Description of the distribution laws of discrete random variable when modelling the information systems by orthogonal series

Yu Yu Gromov¹, O N Morozova¹, V V Zdolnik² and V S Starodubtsev²

¹Department of Information Systems and Information Protection, Tambov State Technical University, Sovetskaya 106, Tambov, Russia
²ET Department, Voronezh Institute Federal Penitentiary Service of Russia, 1a Irkutskaya Street, Voronezh, RU

E-mail: gromovtambov@yandex.ru

Abstract. The decomposition of the distribution laws of a discrete random variable on orthogonal polynomials Kravchuk, Meixner, Charlier, Chebyshev and Khan is obtained and practical recommendations are given on the application of the obtained expansions for approximation of distribution laws.

1. Introduction
To approximate the distribution laws of discrete random variable (RV), orthogonal discrete polynomials, defined on a finite or countable point system, are sometimes used [1, 2]. For example, when approximating the distribution laws close to the Poisson law, Charlier polynomials are often used [3]. However, the basic properties of the discrete polynomials of Kravchuk, Meixner, Charlier, Chebyshev and Khan are not fully described in the literature, and their application is not shown for the representation of the distribution laws of discrete RV by orthogonal series.

2. Statement of the problem
Probabilities of possible values accepted by a discrete RV can be described by using the Charlier orthogonal series

\[ p(x) = \frac{\lambda^x}{x!} \exp(-\lambda) \sum_{n=0}^{\infty} C_n S_n(x), \quad x = 0,1,2, ... \quad (1) \]

where \( S_n(x) \) — discrete Charlier polynomial.

Using the orthogonality conditions and the expression \( h_n = \frac{n!}{\lambda^n} \) for constant \( h_n \) find expansion coefficient \( C_n \) in the formula (1):

\[ C_n = \frac{\int S_n(x) \exp(-\lambda) \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} dx}{\int \exp(-\lambda) \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} dx} \]
\[ C_n = \frac{\lambda^n}{n!} \sum_{x=0}^{\infty} S_n(x) p(x). \]  

(2)

From (2) accounting for \( S_0(x) = 1 \), \( S_1(x) = 1 - \frac{x}{\lambda} \), \( S_2(x) = 1 - \left( 2 + \frac{1}{\lambda} \right) x + \frac{x^2}{\lambda^2} \) it follows that

\[ C_0 = 1; \ C_1 = \lambda - m_1; \ C_2 = 0.5 \left[ m_2 - (1 + 2\lambda)m_1 + \lambda^2 \right] \]  

(3)

Let us consider in which case in the expansion (1) we can be restricted just to the first term, which coincides with the Poisson law. We use the coefficients \( K_1 \) and \( K_2 \), which are determined by the expression

\[ K_1 = \frac{\mu_2}{m_1 + \mu_2}, \ K_2 = 1 - \frac{\mu_1}{m_1 + \mu_2} + \frac{1 - 2 \mu_1 / m_1}{2(m_1 - 1 + \mu_2 / m_2)}. \]

For the Poisson law \( K_1 = 0.5 \) and \( K_2 = 1 \). Then in the expansion (2) \( C_0 = 1 \), \( C_1 = C_2 = C_3 = 0 \). Parameter value \( \lambda \) is found from (3), accounting that \( C_1 = 0 \). Herewith we get \( \lambda = m_1 \).

We give an example of the expansion in the Charlier series the distribution law of discrete RV. Often, it is used the sum distribution of two independent discrete RVs, one of which is distributed according to Poisson law and the other according to the negative binomial law [4, 5]

\[ p(x) = (1 - q)^x \exp(-\lambda) \frac{\lambda^x}{x!} 2F_1 (-x, \alpha, -q / \lambda). \]  

(4)

The Charlier expansion for the distribution (4) has the following form:

\[ p(x) = \frac{\lambda^x}{x!} \exp(-\lambda) \sum_{n=0}^{\infty} \frac{(\alpha)_n}{n!} (-q) \frac{\mu_n}{1-q} S_n(x), \ x = 0, 1, 2, \ldots \]  

(5)

Using Charlier polynomials, we can also get expressions for the distribution law of two Poisson dependent random variables

\[ p(x, y) = p(x)p(y) \sum_{n=0}^{\infty} \frac{(r \lambda)^n}{n!} S_n(x)S_n(y) = \]  

\[ = p(x)p(y)(1-r)^{x+y} \exp(r \lambda) 2F_1 \left( -x - y; \frac{r}{(1-r)^2 \lambda} \right). \]  

(6)

and for conditional probability of Poisson random variable

\[ p(x / y) = p(x, y) / p(y) = \frac{\lambda^x}{x!} (1-r)^{x+y} \exp(r \lambda - \lambda) 2F_1 \left( -x - y; \frac{r}{(1-r)^2 \lambda} \right), \]  

(7)

where \( r \) — correlation coefficient, characterizing the degree of linear connection between random variables.

Probabilities of possible values accepted by a discrete RV can also be described with the help of the orthogonal Kravchuk series

\[ p(x) = \frac{\lambda^x}{(1-p)^{N-x} \sum_{n=0}^{x} C_n K_n(x)} x = 0, 1, 2, \ldots, N, \]  

(8)
где $K_n(x)$ — Kravchuk discrete polynomial.

With the help of orthogonality conditions and the expression $h_n = \frac{n!(1 - p)^n(N - n)!}{p^n N!}$ for constant $h_n$ find the expansion coefficient $C_n$ in the formula (8):

$$C_n = \left(\frac{N}{n}\right) \frac{p^n}{(1 - p)^n} \sum_{x=0}^{N} K_n(x)p(x).$$

(9)

From (9) considering that $K_0(x) = 1$; $K_1(x) = 1 - \frac{x}{pN}$; $K_2(x) = 1 - \frac{2x}{pN} + \frac{x(x - 1)}{p^2N(N - 1)}$ it follows that:

$$C_0 = 1; \quad C_1 = \frac{1}{(1 - p)}(pN - m_1),$$

$$C_2 = \frac{0.5}{(1 - p)^2} \left(m_2 - (1 + 2p(N - 1))m_1 + p^2N(N - 1)\right)$$

(10)

Let us consider in which case in the expansion (8) we can restrict ourselves just to the first term, which coincides with the binomial law (2.16). For the binomial law $0 < K_1 < 0.5$ and $K_2 = 1$. Then in the expansion (2.66) $C_0 = 1, C_1 = C_2 = C_3 = 0$. The parameter values $p$ and $N$ must be capable of requiring the conditions: $C_1 = C_2 = 0$. Equating the right hand sides of expressions (10) for the coefficients $C_1$ and $C_2$ to zero, we obtain a system of equations, in the result of its solution we will get:

$$N = \frac{m_1^2}{m_1 - \mu_2}; \quad p = \frac{m_1}{N}.$$ 

(11)

Using the Kravchuk polynomials, one can obtain an expression for the distribution law of two dependent random variables

$$p(x, y) = p(x)p(y)\sum_{n=0}^{N} \binom{N}{n} \left(\frac{r p}{1 - p}\right)^n K_n(x)K_n(y) =$$

$$\frac{p(x)p(y)}{(1 - r)^N} \left(1 - p(1 - r)\right)^{x+y-N} \left[\begin{array}{c} -x, -y; -N; -r \\ \frac{r}{(1 - r)^2 p(1 - p)} \end{array}\right],$$

(12)

each of which is distributed according to the binomial law. Herewith the conditional probability of a random variable

$$p(x/y) = \frac{N! p^x(1-r)^{x+y}(1-p)^{y-N}}{(N-x)! [1 - p(1 - r)]^{x+y-N}} \left[\begin{array}{c} -x, -y; -N; -r \\ \frac{r}{(1 - r)^2 p(1 - p)} \end{array}\right].$$

(13)

Approximation of the distribution laws of discrete RV can also be performed with the help of the orthogonal Meixner series

$$p(x) = \frac{(\alpha)}{x!} q^{x}(1 - q)^{y} \sum_{n=0}^{\infty} C_n M_n(x), \quad x = 0, 1, 2, \ldots,$$

(14)
where $M_n(x)$ — discrete Meixner polynomial.

Considering the orthogonality conditions and the expression $h(x) = \frac{n!}{(\alpha)_n q^n}$ for constant $h_n$ find the coefficients of expansion $C_n$ in the formula (14):

$$C_n = \frac{(\alpha)_n q^n}{n!} \sum_{x=0}^{\infty} M_n(x)p(x).$$

(15)

From (15) considering $M_0(x) = 1$, $M_1(x) = 1 - \frac{x(1-q)}{\alpha q}$, $M_2(x) = 1 - \frac{2x(1-q)}{\alpha q} + \frac{x(x-1)(1-q)^2}{\alpha(\alpha+1)q^2}$ it is followed that:

$$C_0 = 1; \quad C_1 = \alpha q - m_i(1 - q);$$

$$C_2 = 0.5m_2(1 - q)^2 - m_i(1 - q)[q(\alpha + 1) + 0.5(1 - q)] + 0.5\alpha(\alpha + 1)q^2.$$  (16)

Let us consider in which case in the expansion (14) we can restrict ourselves just to the first term, which coincides with the negative binomial law. For a negative binomial law $0.5 < K_1 < 1$ and $K_2 = 1$. Then in the expansion (15) $C_0 = 1$, $C_1 = C_2 = C_3 = 0$. Parameters values $\alpha$ and $q$ must meet the conditions: $C_1 = C_2 = 0$. Equating the right hand sides of expressions (16) for coefficients $C_1$ and $C_2$ to zero, we obtain a system of equations, in the result of its solution we will get:

$$q = 1 - \frac{m_i}{\mu_2}; \quad \alpha = \frac{m_i(1-q)}{q} = \frac{m_i^2}{\mu_2 - m_i}.$$  (17)

We give an example of the expansion in the Meixner series the distribution law of a discrete RV. The Laguerre distribution [4—6] is widely used:

$$p(x) = \exp(-b)(1-q)^{\alpha} q^\alpha L_k^{\alpha-1}\left(-\frac{b(1-q)}{1}\right),$$

(18)

where $a > 0$, $b \geq 0$, $0 < q < 1$ — distribution parameters.

The expansion in the Meixner series for distribution (18) has the following form:

$$p(x) = \frac{(a)_x q^x(1-q)^{\alpha}}{x!} \sum_{n=0}^{\infty} b^n \frac{n!}{n!} M_n(x), \quad x = 0,1,2, \ldots.$$

Using the Meixner polynomials, one can obtain an expression for the distribution law of two dependent random variables

$$p(x,y) = p(x)p(y) \sum_{n=0}^{\infty} \frac{(a)_x (rq)^n}{n!} M_n(x)M_n(y) =$$

$$= p(x)p(y) \frac{(1-r)^{x+y}}{(1-qr)^{a+x+y}} 2F_1 \left(-x-y, a; \frac{r(1-q)^2}{(1-r)^2 q}\right),$$

(19)
each of which is distributed by negative binomial law. Herewith the conditional probability of a random variable

\[ p(x \mid y) = \frac{a_x q^y (1-q)^{a} (1-r)^{x+y}}{x!(1-qr)^{a+x+y}} F_1 \left( -x, y; a; \frac{r(1-q)^2}{(1-r)^2 q} \right). \]  

(20)

Probabilities of possible values accepted by a discrete RV can be described with the help of the orthogonal Chebyshev series

\[ p(x) = \frac{1}{N+1} \sum_{n=0}^{\infty} C_n T_n(x), \quad x=0, 1, 2, \ldots, N, \]  

(21)

where \( T_n(x) \) — discrete Chebyshev polynomial.

Using the orthogonality condition and the expression \( h_n = \frac{(N+n+1)!(N-n)!}{(2n+1)N!(N+1)!} \) for constant \( h_n \), find the coefficients of expansion \( C_n \) in the formula (21):

\[ C_n = \frac{(2n+1)N!(N+1)!}{(N+n+1)!(N-n)!} \sum_{x=0}^{N} T_n(x) p(x). \]  

(22)

From (22) with regard to \( h_n = \frac{(c+1-N)_{x} (N-x+1)_{x} (b-x+1)_{x} (b+c+1-N)_{x} (c+1-N)_{x}}{x!} \), \( x = 0, 1, 2, \ldots, N \), when \( b=1 \) and \( c=1 \) it is followed that:

\[ C_0 = 1; \quad C_1 = \frac{3}{(N+2)} (N-2m_1); \]

\[ C_2 = \frac{5}{(N+2)(N+3)} [6m_2 - 6Nm_1 + N(N-1)]. \]

Using the Chebyshev polynomials one can obtain an expression for the distribution law of two dependent random variables

\[ p(x, y) = \frac{1}{(N+1)^2} \sum_{n=0}^{N} \binom{-N}{n} \binom{-r}{n} \frac{(2n+1)}{(N+2)} T_n(x) T_n(y). \]  

(23)

each of which is distributed by discrete uniform law.

Herewith the conditional probability of random value

\[ p(x \mid y) = \frac{1}{(N+1)} \sum_{n=0}^{N} \binom{-N}{n} \binom{-r}{n} \frac{(2n+1)}{(N+2)} T_n(x) T_n(y). \]  

(24)

Probabilities of possible values accepted by a discrete RV can also be described by an orthogonal series using the Hahn polynomials of the first kind

\[ p(x) = \frac{(c)_{N} (N-x+1)_{x} (b)_{x}}{(b+c)_{N} (N-x+c)_{x} x!} \sum_{n=0}^{N} C_n P_n(x), \quad x = 0, 1, 2, \ldots, N, \]  

(25)

where \( P_n(x) \) — discrete Hahn polynomials of the first kind.

Using the orthogonality condition and expression
In the formula (21):

\[ h_n = \frac{(n + b + c - 1)(N + b + c)_n (c)_n (N - n)!n!}{(2n + b + c - 1)(b)_n (b + c)_n N!} \]

For constant \( h_n \), find the coefficients of expansion \( C_n \) in the formula (21):

\[ C_n = \left( \begin{array}{c} N \\ n \end{array} \right) \frac{(2n + b + c - 1)(b)_n (b + c)_n}{(n + b + c - 1)(N + b + c)_n (c)_n (n + b + c - 1)} \sum_{r=0}^{N} R_r(x) p(x). \]  

(26)

Using Hahn polynomials of the first kind, one can obtain an expression for the distribution law of two dependent random variables

\[ p(x, y) = p(x) p(y) \sum_{n=0}^{N} \left( \begin{array}{c} N \\ n \end{array} \right) \frac{(b)_n (b + c)_n (2n + b + c - 1)r^n}{(N + b + c)_n (c)_n (n + b + c - 1)} P_n(x) P_n(y), \]

(27)

Each of which has a distribution

\[ p(x) = \frac{(c)_n (N + 1 - x)_n (b)_n}{(b + c)_n (N - x + c)_n x!}, \quad x = 0, 1, 2, \ldots, N. \]

Herewith the conditional probability of a random variable

\[ p(x / y) = p(x) \sum_{n=0}^{N} \left( \begin{array}{c} N \\ n \end{array} \right) \frac{(b)_n (b + c)_n (2n + b + c - 1)r^n}{(N + b + c)_n (c)_n (n + b + c - 1)} P_n(x) P_n(y), \]

(28)

where

\[ p(x) = \frac{(c)_n (N - x + 1)_n (b)_n}{(b + c)_n (N - x + c)_n x!}, \quad x = 0, 1, 2, \ldots, N. \]

(29)

Probabilities of possible values accepted by a discrete RV can also be described by an orthogonal series with the use of Hahn polynomials of the second kind

\[ p(x) = \frac{(1 + c - N)_n (N - x + 1)_n (b - x + 1)_n}{(b + c + 1 - N)_n (1 + c - N)_n x!} \sum_{r=0}^{N} C_n R_r(x), \]

(30)

where \( x = 0, 1, 2, \ldots, N \); \( R_r(x) \) – discrete Hahn polynomials of the second kind.

Using the orthogonality condition and the expression

\[ h_n = \frac{(1 + b + c)(c + 1 - n)_n (b + c + 1 - N - n)_n (N - n)!n!}{(1 + b + c - 2n)(2 + b + c - n)_n (b + 1 - n)_n N!} \]

For constant \( h_n \), find the expansion coefficients \( C_n \) in the formula (30)

\[ C_n = \left( \begin{array}{c} N \\ n \end{array} \right) \frac{(1 + b + c - 2n)(c + 1 - n)_n (b + c + 1 - N - n)_n}{(1 + b + c - n)_n (1 + b + c)_n (1 + b)_n} \sum_{r=0}^{N} R_r(x) p(x). \]  

(31)

Using the Hahn polynomial of the second kind, one can obtain an expression for the distribution law of two dependent random variables

\[ p(x, y) = p(x) p(y) \sum_{n=0}^{N} \left( \begin{array}{c} N \\ n \end{array} \right) \frac{(c + 1)_n (b + c + 1 - N)_n (1 + b + c - 2n)r^n}{(1 + b + c)_n (1 + b)_n (1 + b + c - n)} R_r(x) R_r(y), \]

(32)

each of which is distributed by the law
$$p(x) = \frac{(1+c-N)^N (N+1-x)^x (b+1-x)^x}{(1+b+c-N)^N (1+c-N)^x x!}, \quad x = 0,1,2...N.$$