OCCUPANCY PROBLEMS AND RANDOM ALGEBRAS

by

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

Research Report No. 88-11 2

April 1988

Occupancy Problems and Random Algebras

by

Michael H. Albert and Alan M. Frieze

Department of Mathematics
Carnegie Mellon University
Pittsburgh, PA 15213

April 1988

ABSTRACT

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 note we consider various algebras generated by k randomly chosen subsets of $[n] = \{1,2,\dots,n\}$. As in the study of random graphs (Erdös 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 $A_1, A_2, \dots, A_k$ independently and randomly from $\mathcal{P}_n$ (with replacement). Let $\underline{A}^{(k)}$ denote $A_1, A_2, \dots, A_k$.

We consider

(i) $\mathcal{B}(\underline{A}^{(k)})$ = the Boolean algebra generated by $\underline{A}^{(k)}$.

(ii) $\mathcal{D}(\underline{A}^{(k)})$ = the distributive lattice generated by $\underline{A}^{(k)}$.

(iii) $\mu(\underline{A}^{(k)})$ = the meet semi-lattice generated by $\underline{A}^{(k)}$.

In each case we determine the asymptotic probability that the algebras generated are (a) free, (b) the whole of $\mathcal{P}_n$. We prove the following

Theorem

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

$$\lim_{n\to\infty} P(\mathcal{B}(\underline{A}^{(k)}) \text{ is free}) = 1 \qquad \text{for } k \le \kappa-\epsilon$$

$$\lim_{n\to\infty} P(\mathcal{B}(\underline{A}^{(k)}) \text{ is free}) = 0 \qquad \text{for } k \ge \kappa+\epsilon$$

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

$$\lim_{n\to\infty} P(\mathfrak{B}(\underline{A}^{(k)}) \text{ is all of } \mathcal{F}_n) = \begin{cases} 0 & a_n \to -\infty \\ e^{-2^{-(a+1)}} & a_n \to a \\ 1 & a_n \to \infty \end{cases}$$

(c)

$$\lim_{n\to\infty} P(\mathfrak{D}(\underline{A}^{(k)}) \text{ is free}) = \lim_{n\to\infty} P(\mathfrak{B}(\underline{A}^{(k)}) \text{ is free})$$

(d)

$$\lim_{n\to\infty} P(\mathfrak{D}(\underline{A}^{(k)}) = \mathcal{F}_n) = \lim_{n\to\infty} P(\mathfrak{B}(\underline{A}^{(k)}) = \mathcal{F}_n).$$

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

$$\lim_{n\to\infty} P(\mu(\underline{A}^{(k)}) \text{ is free}) = \begin{cases} 0 & b_n \to -\infty \\ e^{-e^{-b}} & b_n \to b \\ 1 & b_n \to +\infty \end{cases}$$

(f) We deviate from our probabilistic model by assuming the k sets are chosen <u>without</u> replacement. Let now $k = 2^n(1 - \frac{c_n}{n})$ where $c_n \geq 0$. Then

$$\lim_{n\to\infty} P(\mu(\underline{A}^{(k)}) = \mathcal{F}_n) = \begin{cases} 0 & c_n \to \infty \\ e^{-c} & c_n \to c \\ 1 & c_n \to 0 \end{cases}$$

□

§2. Preliminaries

For $S \subseteq [k]$ we define

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

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

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

(2.1) $x_{11}, x_{12}, \ldots, x_{kn}$ form a sequence of independent Bernoulli random variables where for all i, j

$$P(x_{ij} = 0) = P(x_{ij} = 1) = \frac{1}{2}.$$

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

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

and that

(2.3) the random variables $S_1, S_2, \ldots, S_n$ are independent.

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_{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 $\mathfrak{B}(\underline{A}^{(k)})$. We have the following simple result.

Proposition 3.1

$\mathfrak{B}(\underline{A}^{(k)})$ is free if and only if $A_S \neq \phi$ for all $S \subseteq [k]$.

Proof

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

□

It follows from §2 and Proposition 3.1 that

$$P(\mathfrak{B}(\underline{A}^{(k)}) \text{ is free}) = P_{BB}\,(\text{each box is non-empty}).$$

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

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

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

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

$$k \le z - \epsilon \to P(\mathfrak{B}(\underline{A}^{(k)}) \text{ is free}) \to 1$$
$$k \ge z + \epsilon \to P(\mathfrak{B}(\underline{A}^{(k)}) \text{ is free}) \to 0.$$

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

Another simple remark,

Proposition 3.2

$\mathfrak{B}(\underline{A}^{(k)})$ is all of $\mathfrak{P}_n$ if and only if $|A_S| \le 1$ for all $S \subseteq [k]$.

Proof

$\mathfrak{B}(\underline{A}^{(k)})$ is all of $\mathfrak{P}_n$ 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

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

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

<u>Case 1</u>: $a_n \to \infty$.

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

<u>Case 2</u>: $a_n \to a$.

Observe first that

$$E_{BB}(\sum_{t=3}^{n} z_t) \le m\binom{n}{3}\left(\frac{1}{m}\right)^3 = O(m^{-2})$$

and so $P_{BB}(\sum_{t=3}^{n} z_t > 0) = o(1)$.

Thus we only have to show that

$$\lim_{n\to\infty} P_{BB}\ (Z_2 > 0) = e^{-\lambda} \qquad \text{where} \quad \lambda = 2^{-(a+1)}.$$

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

$$\lim_{n\to\infty} E_{BB}((Z_2)_r) = \lambda^r.$$

It follows (see e.g. Bollobás [1], Theorem I.20) that Z is asymptotically Poisson with mean λ. 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 Latices

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

Proposition 4.1

$\mathfrak{D}(\underline{A}^{(k)})$ is free if and only if $A_S \neq \phi$ for $\phi \neq S \subseteq [k]$.

Proof

Assume $\mathfrak{D}(\underline{A}^{(k)})$ is free and $\phi \neq S \subseteq [n]$. Now the two sets

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

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, $A_S \neq \phi$.

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 $A_S \neq \phi$ for $S \neq \phi$, the sets obtained by evaluating these polynomials are distinct and hence $\mathfrak{D}(\underline{A}^{(k)})$ is free.

□

It follows from §2 and Proposition 4.1 that

$$P(\mathcal{D}(\underline{A}^{(k)}) \text{ is free}) = P_{BB}(\text{box } S \text{ is non-empty}, \ \forall\, S \neq \phi)$$

$$= P_{BB}(\text{box } S \text{ is non-empty}, \ \forall\, S) + P_{BB}(\text{box } \phi \text{ is the only empty box}).$$

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

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

But

$$\text{(4.1)} \qquad P_{BB}(\text{box } \phi \text{ is the only empty box}) = (1 - \tfrac{1}{m})^n \, P_{BB}\,(\text{box } S \text{ is non-empty } \forall\, S \neq \phi \,|\, \text{box } \phi \text{ is empty}).$$

Now if $n/m \to \infty$ then $(1 - \frac{1}{m})^n \to 0$ otherwise $(n - (m-1)\log_e(m-1))/(m-1) \to -\infty$ and the conditional probability in (4.1) tends to zero. This completes the proof of (c). Now to part (d) of the theorem.

<u>Proposition 4.2</u>

$\mathcal{D}(\underline{A}^{(k)}) = \mathcal{F}_n$ if and only if $A_\phi = \phi$ and $|A_S| \leq 1$ for all $S \neq \phi$.

<u>Proof</u>

Clearly $\mathcal{D}(A^{(k)}) = \mathcal{F}_n$ if and only if

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

or equivalently

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

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

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

□

Hence

$$P(\mathcal{D}(\underline{A}^{(k)}) = \mathcal{F}_n) = P_{BB}(A_\phi = \phi \text{ and } |A_S| \leq 1 \quad \text{for} \quad S \neq \phi)$$

(4.2) $$= P_{BB}(|A_S| \leq 1,\ \forall\ S) - P_{BB}(|A_S| \leq 1,\ \text{for } S \neq \phi|\ |A_\phi| = 1)P(|A_\phi| = 1)$$

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

§5. Semi-Lattices

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

Proposition 5.1

$\mu(\underline{A}^{(k)})$ is free if and only if $A_{[n]-\{j\}} \neq \phi$ 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

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

So the semilattice $\mu(\underline{A}^{(k)})$ is free if and only if for $\phi \neq I \subseteq [k],\ j \notin I$

$$\bigcap_{i\in I} A_i \not\subseteq A_j.$$

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

$$\bigcap_{i\in[k]-\{j\}} A_i \not\subseteq A_j \qquad \text{for} \quad j \in [k]$$

which is equivalent to the statement in the proposition.

□

Thus

(5.1) $$P(\mu(\underline{A}^{(k)}) \text{ is free}) = P_{BB}(A_{[k]-\{j\}} \neq \phi \quad \text{for} \quad j \in [k]).$$

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

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

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

$$P(\mu(\underline{A}^{(k)}) \text{ is not free}) \leq k(1 - \frac{1}{m})^n$$

$$\leq k\ e^{-n/m}$$

$$(m = 2^k = n/(\log_e \log_2 n + b_n))$$

$$\leq k\ e^{-b_n}/\log_2 n$$

$$= o(1).$$

Case 2: $b_n \to b$.

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

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

and we are done. Now

$$E_{BB}((Z)_r) = (k)_r(1 - \frac{r}{m})^n$$

$$\approx k^r e^{-rn/m}$$

$$\approx k^r (\log_2 n)^{-r} e^{-rb}$$

completing the proof of this case.

Case 3: $b_n \to -\infty$.

Use a monotonicity argument as in Case 3 of (b).

Finally we have

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 $k = 2^n(1 - \frac{c_n}{n})$ sets without replacement. Let $N = 2^n$. It follows from Proposition 5.2 that

$$P(\mu(\underline{A}^{(k)}) = \mathcal{P}_n) = \binom{N-n-1}{k-n-1} / \binom{N}{k}$$

$$\approx \left(\frac{k}{N}\right)^{n+1} \qquad \text{if} \quad c_n = o(N)$$

and the result follows.

□

References

[1] B. Bollobás, **Random Graphs**, Academic Press, London, 1985.

[2] P. Erdös 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**, John Wiley and Sons, New York, 1966.