Counting set systems by weight

Martin Klazar
Institute for Theoretical Computer Science*
and
Department of Applied Mathematics
Charles University, Faculty of Mathematics and Physics
Malostranské náměstí 25
118 00 Prague, Czech Republic
klazar@kam.mff.cuni.cz

Submitted: Jun 21, 2004; Accepted: Jan 27, 2005; Published: Feb 14, 2005
Mathematics Subject Classifications: 05A16, 05C65

Abstract

Applying enumeration of sparse set partitions, we show that the number of set systems $H \subset \exp(\{1,2,\ldots,n\})$ such that $\emptyset \notin H$, $\sum_{E \in H} |E| = n$ and $\bigcup_{E \in H} E = \{1,2,\ldots,m\}$, $m \le n$, equals $(1/\log(2) + o(1))^n b_n$ where b_n is the n-th Bell number. The same asymptotics holds if H may be a multiset. If the vertex degrees in H are restricted to be at most k, the asymptotics is $(1/\alpha_k + o(1))^n b_n$ where α_k is the unique root of $\sum_{i=1}^k x^i/i! - 1$ in $(0,1]$.

1 Introduction

If one wants to count, for a given $n \in \mathbf{N} = \{1,2,\ldots\}$, finite sets H of nonempty finite subsets of $\mathbf{N}$ for which $\sum_{E \in H} |E| = n$, $\bigcup H = \{1,2,\ldots,m\}$ for an $m \le n$, and the sets in H are mutually disjoint, the answer is well known. Such H's are partitions of $\{1,2,\ldots,n\}$ (necessarily $m = n$) and are counted by the n-th Bell number b_n. But how many H's are there if the sets in H may intersect? In other words, what is the number of vertex-labeled simple set systems with n incidences between vertices and edges. In contrast with the case of partitions and Bell numbers, little attention seems to have been paid so far to this natural and basic enumerative problem for general set systems.

We investigate these numbers in [8] and denote them h_n'. By h_n'' we denote the numbers of vertex-labeled set systems with n incidences in which sets may coincide, that is, H is a multiset. We keep this notation here. (The symbol without primes, h_n, denotes in [8] the number of simple vertex-labeled set systems with n vertices.) For example,

*ITI is supported by the project 1M0021620808 of the Ministry of Education of the Czech Republic.

$H_1 = \{\{1\},\{2\},\{3\}\}$, $H_2 = \{\{1,2\},\{3\}\}$, $H_3 = \{\{1,3\},\{2\}\}$, $H_4 = \{\{2,3\},\{1\}\}$, $H_5 = \{\{1,2\},\{1\}\}$, $H_6 = \{\{1,2\},\{2\}\}$, and $H_7 = \{\{1,2,3\}\}$ show that $h_3' = 7$. The three additional multisets $H_8 = \{\{1\},\{1\},\{2\}\}$, $H_9 = \{\{1\},\{2\},\{2\}\}$, and $H_{10} = \{\{1\},\{1\},\{1\}\}$ show that $h_3'' = 10$. See [8] for more values of h_n' and h_n''. In [8] it is shown, among other results, that $b_n \le h_n' \le h_n'' \le 2^{n-1}b_n$. In this note we shall prove the following stronger asymptotic bound.

Theorem. *If $n \to \infty$, h_n' and h_n'' have the asymptotics*

$$((\log 2)^{-1} + o(1))^n \cdot b_n = (1.44269\ldots + o(1))^n \cdot b_n$$

where b_n are Bell numbers (the asymptotics of b_n is reviewed in Proposition 2.6).

We prove the theorem in Section 2. In Section 3 we give concluding comments, point out some analogies and pose some open questions. Now we recall and fix notation. For $n \in \mathbf{N}$, $\{1,2,\ldots,n\}$ is denoted $[n]$. For $A,B \subset \mathbf{N}$, $A < B$ means that $x < y$ for every $x \in A$ and $y \in B$. We use notation $f(n) \ll g(n)$ as synonymous to the $f(n) = O(g(n))$ notation. The coefficient of x^n in a power series $F(x)$ is denoted $[x^n]F$. A *set system* H is here a finite multisubset of $\exp(\mathbf{N})$ whose *edges* $E \in H$ are nonempty and finite. The *vertex set* is $V(H) = \bigcup_{E\in H} E$. The *degree* $\deg(x) = \deg_H(x)$ of a vertex $x \in V(H)$ in H is the number of edges containing x. If there are no multiple edges, we say that H is *simple*. H is a *partition* if its edges are mutually disjoint; in the case of partitions they are usually called *blocks*. The number of partitions H with $V(H) = [n]$ is the Bell number b_n. The *weight* of a set system H is $w(H) = \sum_{v\in V(H)} \deg(v) = \sum_{E\in H} |E|$. H is *normalized* if $V(H) = [m]$ for some m. In the proof of Proposition 2.3 we work with more general set systems H with vertex set contained in the dense linear order of fractions $\mathbf{Q}$. We *normalize* such set system by replacing it by the set system $H' = \{f(E):\ E \in H\}$, $V(H') = [m]$, where $f : V(H) \to [m]$ is the unique increasing bijection.

2 The proof

To estimate h_n' and h_n'' in terms of b_n, we transform a set system H into a set partition with the same weight by splitting each vertex $v \in V(H)$ in $\deg_H(v)$ new vertices which are 1-1 distributed among the edges containing v. The following definitions and Propositions 2.1 and 2.2 make this idea precise.

We call two set partitions P and Q of $[n]$ *orthogonal* if $|A \cap B| \le 1$ for every two blocks $A \in P$ and $B \in Q$. Q is an *interval* partition of $[n]$ if every block of Q is a subinterval of $[n]$. For $n \in \mathbf{N}$ we define $W(n)$ to be the set of all pairs (Q,P) such that Q and P are orthogonal set partitions of $[n]$ and Q is moreover an interval partition. We define a binary relation $\sim$ on $W(n)$ by setting $(Q_1,P_1) \sim (Q_2,P_2)$ iff $Q_1 = Q_2$ and there is a bijection $f : P_1 \to P_2$ such that for every $A \in P_1$ the blocks A and $f(A)$ intersect the same intervals of the partition $Q_1 = Q_2$. It is an equivalence relation.

Proposition 2.1 *For every $n \in \mathbf{N}$, there is a bijection $(Q, P) \mapsto H(Q, P)$ between the set of equivalence classes $W(n)/\sim$ and the set $L(n)$ of normalized set systems H with weight n. In particular, $h''_n = |L(n)| = |W(n)/\sim|$.*

Proof. We transform every $(Q, P) \in W(n)$, where Q consists of the intervals $I_1 < I_2 < \ldots < I_m$, into the set system $H = H(Q, P) = (E_A : \; A \in P)$ where $E_A = \{i \in [m] : A \cap I_i \neq \emptyset\}$. We have $w(H) = n$ and $V(H) = [m]$, so $H \in L(n)$. It is easy to see that equivalent pairs produce the same H and nonequivalent pairs produce distinct elements of $L(n)$.

Let $H \in L(n)$ with $V(H) = [a]$. We split $[n]$ in a intervals $I_1 < I_2 < \ldots < I_a$ so that $|I_i| = \deg_H(i)$. For every $i \in [a]$ we fix arbitrary bijection $f_i : \{E \in H : \; i \in E\} \to I_i$. We define the partitions $Q = (I_1, I_2, \ldots, I_a)$ and $P = (A_E : \; E \in H)$ where $A_E = \{f_i(E) : i \in E\}$. Clearly, $(Q, P) \in W(n)$ and different choices of bijections f_i lead to equivalent pairs. Also, $H(Q, P) = H$. Thus $(Q, P) \mapsto H(Q, P)$ is a bijection between $W(n)/\sim$ and $L(n)$. □

The next proposition summarizes useful properties of the equivalence $\sim$ and the bijection $(Q, P) \mapsto H(Q, P)$. They follow in a straightforward way from the construction and we omit the proof.

Proposition 2.2 *Let $(Q, P) \in W(n)$, $Q = (I_1 < I_2 < \ldots < I_m)$, and $H = H(Q, P)$ (so $V(H) = [m]$). Then $\deg_H(i) = |I_i|$ for every $i \in [m]$. The equivalence class containing (Q, P) has at most $|I_1|! \cdot |I_2|! \cdot \ldots \cdot |I_m|!$ pairs. It has exactly so many pairs if and only if H is simple.*

Proposition 2.3 *For every $n \in \mathbf{N}$, $h'_n \le h''_n \le 2h'_n$.*

Proof. The first inequality is trivial. To prove the second inequality, we construct an injection from the set $N(n)$ of normalized non-simple set systems H with weight n in the set $M(n)$ of normalized simple set system H with weight n. Then $h''_n = |M(n)| + |N(n)| \le 2|M(n)| = 2h'_n$. We say that a vertex $v \in V(H)$ is *regular* if $\deg(v) \ge 2$ or if $v \in E$ for some $E \in H$ with $|E| \ge 2$, else we call v *singular*. Thus v is singular iff $\{v\} \in H$ and $\deg(v) = 1$.

Let $H \in N(n)$. We distinguish two cases. The first case is when every multiple edge of H is a singleton. Then let $k \ge 2$ be the maximum multiplicity of an edge in H and $v = u - 1$ where $u \in V(H)$ is the smallest regular vertex in H; we may have $v = 0$ and then v is not a vertex of H. We have $v < \max V(H)$ and insert between v and u new vertices w_i, $i = 1, 2, \ldots, k-1$ and $v < w_1 < w_2 < \ldots < w_{k-1} < u$. Then we replace every singleton multiedge $\{x\}$ with multiplicity m, $2 \le m \le k$, (we have $x \ge u$) with the new single edge $\{w_1, w_2, \ldots, w_{m-1}, x\}$. Normalizing the resulting set system we get the set system H'. Clearly, $H' \in M(n)$.

The second case is when at least one multiple edge in H is not a singleton. We define k, v, u, and $w_1, \ldots, w_{k-1}$ as in the first case and replace every multiedge E with multiplicity m, $2 \le m \le k$, by the new single edge $\{w_1, w_2, \ldots, w_{m-1}\} \cup E$ (we have $\min E \ge u$). We add between w_{k-1} and u a new vertex s and add a new singleton edge $\{s\}$. This singleton edge is a marker discriminating between both cases and separating the new vertices w_i from those in E. Since $m-1+|E| < m|E|$ if $|E| \geq 2$ and $m \geq 2$, the weight is still at most n. We add in the beginning sufficiently many new singleton edges $\{-r\}, \ldots, \{-1\}, \{0\}$ so that the resulting set system has weight exactly n. Normalizing it, we get the set system H'. Again, $H' \in M(n)$. Note that in both cases the least regular vertex in H' is w_1 and that in both cases the longest interval in $V(H')$ that starts in w_1 and is a proper subset of an edge ends in w_{k-1}.

Given the image $H' \in M(n)$, in order to reconstruct H we let $w \in V(H')$ be the least regular vertex (i.e., w is the first vertex lying in an edge E with $|E| \geq 2$) and let I be the longest interval in $V(H')$ that starts in w and is a proper subset of an edge. If $\max I + 1$ is a singular vertex of H', it must be s and we are in the second case. Else there is no s and we are in the first case. Knowing this and knowing (in the second case) which vertices are the dummy w_i, we uniquely reconstruct the multiedges of H. Thus $H \mapsto H'$ is an injection from $N(n)$ to $M(n)$. $\square$

For $k, n \in \mathbf{N}$ we define $h''_{k,n}$ to be the number of normalized set systems with weight n and maximum vertex degree at most k. The number of such set systems which are simple is $h'_{k,n}$. The next Proposition 2.4 can be proved by an injective argument similar to the previous one and we leave the proof as an exercise for the interested reader. But note that one cannot use the previous injection without change because it creates vertices with high degree.

Proposition 2.4 *For every $k, n \in \mathbf{N}$ we have $h'_{k,n} \leq h''_{k,n} \leq 2h'_{k,n}$.*

For the lower bound on h''_n we need to count sparse partitions. A partition P of $[n]$ is *m-sparse*, where $m \in \mathbf{N}$, if for every two elements $x < y$ of the same block we have $y - x \geq m$. Thus every partition is 1-sparse and 2-sparse partitions are those with no two consecutive numbers in the same block. If $m' < m$, every m-sparse partition is also m'-sparse. The number of m-sparse partitions of $[n]$ is denoted $b_{n,m}$. The following enumeration of sparse partitions was obtained by Prodinger [11] and Yang [15], see also Stanley [14, Problem 1.4.29]. Here we present a simple and nice proof due to Chen, Deng and Du [5].

Proposition 2.5 *Let $m, n \in \mathbf{N}$. For $m > n$ there is only one m-sparse partition of $[n]$. For $m \leq n$ the number $b_{n,m}$ of m-sparse partitions of $[n]$ equals the Bell number b_{n-m+1}.*

Proof. For $m > n$ the only partition in question is that with singleton blocks. Let P be a partition of $[n]$. We represent it by the graph $G = ([n], E)$ where for $x < y$ we set $\{x, y\} \in E$ iff $x, y \in A$ for some block A of P and there is no $z \in A$ with $x < z < y$. The components of G are increasing paths corresponding to the blocks of P. Equivalently, G has the property that each vertex has degree at most 2 and if it has degree 2, it must lie between its two neighbors. Now assume that P is 2-sparse. We transform G into the graph $G' = ([n-1], E')$ where $E' = \{\{x, y-1\} : \{x, y\} \in E, x < y\}$, i.e., we decrease the second vertex of each edge by one. Note that G' is again a graph (no loops arise). The property of G is preserved by the transformation and hence the components of G' are increasing paths and G' describes a partition P' of $[n-1]$. Clearly, P is m-sparse iff P' is $(m-1)$-sparse. Thus $P \mapsto P'$ maps m-sparse partitions of $[n]$ to $(m-1)$-sparse partitions of $[n-1]$. The inverse mapping is obtained by increasing the second vertex of each edge by one. Thus $P \mapsto P'$ is a bijection between the mentioned sets. Iterating it, we obtain the stated identity. $\square$

See [5] for other applications of this bijection. We remark that the representing graphs of partitions (but not the transformation of G into G') were used before by Biane [1] and Simion a Ullman [12].

We need to compare, for fixed m, the growth of b_n and b_{n-m}. The following asymptotics of Bell numbers is due to Moser and Wyman [10].

Proposition 2.6 *For $n \to \infty$,*

$$b_n \sim \frac{\lambda(n)^{n+1/2}}{n^{1/2}\mathrm{e}^{n+1-\lambda(n)}}$$

where the function $\lambda(n)$ is defined by $\lambda(n)\log\lambda(n) = n$.

It follows by a simple calculation that $b_{n-1}/b_n \sim \log n/n$. More generally, we have the following.

Corollary 2.7 *If m fixed and $n \to \infty$,*

$$\frac{b_{n-m}}{b_n} \sim \left(\frac{\log n}{n}\right)^m.$$

In fact, a better approximation is $b_{n-1}/b_n \sim (\log n - \log\log n)/n$. Knuth [9] gives a nice account on the asymptotics of b_n and shows that $b_{n-1}/b_n = (\xi/n)(1 + O(1/n))$ where $\xi \cdot \mathrm{e}^\xi = n$.

We are ready to estimate the numbers of normalized set systems with weight n and maximum degree at most k.

Proposition 2.8 *For fixed $k \in \mathbf{N}$ and $n \to \infty$,*

$$\left(\frac{\log n}{n}\right)^{k-1}\left(\frac{1}{\alpha_k}\right)^n b_n \ll h''_{k,n} \ll \left(\frac{1}{\alpha_k}\right)^n b_n$$

where α_k is the only root of the polynomial $\sum_{i=1}^k x^i/i! - 1$ in $(0, 1]$.

Proof. Let $i_{n,k}$ be the number of interval partitions $Q = (I_1 < I_2 < \ldots < I_m)$ of $[n]$ such that $|I_i| \le k$ for all i and Q is weighted by $(|I_1|! \cdot |I_2|! \cdot \ldots \cdot |I_m|!)^{-1}$. It follows that

$$i_{n,k} = [x^n]\frac{1}{1 - \sum_{i=1}^k x^i/i!} \sim c_k\left(\frac{1}{\alpha_k}\right)^n$$

with some constant $c_k > 0$ because α_k is the only root of the denominator in $(0,1]$ and it is simple. Using Propositions 2.1, 2.2, and 2.4, we obtain the inequalities

$$i_{n,k} b_{n,k} \le h''_{k,n} \le 2 i_{n,k} b_n.$$

In the first inequality we use the fact that if Q is an interval partition of $[n]$ with interval lengths at most k and P is a k-sparse partition of $[n]$, then Q and P are always orthogonal. In the second inequality we neglect orthogonality of the pairs (Q,P) but we count only the corresponding equivalence classes in $W(n)$ with full cardinalities $|I_1|! \cdot |I_2|! \cdot \ldots \cdot |I_m|!$. By Proposition 2.2, this gives an upper bound for $h'_{k,n}$. Using Proposition 2.4, we get an upper bound for $h''_{k,n}$. The explicit lower and upper bounds on $h''_{k,n}$ now follow from the above asymptotics of $i_{n,k}$, Proposition 2.5, and Corollary 2.7. $\square$

Note that $1/\alpha_2 = (1+\sqrt{3})/2$. Thus we have roughly $((1+\sqrt{3})/2)^n b_n = (1.36602\ldots)^n b_n$ normalized set systems with weight n, in which each vertex lies in one or two edges.

Proof of the Theorem. We prove that, for $n \to \infty$,

$$h''_n = \left(\frac{1}{\log 2} + o(1)\right)^n b_n = (1.44269\ldots + o(1))^n b_n.$$

Let i_n be the number of interval partitions Q of $[n]$, weighted as in the previous proof. As in the case of bounded degree, by Propositions 2.1, 2.2, and 2.3 we have the upper bound

$$h''_n \le 2 i_n b_n \sim c\left(\frac{1}{\log 2}\right)^n b_n$$

because

$$i_n = [x^n]\frac{1}{1 - \sum_{i=1}^{\infty} x^i/i!} = [x^n]\frac{1}{2 - \mathrm{e}^x}$$

and $\log 2$ is a simple zero of $2 - \mathrm{e}^x$. As for the lower bound, $h''_n \ge h''_{k,n}$ for every $k, n \in \mathbf{N}$. It is easy to show that $\alpha_k \downarrow \log 2$ for $k \to \infty$. Hence, by the lower bound in Proposition 2.8, for any fixed $\varepsilon > 0$ we have $h''_n > ((\log 2)^{-1} - \varepsilon)^n b_n$ for n big enough. $\square$

3 Concluding remarks

P. Cameron investigates in [4] a family of enumerative problems on 0-1 matrices including h'_n and h''_n as particular cases. He defines $F_{ijkl}(n)$, $i,j,k,l \in \{0,1\}$, to be the number of rectangular 0-1 matrices with no zero row or column and with n 1's, where $i = 0$, resp. $i = 1$, means that matrices differing only by a permutation of rows are identified, resp. are considered as different; $j = 0$, resp. $j = 1$, means that matrices with two equal rows are forbidden, resp. are allowed; and the values of k, l refer to the same (non)restrictions for columns. Notice that $F_{ijkl}(n)$ is nondecreasing in each of the arguments i,j,k,l. Representing set systems by incidence matrices, rows standing for edges and columns for vertices, we see that $h''_n = F_{0111}(n)$ and $h'_n = F_{0011}(n)$. In [4] it is shown that $F_{1111}(n) \sim Ac^{n+1}n!$ where $A = \frac{1}{4}\exp(-(\log 2)^2/2) \approx 0.19661$ and $c = (\log 2)^{-2} \approx 2.08137$. $F_{0101}(n)$ is A049311 of [13], see also Cameron [3]. P. Cameron asks in [2, Problem 3] if there is an effective algorithm to calculate $F_{0101}(n)$; for h'_n and h''_n such algorithms are given in [8].

Interestingly, in the so far derived asymptotics of the functions $F_{ijkl}(n)$ the constant $\log 2 \approx 0.69314$ appears quite often. Our theorem says that intersections of blocks in "partitions" of $[n]$ magnify the counting function by the exponential factor $(\log 2)^{-n}$. The same phenomenon occurs for counting injections and surjections. If i_n is the number of injections from $[n]$ to $\mathbf{N}$ with images normalized to $[n]$ and s_n is the number of all mappings ("injections with intersections") from $[n]$ to $\mathbf{N}$, again with images normalized to $[m]$ (i.e., s_n counts surjections from $[n]$ to $[m]$), then $i_n = n!$ (trivial) and $s_n \sim c(\log 2)^{-n}n!$ where $c = (2\log 2)^{-1}$ (a nice exercise on exponential generating functions, see Flajolet and Sedgewick [6, Chapter 2.3.1]).

Another parallel can be led between sparse partitions and sparse words. We say that a word $u = a_1a_2\dots a_l$ over an alphabet A, $|A| = r$, is k-sparse if $a_i = a_j$, $j > i$, implies $j - i \geq k$. (We remark that k-sparse words are basic objects in the theory of generalized Davenport–Schinzel sequences, see Klazar [7]. Another term for 2-sparse words is *Smirnov words*.) The two notions of sparseness, in fact, coincide: $u = a_1a_2\dots a_l$ defines a partition P of $[l]$ via the equivalence $i \sim j \iff a_i = a_j$ and then, obviously, u is k-sparse if and only if P is k-sparse. A partition of $[l]$ can be defined by many words u (even if A is fixed). The unique canonical defining words are *restricted growth strings*, see [9] for their properties and more references. If v_n is the number of all words over A ($|A| = r$) with length n and $s_{k,n}$ is the number of those which are k-sparse, then $v_n = r^n$ (trivial) and

$$\begin{aligned} s_{k,n} &= r(r-1)\dots(r-k+2)(r-k+1)^{n-k+1} \\ &= \frac{r(r-1)\dots(r-k+2)}{(r-k+1)^{k-1}}\left(1 - \frac{k-1}{r}\right)^n v_n \end{aligned}$$

(simple direct counting, for the generating functions approach see [6, Chapter 3.6.3]). In the case of words over a fixed alphabet sparseness diminishes the counting function by an exponential factor. For partitions the decrease is, fortunately, only by a polynomial factor (Proposition 2.5 and Corollary 2.7).

We conclude with two natural questions. What is the precise asymptotics of $h''_{k,n}$ and h''_n? By Propositions 2.3 and 2.4, $1/2 \leq h'_{k,n}/h''_{k,n} \leq 1$ and $1/2 \leq h'_n/h''_n \leq 1$. Do these ratios go to 1 as $n \to \infty$?

Acknowledgement. I would like to thank Peter Cameron for making [4] available to me and for interesting discussions.

References

[1] P. Biane, Some properties of crossings and partitions, *Discrete Math.,* **175** (1997), 41–53.

[2] P.J. Cameron, Problems on permutation groups, available at http://www.maths.qmul.ac.uk/~pjc/pgprob.html

[3] P.J. Cameron, Sequences realized by oligomorphic permutation groups, *J. Integer Sequences,* **3(1)** (2000), article 00.1.5.

[4] P.J. Cameron, Counting zero-one matrices, draft.

[5] W.Y.C. Chen, Y.-P. Deng and R. Du, Reduction of m-regular noncrossing partitions, *Europ. J. Combin.,* **26** (2005), 237–243.

[6] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, available at http://pauillac.inria.fr/algo/flajolet/Publications/books.html

[7] M. Klazar, Generalized Davenport-Schinzel sequences: results, problems, and applications, *Integers,* **2** (2002), A11, 39 pp.

[8] M. Klazar, Extremal problems for ordered hypergraphs: small patterns and some enumeration, *Discrete Appl. Math.,* **143** (2004), 144–154.

[9] D.E. Knuth, *The Art of Computer Programming*, A Draft of Sections 7.2.1.4-5: Generating All Partitions, available at http://www-cs-faculty.stanford.edu/~knuth/

[10] L. Moser and M. Wyman, An asymptotic formula for the Bell numbers, *Trans. Royal Soc. Can.,* **49** (1955), 49–54.

[11] H. Prodinger, On the number of Fibonacci partitions of a set, *Fibonacci Quart.,* **19** (1981), 463–465.

[12] R. Simion and D. Ullman, On the structure of the lattice of noncrossing partitions, *Discrete Math.,* **98** (1991), 193–206.

[13] N.J.A. Sloane (2000), The On-Line Encyclopedia of Integer Sequences, published electronically at
http://www.research.att.com/~njas/sequences/.

[14] R.P. Stanley, *Enumerative Combinatorics. Volume I,* Wadsworth & Brooks/Cole, Monterey, CA, 1986.

[15] W. Yang, Bell numbers and k-trees, *Discrete Math.,* **156** (1996), 247–252.