MODELING UNBIASED ESTIMATORS WITH GOOD ASYMPTOTIC PROPERTIES FOR THE SUM OF MULTIVARIATE DISCRETE INDEPENDENT RANDOM VARIABLES

A. S. Iskakova
Gumilyov Eurasian National University
Astana, KAZAKHSTAN
e-mail: ayman.astana@gmail.com

Abstract

In a multivariate discrete probability model of distribution of sum discrete random variables is proposed and studied. The concept of the most appropriate of the set of unbiased estimators, which has good asymptotic properties, is introduced.

1 Introduction

Multivariate probabilistic models, as a reflection of the current reality, are absolutely necessary for describe events and situations encountered in daily life. In recent years, a considerable amount of probabilistic models have been developed. However, there are many unsolved problems, for example, in the implementation of monitoring the it is clear only the sum of components, which as a result of observations can not be detected. So far, probabilistic models describing similar situations were not considered.

An exceptional example of the actual use of such a model is the advertising industry, where it is necessary to link the distribution of consumer interests with appropriate advertising in various sources. Similar problems are very common in meteorology and other fields. In this paper we present statistical evaluation of the distribution of sums of unobservable random matrices $\mathbf{L}_1, \ldots, \mathbf{L}_d$ by their amount. Thus, the results of the proposed work can solve many of these problems.

2 Multivariate discrete probability distribution of sum of discrete random variables

Assume that the true image can be represented as a matrix $\mathbf{l}_0 = \|l_{0_{i,j}}\|_{m\times q}$, which imposed distortion, consisting of four factors (matrices) of losses $\mathbf{u} = \|u_{i,j}\|_{m\times q}$, taking values from the set of $\mathbf{l}_1, \ldots, \mathbf{l}_d$.

Obviously, the factors (the matrix), the loss $\mathbf{l}_1, \ldots, \mathbf{l}_d$ are realizations of random matrices $\mathbf{L}_1, \ldots, \mathbf{L}_d$, appearing with probabilities $\mathbf{p} = (p_1, \ldots, p_d)$.

Assume that $V_{\mathbf{u}}$ is the number of possible combinations $r_{1_{v_{\mathbf{u}}}}\mathbf{L}_1, \ldots, r_{d_{v_{\mathbf{u}}}}\mathbf{L}_d$, which together form a matrix of $\mathbf{u}$, where $r_{1_{v_{\mathbf{u}}}}, \ldots, r_{d_{v_{\mathbf{u}}}}$ determine the possible number of balls taken out, which marked the relevant matrices $\mathbf{L}_1, \ldots, \mathbf{L}_d$. In other words, from [2] it follows that $V_{\mathbf{u}}$ is the number of partitions on the part of the matrix $\mathbf{u}$ on $\mathbf{L}_1, \ldots, \mathbf{L}_d$.

Theorem 1. *The distortion is distributed as follows:*

$$P(\mathbf{U}=\mathbf{u})=\sum_{v_{\mathbf{u}}=1}^{V_{\mathbf{u}}} n! \prod_{\alpha=1}^{d} \frac{p_{\alpha}^{r_{\alpha v_{\mathbf{u}}}}}{r_{\alpha_{v_{\mathbf{u}}}}!}. \tag{1}$$

3 Unbiased estimation of the probability distribution of the proposed model

In practice, as a rule, elements of the vector $\mathbf{p}=(p_1,\ldots,p_d)$ are not known. It is also not known matrix $\mathbf{L}_1,\ldots,\mathbf{L}_d$. Consequently, formula (1) does not find the actual application.

Assume that there are photos in the number of k particular locality with the distortions $\mathbf{x}=\{\mathbf{x}_1,...,\mathbf{x}_k\}$. In other words, a number of evidence-x can be interpreted as a realization of a sample of k, whose elements are subject to distribution (1). We denote $\mathbf{r}_{v_\beta}$ vector $(r_{1_{v_\beta}},\ldots,r_{d_{v_\beta}})$, which defines v_β-th solution of equation

$$\begin{cases} \sum\limits_{\alpha=1}^{d} L_\alpha r_{\alpha_{v_\beta}}=\mathbf{u}, \\ \sum\limits_{\alpha=1}^{d} r_{\alpha_{v_\beta}}=n, \end{cases} \tag{2}$$

where $v_\beta=1,\ldots,V_\beta, V_\beta$ is the number of partitions of the matrix $\mathbf{x}_\beta$ on the matrices $\mathbf{L}_1,\ldots,\mathbf{L}_d$. Using the system of equations (2), the matrices $\mathbf{L}_1,\ldots,\mathbf{L}_d$, and the actual data $\mathbf{x}$, we define for each $\beta=1,\ldots,k$ the number of partitions V_β matrix $\mathbf{x}_\beta$ at $\mathbf{L}_1,\ldots,\mathbf{L}_d$, and vectors $\mathbf{r}_{1_\beta},\ldots,\mathbf{r}_{V_\beta}$.

Suppose that for each $j=1,\ldots,\mu$, where $\mu=\prod\limits_{\beta=1}^{k} V_\beta$, there is a vector $\mathbf{z}_j=(z_{1_j},\ldots,z_{d_j})$, defined as $\mathbf{z}_j=\sum\limits_{\beta=1}^{k}\mathbf{r}_{v_\beta}$, and the indices on the right and left side are linked one-to-one correspondence, which is not unique.

Thus, from the above lemma that if some element of the implementation of the sample $\mathbf{x}=(\mathbf{x}_1,\ldots,\mathbf{x}_k)$ of the distribution (1) has more than one partition on the submitted part, it is impossible, using the theorem Rao-Blackwell-Kolmogorov construct an unbiased estimate with minimum variance for the probability distribution (1).

Theorem 2. *The following statistics form an unbiased estimate for the probability distribution (1):*

$$W(\mathbf{u},\mathbf{z}_j)=\frac{\sum\limits_{v_{\mathbf{u}}=1}^{V_{\mathbf{u}}}\prod\limits_{\alpha=1}^{d}\binom{z_{\alpha_j}}{r_{\alpha v_{\mathbf{u}}}}}{\binom{nk}{n}},\ j=1,\ldots,\mu, \tag{3}$$

where $V_{\mathbf{u}}$ is the number of partitions on the part of the matrix $\mathbf{u}$, $\mathbf{L}_1,\ldots,\mathbf{L}_d$; for each partition $r_{1_{v_{\mathbf{u}}}},\ldots,r_{d_{v_{\mathbf{u}}}}$ determine the possible number of matrices $\mathbf{L}_1,\ldots,\mathbf{L}_d$; $k\geq 1$ and $z_{\alpha_j}\geq r_{\alpha_{v_{\mathbf{u}}}}$, when $\alpha=1,\ldots,d$, $v_{\mathbf{u}}=1,\ldots,V_{\mathbf{u}}$.

4 The most suitable unbiased estimates for the probability distribution of the proposed model and their properties

Thus, we have a lot of unbiased estimates of the probability of distortion.

Definition 1. Decision $\mathbf{z}_g$, based on observation, is the most appropriate set of $\mathbf{z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_m\}$, if

$$\prod_{\beta=1}^{k} W(\mathbf{x}_\beta, \mathbf{z}_g) = \max_{j=1,\ldots,\mu} \prod_{\beta=1}^{k} W(\mathbf{x}_\beta, \mathbf{z}_j), \tag{4}$$

where for $\beta = 1, \ldots, k$ elements of $W(\mathbf{x}_\beta, \mathbf{z}) = \{W(\mathbf{x}_\beta, \mathbf{z}_1), \ldots, W(\mathbf{x}_\beta, \mathbf{z}_\mu)\}$ forms an unbiased estimate for the probability distribution (1) defined in (5).

Definition 2. Unbiased estimate of $W(\mathbf{x}_\beta, \mathbf{z}_g)$ for the probability distribution (1) is the most suitable from the entire set of unbiased estimates of $W(\mathbf{x}_\beta, \mathbf{z}) = \{W(\mathbf{x}_\beta, \mathbf{z}_1), \ldots, W(\mathbf{x}_\beta, \mathbf{z}_\mu)\}$ defined in (5), if $\mathbf{z}_g$ is the most appropriate solution, based on observation.

Theorem 3. *The most suitable unbiased estimate of $W(\mathbf{x}_\beta, \mathbf{z}_g)$ for the probability distribution (1) is consistent, asymptotically normal and asymptotically efficient.*

Let us summarize the results:

- proposed and studied a new probability distribution of discrete random variables;
- developed an algorithm for computing the probability and define the generating function for the distribution of the proposed model;
- the set of unbiased estimates for the probability distribution of the proposed model and the variance of these estimates;
- introduced a new concept of the most appropriate evaluation of the set of unbiased estimates, with asymptotic properties.

References

[1] Iskakova A.S. (2014). Construction of the most suitable unbiased estimate distortions of radiation processes from remote sensing data. *Journal of Physics: Conference Series*. Vol. **490**(1), id 012113.

[2] Andrews G. E. (1976). *The Theory of Partitions. Encyclopedia of Mathematics and Its Applications*. Addison-Wesley, London.