VERSITA

TATRA MOUNTAINS
Mathematical Publications
DOI: 10.2478/v10127-012-0018-8
Tatra Mt. Math. Publ. **51** (2012), 183–190

NEW INTEGRATED VIEW AT PARTIAL-SUMS DISTRIBUTIONS

Gejza Wimmer — Ján Mačutek

ABSTRACT. Partial-sums discrete probability distributions occurred in description of many stochastic models. They were used also as a tool for creating new distributions, or as a link between known distributions. It is shown in this paper that every discrete distribution with only non-zero probabilities is a partial-sums distribution, and, moreover, that it has infinitely many parent distributions. The paper generalizes and unifies the concept of partial-sums distribution. Besides, it generalizes some risk models in insurance and revises some approaches to mathematical modelling in quantitative linguistics.

1. Partial-sums probability distributions

Let us have a discrete probability distribution with the probability mass function $\{Q_x\}_{x\in N_0}$ ($N_0 = \{0,1,\dots\}$), its probability generating function being $G(t) = \sum_{x\geq 0} Q_x t^x$, $t \in \langle 0,1\rangle$. Let $f(x,j)$ be a real function. The distribution $\{Q_x\}_{x\in N_0}$ is called a partial-sums distribution with the parent distribution $\{P_x\}_{x\in N_0}$ (or with the parent $\{P_x\}_{x\in N_0}$) when its probability mass function satisfies

$$Q_x = \sum_{j\geq x} f(x,j)P_j, \qquad x \in N_0. \tag{1}$$

Simple special cases of (1) for proper $f(x,j)$ were used in risk models in insurance [7], [8], [9], [14] and [16] and in models in quantitative linguistics and musicology [20]. The partial summation with $f(x,j) = 0$ for $x = 0$ and $f(x,j) = \kappa/j$ for $x = 1,2,\dots$ was analyzed in [19]. Some properties of the geometric distribution under the summation with $f(x,j) = 0$ for $x = j$ and $f(x,j) = \gamma$ for $j > x$ were shown in [12] and [22]. The problem of invariance of discrete distribution under partial summations was solved in general in [10]. Papers [11] and [18]

2010 Mathematics Subject Classification: 60E05, 60E99.
Keywords: partial-sums distribution, Poisson distribution, hypergeometric series.
Supported by the Slovak Research and Development Agency, grant APVV-0096-10 and by the research grant VEGA 2/038/12.

demonstrate that some distributions and families of distributions are connected by partial summations for a particular choice of the function $f(x, j)$ (i.e., a proper function $f(x, j)$ is found for given distributions $\{P_j\}_{j \in N_0}$ and $\{Q_x\}_{x \in N_0}$). Partial--sums distributions are presented also in the comprehensive monograph [4].

2. Main results and examples

The partial-sums distributions are probabilistic models for real stochastic mechanisms but summations (1) were also seen as a tool for creating new discrete distributions (for a given parent distribution $\{P_j^*\}_{j \in N_0}$ some functions $f(x, j)$ were chosen and the resulting partial-sums distributions $\{P_x\}_{x \in N_0}$ were obtained), see, e.g., [15] and [20], or as a link among distributions or distribution families, see [11] and [18]. In this paper we prove that every two discrete probability distributions which are defined on the same support and have only non-zero probabilities are connected by a partial summation. Due to this fact, some applications of partial-sums distributions to mathematical modelling (especially) in quantitative linguistics must be revised. This consequence of our findings will be discussed in Conclusion.

The proof that every two discrete probability distributions defined on the same support with only non-zero probabilities is mathematically trivial, but it has been left unnoticed so far.

Theorem. *Let $\{P_x\}_{x \in \mathtt{N}_0}$ and $\{Q_x\}_{x \in \mathtt{N}_0}$ be discrete probability distributions with $P_x > 0$ for all $x \in \mathtt{N}_0$ and $Q_x > 0$ for all $x \in \mathtt{N}_0$. Then the distributions $\{P_x\}_{x \in \mathtt{N}_0}$ and $\{Q_x\}_{x \in \mathtt{N}_0}$ satisfy* (1) *for*

$$f(x, k) = f(k) = \frac{Q_k - Q_{k+1}}{P_k}, \qquad k \in \mathtt{N}_0. \tag{2}$$

P r o o f. If we replace $f(x, j)$ in (1) with $f(j)$, we have

$$Q_k = \sum_{j \geq k} f(j) P_j, \qquad k \in N_0 \tag{3}$$

and

$$Q_{k+1} = \sum_{j \geq k+1} f(j) P_j, \qquad k \in N_0. \tag{4}$$

(2) is proved by subtracting (4) from (3). □

Similarly, it can be shown that any two discrete distributions which have supports of the same size are connected by a partial summation.

We note that the function $f(j)$ from (2) is a special case of the function $f(x,j)$ from (1), as it is constant with respect to the variable x. Using the function $f(x,j)$ (i.e., the function of two variables) can lead to even a richer system of interrelations among discrete distributions.

Examples.

1. Substituting the geometric and Poisson distribution to (2) we obtain

$$f(j) = \frac{p\,(1-p)^j - p\,(1-p)^{j+1}}{\frac{e^{-\lambda}\lambda^j}{j!}} = \frac{p^2\,(1-p)^j\, j!}{e^{-\lambda}\lambda^j}, \tag{5}$$

which means that the geometric distribution with parameter p is a partial-sums distribution, its parent being the Poisson distribution with parameter λ.

2. Consider the Kemp-hypergeometric family of discrete distributions, that is, the family of distributions with the probability generating function

$$G(t) \;=\; {}_kF_r\big(a_1,\ldots,a_k;\;\; b_1,\ldots,b_r;\;\; \theta(t-1)\big), \quad k,r \geq 1$$

(i.e., the probability generating function of the distributions from the Kemp--hypergeometric family is a generalized hypergeometric function, see [2]). The family contains many well-known distributions, among others the binomial, hypergeometric and negative binomial distributions (see [17, pp. 341–343]). It can be easily seen that

$$\begin{aligned}
G(t) =& 1 + \frac{a_1^{(1)}\ldots a_k^{(1)}}{b_1^{(1)}\ldots b_r^{(1)}}\frac{\theta^1(t-1)^1}{1!} + \frac{a_1^{(2)}\ldots a_k^{(2)}}{b_1^{(2)}\ldots b_r^{(2)}}\frac{\theta^2(t-1)^2}{2!} + \cdots \\
=& \;{}_kF_r(a_1,\ldots,a_k;b_1,\ldots,b_r;-\theta) \\
&+ t\,\left\{\frac{a_1^{(1)}\ldots a_k^{(1)}}{b_1^{(1)}\ldots b_r^{(1)}}\frac{1!}{1!0!}\frac{\theta^1}{1!}(-1)^0 + \frac{a_1^{(2)}\ldots a_k^{(2)}}{b_1^{(2)}\ldots b_r^{(2)}}\frac{2!}{1!1!}\frac{\theta^2}{2!}(-1)^1 + \cdots\right\} \\
&+ t^2\left\{\frac{a_1^{(2)}\ldots a_k^{(2)}}{b_1^{(2)}\ldots b_r^{(2)}}\frac{2!}{2!0!}\frac{\theta^2}{2!}(-1)^0 + \frac{a_1^{(3)}\ldots a_k^{(3)}}{b_1^{(3)}\ldots b_r^{(3)}}\frac{3!}{2!1!}\frac{\theta^3}{3!}(-1)^1 + \cdots\right\} \\
&+ t^3\left\{\frac{a_1^{(3)}\ldots a_k^{(3)}}{b_1^{(3)}\ldots b_r^{(3)}}\frac{3!}{3!0!}\frac{\theta^3}{3!}(-1)^0 + \frac{a_1^{(4)}\ldots a_k^{(4)}}{b_1^{(4)}\ldots b_r^{(4)}}\frac{4!}{3!1!}\frac{\theta^4}{4!}(-1)^1 + \cdots\right\} + \cdots \\
=& \;{}_kF_r(a_1,\ldots,a_k;b_1,\ldots,b_r;-\theta) \\
&+ t^1\frac{a_1^{(1)}\ldots a_k^{(1)}}{b_1^{(1)}\ldots b_r^{(1)}}\frac{\theta^1}{1!}\left\{1 + \frac{(a_1+1)^{(1)}\ldots(a_k+1)^{(1)}}{(b_1+1)^{(1)}\ldots(b_r+1)^{(1)}}\frac{(-\theta)^1}{1!}\right. \\
&\left. + \frac{(a_1+1)^{(2)}\ldots(a_k+1)^{(2)}}{(b_1+1)^{(2)}\ldots(b_r+1)^{(2)}}\frac{(-\theta)^2}{2!} + \cdots\right\}
\end{aligned}$$

$$+\,t^2\frac{a_1^{(2)}\dots a_k^{(2)}}{b_1^{(2)}\dots b_r^{(2)}}\frac{\theta^2}{2!}\left\{1+\frac{(a_1+2)^{(1)}\dots(a_k+2)^{(1)}}{(b_1+2)^{(1)}\dots(b_r+2)^{(1)}}\frac{(-\theta)^1}{1!}\right.$$
$$\left.+\;\frac{(a_1+2)^{(2)}\dots(a_k+2)^{(2)}}{(b_1+2)^{(2)}\dots(b_r+2)^{(2)}}\frac{(-\theta)^2}{2!}+\cdots\right\}+\cdots$$
$$=\sum_{x=0}^{\infty}t^x\frac{a_1^{(x)}\dots a_k^{(x)}}{b_1^{(x)}\dots b_r^{(x)}}\frac{\theta^x}{x!}\ {}_kF_r(a_1+x,\dots,a_k+x;b_1+x,\dots,b_r+x;-\theta)$$

$(x^{(0)}=1,\ x^{(n)}=x(x+1)\dots(x+n-1)$ for $x\in\mathbb{R},\ n\in\mathbb{N}=\{1,2,\dots\})$. Consequently, the probability mass function of the distributions from the Kemp--hypergeometric family can be expressed as

$$P_x=\frac{a_1^{(x)}\dots a_k^{(x)}}{b_1^{(x)}\dots b_r^{(x)}}\frac{\theta^x}{x!}\ {}_kF_r(a_1+x,\dots,a_k+x;\ b_1+x,\dots,b_r+x;-\theta).$$

As it holds

$$\begin{aligned}&e^{\theta}\ {}_kF_r(\alpha_1,\dots,\alpha_k;\beta_1,\dots,\beta_r;-\theta)\\&=\left\{1+\frac{\theta^1}{1!}+\frac{\theta^2}{2!}+\cdots\right\}\left\{1+(-1)^1\frac{\alpha_1^{(1)}\dots\alpha_k^{(1)}}{\beta_1^{(1)}\dots\beta_r^{(1)}}\frac{\theta^1}{1!}\right.\\&\quad\left.+(-1)^2\frac{\alpha_1^{(2)}\dots\alpha_k^{(2)}}{\beta_1^{(2)}\dots\beta_r^{(2)}}\frac{\theta^2}{2!}+\cdots\right\}\\&=\sum_{j=0}^{\infty}\left\{\sum_{i=0}^{j}\binom{j}{i}(-1)^i\frac{\alpha_1^{(i)}\dots\alpha_k^{(i)}}{\beta_1^{(i)}\dots\beta_r^{(i)}}\right\}\frac{\theta^j}{j!},\end{aligned}$$

one can write

$$\begin{aligned}P_x&=e^{-\theta}\frac{a_1^{(x)}\dots a_k^{(x)}}{b_1^{(x)}\dots b_r^{(x)}}e^{\theta}\frac{\theta^x}{x!}\ {}_kF_r(a_1+x,\dots,a_k+x;b_1+x,\dots,b_r+x;-\theta)\\&=\sum_{j=x}^{\infty}\frac{a_1^{(x)}\dots a_k^{(x)}}{b_1^{(x)}\dots b_r^{(x)}}\binom{j}{x}\left\{\sum_{i=0}^{j-x}(-1)^i\binom{j-x}{i}\frac{(a_1+x)^{(i)}\dots(a_k+x)^{(i)}}{(b_1+x)^{(i)}\dots(b_r+x)^{(i)}}\right\}e^{-\theta}\frac{\theta^j}{j!}\\&=\sum_{j=x}^{\infty}\frac{a_1^{(x)}\dots a_k^{(x)}}{b_1^{(x)}\dots b_r^{(x)}}\binom{j}{x}\left\{\sum_{i=0}^{j-x}(-1)^i\binom{j-x}{i}\frac{(a_1+x)^{(i)}\dots(a_k+x)^{(i)}}{(b_1+x)^{(i)}\dots(b_r+x)^{(i)}}\right\}P_j^*.\end{aligned}$$

The last equation is a special case of (1) for

$$f(x,j)=\frac{a_1^{(x)}\dots a_k^{(x)}}{b_1^{(x)}\dots b_r^{(x)}}\binom{j}{x}\left\{\sum_{i=0}^{j-x}(-1)^i\binom{j-x}{i}\frac{(a_1+x)^{(i)}\dots(a_k+x)^{(i)}}{(b_1+x)^{(i)}\dots(b_r+x)^{(i)}}\right\}.$$

We have shown that all distributions from the Kemp-hypergeometric family are partial-sums distributions, having the Poisson distribution as their parent. We have thus substantially generalized Willmot's models from [16].

3. Conclusion

The results achieved have consequences both on the theoretical level and in the applications of partial-sums discrete distributions.

First, according to the theorem proved in the previous section, every discrete distribution with only non-zero probabilities is a partial-sums distribution for some choice of the function $f(x, j)$ and of the parent distribution. Moreover, every discrete distribution with only non-zero probabilities has infinitely many parents. Hence, it does not make any sense to speak about, e.g., a family of partial-sums distributions with Poisson parents—all discrete distributions with only non-zero probabilities on N_0 can be expressed as partial-sums distribution with Poisson parents. There are three elements in the partial summations (1)—the parent distribution, the resulting partial-sums distribution and the function $f(x, j)$ which links them. If one wants to exploit partial summations to build a family of distributions, two of the three elements must be fixed. Otherwise, if one of the elements is fixed, one always obtains all discrete distributions with non-zero probabilities on supports of the same size. Of course, it can happen that two different partial summations (1) with two different parent distributions result in the same partial-sums distribution as shown in the next examples.

Examples.

1. Parent distribution $\{P_j\}_{j\in N_0}$ with probability generating function

   $$\frac{{}_2F_1(2,1;b+3;t)}{{}_2F_1(2,1;b+3;1)} \text{ and probability mass function } \left\{\frac{(j+1)!\, b}{(b+2)^{(j+1)}}\right\}_{j\in N_0},$$

   $b > 0$ (see in [6]) under the summation with $f(x, j) = \frac{1}{j+1}$ is resulting in the same partial-sums distribution $\{Q_x\}_{x\in N_0}$ as parent distribution $\{\tilde{P}_j\}_{j\in N_0}$ with probability generating function

   $$\frac{{}_2F_1(1,1;b+3;t)}{{}_2F_1(1,1;b+3;1)} \text{ and probability mass function } \left\{\frac{j!\, (b+1)}{(b+2)^{(j+1)}}\right\}_{j\in N_0},$$

   $b > 0$ (see in [3], [13]) under the summation with $f(x, j) = \frac{b}{b+1}$ for all j. The resulting Yule distribution $\{Q_x\}_{x\in N_0}$ has probability generating function

$\frac{b}{b+1}\ {}_2F_1(1,1;b+2;t)$ and probability mass function $\left\{\frac{x!\,b}{(b+1)^{(x+1)}}\right\}_{x\in N_0}$,
$b > 0$ (see in [4, p. 287]).

2. Parent distribution $\{P_j\}_{j\in\{0,1,\dots,n\}}$ with probability generating function

$\frac{{}_2F_1(-n,1;-n;t)}{{}_2F_1(-n,1;-n;1)}$ and probability mass function $\left\{\frac{1}{n+1}\right\}_{j\in\{0,1,\dots,n\}}$,

$n \in N_0$ (discrete uniform distribution, see, e.g., in [4]) under the summation with $f(x,j) = \frac{1}{j+1}$ is resulting in the same partial-sums distribution $\{Q_x\}_{x\in\{0,1,\dots,n\}}$ as parent distribution $\{\tilde{P}_j\}_{j\in\{0,1,\dots,n\}}$ with probability generating function

$\frac{{}_3F_2(-n,1,1;-n,2;t)}{{}_3F_2(-n,1,1;-n,2;1)}$ and probability mass function

$$\left\{\frac{1}{(j+1)[\Psi(n+2)-\Psi(1)]}\right\}_{j\in\{0,1,\dots,n\}},$$

Ψ – Digamma function (see in [1], [23]) under the summation with

$$f(x,j) = \frac{1}{1+\mu},$$

μ is the mean of this parent distribution.

The resulting distribution $\{Q_j\}_{j\in\{0,1,\dots,n\}}$ has probability generating function

${}_3F_2(-n,1,1;2,2;1-t)$ and probability mass function

$$\left\{\frac{1}{n+1}\sum_{i=x}^{n}\frac{1}{i+1}\right\}_{x\in\{0,1,\dots,n\}}$$

(see in [5]). Many other examples can be derived from [18].

Second, we mention a suggested application of partial summations in quantitative linguistics. A family of distributions given by the equation

$$\frac{P_{x-1}-P_x}{P_x} = a_0 + \sum_{i=1}^{k_1}\frac{a_{1i}}{(x-b_{1i})^{c1}} + \sum_{i=1}^{k_2}\frac{a_{2i}}{(x-b_{2i})^{c2}} + \cdots \tag{6}$$

was suggested as a general model in this field, see [21]. The equation is interpreted as an equilibrium state between influences of a speaker and a hearer. All particular linguistic laws are supposed to be special cases of this general model. Under this assumption the question appears what to do with models which do not fit the equation (6). The paper [11] tries to answer the question by presenting such models as partial-sums distributions with a parent from the family (6). Here we have shown that this attempt fails, since all discrete distributions with non-zero probabilities can be presented in this way and one would obtain also models which are not linguistically interpretable.

REFERENCES

[1] ESTOUP, J. B.: *Les Gammes Sténographiques.* Gauthier-Villars, Institute Sténographique, Paris, 1916.

[2] GASPER, G.—RAHMAN, R.: *Basic Hypergeometric Series.* Cambridge University Press, Cambridge, 1990.

[3] GIPPS, P. G.: *A queueing model for traffic flow*, J. Roy. Statist. Soc. **39** (1977), 276–282.

[4] JOHNSON, N.L.—KEMP, A.W.—KOTZ, S.: *Univariate Discrete Distributions* (3rd ed.), Wiley, Hoboken, NJ, 2005.

[5] KEMP, C. D.—KEMP, A. W.: *Some distributions arising from an inventory decision problem*, Bull. Int. Statist. Inst. **43** (1969), 367–369.

[6] KRISHNAJI, N.: *A characteristic property of the Yule distribution*, Sankhyā Ser. A **32** (1970), 343–346.

[7] LI, S.: *On a class of discrete time renewal risk models,* Scand. Actuar. J. **4** (2005), 241–260.

[8] LI, S.—LU, Y.—GARRIDO, J.: *A review of discrete-time risk models*, Rev. R. Acad. Cienc. Exactas Fs. Nat., Ser. A Mat. **103** (2009), 321–337.

[9] LIN, X. S.—WILLMOT, G. E.: *Analysis of a defective renewal equation arising in ruin theory*, Insurance Math. Econom. **25** (1999), 63–84.

[10] MAČUTEK, J.: *On two types of partial summations,* Tatra Mt. Math. Publ. **26** (2002), 403–410.

[11] MAČUTEK, J.: *Discrete distributions connected by partial summations,* Glottometrics **11** (2005), 45–52.

[12] MAČUTEK, J.: *A limit property of the geometric distribution,* Theory Probab. Appl. **50** (2006), 316–319.

[13] MILLER, A. J.: *A queueing model for road traffic flow,* J. Roy. Statist. Soc. Ser. B **23** (1961), 64–75.

[14] NAIR, N. U.—HITHA, N.: *Characterization of discrete models by distribution based on their partial sums,* Statist. Probab. Lett. **8** (1989), 335–337.

[15] NARANAN, S.—BALASUBRAHMANYAN, V. K.: *Power Laws in Statistical Linguistics and Related Systems.* Quantitative Linguistics. An International Handbook (Köhler, R., Altmann, G., Piotrowski, R. G., eds.), de Gruyter, Berlin, 2005, pp. 716–738.

[16] WILLMOT, G.: *Mixed compound Poisson distributions,* ASTIN Bull. **16** (1986), 59–79.

[17] WIMMER, G.—ALTMANN, G.: *Thesaurus of Univariate Discrete Probability Distributions.* Stamm, Essen, 1999.

[18] WIMMER, G.—ALTMANN, G.: *On the generalization of the STER distribution applied to generalized hypergeometric parents,* Acta Univ. Palack. Olomuc. Fac. Rerum Natur. Math. **39** (2000), 215–247.

[19] WIMMER, G.—ALTMANN, G.: *A new type of partial-sums distributions,* Statist. Probab. Lett. **52** (2001), 359–364.

[20] WIMMER, G.—ALTMANN, G.: *Models of rank-frequency distributions in language and music,* in: Text as a Linguistic Paradigm: Levels, Constituents, Constructs (Uhlířová, L., Wimmer, G., Altmann, G., Köhler, R., eds.), WVT, Trier, 2001, pp. 283–294.

[21] WIMMER, G.—ALTMANN, G.: *Unified derivation of some linguistic laws,* in: Quantitative Linguistics. An International Handbook (Köhler, R., Altmann, G., Piotrowski, R. G., eds.), de Gruyter, Berlin, 2005, pp. 791–807.

[22] WIMMER, G.—KALAS, J.: *A characterization of the geometric distribution*, Tatra Mt. Math. Publ. **17** (1999), 325–329.

[23] ZÖRNIG, P.—ALTMANN, G.: *Unified representation of Zipf distributions,* Comput. Statist. Data Anal. **19** (1995), 461–473.

Received November 2, 2011

Gejza Wimmer
Institute of Mathematics and Statistics
Masaryk University
Kotlářská 2
CZ–611-37 Brno
CZECH REPUBLIC

Mathematical Institute
Slovak Academy of Sciences
Štefánikova 49
SK–814-73 Bratislava
SLOVAKIA

Faculty of Natural Sciences
Matej Bel University
Tajovského 40
SK–974-01 Banská Bystrica
SLOVAKIA
E-mail: wimmer@mat.savba.sk

Ján Mačutek
Department of Applied Mathematics and Statistics
Faculty of Mathematics, Physics and Informatics
Comenius University
Mlynská dolina
SK–842-48 Bratislava
SLOVAKIA
E-mail: macutek@fmph.uniba.sk