Một phép biến đổi của hàm khối xác suất cho một lớp các biến ngẫu nhiên rời rạc

Lê Thanh Bính *

Khoa Toán và Thống kê, Trường Đại học Quy Nhơn, Việt Nam

**Tác giả liên hệ chính. Email: lethanhbinh@qnu.edu.vn*

TÓM TẮT

Chúng tôi xét một biến ngẫu nhiên rời rạc X chỉ nhận các giá trị nguyên không âm. Ký hiệu miền giá trị của X và hàm khối xác suất của X lần lượt bởi $\mathcal{R}_X$ và $p_X(x)$. Mục đích của bài báo này nhằm đưa ra một phương pháp biến đổi được dùng để biến đổi hàm $p_X(x)$ thành một hàm khối xác suất của một biến ngẫu nhiên rời rạc $\tilde{X}$ với miền giá trị là $\mathcal{R}_{\tilde{X}} = \{k \in \mathbb{N} : k \geq \min \mathcal{R}_X\}$. Chúng tôi tìm thấy một biểu diễn cho hàm đặc trưng của $\tilde{X}$ theo hàm đặc trưng của X. Ngoài ra, tính bảo toàn phân phối của phép biến đổi được chỉ ra trong một số trường hợp cụ thể.

Từ khóa: *Hàm khối xác suất, biến ngẫu nhiên rời rạc, phép biến đổi, hàm đặc trưng.*

A transformation of probability mass functions for a class of discrete random variables

Le Thanh Binh*

Department of Mathematics and Statistics, Quy Nhon University, Vietnam

**Corresponding author. Email: lethanhbinh@qnu.edu.vn*

ABSTRACT

Let us consider a discrete random variable X that takes only non-negative integer values. Let $\mathcal{R}_X$ and $p_X(x)$ denote the range of X and the probability mass function of X, respectively. The aim of this paper is to provide a transformation method used to transform $p_X(x)$ into a probability mass function of a discrete random variable $\tilde{X}$ whose range is $\mathcal{R}_{\tilde{X}} = \{k \in \mathbb{N} : k \geq \min \mathcal{R}_X\}$. We obtain a representation of the characteristic function of $\tilde{X}$ in terms of the characteristic function of X. Moreover, the distribution-preserving property of the transformation is shown in some specific cases.

Keywords: *Probability mass function, discrete random variable, transformation, characteristic function.*

1. INTRODUCTION

In probability theory, a probability distribution is the mathematical function that gives the probabilities of occurrence of different possible outcomes for a random experiment. It is a mathematical description of a random phenomenon in terms of its sample space and the probabilities of events (subsets of the sample space).[1,2] The sample space, often denoted by Ω, is the set of all possible outcomes of a random experiment being observed.

In order to classify probability distributions, we need to define discrete and continuous random variables. A random variable is a function whose domain is a sample space Ω and whose range (i.e., the set of values that it can obtain) is a subset of the real numbers, $\mathbb{R}$. In other words, a random variable assigns real numbers to the outcomes in its sample space. Random variables which take on values from a discrete set of numbers (i.e., whose range is either *finite* or *countably infinite*) are called *discrete random variable*.[3] Otherwise, a random variable is called *contin-*

uous if it ranges over a continuous set of numbers that contains all real numbers between two limits.[3] In other words, a continuous random variable is one that takes an uncountably infinite number of possible values. For instance, a random variable that represents the time between two successive arrivals to a queueing system, or that represents the temperature in a nuclear reactor, is an example of a continuous random variable.[3] It is evident that all random variables defined on a discrete sample space are discrete. However, random variables defined on a continuous sample space may be either discrete or continuous. Probability distributions can be categorized into two main types: discrete and continuous. Discrete distributions deal with the probabilities of specific values for discrete random variables, while continuous distributions handle the probabilities of various values for continuous random variables. Examples of discrete distributions include the *Binomial*, *Poisson*, and *Negative Binomial* distributions. We will introduce these distributions and several other discrete distributions in more detail in Section 3. For continuous distributions, the most popular example is the *normal distribution*. This is also referred to as the *Gaussian distribution*. Some important continuous distributions are often used to build models and to test hypotheses about random variables, such as the student's t-distribution, the *chi-squared distribution* and the *F-distribution*.

The key difference between a discrete probability distribution and a continuous probability distribution is that in a discrete distribution we are able to compute the probability that a random variable can take on a particular value, therefore the probabilities of individual values can be tabulated. Discrete random variables, or discrete distributions, can be completely characterized by their *probability mass functions*. The probability mass function (frequently abbreviated to *pmf*) for a discrete random variable X, gives the probability that the value obtained by X on the outcome of a probability experiment is equal to x ($x \in \mathbb{R}$).[3] In the present paper, we denote it by $p_X(.)$. The formal definition of the probability mass function for a discrete random variable is given in Section 2. Sometimes the term *discrete density function* is used in place of probability mass function. Since a continuous random variable takes an uncountably infinite number of possible values, the probability that it is exactly equal to any one of the infinite possible values is zero. For this reason, the method mentioned above to describe a discrete random variable will not work in the case of a continuous random variable, and then we have to consider the probability of a continuous random variable taking values in an interval. Continuous random variables, or continuous distributions, can be completely characterized by their *probability density functions* (frequently abbreviated to *pdf*). Because the purpose of this study is to concentrate only on discrete distributions, in the article we will ignore the definitions or concepts associated with continuous random variables, and we refer the reader to[1,2,4] for more details.

The starting point of this paper was to study the Binomial distribution (denoted by $\mathrm{Binom}(n, p)$). This distribution has two parameters: the number of trials , $n \in \mathbb{N}^*$, and the probability of success for a single trial, $p \in (0, 1)$. The outcome from a random variable X obeying the Binomial distribution will always be a nonnegative integer with an upper bound at n. By the rules of probability, we can attain that the probability of the event $\{X = k\}$ (i.e., the probability of k successes in n trials) is equal to $\binom{n}{k}p^k(1-p)^{n-k}$. By definition, the quantity $\binom{n}{k}p^k(1-p)^{n-k}$ is the value of the probability mass function of X at k, namely $p_X(k)$. Then, by chance and by intuition, we have found the following equality:

$$\sum_{k=0}^{n}\sum_{i=0}^{k}(np-i)\binom{n}{i}p^i(1-p)^{n-i} = npq,$$

which can be shortly rewritten as

$$\sum_{k=0}^{n}\sum_{i=0}^{k}(\mu - i)p_X(i) = \sigma^2, \qquad (1)$$

where $\mu = np$ and $\sigma^2 = npq$.

At first glance, equality (1) was nothing special. However, it is worth noticing that the quantities $\mu = np$ and $\sigma^2 = npq$ are the *mean* and *variance* of the Binomial random variable X, respectively. Furthermore, the set $\{0; 1; ...; n\}$ is the range of X (denoted by $\mathcal{R}_X$). The definitions of the mean and variance of a discrete random variable are given in Section 2. Then, a question naturally arose in our mind: *Whether equality* (1) *holds true for an arbitrary discrete random variable* X *whose range is a subset of the set of natural numbers, if its mean and variance are finite, or not*? Motivated by this question, we have shown that equality (1) remains true for nonnegative integer-valued random variables satisfying a certain condition. This result is presented in Lemma 3.2. Combining Lemma 3.2 and Lemma 3.1, we then obtain the first main theorem (Theorem 3.2), which gives a way to transform a probability mass function of a nonnegative integer-valued random variable to that of another nonnegative integer-valued random variable. From this result, we achieve the remaining important results as shown in Section 3. Up to the present, there are only a few results on transformations associated with probability mass functions. For instance, the *pignistic transformation* and the *plausibility transformation* are introduced in the research[5]. We briefly recall that these two transformations provide the ways to transform a *basic probability assignment* function to a probability mass function. Notice that a basic probability assignment function (called also mass function) is not a probability mass function. For more detail, see[5].

The rest of the paper is organized as follows. Section 2 revisits key definitions and properties including probability mass function, mean, variance, and characteristic function. Section 3 presents our primary findings. Finally, Section 4 concludes with remarks summarizing the significance of our research outcomes. This systematic approach aids in understanding the framework and contributions of our study.

2. PRELIMINARIES

2.1. Probability mass function, Mean and Variance

From the point of view of understanding the behavior of a discrete random variable, the important thing is to know the probabilities that the random variable takes each value in its range. Such probabilities are described with a *probability mass function*.

Definition 2.1. [4] Let X be a discrete random variable. The *probability mass function* of X, denoted by $p_X(.)$, is defined as

$$p_X(x) = P(X = x) > 0 \quad \text{if } x \in \mathcal{R}_X,$$
$$p_X(x) = 0 \quad \text{if } x \notin \mathcal{R}_X,$$

where $\mathcal{R}_X$ is the range of X.

Obviously, the range of $p_X(.)$ is a subset of the interval $[0, 1]$. Furthermore, by the rules of probability, one can get that the function values add to 1.0 when summed over all possible values of the random variable X. This means that $\sum_{x \in \mathcal{R}_X} p_X(x) = 1$.

Definition 2.2. [4] Let X be a discrete random variable with $\mathcal{R}_X = \{x_k\}_{k \geq 0}$. The *expectation* or the *mean* of the random variable X, denoted by $\mathbb{E}X$, is the number

$$\mathbb{E}X = \sum_{x \in \mathcal{R}_X} x p_X(x) = \sum_{k=0}^{\infty} x_k p_X(x_k), \quad (2)$$

which is defined when $\sum_{k=0}^{\infty} |x_k| p_X(x_k) < \infty$. If the later series diverges, the mean is not defined.

In the case where the mean is defined, its value does not depend on the order of summation. Essentially, the mean $\mathbb{E}X$ denotes a weighted average of the elements in $\mathcal{R}_X$, where the probabilities act as the weights in the discrete setting.

Definition 2.3. Let X be a discrete random variable with $\mathcal{R}_X = \{x_k\}_{k \geq 0}$, and let $\lambda > 0$ be a positive real number (not necessarily integer). The *moment of order* λ of X is defined as

$$\alpha_\lambda = \mathbb{E}X^\lambda = \sum_{k=0}^{\infty} (x_k)^\lambda p_X(x_k).$$

Definition 2.4. [4] Suppose that the mean and the moment of order 2 of the discrete random variable X are finite. The *variance* of X, denoted by Var$X$, is the quantity

$$\begin{aligned}\text{Var}X &= \mathbb{E}(X - \mathbb{E}X)^2 \\ &= \sum_{k=0}^{\infty}(x_k - \mathbb{E}X)^2 p_X(x_k).\end{aligned}$$

The variance characterizes the amount of variation of the random variable from its mean. The following property is commonly useful to compute the variance.

$$\text{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2.$$

The expectation and variance of a random variable are two of the foremost notions in probability theory. For basic properties of expectation and variance, we refer the reader to the studies [1,4,6].

2.2. Characteristic function

In probability theory and mathematical statistics, characteristic functions always play an outstanding role by providing a comprehensive way to describe and analyze probability distributions. They are particularly powerful due to their unique properties and applications in various statistical methodologies.

Definition 2.5. [7] The *characteristic function* of a discrete random variable X is defined as

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = \sum_{k=0}^{\infty} e^{itx_k} p_X(x_k), \qquad (3)$$

where t is any real number and $i = \sqrt{-1}$.

Since $|e^{itx}|$ is a bounded and continuous function for all finite real t and x, the characteristic function always exists. We recall that any characteristic function $\varphi_X(t)$ satisfies the following conditions (see the research [7] Theorem 1.1.1):

1. $\varphi_X(t)$ is uniformly continuous;
2. $\varphi_X(0) = \lim_{t\to 0}\varphi_X(t) = 1$;
3. $|\varphi_X(t)| \le 1$ for all real numbers t.
4. $\varphi_X(-t) = \overline{\varphi_X(t)}$, where the horizontal bar denotes the complex conjugate.

In addition, if the moment of order n exists (where n is a positive integer) then $\varphi_X(t)$ is n times differentiable for all t, and it is related to the $n-th$ derivative of the characteristic function by the formula [7]

$$\alpha_n = (-i)^n \varphi_X^{(n)}(0). \qquad (4)$$

So, the existence of some moments of a random variable ensures the existence of the corresponding derivatives of the characteristic function. We next introduce the following important result (referred to as the uniqueness theorem), which shows that a probability distribution is uniquely determined by its characteristic function.

Proposition 2.1 (Theorem 1.1.2)**.** [7] *Two probability distributions are identical if and only if their characteristic functions are identical.*

For more details on properties of characteristic functions, interested readers could be refer to [7] and the references therein. Thanks to characteristic functions, we arrive at some interesting results as shown in Subsection 3.3.

3. MAIN RESULTS

Let X be a discrete random variable with the range $\mathcal{R}_X \subseteq \mathbb{N}$ (the set $\mathcal{R}_X$ is either finite or countably infinite). Throughout the forthcoming, we always assume that the mean and variance of X exist, and are denoted by μ and σ^2 ($\sigma > 0$) respectively.

3.1. Formulation of transformation

Lemma 3.1. *Let k be a nonnegative integer. If $\mathcal{R}_X$ is infinite, we then get*

$$\sum_{i=0}^{k}(\mu - i)p_X(i) > 0 \Leftrightarrow k \ge \min \mathcal{R}_X.$$

If $\mathcal{R}_{\tilde{X}}$ is fintite with $|\mathcal{R}_X|$ greater than 1, we have

$$\sum_{i=0}^{k}(\mu - i)p_X(i) > 0 \Leftrightarrow \min \mathcal{R}_X \le k \le \max \mathcal{R}_X - 1.$$

Proof. If $\mathcal{R}_X$ is infinite, we have that

$$\sum_{i=0}^{k}(\mu - i)p_X(i) = \sum_{i=\min \mathcal{R}_X}^{k}(\mu - i)p_X(i) > 0$$

for $\min \mathcal{R}_X \le k \le \mu$ (note that $\mu > \min \mathcal{R}_X$). For $k > \mu$, setting $a(i) = (\mu - i)p_X(i)$, we obtain

$$\begin{aligned}\sum_{i=0}^{k}(\mu - i)p_X(i) &= \sum_{i=0}^{[\mu]} a(i) + \sum_{i=[\mu]+1}^{k} a(i)\\ &> \sum_{i=0}^{[\mu]} a(i) + \sum_{i=[\mu]+1}^{\infty} a(i)\\ &= \sum_{i=0}^{\infty} a(i) = \mu - \mu = 0,\end{aligned}$$

where $[.]$ denotes the floor function. Obviously, $\sum_{i=0}^{k}(\mu - i)p_X(i) = 0$ if $k < \min \mathcal{R}_X$.

In the case that $\mathcal{R}_X$ is finite (with $|\mathcal{R}_X| > 1$), due to $\sum_{i=0}^{\max \mathcal{R}_X}(\mu - i)p_X(i) = 0$, we only need to consider k such that $\min \mathcal{R}_X \le k \le \max \mathcal{R}_X - 1$. □

Lemma 3.2. *Assume that*

$$\lim_{n\to\infty} n\sum_{i=0}^{n}(\mu - i)p_X(i) = 0. \tag{5}$$

Then, setting $m = \min \mathcal{R}_X$, *we have*

$$\sum_{k=m}^{\infty}\sum_{i=m}^{k}(\mu - i)p_X(i) = \sigma^2. \tag{6}$$

In the case that $\mathcal{R}_X$ *is finite, equality* (6) *becomes*

$$\sum_{k=m}^{M-1}\sum_{i=m}^{k}(\mu - i)p_X(i) = \sigma^2, \tag{7}$$

where $M := \max \mathcal{R}_X$.

Proof. For each positive integer $n \ge m$, we have

$$\begin{aligned}&\sum_{k=m}^{n}\sum_{i=m}^{k}(\mu - i)p_X(i)\\ &= \sum_{k=0}^{n}\sum_{i=0}^{k}(\mu - i)p_X(i) \quad (p_X(i) = 0 \text{ if } i \notin \mathcal{R}_X)\\ &= \sum_{i=0}^{n}(n+1-i)(\mu - i)p_X(i)\\ &= \sum_{i=0}^{n}[(i-\mu)^2 + n(\mu - i) + (1-\mu)(\mu - i)]p_X(i)\\ &= \sum_{i=0}^{n}(i-\mu)^2 p_X(i) + n\sum_{i=0}^{n}(\mu - i)p_X(i)\\ &\qquad + (1-\mu)\sum_{i=0}^{n}(\mu - i)p_X(i).\end{aligned} \tag{8}$$

It is noteworthy that, by the definitions of μ and σ^2,

$$\sum_{i=0}^{\infty}(i-\mu)^2 p_X(i) = \sigma^2; \sum_{i=0}^{\infty}(\mu - i)p_X(i) = 0. \tag{9}$$

Equality (6) readily follows from (5), (8) and (9).

If R_X is a finite set, by the definition of μ one can easily see that $\sum_{i=m}^{n}(\mu - i)p_X(i) = 0$ for every $n \ge M$. Therefore, condition (5) is always true and we obtain (7). □

Remark 3.1. Lemma 3.2 yields another formula for the variance of a discrete random variable X if its range is a subset of the set of natural numbers, provided that (5) is satisfied. Furthermore, from the proof of Lemma 3.2, we notice that (5) is a necessary and sufficient condition for the validity of (6).

Combining Lemma 3.1 with Lemma 3.2, we immediately attain the first main theorem.

Theorem 3.2. *Assume that* (5) *holds and set* $m = \min \mathcal{R}_X$, $M = \max \mathcal{R}_X$. *Then, there exists a discrete random variable* $\tilde{X}$ *such that*

$$\mathcal{R}_{\tilde{X}} = \begin{cases}\{k \in \mathbb{N} : m \le k\} & \text{if } \mathcal{R}_X \text{ is infinite};\\ \{k \in \mathbb{N} : m \le k < M\} & \text{if } \mathcal{R}_X \text{ is finite};\end{cases}$$

and its probability mass function is given by

$$p_{\tilde{X}}(k) = P(\tilde{X} = k) = \frac{1}{\sigma^2}\sum_{i=m}^{k}(\mu - i)p_X(i), \tag{10}$$

for all $k \in \mathcal{R}_{\tilde{X}}$.

Proof. According to Lemmas 3.1 and 3.2, we have

$$p_{\tilde{X}}(k) > 0 \ (\forall k \in \mathcal{R}_{\tilde{X}}) \quad \text{and} \quad \sum_{k=m}^{\infty} p_{\tilde{X}}(k) = 1,$$

which imply the statement of Theorem 3.2. □

Remark 3.3. In other words, Theorem 3.2 or formula (10) textcolorredprovide a probability transformation which transforms the probability mass function $p_X(.)$ to another probability mass function, $p_{\tilde{X}}(.)$. Also, one can see that the range of $\tilde{X}$ is always a set containing consecutive nonnegative intergers, and has the same minimum value as the one of the initial random variable, $\mathcal{R}_X$.

Let us now consider the following example to more understand the use of the transformation.

Example 3.4. Let X be the random variable with the probability distribution described as follows:

X	0	4	6	8
$p_X(x)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$

By direct calculation, using (10) we get

$$\mu = 6; \sigma^2 = 7;$$
$$p_{\tilde{X}}(0) = p_{\tilde{X}}(1) = p_{\tilde{X}}(2) = p_{\tilde{X}}(3) = \frac{3}{28};$$
$$p_{\tilde{X}}(4) = p_{\tilde{X}}(5) = \frac{1}{7};$$
$$p_{\tilde{X}}(6) = p_{\tilde{X}}(7) = \frac{1}{7}.$$

Clearly, $\sum_{k=0}^{7} p_{\tilde{X}}(k) = 1$ and the corresponding probability distribution of $\tilde{X}$ is given as

$\tilde{X}$	0	1	2	3	4	5	6	7
$p_{\tilde{X}}(k)$	$\frac{3}{28}$	$\frac{3}{28}$	$\frac{3}{28}$	$\frac{3}{28}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$

The next example was intended as an attempt to extend the claim of Theorem 3.2 to the case that X takes (positive) noninteger values. However, we obtain that the claim is no longer true.

Example 3.5. Let X be the random variable with the probability distribution given as

X	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2
$p_X(x)$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$

From (10), we get

$$\mu = \frac{19}{16};\ \sigma^2 = \frac{143}{256};$$
$$p_{\tilde{X}}(0) = \frac{38}{143};$$
$$p_{\tilde{X}}(1) = \frac{88}{143}.$$

We obtain

$$p_{\tilde{X}}(0) + p_{\tilde{X}}(1) = \frac{126}{143} \neq 1.$$

Hence, equality (6) does not hold.

3.2. The characteristic function $\varphi_{\tilde{X}}(.)$

As mentioned in Section 1, it will be very useful to obtain an expression for the characteristic function of $\tilde{X}$. By Proposition 2.1, the fact that every distribution is uniquely determined by its characteristic function allows us to be able to determine the distribution type of $\tilde{X}$, without having to find the mass probability function $p_{\tilde{X}}(.)$.

Theorem 3.6. *With the settings of Theorem 3.2, the characteristic function $\varphi_{\tilde{X}}(.)$ of the random variable $\tilde{X}$ is given by*

$$\varphi_{\tilde{X}}(t) = \frac{\mu\varphi_X(t) + i\varphi'_X(t)}{\sigma^2(1-e^{it})}, \forall t \in \mathbb{R}, \tag{11}$$

where, as before, μ, σ^2 and $\varphi_X(.)$ are respectively the mean, variance and characteristic function of the random variable X.

Proof. For simplicity of notations, throughout the proof, p_k and $\tilde{p}_k$ stand for $p_X(k)$ and $p_{\tilde{X}}(k)$, respectively. From (10) and by grouping the terms appropriately, we attain

$$S_n(t) := \sum_{k=0}^{n} e^{itk}\tilde{p}_k$$
$$= \frac{1}{\sigma^2}\sum_{k=0}^{n}\left[e^{itk}\sum_{j=0}^{k}(\mu-j)p_j\right]$$
$$= \frac{1}{\sigma^2}(\mu-0)p_0\sum_{k=0}^{n}e^{itk} + \frac{1}{\sigma^2}(\mu-1)p_1\sum_{k=1}^{n}e^{itk}$$
$$+ \frac{1}{\sigma^2}(\mu-2)p_2\sum_{k=2}^{n}e^{itk} + \ldots + \frac{1}{\sigma^2}(\mu-n)p_ne^{itn}$$

$$= \frac{1}{\sigma^2} \sum_{j=0}^{n} \Big[(\mu - j) p_j \sum_{k=j}^{n} e^{itk} \Big]$$
$$= \frac{1}{\sigma^2} \left(S_{1,n}(t) - S_{2,n}(t) \right), \qquad (12)$$

where

$$S_{1,n}(t) := \sum_{j=0}^{n} \Big[(\mu - j) p_j \sum_{k=0}^{n} e^{itk} \Big]; \qquad (13a)$$

$$S_{2,n}(t) := \sum_{j=1}^{n} \Big[(\mu - j) p_j \sum_{k=0}^{j-1} e^{itk} \Big]. \qquad (13b)$$

On the other hand, by definition,

$$\varphi_{\tilde{X}}(t) = \lim_{n\to\infty} S_n(t), \qquad (14)$$

we are thus left with the task of determining the limits of $S_{1,n}(t)$ and $S_{2,n}(t)$ as n tends to ∞.

To find the limit of $S_{1,n}(t)$ defined as (13a), it is worth pointing out that

$$0 < |S_{1,n}(t)| = \left| \sum_{k=0}^{n} e^{itk} \right| \left| \sum_{j=0}^{n} (\mu - j) p_j \right|$$
$$\leq \left| n \sum_{j=0}^{n} (\mu - j) p_j \right|.$$

From (5) and the Squeeze Theorem, it immediately follows that

$$\lim_{n\to\infty} S_{1,n}(t) = 0. \qquad (15)$$

In order to arrive at the remaining limit, we first rewrite $S_{n,2}(t)$, given by (13b), as follows

$$S_{2,n}(t) = \frac{1}{1 - e^{it}} \sum_{j=1}^{n} (\mu - j) p_j (1 - e^{itj})$$
$$= \frac{1}{1 - e^{it}} \sum_{j=0}^{n} (\mu - j) p_j (1 - e^{itj})$$
$$= \frac{1}{1 - e^{it}} \Big[\sum_{j=0}^{n} (\mu - j) p_j - \mu \sum_{j=0}^{n} p_j e^{itj}$$
$$+ \sum_{j=0}^{n} j p_j e^{itj} \Big]. \qquad (16)$$

Letting n tend to ∞ in the both sides of (16), we get

$$\lim_{n\to\infty} S_{2,n}(t) = \frac{1}{1 - e^{it}} \Big[-\mu \varphi_X(t) + \frac{\varphi'_X(t)}{i} \Big], \qquad (17)$$

owing to the following simple equalities,

$$\sum_{j=0}^{\infty} (\mu - j) p_j = 0;$$
$$\sum_{j=0}^{\infty} p_j e^{itj} = \varphi_X(t); \quad \sum_{j=0}^{\infty} j p_j e^{itj} = \frac{1}{i} \varphi'_X(t).$$

From (12), (14), (15) and (17), the proof of Theorem 3.6 is completed. □

3.3. Distribution-preserving property

The work of this section contains descriptions of some different well-kown discrete distributions used in probability. By the method of characteristic functions, our aim is to verify whether the random variables X and $\tilde{X}$ are able to belong to the same family of distributions (in other words, whether the distribution family of X can be preserved by the formulated transformation) for each considered case.

- **Binomial distribution**

Binomial distributions correspond to random variables that count the number of successes among n independent trials having the same probability of success. Such trials are called Bernoulli trials. The probabilistic model of Bernoulli trials is applicable in many situations, where it is reasonable to assume independence and constant success probability.

Definition 3.1. [6,8] A random variable X is said to have a *Binomial distribution* with parameters n and p (where $n \in \mathbb{N}^*$, $0 < p < 1$) if

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \qquad (18)$$

for all $k = 0, 1, ..., n$. We write $X \sim \mathrm{B}(n, p)$.

If $X \sim \mathrm{B}(n, p)$, the mean and variance are[6]

$$\mu = np, \;\; \sigma^2 = np(1 - p), \qquad (19)$$

and the characteristic function is given by[7]

$$\varphi_X(t) = (1 - p + pe^{it})^n. \qquad (20)$$

From (19), (20) and (11), we have

$$\varphi_{\tilde{X}}(t) = (1 - p + pe^{it})^{n-1}, \tag{21}$$

which immediately implies that

$$\tilde{X} \sim \mathrm{B}(n-1, p),$$

for all $n \geq 2$.

- **Poisson distribution**

Poisson distributions are applied when the random variables under consideration count the number of events occurring in a specified time period or a spatial area, and the observed processes satisfy the primary conditions of time (or space) homogeneity, independent increments, and no memory of the past.

Definition 3.2. [6,8] A random variable X is said to have a *Poisson distribution* with unique parameter $\lambda > 0$ if

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, ... \tag{22}$$

We then write $X \sim Pois(\lambda)$.

The mean, variance and characteristic function of the Poisson distribution are[7]

$$\mu = \sigma^2 = \lambda, \tag{23}$$
$$\varphi_X(t) = \exp[\lambda(e^{it} - 1)]. \tag{24}$$

First of all, let us prove that assumption (5) is satisfied. Indeed, by (22) and (23), we get

$$\begin{aligned}
&\sum_{k=0}^{n}(\mu - k)p_X(k) \\
&= \sum_{k=0}^{n}(\lambda - k)\frac{e^{-\lambda}\lambda^k}{k!} \\
&= e^{-\lambda}\Big[\sum_{k=0}^{n}\frac{\lambda^{k+1}}{k!} - \sum_{k=1}^{n}\frac{\lambda^k}{(k-1)!}\Big] \\
&= e^{-\lambda}\Big[\sum_{k=0}^{n}\frac{\lambda^{k+1}}{k!} - \sum_{k=0}^{n-1}\frac{\lambda^{k+1}}{k!}\Big] \\
&= e^{-\lambda}\frac{\lambda^{n+1}}{n!}.
\end{aligned}$$

As a result, assumption (5) is equivalent to

$$\frac{\lambda^n}{n!} \longrightarrow 0 \ \text{ as } \ n \to \infty,$$

which is true for all $\lambda > 0$. So, (5) is valid. By Theorem 3.6, (23) and (24), it is straightforward to find the expression for $\varphi_{\tilde{X}}$,

$$\varphi_{\tilde{X}}(t) = \exp[\lambda(e^{it} - 1)] = \varphi_X(t).$$

Thus, we have

$$\tilde{X} \sim \ \mathrm{Pois}(\lambda).$$

- **Negative binomial distribution**

The Negative Binomial distribution is a discrete probability distribution that models the number of failures in a sequence of independent and identically distributed Bernoulli trials before a specified number of successes occurs. In a sequence of independent Bernoulli trials, each trial has two potential outcomes called ”success” and ”failure”. In each trial the probability of success is p $(0 < p < 1)$ and of failure is $1 - p$. One observes this sequence until a number r of successes occurs, where r is a fixed integer.

Definition 3.3. [6,8] Let the random variable X denote the number of observed failures before the r^{th} success occurs. Then

$$P(X = k) = \binom{k + r - 1}{k}(1 - p)^k p^r, \tag{25}$$

for all $k = 0, 1, 2, ...$

In this case, the random variable X is said to have the *Negative Binomial distribution* with parameters r and p. We denote by $X \sim \mathrm{NB}(r, p)$.

If $X \sim \mathrm{NB}(r, p)$, then

$$\mu = \frac{r(1-p)}{p}, \sigma^2 = \frac{r(1-p)}{p^2}, \tag{26}$$

and its characteristic function is given as[7]

$$\varphi_X(t) = \left(\frac{p}{1 - (1-p)e^{it}}\right)^r, \ t \in \mathbb{R}. \tag{27}$$

From (25) and (26), we first remark that

$$\sum_{k=0}^{n}(\mu-k)p_X(k)$$
$$= \sum_{k=0}^{n}\Big(\frac{rq}{p}-k\Big)C_{k+r-1}^{k}q^kp^r \qquad (q:=1-p)$$
$$= rqp^{r-1}+rp^{r-1}\sum_{k=1}^{n}\Big(C_{k+r-1}^{k}q^{k+1}-C_{k+r-1}^{k-1}q^kp\Big)$$
$$= rqp^{r-1}+rp^{r-1}\Big[\sum_{k=1}^{n}\Big(C_{k+r-1}^{k}+C_{k+r-1}^{k-1}\Big)q^{k+1} - \sum_{k=1}^{n}C_{k+r-1}^{k-1}q^k\Big]$$
$$= rqp^{r-1}+rp^{r-1}\Big[\sum_{k=1}^{n}C_{k+r}^{k}q^{k+1}-\sum_{k=0}^{n-1}C_{k+r}^{k}q^{k+1}\Big]$$
$$= rp^{r-1}C_{n+r}^{n}q^{n+1}.$$

Due to $0<q<1$, it is easy to check that

$$nC_{n+r}^{n}q^{n+1}=\frac{n(n+1)...(n+r)}{r!}q^{n+1}\longrightarrow 0$$

as $n\to\infty$. In other words, (5) holds true.

Accordingly, by Theorem 3.6, we attain the characteristic function of $\tilde{X}$ defined by

$$\varphi_{\tilde{X}}(t)=\left(\frac{p}{1-(1-p)e^{it}}\right)^{r+1},$$

which concludes that $\tilde{X}\sim \text{NB}(r+1,p)$.

• Geometric distribution

Consider independent trials such that a certain event may happen at any given trial with probability p. The trials continue until the event occurs for the first time. The number, X, of trials performed before the event occurs has a geometric distribution.[6]

Definition 3.4. [6] A random variable X is said to have a *geometric distribution* with parameter p, where $0<p<1$, if its probability mass function is defined by

$$P(X=k)=(1-p)^kp, \tag{28}$$

for all $k=0,1,2,...$ We then write $X\sim \text{Geo}(p)$.

From (28), it is easy to see that the geometric distribution is the special case of the negative binomial with $r=1$, namely,

$$X\sim \text{Geo}(p)\Leftrightarrow X\sim \text{NG}(1,p).$$

As a consequence, we get that $\tilde{X}\sim \text{NB}(2,r)$ if $X\sim \text{Geo}(p)$.

• Hypergeometric distribution

The hypergeometric distribution is a discrete probability distribution that models the probability of obtaining a specific number of successes in a sample drawn without replacement from a finite population containing two distinct types of elements[6,8] (i.e., a finite population whose elements can be classified into two categories one which possesses a certain characteristic and another which does not possess that characteristic). For instance, suppose an urn contains K white balls and $(N-K)$ black balls. From this, n balls are drawn without replacement. The probability that the sample of size n contains k white balls and $(n-k)$ black balls can be obtained by hypergeometric distribution.

The hypergeometric distribution is characterized by the following parameters:

- N: The total population size.
- K: The number of elements of Type 1 in the population.
- n: The number of draws without replacement (the sample size).

Definition 3.5. Let N, K and n be integers such that $N\geq 1$, $0\leq K\leq N$, and $1\leq n\leq N$. A random variable X is said to have a *hypergeometric distribution* with parameters (N,K,n), written as $X\sim \text{HG}(N,K,n)$, if the corresponding probability mass function is given by

$$P(X=k)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \tag{29}$$

where $k\in\mathbb{Z}$ and

$$\max(0,n+K-N)\leq k\leq \min(n,K).$$

If $X \sim \mathrm{HG}(N, K, n)$, the mean and variance are

$$\mu = n\frac{K}{N}\ ,\ \sigma^2 = n\frac{K(N-K)(N-n)}{N^2(N-1)}, \qquad (30)$$

and its characteristic function is given by[7]

$$\varphi_X(t) = \frac{\binom{N-K}{n} {}_2F_1[-n, -K; N-K-n+1; e^{it}]}{\binom{N}{n}}. \qquad (31)$$

where

$${}_2F_1[a, b; c; z] = 1 + \frac{ab}{c}\frac{z}{1!} + \frac{a(a+1)b(b+1)}{c(c+1)}\frac{z^2}{2!} + \cdots \qquad (32)$$

is the *Gaussian hypergeometric function*.[6]

By virtue of the fact that

$$\frac{\partial {}_2F_1[a, b; c; z]}{\partial z} = \frac{ab}{c} {}_2F_1[a+1, b+1; c+1; z],$$

we then attain

$$\varphi'_X(t) = ie^{it}\frac{nK\binom{N-K}{n} {}_2F_1[\alpha, \beta; \gamma; e^{it}]}{(N-K-n+1)\binom{N}{n}}, \qquad (33)$$

where $\alpha := -n+1, \beta := -K+1$, and $\gamma := N-K-n+2$. Unfortunately, at first we couldn't find the explicit expression for $\varphi_{\tilde{X}}(t)$ by means of formula (11) in Theorem 3.6. Therefore, it is difficult for us to determine the appropriate distribution of the random variable $\tilde{X}$.

However, according to the above results and Theorem 3.2, we have had a reasonable belief that the random variable $\tilde{X}$ should follow a hypergeometric distribution and, futhermore, its support set must be $\{k \in \mathbb{Z} : \max(0, n+K-N) \leq k \leq \min(n-1, K-1)\}$. For this reason, we aim at proving that

$$\tilde{X} \sim \mathrm{HG}(N-2, K-1, n-1), \qquad (34)$$

provided that $N \geq 3$, $K \geq 1$ and $n \geq 2$.

To do this, we first note that (34) is equivalent to

$$\varphi_{\tilde{X}}(t) = \frac{\binom{\tilde{N}-\tilde{K}}{\tilde{n}} {}_2F_1[-\tilde{n}, -\tilde{K}; \tilde{N}-\tilde{K}-\tilde{n}+1; e^{it}]}{\binom{\tilde{N}}{\tilde{n}}}, \qquad (35)$$

where $\tilde{N} := N-2$, $\tilde{K} := K-1$ and $\tilde{n} := n-1$. With the aid of the algebraic computation software (MAPLE), we could easily verify that the following identity

$$\varphi_{\tilde{X}}(t) - \frac{\mu\varphi_X(t) + i\varphi'_X(t)}{\sigma^2(1-e^{it})} \equiv 0 \ (\forall t \in \mathbb{R}),$$

holds true if μ, σ^2, $\varphi_X(t)$, $\varphi'_X(t)$, and $\varphi_{\tilde{X}}(t)$ are given by (30), (31), (33) and (35), respectively. Hence, assertion (34) is true.

• Logarithmic series distribution

The logarithmic series distribution (also known as the the log-series distribution) is a discrete probability distribution derived from the Maclaurin series expansion:

$$\ln(1-p) = -p - \frac{p^2}{2} - \frac{p^3}{3} - \cdots = \sum_{k=1}^{\infty} \frac{-p^k}{k}, \qquad (36)$$

where $0 < p < 1$. From this, we get

$$\sum_{k=1}^{\infty} \frac{-p^k}{k \ln(1-p)} = 1.$$

So, it is easy to see that

$$f(k) = \frac{-p^k}{k \ln(1-p)},\ k = 1, 2, ...,$$

defines a probability mass function on the set of positive integers.

Definition 3.6. [6] A random variable X is said to have a *logarithmic series distribution* with parameter p, where $0 < p < 1$, if its probability mass function is given as

$$P(X = k) = -\frac{p^k}{k \ln(1-p)}, k = 1, 2, 3, ... \qquad (37)$$

We then write $X \sim \mathrm{LogSeries}(p)$.

The logarithmic series distribution is sometimes used to model the number of items of a product purchased by a buyer in a specified interval.

If $X \sim \mathrm{LogSeries}(p)$, the mean and variance are given as

$$\mu = -\frac{p}{(1-p)\ln(1-p)}, \qquad (38a)$$

$$\sigma^2 = -\frac{p^2 + p\ln(1-p)}{(1-p)^2(\ln(1-p))^2}. \qquad (38b)$$

Besides, its characteristic function is as follows[7]

$$\varphi_X(t) = \frac{\ln(1 - pe^{it})}{\ln(1 - p)}. \tag{39}$$

Let us now show that assumption (5) is satisfied when $X \sim \text{LogSeries}(p)$. For any positive integer n, according to (37) and (38a), we derive

$$\begin{aligned} & n\sum_{k=1}^{n}(\mu - k)p_X(k) \\ &= A(p)n\sum_{k=1}^{n}p^k\left((1-p)\ln(1-p) + \frac{p}{k}\right) \\ &= A(p)n\left(\ln(1-p)\sum_{k=1}^{n}(p^k - p^{k+1}) + p\sum_{k=1}^{n}\frac{p^k}{k}\right) \\ &= A(p)n\left(\ln(1-p)(p - p^{n+1}) + p\sum_{k=1}^{n}\frac{p^k}{k}\right) \\ &= A(p)\left(-np^{n+1}\ln(1-p) + pnB_n(p)\right), \end{aligned} \tag{40}$$

where

$$A(p) := \frac{1}{(1-p)(\ln(1-p))^2}, \tag{41a}$$

$$B_n(p) := \ln(1-p) + \sum_{k=1}^{n}\frac{p^k}{k}. \tag{41b}$$

Owing to $\lim_{n\to\infty} np^{n+1} = 0$ for all $p \in (0, 1)$, it follows easily from (40) that assumption (5) holds true if and only if

$$\lim_{n\to\infty} nB_n(p) = 0. \tag{42}$$

To verify (42), it is worth noting that $B_n(p)$ defined as (41b) is exactly equal to the Lagrange remainder of order n (usually denoted by $R_n(.)$) for the Maclaurin series in equation (36). Using the Lagrange remainder formula[9] applied for the function $f(x) = \ln(1 + x)$ at $x = -p$, for each n, we then attain

$$\begin{aligned} B_n(p) &= \frac{(-1)^n(-p)^{n+1}}{(1+\xi_n)^{n+1}(n+1)} \\ &= \frac{-1}{n+1}\left(\frac{p}{1+\xi_n}\right)^{n+1}, \end{aligned} \tag{43}$$

where ξ_n is some number (depending on n) between $-p$ and 0. Thus, owing to (43), limit (42) is equivalent to

$$\lim_{n\to\infty}\left(\frac{p}{1+\xi_n}\right)^{n+1} = 0, \tag{44}$$

which evidently depends on the limit of $\frac{p}{1+\xi_n}$ as n tends to ∞. More specifically, noticing $0 < 1 - p < 1 + \xi_n$ and setting $c := \lim_{n\to\infty}\frac{p}{1+\xi_n}$, if $c \in [0, 1)$ then (44) is true. If $c = 1$, the right hand side of (44) has the indeterminate form 1^∞, and hence we haven't been able to draw an exact conclusion on (44).

Moreover, from the following estimate

$$0 < \frac{p}{1+\xi_n} \leq \frac{p}{1-p} \quad (\forall n \in \mathbb{N}^*),$$

we easily achieve that (44) holds true for all $p \in (0, 1/2)$. However, we haven't yet verified the validity of (44) (equivalently, that of (42)) in the case $p \in [1/2, 1)$. We want to emphasize here that the claims of Lemma 3.2, Theorem 3.2 and Theorem 3.6 are no longer true if (44) does not hold.

Let $p \in (0, 1/2)$. By virtue of Theorem 3.6, and from (38a), (38b), we get the characteristic function of $\tilde{X}$ given as

$$\varphi_{\tilde{X}}(t) = \frac{q\left((1-pe^{it})\ln(1-pe^{it}) - e^{it}q\ln q\right)}{(\ln q + p)(1-e^{it})(1-pe^{it})}, \tag{45}$$

where $q := 1 - p$. We haven't determined the probability distribution family corresponding to the characteristic function defined by (45).

Remark 3.7. By using L'Hospital's rule, we get that $\lim_{t\to 0}\varphi_{\tilde{X}}(t) = 0$ for all $p \in (0, 1)$ (not only for $p \in (0, 1/2)$), where $\varphi_{\tilde{X}}(t)$ is given as (45). This means that a basic property of characteristic functions (as presented in Section 2) is satisfied for all values of p. In addition, with the aid of MAPLE, we have checked by direct calculation that (42) (hence, so is assumption (5)) remains true for many values of p in $[0.5, 1)$ (such as 0.5, 0.6, 0.65, 0.7, and up to $p = 0.78$). Therefore, we can reasonably predict that if $X \sim \text{LogSeries}(p)$, Theorems 3.2 and 3.6 is then true for every $p \in (0, 1)$. We have been trying to prove this.

4. CONCLUSION

In the present study, we propose a novel transformation of probability mass functions associated with nonnegative interger-valued discrete random variables. We also demonstrate that our proposed transformation preserves some well-known families of distributions, such as Poisson distribution, Negative Binomial distribution and Hypergeometric distribution. In the future, we intend to extend our research in two directions. The first one is to continue determining the distribution of the resulting random variable ($\tilde{X}$) when the initial random variable (X) has another discrete distrbution, in addition to the distributions listed in Section 3. This work aims to further verify the distribution-preserving property of the transformation. Besides, we would like to discover its useful applications in various fields. The second direction, and the more difficult, is to construct an analogous transformation of probability density functions in the case of continuous random variables. One of the most important aims of probability theory is to find transformations which can preserve an initial probability distribution in some sense. Consequently, such transformations have attracted a great deal of attention.

Acknowledgment

This study is conducted within the framework of science and technology projects at institutional level of Quy Nhon University under the project code T2023.792.02.

REFERENCES

1. R. B. Ash. *Basic probability theory* (2^{nd} *edition*), Dover, New York, 2008.
2. M. Evans, J. S. Rosenthal. *Probability and statistics: the science of uncertainty*, W. H. Freeman and Co., New York, 2010.
3. W. J. Stewart. *Probability, markov chains, queues, and simulation* (1^{st} *edition*), Princeton University Press, New Jersey, 2009.
4. A. DasGupta. *Fundamentals of probability: a first course*, Springer, New York, 2010.
5. R. P. Srivastava, T. J. Mock. *Belief functions in business decisions (Studies in fuzziness and soft computing)*, Springer, Berlin Heidelberg, 2002.
6. N. L. Johnson, A. W. Kemp, S. Kotz. *Univariate discrete distributions* (3^{rd} *edition*), Wiley, New Jersey, 2005.
7. N. G. Ushakov. *Selected topics in characteristic functions*, De Gruyter, Berlin, 2011.
8. Y. G. Sinai. *Probability theory: an introductory course*, Springer, Berlin Heidelberg, 1992.
9. D. Bock, D. Donovan, S. O. Hockett. *Barron's AP calculus* (14^{th} *edition*), Barron's Educational Series, New York, 2017.