The largest triangular submatrix of a random matrix

Svante Janson
(joint work with Zur Izhakian and John Rhodes)

Consider a random random $n \times n$ matrix $X_n = (x_{ij})$, where the entries are i.i.d. and, for example, $\{0,1\}$-valued with a fixed probability $\mathbb{P}(x_{ij} = 0) = p_0 > 0$.

Problem 1. What is the size T_n of the largest (lower) triangular submatrix of X_n?

A submatrix of order m is defined by selecting subsets of m rows and m columns. We may also allow reordering the rows and columns, and consider the largest permuted triangular submatrix. Moreover, the problem came originally from a question in supertropical algebra; in that setting the matrix elements can take three values, and we are really interested in triangular submatrices with 1's on the diagonal, see [3] and the references there for details. Asymptotically, to the first order treated here, the different versions have the same answer, and we consider for simplicity only the version stated above.

Theorem 2. *Let $Q = 1/p_0 > 1$. Then, as $n \to \infty$,*

$$T_n / \log_Q n \xrightarrow{\text{p}} 2 + \sqrt{2},$$

where $\xrightarrow{\text{p}}$ denotes convergence in probability.

Remark. The corresponding problem of the largest square submatrix with only 0's (or, equivalently, after interchange of 0 and 1, with only 1's) has been studied by several authors, see [6] and the references therein. It is shown in [6] that if S_n is the size of the largest such matrix, then $S_n / \log_Q n \xrightarrow{\text{p}} 2$. This problem can be seen as finding the largest balanced complete subgraph of a random bipartite graph. The analogous problem of finding the largest complete set in a random graph $G(n,p)$ (or, equivalently, the largest independent set in $G(n, 1-p)$) was solved by [2] and [5], see also [1] and [4]; again the size, C_n say, is asymptotically $2 \log_Q n$, where $Q = 1/p$.

Note that $T_n \geq S_n \geq \lfloor T_n/2 \rfloor$, which shows that T_n and S_n are equal within a factor of $2 + o(1)$, and in particular of the same order of magnitude. However, it does not seem possible to get the right constant in front of $\log_Q n$ for one of these problems from the other.

A simple calculation shows that the expected number of triangular submatrices of size $c \log_Q n$ tends to infinity if $c < 4$; hence the first moment method is not useful here. The reason is that triangular submatrices of size $c \log_Q n$ with $2 + \log_Q 2 < c < 4$ are unlikely, but if they occur, they tend to occur in large groups.

The proof is therefore based on studying a truncated version of triangular submatrices, and then extend these to triangular matrices. See [3] for details.

Open problems

For the largest square zero submatrix and the largest cliques in $G(n,p)$, much more precise estimates are known, see [6] and [1, 4]; for example, it follows that if

$$s(n) = 2\log_Q n - 2\log_Q \log_Q n + 2\log_Q(e/2),$$

then for any $\epsilon > 0$, $\lfloor s(n) - \epsilon \rfloor \le S_n \le \lfloor s(n) + \epsilon \rfloor$ and $\lfloor s(n) + 1 - \epsilon \rfloor \le C_n \le \lfloor s(n) + 1 + \epsilon \rfloor$ w.h.p. (and, in fact, almost surely); in particular the sizes are concentrated on one or at most two values. It would be interesting to find similar sharper versions of the result above, which leads to the following open problems.

Problem 3. Find second order terms for T_n.

Problem 4. Is T_n concentrated on at most two values?

Problem 5. Prove a version of Theorem 2 (or a stronger result) with convergence almost surely instead of just in probability.

Problem 6. Find corresponding results when p_0 and p_1 depend on n.

Problem 7. Find corresponding results for rectangular matrices.

Acknowledgement. This work was started during a chance meeting of researchers from two different groups at a supper table during a previous visit to Mathematisches Forschungsinstitut Oberwolfach, Germany, and the work was essentially completed there.

References

[1] B. Bollobás, *Random Graphs*, 2nd ed. Cambridge Univ. Press, Cambridge, 2001.

[2] B. Bollobás and P. Erdős, Cliques in random graphs. *Math. Proc. Camb. Phil. Soc.* **80** (1976), 419–427.

[3] Z. Izhakian, S. Janson and J. Rhodes, Superboolean rank and the size of the largest triangular submatrix of a random matrix. *Proc. Amer. Math. Soc.*, to appear.

[4] S. Janson, T. Łuczak & A. Ruciński, *Random Graphs.* Wiley, New York, 2000.

[5] D. Matula, The largest clique size in a random graph. Tech. Rep., Dept. Comp. Sci., Southern Methodist Univerity, Dallas, Texas, 1976.

[6] X. Sun and A. B. Nobel, On the size and recovery of submatrices of ones in a random binary matrix. *J. Mach. Learn. Res.* **9** (2008), 2431–2453.