Maximizing the Entropy of a Sum of Independent Random Variables

Erick Ordentlich
Computer Systems Laboratory
HP Laboratories Palo Alto
HPL-1999-120
October, 1999

majorization, multiple access channel, time sharing

Let $X_1, \ldots, X_n$ be n independent, symmetric random variables supported on the interval [-1,1] and let $S_n = \sum_{i=1}^{n} X_i$ be their sum. We show that the differential entropy of S_n is maximized when $X_1, \ldots, X_{n-1}$ are Bernoulli taking on +1 or -1 with equal probability and X_n is uniformly distributed. This entropy maximization problem is due to Shlomo Shamai [1] who also conjectured the solution[1].

Internal Accession Date Only

Maximizing the Entropy of a Sum of Independent Random Variables

Erik Ordentlich

October 8, 1999

Abstract

Let $X_1, \ldots, X_n$ be n independent, symmetric random variables supported on the interval [-1,1] and let $S_n = \sum_{i=1}^{n} X_i$ be their sum. We show that the differential entropy of S_n is maximized when $X_1, \ldots, X_{n-1}$ are Bernoulli taking on +1 or -1 with equal probability and X_n is uniformly distributed. This entropy maximization problem is due to Shlomo Shamai [1] who also conjectured the solution[1].

1 Introduction

The differential entropy $h(S)$ of a real valued random variable S is defined as

$$h(S) = -\int_{-\infty}^{\infty} f_S(s) \log f_S(s) ds, \tag{1}$$

where $f_S(s)$ is the density of S with respect to Lebesgue measure. If S has no density then $h(S)$ is taken to be $-\infty$. We obtain a maximum entropy result for the sum of bounded independent symmetric random variables. Specifically, let $X_1, \ldots, X_n$ be any independent random variables symmetrically distributed about zero and bounded between -1 and 1. Let $Z_1, \ldots, Z_{n-1}$ be i.i.d. Bernoulli taking on 1 and -1 with equal probability and U be independent of $Z_1, \ldots, Z_{n-1}$ and uniformly distributed on [-1,1]. We show that the sums

$$S_n = \sum_{i=1}^{n} X_i \text{ and } S_n^* = U + \sum_{i=1}^{n-1} Z_i \tag{2}$$

satisfy

$$h(S_n) \leq h(S_n^*). \tag{3}$$

[1]Shamai conjectured that the present result holds even without the symmetry assumption.

The density of S_n^* is symmetric about zero and piecewise constant. It is given by (up to a set of Lebesgue measure zero)

$$f_{S_n^*}(s) = \begin{cases} 2^{-n} & \text{if} \quad s \in (-n, -n+2) \\ & \vdots \\ \binom{n-1}{j} 2^{-n} & \text{if} \quad s \in (-n+2j, -n+2j+2) \\ & \vdots \\ 2^{-n} & \text{if} \quad s \in (n-2, n) \end{cases} \tag{4}$$

where j runs from 0 to $n-1$.

The differential entropy of S_n^* computed according to (1) from the above expression for $f_{S_n^*}$ simplifies to

$$h(S_n^*) = H(s_0^*, \ldots, s_{n-1}^*) + 1, \tag{5}$$

where

$$H(s_0^*, \ldots, s_{n-1}^*) \triangleq -\sum_{j=0}^{n-1} s_j^* \log s_j^*, \tag{6}$$

and

$$s_j^* = Pr(S_n^* \in [-n+2j, -n+2j+2]) = \binom{n-1}{j} 2^{-(n-1)}, \tag{7}$$

for $j = 0, \ldots, n-1$. Note that $\mathbf{s}^* \triangleq (s_0^*, \ldots, s_{n-1}^*)$ is simply the $(n-1)^{\text{th}}$ order binomial 1/2 probability distribution and that $H(\mathbf{s}^*)$ is the discrete entropy of this distribution.

For $n = 1$ and $n = 2$ it is obvious that $h(S_n^*)$ is maximal. In these cases S_n^* is uniformly distributed on $[-1, 1]$ and $[-2, 2]$ respectively (which are also the support sets of S_n) and it is well known that the differential entropy of a random variable supported on $[-a, a]$ is indeed maximal when it is uniformly distributed.

For $n > 2$ the result is less obvious. The key step in our proof is the following lemma.

Lemma 1 *If $Z_1, \ldots, Z_n$ are i.i.d. Bernoulli taking on values +1 and -1 with equal probability, and if $a_1, \ldots, a_n$ satisfy $0 \le a_i \le 1$, then*

$$Pr(\sum_{i=1}^{n} Z_i a_i \in [-n+2j, n-2j]) \ge Pr(\sum_{i=1}^{n-1} Z_i \in [-n+2j, n-2j]) \tag{8}$$

for all integers j satisfying $0 \le j \le J$ where J is the largest integer such that $n - 2j > 0$.

We prove the result in two parts. First, in Section 2, we reduce the problem to Lemma 1. We then prove Lemma 1 in Section 3.

Motivation The more general problem obtained by removing the symmetry constraints on the X_i is motivated as follows [1]. The problem is related to the maximum achievable throughput (sum of transmission rates of all users) of a peak power constrained additive noise multiple access channel in the limit of low noise power. Specifically, consider an n user multiple access channel with inputs $X_1, \ldots, X_n$ and output $Y = \sum_{i=1}^n X_i + Z$, where Z is Gaussian (for example) with variance σ^2 and the inputs X_i are peak constrained to lie in the intervals $[-1, 1]$. The maximum throughput is obtained by maximizing the mutual information $I(Y; X_1, \ldots, X_n)$ under the constraint that the X_i are independent and supported on the intervals $[-1, 1]$. Since the mutual information can be written as

$$
\begin{aligned}
I(Y; X_1, \ldots, X_n) &= h(Y) - h(Y|X_1, \ldots, X_n) \qquad (9)\\
&= h(Y) - h(Z) \qquad (10)
\end{aligned}
$$

where $h(Y)$ and $h(Z)$ are the differential entropies of Y and Z, the problem reduces to maximizing the differential entropy $h(Y) = h(\sum_{i=1}^n X_i + Z)$ subject to the above constraints on the X_i. In the limit of σ^2 tending to zero, the solution converges to that of simply maximizing $h(\sum_{i=1}^n X_i)$. Of course in this limit the maximum mutual information tends to infinity. The interesting point, however, is that since the maximizing distribution on the X_i is highly asymmetric in the limit, namely one X_i is uniform and the other X_i are Bernoulli, a similar asymmetry must also exist for sufficiently small σ^2. This in turn means that for these σ^2, time sharing *must* be used to achieve the maximum possible throughput with equal transmission rates for all users. In contrast, for the *average* power constrained Gaussian multiple access channel the symmetric maximum throughtput can be achieved without time sharing.

It is conjectured that S_n^* maximizes the differential entropy even without the symmetry constraint on the X_i, however, we have not been able to prove this. This would follow from another conjecture that given any independent random variables $X_1, \ldots, X_n$,

$$h(\sum_{i=1}^n X_i) \leq h(\sum_{i=1}^n Z_i X_i), \qquad (11)$$

where $Z_1, \ldots, Z_n$ are Bernoulli taking on +1 and -1 with equal probability. The result of [2] showing that the discrete entropy of a sum of independent Bernoullis is maximized when they are symmetric suggests the validity of this conjecture.

2 Reduction to Lemma 1

This first part of the proof involves a majorization relationship. Given two probability distributions $\mathbf{p} = (p_0 \geq p_2 \geq \ldots \geq p_m)$ and $\mathbf{q} = (q_0 \geq \ldots \geq q_m)$, $\mathbf{p}$ majorizes $\mathbf{q}$ if for all k

$$\sum_{i=0}^k p_i \geq \sum_{i=0}^k q_i. \qquad (12)$$

The consequence of majorization we use is that if $\mathbf{p}$ majorizes $\mathbf{q}$ then $H(\mathbf{p}) \leq H(\mathbf{q})$ where $H(\cdot)$ is the discrete entropy. See [3] for this and many other results concerning majorization.

Recall that S_n is the sum of n independent random variables X_i which are symmetric and supported on $[-1,1]$. Assume S_n has a density and define the probability distribution $\mathbf{s} = (s_0, \ldots, s_{n-1})$ by setting

$$s_j = Pr(S_n \in [-n+2j, -n+2j+2]) \tag{13}$$

for $j = 0, \ldots, n-1$. The concavity of the function $-x \log x$ and Jensen's inequality imply that the differential entropy of S_n satisfies

$$h(S_n) \leq H(\mathbf{s}) + 1. \tag{14}$$

Recall the distribution $\mathbf{s}^*$ defined in (7). It now suffices to show that $H(\mathbf{s}) \leq H(\mathbf{s}^*)$. Specifically, this relation together with (5) and (14) imply

$$\begin{aligned} h(S_n) &\leq H(\mathbf{s}) + 1 && (15)\\ &\leq H(\mathbf{s}^*) + 1 && (16)\\ &= h(S_n^*), && (17) \end{aligned}$$

thereby proving that $h(S_n) \leq h(S_n^*)$.

To prove $H(\mathbf{s}) \leq H(\mathbf{s}^*)$, we show that for all $\mathbf{s}$ constructed as above, $\mathbf{s}$ majorizes $\mathbf{s}^*$. Our task is simplified slightly by the following lemma.

Lemma 2 *Let J be the largest integer j such that $j \leq n-1-j$. If for $\mathbf{s}^*$ and $\mathbf{s}$ defined in (7) and (13)*

$$\sum_{l=j}^{n-1-j} s_l \geq \sum_{l=j}^{n-1-j} s_l^* \tag{18}$$

for all $j \leq J$, then $\mathbf{s}$ majorizes $\mathbf{s}^*$.

Proof: First note that

$$(s_J^* = s_{n-1-J}^*) \geq \ldots \geq (s_1^* = s_{n-2}^*) \geq (s_0^* = s_{n-1}^*) \tag{19}$$

and that

$$(s_J = s_{n-1-J}), \ \ldots, \ (s_1 = s_{n-2}), \ (s_0 = s_{n-1}). \tag{20}$$

In view of this, (18) holding for all j is almost the definition of majorization.

One problem is that (18) does not account for all the partial sums appearing in the definition of majorization, which additionally requires

$$\sum_{l=j}^{n-1-j-1} s_l \geq \sum_{l=j}^{n-1-j-1} s_l^*, \tag{21}$$

for $j \leq J$ where both sums are taken to be zero if $n-2-j < j$. This is resolved as follows. For $j \leq J-1$ and $j = J$ when $J < n-1-J$ we have from (18)

$$\sum_{l=j}^{n-1-j} s_l \geq \sum_{l=j}^{n-1-j} s_l^*, \tag{22}$$

and

$$\sum_{l=j+1}^{n-1-j-1} s_l \geq \sum_{l=j+1}^{n-1-j-1} s_l^*, \tag{23}$$

so that

$$\frac{1}{2}\left(\sum_{l=j}^{n-1-j} s_l + \sum_{l=j+1}^{n-1-j-1} s_l\right) \geq \frac{1}{2}\left(\sum_{l=j}^{n-1-j} s_l^* + \sum_{l=j+1}^{n-1-j-1} s_l^*\right), \tag{24}$$

which simplifies to

$$\frac{1}{2}\left(s_j + s_{n-1-j}\right) + \sum_{l=j+1}^{n-1-j-1} s_l \geq \frac{1}{2}\left(s_j^* + s_{n-1-j}^*\right) + \sum_{l=j+1}^{n-1-j-1} s_l^*. \tag{25}$$

This is equivalent to (21), since $s_j = s_{n-1-j}$ and $s_j^* = s_{n-1-j}^*$.

Note that the $(s_j = s_{n-1-j})$'s may not be in the same decreasing order as the $(s_j^* = s_{n-1-j}^*)$'s. This is not an issue, however, since if the $(s_j = s_{n-1-j})$ are rearranged in decreasing order, the inequalities (18) and (21) would continue to hold. □

Recall how the probability distributions $\mathbf{s}$ and $\mathbf{s}^*$ are related to the random variables S_n and S_n^*. Proving (18) would show that in this sense S_n^* is less densely distributed about the origin than any S_n obtained as the sum of independent random variables satisfying the boundedness and symmetry assumptions. The random variable S_n^* is maximally "spread out" and hence should have the highest entropy.

A key part of the proof of (18) follows from the symmetry and independence of the random variables X_i. Specifically, conditioned on their absolute values $|X_1| = a_1, \ldots, |X_n| = a_n$, the X_i are independent Bernoulli random variables taking on $-a_i$ and a_i with probability 1/2. This implies the following characterization of $\mathbf{s}$; that

$$\begin{aligned} s_j &= Pr(S_n \in [-n+2j, -n+2j+2]) & (26)\\ &= E\left[Pr(S_n \in [-n+2j, -n+2j+2] \,|\, |X_1| = a_1, \ldots, |X_n| = a_n)\right]. & (27) \end{aligned}$$

Further, from the above observation about the X_i,

$$Pr(S_n \in [-n+2j, -n+2j+2] \,|\, |X_1| = a_1, \ldots, |X_n| = a_n) \tag{28}$$

$$= \; Pr(\sum_{i=1}^{n} Z_i a_i \in [-n+2j, -n+2j+2]), \tag{29}$$

where the Z_i are i.i.d. Bernoulli +1 or -1 with probability 1/2.

The distribution $\mathbf{s}^*$ (7) can also be expressed in terms of the Z_i as

$$s_j^* = Pr(\sum_{i=1}^{n-1} Z_i = -n+2j+1) = Pr(\sum_{i=1}^{n-1} Z_i \in [-n+2j, -n+2j+2]). \tag{30}$$

Therefore, if we could show that

$$Pr(\sum_{i=1}^{n} Z_i a_i \in [-n+2j, n-2j]) \geq Pr(\sum_{i=1}^{n-1} Z_i \in [-n+2j, n-2j]), \tag{31}$$

then (18) would follow by taking expectations of both sides of (31) with respect to the a_i distributed[2] as $|X_i|$. The validity of (31), however, is precisely the claim of Lemma 1 from the introduction. The lemma states that if $Z_1, \ldots, Z_n$ are i.i.d. Bernoulli taking on +1 and -1 with equal probability and $a_1, \ldots, a_n$ satisfy $0 \leq a_i \leq 1$, then (31) holds for all integers j satisfying $0 \leq j \leq J$ where J is the largest j for which $n - 2j > 0$ (or $j \leq n - 1 - j$).

Thus the proof of $h(S_n) \leq h(S_n^*)$ has been reduced to proving Lemma 1, which is carried out in the next section.

3 Proof of Lemma 1

First note that the left hand side of (31) is independent of the ordering of the a_i. Therefore, we will assume that $1 \geq a_1 \geq a_2 \geq \ldots \geq a_n \geq 0$.

The symmetry of the Z_i allows us to prove

$$Pr(\sum_{i=1}^{n} Z_i a_i \in [-n+2j, n-2j]) \geq Pr(\sum_{i=1}^{n-1} Z_i \in [-n+2j, n-2j]) \tag{32}$$

by showing

$$Pr(\sum_{i=1}^{n} Z_i a_i > n-2j) \leq Pr(\sum_{i=1}^{n-1} Z_i > n-2j). \tag{33}$$

[2]The right hand side of (31) is independent of the a_i.

Based on these observations, the validity of the lemma in the special case of $n = 3$ is easily verified. We need only show that $Pr(Z_1 a_1 + Z_2 a_2 + Z_3 a_3 > 1) \leq 1/4$. This follows easily since, assuming $1 \geq a_1 \geq a_2 \geq a_3 \geq 0$, it is obvious that if $\sum_{i=1}^{3} Z_i a_i > 1$ then $Z_1 = 1$ and $Z_2 = 1$ and the probability of this is $1/4$.

We now proceed with the general case. Let

$$\mathcal{S}_j = \{\mathbf{z}^{n-1} : \sum_{i=1}^{n-1} z_i > n - 2j\} \subset \{+1, -1\}^{n-1}, \tag{34}$$

and

$$\mathcal{S}_{j,\mathbf{a}} = \{\mathbf{z}^n : \sum_{i=1}^{n} z_i a_i > n - 2j)\} \subset \{+1, -1\}^n, \tag{35}$$

where $\mathbf{z}^{n-1} = z_1, \ldots, z_{n-1}$ and $\mathbf{z}^n = z_1, \ldots, z_n$ denote sequences of +1 and -1. Then

$$Pr(\sum_{i=1}^{n-1} z_i > n - 2j) = \frac{|\mathcal{S}_j|}{2^{n-1}}, \tag{36}$$

and

$$Pr(\sum_{i=1}^{n} z_i a_i > n - 2j) = \frac{|\mathcal{S}_{j,\mathbf{a}}|}{2^n}. \tag{37}$$

Therefore, the lemma is equivalent to

$$|\mathcal{S}_{j,\mathbf{a}}| \leq 2|\mathcal{S}_j| \tag{38}$$

for all $0 \leq j \leq J$. Note that there is nothing to prove for $j = 0$ since both $\mathcal{S}_0$ and $\mathcal{S}_{0,\mathbf{a}}$ are empty. Also note for future reference that $\mathcal{S}_j$ consists of all $\mathbf{z}^{n-1}$ with $j - 1$ or fewer -1's and therefore its cardinality is

$$|\mathcal{S}_j| = \sum_{k=0}^{j-1} \binom{n-1}{k}. \tag{39}$$

.

The proof of (38) is outlined as follows. First we find disjoint sets $\mathcal{S}'_{j,\mathbf{a}}$ and $\mathcal{S}''_{j,\mathbf{a}}$ satisfying

$$\mathcal{S}_{j,\mathbf{a}} \subset \mathcal{S}'_{j,\mathbf{a}} \cup \mathcal{S}''_{j,\mathbf{a}}, \tag{40}$$

and

$$|\mathcal{S}'_{j,\mathbf{a}}| = 2|\mathcal{S}_j|. \tag{41}$$

Then we exhibit a map ϕ with domain $\mathcal{S}''_{j,\mathbf{a}}$ and range $\mathcal{S}'_{j,\mathbf{a}}$ which is one to one and has the property that if $\mathbf{z}^n \in \mathcal{S}_{j,\mathbf{a}}$ then $\phi(\mathbf{z}^n) \notin \mathcal{S}_{j,\mathbf{a}}$. Therefore

$$\mathcal{S}_{j,\mathbf{a}} \subset \left(\mathcal{S}'_{j,\mathbf{a}} \setminus \phi(\mathcal{S}_{j,\mathbf{a}} \cap \mathcal{S}''_{j,\mathbf{a}})\right) \cup \left(\mathcal{S}_{j,\mathbf{a}} \cap \mathcal{S}''_{j,\mathbf{a}}\right), \tag{42}$$

and since ϕ is one to one, the cardinality of the set on the right equals the cardinality of $\mathcal{S}'_{j,\mathbf{a}}$ which by (41) is $2|\mathcal{S}_j|$ thereby proving (38).

Part 1: In the first part of the proof we specify the sets $\mathcal{S}'_{j,\mathbf{a}}$ and $\mathcal{S}''_{j,\mathbf{a}}$. Consider for the moment the set of sequences $\mathbf{z}^n$ with exactly k -1's. Denote this set by

$$T_k = \{\mathbf{z}^n : \sum_{i=1}^{n} I(z_i = -1) = k\}, \tag{43}$$

where $I(\cdot)$ is the indicator function. Since the $a_i \leq 1$, a sequence having $2j$ or more -1's can not satisfy

$$\sum_{i=1}^{n} z_i a_i > n - 2j, \tag{44}$$

implying that

$$\mathcal{S}_{j,\mathbf{a}} \subset \bigcup_{k=0}^{2j-1} T_k \tag{45}$$

For a sequence $\mathbf{z}^n \in T_k$ let $i_1 < \ldots < i_k$ denote the k indices i for which $z_i = -1$. For such an index i_r, the number of indices $i > i_r$ for which $z_i = +1$ is exactly $n - i_r - (k - r)$. Denote this number by w_r so that

$$w_r = n - i_r - k + r \tag{46}$$

where the dependence on $\mathbf{z}^n$ is assumed.

If, for a sequence $\mathbf{z}^n \in T_k$ and $m \leq k$, it is the case that

$$w_1 \geq m,\ w_2 \geq m - 1,\ \ldots,\ w_m \geq 1, \tag{47}$$

then

$$\sum_{i=1}^{n} z_i a_i \leq n - k - m. \tag{48}$$

The reason is that (47) implies the existence of m distinct indices $i'_1 < \ldots < i'_m$ satisfying $i_r < i'_r$ for $1 \leq r \leq m$ and for which $z_{i'_r} = 1$. Letting $\mathcal{S} = \{i_1, \ldots, i_k, i'_1, \ldots, i'_m\}$ we then have

$$\begin{aligned} \sum_{i=1}^{n} z_i a_i &= \sum_{i \notin \mathcal{S}} z_i a_i + \sum_{r=1}^{m} (a_{i'_r} - a_{i_r}) - \sum_{r=m+1}^{k} a_{i_r} && (49) \\ &\leq \sum_{i \notin \mathcal{S}} a_i && (50) \\ &\leq n - k - m, && (51) \end{aligned}$$

where we use that $a_{i'_r} \leq a_{i_r}$ for each r.

Therefore, any $\mathbf{z}^n$ in T_k for $k \geq j$ which is also in $\mathcal{S}_{j,\mathbf{a}}$ must have

$$w_1 < 2j-k \ \text{ or } \ w_2 < 2j-k-1 \ \text{ or } \ \ldots \ \text{ or } \ w_{2j-k} < 1, \tag{52}$$

since otherwise by (48)

$$\begin{aligned} \sum_{i=1}^{n} z_i a_i &\leq n-k-(2j-k) & (53)\\ &= n-2j. & (54)\end{aligned}$$

Since the w_r's are integers, the constraints (52) are equivalent to

$$w_1 \leq 2j-k-1 \ \text{ or } \ w_2 \leq 2j-k-2 \ \text{ or } \ \ldots \ \text{ or } \ w_{2j-k} \leq 0, \tag{55}$$

with the general form that for some $1 \leq r \leq 2j-k$

$$w_r \leq 2j-k-r. \tag{56}$$

The constraint (55) yields the inclusion

$$\mathcal{S}_{j,\mathbf{a}} \subset \left(\bigcup_{k=0}^{j-1} T_k\right) \cup \left(\bigcup_{k=j}^{2j-1} \{\mathbf{z}^n : \mathbf{z}^n \in T_k \text{ and } w_r \leq 2j-k-r \text{ for some } 1 \leq r \leq 2j-k\}\right) \tag{57}$$

which is tighter than (45).

To obtain $\mathcal{S}'_{j,\mathbf{a}}$ and $\mathcal{S}''_{j,\mathbf{a}}$, we need to further decompose the sets

$$\mathcal{Z}_{j,k} \triangleq \{\mathbf{z}^n : \mathbf{z}^n \in T_k \, \text{and } w_r \leq 2j-k-r \text{ for some } 1 \leq r \leq 2j-k\}. \tag{58}$$

For a particular $\mathbf{z}^n$ in this set, let r^* be the largest r for which $w_r \leq 2j-k-r$. Therefore $w_r > 2j-k-r$ for all $r > r^*$ and $w_{r^*} \leq 2j-k-r^*$. The dependence of r^* on $\mathbf{z}^n$, j, and k is implicit. The set $\mathcal{Z}_{j,k}$ can now be decomposed based on the value of r^*. Specifically express $\mathcal{Z}_{j,k}$ as the disjoint union of the sets

$$\mathcal{Z}_{j,k,l} \triangleq \{\mathbf{z}^n : \mathbf{z}^n \in \mathcal{Z}_{j,k} \text{ and } r^* = l\}, \tag{59}$$

so that

$$\mathcal{Z}_{j,k} = \bigcup_{l=1}^{2j-k} \mathcal{Z}_{j,k,l}. \tag{60}$$

The set $\mathcal{Z}_{j,j,j}$ is particularly simple. Recall that $\mathcal{Z}_{j,j,j}$ consists of all sequences $\mathbf{z}^n$ with j -1's and $r^* = j$, meaning that $w_j \leq 2j-j-j = 0$. Therefore, the last -1 (the j^{th}) must occur at index n or $z_n = -1$. There are no constraints on the indices of the other $j-1$ -1's. The cardinality of $\mathcal{Z}_{j,j,j}$ is easily computed as

$$|\mathcal{Z}_{j,j,j}| = \binom{n-1}{j-1}. \tag{61}$$

The first part of the proof is completed by identifying $\mathcal{S}'_{j,\mathbf{a}}$ as

$$\mathcal{S}'_{j,\mathbf{a}} = \left(\bigcup_{k=0}^{j-1} T_k\right) \cup \mathcal{Z}_{j,j,j} \tag{62}$$

and $\mathcal{S}''_{j,\mathbf{a}}$ as

$$\begin{aligned} \mathcal{S}''_{j,\mathbf{a}} &= \left(\bigcup_{k=j+1}^{2j-1} \mathcal{Z}_{j,k}\right) \cup \left(\mathcal{Z}^c_{j,j,j} \cap \mathcal{Z}_{j,j}\right) && (63) \\ &= \left(\bigcup_{k=j+1}^{2j-1} \bigcup_{l=1}^{2j-k} \mathcal{Z}_{j,k,l}\right) \cup \left(\bigcup_{l=1}^{j-1} \mathcal{Z}_{j,j,l}\right). && (64) \end{aligned}$$

The cardinality property (41) of $\mathcal{S}'_{j,\mathbf{a}}$ follows from

$$\begin{aligned} |\mathcal{S}'_{j,\mathbf{a}}| &= \left|\left(\bigcup_{k=0}^{j-1} T_k\right) \cup \mathcal{Z}_{j,j,j}\right| && (65) \\ &= \sum_{k=0}^{j-1} \binom{n}{k} + \binom{n-1}{j-1} && (66) \\ &= \sum_{k=0}^{j-1} \binom{n-1}{k} + \binom{n-1}{k-1} + \binom{n-1}{j-1} && (67) \\ &= 2\sum_{k=0}^{j-1} \binom{n-1}{k} && (68) \\ &= 2|\mathcal{S}_j|, && (69) \end{aligned}$$

where it is understood that $\binom{n-1}{-1} = 0$.

Part 2: The second part of the proof is the construction of the one to one map ϕ with domain $\mathcal{S}''_{j,\mathbf{a}}$, range $\mathcal{S}'_{j,\mathbf{a}}$ and the property that if

$$\sum_{i=1}^{n} a_i z_i > n - 2j \tag{70}$$

then $\hat{\mathbf{z}}^n = \phi(\mathbf{z}^n)$ satisfies

$$\sum_{i=1}^{n} a_i \hat{z}_i \leq n - 2j. \tag{71}$$

The construction of ϕ uses the following lemma.

Lemma 3 *Given a set of m elements $\mathcal{S} = \{1, 2, \ldots, m\}$ and $l \leq \lfloor m/2 \rfloor$ there exists a one to one map $\psi_{l,m}$ from the collection of subsets of size l to itself with the property that $\psi_{l,m}(\mathcal{A})$ is contained in the complement of $\mathcal{A}$.*

The validity of this lemma can be seen by considering the bi-partite graph with left and right vertex sets corresponding to the subsets of size l of $\mathcal{S}$, and edge set consisting of $\{(\mathcal{A}_l, \mathcal{A}_r) : \mathcal{A}_l \subset \mathcal{A}_r^c\}$. The resulting graph is regular, and therefore exhibits a complete matching (See, for example, [4]). We can thus let $\psi_{l,m}$ correspond to such a complete matching.

We now define ϕ. First, for a particular $\mathbf{z}^n \in \mathcal{Z}_{j,k,l} \subset \mathcal{S}''_{j,\mathbf{a}}$, let

$$\mathcal{S} = \{i_1, i_2, \ldots, i_{l-1}, i_l\}, \tag{72}$$

denote the first $l \leq 2j - k$ indices i where $z_i = -1$. Let ψ_{l,i_l} be the map provided by Lemma 3 (we will show that $l \leq \lfloor i_l/2 \rfloor$) sending subsets of size l of $\{1, 2, \ldots, i_l\}$ into subsets of size l such that $\psi_{l,i_l}(\mathcal{A}) \subset \mathcal{A}^c$.

Definition 1 *Define $\hat{\mathbf{z}}^n = \phi(\mathbf{z}^n)$ as*

- $\hat{z}_i = -z_i$ *for $i \geq i_l + 1$ and for $i \in \mathcal{S} \bigcup \psi_{l,i_l}(\mathcal{S})$.*
- $\hat{z}_i = z_i$ *for all other i.*

To show that ϕ is well defined we must show that for all $\mathcal{Z}_{j,k,l} \subset \mathcal{S}''_{j,\mathbf{a}}$, $l \leq \lfloor i_l/2 \rfloor$. To see this, note first that from (64) the set of (k, l) for which $\mathcal{Z}_{j,k,l} \subset \mathcal{S}''_{j,\mathbf{a}}$ is given by

$$\{k, l : (k \geq j + 1, 1 \leq l \leq 2j - k) \mathbf{or} (k = j, l \leq k - 1)\} \tag{73}$$

so that $l \leq k - 1$ in all cases. Therefore it makes sense to talk about i_{l+1} and w_{l+1}. Now note that from the definition of $\mathcal{Z}_{j,k,l}$ (59), w_l and w_{l+1} satisfy

$$w_l \leq 2j - k - l \text{ and } w_{l+1} \geq 2j - k - l \tag{74}$$

and since by definition $w_l \geq w_{l+1}$, it must be that

$$w_l = w_{l+1} = 2j - k - l. \tag{75}$$

On the other hand, from (46) we have $w_l = n - i_l - k + l$, so that

$$\begin{aligned} i_l &= n - w_l - k + l && (76)\\ &= n - (2j - k - l) - k + l && (77)\\ &= n - 2j + 2l && (78)\\ &> 2l, && (79) \end{aligned}$$

thereby proving that $l \leq \lfloor i_l/2 \rfloor$.

We now have to prove that ϕ has the desired properties which are

1. $\phi(\mathcal{S}''_{j,\mathbf{a}}) \subset \mathcal{S}'_{j,\mathbf{a}}$
2. ϕ is one to one from $\mathcal{S}''_{j,\mathbf{a}}$ to $\mathcal{S}'_{j,\mathbf{a}}$.
3. If $\sum_{i=1}^{n} a_i z_i > n - 2j$ then $\hat{\mathbf{z}}^n = \phi(\mathbf{z}^n)$ satisfies $\sum_{i=1}^{n} a_i \hat{z}_i \leq n - 2j$.

The first property, that $\phi(\mathcal{S}''_{j,\mathbf{a}}) \subset \mathcal{S}'_{j,\mathbf{a}}$, follows from two facts: first, for $k \geq j$, $\phi(\mathcal{Z}_{j,k,l}) \subset T_{2j-k}$ and $T_{2j-k} \subset \mathcal{S}'_{j,\mathbf{a}}$ for $k \geq j+1$; second, for $l \leq j-1$, $\phi(\mathcal{Z}_{j,j,l}) \subset \mathcal{Z}_{j,j,j} \subset \mathcal{S}'_{j,\mathbf{a}}$. The first fact holds because for $\mathbf{z}^n$ in $\mathcal{Z}_{j,k,l} \subset \mathcal{S}''_{j,\mathbf{a}}$ the number of -1's in $\phi(\mathbf{z}^n)$ equals $w_l = 2j - k - l$ (the number of +1's in positions greater than i_l given by (75)) plus l (the number of -1's in positions less than or equal to i_l) giving a total of $2j - k$. To see the second fact note that from the definition of $\mathcal{Z}_{j,j,l}$, $w_j > 0$ for $l \leq j - 1$, meaning that $z_n = 1$. Therefore, for $\mathbf{z}^n$ in these sets, $\hat{\mathbf{z}}^n = \phi(\mathbf{z}^n)$ satisfies $\hat{z}_n = -1$ and since from fact one $\hat{\mathbf{z}}^n$ has $(2j - j = j)$ -1's, $\phi(\mathbf{z}^n) \in \mathcal{Z}_{j,j,j}$.

That ϕ is one to one also follows partially from the above discussion which shows that for $k_1 < k_2$ and all l_1, l_2

$$\phi(\mathcal{Z}_{j,k_1,l_1}) \cap \phi(\mathcal{Z}_{j,k_2,l_2}) = \emptyset. \qquad (80)$$

It is also the case that for $l_1 < l_2$

$$\phi(\mathcal{Z}_{j,k,l_1}) \cap \phi(\mathcal{Z}_{j,k,l_2}) = \emptyset. \qquad (81)$$

This is true since it follows from the definition of the sets $\mathcal{Z}_{j,k,l}$ that if $\mathbf{z}_1^n \in \mathcal{Z}_{j,k,l_1}$ and $\mathbf{z}_2^n \in \mathcal{Z}_{j,k,l_2}$, then $z_{1i} \neq z_{2i}$ for at least one $i \geq i_{l_2}$. Note that $i_{l_1} \leq i_{l_2}$ by (78). For this range of $i \geq i_{l_2}$, $\phi(\mathbf{z}^n)$ flips sign, and hence $\phi(\mathbf{z}_1^n) \neq \phi(\mathbf{z}_2^n)$.

The only thing left to demonstrate that ϕ is one to one is to show that if $\mathbf{z}_1^n \neq \mathbf{z}_2^n$ but both $\mathbf{z}_1^n$ and $\mathbf{z}_2^n$ are in $\mathcal{Z}_{j,k,l}$, then $\phi(\mathbf{z}_1^n) \neq \phi(\mathbf{z}_2^n)$. Note that by (78) i_l, the index of the l^{th} -1, is the same for all $\mathbf{z}^n \in \mathcal{Z}_{j,k,l}$. If $\mathbf{z}_1^n$ and $\mathbf{z}_2^n$ differ in some position i satisfying $i_l + 1 \leq i \leq n$ then, since ϕ just flips sign for this range of positions, it follows that $\phi(\mathbf{z}_1^n) \neq \phi(\mathbf{z}_2^n)$. On the other hand, suppose $\mathbf{z}_1^n$ and $\mathbf{z}_2^n$ differ for i satisfying $1 \leq i \leq i_l$. Then, if $\mathcal{S}_1$ is the set of i such that $i \leq i_l$ where $\mathbf{z}_1^n$ is -1, and $\mathcal{S}_2$ is the corresponding set for $\mathbf{z}_2^n$, it must be that $\mathcal{S}_1 \neq \mathcal{S}_2$. Therefore, according to Definition 1 and the one to one property of the subset map ψ_{l,i_l}, the subset of indices $i \leq i_l$ where $\phi(\mathbf{z}_1^n)$ is -1 and the corresponding subset of indices for $\phi(\mathbf{z}_2^n)$ must differ. This completes the proof that ϕ is one to one.

It remains to prove the final property of ϕ that if $\sum_{i=1}^n a_i z_i > n-2j$ then $\hat{\mathbf{z}}^n = \phi(\mathbf{z}^n)$ satisfies $\sum_{i=1}^n a_i \hat{z}_i \leq n-2j$. To see this, define the subsets of indices

$$\mathcal{S}^{(1)} = \{i_1, \ldots, i_l\} \text{ and } \mathcal{S}^{(2)} = \{i : i_l + 1 \leq i \leq n\}\ , \tag{82}$$

and

$$\mathcal{S} = \mathcal{S}^{(1)} \cup \psi_{l,i_l}(\mathcal{S}^{(1)}) \cup \mathcal{S}^{(2)}. \tag{83}$$

The cardinality of $\mathcal{S}$ is

$$\begin{aligned} |\mathcal{S}| &= n - i_l + 2l && (84)\\ &= n - (n - 2j + 2l) + 2l && (85)\\ &= 2j && (86) \end{aligned}$$

so that $|\mathcal{S}^c| = n - 2j$. If

$$\begin{aligned} \sum_{i=1}^n a_i z_i &= \sum_{i \in \mathcal{S}} a_i z_i + \sum_{i \in \mathcal{S}^c} a_i z_i && (87)\\ &> n - 2j, && (88) \end{aligned}$$

then

$$\begin{aligned} \sum_{i \in \mathcal{S}} a_i z_i &> n - 2j - \sum_{i \in \mathcal{S}^c} a_i z_i \\ &\geq n - 2j - |\mathcal{S}^c| \\ &= 0. && (89) \end{aligned}$$

From the definition of ϕ (Definition 1), however, $\hat{\mathbf{z}}^n = \phi(\mathbf{z}^n)$ satisfies

$$\hat{z}_i = -z_i \text{ for } i \in \mathcal{S} \text{ and } \hat{z}_i = z_i \text{ for } i \in \mathcal{S}^c. \tag{90}$$

Therefore, (89) implies that $\sum_{i \in \mathcal{S}} a_i \hat{z}_i < 0$, and hence

$$\begin{aligned} \sum_{i=1}^n a_i \hat{z}_i &= \sum_{i \in \mathcal{S}} a_i \hat{z}_i + \sum_{i \in \mathcal{S}^c} a_i \hat{z}_i && (91)\\ &< \sum_{i \in \mathcal{S}^c} a_i \hat{z}_i && (92)\\ &\leq |\mathcal{S}^c| && (93)\\ &= n - 2j. && (94) \end{aligned}$$

This completes the proof.

4 References

[1] S. Shamai (Shitz), S. Verdu, *BSF grant application,* Appendix, November 1992.

[2] L. A. Shepp, I. Olkin, *Entropy of the sum of independent bernoulli random variables and of the multinoial distribution,* Technical Report 131, Department of Statistics, Stanford University, Stanford, California, 1978.

[3] A. W. Marshall, I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, volume 143 of *Mathematics in Science and Engineering*, Academic Press, London, 1979.

[4] N. L. Biggs, *Discrete Mathematics,* Oxford University Press, Oxford, 1985.