analysis and Number Theory, vol. 8, no. 1 (2020): 6-15.
[8] F.M. Dekking C. Kraaikamp H.P. Lopuhaa L.E. Meester, A Modern Introduction to Probability and Statistics, Understanding Why and How, Delft Institute of Applied Mathematics Delft University of Technology Mekelweg 4 2628 CD Delft The Netherlands.
[9] Charles M. Grinstead, J. Laurie Snell, Introduction to Probability, Swarthmore College and Dartmouth College, USA.
[10] Aziz ATTA, "Twin Polynomials and Kernels Matrix." Turkish Journal of Analysis and Number Theory, vol. 8, no. 3 (2020): 57-69.
[11] Andreas N. Philippou, Demetrios L. Antzoulakos, Binomial Distribution, Chapter · January 2011, researchgate.net/
[12] Matthew Aldridge, Measure Theory and Integration: Measures, MA40042, UK.
[13] M. Samuel Fiorini, MATH-F-307 Mathématiques discrètes, Version du 5 octobre 2012.