Entropy and Enumeration

Nicholas Pippenger*
(nicholas@cs.ubc.ca)

Department of Computer Science
The University of British Columbia
Vancouver, British Columbia V6T 1Z4
CANADA

Abstract: Shannon's notion of the entropy of a random variable is used to give simplified proofs of asymptotic formulas for the logarithms of the numbers of monotone Boolean functions and Horn functions, and for equivalent results concerning families of sets and closure operations.

* The work reported here was supported by an NSERC Research Grant.

1. Introduction

The goal of this paper is to show how Shannon's notion of the entropy of a discrete random variable (see Shannon [S]) can be used to solve some problems of combinatorial enumeration. All of the results we obtain (and in fact stronger ones) have previously been established by direct combinatorial arguments. The proofs given here are simpler, however, and we believe they show the relevance of information-theoretic methods for enumerative problems. Information-theoretic methods have been used in the past to prove a number of combinatorial results (see for example Lindström [L], Pippenger [P], Chung, Graham, Frankl and Shearer [C] and Newman and Wigderson [N]), but we know of only one previous result, due to Massey [M], that has a natural enumerative interpretation. Since we shall need Massey's result later, we shall give the argument here.

Proposition 1.1: Let $\mathcal{B}$ denote the set of subsets of an n-element set $[n] = \{1, \ldots, n\}$ having cardinality at most m, and let $B = \#\mathcal{B}$ denote the number of such subsets. Then we have

$$B \leq 2^{n h_1(m/n)},$$

where

$$h_1(q) = \begin{cases} -q \log_2 q - (1-q)\log_2(1-q), & \text{if } 0 \leq q \leq 1/2; \\ 1, & \text{if } 1/2 \leq q \leq 1. \end{cases}$$

Proof: Let X be a random set uniformly distributed in $\mathcal{B}$. We have

$$\log_2 B = H(X). \tag{1.1}$$

We may encode X as a sequence $x = (x_1, \ldots, x_n)$ of random variables, where x_i assumes the value 1 if $i \in X$ and 0 if $i \notin X$. Let $p = \Pr(x_i = 1)$. (This probability is independent of i by symmetry.) Then $p \leq m/n$, since the sum of the cardinalities of the sets in $\mathcal{B}$ is pnB, but also at most mB. The entropy of a binary random variable that assumes the value 1 with probability r is

$$h(r) = -r \log_2 r - (1-r)\log_2(1-r).$$

Thus the entropy of a binary random variable that assumes the value 1 with probability at most q is $\max_{0 \leq r \leq q} h(r) = h_1(q)$. Thus we have

$$H(X) = H(x) \leq \sum_{1 \leq i \leq n} H(x_i) \leq n\, h_1(m/n).$$

Combining this estimate with (1.1) completes the proof. $\triangle$

This result is of course well known by other arguments, such as the Chernoff bound (see Jelinek [J], Section 5.2, for example). The argument given above, however, presents in embryonic form the method we shall use below.

It is worthwhile to consider carefully what is involved in Massey's argument. The main ingredient of course is Shannon's notion of the entropy $H(X)$ of a random variable X that assumes values $v_1, v_2, \ldots, v_B$ with probabilities $p_1, p_2, \ldots, p_B$, respectively:

$$H(X) = - \sum_{1 \leq k \leq B} p_k \log_2 p_k.$$

It is a straightforward result that $H(X)$ assumes its maximum value $\log_2 B$ when X has the uniform distribution $p_1 = p_2 = \cdots = p_B = 1/B$ (Jelinek [J], Lemma 4.9). Another ingredient is the inequality

$$H(X_1, \ldots, X_n) \leq H(X_1) + \cdots + H(X_n),$$

which follows by induction from the two-component version

$$H(X, Y) \leq H(X) + H(Y)$$

(Jelinek [J], Lemma 4.14). A final ingredient, which is not needed in Massey's proof but will be needed in ours, is the inequality

$$H(X, Y) \geq H(X) \tag{1.2}$$

(Jelinek [J], Lemmas 4.12 and 4.13). We shall use (1.2) in the following form: if X is a deterministic function of Y, then $H(Y) = H(X, Y) \geq H(X)$.

2. Monotone Boolean Functions

The objects that we enumerate can be regarded as either Boolean functions or families of sets. A *Boolean function* of n arguments is a map $f : \{0,1\}^n \to \{0,1\}$. A Boolean n-tuple $x = (x_1, \ldots, x_n)$ of arguments for such a function will be called a *point*. A *family of sets* over n elements is a family $F \subseteq \mathcal{P}([n])$, where $[n] = \{1, \ldots, n\}$ and $\mathcal{P}([n])$ denotes the power set (set of all subsets) of $[n]$. Given a Boolean function f of n arguments, we can associate with it a family F_f of sets over n elements by associating with each $x = (x_1, \ldots, x_n) \in \{0,1\}^n$ a set $X_x = \{k : x_k = 1\} \subseteq [n]$ and then taking $F_f = \{X_x : f(x) = 1\}$.

This association is clearly a one-to-one correspondence, and we shall use function-theoretic and set-theoretic terminology interchangeably. If f is a function, we define its *complement* $\overline{f}$ by $\overline{f}(x) = \overline{f(x)}$, and we define its *reflection* f' by $f'(x) = f(\overline{x})$ (where $\overline{x} = (\overline{x_1}, \ldots, \overline{x_n})$ denotes the component-wise complement of x).

The first problem to which we apply our method is the enumeration of monotone Boolean functions. A Boolean function f is *monotone* (or "positive") if the inequalities $x_1 \leq y_1, \ldots, x_n \leq y_n$ together imply $f(x_1, \ldots, x_n) \leq f(y_1, \ldots, y_n)$. A function is the complement of a monotone function if and only if it is the reflection of a monotone function; the resulting functions are called *antitone* (or "negative") and they are in one-to-one correspondence (in two ways) with the monotone functions. Monotone functions correspond to families of sets that are *saturated upward* (or closed under taking supersets). The minimal sets in such a family form an *inclusion-free* family (or "antichain", or "clutter"). Antitone functions correspond to families of sets that are *saturated downward* (or closed under taking subsets). The maximal sets in such families also form inclusion-free families. Let $\psi(n)$ denote the number of monotone functions of n arguments. (Some authors exclude the constant functions, and thus define $\psi(n)$ to be smaller by 2.) In estimating $\psi(n)$, we shall also be enumerating antitone functions, upward-saturated families, downward-saturated families and inclusion-free families.

The problem of determining $\psi(n)$ was posed by Dedekind [D], but apart from results for specific small values of n, all exact results amount to paraphrases of one of the definitions. There is a long sequence of works aimed at estimating the behavior of $\psi(n)$ for large n, and we shall not recount it here. The particular result we propose to derive is due to Kleitman [K1], who was the first to give an asymptotic formula for the logarithm of $\psi(n)$,

$$\log_2 \psi(n) \sim \binom{n}{\lfloor \frac{n}{2} \rfloor}. \tag{2.1}$$

The best result currently known is due to Korshunov [K2], who gives an asymptotic formula for $\psi(n)$ itself,

$$\psi(n) \sim 2^{\binom{n}{\frac{n}{2}}} \exp\left[\binom{n}{\frac{n-2}{2}} \left(2^{-n/2} + n^2\, 2^{-n-5} - n\, 2^{-n-4}\right)\right]$$

for n even and

$$\psi(n) \sim 2 \cdot 2^{\binom{n}{\frac{n-1}{2}}} \exp\left[\binom{n}{\frac{n-3}{2}} \left(2^{-(n+3)/2} - n^2\, 2^{-n-6} - n\, 2^{-n-3}\right) + \binom{n}{\frac{n-1}{2}} \left(2^{-(n+1)/2} + n^2\, 2^{-n-4}\right)\right]$$

for n odd. This last result lies far beyond the reach of our information-theoretic method. Indeed, Kleitman's proof yields

$$\log_2 \psi(n) = \binom{n}{\lfloor \frac{n}{2} \rfloor} \left(1 + O\left(\frac{\log n}{n^{1/2}}\right)\right),$$

whereas we establish (2.1) only in the following weaker form.

Theorem 2.1: As $n \to \infty$, we have

$$\log_2 \psi(n) = \binom{n}{\lfloor \frac{n}{2} \rfloor} \left(1 + O\left(\frac{(\log n)^{3/2}}{n^{1/4}}\right)\right).$$

Proof: Set $N = \binom{n}{\lfloor \frac{n}{2} \rfloor}$. The lower bound

$$\log_2 \psi(n) \geq N$$

is immediate, since there are 2^N monotone Boolean functions f for which $f(x) = 0$ for $\|x\| < \lfloor n/2 \rfloor$ and $f(x) = 1$ for $\|x\| > \lfloor n/2 \rfloor$, where $\|x\| = \sum_{1 \leq i \leq n} x_i$. Thus our task is to obtain an asymptotically matching upper bound.

Let $\mathcal{C} = \{C_1, \ldots, C_N\}$ be a partition of $\{0,1\}^n$ into N disjoint chains. (The existence of such partitions is well known; one explicit construction is described in the Appendix.) Say a point x is *low* if $\|x\| < n/4$, and say that a chain C_j is *low* if it contains a low point.

Let $\mathcal{M}_n$ denote the set of all monotone Boolean functions of n arguments, and let f be a random function uniformly distributed on $\mathcal{M}_n$. We have

$$\log_2 \psi(n) = H(f).$$

We shall obtain an upper bound for $H(f)$.

Given f, we shall construct a random variable $\delta = (\tilde{\delta}, \hat{\delta})$ as follows. For any function $g \in \mathcal{M}_n$, let

$$\gamma_j(g) = \#\{x \in C_j : g(x) = 1\}.$$

Set $p = (\log n)^{1/2}/n^{1/4}$ and let $\upsilon_1, \ldots, \upsilon_N$ be independent random variables defined as follows. If C_j is low, then $\upsilon_j = 1$. Otherwise, υ_j assumes the value 1 with probability p and the value 0 with probability $1-p$. Take $\tilde{\delta} = (\tilde{\delta}_1, \ldots, \tilde{\delta}_N)$, where $\tilde{\delta}_j = \upsilon_j \, \gamma_j(f)$. Let $\tilde{f}$ be the smallest monotone function (the conjunction of all monotone functions) such that $\gamma_j(\tilde{f}) \geq \tilde{\delta}_j$ for all $1 \leq j \leq N$. Clearly we have $\gamma_j(f) \geq \gamma_j(\tilde{f})$ for all $1 \leq j \leq N$.

Take $\hat{\delta} = (\hat{\delta}_1, \ldots, \hat{\delta}_N)$, where $\hat{\delta}_j = \gamma_j(f) - \gamma_j(\tilde{f})$. Clearly f is determined by δ, so that $H(f) \leq H(\delta)$ and

$$\log_2 \psi(n) \leq H(\delta).$$

We shall obtain an upper bound for $H(\delta)$.

Lemma 2.2: Suppose that the random variable K takes values in $\{0, \ldots n\}$, and that for some $k \geq 1$ and $0 \leq q \leq 1$,

$$\Pr(K \geq k) \leq q.$$

Then

$$H(K) \leq h_1(q) + \log_2 k + q \log_2 n.$$

Proof: For any event E, we have

$$H(K) \leq H(E) + \Pr(\overline{E})\, H(K \mid \overline{E}) + \Pr(E)\, H(K \mid E).$$

The lemma follows by taking E to be the event "$K \geq k$", so that $H(E) \leq h_1(q)$. $\triangle$

We have

$$H(\tilde{\delta}) \leq \sum_{1 \leq j \leq N} H(\tilde{\delta}_j).$$

Since $\tilde{\delta}_j \geq 1$ only if $v_j = 1$, we can apply Lemma 2.2 with $k = 1$ and $q = \Pr(v_j = 1)$ to each term of this sum. If C_j is low, we have $q = 1$. Otherwise, we have $q = p$. Letting M denote the number of low chains, we have

$$H(\tilde{\delta}) \leq M(1 + \log_2 n) + (N - M)\big(h_1(p) + p \log_2 n\big).$$

The number of low chains is at most the number of low points, which can be bounded by applying Proposition 1.1 with $m = n/4$ to yield

$$M \leq 2^{n\, h_1(1/4)} \leq \frac{2^n}{n}$$

(for n is sufficiently large). Since $N \sim (2/\pi n)^{1/2}\, 2^n < 2^n / n^{1/2}$ (for n sufficiently large) and $h_1(p) \leq -2 \log_2 p$ (for $p \leq 1/2$), we obtain

$$\begin{aligned} H(\tilde{\delta}) &\leq \frac{2^n \log_2 n}{n} + \frac{3 \cdot 2^n (\log_2 n)^{3/2}}{n^{3/4}} \\ &\leq \frac{4 \cdot 2^n (\log_2 n)^{3/2}}{n^{3/4}} \qquad (2.2) \end{aligned}$$

(for n sufficiently large).

Next we have

$$H(\hat{\delta}) \le \sum_{1\le j\le N} H(\hat{\delta}_j).$$

This time we shall apply Lemma 2.1 with $k=2$. Set $q_j = \Pr(\hat{\delta}_j \ge 2)$ and $Q = \sum_{1\le j\le N} q_j$. Then we have

$$\begin{aligned} H(\hat{\delta}) &\le \sum_{1\le j\le N} h_1(q_j) + 1 + q_j \log_2 n \\ &\le N h_1(Q/N) + N + Q\log_2 n, \end{aligned} \tag{2.3}$$

since $h_1(q)$ is concave in q.

Say that a chain C_j is *bad* if $\hat{\delta}_j \ge 2$. The quantity Q is the expected number of bad chains. Say that a point $x \in \{0,1\}^n$ is *bad* if (1) x is not low, (2) the chain C_j that contains x also contains some point y with $\|y\| = \|x\| - 1$ and $f(y) = 1$, and (3) $\tilde{f}(x) = 0$. Let r_x denote the probability that x is bad, and let $R = \sum_{x\in\{0,1\}^n} r_x$. A chain C_j is bad only if some $x \in C_j$ is bad, so we have $Q \le R$.

We shall now estimate r_x. If x is low, then x cannot be bad. Thus we may suppose that x is not low. Set $s = 2n^{1/4}(\log n)^{1/2}$. Say x is *heavy* if there are at least s values of $y \in \{0,1\}^n$ such that $\|y\| = \|x\| - 1$ and $f(y) = 1$. If x is heavy, then $\tilde{f}(x) = 0$ only if $v_j = 0$ for each of the s or more chains C_j that contain some y with $\|y\| = \|x\| - 1$ and $f(y) = 1$. Thus the probability that a heavy x is bad is at most $(1-p)^s \le e^{-ps} \le 1/n$.

The group $\mathrm{Sym}(n)$ of permutations of the set $[n]$ acts on points by $\sigma(x) = (x_{\sigma^{-1}(1)}, \ldots, x_{\sigma^{-1}(n)})$, and on functions by $\sigma(f) = f\big(\sigma^{-1}(x)\big)$. The subgroup $\mathrm{Stab}(x)$ of $\mathrm{Sym}(x)$ that fixes x acts transitively on the points y such that $\|y\| = \|x\| - 1$.

Whether x is heavy depends on f only through the orbit of f under the group $\mathrm{Stab}(x)$. If x is not heavy, the probability (averaging over this orbit) that the chain C_j that contains x also contains some y with $\|y\| = \|x\| - 1$ and $f(y) = 1$ is at most $s/(n/4) = 8(\log_2 n)^{1/2}/n^{3/4}$. Thus the probability that x is bad is at most $\max\{1/n, 8(\log_2 n)^{1/2}/n^{3/4}\} = 8(\log_2 n)^{1/2}/n^{3/4}$ (for n sufficiently large). Thus the expected number of bad x that are not heavy is at most $8(\log_2 n)^{1/2} 2^n/n^{3/4}$ Thus we obtain

$$Q \le R \le \frac{8\cdot 2^n (\log_2 n)^{1/2}}{n^{3/4}}.$$

Since $h_1(q)$ is non-decreasing in q, substituting this bound in (2.3) yields

$$H(\hat{\delta}) \le N + \frac{16\cdot 2^n (\log_2 n)^{3/2}}{n^{3/4}}.$$

Combining this bound with (2.2) yields

$$H(\delta) \leq N + \frac{20 \cdot 2^n (\log_2 n)^{3/2}}{n^{3/4}}.$$

Since $N \sim (2/\pi n)^{1/2}\, 2^n > 2^n/2n^{1/2}$ (for n sufficiently large), the proof of Theorem 2.1 is complete.

3. Horn Functions

The second problem to which we apply our method is the enumeration of families of sets closed under taking unions. These are of course in one-to-one correspondence under reflection with families of sets closed under taking intersections. They also correspond to functions f that satisfy the condition that $f(x) \wedge f(y) \leq f(x \vee y)$ (where $x \vee y$ denotes the component-wise disjunction $(x_1 \vee y_1, \ldots, x_n \vee y_n)$ of $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$), or the condition $f(x) \wedge f(y) \leq f(x \wedge y)$ (where $x \wedge y$ denotes the component-wise conjunction $(x_1 \wedge y_1, \ldots, x_n \wedge y_n)$ of $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$). Functions satisfying the latter condition (or sometimes their complements) are called Horn functions (see Horn [H]). Let $\alpha(n)$ denote the number of Horn functions of n arguments.

The sets in a union-closed (respectively, intersection-closed) family of sets that cannot be expressed as the union (respectively, intersection) of two other sets in the family form a *union-free* (respectively, *intersection-free*) family of sets. If one adjoins to a union-free (respectively, intersection-free) family of sets all sets that can be obtained by taking unions (respectively, intersections) one recovers the original union-closed (respectively, intersection closed) family. Thus union-free and intersection-free families will be enumerated in this section as well.

A map $C : \mathcal{P}([n]) \to \mathcal{P}([n])$ is a *closure operation* on n elements if it is (1) inflationary ($C(X) \supseteq X$), (2) non-decreasing ($X \subseteq Y$ implies $C(X) \subseteq C(Y)$), and (3) idempotent ($C\big(C(X)\big) = C(X)$). If C is a closure operation on n elements, the *closed* sets (the sets X such that $C(X) = X$) form a family of sets closed under intersections and containing $[n]$. Conversely, if F is a family of sets closed under intersections and containing $[n]$, one recovers the original closure operation by taking $C(X)$ to be the intersection of all sets Y in F that such that $X \subseteq Y$. Since $[n]$ can be added to or deleted from an intersection-closed family over n elements without affecting the property of being intersection-closed, the number of closure operations on n elements is just $\alpha(n)/2$. This factor of 2 will be negligible compared with the error terms in our estimates, so closure operations will be enumerated in this section as well.

We shall say that a closure operation C on n elements is *proper* if $C(\emptyset) = \emptyset$. Let $\beta(n)$ denote the number of proper closure operations on n elements. The closure operations C on n elements with $C(\emptyset) = \{k+1, \ldots, n\}$ are in one-to-one correspondence with the proper closure operations C' on k elements via the correspondence $C(X) = C'(X \cap [k]) \cup \{k+1, \ldots, n\}$, $C'(X) = C(X) \cap [k]$. This implies $\alpha(n)/2 = \sum_{0 \leq k \leq n} \binom{n}{k} \beta(k)$, which by the principle of inclusion-exclusion implies $\beta(n) = \sum_{0 \leq k \leq n} \binom{n}{k} (-1)^{n-k} \alpha(k)/2$. It will follow from our estimates that all the terms in this sum except for the one in which $k = n$ are negligible, so that $\beta(n) \sim \alpha(n)/2$. Thus proper closure operations will be asymptotically enumerated in this section as well.

We shall take as the objects to be enumerated the functions satisfying $f(x) \wedge f(y) \leq f(x \vee y)$, since they are the most similar to monotone functions, and will thus require the least modification of the proof in Section 2. We shall call such functions *Horn* functions (though it is their reflections, or the complements of their reflections, that are usually so called).

Borosch *et al.* [B2] have shown that

$$\binom{n}{\lfloor \frac{n}{2} \rfloor} \leq \log_2 \beta(n) \leq 2\sqrt{2} \binom{n}{\lfloor \frac{n}{2} \rfloor} \left(1 + O\left(\frac{\log n}{n}\right)\right),$$

which was improved by Alekseyev [A1] to

$$\log_2 \alpha(n) = \binom{n}{\lfloor \frac{n}{2} \rfloor} \left(1 + O\left(\frac{\log n}{n^{1/4}}\right)\right).$$

We shall the following, slightly weaker, result.

Theorem 3.1: As $n \to \infty$, we have

$$\log_2 \alpha(n) = \binom{n}{\lfloor \frac{n}{2} \rfloor} \left(1 + O\left(\frac{(\log n)^{3/2}}{n^{1/4}}\right)\right).$$

Proof: Set $N = \binom{n}{\lfloor \frac{n}{2} \rfloor}$. The lower bound

$$\log_2 \alpha(n) \geq N$$

is immediate, since every monotone function is a Horn function, so $\log_2 \alpha(n) \geq \log_2 \psi(n) \geq N$. Thus our task is to obtain an asymptotically matching upper bound.

Let $\mathcal{N}_n$ denote the set of all Horn functions of n arguments, and let f be a random function uniformly distributed on $\mathcal{N}_n$. We have

$$\log_2 \alpha(n) = H(f).$$

We shall obtain an upper bound for $H(f)$.

Given f, we shall construct a random variable $\delta = (\mathcal{F}, \mathcal{G}, \tilde{\delta}, \hat{\delta})$ as follows. Say that a point $x \in \{0,1\}^n$ is a *fall* for f if (1) $f(x) = 0$ and (2) the chain C_j that contains x also contains a point y with $\|y\| = \|x\| - 1$ and $f(y) = 1$. We shall let $\mathcal{F} \subseteq \{0,1\}^n$ be the set of falls for f. Say that a point $x \in \{0,1\}^n$ is a *rise* for f if (1) $f(x) = 1$ and (2) if the chain C_j that contains x also contains a point y with $\|y\| = \|x\| - 1$, then $f(y) = 0$. We shall let $\mathcal{G} \subseteq \{0,1\}^n$ be the set of rises for f that appear in chains that also contain falls. Since rises and falls alternate in a chain (proceeding from bottom to top), the random variables $\mathcal{F}$ and $\mathcal{G}$ specify f at all points that appear in chains that contain falls. Thus f is monotone on any chain that does not contain a fall. We can therefore complete the encoding of f by defining $\tilde{\delta}$ and $\hat{\delta}$ as in Section 2, with the following changes: (1) we take $\tilde{\delta}_j = 0$ and $\hat{\delta}_j = 0$ for all j for which the chain C_j contains a fall, and (2) we define $\hat{f}$ to be the smallest Horn function $\hat{f}$ that agrees with f on the chains that contain falls and for which $\gamma_j(\hat{f}) \geq \tilde{\delta}_j$ for all j such that C_j contains no fall.

To obtain an upper bound for $H(\mathcal{F}, \mathcal{G}, \tilde{\delta}, \hat{\delta})$, we shall need a lemma.

Lemma 3.2: Suppose that the random variable $\mathcal{K}$ takes values in $\mathcal{P}([n])$, has cardinality $K = \#\mathcal{K} < 2^n$ and expected cardinality $E = \mathrm{Ex}(K)$. Then

$$H(\mathcal{K}) \leq n + E \log_2 \left(\frac{e2^n}{E} \right).$$

Proof: If $p_k = \Pr(K = k)$, then

$$E = \sum_{0 \leq k < 2^n} p_k \, k$$

and

$$\begin{aligned} H(\mathcal{K}) &= H(K) + H(\mathcal{K} \mid K) \\ &\leq n + \sum_{0 \leq k < 2^n} p_k \log_2 \binom{2^n}{k} \\ &\leq n + \sum_{0 \leq k < 2^n} p_k \, k \log_2 \left(\frac{e2^n}{k} \right). \end{aligned}$$

Since $k \log_2(e2^n/k)$ is convex in k, the inequality of the lemma follows. $\triangle$

To bound $H(\mathcal{F})$ we apply Lemma 3.2 with $\mathcal{K} = \mathcal{F}$. We observe that the point $x = (0, \ldots, 0)$ with $\|x\| = 0$ cannot be a fall, so that $\#\mathcal{F} < 2^n$. To estimate $E = \mathrm{Ex}(\#\mathcal{F})$, we shall estimate the probability that a point x with $\|x\| \geq 1$ is a fall. If x is a fall, then we must have $f(x) = 0$, and there must be exactly one point y with $\|y\| = \|x\| - 1$ for which $f(y) = 1$. (If there were more than one such point, say y and z, then $f(x) = f(y \vee z) = 1$

would follow from $f(y) = 1$ and $f(z) = 1$ for the Horn function f.) The probability that this unique point y is in the same chain as x is (averaging over the orbit of f under $\mathrm{Stab}(x)$) just $1/\|x\|$. Thus we have

$$\begin{aligned} E &= \sum_{1\le k\le n} \frac{1}{k}\binom{n}{k} \\ &\le \sum_{1\le k\le n} \frac{2}{k+1}\binom{n}{k} \\ &= \frac{2}{n+1}\sum_{1\le k\le n}\binom{n+1}{k+1} \\ &\le \frac{4\cdot 2^n}{n+1}. \end{aligned}$$

Since $E\log_2(e2^n/E)$ is increasing in E for $E \le 2^n$, we obtain

$$\begin{aligned} H(\mathcal{F}) &\le n + \frac{4\cdot 2^n}{n+1}\log_2\frac{e(n+1)}{4} \\ &\le \frac{4\cdot 2^n \log_2 n}{n} \end{aligned} \tag{3.1}$$

(for n sufficiently large). Since rises and falls alternate in a chain that contains a fall, we have $\#\mathcal{G} \le 2\#\mathcal{F}$, and thus

$$H(\mathcal{G}) \le \frac{8\cdot 2^n\log_2 n}{n}. \tag{3.2}$$

To complete the proof, we estimate $H(\tilde\delta)$ and $H(\hat\delta)$ as in Section 2. The only modification needed is in the estimate for the probability that a point x that is heavy is also bad. In the present case this can occur only if $v_j = 0$ for all but at most one of the s or more chains that contain points y with $\|y\| = \|x\| - 1$ and for which $f(y) = 1$ (else we would have $\tilde f(x) = 1$, and x would not be bad). If there are $t \ge s$ chains that contain points y with $\|y\| = \|x\| - 1$ and for which $f(y) = 1$, then the probability of this event is $(1-p)^t + tp(1-p)^{t-1} \le (1-p)^s + sp(1-p)^{s-1} \le 1/n$, as before. We can therefore continue as in Section 2 to obtain the estimate

$$H(\tilde\delta,\hat\delta) \le N + \frac{20\cdot 2^n(\log_2 n)^{3/2}}{n^{3/4}}.$$

Combining this with the estimates (3.1) and (3.2) yields

$$H(\delta) \le N + \frac{32\cdot 2^n(\log_2 n)^{3/2}}{n^{3/4}}.$$

Since $N \sim (2/\pi n)^{1/2}\, 2^n > 2^n/2n^{1/2}$ (for n sufficiently large), the proof of Theorem 3.1 is complete.

4. Conclusion

The results of Sections 2 and 3 show how the notion of entropy can be used to simplify the proofs of some results in combinatorial enumeration. Our proofs have departed from the pattern given by Massey in one important respect: whereas in Massey's proof, the "encoding" of an object to be enumerated is a deterministic function of the object, in our proofs the encoding is constructed by means of "auxiliary randomization". This randomization occurs through the random variables $\upsilon_1, \ldots, \upsilon_N$, and it is responsible for the need to use the inequality (1.2). It would be possible to eliminate this auxiliary randomization in the following way. We could choose a *fixed* partition of the chains into two classes: a class $\mathcal{P}$ of chains C_j for which the "absolute" parameters $\gamma_j(f)$ are specified, and a class $\mathcal{Q}$ for which the "relative" parameters $\gamma_j(f) - \gamma_j(\tilde{f})$ are specified. This partition must have two properties: (1) any chain that contains a low point must appear in $\mathcal{P}$, and (2) for any point x that appears in a chain in $\mathcal{Q}$, there must be at least r points y with $\|y\| = \|x\| - 1$ that appear in chains in $\mathcal{P}$, where $r = 2n^{3/4}(\log_2 n)^{1/2}$. We want $\mathcal{P}$ to be as small as possible. A standard argument involving choosing chains at random, shows that (1) and (2) can be satisfied by some $\mathcal{P}$ containing $O(Nr/n) = O\big(2^n(\log n)^{1/2}/n^{3/4}\big)$ chains. Then, to bound the probability that a heavy point x is bad, we again average f over orbits of $\mathrm{Stab}(x)$ (as we did for x not heavy). The argument yields the same order of error term as the one obtained in Sections 2 and 3, but the proof is if anything slightly more complicated, which is why we have used auxiliary randomization.

It may also be worthwhile to compare our proofs with the original proofs given by Kleitman [K1] and Alekseyev [A1]. Their proofs encode an arbitrary function of the relevant type, but choose the encoding from a set of alternative encodings so as to minimize the length of the encoding. The argument bounds the minimum by the average over the set of alternative encodings, and this averaging plays a role similar to our auxiliary randomization. The averaging occurs in two parts of the proof. First, the function encoded is not necessarily the arbitrarily given function f, but rather a version $\sigma(f)$ of f with its arguments permuted. In the original proofs, this requires σ to be encoded as well as $\sigma(f)$, which increases the bound in a negligible way. In our proofs, the averaging over orbits of f under the groups $\mathrm{Stab}(x)$ plays the same role, and does not occasion any increase in the bound. Second, in the original proofs, the chains are partitioned into "blocks", and averaging is done over permutations of the blocks. This requires that a permutation of the blocks be encoded, and gives rise to another negligible increase in the bound. This averaging is eliminated entirely in our proofs, being replaced by the "sampling" effected by the random variables $\upsilon_1, \ldots, \upsilon_N$ (in the form we have given, with auxiliary randomization), or by the fixed partition $\mathcal{C} = \mathcal{P} \cup \mathcal{Q}$ described in the preceding paragraph. A precursor of this sampling appears in the proof of Alekseyev [A1], as well as in an earlier proof of Andreyev [A2] of a result concerning the computational complexity of monotone Boolean functions (which implicitly gives another proof of (2.1)).

5. References

[A1] V. B. Alekseyev, "O Chisle Semeĭstv Podmnozhestv, Zamknutykh Otnositel'no Peresecheniya", *Diskr. Mat.*, 1 (1989) 129–136.

[A2] A. E. Andreyev, "O Slozhnosti Monotonnykh Funktsiĭ", *Vestnik Moskov. Univ. Ser. I*, (1985) 83–87.

[B1] N. G. de Bruijn, C. van Ebbenhorst Tengbergen and D. Kruyswijk, "On the Set of Divisors of a Number", *Nieuw Arch. Wiskunde*, 23 (1951) 191–193.

[B2] G. Burosch, J. Demetrovics, G. O. H. Katona, D. J. Kleitman and A. A. Sapozhenko, "On the Number of Databases and Closure Operations", *Theoretical Computer Science*, 78 (1991) 377–381.

[C] F. R. K. Chung, R. L. Graham, P. Frankl and J. B. Shearer, "Some Intersection Theorems for Ordered Sets and Graphs", *J. Combinatorial Theory A*, 43 (1986) 23–37.

[D] R. Dedekind, "Über Zerlegungen von Zahlen durch ihre grössten geminsamen Teiler", *Festschrift Hoch. Braunschweig*, 2 (1897) 103–148.

[G] C. Greene and D. J. Kleitman, "Strong Versions of Sperner's Theorem", *J. Combinatorial Theory (A)*, 20 (1976) 80–88.

[H] A. Horn, "On Sentences Which Are True of Direct Unions of Algebras", *J. Symbolic Logic*, 16 (1951) 14–21.

[J] F. Jelinek, *Probabilistic Information Theory*, McGraw-Hill, 1968.

[K1] D. Kleitman, "On Dedekind's Problem: The Number of Monotone Boolean Functions", *Proc. Amer. Math. Soc.*, 21 (1969) 677–682.

[K2] A. D. Korshunov, "O Chisle Monotonnykh Bulevykh Funktsiĭ", *Problemy Kibernetiki*, 38 (1980) 5–108.

[L] B. Lindström, "On a Combinatory Detection Problem", *Publ. Math. Inst. Hungarian Acad. Sci.*, 9 (1964) 195–207.

[M] J. L. Massey, "On the Fractional Weight of Distinct Binary n-Tuples", *IEEE Trans. on Info. Theory*, 20 (1974) 131.

[N] I. Newman and A. Wigderson, "Lower Bounds on Formula Size of Boolean Functions Using Hypergraph Entropy", *SIAM J. Discrete Math.*, 8 (1995) 536–542.

[P] N. Pippenger, "An Information-Theoretic Method in Combinatorial Theory", *J. Combinatorial Theory A*, 23 (1977) 99–104.

[S] C. E. Shannon, "A Mathematical Theory of Communication", *Bell System Tech. J.*, 27 (1948) 379–423, 623–656.

A. Appendix

The following description of a partition of $\{0,1\}^n$ into $N = \binom{n}{\lfloor \frac{n}{2} \rfloor}$ chains is due to Greene and Kleitman [G], who attribute the underlying partition to de Bruijn, van Ebbenhorst Tengbergen and Kruyswijk [B1]. Imagine that in each Boolean n-tuple $(x_1, \ldots, x_n)$, 0s are replaced by left parentheses and 1s by right parentheses. In any such sequence of parentheses, regard a left parenthesis as matched with a following right parentheses if they are adjacent, or if any intervening parentheses are matched (by recursive application of this rule). There may then be some unmatched parentheses, but any unmatched right parentheses must appear before any unmatched left parentheses, else some additional pair would match. Partition the n-tuples into classes by putting into the same class all n-tuples that have the same matching parentheses in the same positions. Then all n-tuples in the same class form a chain, ordered from smallest (corresponding to all unmatched parentheses being left parentheses) to largest (corresponding to all unmatched parentheses being right parentheses). And there are exactly N such chains, since each chain contains just one n-tuple with $\lfloor \frac{n}{2} \rfloor$ right (and $\lceil \frac{n}{2} \rceil$ left) parentheses.