ON CONCENTRATION OF PROBABILITY

SVANTE JANSON

Abstract. We give a survey of several methods to obtain sharp concentration results, typically with exponentially small error probabilities, for random variables occuring in combinatorial probability.

Introduction

In probabilistic combinatorics, it is often important to show that a random variable is sharply concentrated about its mean. A well-known, simple, but rather weak result of this type is Chebyshev's inequality

$$\mathbb{P}(|X - \mathbb{E}\,X| \geq t) \leq \operatorname{Var}(X)/t^2,$$

which holds for any random variable X with finite variance and any $t > 0$. This inequality is very general (and very useful), but is in many cases too weak. The purpose of this paper is to give a survey of some more or less recent stronger inequalities, which under suitable assumptions yield estimates that decrease exponentially as $t \to \infty$.

1. Sums of independent variables

An important case is when the variable X can be written as a sum $\sum_1^n X_i$ of independent random variables. We consider here only the case when each X_i is an indicator variable; thus $X_i \in \operatorname{Be}(p_i)$ where $p_i = \mathbb{P}(X_i = 1) = \mathbb{E}\,X_i$. (The results hold more generally, assuming only that $0 \leq X_i \leq 1$.) Let $\lambda = \mathbb{E}\,X = \sum_1^n p_i$.

Consider first the case of a binomially distributed variable $X \in \operatorname{Bi}(n,p)$; this is of the type above with all $p_i = p$, and thus $\lambda = np$. Applying Markov's inequality to e^{uX}, one finds

$$\mathbb{P}(X \geq \lambda + t) \leq e^{-u(\lambda+t)}\,\mathbb{E}\,e^{uX} = e^{-u(\lambda+t)}(1-p+pe^u)^n, \qquad u \geq 0. \tag{1}$$

The right hand side attains its minimum at $e^u = (\lambda+t)(1-p)/(n-\lambda-t)p$, assuming $0 < \lambda + t < n$, which yields

$$\mathbb{P}(X \geq \mathbb{E}\,X + t) \leq \Big(\frac{\lambda}{\lambda+t}\Big)^{\lambda+t}\Big(\frac{n-\lambda}{n-\lambda-t}\Big)^{n-\lambda-t}, \qquad 0 \leq t \leq n-\lambda; \tag{2}$$

for $t > n - \lambda$ the probability is 0. This bound is implicit in Chernoff [8] and appears explicitly in Okamoto [19]. For applications, it is usually convenient to

---

Date: November 5, 1998; revised August 15, 2000; typo corrected May 25, 2010.

Key words and phrases. Concentration of probability, Azuma's inequality, Chernoff bound, Janson's inequality, Suen's inequality, Talagrand's inequality.

This paper is based on lectures at the workshop "Probabilistic combinatorics" at the Paul Erdős center in Budapest 1998.

replace the right hand side of (2) by a larger but simpler bound. Our favourites are the following, which follow from (2) by some calculus.

Theorem 1. *If* $X \in \mathrm{Bi}(n,p)$ *and* $\lambda = np$*, then*

$$\mathbb{P}(X \geq \mathbb{E}\,X + t) \leq \exp\Bigl(-\frac{t^2}{2(\lambda + t/3)}\Bigr), \qquad t \geq 0; \tag{3}$$

$$\mathbb{P}(X \leq \mathbb{E}\,X - t) \leq \exp\Bigl(-\frac{t^2}{2\lambda}\Bigr), \qquad t \geq 0. \tag{4}$$

Remark. The inequality (3) would not hold in general without the term $t/3$ in the denominator.

We now return to the case when $X_i \in \mathrm{Be}(p_i)$ with different p_i. Using Jensen's inequality, it is easily seen that (1) holds in this case too, with $p = \lambda/n = \sum p_i/n$. Consequently, the following generalization of Theorem 1 holds.

Theorem 2. *If* $X_i \in \mathrm{Be}(p_i)$*,* $i = 1, \dots, n$*, are independent and* $X = \sum_1^n X_i$*, then* (2)*,* (3) *and* (4) *hold, with* $\lambda = \mathbb{E}\,X$*.*

Many other similar bounds, often sharper than (3) and (4), have been proved under these, or more general, assumptions by the same method (which goes back at least to Bernstein [4]); see e.g. [3], [11] and [1, Appendix A].

Remark. Much more precise estimates of the tail probabilities for sums of independent, but not necessarily identically distributed, random variables were obtained by Feller [10] using different methods. Feller's result implies, for the cases above and with $\sigma^2 = \operatorname{Var} X$,

$$\mathbb{P}(X \geq \mathbb{E}\,X + t) = e^{\theta_1 \frac{t}{\sigma^2} x^2}\Bigl(1 - \Phi(x) + \frac{\theta_2}{\sigma} e^{-x^2/2}\Bigr), \qquad 0 < t < \sigma^2/12,$$

where $x = t/\sigma$, Φ is the normal distribution function, $|\theta_1| \leq 6/7(1 - 12t/\sigma^2)$ and $|\theta_2| < 9$.

Very precise asymptotic results for the binomial distribution are given by Littlewood [16].

Some related distributions. The Poisson distribution $\mathrm{Po}(\lambda)$ is a limit of binomial distributions, and thus Theorem 1 implies the same bounds for the Poisson case.

Theorem 3. *If* $X \in \mathrm{Po}(\lambda)$*, then* (3) *and* (4) *hold.*

Furthermore, as a special case of a result in [22], a hypergeometric distribution is the distribution of a certain sum of independent indicator variables. (The proof is algebraic; there is no known probabilistic interpretation of these indicators, which in general have irrational expectations.) Consequently we can apply Theorem 2.

Theorem 4. *Let* X *have a hypergeometric distribution with parameters* N*,* n *and* m*. Then* (3) *and* (4) *hold, with* $\lambda = \mathbb{E}\,X = mn/N$*.*

(This was proved more directly by Hoeffding [11].)

2. Sums of dependent variables

In this section we discuss random variables of the type $X = \sum_{\alpha\in\mathcal{A}} I_\alpha$, where $I_\alpha \in \mathrm{Be}(p_\alpha)$ are indicator variables as above, but now they may be dependent. (However, the inequalities below are useful only in cases when "most" variables are independent.) In order to quantify the dependence, we say that a graph Γ is a *dependency graph* for the family $\{I_\alpha\}_{\alpha\in\mathcal{A}}$ if the vertex set $V(\Gamma) = \mathcal{A}$ and two subfamilies $\{I_\alpha\}_{\alpha\in A}$ and $\{I_\alpha\}_{\alpha\in B}$ are independent whenever A and B are two disjoint subsets of $\mathcal{A}$ such that there is no edge in Γ connecting A and B.

Remark. A slightly weaker definition of dependency graph is commonly used for the Lovasz local lemma [1].

We concentrate for simplicity on bounds for $\mathbb{P}(X = 0)$, i.e., on the probability that $I_\alpha = 0$ for every $\alpha \in \mathcal{A}$. Similar bounds exist for the lower tail probabilities $\mathbb{P}(X \le \mathbb{E}\,X - t)$, $t \ge 0$, but not for the upper tail [12, 13].

A special case, which has had many applications in the last decade, is when there is an underlying family $\{J_j\}_{j\in\mathcal{J}}$ of *independent* random indicator variables, each index α is a (nonempty) subset of $\mathcal{J}$ and $I_\alpha = \prod_{j\in\alpha} J_j$. In other words, if W denotes the random set $\{j \in \mathcal{J} : J_j = 1\}$, then I_α is the indicator of the event $\alpha \subseteq W$, where $\alpha \in \mathcal{A}$, an arbitrary family of nonempty subsets of $\mathcal{J}$. In this case, we define a dependency graph Γ by taking $V(\Gamma) = \mathcal{A}$ and letting the edge set consist of the edges $\alpha\beta$ for all pairs $\alpha, \beta \in \mathcal{A}$ with $\alpha \ne \beta$ and $\alpha \cap \beta \ne \emptyset$.

We fix a dependency graph Γ for $\{I_\alpha\}$ and write $\alpha \sim \beta$ if $\alpha, \beta \in \mathcal{A}$ and there is an edge in Γ between α and β. (In particular, then $\alpha \ne \beta$.) We further write $\alpha \sim \{\beta, \gamma\}$ if $\alpha \sim \beta$ or $\alpha \sim \gamma$, and define

- $\lambda = \mathbb{E}\,X = \sum_{\alpha\in\mathcal{A}} p_\alpha$
- $\Delta = \sum_{\{\alpha,\beta\}:\alpha\sim\beta} \mathbb{E}(I_\alpha I_\beta)$, summing over *unordered* pairs $\{\alpha, \beta\}$, i.e. over the edges in Γ. As a sum over ordered pairs, $\Delta = \frac{1}{2}\sum_{\alpha\in\mathcal{A}}\sum_{\beta\sim\alpha} \mathbb{E}(I_\alpha I_\beta)$.
- $\delta_\alpha = \sum_{\beta\sim\alpha} p_\beta$.
- $\delta = \max_{\alpha\in\mathcal{A}} \delta_\alpha$.
- $\varepsilon = \max_{\alpha\in\mathcal{A}} p_\alpha$.

In the special case we thus have $\Delta = \frac{1}{2}\sum\sum_{\alpha\ne\beta,\,\alpha\cap\beta\ne\emptyset} \mathbb{E}(I_\alpha I_\beta)$. Then, the following bounds hold [14].

Theorem 5. *In the special case, with $X = \sum I_A$, $\lambda = \mathbb{E}\,X$ and Δ as above,*

$$\mathbb{P}(X = 0) \le \exp(-\lambda + \Delta); \tag{5}$$

$$\mathbb{P}(X = 0) \le \exp\Bigl(-\frac{\lambda^2}{\lambda + 2\Delta}\Bigr) = \exp\left(-\frac{\lambda^2}{\sum\sum_{\alpha\cap\beta\ne\emptyset} \mathbb{E}(I_\alpha I_\beta)}\right). \tag{6}$$

The first bound is best for $\Delta \le \lambda/2$, while the second is better for larger Δ.

Remark. Boppana and Spencer [6] proved a slightly different version of (5), viz.

$$\mathbb{P}(X = 0) \le e^{\Delta/(1-\varepsilon)} \prod_{\alpha\in\mathcal{A}} (1 - p_\alpha), \tag{7}$$

which has the advantage of being exact when $\Delta = 0$ (i.e. when the variables I_α are independent), but otherwise seems slightly less convenient for applications. Although the upper bounds (5) and (7) are quite close when $\varepsilon = \max p_\alpha$ is small, neither of them dominates the other. It is intriguing to note that the conceivable common improvement $e^\Delta \prod_\alpha (1 - p_\alpha)$ fails to be an upper bound to $\mathbb{P}(X = 0)$; this is seen by the simple example where $X = I_1 + I_2$ with $I_1 = I_2 \in \mathrm{Be}(p)$, for which $\Delta = p$ and $\mathbb{P}(X = 0) = 1 - p > e^p(1 - p)^2$, cf. [13].

Note also that in the special case, by the FKG inequality, we have the lower bound

$$\mathbb{P}(X = 0) \geq \prod_{A \in \mathcal{A}} (1 - p_A) \geq \exp\Big(-\frac{\lambda}{1 - \varepsilon}\Big),$$

which shows that the bounds above are quite sharp when Δ and ε are small.

The inequalities (5)–(7) have been widely used, although they apply only to the sum of indicator variables with a very special structure. For example, they apply to the number of copies of a given graph in $\mathbb{G}(n, p)$, but they do not apply to the number of induced copies.

In a paper which has too long been neglected, Suen [20] proved, at about the same time as [14, 6] and independently of them, a similar inequality for the general case. The Suen inequality and later versions of it [13] yield almost as good bounds as (5)–(7), but are thus much more widely applicable.

We do not give Suen's original inequality [20] here, but rather a slight improvement of it (8) followed by two consequences of it [13], which are more convenient for applications and closely resemble (5) and (6) for the special case above.

Theorem 6. *Let $X = \sum_\alpha I_\alpha$, where $\{I_\alpha\}_{\alpha \in \mathcal{A}}$ is a finite family of Bernoulli random variables having a dependency graph Γ. Then, with the notation above,*

$$\mathbb{P}(X = 0) \leq \exp\Big(\sum_{\{\alpha,\beta\}:\alpha\sim\beta} \mathbb{E}(I_\alpha I_\beta) \prod_{\gamma\sim\{\alpha,\beta\}} (1 - p_\gamma)^{-1} \Big) \prod_{\nu \in \mathcal{A}} (1 - p_\nu); \tag{8}$$

$$\mathbb{P}(X = 0) \leq \exp(-\lambda + \Delta e^{2\delta}); \tag{9}$$

$$\mathbb{P}(X = 0) \leq \exp\left(- \min\Big(\frac{\lambda^2}{8\Delta}, \frac{\lambda}{2}, \frac{\lambda}{6\delta}\Big)\right) = e^{-\lambda^2/\max(8\Delta, 2\lambda, 6\delta\lambda)}. \tag{10}$$

The bounds (9) and (10) are similar to the ones in (5) and (6), but with extra terms involving δ and somewhat worse constants in (10). In applications, δ is usually small and the bounds (9) and (10) are often as useful as (5) and (6). Nevertheless, in at least one application (a recent bound for the 3-SAT problem by Kirousis, Kranakis, Krizanc and Stamatiou [15], where $\delta \approx 0.089$), the extra factor $e^{2\delta}$ in (9) affects the final result significantly, and it is desirable to reduce it as much as possible. It is shown in [13], using a proof by Spencer (personal communication), that if $0 \leq \delta + \varepsilon$, this factor $e^{2\delta}$ can be replaced by the smallest root φ of $\varphi = e^{(\delta+\varepsilon)\varphi}$; this is the version used in [15].

It is an open problem whether the factor $e^{2\delta}$ in (9) can be eliminated completely, i.e. whether (5) (or (6), (7)) holds in the general case too.

3. Azuma's inequality

The final inequalitites that we consider apply to random variables of the form $X = f(Z_1, \dots, Z_N)$, where $Z_1, \dots, Z_N$ are independent random variables; here the underlying random variables Z_k can take values in arbitrary (and possibly different) spaces Λ_k. Our basic assumption is the following Lipschitz condition, for some numbers $c_1, \dots, c_N$. (In most applications, $c_k = 1$.)

(L) If the vectors $z, z' \in \prod_1^N \Lambda_i$ differ only in the kth coordinate, then $|f(z) - f(z')| \le c_k$, $k = 1, \dots, N$.

In combinatorial applications, one often considers a random subset W of a set $A = \{a_1, \dots, a_N\}$, and sets $Z_k = \mathbf{1}[a_k \in W]$, the indicator variable of the event $\{a_k \in W\}$.

More generally, we may partition the ground set A into blocks B_k and let Z_k be the vector of the indicator variables $(\mathbf{1}[a_i \in W])_{i \in B_k}$ (thus Z_k takes values in the product space $\{0,1\}^{B_k}$). A common instance of this construction is "vertex exposure" for a random graph $\mathbb{G}(n,p)$, where $A = \{\{i,j\} : 1 \le i < j \le n\}$ and $B_k = \{\{i,k\} : 1 \le i < k\}$, $k = 1, \dots, n$.

The first result for this situation is based on the following martingale inequality, known as Azuma's inequality [2], although it was earlier proved by Hoeffding [11].

Theorem 7. *If $(X_k)_0^N$ is a martingale, and there exist constants $c_k \ge 0$ such that $|X_k - X_{k-1}| \le c_k$, $k = 1, \dots, N$, then*

$$\mathbb{P}(X_N \ge X_0 + t) \le \exp\Bigl(-t^2/2\sum_1^N c_k^2\Bigr), \qquad t \ge 0.$$

For applications to our situation, let $X_k = \mathbb{E}(X \mid Z_1, \dots, Z_k)$; then the assumptions of Theorem 7 are satisfied, and thus

$$\mathbb{P}(X \ge \mathbb{E}\, X + t) \le \exp\bigl(-t^2/2\sum_1^N c_k^2\bigr).$$

In fact, the exponent here (but not in Theorem 7) may be improved by a factor 4 [17].

Theorem 8. *Let $Z_1, \dots, Z_N$ be independent random variables, with Z_k taking values in a set Λ_k. Assume that a (measurable) function $f : \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_N \to \mathbb{R}$ satisfies the Lipschitz condition* (L). *Then, the random variable $X = f(Z_1, \dots, Z_N)$ satisfies, for any $t \ge 0$,*

$$\mathbb{P}(X \ge \mathbb{E}\, X + t) \le \exp\Bigl(-2t^2 \Bigm/ \sum_1^N c_k^2\Bigr), \tag{11}$$

$$\mathbb{P}(X \le \mathbb{E}\, X - t) \le \exp\Bigl(-2t^2 \Bigm/ \sum_1^N c_k^2\Bigr). \tag{12}$$

For further related results, including an extension to unbounded martingale differences, see the survey by Bollobás [5]. For another extension (weakening (L) by a truncation), see [7].

4. Talagrand's inequality

Talagrand [21] has given several inequalities on concentration of measure, with various applications. We will here treat one of his inequalities only; it yields results similar to those obtained by Azuma's inequality, although often much stronger, and it has already had important combinatorial applications.

Remark. The interested reader should study also the other inequalities by Talagrand in [21] and subsequent papers; it seems reasonable to guess that some of them too will be useful in combinatorial applications. Moreover, there are other proofs of the results below, with extensions in different directions, by Marton [18] and Dembo [9].

We continue with the assumptions of the preceding section, in particular condition (L), and assume furthermore the following, for some function ψ and the same c_k as in (L).

(C) If $z = (z_k)_1^N \in \prod_1^N \Lambda_k$ and $r \in \mathbb{R}$ with $f(z) \geq r$, then there exists a set $J \subseteq \{1, \dots, N\}$ with $\sum_{k \in J} c_k^2 \leq \psi(r)$, such that for all $y = (y_k)_1^N \in \prod_1^N \Lambda_k$ with $y_k = z_k$ when $k \in J$, we have $f(y) \geq r$.

In other words, there exists a vector $(z_k)_{k \in J}$ (called a *certificate*), which forces $f \geq r$, such that the index set J is not too large. (The set J generally depends on z and r.)

Every function f trivially satisfies (C) with $\psi(r) = \sum_1^N c_k^2$ for all r; just take $J = \{1, \dots, N\}$. This choice of ψ in the theorem below yields an inequality similar to the one in Theorem 8, but generally somewhat weaker. In many interesting applications, however, (C) holds with a much smaller ψ; this leads to stronger estimates that significantly surpass Azuma's inequality. (It is immaterial that we obtain inequalities for the deviation from the median instead of the mean.)

Example. Let X be the order of the largest independent set in the random graph $\mathbb{G}(n, p)$. Then, using vertex exposure as described above, (L) holds with $c_k = 1$, and (C) holds with $\psi(r) = \lceil r \rceil$ for $r \geq 0$; the certificate is just any independent set of order $\lceil r \rceil$.

Talagrand's inequality may then be stated as follows.

Theorem 9. *Suppose that $Z_1, \dots, Z_N$ are independent random variables taking their values in some sets $\Lambda_1, \dots, \Lambda_N$, respectively. Suppose further that $X = f(Z_1, \dots, Z_N)$, where $f : \Lambda_1 \times \cdots \times \Lambda_N \to \mathbb{R}$ is a (measurable) function such that (L) and (C) hold for some constants c_k, $k = 1, \dots, N$, and some function ψ. Then, for every $r \in \mathbb{R}$ and $t \geq 0$,*

$$\mathbb{P}(X \leq r - t)\, \mathbb{P}(X \geq r) \leq e^{-t^2/4\psi(r)}. \tag{13}$$

In particular, if m is a median of X, then for every $t \geq 0$,

$$\mathbb{P}(X \leq m - t) \leq 2e^{-t^2/4\psi(m)} \tag{14}$$

and

$$\mathbb{P}(X \geq m + t) \leq 2e^{-t^2/4\psi(m+t)}. \tag{15}$$

Actually, the theorem above is a simple corollary of the following more general inequality due to Talagrand [21].

Theorem 10. *Suppose that A and B are two (measurable) subsets of $\prod_1^N \Lambda_k$ such that for some $t \geq 0$ the following separation condition holds: for every $z \in B$, there exists a non-zero vector $\alpha = (\alpha_i)_1^N \in \mathbb{R}^N$ such that for every $y \in A$,*

$$\sum_{i:y_i \neq z_i} \alpha_i \geq t\Big(\sum_1^N \alpha_i^2\Big)^{1/2}.$$

Then

$$\mathbb{P}(A)\,\mathbb{P}(B) \leq e^{-t^2/4}.$$

Remark. Talagrand [21, Corollary 4.2.5] further showed that the conclusion of Theorem 10 can be improved to

$$\sqrt{\log(1/\,\mathbb{P}(A))} + \sqrt{\log(1/\,\mathbb{P}(B))} \geq t/\sqrt{2},$$

which implies for example that (14) can be improved to the smaller, but more complicated, bound

$$\mathbb{P}(X \leq m - t) \leq \exp\Big(-\frac{1}{2\psi(m)}\Big(t - \sqrt{2\log 2\,\psi(m)}\Big)^2\Big), \quad t \geq \sqrt{2\log 2\,\psi(m)}.$$

Acknowledgement

This survey is based on a chapter in a forthcoming book on random graphs that I am writing together with Tomasz Łuczak and Andrzej Ruciński. I thank these my coauthors for helpful discussions, which have contributed to this paper too.

References

[1] N. Alon & J. Spencer, *The Probabilistic Method.* Wiley, New York 1992.

[2] K. Azuma, Weighted sums of certain dependent variables. *Tôhoku Math. J.* **3** (1967), 357–367.

[3] G. Bennett, Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** (1962), 33–45.

[4] S. Bernstein, On a modification of Chebyshev's inequality and of the error formula of Laplace. *Ann. Sci. Inst. Savantes Ukraine, Sect. Math.* **1** (1924), 38–49. (Russian)

[5] B. Bollobás, Martingales, isoperimetric inequalities and random graphs. In *Combinatorics* (Eger 1987), Colloq. Math. Soc. János Bolyai, 52, North-Holland, Amsterdam 1988, 113–139.

[6] R. Boppana & J. Spencer, A useful elementary correlation inequality. *J. Combin. Th. Ser. A* **50** (1989), 305–307.

[7] T.K. Chalker, A.P. Godbole, P. Hitczenko, J. Radcliff & O.G. Ruehr, On the size of a random sphere of influence graph. *Adv. Appl. Probab.* **31** (1999), 596–609.

[8] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** (1952), 493–507.

[9] A. Dembo, Information inequalities and concentration of measure. *Ann. Probab.* **25** (1997), 927–939.

[10] W. Feller, Generalization of a probability limit theorem of Cramér. *Trans. Amer. Math. Soc.* **54** (1943), 361–372.

[11] W. Hoeffding, Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** (1963), 13–30.

[12] S. Janson, Poisson approximation for large deviations. *Random Struct. Alg.* **1** (1990), 221–229.

[13] S. Janson, New versions of Suen's correlation inequality. *Random Struct. Alg.* **13** (1998), 467–483.

[14] S. Janson, T. Łuczak & A. Ruciński, An exponential bound for the probability of nonexistence of a specified subgraph in a random graph. In *Random Graphs '87* (Poznań 1987), Wiley, Chichester 1990, 73–87.

[15] L.M. Kirousis, E. Kranakis, D. Krizanc & Y.C. Stamatiou, Approximating the unsatisfiability threshold of random formulas. *Random Struct. Alg.* **12** (1998), 253–269.

[16] J. E. Littlewood, On the probability in the tail of a binomial distribution. *Adv. Appl. Probab.* **1** (1969), 43–72.

[17] C. McDiarmid, On the method of bounded differences. In *Surveys in Combinatorics* (Norwich 1989), London Math. Soc. Lect. Not. 141, Cambridge Univ. Press, Cambridge 1989, 148–188.

[18] K. Marton, A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.* **6** (1996), 556–571. Erratum, *Geom. Funct. Anal.* **7** (1997), 609–613.

[19] M. Okamoto, Some inequalities relating to the partial sum of binomial probabilities. *Ann. Inst. Statist. Math.* **10** (1958), 29–35.

[20] W.C.S. Suen, A correlation inequality and a Poisson limit theorem for nonoverlapping balanced subgraphs of a random graph. *Random Struct. Alg.* **1** (1990), 231–242.

[21] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.* **81** (1995), 73–205.

[22] V.A. Vatutin & V.G. Mikhaĭlov, Limit theorems for the number of empty cells in an equiprobable scheme for group allocation of particles. *Teor. Veroyatnost. i Primenen.* **27** (1982), 684–692 (Russian); English transl. *Theor. Probab. Appl.* **27** (1982), 734–743.

Department of Mathematics, Uppsala University, PO Box 480, S-751 06 Uppsala, Sweden

E-mail address: `svante.janson@math.uu.se`