Теорія Ймовір. та Матем. Статист.
Вип. 76, 2007

Theor. Probability and Math. Statist.
No. 76, 2008, Pages 103–116
S 0094-9000(08)00735-7
Article electronically published on July 14, 2008

A THEOREM ON THE DISTRIBUTION OF THE RANK OF A SPARSE BOOLEAN RANDOM MATRIX AND SOME APPLICATIONS

UDC 519.21

V. I. MASOL AND S. V. POPERESHNYAK

Abstract. We consider some estimates of the rate of convergence of the distribution of a sparse Boolean random matrix to the Poisson distribution. The results obtained in the paper are applied to estimate the probability that a nonhomogeneous system of Boolean random linear equations is consistent.

1. Introduction

Different methods are used in [1] and [2] to prove that the distribution of the rank $r(A)$ of a sparse Boolean random matrix A approaches the Poisson distribution as

$$T = T(n) \to \infty \quad \text{and } n \to \infty$$

(here n and T mean the number of columns and rows in the matrix A, respectively). If T and n are finite, then the distribution of the rank $r(A)$ can be expressed in terms of the factorial moments of the random variable $r(A)$. These expressions are found in [3]. The asymptotic behavior as $n \to \infty$ of the factorial moments is given in [4, Theorem 4] under certain assumptions, say if $n - T = \text{const}$ and if the distribution of entries of the matrix A depends on the number of the column only.

We are interested in the bounds of the rate of convergence of the distribution of the rank of the matrix A to the Poisson distribution as n grows and in applications of these bounds for an estimation of the probability that the following nonhomogeneous system of linear equations is consistent:

$$AX = B, \tag{1}$$

where the entries of the matrix A of coefficients belong to the field $GF(2)$ consisting of two elements, the vector column $B = (b_1, \dots, b_T)'$ is formed by the random variables $b_1, \dots, b_T$ that do not depend on the entries of the matrix A and moreover $b_1, \dots, b_T$ are jointly independent random variables assuming values in $GF(2)$ and having known distributions, and where $X = (x_1, x_2, \dots, x_n)'$ is an n-dimensional column vector such that $x_i \in GF(2)$, $i = 1, \dots, n$.

In contrast to the papers [1, 2], we consider random matrices A whose entries have the distributions that may depend on their position in the matrix A.

2000 *Mathematics Subject Classification.* Primary 68U20; Secondary 60G10.

Key words and phrases. Boolean random matrix, rank of a matrix, the probability that a system is consistent, the rate of convergence of distributions.

2. Main results

Let the entries of a $T \times n$ matrix $A = \|a_{ij}\|$, $i = 1, \ldots, T$, $j = 1, \ldots, n$, be independent random variables assuming values in $GF(2)$ and whose distributions are given by

$$\mathsf{P}\{a_{ij} = 1\} = 1 - \mathsf{P}\{a_{ij} = 0\} = \frac{\ln n + x_{ij}}{n}, \tag{2}$$

where

$$|x_{ij}| \le c, \qquad c = \text{const},\ i = 1, \ldots, T,\ j = 1, \ldots, n. \tag{3}$$

We assume that the matrix A has at least $n_0 > 1$ columns; thus the distributions (2) are well defined for $n \ge n_0$. Note that a matrix A whose entries have the distributions given by (2)–(3) is called a sparse Boolean matrix.

Denote by $r(A)$ the rank of the matrix A and put

$$\lambda = \frac{1}{n} \sum_{i=1}^{T} \exp\left\{-\frac{1}{n} \sum_{j=1}^{n} x_{ij}\right\}. \tag{4}$$

Theorem 2.1. *Let conditions* (2) *and* (3) *hold. If*

$$\frac{T}{n} \le 1 - \frac{\log_2 \ln n}{(\ln n)^q}, \qquad q = \text{const},\ 0 < q < 1, \tag{5}$$

then

$$\left|\mathsf{P}\{r(A) = T - k\} - \frac{e^{-\lambda} \cdot \lambda^k}{k!}\right| \le 2\,(1 + \delta)\, c(n, k) \frac{\ln^4 n}{n\,(\ln \ln n)^2}$$

for $k = 0, 1, 2, \ldots, n$.

If, additionally,

$$\varliminf_{n \to \infty} \frac{T}{n} > 0, \tag{6}$$

then

$$1 < \varliminf_{n \to \infty} c(n, k) \le \varlimsup_{n \to \infty} c(n, k) \le e^{e^c},$$

where $\delta = \text{const} > 0$. *The explicit expression for the coefficient* $c(n, k)$ *is given by equality* (40).

Further we assume that

$$\mathsf{P}\{a_{ij} = 1\} = 1 - \mathsf{P}\{a_{ij} = 0\} = \frac{\ln T + x_{ij}}{T} \tag{7}$$

and that restrictions (3) are satisfied. Put

$$\lambda_1 = \frac{1}{T} \sum_{j=1}^{n} \exp\left\{-\frac{1}{T} \sum_{i=1}^{T} x_{ij}\right\}.$$

Theorem 2.2. *Let conditions* (3) *and* (7) *hold. If*

$$1 + \frac{\log_2 \ln T}{(\ln T)^q} \le \frac{T}{n}, \qquad q = \text{const},\ 0 < q < 1,$$

then

$$\left|\mathsf{P}\{r(A) = n - k\} - \frac{e^{-\lambda_1} \cdot \lambda_1^k}{k!}\right| \le 2\,(1 + \delta)\, c(T, k) \frac{\ln^4 T}{T\,(\ln \ln T)^2}$$

for $k = 0, 1, 2, \ldots, T$ *where* $\delta = \text{const} > 0$.

If, additionally,

$$\varlimsup_{n\to\infty} \frac{T}{n} < \infty,$$

then

$$1 < \varliminf_{n\to\infty} c(T,k) \le \varlimsup_{n\to\infty} c(T,k) \le e^{e^c}.$$

Consider system (1) where the entries of the $T \times n$ matrix $A = \|a_{ij}\|$ are independent random variables whose distributions satisfy conditions (2)–(3) for $i = 1, \dots, T$, $j = 1, \dots, n$.

The column vector $B = (b_1, \dots, b_T)'$ consists of random variables $b_1, \dots, b_T$ that are jointly independent, do not depend on the entries of the matrix A, and assume values 0 and 1 with probabilities

$$\mathsf{P}\{b_i = 0\} = \frac{1}{2}(1 + \varepsilon_i(n)), \qquad i = 1, 2, \dots, T. \tag{8}$$

Denote by $P_{n,T}$ the probability that system (1) is consistent.

Theorem 2.3. *Let*

$$\frac{T}{n} \le 1 - \frac{\log_2 \ln n}{(\ln n)^q}, \qquad q = \text{const},\ 0 < q < 1,$$

for all $n \ge n_0 > 1$. Assume that conditions (2), (3), *and* (8) *hold. If*

$$\tilde{\varepsilon}(n) \le \beta < 1 \quad \textit{and} \quad \beta = \text{const}, \tag{9}$$

where

$$\tilde{\varepsilon}(n) = \max_{1 \le k \le T} |\varepsilon_k(n)|, \tag{10}$$

then

$$-\alpha_2 f(n) - \tilde{\varepsilon}(n) \le P_{n,T} - \exp\left\{-\frac{\lambda}{2}\right\} \le \alpha_1 f(n) + \tilde{\varepsilon}(n). \tag{11}$$

Here

$$f(n) = 4(1+\delta)\, e^{e^c} \frac{\ln^4 n}{n(\ln\ln n)^2}, \qquad \alpha_1 = 1 + \frac{1}{1 - \tilde{\varepsilon}(n)}, \qquad \alpha_2 = 1 + \Delta(n,T) + \frac{1}{1 + \tilde{\varepsilon}(n)},$$

$\delta = \text{const} > 0$, and

$$0 \le \Delta(n,T) \le \left(\frac{\lambda e}{2T}\right)^{T+1} \frac{\sqrt{2T}}{2T - \lambda e} (f(n))^{-1}. \tag{12}$$

If, additionally, condition (6) *holds, then*

$$\Delta(n,T) \to 0 \quad \textit{as } n \to \infty.$$

3. Auxiliary results for the proof of Theorem 2.1

Denote by $\xi_{n,T}$ the number of nonzero rows in the matrix A.

Lemma 3.1. *If conditions* (2) *and* (3) *hold, then*

$$\left| \mathsf{P}\{\xi_{n,T} = k\} - \frac{e^{-\lambda} \cdot \lambda^k}{k!} \right| \le Q(n,k) \tag{13}$$

for $k = 0, 1, 2, \dots, n$, where

$$Q(n,k) = \frac{\ln^2 n}{n} c_1(n,k) \tag{14}$$

and the coefficient $c_1(n,k)$ is given by equality (18).

If, additionally,

$$\frac{T}{n} \leq 1, \tag{15}$$

then

$$0 < \varliminf_{n\to\infty} c_1(n,k) \leq \varlimsup_{n\to\infty} c_1(n,k) \leq e^{-1}$$

for $k = 0$, *and*

$$0 < \varliminf_{n\to\infty} c_1(n,k) \leq \varlimsup_{n\to\infty} c_1(n,k) \leq \frac{k^k}{k!} e^{-k} \max(k, e^c)$$

for $k \geq 1$.

Proof. It is clear that the probability $p_n^{(i)}$ that a row i of the matrix A is constituted of zeros is equal to

$$p_n^{(i)} = \prod_{j=1}^{n} \left(1 - \frac{\ln n + x_{ij}}{n}\right), \qquad i = 1, \dots, T.$$

Put $a = p_n^{(1)} + p_n^{(2)} + \cdots + p_n^{(T)}$.

According to the Poisson theorem [5],

$$\left| \mathsf{P}\{\xi_{n,T} = k\} - \frac{e^{-a} \cdot a^k}{k!} \right| \leq \frac{1}{n^2} \sum_{i=1}^{T} \exp\left\{ -\frac{2}{n} \sum_{j=1}^{n} p_{ij} \right\}. \tag{16}$$

Using (2) and (3) we find that

$$\left| \frac{\lambda^k e^{-\lambda}}{k!} - \frac{a^k e^{-a}}{k!} \right| \leq \frac{\lambda^k e^{-\lambda} \gamma_n}{k!} \max\left\{ k, \lambda \left(1 + \frac{\lambda \gamma_n}{2} e^{\lambda \gamma_n} \right) \right\}, \tag{17}$$

where

$$\gamma_n = \left(\frac{\ln^2 n}{2n} \right) \frac{\left(1 - c(\ln n)^{-1}\right)^2}{1 - (\ln n - c)\, n^{-1}}.$$

Combining (16) and (17) we see that

$$\left| \mathsf{P}\{\xi_{n,T} = k\} - \frac{e^{-\lambda} \cdot \lambda^k}{k!} \right| \leq Q(n,k),$$

where

$$Q(n,k) = \frac{\ln^2 n}{n} c_1(n,k)$$

and

$$\begin{aligned} c_1(n,k) = {} & \frac{1}{n \ln^2 n} \sum_{i=1}^{T} \exp\left\{ -\frac{2}{n} \sum_{j=1}^{n} p_{ij} \right\} \\ & + \frac{\lambda^k e^{-\lambda}}{2k!} \frac{\left(1 - c(\ln n)^{-1}\right)^2}{1 - (\ln n - c)\, n^{-1}} \max\left\{ k, \lambda \left(1 + \frac{\lambda \gamma_n}{2} e^{\lambda \gamma_n} \right) \right\}. \end{aligned} \tag{18}$$

One can use (2), (3), and (15) to estimate the coefficient $c_1(n,k)$; namely, for $k = 0$,

$$0 < \varliminf_{n\to\infty} c_1(n,k) \leq \varlimsup_{n\to\infty} c_1(n,k) \leq e^{-1}$$

and for $k \geq 1$,

$$0 < \varliminf_{n\to\infty} c_1(n,k) \leq \varlimsup_{n\to\infty} c_1(n,k) \leq \frac{k^k}{k!} e^{-k} \max(k, e^c).$$

Lemma 3.1 is proved. □

Denote by $S_1(A)$ the total number of independent critical sets of the matrix A such that each of them has at least one nonzero row. (Recall [2] that a set $C = \{t_1, \dots, t_m\}$, $m \in \{1, 2, \dots, T\}$, of indices of rows in a matrix A is called a critical set if

$$a_{t_1} \oplus \cdots \oplus a_{t_m} = 0,$$

where the sum is understood coordinate-wise, $a_t = (a_{t1}, \dots, a_{tn})$, and $\oplus$ is the symbol of addition in the field $GF(2)$. Critical sets $C_1, \dots, C_s$ are called independent if the equality $\varepsilon_1 C_1 \Delta \varepsilon_2 C_2 \Delta \cdots \Delta \varepsilon_s C_s = 0$ is equivalent to $\varepsilon_1 = \varepsilon_2 = \cdots = \varepsilon_s = 0$, where $\varepsilon_i \in \{0, 1\}$, $i, j \in \{1, 2, \dots, s\}$, and if

$$C_i \Delta C_j = (C_i \cup C_j) \setminus (C_i \cap C_j)$$

for different critical sets C_i and C_j.

Lemma 3.2. *If condition* (2) *holds, then the mathematical expectation of the random variable* $S_1(A)$ *is given by*

$$\mathsf{E}\, S_1(A) = \sum_{k=0}^{T} \sum_{1 \le i_1 < \cdots < i_k \le T} \frac{1}{2^n} \prod_{j=1}^{n} \left(1 + \prod_{s=1}^{k} \left(1 - \frac{2\,(\ln n + x_{i_s j})}{n}\right)\right)$$
$$- \sum_{k=0}^{T} \sum_{1 \le i_1 < \cdots < i_k \le T} \prod_{s=1}^{k} \prod_{j=1}^{n} \left(1 - \frac{\ln n + x_{i_s j}}{n}\right).$$

Proof. Lemma 3.2 is proved similarly to the corresponding assertion in [2], where the expression for the mathematical expectation of $\mathsf{E}\, S_1(A)$ is obtained for the case of

$$x_{ij} = x, \qquad i = 1, \dots, T,\ j = 1, \dots, n. \qquad \square$$

Put

$$f(n) = \left[(\ln \ln n)^{-1}\,(1 + \delta_1) \ln n\right],$$

where $\delta_1 = \text{const}, \delta_1 > 0$, and

$$\mu(n) = \sum_{k=0}^{f(n)} \sum_{1 \le i_1 < \cdots < i_k \le T} \exp\left\{-\sum_{s=1}^{k} \sum_{j=1}^{n} \frac{\ln n + x_{i_s j}}{n}\right\}.$$

Here $[\cdot]$ denotes the integer part.

Lemma 3.3. *Assume that conditions* (2), (3), *and* (6) *are valid. If*

$$T \le n, \tag{19}$$

then

$$1 < \varliminf_{n \to \infty} \mu(n) \le \varlimsup_{n \to \infty} \mu(n) \le e^{e^c}. \tag{20}$$

Proof. First we estimate $\mu(n)$ from above. The sum $\mu(n)$ can be represented in the following form:

$$\mu(n) = \theta_1(n) - \theta_2(n), \tag{21}$$

where

$$\theta_1(n) = \sum_{k=0}^{T} \sum_{1 \le i_1 < \cdots < i_k \le T} \exp\left\{-\sum_{s=1}^{k} \sum_{j=1}^{n} \frac{\ln n + x_{i_s j}}{n}\right\},$$
$$\theta_2(n) = \sum_{k=f(n)+1}^{T} \sum_{1 \le i_1 < \cdots < i_k \le T} \exp\left\{-\sum_{s=1}^{k} \sum_{j=1}^{n} \frac{\ln n + x_{i_s j}}{n}\right\}.$$

Then

$$\theta_1(n) \le e^{\lambda}. \tag{22}$$

Since $\theta_2(n) \ge 0$, we derive from (21) and (22) that

$$\mu(n) \le e^{\lambda}. \tag{23}$$

Now we estimate $\mu(n)$ from below. We have

$$\theta_1(n) \ge e^{\lambda} \exp\left\{-\frac{1}{2n^2}\sum_{i=1}^{T}\exp\left(-\frac{2}{n}\sum_{j=1}^{n} x_{ij}\right)\right\}.$$

Taking into account representation (21) we obtain

$$\mu(n) \ge e^{\lambda} \exp\left\{-\frac{1}{2n^2}\sum_{i=1}^{T}\exp\left(-\frac{2}{n}\sum_{j=1}^{n} x_{ij}\right)\right\} - \theta_2(n).$$

The latter inequality together with (23) yields

$$e^{\lambda} \exp\left\{-\frac{1}{2n^2}\sum_{i=1}^{T}\exp\left(-\frac{2}{n}\sum_{j=1}^{n} x_{ij}\right)\right\} - \theta_2(n) \le \mu(n) \le e^{\lambda}. \tag{24}$$

Thus

$$\theta_2(n) \to 0 \quad \text{as } n \to \infty. \tag{25}$$

Now we pass to the limit as $n \to \infty$ in inequality (24) and get (20) by relation (25) and conditions (3), (6), and (19).

Lemma 3.3 is proved. □

Put

$$\Gamma(n) = u\left(1 + \frac{u(1+\delta_1)}{2} e^{u(1+\delta_1)}\right), \qquad u = \frac{\ln^4 n}{n\,(\ln\ln n)^2}.$$

Lemma 3.4. *If conditions* (2) *and* (3) *are valid, then*

$$\begin{aligned}\sum_{k=0}^{f(n)} \sum_{1\le i_1<\cdots<i_k\le T} \frac{1}{2^n}\prod_{j=1}^{n}\left(1+\prod_{s=1}^{k}\left(1-\frac{2\,(\ln n + x_{i_s j})}{n}\right)\right)& \\ \le \mu(n) + \mu(n)(1+\delta)\Gamma\,(n)\,.&\end{aligned} \tag{26}$$

Proof. We estimate the left hand side of relation (26) as follows:

$$\begin{aligned}&\sum_{k=0}^{f(n)} \sum_{1\le i_1<\cdots<i_k\le T} \frac{1}{2^n}\prod_{j=1}^{n}\left(1+\prod_{s=1}^{k}\left(1-\frac{2\,(\ln n + x_{i_s j})}{n}\right)\right) \\ &\quad\le \sum_{k=0}^{f(n)} \sum_{1\le i_1<\cdots<i_k\le T} \prod_{j=1}^{n}\left(1-\sum_{s=1}^{k}\frac{(\ln n + x_{i_s j})}{n} + \left(\sum_{s=1}^{k}\frac{(\ln n + x_{i_s j})}{n}\right)^2\right).\end{aligned}$$

The latter inequality with the help of (3) implies that

$$\begin{aligned}\sum_{k=0}^{f(n)} \sum_{1\le i_1<\cdots<i_k\le T} \frac{1}{2^n}\prod_{j=1}^{n}\left(1+\prod_{s=1}^{k}\left(1-\frac{2\,(\ln n + x_{i_s j})}{n}\right)\right)& \\ \le \mu(n)\cdot\exp\left\{\frac{f^2(n)\,(\ln n + c)^2}{n}\right\}.&\end{aligned} \tag{27}$$

Further

$$\exp\left\{\frac{f^2(n)\,(\ln n + c)^2}{n}\right\} \le 1 + u\,(1+\delta)\left(1 + \frac{u}{2}\,(1+\delta)\cdot\exp\left\{u\,(1+\delta)\right\}\right). \tag{28}$$

Relations (27) and (28) imply inequality (26).

Lemma 3.4 is proved. □

Lemma 3.5. *Let conditions* (2), (3), (6), *and* (19) *hold. Then*

$$\sum_{k=0}^{f(n)} \sum_{1\le i_1<\cdots<i_k\le T} \prod_{s=1}^{k}\prod_{j=1}^{n}\left(1 - \frac{\ln n + x_{i_s j}}{n}\right) \ge \mu\,(n) - c_2(n)\frac{\ln^3 n}{n\cdot\ln\ln n},$$

where

$$\frac{(1+\delta_1)}{2} < \varliminf_{n\to\infty} c_2(n) \le \varlimsup_{n\to\infty} c_2\,(n) \le \frac{e^{e^c}}{2}\,(1+\delta_1)\,.$$

Proof. Using the notation introduced above we obtain from relation (3) that

$$\sum_{k=0}^{f(n)} \sum_{1\le i_1<\cdots<i_k\le T} \prod_{s=1}^{k}\prod_{j=1}^{n}\left(1 - \frac{\ln n + x_{i_s j}}{n}\right) \ge \mu(n) - c_2(n)\cdot\frac{\ln^3 n}{n\cdot\ln\ln n},$$

where

$$c_2(n) = \frac{\mu(n)}{2}\,(1+\delta_1)\left(1+\frac{c}{\ln n}\right)^2 \frac{1}{1-\frac{\ln n + c}{n}}. \tag{29}$$

Taking into account inequalities (20) and passing to the limit as $n\to\infty$ we obtain

$$\frac{(1+\delta_1)}{2} < \varliminf_{n\to\infty} c_2(n) \le \varlimsup_{n\to\infty} c_2(n) \le \frac{1}{2}e^{e^c}\,(1+\delta_1)\,.$$

Lemma 3.5 is proved. □

Lemma 3.6. *Let conditions* (2), (3), (6), *and* (19) *hold. Then*

$$\sum_{k:\frac{n}{2}\left(1-\frac{1}{\ln n}\right)\le k\le\frac{n}{2}\left(1+\frac{1}{\ln n}\right)} \binom{n}{k}\frac{1}{2^n}\sum_{l=f(n)}^{T}\binom{T}{l}\left(1-\frac{2\,(\ln n - c)}{n}\right)^{kl} \le \left(\frac{c_3(n)}{f(n)}\right)^{f(n)}\cdot\frac{1}{\sqrt{f(n)}}\cdot c_4(n),$$

where $0 < \varliminf_{n\to\infty} c_3(n) \le \varlimsup_{n\to\infty} c_3(n) \le e^{2+c}$ *and* $\lim_{n\to\infty} c_4(n) = (2\pi)^{-1/2}$.

Proof. For $(n/2)(1-1/\ln n) \le k \le (n/2)(1+1/\ln n)$, we have

$$\left(1-\frac{2\,(\ln n - c)}{n}\right)^k \le e^{-2k(\ln n - c)/n} \le \frac{1}{n}e^{c+1-c/\ln n}.$$

Using the above estimate and applying the Stirling formula we obtain

$$\sum_{k:\frac{n}{2}\left(1-\frac{1}{\ln n}\right)\le k\le\frac{n}{2}\left(1+\frac{1}{\ln n}\right)} \binom{n}{k}\frac{1}{2^n}\sum_{l=f(n)}^{T}\binom{T}{l}\left(1-\frac{2\,(\ln n - c)}{n}\right)^{kl} \le \left(\frac{c_3(n)}{f(n)}\right)^{f(n)}\cdot\frac{1}{\sqrt{f(n)}}\cdot c_4(n),$$

where

$$c_3(n) = \frac{T}{n} \cdot e^{2+c-c/\ln n}, \tag{30}$$

$$c_4(n) = \frac{1}{\sqrt{2\pi}} \cdot \left(1 - \frac{c_3(n)}{f(n)}\right)^{-1}. \tag{31}$$

Considering (6) and (19) we get

$$0 < \varliminf_{n\to\infty} c_3(n) \le \varlimsup_{n\to\infty} c_3(n) \le e^{2+c}, \qquad \lim_{n\to\infty} c_4(n) = (2\pi)^{-1/2}.$$

Lemma 3.6 is proved. □

Lemma 3.7. *Let conditions* (2), (3), *and* (5) *hold. Then*

$$\begin{aligned}\sum_{k=f(n)+1}^{T} \sum_{1\le i_1<\cdots<i_k\le T} \frac{1}{2^n} \prod_{j=1}^{n} \left(1 + \prod_{s=1}^{k} \left(1 - \frac{2\left(\ln n + x_{i_s j}\right)}{n}\right)\right) \\ \le c_5(n) \cdot (\ln n)^{-(n\cdot c_6(n))/(\ln n)^q} \\ + \left(\frac{c_3(n)}{f(n)}\right)^{f(n)} \cdot \frac{1}{\sqrt{f(n)}} \cdot c_4(n) + c_7(n) \cdot \exp\left\{-\frac{n}{2\ln^2 n} c_8(n)\right\},\end{aligned}$$

where

$$c_5(n) = \frac{1}{\sqrt{2\pi}}\left(1 + \frac{(\ln n)^q}{n}\right), \tag{32}$$

$$c_6(n) = 1 - q - \frac{1}{\ln\ln n}\left(1 - \ln\left(1 - \frac{(\ln n)^q}{n}\right) + \frac{1}{2n}(\ln n)^{1+q} - \frac{q(\ln n)^q \ln\ln n}{2n}\right), \tag{33}$$

$$c_7(n) = \frac{1}{\sqrt{2\pi}\left(1 - \frac{1}{\ln^2 n}\right)}\left(\exp\left\{\frac{T}{n} e^{c+1-\frac{c}{\ln n}+2\frac{\ln n - c}{n}}\right\} + \exp\left\{\frac{T}{n} e^{c-1+\frac{c}{\ln n}}\right\}\right), \tag{34}$$

$$c_8(n) = \left(1 - \frac{1}{2\ln n} - \frac{1}{2\ln^2 n} - \frac{\ln^3 n}{n}\right). \tag{35}$$

If, additionally, condition (6) *holds, then*

$$\lim_{n\to\infty} c_5(n) = (2\pi)^{-1/2}, \qquad \lim_{n\to\infty} c_6(n) = 1 - q, \qquad \lim_{n\to\infty} c_8(n) = 1,$$

$$\sqrt{\frac{2}{\pi}} < \varliminf_{n\to\infty} c_7(n) \le \varlimsup_{n\to\infty} c_7(n) \le \frac{e^{c+1} + e^{c-1}}{\sqrt{2\pi}}.$$

Proof. The proof of Lemma 3.7 uses the result of Lemma 3.6 and follows the lines of the proof of Lemma 3.3.2 in [2]. □

Denote by $S(A)$ the maximal number of independent critical sets of a matrix A.

Lemma 3.8. *Assume that conditions* (2) *and* (3) *hold for* $k = 0, 1, 2, \ldots, n$. *Then*

$$\left|\mathsf{P}\left\{S(A) = k\right\} - \frac{e^{-\lambda}\cdot\lambda^k}{k!}\right| \le 2\,\mathsf{E}\,S_1(A) + Q(n,k),$$

where $\mathsf{E}\,S_1(A)$ *is found in Lemma* 3.2, *while* $Q(n,k)$ *is obtained in Lemma* 3.1.

Proof. It follows from (13) that

$$\begin{aligned}\left|\mathsf{P}\{S(A) = k\} - \frac{e^{-\lambda}\cdot\lambda^k}{k!}\right| &\le \left|\mathsf{P}\{S(A) = k\} - \mathsf{P}\{\xi_{n,T} = k\}\right| + Q(n,k) \\ &= \left|\mathsf{P}\{S_1(A) + \xi_{n,T} = k\} - \mathsf{P}\{\xi_{n,T} = k\}\right| + Q(n,k).\end{aligned}$$

Applying the following relations

$$\mathsf{P}\{S_1(A)+\xi_{n,T}=k\}=\sum_{l=0}^{k}\mathsf{P}\{S_1(A)=l,\xi_{n,T}=k-l\},$$
$$\big|\mathsf{P}\{S_1(A)=0,\xi_{n,T}=k\}-\mathsf{P}\{\xi_{n,T}=k\}\big|\le \mathsf{P}\{S_1(A)\ge 1\}$$

and Chebyshev's inequality we prove that

$$\left|\mathsf{P}\{S(A)=k\}-\frac{e^{-\lambda}\cdot\lambda^k}{k!}\right|\le 2\,\mathsf{P}\{S_1(A)\ge 1\}+Q(n,k)\le 2\,\mathsf{E}\,S_1(A)+Q(n,k).$$

Lemma 3.8 is proved. □

4. Proof of Theorem 2.1

First we obtain a bound for $\mathsf{E}\,S_1(A)$. For $k\ge 0$, put

$$\Delta_1(k)=\sum_{1\le i_1<\dots<i_k\le T}\frac{1}{2^n}\prod_{j=1}^{n}\left(1+\prod_{s=1}^{k}\left(1-\frac{2(\ln n+x_{i_sj})}{n}\right)\right),$$
$$\Delta_2(k)=\sum_{1\le i_1<\dots<i_k\le T}\prod_{s=1}^{k}\prod_{j=1}^{n}\left(1-\frac{\ln n+x_{i_sj}}{n}\right).$$

Lemma 3.2 implies that

$$\mathsf{E}\,S_1(A)\le\sum_{k=0}^{f(n)}\Delta_1(k)+\sum_{k=f(n)+1}^{T}\Delta_1(k)-\sum_{k=0}^{f(n)}\Delta_2(k).$$

By Lemmas 3.4 and 3.5 we get

$$\sum_{k=0}^{f(n)}\Delta_1(k)-\sum_{k=0}^{f(n)}\Delta_2(k)\le\mu(n)(1+\delta)\Gamma(n)+c_2(n)\frac{\ln^3 n}{n\cdot\ln\ln n}. \tag{36}$$

Combining (36) and the estimate for the sum $\sum_{k=f(n)+1}^{T}\Delta_1(k)$ obtained in Lemma 3.7 we derive the inequality

$$\mathsf{E}\,S_1(A)\le\mu(n)\,(1+\delta)\,\Gamma(n)+F(n), \tag{37}$$

where

$$\begin{aligned}F(n)&=c_2(n)\frac{\ln^3 n}{n\cdot\ln\ln n}+c_5(n)\cdot(\ln n)^{-n\cdot c_6(n)/(\ln n)^q}\\&\quad+\left(\frac{c_3(n)}{f(n)}\right)^{f(n)}\cdot\frac{1}{\sqrt{f(n)}}\cdot c_4(n)+c_7(n)\cdot\exp\left\{-\frac{n}{2\ln^2 n}c_8(n)\right\}\end{aligned} \tag{38}$$

and where the coefficients $c_2(n),c_3(n),\dots,c_8(n)$ are defined by equalities (29)–(35), respectively.

According to (37) and Lemma 3.8,

$$\left|\mathsf{P}\{S(A)=k\}-\frac{e^{-\lambda}\cdot\lambda^k}{k!}\right|\le 2\big\langle\mu(n)\,(1+\delta)\,\Gamma(n)+F(n)\big\rangle+Q(n,k).$$

Considering the explicit expression for $\Gamma(n)$ we write

$$\left|\mathsf{P}\{S(A)=k\}-\frac{e^{-\lambda}\cdot\lambda^k}{k!}\right|\le 2\,(1+\delta)\,c_{11}(n)u+2F(n)+Q(n,k),$$

where

$$c_{11}(n) = \mu(n) + \mu(n)\frac{u}{2}\,(1+\delta)\exp\{u\,(1+\delta)\}\,. \tag{39}$$

The latter inequality implies

$$\left|\mathsf{P}\{r(A) = T-k\} - \frac{e^{-\lambda}\cdot\lambda^k}{k!}\right| \le 2\,(1+\delta)\,c(n,k)\frac{\ln^4 n}{n\,(\ln\ln n)^2},$$

since $r(A) + S(A) = T$ (the latter equality is proved in [2]), where

$$c(n,k) = c_{11}(n) + \frac{F(n)}{(1+\delta)u} + \frac{Q(n,k)}{2u(1+\delta)} \tag{40}$$

and $c_{11}(n)$, $F(n)$, and $Q(n,k)$ are defined by equalities (39), (38), and (14), respectively.

Theorem 2.1 is proved.

5. Proof of Theorem 2.2

Let A be the matrix defined in Theorem 2.2 and let A' denote the transpose matrix of A. Theorem 2.1 can be applied to A', since the analogue of condition (5) written for A' is

$$\frac{n}{T} \le 1 - \upsilon + \frac{\upsilon^2}{1-\upsilon}, \tag{41}$$

where

$$\upsilon = \frac{\log_2 \ln T}{(\ln T)^q}.$$

Note that the above condition follows from $T/n \ge 1+\upsilon$.

The term $\upsilon^2(1-\upsilon)^{-1}$ on the right hand side of (41) changes the coefficient $c_6(n)$ in Theorem 2.2. The changed coefficient is equal to

$$c_6(T) + \frac{\upsilon}{1-\upsilon},$$

where $c_6(n)$ is defined by equality (33).

Theorem 2.2 is proved.

6. Proof of Theorem 2.3

Denote by $\mu_{n,T}$ the number of solutions of system (1). The probability $P_{n,T}$ that system (1) is consistent equals the probability that the system has at least one solution, that is, $P_{n,T} = \mathsf{P}\{\mu_{n,T} > 0\}$.

We find an explicit expression for the probability of the event $\{\mu_{n,T} > 0\}$ given that the rank $r(A)$ of the matrix A is equal to r. Without loss of generality we assume that the rows $1, 2, \dots, r$ are linearly independent in the matrix A. Then each of the rows $r+1, \dots, T$ is a linear combination of the first r rows. In order that the system be consistent, its right hand sides $b_{r+1}, \dots, b_T$ should be such that

$$\varepsilon^*_{1i} b_1 \oplus \cdots \oplus \varepsilon^*_{ri} b_r = b_i, \qquad i = r+1, \dots, T, \tag{42}$$

where $\varepsilon^*_{1i}, \dots, \varepsilon^*_{ri}$ assume only the values 0 or 1.

The probability of an arbitrary relation among (42) is equal to

$$\mathsf{P}\left\{b_{\nu_1} \oplus \cdots \oplus b_{\nu_s} = b_i\right\} = \frac{1}{2}\left(1 + \left(\prod_{j=1}^{s(i)} \varepsilon_{\nu_j}(n)\right)\varepsilon_i(n)\right), \tag{43}$$

where $1 \le s \le r$, $s = s(i)$, $i = r+1, \dots, T$, and $s(i)$ is the number of nonzero elements among $\varepsilon^*_{1i}, \dots, \varepsilon^*_{ri}$ on the left hand side of (42). Equality (43) can be checked by induction.

Taking into account (9), (10), (43), and

$$P_{n,T} = \sum_{k=0}^{T} \mathsf{P}\{r(A) = T - k\} \cdot \mathsf{P}\left\{\mu_{n,T} > 0 \;/\; r(A) = T - k\right\}$$

we obtain

$$\sum_{k=0}^{T} \mathsf{P}\left\{r\left(A\right) = T - k\right\} \left(\frac{1 - \tilde{\varepsilon}(n)}{2}\right)^{k} \le P_{n,T} \le \sum_{k=0}^{T} \mathsf{P}\left\{r(A) = T - k\right\} \left(\frac{1 + \tilde{\varepsilon}(n)}{2}\right)^{k}. \tag{44}$$

Further we show that

$$P_{n,T} - e^{-\lambda/2} \le \alpha_1 f\left(n\right) + \tilde{\varepsilon}(n). \tag{45}$$

First we estimate from above the difference $P_{n,T} - e^{-\lambda/2}$ by using the right hand side of relation (44):

$$P_{n,T} - e^{-\lambda/2} \le \sum_{k=0}^{T} \left(\mathsf{P}\left\{r(A) = T - k\right\} - \frac{e^{-\lambda}\lambda^{k}}{k!}\right) \frac{1}{2^{k}} + \sum_{k=0}^{T} \frac{e^{-\lambda}\lambda^{k}}{k!} \frac{1}{2^{k}} - e^{-\lambda/2} + \theta,$$

where

$$\theta = \sum_{k=0}^{T} \left(\mathsf{P}\left\{r(A) = T - k\right\} - \frac{e^{-\lambda}\lambda^{k}}{k!}\right) \frac{\left(1 + \tilde{\varepsilon}(n)\right)^{k} - 1}{2^{k}} + \sum_{k=0}^{T} \frac{e^{-\lambda}\lambda^{k}}{k!} \frac{\left(1 + \tilde{\varepsilon}(n)\right)^{k} - 1}{2^{k}}.$$

Now we use Theorem 2.1 and get

$$P_{n,T} - e^{-\lambda/2} \le f(n) + \theta. \tag{46}$$

We again use Theorem 2.1 to estimate the sum θ from above:

$$\theta \le \frac{f(n)}{2} \sum_{k=0}^{T} \frac{\left(1 + \tilde{\varepsilon}(n)\right)^{k}}{2^{k}} + e^{-\lambda} \sum_{k=0}^{T} \frac{\lambda^{k}}{k!} \frac{\tilde{\varepsilon}(n) 2^{k}}{2^{k}},$$

whence

$$\theta \le \frac{f(n)}{2} \frac{1}{1 - (1 + \tilde{\varepsilon}(n))/2} + \tilde{\varepsilon}(n). \tag{47}$$

Bounds (46) and (47) imply (45).

Now we show that

$$P_{n,T} - e^{-\lambda/2} \ge -\alpha_2 f\left(n\right) - \tilde{\varepsilon}(n). \tag{48}$$

Indeed, using the left hand side of relation (44) we estimate the difference $P_{n,T} - e^{-\lambda/2}$ from below:

$$P_{n,T} - e^{-\lambda/2} \ge \sum_{k=0}^{T} \left(\mathsf{P}\left\{r(A) = T - k\right\} - \frac{e^{-\lambda}\lambda^{k}}{k!}\right) \frac{1}{2^{k}} + \sum_{k=0}^{T} \frac{e^{-\lambda}\lambda^{k}}{k!} \frac{1}{2^{k}} - e^{-\lambda/2} + \theta,$$

where

$$\theta = \sum_{k=0}^{T} \left(\mathsf{P}\{r(A) = T - k\} - \frac{e^{-\lambda}\lambda^{k}}{k!}\right) \frac{\left(1 - \tilde{\varepsilon}(n)\right)^{k} - 1}{2^{k}} + \sum_{k=0}^{T} \frac{e^{-\lambda}\lambda^{k}}{k!} \frac{\left(1 - \tilde{\varepsilon}(n)\right)^{k} - 1}{2^{k}}.$$

Theorem 2.1 implies that

$$P_{n,T} - e^{-\lambda/2} \ge -f(n) + \sum_{k=0}^{T} \frac{e^{-\lambda}\lambda^{k}}{k!} \frac{1}{2^{k}} - e^{-\lambda/2} + \theta,$$

whence

$$P_{n,T} - e^{-\lambda/2} \geq -f(n) - \sum_{k=T+1}^{\infty} \frac{e^{-\lambda}(\lambda/2)^k}{k!} + \theta \geq -f(n) - e^{-\lambda} \sum_{k=T+1}^{\infty} \left(\frac{\lambda e}{2k}\right)^k \frac{1}{\sqrt{2\pi k}} + \theta.$$

It follows from (3) and (4) that the inequality $\lambda e/(2T) < 1$ holds for $n > 2^{-1}e^{1+c}$. For this n, we get

$$P_{n,T} - e^{-\lambda/2} \geq -f(n)\left(1 + \Delta\left(n, T\right)\right) + \theta, \tag{49}$$

where $\Delta(n, T)$ satisfies the bounds (12).

The sum θ is estimated from below with the help of Theorem 2.1:

$$\theta \geq -\frac{f(n)}{2} \sum_{k=0}^{T} \frac{\left(1 - \tilde{\varepsilon}(n)\right)^k}{2^k} - e^{-\lambda} \sum_{k=0}^{T} \frac{\lambda^k}{k!} \frac{\tilde{\varepsilon}(n) 2^k}{2^k},$$

whence

$$\theta \geq -\frac{f(n)}{2} \frac{1}{1 - \frac{1-\tilde{\varepsilon}(n)}{2}} - \tilde{\varepsilon}(n). \tag{50}$$

Relation (49) together with (12) and (50) implies (48).

The upper and lower bounds (45) and (48) prove relation (11).

Theorem 2.3 is proved.

7. The limit distribution of the rank of a sparse Boolean random matrix

Theorem 7.1. ***I***. *Let the assumptions of Theorem* 2.1 *hold. If*

$$\lambda \to \omega, \qquad n \to \infty, \tag{51}$$

and $0 < \omega < \infty$, *then*

$$\mathsf{P}\left\{r(A) = T - k\right\} \to e^{-\omega} \frac{\omega^k}{k!}, \qquad n \to \infty, \tag{52}$$

for a fixed $k = 0, 1, 2, \dots$.

II. *Let conditions* (2), (3), (5), *and* (51) *hold. If* $\omega > 0$, *then* $\omega < \infty$ *and relation* (52) *is valid.*

The proof of Theorem 7.1 is easy in view of Theorem 2.1.

Corollary 7.1. *Let condition* (2) *hold for*

$$x_{ij} = x, \qquad i = 1, \dots, T, \ j = 1, \dots, n,$$

where x *is a fixed number. If* $T/n \to \alpha$ *as* $n \to \infty$ *for some* $0 < \alpha < 1$, *then relation* (52) *is valid for* $\omega = \alpha e^{-x}$.

Theorem 7.2. ***I***. *Let the assumptions of Theorem* 2.2 *hold. If*

$$\lambda_1 \to \omega_1, \qquad n \to \infty, \tag{53}$$

and $0 < \omega_1 < \infty$, *then*

$$\mathsf{P}\left\{r(A) = n - k\right\} \to e^{-\omega_1} \frac{\omega_1^k}{k!}, \qquad n \to \infty, \tag{54}$$

for a fixed $k = 0, 1, 2, \dots$.

II. *Let conditions* (3), (7), *and* (53) *hold. If*

$$1 + \frac{\log_2 \ln T}{(\ln T)^q} \leq \frac{T}{n},$$

$q = \text{const}$, $0 < q < 1$, *and* $\omega_1 > 0$, *then* $\omega_1 < \infty$ *and relation* (54) *is valid.*

The proof of Theorem 7.2 is easy in view of Theorem 2.2.

Corollary 7.2. *Let condition* (2) *hold for*

$$x_{ij} = x, \qquad i = 1, \dots, T,\ j = 1, \dots, n,$$

where x *is a fixed number. If* $T/n \to \alpha$ *as* $n \to \infty$ *for some* $1 < \alpha < \infty$*, then relation* (54) *is valid for* $\omega_1 = \alpha^{-1}e^{-x}$.

The proof of Corollary 7.1 (Corollary 7.2) follows from the observation that the assumptions of the first (second) statement of Theorem 7.1 (Theorem 7.2) obviously hold under the assumptions of the corresponding corollary.

Remark 7.1. Corollaries 7.1 and 7.2 are proved in [2, Theorem 3.3.1] and [2, Theorem 3.3.2], respectively.

8. The asymptotic behavior of the probability $P_{n,T}$

Theorem 8.1. *Let conditions* (2), (3), (5), (6), *and* (8) *hold. If*

$$\lambda \to \lambda^* \tag{55}$$

and

$$\tilde{\varepsilon}(n) \to 0 \tag{56}$$

as $n \to \infty$*, then*

$$0 < \lambda^* < \infty \tag{57}$$

and moreover

$$P_{n,T} \to \exp\left\{-\frac{\lambda^*}{2}\right\}, \qquad n \to \infty. \tag{58}$$

Inequality (57) can easily be derived from (3), (5), and (6). Relation (58) follows from Theorem 2.3 by (55) and (56).

Corollary 8.1. *Assume that condition* (2) *is valid for*

$$x_{ij} = x, \qquad i = 1, \dots, T,\ j = 1, \dots, n,$$

where x *is a fixed number and* $T/n \to \alpha$ *as* $n \to \infty$ *for some* $0 < \alpha < 1$. *Then relation* (58) *holds for* $\lambda^* = \alpha e^{-x}$.

Bibliography

1. G. V. Balakin, *The distribution of the rank of random matrices over a finite field*, Teor. Verojatnost. i Primenen. **XIII** (1968), no. 4, 631–641; English transl. in Theor. Probab. Appl. **13** (1968), no. 4, 594–605. MR0243571 (39:4892)
2. V. F. Kolchin, *Random Graphs*, Fizmatlit, Moscow, 2000, 256 pp. (Russian) MR1812261 (2002e:60014)
3. V. I. Masol, *Moments of the number of solutions of a system of random Boolean equations*, Random Oper. Stochastic Equations **1** (1993), no. 2, 171–179. MR1254185 (94h:60089)

4. V. I. Masol, *Invariance theorems for systems of random Boolean equations*, Sixth Intern. Vilnius Conf. of Probability Theory and Math. Statist., Abstracts of Communications, 1993, pp. 19–20.
5. B. A. Sevast'yanov, *A Course in Probability Theory and Mathematical Statistics*, "Nauka", Moscow, 1982. (Russian) MR712294 (85a:60006)

Department of Probability Theory and Mathematical Statistics, Faculty for Mechanics and Mathematics, National Taras Shevchenko University, Academician Glushkov Avenue 6, Kyiv 03127, Ukraine
E-mail address: vimasol@ukr.net

Department of Probability Theory and Mathematical Statistics, Faculty for Mechanics and Mathematics, National Taras Shevchenko University, Academician Glushkov Avenue 6, Kyiv 03127, Ukraine
E-mail address: Popereshnyak_sv@mail.ru

Received 27/DEC/2005

Translated by N. SEMENOV