Discrete Mathematics 87 (1991) 1–8
North-Holland 

# Occupancy problems and random algebras

Michael H. Albert and Alan M. Frieze
*Department of Mathematics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

Received 22 April 1988
Revised 26 January 1989

*Abstract*

Albert, M.H. and A.M. Frieze, Occupancy problems and random algebras, Discrete Mathematics 87 (1991) 1–8.

For $k$ randomly chosen subsets of $[n]=\{1, 2, \ldots, n\}$ we consider the probability that the Boolean algebra, distributive lattice, and meet semilattice which they generate are respectively free, or all of $2^{[n]}$. In each case we describe a threshold function for the occurrence of these events. The threshold functions for freeness are close to their theoretical maximum values.

## 1. Introduction

In this paper we consider various algebras generated by $k$ randomly chosen subsets of $[n]=\{1, 2, \ldots, n\}$. As in the study of random graphs (Erdős and Rényi [2], Bollobás [1]) we focus on the threshold for the occurrence of various events.

To be specific consider $\mathcal{P}_n = 2^{[n]}$ to be a probability space in which each subset of $[n]$ has the same probability $2^{-n}$. Now select $A_1, A_2, \ldots, A_k$ independently and randomly from $\mathcal{P}_n$ (*with replacement*). Let $\underset{\sim}{A}^{(k)}$ denote $A_1, A_2, \ldots, A_k$.

We consider

(i) $\mathcal{B}(\underset{\sim}{A}^{(k)})$ = the Boolean subalgebra of $\mathcal{P}_n$ generated by $\underset{\sim}{A}^{(k)}$,
(ii) $\mathcal{D}(\underset{\sim}{A}^{(k)})$ = the distributive sublattice of $\mathcal{P}_n$ generated by $\underset{\sim}{A}^{(k)}$,
(iii) $\mu(\underset{\sim}{A}^{(k)})$ = the meet sub-semi-lattice of $\mathcal{P}_n$ generated by $\underset{\sim}{A}^{(k)}$.

In each case we determine the asymptotic probability that the algebras generated are (a) freely generated by $\underset{\sim}{A}^{(k)}$, or, (b) the whole of $\mathcal{P}_n$.

For example, to say that $\underset{\sim}{A}^{(k)}$ freely generates $\mathcal{B}(\underset{\sim}{A}^{(k)})$ means that for any two Boolean polynomials $p$ and $q$ in variables $x_1, x_2, \ldots, x_k$ if

$$p(A_1, A_2, \ldots, A_k) = q(A_1, A_2, \ldots, A_k)$$

then $p(x_1, x_2, \ldots, x_k) = q(x_1, x_2, \ldots, x_k)$ is an identity true in all Boolean algebras. Since there is in fact a normal form for Boolean polynomials it is equivalent to demand that polynomials in $k$ variables with distinct normal forms evaluate at $A_1, A_2, \ldots, A_k$ to distinct subsets of $[n]$. Similar criteria apply to $\mathscr{D}(\underset{\sim}{A}^{(k)})$ and $\mu(\underset{\sim}{A}^{(k)})$.

We prove the following.

**Theorem.** (a) *Let* $\varepsilon > 0$ *be fixed and let* $\kappa = \log_2 n - \log_2 \log_e n + \log_2 \log_e 2$. *Then*

$$\lim_{n\to\infty} P(\mathscr{B}(\underset{\sim}{A}^{(k)}) \text{ is freely generated by } \underset{\sim}{A}^{(k)}) = 1 \quad \text{for } k \leqslant \kappa - \varepsilon,$$

$$\lim_{n\to\infty} P(\mathscr{B}(\underset{\sim}{A}^{(k)}) \text{ is freely generated by } \underset{\sim}{A}^{(k)}) = 0 \quad \text{for } k \geqslant \kappa + \varepsilon.$$

(b) *Let* $k = 2\log_2 n + a_n$. *Then*

$$\lim_{n\to\infty} P(\mathscr{B}(\underset{\sim}{A}^{(k)}) \text{ is all of } \mathscr{P}_n) = \begin{cases} 0 & a_n \to -\infty \\ e^{-2^{-(a+1)}} & a_n \to a, \\ 1 & a_n \to +\infty. \end{cases}$$

(c)

$$\lim_{n\to\infty} P(\mathscr{D}(\underset{\sim}{A}^{(k)}) \text{ is freely generated by } \underset{\sim}{A}^{(k)})$$
$$= \lim_{n\to\infty} P(\mathscr{B}(\underset{\sim}{A}^{(k)}) \text{ is freely generated by } \underset{\sim}{A}^{(k)})$$

(d) *Under the assumptions on* $k$, $a_n$ *of part* (b),

$$\lim_{n\to\infty} P(\mathscr{D}(\underset{\sim}{A}^{(k)}) = \mathscr{P}_n) = \lim_{n\to\infty} P(\mathscr{B}(\underset{\sim}{A}^{(k)}) = \mathscr{P}_n).$$

(e) *Let* $k = \log_2 n - \log_2(\log_e \log_2 n + b_n)$. *Then*

$$\lim_{n\to\infty} P(\mu(\underset{\sim}{A}^{(k)}) \text{ is freely generated by } \underset{\sim}{A}^{(k)}) = \begin{cases} 0 & b_n \to -\infty \\ e^{-e^{-b}} & b_n \to b, \\ 1 & b_n \to +\infty. \end{cases}$$

(f) *We deviate from our probabilistic model by assuming the k sets are chosen* without *replacement. Let now* $k = 2^n(1 - c_n/n)$ *where* $c_n \geqslant 0$. *Then*

$$\lim_{n\to\infty} P(\mu(\underset{\sim}{A}^{(k)}) = \mathscr{P}_n) = \begin{cases} 0 & c_n \to \infty, \\ e^{-c} & c_n \to c, \\ 1 & c_n \to 0. \end{cases}$$

## 2. Preliminaries

For $S \subseteq [k]$ we define

$$A_S = \bigcap_{i\in S} A_i \cap \bigcap_{i\notin S} \bar{A}_i, \quad \text{where } \bar{A}_i = [n] \setminus A_i$$

and note that the sets $A_S$, $S \subseteq [k]$ partition $[n]$. Thus, in particular:

$$A_\emptyset = \bigcap_{i\in[k]} \bar{A}_i \quad \text{and} \quad A_{[k]} = \bigcap_{i\in[k]} A_i.$$

It is useful to consider the $k \times n$ 0–1 matrix $X = \|x_{ij}\|$ where $x_{ij} = 1\,(0)$ whenever $j \in A_i$ ($j \notin A_i$). Our probability assumption is equivalent to

$$x_{11}, x_{12}, \ldots, x_{kn} \text{ form a sequence of independent Bernoulli random variables where for all } i, j\ P(x_{ij}=0) = P(x_{ij}=1) = \tfrac{1}{2}. \tag{2.1}$$

Now let $S_j = \{i \in [k]\colon j \in A_i\}$. It follows from (2.1) that:

$$P(S_j = S) = 2^{-k} \quad \text{for all } S \subseteq [k]. \tag{2.2}$$

$$\text{The random variables } S_1, S_2, \ldots, S_n \text{ are independent.} \tag{2.3}$$

Now we can view the construction of $A_1, A_2, \ldots, A_k$ as the construction of $S_1, S_2, \ldots, S_n$. Then, since $j \in A_{S_j}$, we have the following situation.

We have $m = 2^k$ boxes each labelled by a distinct subset of $[k]$. We have distinct balls labelled $1, 2, \ldots, n$ which are independently placed randomly into boxes. (We keep $m = 2^k$ throughout the paper.)

Placing $j$ into box $S$ is to be interpreted as putting $S_j = S$.

We refer to this as the Balls-in-Boxes construction and use $P_{\text{BB}}$ to refer to probabilities defined on this space.

It follows from (2.2) and (2.3) that in this space we determine a matrix $X$ with the same distribution as in (2.1).

## 3. Boolean algebras

Let us now consider $\mathcal{B}(\underset{\sim}{A}^{(k)})$. We have the following simple result.

**Proposition 3.1.** *$\mathcal{B}(\underset{\sim}{A}^{(k)})$ is freely generated by $\underset{\sim}{A}^{(k)}$ if and only if $A_S \neq \emptyset$ for all $S \subseteq [k]$.*

**Proof.** If $A_S = \emptyset$ for some $S \subseteq [k]$ then clearly $\mathcal{B}(\underset{\sim}{A}^{(k)})$ is not free. Conversely, suppose $\mathcal{B}(\underset{\sim}{A}^{(k)})$ is not free. Then there exist $S, T \subseteq [k]$, $S \cap T = \emptyset$ such that $\emptyset = \bigcap_{i\in S} A_i \cap \bigcap_{i\in T} \bar{A}_i \supseteq A_S$. □

It follows from Section 2 and Proposition 3.1 that

$$P(\mathcal{B}(\underset{\sim}{A}^{(k)}) \text{ is freely generated by } \underset{\sim}{A}^{(k)}) = P_{\text{BB}}(\text{each box is non-empty}).$$

Now the latter probability has been studied under the guise of the Coupon Collector Problem (Feller [3]).

Assuming $k = k(n)$ let $d(n) = (n - m \log_e m)/m$. (Recall $m = 2^k$.) It is well known that

$$\lim_{n\to\infty} P_{\mathrm{BB}}(\text{each box is non-empty}) = \begin{cases} 0 & d(n)\to -\infty, \\ \mathrm{e}^{-\mathrm{e}^{-d}} & d(n)\to d, \\ 1 & d(n)\to +\infty. \end{cases}$$

Thus if $z$ satisfies $n = 2^z z \log_e 2$ and $\varepsilon > 0$ is fixed then

$$k \leqslant z - \varepsilon \;\Rightarrow\; P(\mathscr{B}(\underset{\sim}{A}^{(k)}) \text{ is freely generated by } \underset{\sim}{A}^{(k)}) \to 1,$$

$$k \geqslant z + \varepsilon \;\Rightarrow\; P(\mathscr{B}(\underset{\sim}{A}^{(k)}) \text{ is freely generated by } \underset{\sim}{A}^{(k)}) \to 0.$$

Since $z = (\log_e n - \log_e \log_e n + \log_e \log_e 2)/\log_e 2 + \mathrm{o}(1)$ we have part (a) of the Theorem.

Another simple remark.

**Proposition 3.2.** *$\mathscr{B}(\underset{\sim}{A}^{(k)})$ is all of $\mathscr{P}_n$ if and only if $|A_S| \leqslant 1$ for all $S \subseteq [k]$.*

**Proof.** $\mathscr{B}(\underset{\sim}{A}^{(k)})$ is all of $\mathscr{P}_n$ if and only if there exist $S_j$, $T_j$, $j = 1, 2, \ldots, n$ such that $\{j\} = \bigcap_{i\in S_j} A_i \cap \bigcap_{i\in T_j} \bar{A}_i$. This implies the proposition. □

Thus

$$P(\mathscr{B}(\underset{\sim}{A}^{(k)}) \text{ is all of } \mathscr{P}_n) = P_{\mathrm{BB}}(\text{each box contains at most one ball}).$$

We now prove part (b) of the Theorem.

Let $z_t$ = the number of boxes containing exactly $t$ balls. Let $k = 2\log_2 n + a_n$ so that $m = 2^{a_n} n^2$.

*Case* 1: $a_n \to \infty$.

$$E_{\mathrm{BB}}\Big(\sum_{t=2}^{n} z_t\Big) \leqslant m\binom{n}{2}\Big(\frac{1}{m}\Big)^2 \leqslant 2^{-(a_n+1)} \to 0.$$

*Case* 2: $a_n \to a$.

Observe first that

$$E_{\mathrm{BB}}\Big(\sum_{t=3}^{n} z_t\Big) \leqslant m\binom{n}{3}\Big(\frac{1}{m}\Big)^3 = \mathrm{O}(n^{-1})$$

and so $P_{\mathrm{BB}}(\sum_{t=3}^{n} z_t > 0) = \mathrm{o}(1)$. Thus we only have to show that

$$\lim_{n\to\infty} P_{\mathrm{BB}}(z_2 = 0) = \mathrm{e}^{-\lambda} \quad \text{where } \lambda = 2^{-(a+1)}.$$

Let $r \geqslant 0$ be a fixed integer. We show

$$\lim_{n\to\infty} E_{\mathrm{BB}}((z_2)_r) = \lambda^r.$$

It follows (see e.g. Bollobás [1, Theorem I.20]) that $z_2$ is asymptotically Poisson with mean $\lambda$. This will complete this case.

Now

$$E_{BB}((z_2)_r) = (m)_r \binom{n}{2r} \frac{(2r)!}{2^r} \left(\frac{1}{m}\right)^{2r} \left(1 - \frac{r}{m}\right)^{n-2r} \approx \lambda^r$$

and we are done.

*Case* 3: $a_n \to -\infty$.

This follows from Case 2 by a simple monotonicity argument. (Ultimately we are throwing $n$ balls into more boxes than the case of any fixed $a$.)

## 4. Distributive lattices

Let us now consider $\mathscr{D}(\underset{\sim}{A}^{(k)})$. We have the following:

**Proposition 4.1.** *$\mathscr{D}(\underset{\sim}{A}^{(k)})$ is freely generated by $\underset{\sim}{A}^{(k)}$ if and only if $A_S \neq \emptyset$ for $\emptyset \neq S \subseteq [k]$.*

**Proof.** Assume $\mathscr{D}(\underset{\sim}{A}^{(k)})$ is freely generated by $\underset{\sim}{A}^{(k)}$ and $\emptyset \neq S \subseteq [n]$. Now the two sets

$$C = \bigcap_{i \in S} A_i, \qquad D = C \cap \bigcup_{j \notin S} A_j$$

must be distinct. That is, there exists an element belonging to $\bigcap_{i \in S} A_i$ but not to any $A_j$, for $j \notin S$. Put another way, $A_S \neq \emptyset$.

Conversely, given any two distributive lattice polynomials in $k$ variables which have different disjunctive normal forms, then their symmetric difference (as a Boolean polynomial) contains a term with at least one positive instance of a variable. Thus if $A_S \neq \emptyset$ for $S \neq \emptyset$, the sets obtained by evaluating these polynomials are distinct and hence $\mathscr{D}(\underset{\sim}{A}^{(k)})$ is freely generated by $\underset{\sim}{A}^{(k)}$. □

It follows from Section 2 and Proposition 4.1 that

$$\begin{aligned} P(\mathscr{D}(\underset{\sim}{A}^{(k)}) \text{ is freely generated by } \underset{\sim}{A}^{(k)}) &= P_{BB}(\text{box } S \text{ is non-empty, } \forall S \neq \emptyset) \\ &= P_{BB}(\text{box } S \text{ is non-empty, } \forall S) + P_{BB}(\text{box } \emptyset \text{ is the only empty box}). \end{aligned}$$

Thus, by Proposition 3.1, in order to prove (c) we need only show that

$$\lim_{n \to \infty} P_{BB}(\text{box } \emptyset \text{ is the only empty box}) = 0 \quad \text{for all } k \geq 0.$$

But

$$P_{BB}(\text{box } \emptyset \text{ is the only empty box}) = \left(1 - \frac{1}{m}\right)^n P_{BB}(\text{box } S \text{ is non-empty } \forall S \neq \emptyset \mid \text{box } \emptyset \text{ is empty}). \tag{4.1}$$

Moreover, since there are $m$ boxes, by symmetry we obtain:

$$P_{BB}(\text{box } \emptyset \text{ is the only empty box}) \leq \frac{1}{m}.$$

Therefore

$$P_{BB}(\text{box } \emptyset \text{ is the only empty box}) \leq \max\left(\left(1 - \frac{1}{m}\right)^n, \frac{1}{m}\right).$$

Fix $\varepsilon > 0$, If $n > 1/\varepsilon^2$ then either

$$\frac{n}{m} > \frac{1}{\varepsilon}, \quad \text{or} \quad m > \frac{1}{\varepsilon}.$$

In the first case $1/m > 1/\varepsilon n$ and hence

$$\left(1 - \frac{1}{m}\right)^n < \left(1 - \frac{1}{\varepsilon n}\right)^n < e^{-1/\varepsilon} < \varepsilon$$

while in the second case $1/m < \varepsilon$. Hence, regardless of the values of $k$,

$$\lim_{n\to\infty} P_{\mathrm{BB}}(\text{box } \emptyset \text{ is the only empty box}) = 0.$$

This completes the proof of (c). Now to part (d) of the theorem.

**Proposition 4.2.** *$\mathscr{D}(\underline{A}^{(k)}) = \mathscr{P}_n$ if and only if $A_\emptyset = \emptyset$ and $|A_S| \leq 1$ for all $S \neq \emptyset$.*

**Proof.** Clearly $\mathscr{D}(\underline{A}^{(k)}) = \mathscr{P}_n$ if and only if

$$\{j\} = \bigcap_{j\in A_i} A_i \quad \text{for all } j \in [n],$$

or equivalently

$$\forall j \in [n] \quad \{j\} = A_{\{i:\, j\in A_i\}} \quad \text{and} \quad \{i : j \in A_i\} \neq \emptyset.$$

As the sets $A_S$ partition $[n]$ this condition is realised if and only if

$$A_\emptyset = \emptyset \quad \text{and} \quad |A_S| \leq 1 \quad \text{for } S \neq \emptyset. \qquad \square$$

Hence

$$\begin{aligned} P(\mathscr{D}(\underline{A}^{(k)}) = \mathscr{P}_n) &= P_{\mathrm{BB}}(A_\emptyset = \emptyset \text{ and } |A_S| \leq 1 \text{ for } S \neq \emptyset) \\ &= P_{\mathrm{BB}}(|A_S| \leq 1,\ \forall S) - P_{\mathrm{BB}}(|A_S| \leq 1, \text{ for } S \neq \emptyset \mid |A_\emptyset| = 1) P(|A_\emptyset| = 1). \end{aligned} \tag{4.2}$$

Now $P(|A_\emptyset| = 1) = (n/m)(1 - 1/m)^{n-1}$ and this tends to zero if $n/m \to 0$ or $\infty$. But if $n/m \to c > 0$ then the conditional probability in (4.2) goes to zero in view of (b). This completes the proof of (d).

## 5. Semi-lattices

A semi-lattice is simply a set together with a single idempotent, associative, and commutative operation. In $\mathscr{P}_n$ we take this operation to be intersection, hence $\mu(A^{(k)})$ is simply the smallest subset of $\mathscr{P}_n$ containing $A^{(k)}$ and closed under intersection.

We now consider $\mu(\underset{\sim}{A}^{(k)})$. We have the following:

**Proposition 5.1.** *$\mu(\underset{\sim}{A}^{(k)})$ is freely generated by $\underset{\sim}{A}^{(k)}$ if and only if $A_{[n]-\{j\}} \neq \emptyset$ for all $j \in [n]$.*

**Proof.** The covering pairs in a free semilattice generated by $x_1, x_2, \ldots, x_k$ are exactly those pairs

$$\bigwedge_{i \in I} x_i > \Big(\bigwedge_{i \in I} x_i\Big) \wedge x_j \quad \emptyset \neq I \subseteq [k],\ j \notin I.$$

So the semilattice $\mu(\underset{\sim}{A}^{(k)})$ is freely generated by $\underset{\sim}{A}^{(k)}$ if and only if for $\emptyset \neq I \subseteq [k]$, $j \notin I$

$$\bigcap_{i \in I} A_i \nsubseteq A_j.$$

For this to be true, it is necessary and sufficient that

$$\bigcap_{i \in [k]-\{j\}} A_i \nsubseteq A_j \quad \text{for } j \in [k]$$

which is equivalent to the statement in the proposition. $\square$

Thus

$$P(\mu(\underset{\sim}{A}^{(k)}) \text{ is freely generated by } \underset{\sim}{A}^{(k)}) = P_{\mathrm{BB}}(A_{[k]-\{j\}} \neq \emptyset \text{ for } j \in [k]). \tag{5.1}$$

Now for $T \subseteq [k]$, $|T| = t$ we have

$$P(A_{[k]-\{j\}} = \emptyset \text{ for } j \in T) = \Big(1 - \frac{t}{m}\Big)^n. \tag{5.2}$$

Recall that $k = \log_2 n - \log_2(\log_e \log_2 n + b_n)$.

*Case* 1: $b_n \to +\infty$.

(5.1) and (5.2) imply

$$P(\mu(\underset{\sim}{A}^{(k)}) \text{ is not free}) \leq k\Big(1 - \frac{1}{m}\Big)^n \leq k\, \mathrm{e}^{-n/m}$$
$$\leq k\, \mathrm{e}^{-b_n}/\log_2 n = \mathrm{o}(1).$$

*Case* 2: $b_n \to b$.

Let $Z$ = the number of boxes $[k] - \{j\}$ which are empty, and let $\tau = \mathrm{e}^{-b}$ and let $r \geq 1$ be a fixed integer. Proceeding as in Case 2 of (b) we prove

$$\lim_{n \to \infty} E_{\mathrm{BB}}((Z)_r) = \tau^r$$

and we are done. Now

$$E_{\mathrm{BB}}((Z)_r) = (k)_r\Big(1 - \frac{r}{m}\Big)^n \approx k^r\, \mathrm{e}^{-rn/m} \approx k^r(\log_2 n)^{-r}\, \mathrm{e}^{-rb}$$

completing the proof of this case.

*Case* 3: $b_n \to -\infty$.
Use a monotonicity argument as in Case 3 of (b).

An element $x$ of a semilattice $(S, \wedge)$ is called *meet-irreducible* if $x = y \wedge z$ implies $x = y$ or $x = z$.

**Proposition 5.2.** *$\mu(\underline{A}^{(k)}) = \mathcal{P}_n$ if and only if*

$$\{[n]\} \cup \{[n] - \{j\} : j \in [n]\} \subseteq \{A_i : i \in [k]\}.$$

**Proof.** The sets $[n]$ and $[n] - \{j\}$, $j \in [k]$, are the meet-irreducibles of $\mathcal{P}_n$, which must be contained in any set which generates $\mathcal{P}_n$ as a meet-semilattice. □

Suppose now that we choose $k = 2^n(1 - c_n/n)$ sets *without* replacement. Let $N = 2^n$. It follows from Proposition 5.2 that

$$P(\mu(\underline{A}^{(k)}) = \mathcal{P}_n) = \binom{N-n-1}{k-n-1} \Big/ \binom{N}{k} \approx \left(\frac{k}{N}\right)^{n+1} \quad \text{if } c_n = o(n)$$

and the results follows. □ (Theorem)

## References

[1] B. Bollobás, Random Graphs (Academic Press, London, 1985).
[2] P. Erdős and A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hungar. Acad. Sci. 5 (1960) 17–61.
[3] W. Feller, An Introduction to Probability Theory and its Applications (Wiley, New York, 1966).