Lectures 4-5: Entropy as a Counting Tool

Algebraic and Analytical Methods in Coding Theory

September 2026

Abstract

The goal is to show how Shannon entropy can be used as a counting tool in combinatorics and coding theory. We start from a card puzzle, prove the entropy inequalities that will be needed, derive Shearer's lemma, and apply it to projections, triangles, and independent sets. We then study two coding problems for the binary adder channel. For two different codes we prove $R_1 + R_2 \leq 3/2$. For a single binary B_2 code we first obtain the direct bound $R \leq 3/4$, and then use a prefix–suffix argument to improve it to $R \leq 3/5$.

Contents

1 A puzzle about transmission of information

Let

$$D = \{1, 2, \ldots, n\}$$

be a deck of n labelled cards. A set of k cards is given to Alice. She chooses one card to hide and shows the other $k-1$ cards to Bob, one at a time, in an order of her choice. Alice and Bob can agree on a strategy before the game, but they cannot communicate during the game.

Bob must always determine the hidden card.

1.1 An informal information count

Alice can choose which card to hide in k ways. She can order the other $k-1$ cards in $(k-1)!$ ways. Hence she has

$$k(k-1)! = k!$$

possible actions. If we measure information in bits, this is at most

$$\log(k!)$$

bits.

After Bob has seen $k-1$ cards, there are

$$n - k + 1$$

possible labels for the missing card. Distinguishing these possibilities requires at least

$$\log(n - k + 1)$$

bits. This suggests

$$\log(n - k + 1) \leq \log(k!),$$

or

$$\boxed{n - k + 1 \leq k!.}$$

1.2 A formal counting proof

There are

$$\binom{n}{k}$$

possible hands. A message sent by Alice is an ordered sequence of $k-1$ distinct cards. The number of such sequences is

$$n(n-1)\cdots(n-k+2) = \frac{n!}{(n-k+1)!}.$$

If two different hands produced the same ordered sequence, Bob would see the same message in two different situations and could not know which hidden card was correct. Therefore the strategy must map different hands to different messages. Hence

$$\binom{n}{k} \leq \frac{n!}{(n-k+1)!}.$$

Now

$$\frac{n!}{(n-k+1)!} = \binom{n}{k} \frac{k!}{n-k+1}.$$

After cancelling $\binom{n}{k}$, we obtain

$$\boxed{n-k+1 \leq k!.}$$

The important point is not the puzzle itself. It is the idea that counting possible messages is the same as measuring how much information can be transmitted. Shannon entropy gives a precise language for this idea.

2 Shannon entropy

Throughout the notes, all logarithms are in base 2.

Definition 2.1 (Entropy)**.** *Let X be a random variable with finite alphabet $\mathcal{X}$. Its entropy is*

$$H(X) := \mathbb{E}[-\log \mathbb{P}(X)] = -\sum_{x \in \mathcal{X}} \mathbb{P}(X = x) \log \mathbb{P}(X = x).$$

If X is uniform on a finite set $\mathcal{F}$, then

$$\mathbb{P}(X = x) = \frac{1}{|\mathcal{F}|}$$

for every $x \in \mathcal{F}$, so

$$H(X) = \log |\mathcal{F}|.$$

This simple identity is the bridge from entropy to counting.

Definition 2.2 (Conditional entropy)**.** *For random variables X, Y,*

$$H(X \mid Y) := -\sum_{x,y} \mathbb{P}(X = x, Y = y) \log \mathbb{P}(X = x \mid Y = y).$$

Equivalently,

$$H(X \mid Y) = \sum_{y} \mathbb{P}(Y = y) H(X \mid Y = y).$$

2.1 The chain rule

Using

$$p(x, y) = p(x)p(y \mid x),$$

we have

$$H(X, Y) = -\sum_{x,y} p(x, y) \log p(x, y)$$

$$= -\sum_{x,y} p(x,y) \log p(x) - \sum_{x,y} p(x,y) \log p(y \mid x)$$
$$= H(X) + H(Y \mid X).$$

Thus

$$\boxed{H(X,Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y).}$$

Repeatedly applying the chain rule gives

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i \mid X_1, \ldots, X_{i-1}).$$

2.2 Four basic properties

Proposition 2.3 (Maximality)**.** *If X takes at most m values, then*

$$H(X) \leq \log m.$$

Equality holds if and only if X is uniform on its support of size m.

Proof. Write $p_x = \mathbb{P}(X = x)$. Since log is concave, Jensen's inequality gives

$$\begin{aligned} H(X) &= \sum_x p_x \log \frac{1}{p_x} \\ &\leq \log \left(\sum_x p_x \frac{1}{p_x} \right) \\ &= \log m. \end{aligned}$$

Equality in Jensen holds exactly when all positive p_x are equal. □

Proposition 2.4 (Conditioning reduces entropy)**.** *For any finite random variables X, Y,*

$$\boxed{H(X \mid Y) \leq H(X).}$$

Equality holds if X and Y are independent.

Proof. For each value y, let $p^{(y)}$ be the conditional distribution of X given $Y = y$. The marginal distribution of X is the average

$$p_X = \sum_y \mathbb{P}(Y = y) p^{(y)}.$$

The entropy function is concave because $-t \log t$ is concave. Therefore

$$H(X) = H\left(\sum_y \mathbb{P}(Y = y) p^{(y)} \right) \geq \sum_y \mathbb{P}(Y = y) H(p^{(y)}) = H(X \mid Y).$$

□

Corollary 2.5 (Subadditivity)**.**

$$\boxed{H(X_1, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i).}$$

Proof. Use the chain rule and then remove conditioning:

$$\begin{aligned} H(X_1, \dots, X_n) &= \sum_{i=1}^{n} H(X_i \mid X_1, \dots, X_{i-1}) \\ &\leq \sum_{i=1}^{n} H(X_i). \end{aligned}$$

□

Proposition 2.6 (A deterministic map cannot increase entropy)**.** *If $Z = f(X)$, then*

$$\boxed{H(Z) \leq H(X).}$$

If f is injective on the support of X, then equality holds.

Proof. Since Z is determined by X,

$$H(Z \mid X) = 0.$$

Hence

$$H(X, Z) = H(X).$$

By the chain rule in the other direction,

$$H(X, Z) = H(Z) + H(X \mid Z) \geq H(Z).$$

Thus $H(Z) \leq H(X)$. If f is injective, then X is also determined by Z, so $H(X \mid Z) = 0$ and equality follows. □

Main idea. To count a finite family, choose an object uniformly from the family. Then its entropy is exactly the logarithm of the size of the family. Bound the same entropy in another way.

3 Shearer's lemma

3.1 A geometric example: projections of points in $\mathbb{R}^3$

Suppose $\mathcal{X} \subset \mathbb{R}^3$ consists of N distinct points. Let

$$N_{12}, \qquad N_{13}, \qquad N_{23}$$

be the numbers of distinct projections on the xy, xz, and yz planes.

Proposition 3.1.

$$\boxed{N^2 \leq N_{12} N_{13} N_{23}.}$$

Proof. Choose

$$X = (X_1, X_2, X_3)$$

uniformly from the N points. Then

$$H(X_1, X_2, X_3) = \log N.$$

The projection (X_1, X_2) takes at most N_{12} values, so maximality gives

$$H(X_1, X_2) \leq \log N_{12}.$$

Similarly,

$$H(X_1, X_3) \le \log N_{13}, \qquad H(X_2, X_3) \le \log N_{23}.$$

We now prove

$$2H(X_1, X_2, X_3) \le H(X_1, X_2) + H(X_1, X_3) + H(X_2, X_3).$$

Expand the right-hand side:

$$\begin{aligned} &H(X_1, X_2) + H(X_1, X_3) + H(X_2, X_3) \\ &= H(X_1) + H(X_2 \mid X_1) + H(X_1) + H(X_3 \mid X_1) + H(X_2) + H(X_3 \mid X_2). \end{aligned}$$

Now use

$$H(X_2) \ge H(X_2 \mid X_1),$$

$$H(X_3 \mid X_1) \ge H(X_3 \mid X_1, X_2), \qquad H(X_3 \mid X_2) \ge H(X_3 \mid X_1, X_2).$$

The result is at least

$$2H(X_1) + 2H(X_2 \mid X_1) + 2H(X_3 \mid X_1, X_2) = 2H(X_1, X_2, X_3).$$

Consequently,

$$2 \log N \le \log N_{12} + \log N_{13} + \log N_{23},$$

which is equivalent to

$$N^2 \le N_{12} N_{13} N_{23}.$$

□

Example 3.2 (Equality)**.** *Take*

$$\mathcal{X} = \{0, 1\}^3.$$

There are $N = 8$ points. Each two-dimensional projection contains the four points of $\{0, 1\}^2$, so

$$N_{12} = N_{13} = N_{23} = 4.$$

Thus

$$8^2 = 4 \cdot 4 \cdot 4.$$

More generally, if $\mathcal{X} = A \times B \times C$, then

$$N = |A||B||C|, \quad N_{12} = |A||B|, \quad N_{13} = |A||C|, \quad N_{23} = |B||C|,$$

and equality holds. In entropy language, the three coordinates are independent. The converse is also true: equality in the complete chain above forces this product structure; see Exercise 2.

3.2 The general lemma

Let

$$X = (X_1, \dots, X_n)$$

and let

$$\mathcal{A} = \{A_i\}_{i \in I}, \qquad A_i \subseteq [n].$$

For $A \subseteq [n]$, write

$$X_A = (X_j : j \in A).$$

Theorem 3.3 (Shearer's lemma)**.** *Assume that every coordinate $j \in [n]$ belongs to at least k members of $\mathcal{A}$. Then*

$$\boxed{\sum_{i \in I} H(X_{A_i}) \ge kH(X).}$$

Proof. Fix $A \subseteq [n]$. List the coordinates of A in increasing order and use the chain rule:

$$H(X_A) = \sum_{j \in A} H\big(X_j \mid X_{A \cap [j-1]}\big).$$

Since

$$A \cap [j-1] \subseteq [j-1],$$

conditioning on all previous coordinates can only reduce entropy. Therefore

$$H\big(X_j \mid X_{A \cap [j-1]}\big) \geq H(X_j \mid X_1, \ldots, X_{j-1}).$$

Hence

$$H(X_A) \geq \sum_{j \in A} H(X_j \mid X_1, \ldots, X_{j-1}).$$

Now sum this inequality over $A_i \in \mathcal{A}$:

$$\sum_{i \in I} H(X_{A_i}) \geq \sum_{i \in I} \sum_{j \in A_i} H(X_j \mid X_1, \ldots, X_{j-1}).$$

Every coordinate j appears in at least k sets A_i. Thus

$$\begin{aligned} \sum_{i \in I} H(X_{A_i}) &\geq k \sum_{j=1}^{n} H(X_j \mid X_1, \ldots, X_{j-1}) \\ &= kH(X_1, \ldots, X_n). \end{aligned}$$

□

4 Two applications of Shearer's lemma

4.1 Triangles in a graph

Let G be a graph with m edges and T triangles.

Proposition 4.1.

$$\boxed{T \leq \frac{(2m)^{3/2}}{6} = \frac{\sqrt{2}}{3} m^{3/2}.}$$

Proof. Choose an ordered triangle

$$(X_1, X_2, X_3)$$

uniformly. Every triangle has $3! = 6$ orderings, so there are $6T$ possible values and

$$H(X_1, X_2, X_3) = \log(6T).$$

Apply Shearer's lemma to

$$A_1 = \{1, 2\}, \qquad A_2 = \{1, 3\}, \qquad A_3 = \{2, 3\}.$$

Each coordinate appears in exactly two sets, hence

$$2H(X_1, X_2, X_3) \leq H(X_1, X_2) + H(X_1, X_3) + H(X_2, X_3).$$

Each ordered pair (X_i, X_j) is an oriented edge of G. There are $2m$ oriented edges, therefore

$$H(X_i, X_j) \leq \log(2m).$$

It follows that

$$2\log(6T) \le 3\log(2m).$$

Exponentiating,

$$(6T)^2 \le (2m)^3,$$

and taking square roots gives the claim. □

For a complete graph, $m \sim N^2/2$ and $T \sim N^3/6$, so the constant is asymptotically sharp.

5 A local entropy calculation for the binary adder

The coding applications below use the same one-coordinate inequality.

Lemma 5.1. *Let*

$$U \sim \mathrm{Bern}(p), \qquad V \sim \mathrm{Bern}(q)$$

be independent. Then

$$\boxed{H(U+V) \le \frac{3}{2}.}$$

Equality holds exactly at $p = q = 1/2$.

Proof. We give a short symmetrization argument.

Step 1: keep $p+q$ fixed. Put

$$s = p + q, \qquad r = pq.$$

The distribution of $S = U + V$ is

$$\mathbb{P}(S=0) = 1 - s + r, \quad \mathbb{P}(S=1) = s - 2r, \quad \mathbb{P}(S=2) = r.$$

For fixed s, regard the entropy as a function of r. Since the derivatives of the three probabilities add to zero, direct differentiation gives

$$\frac{d}{dr}H(S) = \log\frac{(s-2r)^2}{r(1-s+r)}.$$

Now

$$s - 2r = p(1-q) + (1-p)q.$$

By the arithmetic–geometric mean inequality,

$$\begin{aligned} p(1-q) + (1-p)q &\ge 2\sqrt{p(1-q)(1-p)q} \\ &= 2\sqrt{pq(1-p)(1-q)} \\ &= 2\sqrt{r(1-s+r)}. \end{aligned}$$

Therefore

$$\frac{(s-2r)^2}{r(1-s+r)} \ge 4,$$

and hence

$$\frac{d}{dr}H(S) > 0$$

in the interior. So, for fixed $p+q$, the entropy increases with pq. The product pq is maximal when $p = q$. We may therefore set

$$p = q = t.$$

Step 2: maximize along the diagonal. Now U, V are i.i.d. Bernoulli(t). Since $S = U + V$ is a deterministic function of (U, V),

$$H(S) = H(U, V) - H(U, V \mid S).$$

Independence gives

$$H(U, V) = 2h(t).$$

If $S = 0$, then $(U, V) = (0, 0)$. If $S = 2$, then $(U, V) = (1, 1)$. Only when $S = 1$ are there two possible ordered pairs, $(0, 1)$ and $(1, 0)$, and they are equally likely. Hence

$$H(U, V \mid S) = \mathbb{P}(S = 1) = 2t(1 - t).$$

Thus

$$g(t) := H(S) = 2h(t) - 2t(1 - t).$$

This function is symmetric:

$$g(t) = g(1 - t).$$

Moreover,

$$g''(t) = 4 - \frac{2}{(\ln 2)t(1 - t)} < 0,$$

because $t(1 - t) \leq 1/4$. Therefore g is strictly concave and symmetric, so its maximum is at $t = 1/2$. At this point

$$S \sim \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right),$$

and

$$H(S) = -\frac{1}{4}\log\frac{1}{4} - \frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} = \frac{3}{2}.$$

□

6 Two sum-distinct codes

Consider two binary codes

$$C_1, C_2 \subseteq \{0, 1\}^n.$$

We add vectors over the integers, so the output alphabet of one coordinate is $\{0, 1, 2\}$.

Definition 6.1 (Uniquely decodable pair)**.** *The pair (C_1, C_2) is uniquely decodable for the binary adder channel if*

$$(x, y) \longmapsto x + y$$

is injective on $C_1 \times C_2$.

Define

$$R_1 = \frac{1}{n}\log|C_1|, \qquad R_2 = \frac{1}{n}\log|C_2|.$$

Theorem 6.2. *Every uniquely decodable pair satisfies*

$$\boxed{R_1 + R_2 \leq \frac{3}{2}.}$$

Proof. Choose X uniformly from C_1 and Y uniformly from C_2, independently. Then

$$H(X) = \log |C_1|, \qquad H(Y) = \log |C_2|,$$

and independence gives

$$H(X, Y) = \log |C_1| + \log |C_2|.$$

Put

$$Z = X + Y.$$

Because the sum map is injective on $C_1 \times C_2$, Z determines (X, Y). Thus

$$H(Z) = H(X, Y).$$

By subadditivity,

$$H(Z) \leq \sum_{i=1}^{n} H(Z_i).$$

For every coordinate i, the random variables X_i, Y_i are independent Bernoulli random variables, although their parameters can depend on i and on the code. The local lemma gives

$$H(Z_i) = H(X_i + Y_i) \leq \frac{3}{2}.$$

Therefore

$$\log |C_1| + \log |C_2| = H(Z) \leq \frac{3}{2} n.$$

Dividing by n gives

$$R_1 + R_2 \leq \frac{3}{2}.$$

□

Main idea. The proof has only two ingredients: global injectivity of the sum map and a one-coordinate entropy bound. This is the cleanest form of the entropy method in coding theory.

7 From two codes to one binary B_2 code

We now use one code instead of two different codes.

Definition 7.1 (Binary B_2 code)**.** *A code*

$$C \subseteq \{0, 1\}^n$$

is a binary B_2 code if the integer sums $x + y$ are pairwise distinct when indexed by unordered pairs $\{x, y\}$ of codewords, allowing $x = y$. Equivalently, $x + y$ determines the unordered pair $\{x, y\}$.

Let

$$M = |C|, \qquad R = \frac{1}{n} \log M.$$

The connection with the previous problem is immediate. If we tried to set

$$C_1 = C_2 = C,$$

the sum map would no longer be injective on ordered pairs because

$$x + y = y + x.$$

The output sees the unordered pair but not its order.

7.1 First attempt: a direct entropy bound

Let X, Y be independent and uniform on C and define

$$S = X + Y.$$

The B_2 property says that S determines $\{X, Y\}$. Thus, given S, there are at most two possible ordered pairs: (X, Y) and (Y, X). Therefore

$$H(X, Y \mid S) \leq 1.$$

Since S is a function of (X, Y),

$$H(X, Y) = H(S) + H(X, Y \mid S).$$

Hence

$$2 \log M = H(X, Y) \leq H(S) + 1.$$

By subadditivity and the Bernoulli-sum lemma,

$$H(S) \leq \sum_{i=1}^{n} H(X_i + Y_i) \leq \frac{3}{2} n.$$

Consequently,

$$2 \log M \leq \frac{3}{2} n + 1.$$

Dividing by n and letting n grow gives

$$\boxed{R \leq \frac{3}{4}.}$$

One can make the finite calculation slightly sharper. If $X \neq Y$, then there are exactly two equiprobable ordered pairs compatible with the sum; if $X = Y$, there is only one. Since $\mathbb{P}(X = Y) = 1/M$,

$$H(X, Y \mid S) = 1 - \frac{1}{M}.$$

The asymptotic rate bound is the same.

Remark 7.2. *The direct argument loses only one bit because of the swap ambiguity. This is not enough to obtain the stronger* $3/5$ *bound. The next argument changes the random experiment so that order is recovered from a difference, except when the two codewords are equal.*

8 The prefix–suffix method for binary B_2 codes

8.1 Step 1: group codewords by a prefix

Split every codeword into a prefix of length e and a suffix of length

$$f = n - e.$$

Write

$$c_i = (p_i, w_i).$$

For each possible prefix $\ell \in \{0, 1\}^e$, define

$$P_\ell = \{(\ell, w) : w \in W_\ell\}, \qquad M_\ell = |W_\ell|.$$

There are at most 2^e prefix classes and

$$\sum_{\ell} M_\ell = M.$$

Define

$$S_2 := \sum_{\ell} M_\ell^2.$$

Cauchy–Schwarz gives

$$M^2 = \left(\sum_{\ell} M_\ell\right)^2 \le 2^e \sum_{\ell} M_\ell^2 = 2^e S_2.$$

Therefore

$$\boxed{S_2 \ge \frac{M^2}{2^e}.}$$

Now choose an ordered pair of codewords uniformly among all ordered pairs that have the same prefix. There are exactly S_2 such pairs. Let (X, Y) be the corresponding pair of suffixes. Because the pair is uniform,

$$\boxed{H(X, Y) = \log S_2.}$$

Notice an important detail. The prefix class ℓ is not chosen uniformly. It is chosen with probability

$$\mathbb{P}(I = \ell) = \frac{M_\ell^2}{S_2}.$$

Conditioned on $I = \ell$, the two suffixes X, Y are independent and uniform on W_ℓ.

8.2 Step 2: use the difference of the suffixes

Define

$$Z = X - Y \in \{-1, 0, 1\}^f.$$

The difference has an advantage over the sum: it remembers the order, since swapping X and Y changes Z to $-Z$. We now show that the B_2 property makes the difference map injective away from the diagonal.

Lemma 8.1 (Off-diagonal injectivity)**.** *The map*

$$(w_1, w_2) \longmapsto w_1 - w_2$$

is injective on the ordered pairs with $w_1 \neq w_2$ taken from prefix classes.

Proof. Suppose

$$w_1 - w_2 = w_3 - w_4,$$

where w_1, w_2 belong to the prefix class ℓ and w_3, w_4 belong to the prefix class m. Define the corresponding full codewords

$$a = (\ell, w_1), \quad b = (\ell, w_2), \quad c = (m, w_3), \quad d = (m, w_4).$$

The equality of differences is equivalent to

$$w_1 + w_4 = w_2 + w_3.$$

The prefix of $a + d$ is $\ell + m$, and the prefix of $b + c$ is also $\ell + m$. Therefore

$$a + d = b + c.$$

Since C is a B_2 code, the sum determines the unordered pair. Hence

$$\{a, d\} = \{b, c\}.$$

There are two possible matchings. The matching $a = b$ and $d = c$ would imply $w_1 = w_2$, contradicting the off-diagonal assumption. Therefore the other matching must hold:

$$a = c, \qquad d = b.$$

This gives

$$\ell = m, \qquad w_1 = w_3, \qquad w_2 = w_4.$$

Thus the ordered pairs are equal. □

Every nonzero value of Z therefore has exactly one preimage. The only value with many preimages is

$$Z = 0,$$

which occurs exactly for diagonal pairs $X = Y$.

Since Z is a deterministic function of (X, Y),

$$H(X, Y) = H(Z) + H(X, Y \mid Z).$$

For $Z \neq 0$, the conditional entropy is zero by injectivity. Hence

$$\boxed{H(X, Y) = H(Z) + \mathbb{P}(Z = 0)H(X, Y \mid Z = 0).}$$

8.3 Step 3: control the diagonal term

There is one diagonal ordered pair for each codeword. Hence exactly M of the S_2 ordered pairs satisfy $Z = 0$. Therefore

$$\mathbb{P}(Z = 0) = \frac{M}{S_2}.$$

Conditioned on $Z = 0$, the M diagonal pairs are equiprobable, so

$$H(X, Y \mid Z = 0) = \log M.$$

Thus

$$\mathbb{P}(Z = 0)H(X, Y \mid Z = 0) = \frac{M}{S_2} \log M.$$

Using

$$S_2 \geq \frac{M^2}{2^e},$$

we obtain

$$\boxed{\mathbb{P}(Z = 0)H(X, Y \mid Z = 0) \leq \frac{2^e}{M} \log M.}$$

The idea will be to choose e slightly smaller than $\log M$, making this term bounded while keeping the prefix long enough to be useful.

8.4 Step 4: bound the entropy of one difference coordinate

Write

$$Z = (Z_1, \dots, Z_f).$$

Fix a coordinate j. Conditioned on the prefix class $I = \ell$, the random variables X_j, Y_j are independent and identically distributed Bernoulli variables. Let

$$p_{\ell,j} = \mathbb{P}(X_j = 1 \mid I = \ell).$$

Then

$$\begin{aligned}
\mathbb{P}(Z_j = 0 \mid I = \ell) &= \mathbb{P}(X_j = Y_j \mid I = \ell) \\
&= p_{\ell,j}^2 + (1 - p_{\ell,j})^2 \\
&= \frac{1}{2} + 2\left(p_{\ell,j} - \frac{1}{2}\right)^2 \\
&\geq \frac{1}{2}.
\end{aligned}$$

Averaging over ℓ gives

$$\boxed{\mathbb{P}(Z_j = 0) \geq \frac{1}{2}.}$$

The random variable Z_j takes only three values $-1, 0, 1$. Under the constraint that the middle probability is at least $1/2$, entropy is maximized when

$$\mathbb{P}(Z_j = 0) = \frac{1}{2}$$

and the remaining probability is split equally between -1 and 1. Thus the maximizing distribution is

$$\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right),$$

and

$$\boxed{H(Z_j) \leq \frac{3}{2}.}$$

This is the same local constant that appeared for the binary adder channel.

By subadditivity,

$$H(Z) \leq \sum_{j=1}^{f} H(Z_j) \leq \frac{3}{2} f.$$

Since $f = n - e$,

$$\boxed{H(Z) \leq \frac{3}{2}(n - e).}$$

8.5 Step 5: choose the prefix length and finish the proof

Assume that M grows exponentially with n. If not, the asymptotic rate is zero and there is nothing to prove. Put

$$L = \log M$$

and choose

$$\boxed{e = \lfloor L - \log L \rfloor.}$$

Then

$$2^e \le \frac{M}{\log M},$$

so the diagonal contribution satisfies

$$\mathbb{P}(Z=0)H(X,Y \mid Z=0) \le 1.$$

Therefore

$$\begin{aligned}\log S_2 &= H(X,Y)\\ &= H(Z) + \mathbb{P}(Z=0)H(X,Y \mid Z=0)\\ &\le \frac{3}{2}(n-e)+1.\end{aligned}$$

On the other hand,

$$S_2 \ge \frac{M^2}{2^e}$$

implies

$$\log S_2 \ge 2\log M - e = 2L - e.$$

Combining the two estimates,

$$2L - e \le \frac{3}{2}(n-e)+1.$$

Move the e terms to the left:

$$2L + \frac{1}{2}e \le \frac{3}{2}n + 1.$$

Our choice of e gives

$$e = L - \log L + O(1) = L + o(n).$$

Hence

$$\frac{5}{2}L \le \frac{3}{2}n + o(n).$$

Divide by n:

$$\frac{5}{2}R \le \frac{3}{2} + o(1).$$

Taking the limit superior gives

$$\boxed{R \le \frac{3}{5}.}$$

9 What the two coding proofs have in common

The coding part of the lecture can be summarized by the following table.

Problem	Main injective object	Entropy conclusion
Two codes C_1, C_2	$(x,y) \mapsto x+y$	$R_1 + R_2 \le 3/2$
One binary B_2 code, direct	sum determines $\{x,y\}$	$R \le 3/4$
One binary B_2 code, refined	suffix difference, off diagonal	$R \le 3/5$

The same local distribution

$$\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$$

and the same entropy value

$$\frac{3}{2}$$

appear in both problems. The difference is global.

For two separate codes, the two users are labelled: the output sum uniquely identifies the ordered pair. For one code, the sum is symmetric and forgets the order. A direct treatment of this symmetry gives the 3/4 bound. The prefix–suffix method is stronger because the difference of two suffixes remembers the order. The only point where the difference loses information is $X = Y$, and the prefix length is chosen so that this diagonal event contributes only $O(1)$ bits.

Main idea. Entropy methods often have two layers. First find a local entropy bound for one coordinate. Then use combinatorial injectivity to turn that local bound into a global bound on the size of a code or a family.

Exercises

1. Show directly that every Cartesian product $A \times B \times C \subset \mathbb{R}^3$ gives equality in the projection inequality.

2. In the projection proof, identify all the places where equality is needed. Use them to show that equality forces the three coordinates to be mutually independent.

3. Verify the triangle bound for $G = K_N$ and check that the ratio tends to 1 as $N \to \infty$.

4. For a binary B_2 code, prove the exact identity

$$H(X, Y \mid X + Y) = 1 - \frac{1}{M}$$

when X, Y are independent and uniform on the code.

5. In the prefix–suffix proof, replace $e = \lfloor \log M - \log\log M \rfloor$ by $e = \lfloor \log M - c \log\log M \rfloor$ for a fixed $c > 0$. Determine for which c the diagonal term is $o(n)$ and check that the asymptotic rate bound remains 3/5.

References

[1] J. Radhakrishnan, *Entropy and Counting*, lecture notes.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley, 2006.

[3] J. Kahn, An entropy approach to the hard-core model on bipartite graphs, *Combinatorics, Probability and Computing* **10** (2001), 219–237.

[4] B. Lindström, Determination of two vectors from the sum, *Journal of Combinatorial Theory* **6** (1969), 402–407.

[5] S. Della Fiore and M. Dalai, A note on $\overline{2}$-separable codes and B_2 codes, *Discrete Mathematics* **345** (2022), 112751.