Models of multidimensional discrete distribution of probabilities of random variables in information systems

To cite this article: Yu Yu Gromov et al 2018 J. Phys.: Conf. Ser. 973 012034

View the article online for updates and enhancements.

Related content
- Moments of two noncommutative random variables in terms of their joint quantum operators
  Gabriela Popa and Aurel I Stan
- Gamma distributed random variables and their semi-quantum operators
  Gabriela Popa and Aurel I Stan
- Letter
  Saralees Nadarajah
Models of multidimensional discrete distribution of probabilities of random variables in information systems

Yu Yu Gromov, Yu V Minin, O G Ivanova and O N Morozova

Department of Information Systems and Information Protection, Tambov State Technical University, Sovetskaya 106, Tambov, Russia

E-mail: gromovtambov@yandex.ru

Abstract. Multidimensional discrete distributions of probabilities of independent random values were received. Their one-dimensional distribution is widely used in probability theory. Producing functions of those multidimensional distributions were also received.

1. Introduction
Any model of information system is an abstract, formally described object. It describes the formalized process of information system functioning and it is capable to cover only its basic, specific characteristics, while leaving aside minor unnecessary factors.

The formalization of information process is preceded by the structure study, in this way it is developed a meaningful description of the process, i.e. the first attempt to get across the specific characteristics of investigated information process and to set a task. This gives information about the physical nature and quantitative characteristics of the process elementary phenomenon, the nature of their interaction, the place of each phenomenon in general process.

A meaningful description serves to construct a formalized scheme and a model of information process.

A formalized scheme of the process is elaborated in case when, due to the complexity of the process or the difficulties of formalizing some of its elements, a direct transition from a meaningful description to a model is impossible or inexpedient. To construct a formalized scheme, it is necessary to select the characteristics of the process; establish a system of parameters providing the process; to determine all dependencies between characteristics and parameters in view of factors considering when formalizing.

As a rule, the most difficult stage of simulation process is the conversion of the identified essential factors into the language of mathematical concepts and determination of interactions between these quantities. The fact is that the requirements of content and deductivity of the model are contradictory in their essence. To satisfy the requirement of content, it is necessary to consider in the model as many factors of information process as possible. Herewith, naturally, the model becomes more complex, which makes it difficult to study and obtain meaningful results. However, the desire to obtain a result in easier way leads to the need to simplify the model, thereby reducing its content. Thus, it is required to achieve a reasonable compromise, ensuring the possibility of obtaining non-trivial results and not missing the essence of the real process. In this case, an accurate set of all initial data, known parameters and initial conditions is attached.
A meaningful description may not provide the necessary information for constructing a formalized scheme, and then additional experiments and observations of the process under investigation are necessary. But in this case, when developing a formalized scheme, they should be fully used.

Further transformation of the formalized scheme into a model is fulfilled without inputting the additional information.

Depending on the complexity of the analytical description of the system, the following main methods of using mathematical models are distinguished: analytical research; qualitative research; research using numerical methods; simulation modelling.

In most cases, when modeling the information systems, there is no complete mathematical formulation of the problem, and even though there are analytical methods and approaches, but they are very complex and laborious, and besides evaluating certain parameters it is necessary to implement monitoring of the process during a certain period, and in some cases, it requires slowing down or accelerating the model time. In this context, simulation of information systems is simpler way of research than analytical.

For qualitative evaluation of a complex system, it is convenient to use the results of the theory of random processes. The experience of observing objects shows that they function under the conditions of many random factors, so the prediction of the behavior of complex system can be meaningful within the framework of probability categories. In other words, only the probabilities of their occurrence can be indicated for expected events, and in respect to some values it is necessary to confine to the laws of their distribution or other probability characteristics (for example, mean values, variances, etc.).

To study the process of functioning of each complex system, considering random factors, it is necessary to have a clear idea of the sources of random effects and very reliable data on their quantitative characteristics. Therefore, any calculation or theoretical analysis associated with the study of a complex system is preceded by an experimental accumulation of statistical material characterizing the behavior of individual elements and the whole system in real conditions. This material processing allows to obtain the initial data for calculation and analysis.

The main sources of random influences are the factors of the "external" environment and deviations from the normal operation modes (errors, noises, etc.) occurring within the system.

It follows from what has been said that a very serious attention must be given to the consideration of random factors while investigating the complex systems using random numbers with specified probabilistic characteristics. In this case, the results obtained with one-time modeling should be regarded only as an implementation of a random process. Each of these realizations separately cannot serve as an objective characteristic of system under study. The required values are usually determined by averaging and statistical processing of the data of many realizations, so this approach is also called the statistical modeling method.

2. Statement of the problem

Nowadays multidimensional discrete distributions are presented more often in tables, or matrixes [1,2], i.e. analytic expressions for them are used little in the theory of probability. Among them, we should mention first of all the polynomial probability distribution of dependent random variables (RV) $X_1, \ldots, X_k$, which take integer non-negative values $n_1, \ldots, n_k$ with probabilities [1-3]

$$P\{X_1 = n_1, \ldots, X_k = n_k\} = \frac{n!}{n_1! \ldots n_k!} p_1^{n_1} \ldots p_k^{n_k}, \quad (1)$$

where $0 < p_j < 1; \quad p_1 + \ldots + p_k = 1; \quad n_1 + \ldots + n_k = n; \quad k \geq 2.$

Producing function of distribution (1) is defined by an expression

$$G(z_1, \ldots, z_k) = (p_1 z_1 + \ldots + p_k z_k)^n, \quad (2)$$
where $z_j$ is complex number $|z_j| \leq 1$.

In paper [3] it is noted, that distribution (1) is $(k-1)$-dimensional, i.e. in dimension $R^k$ it is degenerated. That’s why for $k$-dimensional polynomial probability distribution of dependent RV $X_1, \ldots, X_k$ instead of expression (1) is offered to use ratio

$$P\{X_i = n_1, \ldots, X_k = n_k\} = \frac{n!}{n_1! \cdots n_k!} p_0^{n_0} p_1^{n_1} \cdots p_k^{n_k}. \quad (3)$$

Distribution (3) can be considered as a modified polynomial distribution. Its producing function is defined by an expression.

$$G(z_1, \ldots, z_k) = (p_0 + p_1 z_1 + \cdots + p_k z_k)^n \quad (4)$$

Negative polynomial probability distribution of dependent RV $X_1, \ldots, X_k$, taking integer non-negative values $m_1, \ldots, m_k$ with probabilities [3] is used sometimes

$$P\{X_i = m_1, \ldots, X_k = m_k\} = \frac{\Gamma(a+m_1 + \cdots + m_k)}{\Gamma(a) m_1! \cdots m_k!} p_0^{m_0} p_1^{m_1} \cdots p_k^{m_k}, \quad (5)$$

where $\Gamma(z)$ is gamma-function; $\alpha > 0; \ 0 < p_j < 1; \ p_0 + \cdots + p_k = 1; \ k \geq 1$.

Producing function for distribution (5) is defined by an expression

$$G(z_1, \ldots, z_k) = p_0^n (1 - p_1 z_1 - \cdots - p_k z_k)^{\alpha}. \quad (6)$$

The main purpose of the work is to receive models of multidimensional discrete probability distributions of dependent RV, similar to probability models (3) and (5)

3. The solution of the problem
Table 1 shows the expressions of one-dimensional discrete distributions, widely used in probability theory [1-4]. Table 2 shows expressions for producing functions of these distributions.

| № | Type of distribution | Expression for distribution $p(x)$ |
|---|----------------------|------------------------------------|
| 1 | Binomial             | $p(x) = \frac{n!}{(n-x)!x!} (1-p)^{n-x} p^x$ $x = 0, 1, \ldots, n; \ 0 < p < 1, \ n \geq 1$ |
| 2 | Negative binomial    | $p(x) = \frac{\Gamma(a+x)}{\Gamma(a) x!} (1-q)^x q^a, \quad x = 0, 1, \ldots, \infty; \ 0 < q < 1, \ \alpha > 0$ |
| 3 | Poisson’s            | $p(x) = \frac{\lambda^x}{x!} \exp(-\lambda), \ x = 0, 1, \ldots, \infty$ |
| 4 | Hypergeometric       | $p(x) = \frac{n! N_0! N_1! (N-n)!}{(n-x)! x! (N_0-n+x)! (N_1-x)! N!}$ $x = 0, 1, \ldots, n; \ N_0 \geq n, \ N_1 \geq n, \ n \geq 2; \ N_0 + N_1 = N$ |
5 Negative hypergeometric

\[ p(x) = \frac{n! (b_1)_{x} (b_0)_{n-x}}{(n-x)!x! (b)_n}, \]

\[ x = 0, 1, \ldots, n; b_0 > 0, b_1 > 0, n \geq 2; b_0 + b_1 = b \]

6 Negative beta-binomial

\[ p(x) = \frac{\Gamma(a + b_0 + 1)\Gamma(b + 1)(a)_{x} (b_1)_{x}}{\Gamma(b_0 + 1)\Gamma(a + b + 1)(a + b + 1)_{x} x!}, \]

\[ x = 0, 1, \ldots, \infty; a > 0, b_0 > 0; b_1 \geq a; b_0 + b_1 = b \]

| № | Type of distribution | Expression for producing function G(z) |
|---|---------------------|----------------------------------------|
| 1 | Binomial            | \[ G(z) = (q + p z)^a \] \[ 0 < p < 1, \ q = 1 - p; \ n \geq 1 \] |
| 2 | Negative binomial   | \[ G(z) = q^a (1 - p z)^a \] \[ 0 < p < 1; q = 1 - p; \ \alpha > 0 \] |
| 3 | Poisson’s           | \[ G(z) = \exp(\lambda z - \lambda); \ \lambda > 0 \] |
| 4 | Hypergeometric      | \[ G(z) = \frac{N_0! (N - n)!}{(N_0 - n)! N!} x F_1(-n, -N_1; 1 + N_0 - n; z) \] |
|   |                     | \[ N_0 \geq n, N_1 \geq n, n \geq 2; N_0 + N_1 = N \] |
| 5 | Negative hypergeometric | \[ G(z) = \left( \frac{b_0}{b_0 + b_1} \right)^x x F_1(-n, b_1; 1 + b_0 - n; z) \] |
|   |                     | \[ b_0 > 0, b_1 > 0, n \geq 2; b_0 + b_1 = b \] |
| 6 | Negative beta-binomial | \[ G(z) = \frac{\Gamma(a + b_0 + 1)\Gamma(b + 1)}{\Gamma(b_0 + 1)\Gamma(a + b + 1)} x F_1(a, b_1; a + b + 1; z) \] |
|   |                     | \[ a > 0, b_0 > 0; b_1 \geq a; b_0 + b_1 = b \] |

If we put \( k = 1, n_1 = x, n_0 = n - x, p_1 = p, p_0 = q = 1 - p \) in ratios (3) and (4), then they will transform respectively into one-dimensional binomial distribution and its producing function. Their expressions are shown in tables 1 and 2 (1 line). That’s why, in the role of \( k \)-dimensional binomial probability distribution of dependent RV-s \( X_1, \ldots, X_k \) and its producing function we can use respectively \( k \) -dimensional modified polynomial distribution (3) and producing function (4).

Analogically we can make sure, that if in ratios (5) and (6) to accept \( k = 1, m_1 = x, p_1 = p, p_0 = q = 1 - p \), then they will transform respectively in one-dimensional negative binomial distribution and its producing function. Their expressions are shown in tables 1 and 2 (line 2).

That’s why, in the role of \( k \) -dimensional binomial probability distribution of dependent RV-s \( X_1, \ldots, X_k \) and its producing function we can use respectively \( k \)-dimensional modified polynomial distribution (5) and producing function (6).

It’s known from probability theory, that the limiting case of one-dimensional binomial distribution, when \( p \to 0 \) and \( n \to \infty \), and product \( n p \) aims to some positive constant value \( \lambda \) (\( n p \to \lambda \)), is one-dimensional Poisson distribution [1-3]. Expressions to Poisson distribution and its
producing function are shown in the tables 1 and 2 (line 3). Respectively, the limiting case for k-dimensional binomial distribution (3) is k-dimensional Poisson’s distribution [5]

\[ P\{X_i = n_i, \ldots, X_k = n_k\} = \exp\left(-\sum_{i=1}^{k} \lambda_i\right) \prod_{i=1}^{k} \frac{\lambda_i^{n_i}}{n_i!}. \]  

(7)

It follows from analysis of table 2 (line 4-6), that the basis of producing functions hypergeometric, negative hypergeometric and negative beta-binomial distributions is a hypergeometric Gaussian function \( \mathbb{F}_0(a, b; c; z) \) [6]. That’s why for these distributions expression of producing function can be presented in general view

\[ G(z) = K_u \mathbb{F}_0(a, b; c; z) = K_u \sum_{x=0}^{\infty} \frac{(a)_x (b)_x}{(c)_x} x^n z^x = \sum_{x=0}^{\infty} p(x) z^x. \]  

(8)

where \( (a)_x = a(a+1)\ldots(a+k-1) \) – symbol of Poghammer; \( K_u = \frac{\Gamma(c-a)\Gamma(c-b)}{\Gamma(c)\Gamma(c-a-b)} \) normalisation coefficient.

From (8) it is easy to receive in general view the expression for one-dimensional probability distribution of shown distribution laws

\[ p(x) = \frac{\Gamma(c-a)\Gamma(c-b)}{\Gamma(c)\Gamma(c-a-b)} \frac{(a)_x (b)_x}{(c)_x} x^n, \]  

(9)

It’s one of comfortabilities of using producing functions.

In the analogy with one-dimensional case (8) producing function of multidimensional function must contain hypergeometric function of some dependent values. Such function can be Laurichell function \( \mathbb{F}_0(a_1, b_1 \ldots b_k; c; z_1, \ldots, z_k) \), which in the case of 1 variable \( k=1 \) transforms in hypergeometric Gaussian function [6]. In this producing function either

\[ G(z_1, \ldots, z_k) = K_u \mathbb{F}_0^{(a_1, b_1 \ldots b_k; c; z_1, \ldots, z_k)} \]  

(10)

or

\[ G(z_1, \ldots, z_k) = K_u \sum_{x_1, \ldots, x_k = 0}^{\infty} \frac{(a)_{x_1} \ldots (a)_{x_k}}{(c)_{x_1} \ldots (c)_{x_k}} \frac{(b_1)_{x_1} \ldots (b_k)_{x_k}}{(c)_{x_1} \ldots (c)_{x_k}} z_1^{x_1} \ldots z_k^{x_k} n_1! \ldots n_k!, \]  

(11)

where \( b = b_0 + \ldots + b_k \); \( K_u = 1/F_0^{(a, b_1 \ldots b_k; c; 1)} \).

From (11) follows in general case an expression for k-dimensional probability distribution of dependent RV-s \( X_1, \ldots, X_k \)

\[ p(n_1, \ldots, n_k) = K_u \frac{(a)_{n_1} \ldots (a)_{n_k}}{(c)_{n_1} \ldots (c)_{n_k}} (b_1)_{n_1} \ldots (b_k)_{n_k} n_1! \ldots n_k!. \]  

(12)

If \( a = -n, b_i = -N_i, \ldots, b_k = -N_k \), \( b = -N, c = 1+N_0-n \), then multidimensional distribution (12) transforms in k-dimensional hypergeometric distribution, defined by an expression

\[ p(n_1, \ldots, n_k) = \binom{N_0}{n_0} \binom{N_1}{n_1} \ldots \binom{N_k}{n_k} \binom{N}{n}, \]  

(13)
where \( n_0 = n - n_1 - \ldots - n_k \); 
\[
\binom{N}{n} = \frac{N!}{(N-n)!n!} - \text{binomial coefficients.}
\]

If \( a = n, \ b = b_0 + \ldots + b_k, \ c = 1 - b_0 - n \), then multidimensional distribution (12) transforms in k-dimensional negative hypergeometric distribution, defined by an expression

\[
p(n_1, \ldots, n_k) = \frac{n!}{(b)_n} \frac{(b_0)_{n_1} (b_1)_{n_1} \ldots (b_k)_{n_k}}{n_0! n_1! \ldots n_k!}.
\]

If \( b = b_0 + \ldots + b_k, \ c = 1 + a + b \), then multidimensional distribution (12) transforms in \( k \)-dimensional negative beta-binomial distribution, defined by an expression

\[
p(n_1, \ldots, n_k) = \Gamma(b + 1) \Gamma(a + b_0 + 1) \frac{(a)_n \ldots (b_k)_n}{\Gamma(a + b + 1) \Gamma(b_0 + 1)} \frac{(b_0)_{n_1} \ldots (b_k)_{n_k}}{n_0! n_1! \ldots n_k!}.
\]

In practice, among the multidimensional discrete distributions, two-dimensional distributions can most often be used. Table 3 shows the expressions for two-dimensional discrete distributions of dependent random variables one-dimensional distributions of which are presented in Table 1.

Table 4 shows expressions for producing functions of these distributions.

**Table 3. Two-dimensional discrete distributions.**

| \( n \) | Type of distribution | Expression for distribution \( p(x,y) \) |
|---|---|---|
| 1 | Binomial | \[
p(x, y) = \frac{n!}{(n-x-y)!x!y!} (1 - p_1 - p_2)^{n-x-y} p_1^x p_2^y
\]
| 2 | Negative binomial | \[
p(x, y) = \frac{\Gamma(\alpha + x + y)}{\Gamma(\alpha) x! y!} p_1^x p_2^y (1 - p_1 - p_2)^\alpha,
\]
| 3 | Poisson’s | \[
p(x, y) = \frac{\lambda_1^x \lambda_2^y}{x!y!} \exp(-\lambda_1 - \lambda_2), \ 0 \leq x < \infty; \ 0 \leq y < \infty.
\]
| 4 | Hypergeometric | \[
p(x, y) = \frac{n!}{(n-x-y)!x!y! (N_0 - n + x + y)! (N_1 - x)! (N_2 - y)! N!}
\]
| 5 | Negative hypergeometric | \[
p(x, y) = \frac{n!}{(n-x-y)!x!y! (n + b_0 - x - y)! (b_1)_x (b_2)_y (b_0)_n}
\]
| 6 | Negative beta-binomial | \[
p(x, y) = \frac{\Gamma(a + b_0 + 1) \Gamma(b + 1) (b_1)_x (b_2)_y}{\Gamma(b_0 + 1) \Gamma(a + b + 1) (a + b_1 + b_2 + b)_y x!y!}.
\]

| \( a > 0, \ b_0 > 0; \ b_1 \geq a; \ b_2 \geq a; \ b_0 + b_1 + b_2 = b \) |
Table 4. Producing functions of two-dimensional discrete distributions.

| № | Type of distribution | Expression for distribution $G(z_1,z_2)$ |
|---|----------------------|-----------------------------------------|
| 1 | Binomial             | $G(z_1,z_2) = (q + p_1 z_1 + p_2 z_2)^n$ | $0 < p_i < 1, \ q = 1 - p_1 - p_2 ; \ n \geq 1$ |
| 2 | Negative binomial    | $G(z_1,z_2) = q^a (1 - p_1 z_1 - p_2 z_2)^{n_a}$ | $0 < p_i < 1; \ q = 1 - p_1 - p_2; \ a > 0$ |
| 3 | Poisson’s            | $G(z_1,z_2) = \exp(\lambda_1 z_1 + \lambda_2 z_2 - \lambda_1 - \lambda_2); \ \lambda_1 > 0; \ \lambda_2 > 0$ |
| 4 | Hypergeometric       | $G(z_1,z_2) = \binom{N_1}{n} \binom{N_2}{n} F^{(2)}_D (-n, -N_1; -N_2; 1 + N_0 - n; z_1, z_2)$ | $N_0 \geq n, \ N_1 \geq n, \ N_2 \geq n, \ n \geq 2; \ N_0 + N_1 + N_2 = N$ |
| 5 | Negative hypergeometric | $G(z_1,z_2) = \binom{b_0}{n} F^{(2)}_D (-n, b_1, b_2; 1 - b_0 - n; z_1, z_2)$ | $b_0 > 0, \ b_1 > 0, \ b_2 > 0, \ n \geq 2; \ b_0 + b_1 + b_2 = b$ |
| 6 | Negative beta-binomial | $G(z_1,z_2) = \frac{\Gamma(a+b_0+1)\Gamma(b+1)}{\Gamma(b_0+1)\Gamma(a+b+1)} F^{(2)}_D (a, b_1, b_2; a + b + 1; z_1, z_2)$ | $a > 0, \ b_0 > 0; \ b_1 \geq a; \ b_2 \geq a; \ b_0 + b_1 + b_2 = b$ |

Thus, the proposed approach makes it possible to obtain expressions for multidimensional discrete distributions of probabilities of independent random values, as well as their producing functions.

References

[1] Gnedenko B V 1988 Course of the theory of probability (Moscow: Science) p 448
[2] Fedotkin M A 2012 Models in the theory of probability (Moscow: FIZMATLIT) p 608
[3] Probability and mathematical statistics: encyclopedia ed Yu V Prokhorov 1999 (Moscow: The Great Russian Encyclopedia) p 910
[4] Johnson N L, Kots S and Kemp A 2010 One-dimensional discrete distributions (Moscow: BINOM) p 559
[5] Karpov I G 1998 The laws of distribution of discrete random variables and their identification IV International Scientific and Technical Conference "Radiolocation, Navigation and Communication" I pp 115-125
[6] Prudnikov A P, Brychkov Yu A and Marichev O I 1986 Integrals and series. Additional chapters (Moscow: Nauka) p 800