Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

Mgr. Michaela Koščová

Autoreferát dizertačnej práce

Properties of partial-sum discrete probability distributions

na získanie akademického titulu *philosophiae doctor*

v odbore doktorandského štúdia:

1114 Aplikovaná matematika

Bratislava 2020

Dizertačná práca bola vypracovaná:
v dennej forme doktorandského štúdia na Katedre aplikovanej matematiky a štatistiky Fakulty matematiky, fyziky a informatiky Univerzity Komenského v Bratislave.

Predkladateľ: Mgr. Michaela Koščová
Katedra aplikovanej matematiky a štatistiky
Fakulta matematiky, fyziky a informatiky
Univerzita Komenského v Bratislave
Mlynská dolina
842 48 Bratislava

Školiteľ: doc. Mgr. Ján Mačutek, PhD.
Katedra aplikovanej matematiky a štatistiky
Fakulta matematiky, fyziky a informatiky
Univerzita Komenského v Bratislave
Mlynská dolina
842 48 Bratislava

Obhajoba dizertačnej práce sa koná .... o .... h
pred komisiou pre obhajobu dizertačnej práce v odbore doktorandského štúdia vymenovanou predsedom odborovej komisie dňa ....
v študijnom odbore 1114 Aplikovaná matematika
na Fakulte matematiky, fyziky a informatiky Univerzity Komenského v Bratislave, Mlynská dolina, 842 48 Bratislava

Predseda odborovej komisie:
prof. RNDr. Daniel Ševčovič, DrSc.
Katedra aplikovanej matematiky a štatistiky
Fakulta matematiky, fyziky a informatiky
Univerzita Komenského v Bratislave
Mlynská dolina
842 48 Bratislava

Introduction

Let $\{P_j^*\}_{j=0}^\infty$ and $\{P_j\}_{j=0}^\infty$ be discrete probability distributions defined on the set of non-negative integers. The distribution $\{P_j\}_{j=0}^\infty$ is the result of a partial summation applied to $\{P_j^*\}_{j=0}^\infty$ if

$$P_x = \sum_{j=x}^{\infty} u(x,j) P_j^*, \qquad x = 0, 1, 2, \dots\ , \tag{1}$$

where $u(\cdot,\cdot)$ is a real function. The distributions $\{P_j^*\}_{j=0}^\infty$ and $\{P_j\}_{j=0}^\infty$ are called the parent and the descendant, respectively. A brief overview of partial-sum distributions can be found in [3], pp. 508-512. Summation (1) was used mainly as a tool for creating new discrete distributions (see, e.g., [13] or [18]). However, other applications of partial summations are available in literature.

Different types of partial summation can be obtained by different choices of $u(\cdot,\cdot)$[1]. Two types of partial summation that are special cases of (1) were proposed in [10], namely

$$P_x = \sum_{j=x}^{\infty} g(j) P_j^*, \qquad x = 0, 1, 2, \dots \tag{2}$$

and

$$P_x = h(x) \sum_{j=x}^{\infty} P_j^*, \qquad x = 0, 1, 2, \dots\ , \tag{3}$$

where $g(\cdot)$ and $h(\cdot)$ are real functions.

There is a one-to-one correspondence between discrete distributions and partial summations (2). For every distribution, there exists one and only one partial summation defined by (2), which results in the descendant identical with its parent, i.e.,

$$P_x = P_x^*, \qquad x = 0, 1, 2, \dots\ .$$

The same is true for summation (3). Then we say that the distribution $\{P_j^*\}_{j=0}^\infty$ is invariant with respect to summation (2), or (3). Hence, the functions $g(\cdot)$ and $h(\cdot)$ can be considered new characteristics of discrete distributions.

In [20], these characteristics are derived for the geometric distribution. The geometric distribution is invariant with respect to both of summations (2) and (3) if and only if the function $g(\cdot)$ (as well as $h(\cdot)$) is equal to p, which is the parameter of geometric distribution, i.e., if

$$P_x = p \sum_{j=x}^{\infty} P_j^*, \qquad x = 0, 1, 2, \dots\ . \tag{4}$$

[1] We remind that the function $u(\cdot,\cdot)$ does not have to depend on both of its arguments.

The partial summation (4) can be also called the geometric partial summation (similarly, the partial summation, with respect to which the Poisson distribution is invariant, is called the Poisson partial summation, etc.). The geometric partial summation is widely used in reliability analysis, mostly to determine the residual life of a component in a system, see e.g. [14] or [15]. An application of this type of partial summation in risk models in insurance is presented in [16]. The geometric distribution is somewhat special among discrete distributions, being invariant to the partial summation defined by $g(j) = h(x) = p$. In general, the functions $g(\cdot)$ and $h(\cdot)$ can differ from each other.

The works [8] and [19] (the latter inspired by certain problems in quantitative modelling in linguistics and musicollogy) offer the conditions of invariance for some other specific types of partial summations. Finally, [10] solves the problem of invariance under summations (2) and (3) in general. A method for deriving $g(\cdot)$ and $h(\cdot)$ for a wide class of discrete distributions is presented as follows. Let the parent distribution satisfy the recurrence relation

$$P^*_{x+1} = f(x+1)P^*_x \qquad x = 0, 1, 2, \dots \ .$$

The distribution is invariant with respect to the summation (2) if and only if

$$g(x) = 1 - f(x+1), \qquad x = 0, 1, 2, \dots \ .$$

On the other hand, the distribution is invariant with respect to the summation (3) if and only if

$$h(x) = \left(1 + \sum_{j=1}^{\infty} \prod_{k=1}^{j} f(x+k)\right)^{-1}, \qquad x = 0, 1, 2, \dots \ .$$

The relations between probability generating functions of the parent and descendant distributions are presented in [10], for arbitrary meaningful function $g(\cdot)$ (or $h(\cdot)$).

Generally, in the partial summation (1) there are three elements:

- the parent distribution $\{P^*_j\}_{j=0}^{\infty}$,
- the descendant distribution $\{P_j\}_{j=0}^{\infty}$, and
- the function $u(\cdot,\cdot)$ (or in special cases (2) and (3) the functions $g(\cdot)$ and $h(\cdot)$, respectively) - the link between the parent and the descendant.

If we fix two of the three elements, the remaining one is (almost[2]) uniquely given by (1). What is more, [21] prove that any two discrete distributions defined on the same

[2]The non-uniqueness concerns only parent distributions for some modifications of partial summation (3) and is caused by the normalization constant.

support with only non-zero probabilities are connected by a partial summation for some choice of the function $u(\cdot,\cdot)$.

A special case of (2), known as the STER summation (Sums successively Truncated from the Expectation of the Reciprocal of a random variable having the parent distribution), can be found in [1]. According to [22], the Yule distribution is the only one for which the parent and the descendant are identical under the STER summation. Generalizations of the STER summation are proposed by [17] and [9]. Another special case of (2) finds its application in economic modelling, see e.g. [7], [5], and [6]. In the latter two works, the partial summation is used to characterize a class of renewal risk models. The partial summation is understood more generally there, as an operator on any real-valued function. One more application of partial summation (2) is in finding continuous analogues of discrete distributions (and vice versa) provided by [12].

It is possible to apply partial summation iteratively as follows:

$$P_x^{(1)} = C_1 \sum_{j=x}^{\infty} u(x,j) P_j^*, \qquad x = 0,1,2,\dots\ ,$$

$$P_x^{(k)} = C_k \sum_{j=x}^{\infty} u(x,j) P_j^{(k-1)}, \qquad x = 0,1,2,\dots\ , \qquad k = 2,3,\dots$$

C_k is a normalization constant which ensures that $\{P_x^{(k)}\}_{x=0}^{\infty}$ is a proper distribution for $k = 1,2,3,\dots$.

In [11], the iterated partial summation with

$$u(x,j) = p,$$

$p \in (0,1)$ is scrutinized. We note that the geometric distribution is invariant with respect to this partial summation, see [20]. Using the convergence of the probability generating functions, [11] proves that if the parent distribution $\{P_j^*\}_{j=0}^{\infty}$ satisfies

$$p = \lim_{j\to\infty} \frac{P_{j+1}^*}{P_j^*}, \tag{5}$$

the limit distribution is the geometric distribution with the parameter value equal to p in (5), i.e.,

$$\lim_{i\to\infty} P_x^{(i)} = p(1-p)^x, \qquad x = 0,1,2,\dots\ .$$

Objectives

The dissertation thesis focuses on two main topics, namely invariance of partial summation and iterated partial summation.

1. For each discrete probability distribution there exists one and only one summation under which the distribution is invariant (the parent distribution is the same as descendant). If the summations are parametrized, the parent distribution remains invariant in some cases, but sometimes we obtain another distribution as the descendant. The two families of partial summations are characterized in the thesis.

2. Partial summations can be applied iteratively. The question is whether there exists a limit descendant distribution for the iterated summations. So far, the only case where the limit distribution (geometric) was known is the most simple partial summation in which the summed probabilities are multiplied by a constant. This result are generalized for other types of partial summations.

1 Parametrization of partial summation

In this chapter we consider the partial summation

$$P_x = \sum_{j=x}^{\infty} g(j) P_j^*, \qquad x = 0, 1, 2, \dots \ . \tag{6}$$

Let $f(x)$ be a function given by

$$P_{x+1}^* = f(x+1) P_x^*, \qquad x = 0, 1, 2, \dots \ .$$

Then (according to [10]) the function $g(\cdot)$, which leaves the parent distribution unaltered under summation (6), is

$$g(x) = 1 - f(x+1), \qquad x = 0, 1, 2, \dots \ . \tag{7}$$

For the sake of simplicity, in the following considerations we limit ourselves to discrete distributions with one parameter only.

Denote a the parameter of the distribution $\{P_j^*\}_{j=0}^{\infty}$. In order to emphasize the role of the parameter, we can use the notation $\{P_j^*(a)\}_{j=0}^{\infty}$ and condition of invariance (7) can be rewritten as

$$g(x; a) = 1 - f(x+1; a), \qquad x = 0, 1, 2, \dots \ . \tag{8}$$

The descendant distribution is uniquely given by $\{P_j^*(a)\}_{j=0}^{\infty}$ and $g(x;a),\ x=0,1,\dots$, i.e., it also depends solely on the parameter a. Hence the distribution resulting from (6) is in fact $\{P_j(a)\}_{j=0}^{\infty}$:

$$P_x(a) = \sum_{j=x}^{\infty} g(j;a) P_j^*(a), \qquad x = 0,1,2,\dots \ . \tag{9}$$

Let us consider a modification of summation (9), namely,

$$P_x = c \sum_{j=x}^{\infty} g(j;\lambda) P_j^*(a), \qquad x = 0,1,2,\dots \ , \tag{10}$$

with

$$g(x;\lambda) = 1 - f(x+1;\lambda), \qquad x = 0,1,2,\dots \ , \tag{11}$$

where the formula for the function from (8) is kept, but parameter a was replaced with λ; c is a proper constant (with respect to x) which should ensure that $\{P_j(a)\}_{j=0}^{\infty}$ sums to 1. Our aim is to investigate consequences of the above mentioned change of the parameter value.

1.1 Examples

Geometric distribution: Let $\{P_j^*(a)\}_{j=0}^{\infty}$ be the geometric distribution with parameter $a \in (0;1\rangle$, i.e.,

$$P_j^*(a) = a(1-a)^j, \qquad j = 0,1,2,\dots \ .$$

Summation (10) yields

$$P_x = c \sum_{j=x}^{\infty} g(j;\lambda) a(1-a)^j, \qquad x = 0,1,2,\dots \ ,$$

and function $g(\cdot)$ is determined by (11) as

$$g(x;\lambda) = 1 - f(x+1;\lambda) = 1 - \frac{P_{x+1}^*(\lambda)}{P_x^*(\lambda)} = \lambda, \qquad x = 0,1,2,\dots \ .$$

To find the descendant distribution $\{P_j\}_{j=0}^{\infty}$ it is necessary to find the normalization constant c,

$$c = \left[\sum_{x=0}^{\infty} \sum_{j=x}^{\infty} g(j;\lambda) P_j^*(a) \right]^{-1} = \frac{a}{\lambda}.$$

Hence the descendant is

$$P_x = \frac{a}{\lambda} \sum_{x=j}^{\infty} \lambda a(1-a)^j = a(1-a)^x, \qquad x = 0,1,2,\dots \ .$$

For the geometric distribution, the change of the parameter value from a to λ in summation (10) does not affect the resulting descendant distribution, as the new parameter value is eliminated by the constant c. The descendant distribution is identical with its parent distribution.

Poisson distribution: If $\{P_j^*(a)\}_{j=0}^{\infty}$ is the Poisson distribution with parameter $a \geq 0$,

$$P_j^*(a) = \frac{e^{-a}a^j}{j!}, \qquad j = 0, 1, 2, \dots ,$$

the summation (10) has the form

$$P_x = c\sum_{j=x}^{\infty} g(j;\lambda)\frac{e^{-a}a^j}{j!}, \qquad x = 0, 1, 2, \dots .$$

The function $g(\cdot)$ is determined by (11), i.e.,

$$g(x;\lambda) = 1 - f(x+1;\lambda) = 1 - \frac{P_{x+1}^*(\lambda)}{P_x^*(\lambda)} = 1 - \frac{\lambda}{x+1}, \qquad x = 0, 1, 2, \dots .$$

The normalization constant is

$$c = \left[\sum_{x=0}^{\infty}\sum_{j=x}^{\infty} g(j;\lambda)P_j^*(a)\right]^{-1} = \frac{1}{1-\lambda+a}.$$

Therefore, the descendant distribution is given by

$$P_x = \frac{1}{1-\lambda+a}\sum_{x=j}^{\infty}\left(1 - \frac{\lambda}{j+1}\right)\frac{e^{-a}a^j}{j!}, \qquad x = 0, 1, 2, \dots .$$

The resulting descendant distribution for the Poisson parent is a two-parameter distribution (i.e. it differs from its parent, which has only one parameter). As both parameters a and λ are parameters of certain Poisson distributions, they must not be negative. Moreover, it must hold $a + 1 > \lambda$, because the normalization constant c is supposed to be positive.

1.2 Two families of discrete distributions

We use the examples to demonstrate that the descendant distribution of summation (10) is either a one-parameter distribution identical with its parent or a two-parameter distribution. This behaviour defines two families of discrete distributions. The first family consists of those distributions that yield a two-parameter descendant under summation (10). In the following, this family of distributions is denoted by the term family of sensitive distributions (or the sensitive family). The name of this family is chosen to emphasize the sensitivity of the parent to the change of summation parameter value. On the other hand, the group of distributions that remain unaltered under summation (10) is henceforward denoted as the family of resistant distributions (or the resistant family). The members of the resistant family are, in fact, resistant to the change of summation parameter value, when summation (10) is applied to them. In the following we provide the theoretical background.

Theorem 1. *A one-parameter discrete distribution* $\{P_j^*(a)\}_{j=0}^{\infty}$ *belongs to the resistant family if and only if*

$$\frac{1 - f(j+1;\lambda)}{1 - f(1;\lambda)P_0^*(a) + \sum_{s=1}^{\infty}[s-(s+1)f(s+1;\lambda)]P_s^*(a)} = 1 - f(j+1;a), \qquad (12)$$

$$j = 0, 1, 2, \ldots \,.$$

Theorem 1 offers a rule which strictly distinguishes between the family of resistant distributions and the family of sensitive distributions.

Theorem 2. *If the distribution* $\{P_j^*\}$ *belongs to the resistant family then the ratio*

$$\frac{1 - \dfrac{P_{j+1}^*(\lambda)}{P_j^*(\lambda)}}{1 - \dfrac{P_{j+1}^*(a)}{P_j^*(a)}} \qquad (13)$$

does not depend on j.

Theorem 2 presents a useful necessary condition which can be used to classify distributions to the sensitive family. Discrete distributions with one parameter for which (13) is dependent on j are certainly from the family of sensitive distributions. Those for which (13) is independent of j are candidates for the resistant family and validity of the condition (12) comes into question.

The family of resistant distributions seems to be significantly smaller than the family of sensitive distributions. So far only the geometric and the Salvia-Bolinger distribution are known to be from that family.

2 Iterated partial summations

We propose some solutions regarding iterated partial summations. It is possible to apply partial summation (2) repeatedly, i.e.,

$$\begin{aligned}
P_x^{(1)} =& C_1 \sum_{j=x}^{\infty} g(j) P_j^*, \qquad x = 0, 1, 2, \dots \ , \\
P_x^{(2)} =& C_2 \sum_{j=x}^{\infty} g(j) P_j^{(1)}, \qquad x = 0, 1, 2 \dots \ , \\
&\vdots \\
P_x^{(n)} =& C_n \sum_{j=x}^{\infty} g(j) P_j^{(n-1)}, \qquad x = 0, 1, 2 \dots \ , \\
&\vdots
\end{aligned}$$

C_n is a normalization constant which ensures that $\{P_x^{(n)}\}_{x=0}^{\infty}$, also called the n-th descendant, is a proper distribution for $n = 1, 2, 3, \dots$. The distribution $\{P_x^*\}_{x=0}^{\infty}$ is hereafter called the original parent. In [11] it is proven that if there exists

$$p = \lim_{j\to\infty} \frac{P_{j+1}^*}{P_j^*},$$

the iterated geometric partial summation yields the geometric distribution with the parameter p as the limit distribution. One of the aims of our research is to examine the behaviour of other types of iterated partial summations. At first we perform a computational study for iterated partial summations with finite-support original parents. Then we show the relation between the iterated partial summations and the power method, which is an apparatus from the matrix theory designed to find dominant eigenvalues and eigenvectors of matrices. The use of the power method is applied to the Katz family of partial summations. The conditions under which the power method can be applied, and their impact on the iterated partial-sums problem are discussed in the dissertation thesis.

2.1 Computational study

The computational study inquires about the limit distribution

$$\lim_{n\to\infty} P_x^{(n)}, \qquad x = 0, 1, 2, \dots \ .$$

In order to obtain some preliminary insight in this field, a computational study using the R software was performed. Basically, in the R software it is possible to work only with objects of finite length. Therefore, the computational study is restricted by the choice of parent distribution. Only parent distributions with finite supports are considered, which ensures that the nth descendant has also finite support, $n = 1, 2, 3, \ldots$.

We show the results of the computational study with the binomial distribution playing the role of the original parent. We use two different types of partial summation, the Salvia-Bolinger and the Poisson partial summation. The iterated Salvia-Bolinger partial summation yields the deterministic limit descendant distribution for many different combinations of parameters. The limit descendant of the Poisson partial summation exists too, however, is not deterministic.

2.2 Power method

In this section we present analysis of the power method's use in the field of iterated partial summations, which is also the scope of our recently published paper [4].

Consider the partial summation

$$P_x = c \sum_{j=x}^{\infty} g(j) P_j^*, \qquad x = 0, 1, 2, \ldots, \tag{14}$$

where $g(j)$ is a real function and c a normalization constant (which ensures that the sequence $\{P_j\}_{j=0}^{\infty}$ is a proper probability distribution, i.e., it sums to 1). The distributions $\{P_j^*\}_{j=0}^{\infty}$ and $\{P_j\}_{j=0}^{\infty}$ are called parent and descendant, respectively. Hereafter, we restrict ourselves to parent distributions

$$\{P_0^*, P_1^*, P_2^*, \ldots, P_{S-1}^*\},$$

i.e. to discrete distributions defined on a finite support of the size S. As the probabilities P_j^* in (14) are zero for $j \geq S$ in this case, the partial summation (14) can be written as

$$P_x = c \sum_{j=x}^{S-1} g(j) P_j^*, \qquad x = 0, 1, \ldots, S-1,$$

or equivalently as

$$\mathbb{P} = cA\mathbb{P}^*, \tag{15}$$

where $\mathbb{P}$ and $\mathbb{P}^*$ are the vectors of probabilities

$$(P_0, P_1, \ldots, P_{S-1})^\top$$

and

$$(P_0^*, P_1^*, \ldots, P_{S-1}^*)^\top,$$

respectively. Matrix A is of dimension $S \times S$, with the following structure:

$$A = \begin{pmatrix} g(0) & g(1) & g(2) & g(3) & \dots & g(S-1) \\ 0 & g(1) & g(2) & g(3) & \dots & g(S-1) \\ 0 & 0 & g(2) & g(3) & \dots & g(S-1) \\ 0 & 0 & 0 & g(3) & \dots & g(S-1) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & g(S-1) \end{pmatrix} . \tag{16}$$

The partial summation (14) can be applied iteratively as we describe in the beginning of this chapter, however here we are restricted to finite-support parents only. The descendant distribution becomes a parent of another distribution, i.e.

$$\begin{aligned} P_x^{(1)} =& c_1 \sum_{j=x}^{S-1} g(j) P_j^*, \qquad x = 0, 1, \dots, S-1 \ , \\ P_x^{(2)} =& c_2 \sum_{j=x}^{S-1} g(j) P_j^{(1)}, \qquad x = 0, 1, \dots, S-1 \ , \\ &\vdots \\ P_x^{(n)} =& c_n \sum_{j=x}^{S-1} g(j) P_j^{(n-1)}, \qquad x = 0, 1, \dots, S-1 \ , \\ &\vdots \end{aligned}$$

with c_n, $n = 1, 2, 3, \dots$ being normalization constants. The distribution

$$\{P_x^*\}_{x=0}^{S-1}$$

will be called the original parent. We will now investigate properties of the sequence of the descendant distributions, especially the question under which conditions the limit of this sequence exists. We remind that the existence of the limit for iterated partial summations applied to discrete distributions with infinite supports for a constant function $g(j)$ was proven in [11].

In the following, we will not consider the normalization constants. Then the matrix notation (see (15)) of the iterated partial summations is

$$\begin{aligned}
\mathbb{Q}^{(1)} =& A\mathbb{P}^*,\\
\mathbb{Q}^{(2)} =& A\mathbb{Q}^{(1)} = AA\mathbb{P}^* = A^2\mathbb{P}^*,\\
\mathbb{Q}^{(3)} =& A\mathbb{Q}^{(2)} = AA^2\mathbb{P}^* = A^3\mathbb{P}^*,\\
&\vdots\\
\mathbb{Q}^{(n)} =& A\mathbb{Q}^{(n-1)} = A^n\mathbb{P}^*,\\
&\vdots
\end{aligned}$$

The i-th descendant probability distribution can be obtained by the normalization of the vector $\mathbb{Q}^{(i)} = (Q_0^{(i)}, Q_1^{(i)}, \dots, Q_{S-1}^{(i)})^\top$.

Denote $\|U\|_1$, $\|U\|_2$ the L1-norm and the L2-norm of vector U, respectively. If the limit of the sequence of the descendant distributions exists, it can be written as

$$\mathbb{P}^{(\infty)} = \lim_{n\to\infty} \frac{\mathbb{Q}^{(n)}}{\|\mathbb{Q}^{(n)}\|_1} = \lim_{n\to\infty} \frac{A^n\mathbb{P}^*}{\|A^n\mathbb{P}^*\|_1}. \tag{17}$$

The power method is one of computational approaches to the problem of finding matrix eigenvalues (see e.g. [2], pp. 330-332). We apply it to matrix A (denote its eigenvalues by $\lambda_0, \dots, \lambda_{S-1}$; we remind that in general they need not be distinct) and vector $\mathbb{P}^*$ from (15). To satisfy the conditions of the method, suppose that A is diagonalizable and that it has a unique dominant eigenvalue λ_k (i.e., there exists k such that $|\lambda_k| > |\lambda_i|$, $i \neq k$).

If $\mathbb{P}^*$ is not a non-dominant eigenvector of A and, in addition, if $\mathbb{P}^*$ is such a linear combination of the eigenvectors of A that the coefficient corresponding to the dominant eigenvector is non-zero, then

$$\lim_{n\to\infty} \frac{A^n\mathbb{P}^*}{\|A^n\mathbb{P}^*\|_2} = V,$$

where V is the dominant eigenvector of A (i.e., the one which corresponds to the dominant eigenvalue). Under these conditions, the power method implies the existence of $\lim_{n\to\infty} \mathbb{P}^{(n)}$, see (17), with

$$\mathbb{P}^{(\infty)} = \lim_{n\to\infty} \mathbb{P}^{(n)} = \frac{V}{\|V\|_1}.$$

The matrix A from (16) is an upper triangular matrix, which means that its eigenvalues are its diagonal entries, i.e.

$$\lambda_j = g(j), \qquad j = 0, 1, \dots, S-1.$$

Consequently, to determine the dominant eigenvalue λ_D of A it is necessary to find

$$D = \arg \max_{j \in \{0,1,\dots,S-1\}} |g(j)|.$$

Let $D = k$, i.e., let the dominant eigenvalue be $\lambda_k = g(k)$. The eigenvector corresponding to the dominant eigenvalue λ_k is the solution of the of linear equation

$$AV = \lambda_k V,$$

or, equivalently,

$$(A - \lambda_k I)V = 0.$$

This system of linear equations yields the solution

$$V = \begin{pmatrix} t \\ t\left(1 - \frac{g(0)}{g(k)}\right) \\ t\left(1 - \frac{g(0)}{g(k)}\right)\left(1 - \frac{g(1)}{g(k)}\right) \\ \vdots \\ t\prod_{j=1}^{k}\left(1 - \frac{g(j-1)}{g(k)}\right) \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad t \in \mathbb{R}.$$

In the dissertation thesis, we present detailed results for iterated summations applied to the distributions from the Katz family.

The power method solves the question of the limit of iterated partial summations for a wide spectrum of the partial summation types and of the original parent distributions. However, the power method has its conditions under which it can be applied. We discuss them in the thesis, together with some examples.

Summary

The field of partial summations within the theory of discrete probability distributions offers a wide spectrum of unsolved problems. Our research managed to solve some of them.

First, we bring an answer to the problem of invariance when the partial summation is parametrized. The parametrization of partial summations defines the family of resistant discrete distributions and the family of sensitive discrete distributions. We formulate the sufficient and necessary condition for a distribution to belong to the family of resistant distributions. So far only the geometric and the Salvia-Bolinger distributions are known to belong to this family. We also provide a useful necessary condition for a distribution to belong to the resistant family. It is easy to evaluate it and therefore to identify or to reject other candidates. Finding more distributions from the resistant family, or proving that the resistant family consists of the two above mentioned distributions only, is one of the possible directions of future research in this field.

Second, we suggest an approach to solve the problem of the existence of the limit for iterated partial summation if the original parent has a finite support. Results of our computational study revealed that the iterated application of the Poisson partial summation to a parent with a finite support yields a limit distribution which is not deterministic. This fact encouraged a deeper research on the existence of the limit descendant distribution for finite-supported parent distributions. The idea to reformulate the whole partial summation process as a matrix multiplication allowed us to use the power method. It proves the existence of the limit descendant distribution for a very wide class of parent distributions only with minor exceptions, which we also define. A detailed analysis of the iterated Katz partial summation is provided. We also identify some specific cases, when the power method cannot be applied. The existence of the limit in these cases remains questionable. In future, these results will be generalized to parent distributions with the infinite support.

References

[1] B. H. Bissinger. A type resisting distribution generated from consideration of an inventory decision model. In G. P. Patil, editor, *Classical and Contagious Discrete Distributions*, pages 15–17. Calcutta: Statistical Publishing House and Oxford: Pergamon Press, 1965.

[2] G. H. Golub and C. F. Van Loan. *Matrix Computations.* Baltimore, London: The John Hopkins University Press, 1996.

[3] N. L. Johnson, A. W. Kemp, and S. Kotz. *Univariate Discrete Distributions, 3rd edition.* Hoboken: Wiley, 2005.

[4] M. Koščová, J. Mačutek, and R. Harman. Iterated partial summations applied to finite-support discrete distributions. *Mathematica Slovaca*, 70:489–496, 2020.

[5] S. Li. On a class of discrete time renewal risk models. *Scandinavian Actuarial Journal*, 4:241–260, 2005.

[6] S. Li, Y. Lu, and J. Garrido. A review of discrete-time risk models. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas*, 103/2:321–337, 2009.

[7] X. S. Lin and G. E. Willmot. Analysis of a defective renewal equation arising in ruin theory. *Insurance: Mathematics and Economics*, 25:63–84, 1999.

[8] J. Mačutek. Katz partial summations family. *Tatra Mountains Mathematical Publications*, 22:143–148, 2001.

[9] J. Mačutek. Discrete probability distributions generated by the generalized STER summation. *Brazilian Journal of Probability and Statistics*, 16:1–6, 2002.

[10] J. Mačutek. On two types of partial summations. *Tatra Mountains Mathematical Publications*, 26:403–410, 2003.

[11] J. Mačutek. A limit property of the geometric distribution. *Theory of Probability and Its Applications*, 50/2:316–319, 2006.

[12] J. Mačutek and G. Altmann. Discrete and continuous modelling in quantitative linguistics. *Journal of Quantitative Linguistics*, 14/1:81–94, 2007.

[13] S. Naranan and V. K. Balasubrahmanyan. Power laws in statistical linguistics and related systems. In R. Köhler, G. Altmann, and R. G. Piotrowski, editors, *Quantitative Linguistics. An International Handbook*, pages 716–738. Berlin: Mouton de Gruyter, 2005.

[14] N. Unnikrishnan Nair and N. Hitha. Characterization of discrete models by distribution based on their partial sums. *Statistics & Probability Letters*, 8:335–337, 1989.

[15] N. Unnikrishnan Nair, P. G. Sankaran, and M. Preeth. Reliability aspects of discrete equilibrium distributions. *Communications in Statistics - Theory and Methods*, 41/3:500–515, 2012.

[16] G. Willmot. Mixed compound Poisson distribution. *ASTIN Bulletin*, 16:59–79, 1986.

[17] G. Wimmer and G. Altmann. On the generalization of the STER distribution applied to generalized hypergeometric parents. *Acta Universitatis Palackianae Olomoucensis. Facultas Rerum Naturalium. Mathematica*, 39/1:215–247, 2000.

[18] G. Wimmer and G. Altmann. Models of rank-frequency distributions in language and music. In L. Uhlířová, G. Wimmer, G. Altmann, and R. Köhler, editors, *Text as a Linguistic Paradigm: Levels, Constituents, Constructs*, pages 283–294. Trier: WVT, 2001.

[19] G. Wimmer and G. Altmann. A new type of partial-sums distributions. *Statistics & Probability Letters*, 52:359–364, 2001.

[20] G. Wimmer and J. Kalas. A characterization of the geometric distribution. *Tatra Mountains Mathematical Publications*, 17:325–329, 1999.

[21] G. Wimmer and J. Mačutek. New integrated view at partial-sums distributions. *Tatra Mountains Mathematical Publications*, 51:183–190, 2012.

[22] E. Xekalaki. A property of the Yule distribution and its applications. *Communications in Statistics - Theory and Methods*, 12/10:1181–1189, 1983.

Zoznam publikácií a ohlasov

- **Koščová, M.**, Harman, R., Mačutek, J. (2020). Iterated partial summations applied to finite-support discrete distributions. *Mathematica Slovaca* 70:2, 489–496.

  citované v:

  - Leššová. L. (2019). Oscillating sequences of partial-sums discrete probability distributions. In: Milošević, B., Obradović, M. (eds.), *Proceedings of the 21st European Young Statisticians Meeting* (pp. 41-45). Beograd: Faculty of Mathematics.

- Mačutek, J., Chromý, J., **Koščová, M.** (2019). Menzerath-Altmann law and prothetic /v/ in spoken Czech. *Journal of Quantitative Linguistics* 26:1, 66-80.

  citované v:

  - Hernández-Fernández, A., Torre, I. G., Garrido, J. M., Lacasa, L. (2019). Linguistic laws in speech: The case of Catalan and Spanish. *Entropy* 21, 1153.
  - Torre, I. G., Luque, B., Lacasa, L., Kelo, C. T., Henández-Fernández, A. (2019). On the physical origins of linguistic laws and lognormality in speech. *Royal Society Open Science* 6, 191023.

- Čech, R., Milička, J., Mačutek, J., **Koščová, M.**, Lopatková, M. (2018). Quantitative analysis of syntactic dependency in Czech. In: Jiang, J., Liu, H. (eds.), *Quantitative Analysis of Dependency Structures* (pp. 53-70). Berlin/Boston: Walter de Gruyter.

- **Koščová, M.**, Čech, R., Mačutek, J. (2017). Shluková analýza. In: Karlík, P., Nekula, M., Pleskalová J. (eds.), *CzechEncy – Nový encyklopedický slovník češtiny.* URL: https://www.czechency.org/slovnik/SHLUKOVÁ ANALÝZA

- **Koščová, M.**, Mačutek, J., Kelih, E. (2016). A data-based classification of Slavic languages: Indices of qualitative variation applied to grapheme frequencies. *Journal of Quantitative Linguistics* 23:2, 177-190.

  citované v:

  - Rovenchak, A., Vydrin, V. (2020). Syllable frequencies in Manding: Examples from periodicals in Bamana and Maninka. *Glottometrics* 48, 17-36.

  - Rovenchak, A., Riley, C., Sherman, T. (2018). The diary of Boima Kiakpombgo from Mando town (Liberia): a quantitative study of a Vai text. *Journal of Quantitative Linguistics* 25, 271-287.

- Mačutek, J., **Koščová, M.**, Čech, R. (2016). Lexical compactness across genres in works by Karel Čapek. In: Mayaffre, D., Poudat, C., Vanni, L., Magri, V., Follette, P. (eds.), *Statistical Analysis of Textual Data* (pp. 825-832). Nice: University Nice Sophia Antipolis.

- Čech, R., Mačutek, J., **Koščová, M.** (2015). On the relation between verb full valency and synonymy. In: Hajičová, E., Nivre, J. (eds.), *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)* (pp. 68-73). Uppsala: Uppsala University.

  citované v:

  - Urešová, Z., Fučíková, E., Hajič, J., Hajičová, E. (2018). Defining verbal synonymy: Between syntax and semantics. In: Haug, D., Oepen, S., Øvrelid, L., Candito, M., Hajič, J. (eds.), *Proceedings of the 17th International Workshop on Treebanks and Linguistic Theory (TLT 2018)* (pp. 75-90). Linköping: Linköping University Electronic Press.

- **Cehlárová, M.** (2013). Modifikácia Ordovho grafu a jeho aplikácia na frekvencie grafém v slovanských jazykoch. *Výpočtová štatistika 2013: Prehliadka prác mladých štatistikov a demografov*, 18-23. Bratislava: Slovenská štatistická a demografická spoločnosť.

Účasť na konferenciách

- ROBUST 2016 - Letní škola Jednoty českých matematiků a fyziků, Loučná nad Desnou, Česká republika. Poster a prednáška s názvom *Parametrizované parciálne sumácie* (spoluautori J. Mačutek, G. Wimmer)

- QUALICO 2016 - International Quantitative Linguistics Conference, Trier, Nemecko. Prednáška s názvom *Quantitative Analysis of Syntactic Dependency in Czech* (spoluautori R. Čech, J. Milička, J. Mačutek, M. Lopatková)

- JADT 2016 - 13th International Conference on Statistical Analysis of Textual Data, Nice, Francúzsko. Poster s názvom *Lexical compactness across genres in works by Karel Čapek* (spoluautori J. Mačutek, R. Čech)

- SSoL 2015 - The 9th Summer School of Linguistics, Dačice, Česko. Workshop s názvom *Cluster analysis and its application in the classification of languages/texts* (spoluautor J. Mačutek)
- DEPLING 2015 - the Third International Conference on Dependency Linguistics, Uppsala, Švédsko. Poster s názvom *On the relation between verb full valency and synonymy* (spoluautori R. Čech, J. Mačutek)
- PROBASTAT 2015 - The Seventh International Conference on Probability and Statistics, Smolenice, Slovensko. Poster s názvom *Partial summations divide discrete distributions into two classes* (spoluautori J. Mačutek, G. Wimmer)
- ODAM 2015 - Olomoucian Days of Applied Mathematics, Olomouc, Česko. Prednáška s názvom *Indices of qualitative variation used for graphical representation of categorical data* (spoluautor J. Mačutek)
- Výpočtová štatistika 2013 - Prehliadka prác mladých štatistikov a demografov, Bratislava, Slovensko. Prednáška s názvom *Modification of the Ord's graph and its application to the grapheme frequencies in Slavic languages*

Granty

- Nové štatistické metódy pre špeciálne triedy rozdelení a ich aplikácie (spoluriešiteľ)
  Grant VEGA 2/0054/18
  Hlavný riešiteľ: prof. RNDr. Gejza Wimmer, DrSc.
- Aplikácie matematickej štatistiky v lingvistike (hlavný riešiteľ)
  Grant UK/138/2016
- Štatistiké analýzy lingvistických dát (hlavný riešiteľ)
  Grant UK/83/2015
- Diskrétne a spojité pravdepodobnostné modely a ich aplikácie (spoluriešiteľ)
  Grant VEGA 2/0047/15
  Hlavný riešiteľ: prof. RNDr. Gejza Wimmer, DrSc.