

Discrete probability distributions generated by partial summations

Ján Mačutek

Mathematical Institute, Slovak Academy of Sciences

Department of Mathematics, Constantine the Philosopher
University in Nitra

28 February 2023

Outline

- 1 Partial-sums discrete probability distributions
- 2 Repeated partial summations
 - Repeated geometric partial summations
 - Computational study in R
- 3 Matrix eigenvalues and eigenvectors, and repeated partial summations
 - Matrix notation of partial-sums distributions
 - Repeated Poisson partial summations
- 4 Parametrized partial summations

- 1 Partial-sums discrete probability distributions
- 2 Repeated partial summations
 - Repeated geometric partial summations
 - Computational study in R
- 3 Matrix eigenvalues and eigenvectors, and repeated partial summations
 - Matrix notation of partial-sums distributions
 - Repeated Poisson partial summations
- 4 Parametrized partial summations

Partial summations - definition

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

- $\{P_j^*\}_{j=0}^{\infty}$ - discrete probability distribution (parent)
- $\{P_x\}_{x=0}^{\infty}$ - discrete probability distribution (descendant)
- $g(j)$ - real function (which specifies the partial summation)
- some special cases mentioned in the monograph Univariate Discrete Distributions (Johnson, N.L., Kemp, A.W, Kotz, S., 2005, Wiley)

Partial summations - more in detail

$$P_0 = g(0)P_0^* + g(1)P_1^* + g(2)P_2^* + \dots$$

$$P_1 = \quad \quad \quad g(1)P_1^* + g(2)P_2^* + \dots$$

$$P_2 = \quad \quad \quad \quad \quad \quad g(2)P_2^* + \dots$$

⋮

Partial summations - history

- some special cases have real-world motivations
 - $g(j) = c$ - reliability theory
 - $g(j) = c/j$ - economy (income underreporting), management (ideal time to order new supplies in a shop)
- from purely mathematical point of view, in the (not so distant) past partial summations were seen as a tool for creating new distributions and/or establishing links between distributions
- relations between properties (moments, probability generating functions) of parents and descendants derived¹

¹Mačutek, J. (2003), On two types of partial summations. *Tatra Mountains Mathematical Publications* 26, 403-410.

Partial summations - present

- any two discrete distributions defined on the same support are connected by partial summation²
- this partial summation is uniquely determined

²Wimmer, G., Mačutek, J. (2012). New integrated view at partial-sums distributions. *Tatra Mountains Mathematical Publications* 51, 183-190.

Invariance with respect to partial summations

- for all discrete distributions there is $g(j)$ such that the parent and the descendant distributions are the same, i.e., $P_x^* = P_x$ for $x = 0, 1, 2, \dots$
- hence, the distribution $\{P_j^*\}_{j=0}^{\infty}$ is invariant with respect to the partial summation given by $g(j)$
- function $g(j)$ is uniquely determined
- function $g(j)$ is a (new) characteristic of discrete distributions

Invariance - examples

Poisson distribution

$$P_x^* = \sum_{j=x}^{\infty} \frac{j - \lambda + 1}{j + 1} P_j^*, \quad x = 0, 1, 2, \dots$$

geometric distribution³

$$P_x^* = \sum_{j=x}^{\infty} p P_j^*, \quad x = 0, 1, 2, \dots$$

³Wimmer, G., Kalas, J. (1999). A characterization of the geometric distribution. *Tatra Mountains Mathematical Publications* 17, 325-329.

$$g(j) = 1 - \frac{P_{j+1}^*}{P_j^*}, \quad j = 0, 1, 2, \dots$$

⁴Mačutek, J. (2003). On two types of partial summations. *Tatra Mountains Mathematical Publications* 26, 403-410.

Partial summations - summary

- the world of discrete distributions seen from a new point of view
- new properties of distributions and relations among them appear which are difficult (impossible?) to notice from 'old' perspectives

Partial summations - open problem no. 1

- for every pair of distributions P_x and P_x^* there exists function $g(j)$ such that

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

- but function $g(j)$ uniquely determines another distribution (the third one)
- hence, we have a 'family' consisting of three distributions - in fact, many such families
- what exactly is the relation among the three distributions, in which sense are they related?

- 1 Partial-sums discrete probability distributions
- 2 Repeated partial summations
 - Repeated geometric partial summations
 - Computational study in R
- 3 Matrix eigenvalues and eigenvectors, and repeated partial summations
 - Matrix notation of partial-sums distributions
 - Repeated Poisson partial summations
- 4 Parametrized partial summations

Repeated partial summations

$$P_x^{(1)} = c_1 \sum_{j=x}^{\infty} g(j) P_j^*, \quad x = 0, 1, 2, \dots$$

$$P_x^{(2)} = c_2 \sum_{j=x}^{\infty} g(j) P_j^{(1)}, \quad x = 0, 1, 2, \dots$$

⋮

$$P_x^{(n)} = c_n \sum_{j=x}^{\infty} g(j) P_j^{(n-1)}, \quad x = 0, 1, 2, \dots$$

⋮

- we want to find $\lim_{n \rightarrow \infty} P_x^{(n)}$, $x = 0, 1, 2, \dots$, if it exists

- 1 Partial-sums discrete probability distributions
- 2 Repeated partial summations
 - Repeated geometric partial summations
 - Computational study in R
- 3 Matrix eigenvalues and eigenvectors, and repeated partial summations
 - Matrix notation of partial-sums distributions
 - Repeated Poisson partial summations
- 4 Parametrized partial summations

Repeated geometric partial summations

$$P_x^{(1)} = c_1 \sum_{j=x}^{\infty} P_j^*, \quad x = 0, 1, 2, \dots$$

$$P_x^{(2)} = c_2 \sum_{j=x}^{\infty} P_j^{(1)}, \quad x = 0, 1, 2, \dots$$

⋮

$$P_x^{(n)} = c_n \sum_{j=x}^{\infty} P_j^{(n-1)}, \quad x = 0, 1, 2, \dots$$

⋮

Repeated geometric partial summations

Theorem⁵

If

$$\lim_{x \rightarrow \infty} \frac{P_{x+1}^*}{P_x^*} = q \in (0, 1),$$

then

$$\lim_{n \rightarrow \infty} P_x^{(n)} = (1 - q)q^x, \quad x = 0, 1, 2, \dots$$

⁵Mačutek, J. (2006). A limit property of the geometric distribution. *Theory of Probability and Its Application* 50, 316-319.

Repeated geometric partial summations

Theorem⁵

If

$$\lim_{x \rightarrow \infty} \frac{P_{x+1}^*}{P_x^*} = q \in (0, 1),$$

then

$$\lim_{n \rightarrow \infty} P_x^{(n)} = (1 - q)q^x, \quad x = 0, 1, 2, \dots$$

- The proof uses probability generating functions

$$G_n(t) = \sum_{j=0}^{n-1} (-1)^j \frac{t^j \prod_{x=0}^j c_{n-x}}{(1-t)^{j+1}} + (-1)^n \frac{t^n \prod_{j=1}^n c_j}{(1-t)^n} G^*(t)$$

⁵Mačutek, J. (2006). A limit property of the geometric distribution. *Theory of Probability and Its Application* 50, 316-319.

- 1 Partial-sums discrete probability distributions
- 2 Repeated partial summations
 - Repeated geometric partial summations
 - Computational study in R
- 3 Matrix eigenvalues and eigenvectors, and repeated partial summations
 - Matrix notation of partial-sums distributions
 - Repeated Poisson partial summations
- 4 Parametrized partial summations

Computational study in R

- parent distribution must have finite support
- we will present results for binomial distribution as the parent
- many different partial summations (i.e., many different functions $g(j)$) applied to the binomial distribution

Salvia-Bolinger partial summation

$$Q_j = \frac{a \prod_{k=1}^j (k - a)}{(j + 1)!} \implies g(j) = 1 - \frac{Q_{j+1}}{Q_j} = \frac{a + 1}{j + 2}$$

$$P_x^{(n)} = c_n \sum_{j=x}^{\infty} \frac{a + 1}{j + 2} P_j^{(n-1)}, \quad x = 0, 1, 2, \dots$$

$$n = 1, 2, 3, \dots$$

Salvia-Bolinger partial summation

$$P_x^{(1)} = c_1 \sum_{j=x}^{\infty} \frac{a+1}{j+2} P_j^*, \quad x = 0, 1, 2, \dots$$

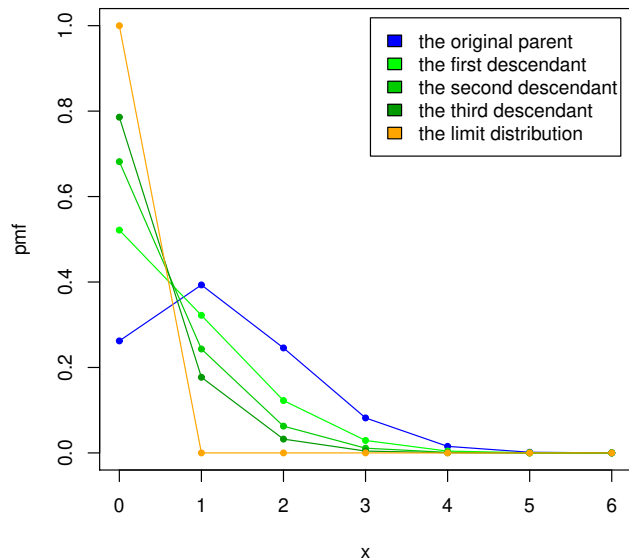
$$P_x^{(2)} = c_2 \sum_{j=x}^{\infty} \frac{a+1}{j+2} P_j^{(1)}, \quad x = 0, 1, 2, \dots$$

⋮

$$P_x^{(n)} = c_n \sum_{j=x}^{\infty} \frac{a+1}{j+2} P_j^{(n-1)}, \quad x = 0, 1, 2, \dots$$

⋮

Salvia-Bolinger partial summation



- parent - binomial
 $n = 6, p = 0.2$
- summation
Salvia-Bolinger,
 $a = 0.5$
- limit
deterministic??

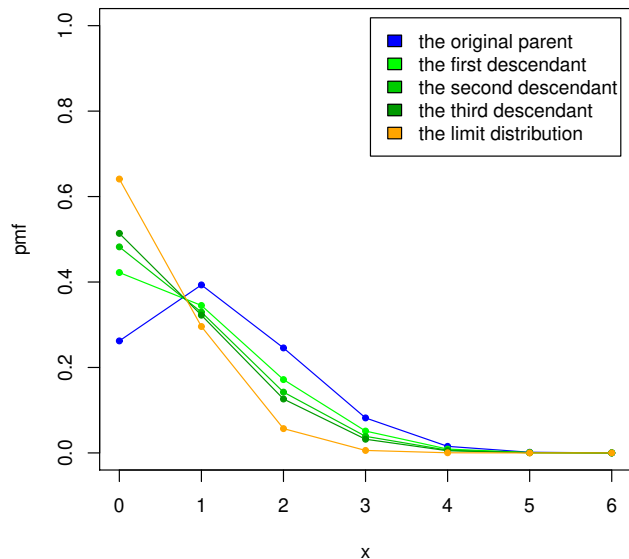
Poisson partial summation

$$Q_j = \frac{e^{-a} a^j}{j!} \quad \Rightarrow \quad g(j) = 1 - \frac{Q_{j+1}}{Q_j} = \frac{j+1-a}{j+1}$$

$$P_x^{(n)} = c_n \sum_{j=x}^{\infty} \frac{j+1-a}{j+1} P_j^{(n-1)}, \quad x = 0, 1, 2, \dots$$

$$n = 1, 2, 3, \dots$$

Poisson partial summation



- parent - binomial
 $n = 6, p = 0.2$
- summation
Poisson
 $a = 0.5$
- limit
not deterministic??

- 1 Partial-sums discrete probability distributions
- 2 Repeated partial summations
 - Repeated geometric partial summations
 - Computational study in R
- 3 Matrix eigenvalues and eigenvectors, and repeated partial summations
 - Matrix notation of partial-sums distributions
 - Repeated Poisson partial summations
- 4 Parametrized partial summations

Power method

- let all coordinates of vector \mathbb{P}^* be nonzero
- let matrix A have only one dominant eigenvalue λ_D , let $\lambda_D \in \mathbb{R}$
- let v be the eigenvector associated with eigenvalue λ_D

$$\lim_{n \rightarrow \infty} \frac{A^n \mathbb{P}^*}{\|A^n \mathbb{P}^*\|} = v$$

$$\lim_{n \rightarrow \infty} \left(A \frac{A^n \mathbb{P}^*}{\|A^n \mathbb{P}^*\|} \right)^\top \left(\frac{A^n \mathbb{P}^*}{\|A^n \mathbb{P}^*\|} \right) = \lambda_D$$

Remarks on power method

- known also as von Mises iteration algorithm
- used by Google search engine
- convergence depends on $\frac{|\lambda_D|}{|\lambda_2|}$ - the higher the ratio, the better
- different norms can be used

- 1 Partial-sums discrete probability distributions
- 2 Repeated partial summations
 - Repeated geometric partial summations
 - Computational study in R
- 3 Matrix eigenvalues and eigenvectors, and repeated partial summations
 - **Matrix notation of partial-sums distributions**
 - Repeated Poisson partial summations
- 4 Parametrized partial summations

Matrix notation of partial-sums distributions

we consider only distributions defined on finite support

$$P_x = \sum_{j=x}^{q-1} g(j)P_j^*, \quad x = 0, 1, \dots, q-1 \quad \iff \quad \mathbb{P} = A\mathbb{P}^*$$

$$\begin{aligned} \mathbb{P}^* &= (P_0^*, P_1^*, \dots, P_{q-1}^*)^\top \\ \mathbb{P} &= (P_0, P_1, \dots, P_{q-1})^\top \end{aligned} \quad A = \begin{pmatrix} g(0) & g(1) & \dots & g(q-1) \\ 0 & g(1) & \dots & g(q-1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & g(q-1) \end{pmatrix}$$

Matrix notation of repeated partial summations

- neglect for a while normalization constants

$$\mathbb{P}^{(1)} = A\mathbb{P}^*$$

$$\mathbb{P}^{(2)} = A\mathbb{P}^{(1)} = AA\mathbb{P}^* = A^2\mathbb{P}^*$$

$$\vdots$$

$$\mathbb{P}^{(n)} = A\mathbb{P}^{(n-1)} = A^n\mathbb{P}^*$$

$$\vdots$$

- the n -th descendant has probability mass function $\frac{\mathbb{P}^{(n)}}{\|\mathbb{P}^{(n)}\|_1}$

$$\mathbb{P}^{(\infty)} = \lim_{n \rightarrow \infty} \frac{\mathbb{P}^{(n)}}{\|\mathbb{P}^{(n)}\|_1} = \lim_{n \rightarrow \infty} \frac{A^n\mathbb{P}^*}{\|A^n\mathbb{P}^*\|_1}$$

Relation to power method

Power method

$$\lim_{n \rightarrow \infty} \frac{A^n \mathbb{P}^*}{\|A^n \mathbb{P}^*\|_2} = \mathbf{v}$$

Repeated partial summations

$$\mathbb{P}^{(\infty)} = \lim_{n \rightarrow \infty} \frac{\mathbb{P}^{(n)}}{\|\mathbb{P}^{(n)}\|_1} = \lim_{n \rightarrow \infty} \frac{A^n \mathbb{P}^*}{\|A^n \mathbb{P}^*\|_1}$$

↓

$$\mathbb{P}^{(\infty)} = \frac{\mathbf{v}}{\|\mathbf{v}\|_1}$$

How to find limit distribution?

- I. find dominant eigenvalue of matrix A

$$A = \begin{pmatrix} g(0) & g(1) & \dots & g(q-1) \\ 0 & g(1) & \dots & g(q-1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & g(q-1) \end{pmatrix}$$

$$\lambda_D = \max_{j \in \{0,1,\dots,q-1\}} |\lambda_j| = \max_{j \in \{0,1,\dots,q-1\}} |g(j)|, \quad \text{if it exists}$$

- II. find eigenvector v associated with eigenvalue λ_D
- III. normalize v so that $\sum v_i = 1$, i.e.,

$$\frac{v}{\|v\|_1}$$

- 1 Partial-sums discrete probability distributions
- 2 Repeated partial summations
 - Repeated geometric partial summations
 - Computational study in R
- 3 Matrix eigenvalues and eigenvectors, and repeated partial summations
 - Matrix notation of partial-sums distributions
 - Repeated Poisson partial summations
- 4 Parametrized partial summations

Poisson partial summation

$$Q_j = \frac{e^{-a} a^j}{j!} \quad \Longrightarrow \quad g(j) = 1 - \frac{Q_{j+1}}{Q_j} = 1 - \frac{a}{j+1}$$

$$P_x^{(n)} = c_n \sum_{j=x}^{\infty} \left(1 - \frac{a}{j+1}\right) P_j^{(n-1)}, \quad x = 0, 1, 2, \dots$$

$$n = 1, 2, 3, \dots$$

Repeated Poisson partial summations

1. find dominant eigenvalue of matrix A

$$\lambda_D = \max_{j \in \{0, 1, \dots, q-1\}} |g(j)| = \max_{j \in \{0, 1, \dots, q-1\}} \left| 1 - \frac{a}{j+1} \right|, \quad \text{if it is unique}$$

- function $1 - \frac{a}{j+1}$ is monotonous in j for all a
- dominant eigenvalue of matrix A is either $g(0)$ or $g(q-1)$

$$\lambda_D = \begin{cases} g(0) = 1 - a & \text{if } a > \frac{2q}{q+1} \\ g(q-1) = 1 - \frac{a}{q} & \text{if } a < \frac{2q}{q+1} \\ \text{does not exist} & \text{if } a = \frac{2q}{q+1} \end{cases}$$

Repeated Poisson partial summations

$$a > \frac{2q}{q+1}$$

II. find eigenvector v associated with eigenvalue λ_D

$$\lambda_D = g(0) = 1 - a$$

■ solve system of linear equations

$$(A - \lambda_D I)v = 0$$

$$\begin{pmatrix} 0 & 1 - \frac{a}{2} & 1 - \frac{a}{3} & \dots & 1 - \frac{a}{q} \\ 0 & \frac{a}{2} & 1 - \frac{a}{3} & \dots & 1 - \frac{a}{q} \\ 0 & 0 & \frac{2a}{3} & \dots & 1 - \frac{a}{q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{(q-1)a}{q} \end{pmatrix} \begin{pmatrix} v_0 \\ v_1 \\ v_2 \\ \vdots \\ v_{q-1} \end{pmatrix} = 0$$

■ we obtain

$$v = (t, 0, \dots, 0)^T, \quad t \in \mathbb{R}$$

Repeated Poisson partial summations

$$a > \frac{2q}{q+1}$$

III. normalize v

$$\mathbb{P}^{(\infty)} = \frac{v}{\|v\|_1} = \frac{(t, 0, \dots, 0)^\top}{t + 0 + \dots + 0} = (1, 0, \dots, 0)^\top$$

Consider parent distributions defined on support $0, 1, \dots, q - 1$; the limit distribution for repeated Poisson partial summations with parameter $a > \frac{2q}{q+1}$ is the deterministic distribution.

Repeated Poisson partial summations

$$a < \frac{2q}{q+1}$$

II. find eigenvector v associated with eigenvalue λ_D

$$\lambda_D = g(q-1) = 1 - \frac{a}{q}$$

■ solve system of linear equations

$$(A - \lambda_D I)v = 0$$

$$\begin{pmatrix} \frac{1-q}{q}a & 1 - \frac{a}{2} & \dots & 1 - \frac{a}{q-1} & 1 - \frac{a}{q} \\ 0 & \frac{2-q}{2q}a & \dots & 1 - \frac{a}{q-1} & 1 - \frac{a}{q} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{(q-1)-q}{(q-1)q}a & 1 - \frac{a}{q} \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \begin{pmatrix} v_0 \\ v_1 \\ v_2 \\ \vdots \\ v_{q-1} \end{pmatrix} = 0$$

■ we obtain $v = (v_0, v_1, \dots, v_{q-1})^T$

$$v_k = \binom{q-1}{k} \left(\frac{a}{q-a} \right)^k t, \quad k = 0, 1, \dots, q-1, \quad t \in \mathbb{R}$$

Repeated Poisson partial summations

$$a < \frac{2q}{q+1}$$

III. normalize v

$$v_k = \binom{q-1}{k} \left(\frac{a}{q-a}\right)^k t, \quad k = 0, 1, \dots, q-1, \quad t \in \mathbb{R}$$

$$t = \left(\sum_{k=0}^{q-1} \binom{q-1}{k} a^k (q-a)^{-k} \right)^{-1} = \left(\frac{q-a}{q} \right)^{q-1}$$

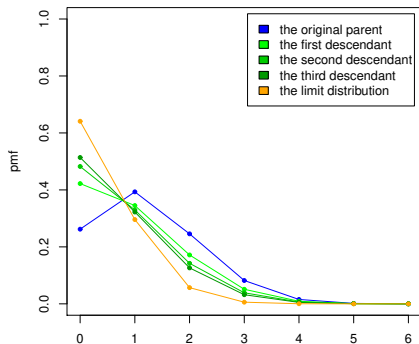
■ limit distribution is

$$\begin{aligned} P_k^{(\infty)} &= \binom{q-1}{k} \left(\frac{a}{q-a}\right)^k \left(\frac{q-a}{q}\right)^{q-1} \\ &= \binom{q-1}{k} \left(\frac{a}{q}\right)^k \left(1 - \frac{a}{q}\right)^{q-1-k}, \quad k = 0, 1, \dots, q-1 \end{aligned}$$

Repeated Poisson partial summations

$$a < \frac{2q}{q+1}$$

Consider parent distributions defined on support $0, 1, \dots, q - 1$; the limit distribution for repeated Poisson partial summations with parameter $a < \frac{2q}{q+1}$ is the binomial distribution with parameters $q - 1, \frac{a}{q}$.



- parent - binomial
 $n = 6, p = 0.2$
- summation
Poisson
 $a = 0.5$
- limit
binomial
 $Bin(6, \frac{0.5}{7})$

The binomial or deterministic distribution is the limit distribution for $g(j)$ given by all distributions from Katz family (binomial, negative binomial with geometric as its special case, Poisson).

Repeated partial summations in general

- I. let the dominant eigenvalue exist,

$$\lambda_D = g(r), \quad r \in \{0, 1, \dots, q-1\}$$

- II. find eigenvector associated with λ_D

$$(A - \lambda_D I)v = 0$$

$$\begin{pmatrix} g(0) - g(r) & g(1) & \dots & g(q-1) \\ 0 & g(1) - g(r) & \dots & g(q-1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & g(q-1) - g(r) \end{pmatrix} \begin{pmatrix} v_0 \\ v_1 \\ \vdots \\ v_{q-1} \end{pmatrix} = 0$$

Repeated partial summations in general

- solution of the system of equations:

$$v_k = \begin{cases} \frac{t}{(g(r))^k} \prod_{j=1}^k (g(r) - g(j-1)), & t \in \mathbb{R} \\ & \text{if } k = 0, 1, \dots, r \\ 0 & \text{if } k = r+1, r+2, \dots, q-1 \end{cases}$$

Partial summations - open problem no. 2 and 3

- how to find the limit distribution if there is no dominant eigenvalue?
- how to find the limit distribution for parents with infinite support?

- 1 Partial-sums discrete probability distributions
- 2 Repeated partial summations
 - Repeated geometric partial summations
 - Computational study in R
- 3 Matrix eigenvalues and eigenvectors, and repeated partial summations
 - Matrix notation of partial-sums distributions
 - Repeated Poisson partial summations
- 4 Parametrized partial summations

Parametrized partial summations

- consider only discrete distributions with one parameter
- recall invariance - for every distribution P_* there is function $g(j)$ such that

$$P_x^* = \sum_{j=x}^{\infty} g(j) P_j^*$$

- emphasizing the parameter, one can write

$$P_x^*(a) = \sum_{j=x}^{\infty} g(j, a) P_j^*(a)$$

Parametrized partial summations

- change now the parameter value while keeping the formula for function $g(j)$, i.e., consider summation

$$P_x = \sum_{j=x}^{\infty} g(j, \lambda) P_j^*(a)$$

Parametrized partial summations - Poisson summation

- Poisson distribution with parameter a is invariant with respect to summation

$$P_x^*(a) = \sum_{j=x}^{\infty} \frac{j - a + 1}{j + 1} P_j^*(a)$$

- consider summation

$$P_x = c \sum_{j=x}^{\infty} \frac{j - \lambda + 1}{j + 1} \frac{e^{-a} a^j}{j!}$$

- we have

$$P_0 = \frac{a - \lambda + \lambda e^{-a}}{a(a + 1 - \lambda)}$$

- descendant distribution has two parameters

Parametrized partial summations - geometric summation

- geometric distribution with parameter a is invariant with respect to summation

$$P_x^*(a) = \sum_{j=x}^{\infty} a P_j^*(a)$$

- consider summation

$$P_x = c \sum_{j=x}^{\infty} \lambda a (1-a)^j$$

- in this case, normalization constant c 'cancels' the new parameter λ , which means that parent and descendant are the same.

- distributions are either sensitive (e.g., Poisson) or resistant (e.g., geometric) with respect to a change of parameter value in function $g(j)$
- many distributions are sensitive
- open problem no. 5 - what is the common property of resistant distributions?

Ďakujem za pozornosť