Chapter 24

Entropy and Shannon's Theorem

NEW CS 473: Theory II, Fall 2015
November 18, 2015

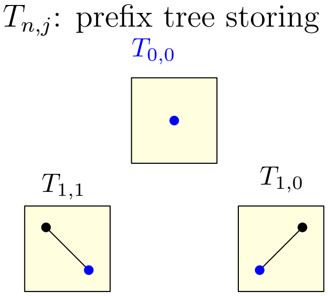
24.1 Entropy

24.2 Extracting randomness

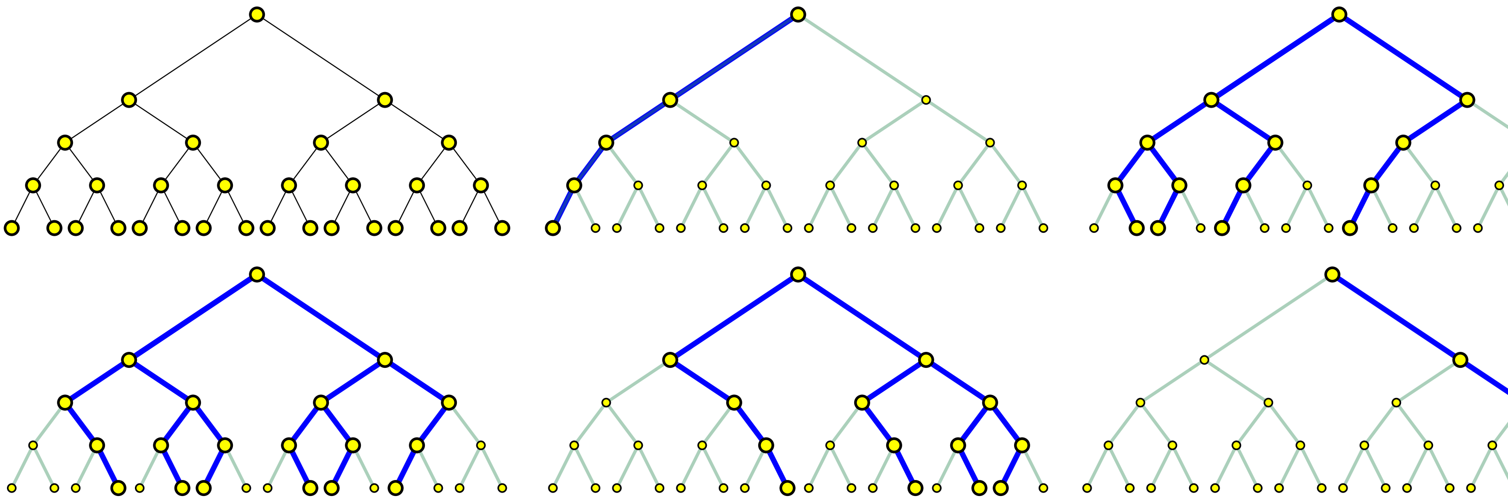
24.2.1 Enumerating binary strings with j ones

24.2.1.1 Storing all strings of length n and j bits on

(A) $S_{n,j}$: set of all strings of length n with j ones in them.
(B) $T_{n,j}$: prefix tree storing all $S_{n,j}$.



24.2.1.2 Binary strings of length 4



(A) $S_{4,0} = \{0000\} \implies \#(0000) = 0.$

(B) $S_{4,1} = \{0001, 0010, 0100, 1000\}$

$\implies \#(0001) = 0.$

$\#(0010) = 1.$

$\#(0100) = 2.$

$\#(1000) = 3.$

(C) $S_{4,2} = \{0011, 0101, 0110, 1001, 1010, 1100\}$

$\implies$

$\#(0011) = 0.$

$\#(0101) = 1.$

$\#(0110) = 2.$

$\#(1001) = 3.$

$\#(1010) = 4.$

$\#(1100) = 5.$

(D) $S_{4,3} = \{0111, 1011, 1101, 1110\}$

$\implies$

$\#(0111) = 0.$

$\#(1011) = 1.$

$\#(1101) = 2.$

$\#(1110) = 3.$

(E) $S_{4,4} = \{1111\}$

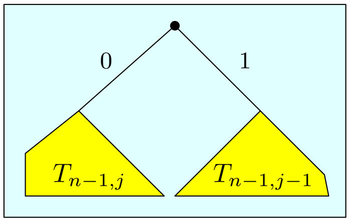
$\implies$

$\#(1111) = 0.$

(F)

24.2.1.3 Prefix tree $\forall$ binary strings of length n with j ones

$T_{n,j}$

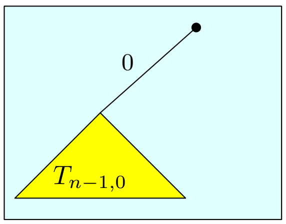


# of leafs:
$|T_{n,j}| = |T_{n-1,j}| + |T_{n-1,j-1}|$

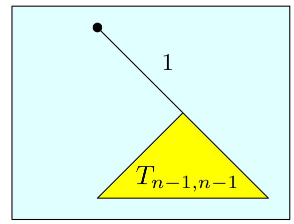
$\binom{n}{j} = \binom{n-1}{j} + \binom{n-1}{j-1}$

$\implies |T_{n,j}| = \binom{n}{j}$.

$T_{n,0}$



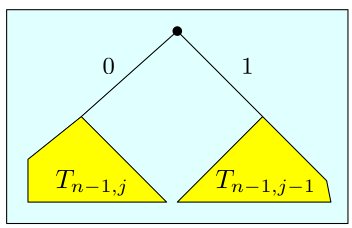
$T_{n,n}$



24.2.1.4 Encoding a string in $S_{n,j}$

(A) $T_{n,j}$ leafs corresponds to strings of $S_{n,j}$.
(B) Order all strings of $S_{n,j}$ order in lexicographical ordering
(C) $\equiv$ ordering leafs of $T_{n,j}$ from left to right.

$T_{n,j}$

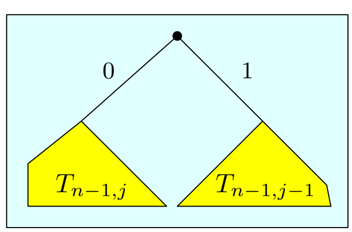


(D) Input: $s \in S_{n,j}$: compute index of s in sorted set $S_{n,j}$.
(E) **EncodeBinomCoeff**(s) denote this polytime procedure.

24.2.1.5 Decoding a string in $S_{n,j}$

(A) $T_{n,j}$ leafs corresponds to strings of $S_{n,j}$.
(B) Order all strings of $S_{n,j}$ order in lexicographical ordering
(C) $\equiv$ ordering leafs of $T_{n,j}$ from left to right.

$T_{n,j}$



(D) $x \in \left\{1, \ldots, \binom{n}{j}\right\}$: compute xth string in $S_{n,j}$ in polytime.
(E) **DecodeBinomCoeff** (x) denote this procedure.

24.2.1.6 Encoding/decoding strings of $S_{n,j}$

Lemma 24.2.1. *$S_{n,j}$: Set of binary strings of length n with j ones, sorted lexicographically.*

(A) **EncodeBinomCoeff**(α)*: Input is string $\alpha \in S_{n,j}$, compute index x of α in $S_{n,j}$ in polynomial time in n.*

(B) **DecodeBinomCoeff**(x)*: Input index $x \in \left\{1, \ldots, \binom{n}{j}\right\}$. Output xth string α in $S_{n,j}$, in time $O(\operatorname{polylog} n + n)$.*

24.2.2 Extracting randomness

24.2.2.1 Extracting randomness

Theorem 24.2.2. *Consider a coin that comes up heads with probability $p > 1/2$. For any constant $\delta > 0$ and for n sufficiently large:*

(A) One can extract, from an input of a sequence of n flips, an output sequence of $(1-\delta)n\mathbb{H}(p)$ (unbiased) independent random bits.

(B) One can not extract more than $n\mathbb{H}(p)$ bits from such a sequence.

24.2.2.2 Proof...

(A) There are $\binom{n}{j}$ input strings with exactly j heads.

(B) each has probability $p^j(1-p)^{n-j}$.

(C) map string s like that to index number in the set $S_j = \left\{1, \ldots, \binom{n}{j}\right\}$.

(D) Given that input string s has j ones (out of n bits) defines a uniform distribution on $S_{n,j}$.

(E) $x \leftarrow$ **EncodeBinomCoeff**(s)

(F) x uniform distributed in $\{1, \ldots, N\}$, $N = \binom{n}{j}$.

(G) Seen in previous lecture...

(H) ... extract in expectation, $\lfloor \lg N \rfloor - 1$ bits from uniform random variable in the range $1, \ldots, N$.

(I) Extract bits using **ExtractRandomness**(x, N):.

24.2.2.3 Exciting proof continued...

(A) Z: random variable: number of heads in input string s.

(B) B: number of random bits extracted.

$$\mathbf{E}\big[B\big] = \sum_{k=0}^{n} \mathbf{Pr}[Z = k]\, \mathbf{E}\left[B \,\middle|\, Z = k\right],$$

(C) Know: $\mathbf{E}\left[B \,\middle|\, Z = k\right] \geq \left\lfloor \lg \binom{n}{k} \right\rfloor - 1$.

(D) $\varepsilon < p - 1/2$: sufficiently small constant.

(E) $n(p-\varepsilon) \leq k \leq n(p+\varepsilon)$:

$$\binom{n}{k} \geq \binom{n}{\lfloor n(p+\varepsilon) \rfloor} \geq \frac{2^{n\mathbb{H}(p+\varepsilon)}}{n+1},$$

(F) ... since $2^{n\mathbb{H}(p)}$ is a good approximation to $\binom{n}{np}$ as proved in previous lecture.

24.2.2.4 Super exciting proof continued...

$\mathbf{E}\big[B\big] = \sum_{k=0}^{n} \mathbf{Pr}[Z = k]\, \mathbf{E}\Big[B \;\Big|\; Z = k\Big]$.

$$\begin{aligned}
\mathbf{E}\Big[B\Big] &\geq \textstyle\sum_{k=\lfloor n(p-\varepsilon)\rfloor}^{\lceil n(p+\varepsilon)\rceil} \mathbf{Pr}\Big[Z = k\Big]\, \mathbf{E}\Big[B \;\Big|\; Z = k\Big] \\
&\geq \sum_{k=\lfloor n(p-\varepsilon)\rfloor}^{\lceil n(p+\varepsilon)\rceil} \mathbf{Pr}\big[Z = k\big] \left(\left\lfloor \lg \binom{n}{k} \right\rfloor - 1 \right) \\
&\geq \sum_{k=\lfloor n(p-\varepsilon)\rfloor}^{\lceil n(p+\varepsilon)\rceil} \mathbf{Pr}\Big[Z = k\Big] \left(\lg \frac{2^{n\mathbb{H}(p+\varepsilon)}}{n+1} - 2 \right) \\
&= \Big(n\mathbb{H}(p+\varepsilon) - \lg(n+1) - 2\Big)\, \mathbf{Pr}[|Z - np| \leq \varepsilon n] \\
&\geq \Big(n\mathbb{H}(p+\varepsilon) - \lg(n+1) - 2\Big) \left(1 - 2\exp\left(-\frac{n\varepsilon^2}{4p}\right)\right),
\end{aligned}$$

since $\mu = \mathbf{E}[Z] = np$ and $\mathbf{Pr}\Big[|Z - np| \geq \frac{\varepsilon}{p} pn\Big] \leq 2\exp\left(-\frac{np}{4}\left(\frac{\varepsilon}{p}\right)^2\right) = 2\exp\left(-\frac{n\varepsilon^2}{4p}\right)$, by the Chernoff inequality.

24.2.2.5 Hyper super exciting proof continued...

(A) Fix $\varepsilon > 0$, such that $\mathbb{H}(p+\varepsilon) > (1 - \delta/4)\mathbb{H}(p)$, p is fixed.
(B) $\implies n\mathbb{H}(p) = \Omega(n)$,
(C) For n sufficiently large: $-\lg(n+1) \geq -\frac{\delta}{10} n\mathbb{H}(p)$.
(D) ... also $2\exp\left(-\frac{n\varepsilon^2}{4p}\right) \leq \frac{\delta}{10}$.
(E) For n large enough;

$$\begin{aligned}
\mathbf{E}[B] &\geq \left(1 - \frac{\delta}{4} - \frac{\delta}{10}\right) n\mathbb{H}(p) \left(1 - \frac{\delta}{10}\right) \\
&\geq (1 - \delta) n\mathbb{H}(p),
\end{aligned}$$

24.2.2.6 Hyper super duper exciting proof continued...

(A) Need to prove upper bound.
(B) If input sequence x has probability $\mathbf{Pr}[X = x]$, then $y = \mathbf{Ext}(x)$ has probability to be generated $\geq \mathbf{Pr}[X = x]$.
(C) All sequences of length $|y|$ have equal probability to be generated (by definition).
(D) $2^{|\mathbf{Ext}(x)|}\, \mathbf{Pr}[X = x] \leq 2^{|\mathbf{Ext}(x)|}\, \mathbf{Pr}[y = \mathbf{Ext}(x)] \leq 1$.
(E) $\implies |\mathbf{Ext}(x)| \leq \lg(1/\mathbf{Pr}[X = x])$
(F) $\mathbf{E}\Big[B\Big] = \sum_x \mathbf{Pr}\Big[X = x\Big]\, |\mathbf{Ext}(x)| \leq \sum_x \mathbf{Pr}\Big[X = x\Big] \lg \frac{1}{\mathbf{Pr}[X=x]} = \mathbb{H}(X)$.

■

24.3 Coding: Shannon's Theorem

24.3.0.1 Shannon's Theorem

Definition 24.3.1. (A) ***binary symmetric channel*** with parameter p
(B) sequence of bits $x_1, x_2, \ldots$, an

(C) output: $y_1, y_2, \ldots,$
a sequence of bits such that...
(D) $\mathbf{Pr}[x_i = y_i] = 1 - p$ independently for each i.

24.3.0.2 Encoding/decoding with noise

Definition 24.3.2. (A) (k, n) ***encoding function*** $\texttt{Enc} : \{0,1\}^k \to \{0,1\}^n$ takes as input a sequence of k bits and outputs a sequence of n bits.
(B) (k, n) ***decoding function*** $\texttt{Dec} : \{0,1\}^n \to \{0,1\}^k$ takes as input a sequence of n bits and outputs a sequence of k bits.

24.3.0.3 Claude Elwood Shannon

Claude Elwood Shannon (April 30, 1916 - February 24, 2001), an American electrical engineer and mathematician, has been called "the father of information theory".

His master thesis was how to building boolean circuits for any boolean function.

24.3.0.4 Shannon's theorem (1948)

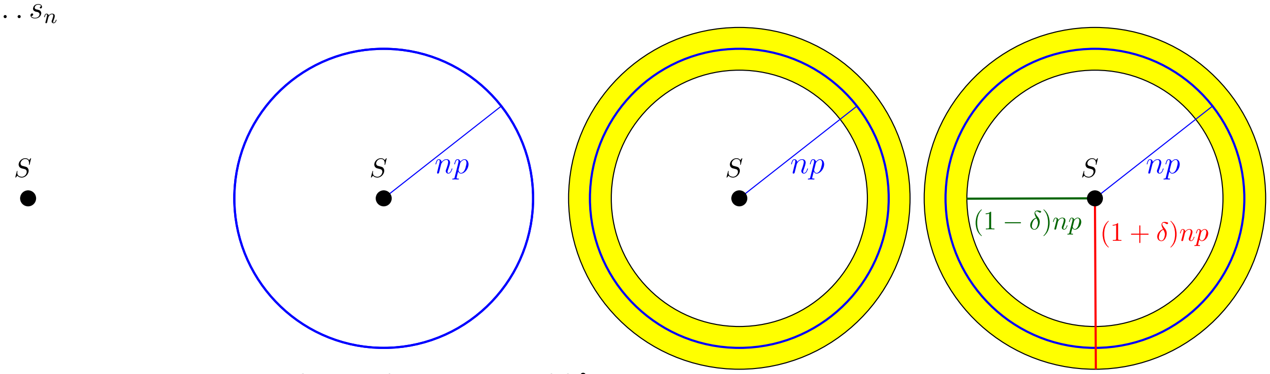
Theorem 24.3.3 (Shannon's theorem). *For a binary symmetric channel with parameter $p < 1/2$ and for any constants $\delta, \gamma > 0$, where n is sufficiently large, the following holds:*

(i) For an $k \leq n(1 - \mathbb{H}(p) - \delta)$ there exists (k, n) encoding and decoding functions such that the probability the receiver fails to obtain the correct message is at most γ for every possible k-bit input messages.

(ii) There are no (k, n) encoding and decoding functions with $k \geq n(1 - \mathbb{H}(p) + \delta)$ such that the probability of decoding correctly is at least γ for a k-bit input message chosen uniformly at random.

24.3.0.5 Intuition behind Shanon's theorem
24.3.0.6 When the sender sends a string...

$S = s_1 s_2 \ldots s_n$



One ring to rule them all!

24.3.0.7 Some intuition...

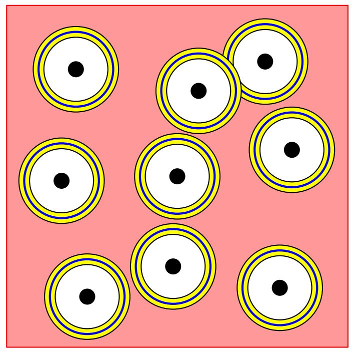
(A) senders sent string $S = s_1 s_2 \ldots s_n$.
(B) receiver got string $T = t_1 t_2 \ldots t_n$.
(C) $p = \mathbf{Pr}[t_i \neq s_i]$, for all i.
(D) U: Hamming distance between S and T: $U = \sum_i [s_i \neq t_i]$.
(E) By assumption: $\mathbf{E}[U] = pn$, and U is a binomial variable.

(F) By Chernoff inequality: $U \in \big[(1-\delta)np, (1+\delta)np\big]$ with high probability, where δ is tiny constant.
(G) T is in a ring R centered at S, with inner radius $(1-\delta)np$ and outer radius $(1+\delta)np$.
(H) This ring has

$$\sum_{i=(1-\delta)np}^{(1+\delta)np} \binom{n}{i} \leq 2\binom{n}{(1+\delta)np} \leq \alpha = 2 \cdot 2^{n\mathbb{H}((1+\delta)p)}.$$

strings in it.

24.3.0.8 Many rings for many codewords...



24.3.0.9 Some more intuition...

(A) Pick as many disjoint rings as possible: $R_1, \ldots, R_\kappa$.
(B) If every word in the hypercube would be covered...
(C) ... use 2^n codewords $\implies \kappa \geq$

$$\kappa \geq \frac{2^n}{|R|} \geq \frac{2^n}{2 \cdot 2^{n\mathbb{H}((1+\delta)p)}} \approx 2^{n(1-\mathbb{H}((1+\delta)p))}.$$

(D) Consider all possible strings of length k such that $2^k \leq \kappa$.
(E) Map ith string in $\{0,1\}^k$ to the center C_i of the ith ring R_i.
(F) If send $C_i \implies$ receiver gets a string in R_i.
(G) Decoding is easy - find the ring R_i containing the received string, take its center string C_i, and output the original string it was mapped to.
(H) How many bits? $k = \lfloor \log \kappa \rfloor = n\Big(1 - \mathbb{H}\Big((1+\delta)p\Big)\Big) \approx n(1 - \mathbb{H}(p))$,

24.3.0.10 What is wrong with the above?
24.3.0.11 What is wrong with the above?

(A) Can not find such a large set of disjoint rings.
(B) Reason is that when you pack rings (or balls) you are going to have wasted spaces around.
(C) Overcome this: allow rings to overlap somewhat.
(D) Makes things considerably more involved.
(E) Details in class notes.