

ON RANDOM SUBSETS OF THE VERTICES OF THE $\backslash(n\backslash)$ -DIMENSIONAL CUBE

M. V. Kozlov

1970

SovietRxiv

View the original and related papers at <https://sovietsrxiv.org/items/ru-197001.21418>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

UDC 519.214.9

MATHEMATICS

M. V. KOZLOV

ON RANDOM SUBSETS OF THE VERTICES OF THE n -DIMENSIONAL CUBE

(Presented by Academician A. N. Kolmogorov, 26 V 1969)

In the set of all vertices of the n -dimensional unit cube, select a random subset $\{\mathbf{a}_i, i = 1, \dots, m\}$, putting $\mathbf{a}_i = (a_i^1, \dots, a_i^n)$, where $a_i^j, i = 1, \dots, m; j = 1, \dots, n$, are mutually independent random variables taking the values 1 and 0 with equal probabilities. Define the distance r_{ij} between the vertices $\mathbf{a}_i$ and $\mathbf{a}_j$ as the number of nonzero coordinates of the vector $\mathbf{a}_i \oplus \mathbf{a}_j$, where $\oplus$ denotes coordinatewise summation modulo 2. In the present note we study some probabilistic characteristics of the family of random variables r_{ij} .

Let $-1 \leq t_0 < t_1 < \dots < t_s$ be arbitrary integers. Introduce random variables $\xi_\alpha, \alpha = 1, \dots, s$, equal to the number of such $r_{ij}, i < j$, that $t_{\alpha-1} < r_{ij} \leq t_\alpha$. Denote by $x_{ij}(\alpha)$ the characteristic function of the event $t_{\alpha-1} < r_{ij} \leq t_\alpha$; we have

$$\xi_\alpha = \sum_{i < j} x_{ij}(\alpha).$$

The mathematical expectation $E\xi_\alpha$ is equal to $\frac{1}{2}m(m-1)\mu_\alpha$, where

$$\mu_\alpha = Ex_{ij}(\alpha) = 2^{-n} \sum_{t_{\alpha-1} < k \leq t_\alpha} C_n^k,$$

the variance of ξ_α is equal to $\frac{1}{2}m(m-1)\mu_\alpha(1-\mu_\alpha)$ (see Lemma 2).

Theorem 1. *Suppose that, as $n, m \rightarrow \infty$, the variables $t_0, t_1, \dots, t_s$ (s fixed) satisfy the condition*

$$\frac{1}{2}m(m-1)\mu_\alpha \rightarrow \lambda_\alpha < \infty, \quad \alpha = 1, \dots, s. \quad (1)$$

Then the quantities $\xi_\alpha, \alpha = 1, \dots, s$, are asymptotically independent Poisson variables with parameters $\lambda_\alpha, \alpha = 1, \dots, s$.

Theorem 2. *If, in the assumptions of Theorem 1, condition (1) is replaced by the requirement*

$$m^2 \mu_1 \rightarrow \infty, \quad \mu_s \rightarrow 0 \quad ((n - 2t_s)/\sqrt{n} \rightarrow \infty),$$

then the system of quantities

$$\eta_\alpha = \frac{\xi_\alpha - \frac{1}{2}m(m-1)\mu_\alpha}{\left(\frac{1}{2}m(m-1)\mu_\alpha\right)^{1/2}} \quad (2)$$

is asymptotically normal with the identity covariance matrix.

The proof of both theorems is based on estimating the multidimensional moments of the quantities ξ_α with subsequent application of generating functions in the case of Theorem 1 (cf. (1)) and of the second convergence theorem (2) in the case of Theorem 2. Below the proof of Theorem 2 is given.

It will be useful for us, with each collection of random variables $x_{i_1 j_1}(\alpha_1), \dots, x_{i_r j_r}(\alpha_r)$, where $i_\nu < j_\nu$, but among the pairs (i_ν, j_ν) there may also be identical ones, to associate the graph $\Gamma(i_1, j_1; \dots; i_r, j_r)$ of r edges (i_ν, j_ν) , connecting vertices with the corresponding numbers (so that, in this connection, ...

multiple connections of vertices). Denote by $\chi(\Gamma)$ the number of connected components of the graph Γ ; call the index $I(\Gamma)$ a nonnegative integer such that: a) in the graph $\Gamma(i_1, j_1; \dots; i_r, j_r)$ there are $I(\Gamma)$ edges whose deletion from Γ leads to a graph without cycles (closed paths of edges); b) this cannot be achieved by deleting a smaller number of edges.

Lemma 1. *If $I(\Gamma(i_1, j_1; \dots; i_r, j_r)) = 0$, then the random variables*

$$x_{i_1 j_1}(\alpha_1), \dots, x_{i_r j_r}(\alpha_r),$$

where $(\alpha_1, \dots, \alpha_r)$ is an arbitrary set of numbers $0, 1, \dots, s$, are mutually independent.

To prove the lemma it suffices to show that the variables

$$\mathbf{a}_{i_\nu} \oplus \mathbf{a}_{j_\nu}, \quad \nu = 1, \dots, r,$$

are mutually independent. Choose in the graph Γ a vertex from which exactly one edge issues. Let this be, say, the vertex numbered j_r . Then the quantity $\mathbf{a}_{j_r}$ does not enter into any other two-term sums

$$\mathbf{a}_{i_\nu} \oplus \mathbf{a}_{j_\nu}, \quad \nu < r.$$

But this means that the conditional probability that the quantity

$$a_{i_r}^1 \oplus a_{j_r}^1$$

takes the value 1, for any fixed values of

$$a_{i_\nu}^1 \oplus a_{j_\nu}^1, \quad \nu < r,$$

is unchanged and equal to $1/2$; the same is true for the other coordinates of the vector $\mathbf{a}_{i_r} \oplus \mathbf{a}_{j_r}$. Applying induction on the number of edges, we obtain the assertion of Lemma 1.

In what follows we shall denote centering of a random variable by mathematical expectation with the sign $\wedge$:

$$\hat{x}_{ij}(\alpha) = x_{ij}(\alpha) - \mu_\alpha, \quad \hat{\xi}_\alpha = \xi_\alpha - \frac{1}{2}m(m-1)\mu_\alpha,$$

and so on.

Lemma 2. *If $I(\Gamma(i_1, j_1; \dots; i_r, j_r)) = l > 0$, then*

$$\mathbf{E}(\hat{x}_{i_1 j_1}(\alpha) \dots \hat{x}_{i_r j_r}(\alpha)) \leq 2^r \mu_\alpha^{r-l}. \quad (3)$$

Indeed, suppose that the edges to be deleted are

$$(i_{r-l+1}, j_{r-l+1}), \dots, (i_r, j_r).$$

Substitute in (3)

$$\hat{x}_{i_\nu j_\nu}(\alpha) = x_{i_\nu j_\nu}(\alpha) - \mu_\alpha$$

and expand the brackets. As a result we obtain 2^r terms, of which the first is

$$\mathbf{E}x_{i_1 j_1}(\alpha) \dots x_{i_r j_r}(\alpha) \leq \mathbf{E}x_{i_1 j_1}(\alpha) \dots x_{i_{r-l} j_{r-l}}(\alpha) = \mu_\alpha^{r-l}.$$

Every other term is obtained from the first by replacing some $x_{i_\nu j_\nu}(\alpha)$ by μ_α . Therefore, estimating by unity all the remaining

$$x_{i_\nu j_\nu}(\alpha), \quad r-l+1 \leq \nu \leq r,$$

we again arrive at the upper bound μ_α^{r-l} , which gives (3).

Let us now turn to the estimation of the moments of $\hat{\xi}_\alpha$. Obviously,

$$\mathbf{E}\hat{\xi}_\alpha^r = \sum_{i_1 < j_1, \dots, i_r < j_r} \mathbf{E}\hat{x}_{i_1 j_1}(\alpha) \dots \hat{x}_{i_r j_r}(\alpha). \quad (4)$$

We split the sum in (4) into parts Σ_l^k , consisting of all terms

$$\mathbf{E}\hat{x}_{i_1 j_1}(\alpha) \dots \hat{x}_{i_r j_r}(\alpha),$$

for which

$$\chi(\Gamma(i_1, j_1; \dots; i_r, j_r)) = k, \quad I(\Gamma) = l.$$

Note that always $k + l - r \leq 0$, since the number of edges $r - l$ of the reduced graph is not less than the number of its connected components, which coincides with $\chi(\Gamma) = k$.

By virtue of Lemma 1, $\Sigma_l^k = 0$ for all k . Moreover, $\Sigma_l^k = 0$ for $l < k$, since in this case at least one of the k connected components of the graph Γ contains no cycles. For $l \geq k$, estimate every term in Σ_l^k by inequality (3). The number of terms in Σ_l^k does not exceed $c_r m^{r-l+k}$, where c_r is some constant; this estimate follows easily from the fact that $r - l + k$ is the number of vertices of a graph Γ with r edges, for which $\chi(\Gamma) = k$, $I(\Gamma) = l$. Thus,

$$\left(\frac{1}{2} m(m-1)\mu_\alpha\right)^{-r/2} \Sigma_l^k \leq c_r m^{k-l} \mu_\alpha^{r/2-l} = c_r m^{k+l-r} / (m^2 \mu_\alpha)^{l-r/2}. \quad (5)$$

Taking into account both forms of the right-hand side of (5) and considering that

$$k - l \leq 0, \quad k + l - r \leq 0, \quad \mu_\alpha \rightarrow 0, \quad m^2 \mu_\alpha \rightarrow \infty,$$

we obtain that, for $l \neq r/2$

the left-hand side of (5) tends to zero. For odd r this immediately leads to the corresponding result for the odd moments (2). The asymptotic value of the even moment (2) coincides with

$$\lim_{n \rightarrow \infty} \left(\frac{1}{2} m(m-1)\mu_\alpha\right)^{-r/2} \Sigma_{r/2}^{r/2} = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (r-1).$$

Thus we have proved that each of the quantities η_α is asymptotically normal with parameters $(0, 1)$. Further, for $\alpha \neq \beta$, $x_{ij}(\alpha)x_{ij}(\beta) = 0$, and therefore

$$\mathbf{E}\hat{x}_{ij}(\alpha)\hat{x}_{ij}(\beta) = -\mu_\alpha\mu_\beta, \quad \mathbf{E}\hat{x}_{ij}(\alpha)^2 = \mu_\alpha(1 - \mu_\alpha). \quad (6)$$

From this and from (2) it is not difficult to derive the assertion of Theorem 2 concerning the covariance matrix.

The proof of Theorem 2 will be completed if we show that, for any real constants $\lambda_1, \dots, \lambda_s$, the random variable $\lambda_1\eta_1 + \dots + \lambda_s\eta_s$ is asymptotically normal. Bearing in mind again to apply the method of moments, we write:

$$\mathbf{E}(\lambda_1\eta_1 + \dots + \lambda_s\eta_s)^r = \sum_{\alpha_1, \dots, \alpha_r} \lambda_{\alpha_1} \dots \lambda_{\alpha_r} \mathbf{E}\eta_{\alpha_1} \dots \eta_{\alpha_r}. \quad (7)$$

In view of the asymptotic normality of each of the quantities η_α , it is enough for us to show that, for any integers $r_1 \geq 0, \dots, r_s \geq 0$, $r_1 + \dots + r_s = r$,

$$\mathbf{E}(\eta_1^{r_1} \dots \eta_s^{r_s}) / \mathbf{E}\eta_1^{r_1} \dots \mathbf{E}\eta_s^{r_s} \rightarrow 1, \quad n \rightarrow \infty, \quad (8)$$

if all r_α are even, and $\mathbf{E}\eta_1^{r_1} \cdots \eta_s^{r_s} \rightarrow 0$ otherwise. We represent the mathematical expectation of interest to us in the form

$$\frac{1}{\left(\frac{1}{2}m(m-1)\right)^{r/2}} \sum_{i_1 < j_1, \dots, i_r < j_r} E \prod_{\nu \leq r_1} \frac{\hat{x}_{i_\nu j_\nu}^{(1)}}{(\mu_1)^{1/2}} \cdots \prod_{r_1 + \dots + r_{s-1} < \nu \leq r_1 + \dots + r_s} \frac{\hat{x}_{i_\nu j_\nu}^{(s)}}{(\mu_s)^{1/2}}. \quad (9)$$

The estimate of the sum (9) is carried out by partitioning it into parts Σ_l^k and subsequently estimating these parts. For this we shall need the following

Lemma 3. If $I(\Gamma(i_1, j_1; \dots; i_r, j_r)) = l > 0$, then

$$\mathbf{E}(\hat{x}_{i_1 j_1}(\alpha_1) \cdots \hat{x}_{i_r j_r}(\alpha_r)) \leq 2^r \mu_{\beta_1} \cdots \mu_{\beta_{r-l}}, \quad (10)$$

where $(\alpha_1, \dots, \alpha_r)$ is an arbitrary set of numbers $0, 1, \dots, s$, and the set $(\beta_1, \dots, \beta_{r-l})$ is obtained from $(\alpha_1, \dots, \alpha_r)$ by deleting certain components. In this case the following property is fulfilled: at least one of the deleted elements is not smaller than some β_ν .

Just as in the proof of Lemma 2, we come to the consideration of the mathematical expectation $\mathbf{E}x_{i_1 j_1}(\alpha_1) \cdots x_{i_r j_r}(\alpha_r)$. We shall now exclude from the graph Γ those l edges whose existence is guaranteed by the condition of the lemma. Removing one edge at each step, after the $(l-1)$ -st step we arrive at a graph with one cycle. The operation of the last step ensures the fulfillment of the additional property in Lemma 3.

To estimate the sum Σ_l^1 , we multiply the inequality (10) by $\left(\frac{1}{2}m(m-1)\right)^{-r/2}$ and transform its right-hand side into the form

$$c_r'' \frac{1}{m^{r-l+1}} \frac{(\mu_{\alpha_1} \cdots \mu_{\alpha_{r-l-1}})^{1/2} m(\mu_{\alpha_{r-l}})^{1/2}}{m^l (\mu_{\alpha_{r-l+1}} \cdots \mu_{\alpha_r})^{1/2}}, \quad (11)$$

where, for convenience, we have assumed that $\beta_\nu = \alpha_\nu$, $\nu = 1, \dots, r-l$, and α_{r-l} does not exceed some α_ν , $\nu = r-l+1, \dots, r$. In the only case $r=2$, $l=1$, nothing more can be said about the second factor in (11) than that it is bounded. In all other cases it tends to zero (uniformly over all terms of the sum Σ_l^1 , the number of which is $\leq c_r m^{r-l+1}$).

When considering the other Σ_l^k , it is necessary to represent each term of Σ_l^k as a product of k mathematical expectations of products of the quantities

$$\hat{x}_{i_\nu j_\nu}(\alpha)/(\mu_\alpha)^{1/2},$$

corresponding to one and the same connected component of the graph $\Gamma(i_1, j_1; \dots; i_r, j_r)$; then, for each of these k factors, write an estimate of the form (11), in which the roles of r and l will be played, respectively, by the

number of edges of the connected component and its index. As a result, if r is odd, then at least one of the components contains an odd (and hence > 2) number of edges; therefore all odd moments (7) tend to zero.

For even r in the sum (9) it is necessary to retain only the terms corresponding to a graph consisting of $r/2$ connected components, each of which is a cycle of two edges. Moreover, by virtue of (6) one may exclude all terms for which at least one two-edge cycle joins edges corresponding to the quantities

$$\hat{x}_{i,j_\nu}(\alpha)/(\mu_\alpha)^{1/2}$$

from (9) with different α . From this it is easy to derive the result (8) that we need.

The author expresses sincere gratitude to A. N. Kolmogorov for his guidance.

Moscow State University
named after M. V. Lomonosov

Received
22 V 1969

REFERENCES

- ¹ G. S. Plesnevich, DAN, 182, No. 1 (1968). ² M. Loève, *Probability Theory*, IL, 1962.

Note: Figure translations are in progress. See original paper for figures.

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.