

Chapter 23

Entropy, Randomness, and Information

NEW CS 473: Theory II, Fall 2015

November 13, 2015

23.1 Entropy

23.1.0.1 Quote

“If only once - only once - no matter where, no matter before what audience - I could better the record of the great Rastelli and juggle with thirteen balls, instead of my usual twelve, I would feel that I had truly accomplished something for my country. But I am not getting any younger, and although I am still at the peak of my powers there are moments - why deny it? - when I begin to doubt - and there is a time limit on all of us.”

–Romain Gary, The talent scout.

23.2 Entropy

23.2.0.1 Entropy: Definition

Definition 23.2.1. The *entropy* in bits of a discrete random variable X is

$$\mathbb{H}(X) = - \sum_x \Pr[X = x] \lg \Pr[X = x].$$

Equivalently, $\mathbb{H}(X) = \mathbf{E} \left[\lg \frac{1}{\Pr[X]} \right]$.

23.2.0.2 Entropy intuition...

Intuition... $\mathbb{H}(X)$ is the number of *fair* coin flips that one gets when getting the value of X .

Interpretation from last lecture... Consider a (huge) string $S = s_1 s_2 \dots s_n$ formed by picking characters independently according to X . Then

$$|S| \mathbb{H}(X) = n \mathbb{H}(X)$$

is the minimum number of bits one needs to store the string S .

23.2.0.3 Binary entropy

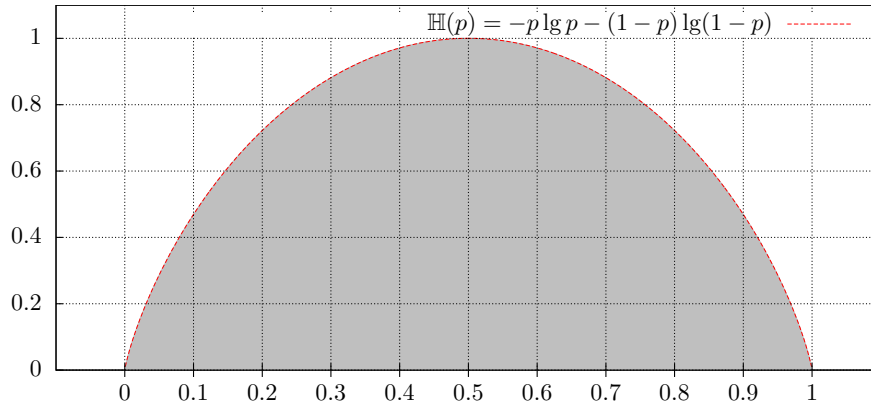
$$\mathbb{H}(X) = - \sum_x \Pr[X = x] \lg \Pr[X = x]$$

\implies

Definition 23.2.2. The **binary entropy** function $\mathbb{H}(p)$ for a random binary variable that is 1 with probability p , is $\mathbb{H}(p) = -p \lg p - (1-p) \lg(1-p)$. We define $\mathbb{H}(0) = \mathbb{H}(1) = 0$.

Q: How many truly random bits are there when given the result of flipping a single coin with probability p for heads?

23.2.0.4 Binary entropy: $\mathbb{H}(p) = -p \lg p - (1-p) \lg(1-p)$



- (A) $\mathbb{H}(p)$ is a concave symmetric around $1/2$ on the interval $[0, 1]$.
- (B) maximum at $1/2$.
- (C) $\mathbb{H}(3/4) \approx 0.8113$ and $\mathbb{H}(7/8) \approx 0.5436$.
- (D) \implies coin that has $3/4$ probably to be heads have higher amount of “randomness” in it than a coin that has probability $7/8$ for heads.

23.2.0.5 And now for some unnecessary math

- (A) $\mathbb{H}(p) = -p \lg p - (1-p) \lg(1-p)$
- (B) $\mathbb{H}'(p) = -\lg p + \lg(1-p) = \lg \frac{1-p}{p}$
- (C) $\mathbb{H}''(p) = \frac{p}{1-p} \cdot \left(-\frac{1}{p^2}\right) = -\frac{1}{p(1-p)}$.
- (D) $\implies \mathbb{H}''(p) \leq 0$, for all $p \in (0, 1)$, and the $\mathbb{H}(\cdot)$ is concave.
- (E) $\mathbb{H}'(1/2) = 0 \implies \mathbb{H}(1/2) = 1$ max of binary entropy.
- (F) \implies balanced coin has the largest amount of randomness in it.

23.2.1 Task at hand: Squeezing good random bits...

23.2.1.1 ...out of bad random bits...

- (A) b_1, \dots, b_n : result of n coin flips...
- (B) From a faulty coin!
- (C) p : probability for head.
- (D) We need fair bit coins!
- (E) Convert $b_1, \dots, b_n \implies b'_1, \dots, b'_m$.
- (F) **New bits must be truly random:** Probability for head is $1/2$.
- (G) **Q:** How many truly random bits can we extract?

23.2.2 Intuitively...

23.2.2.1 Squeezing good random bits out of bad random bits...

Question... Given the result of n coin flips: b_1, \dots, b_n from a faulty coin, with head with probability p , how many truly random bits can we extract?

If believe intuition about entropy, then this number should be $\approx n\mathbb{H}(p)$.

23.2.2.2 Back to Entropy

(A) **entropy** of X is $\mathbb{H}(X) = -\sum_x \Pr[X = x] \lg \Pr[X = x]$.

(B) Entropy of uniform variable..

Example 23.2.3. A random variable X that has probability $1/n$ to be i , for $i = 1, \dots, n$, has entropy $\mathbb{H}(X) = -\sum_{i=1}^n \frac{1}{n} \lg \frac{1}{n} = \lg n$.

(C) Entropy is oblivious to the exact values random variable can have.

(D) \implies random variables over $-1, +1$ with equal probability has the same entropy (i.e., 1) as a fair coin.

23.2.2.3 Lemma: Entropy additive for independent variables

23.2.2.4 Lemma: Entropy additive for independent variables

Lemma 23.2.4. Let X and Y be two independent random variables, and let Z be the random variable (X, Y) . Then $\mathbb{H}(Z) = \mathbb{H}(X) + \mathbb{H}(Y)$.

23.2.2.5 Proof

In the following, summation are over all possible values that the variables can have. By the independence of X and Y we have

$$\begin{aligned} \mathbb{H}(Z) &= \sum_{x,y} \Pr[(X, Y) = (x, y)] \lg \frac{1}{\Pr[(X, Y) = (x, y)]} \\ &= \sum_{x,y} \Pr[X = x] \Pr[Y = y] \lg \frac{1}{\Pr[X = x] \Pr[Y = y]} \\ &= \sum_x \sum_y \Pr[X = x] \Pr[Y = y] \lg \frac{1}{\Pr[X = x]} \\ &\quad + \sum_y \sum_x \Pr[X = x] \Pr[Y = y] \lg \frac{1}{\Pr[Y = y]} \end{aligned}$$

23.2.2.6 Proof continued

$$\begin{aligned}
\mathbb{H}(Z) &= \sum_x \sum_y \Pr[X = x] \Pr[Y = y] \lg \frac{1}{\Pr[X = x]} \\
&\quad + \sum_y \sum_x \Pr[X = x] \Pr[Y = y] \lg \frac{1}{\Pr[Y = y]} \\
&= \sum_x \Pr[X = x] \lg \frac{1}{\Pr[X = x]} \\
&\quad + \sum_y \Pr[Y = y] \lg \frac{1}{\Pr[Y = y]} \\
&= \mathbb{H}(X) + \mathbb{H}(Y).
\end{aligned}$$

■

23.2.2.7 Bounding the binomial coefficient using entropy

23.2.2.8 Bounding the binomial coefficient using entropy

Lemma 23.2.5. $q \in [0, 1]$

nq is integer in the range $[0, n]$.

Then

$$\frac{2^{n\mathbb{H}(q)}}{n+1} \leq \binom{n}{nq} \leq 2^{n\mathbb{H}(q)}.$$

23.2.2.9 Proof

Holds if $q = 0$ or $q = 1$, so assume $0 < q < 1$. We have

$$\binom{n}{nq} q^{nq} (1-q)^{n-nq} \leq (q + (1-q))^n = 1.$$

We also have: $q^{-nq} (1-q)^{-(1-q)n} = 2^{n(-q \lg q - (1-q) \lg(1-q))} = 2^{n\mathbb{H}(q)}$, we have

$$\binom{n}{nq} \leq q^{-nq} (1-q)^{-(1-q)n} = 2^{n\mathbb{H}(q)}.$$

23.2.3 Proof continued

23.2.3.1 Other direction...

- (A) $\mu(k) = \binom{n}{k} q^k (1-q)^{n-k}$
- (B) $\sum_{i=0}^n \binom{n}{i} q^i (1-q)^{n-i} = \sum_{i=0}^n \mu(i)$.
- (C) Claim: $\mu(nq) = \binom{n}{nq} q^{nq} (1-q)^{n-nq}$ largest term in $\sum_{k=0}^n \mu(k) = 1$.
- (D) $\Delta_k = \mu(k) - \mu(k+1) = \binom{n}{k} q^k (1-q)^{n-k} \left(1 - \frac{n-k}{k+1} \frac{q}{1-q}\right)$,
- (E) sign of Δ_k = size of last term...
- (F) $\text{sign}(\Delta_k) = \text{sign}\left(1 - \frac{(n-k)q}{(k+1)(1-q)}\right)$
 $= \text{sign}\left(\frac{(k+1)(1-q) - (n-k)q}{(k+1)(1-q)}\right)$.

23.2.3.2 Proof continued

- (A) $(k+1)(1-q) - (n-k)q = k+1 - kq - q - nq + kq = 1 + k - q - nq$.
- (B) $\implies \Delta_k \geq 0$ when $k \geq nq + q - 1$
 $\Delta_k < 0$ otherwise.
- (C) $\mu(k) = \binom{n}{k} q^k (1-q)^{n-k}$
- (D) $\mu(k) < \mu(k+1)$, for $k < nq$, and $\mu(k) \geq \mu(k+1)$ for $k \geq nq$.
- (E) $\implies \mu(nq)$ is the largest term in $\sum_{k=0}^n \mu(k) = 1$.
- (F) $\mu(nq)$ larger than the average in sum.
- (G) $\implies \binom{n}{k} q^k (1-q)^{n-k} \geq \frac{1}{n+1}$.
- (H) $\implies \binom{n}{nq} \geq \frac{1}{n+1} q^{-nq} (1-q)^{-(n-nq)} = \frac{1}{n+1} 2^{n\mathbb{H}(q)}$. ■

23.2.3.3 Generalization...

Corollary 23.2.6. *We have:*

$$(i) q \in [0, 1/2] \implies \binom{n}{\lfloor nq \rfloor} \leq 2^{n\mathbb{H}(q)}. \quad (ii) q \in [1/2, 1] \binom{n}{\lceil nq \rceil} \leq 2^{n\mathbb{H}(q)}.$$

$$(iii) q \in [1/2, 1] \implies \frac{2^{n\mathbb{H}(q)}}{n+1} \leq \binom{n}{\lfloor nq \rfloor}. \quad (iv) q \in [0, 1/2] \implies \frac{2^{n\mathbb{H}(q)}}{n+1} \leq \binom{n}{\lceil nq \rceil}.$$

Proof is straightforward but tedious.

23.2.3.4 What we have...

- (A) Proved that $\binom{n}{nq} \approx 2^{n\mathbb{H}(q)}$.
- (B) Estimate is loose.
- (C) Sanity check...
 - (I) A sequence of n bits generated by coin with probability q for head.
 - (II) By Chernoff inequality... roughly nq heads in this sequence.
 - (III) Generated sequence Y belongs to $\binom{n}{nq} \approx 2^{n\mathbb{H}(q)}$ possible sequences .
 - (IV) ...of similar probability.
 - (V) $\implies \mathbb{H}(Y) = n\mathbb{H}(q) \approx \lg \binom{n}{nq}$.

23.2.4 Extracting randomness

23.2.4.1 Just one bit...

question Given a coin C with:

p : Probability for head.

$q = 1 - p$: Probability for tail.

Q: How to get one true random bit, by flipping C .

Describe an algorithm!

23.2.4.2 Extracting randomness...

Entropy can be interpreted as the amount of unbiased random coin flips can be extracted from a random variable.

Definition 23.2.7. An extraction function **Ext** takes as input the value of a random variable X and outputs a sequence of bits y , such that $\Pr[\mathbf{Ext}(X) = y \mid |y| = k] = \frac{1}{2^k}$, whenever $\Pr[|y| = k] > 0$, where $|y|$ denotes the length of y .

23.2.4.3 Extracting randomness...

- (A) X : uniform random integer variable out of $0, \dots, 7$.
- (B) $\text{Ext}(X)$: binary representation of x .
- (C) Def. subtle: all extracted seqs of same len have same probability.
- (D) Another example of extraction scheme:
 - (A) X : uniform random integer variable $0, \dots, 11$.
 - (B) $\text{Ext}(x)$: output the binary representation for x if $0 \leq x \leq 7$.
 - (C) If x is between 8 and 11?
 - (D) Idea... Output binary representation of $x - 8$ as a two bit number.
- (E) A valid extractor...

$$\Pr[\text{Ext}(X) = 00 \mid |\text{Ext}(X)| = 2] = \frac{1}{4},$$

23.2.4.4 Technical lemma

The following is obvious, but we provide a proof anyway.

Lemma 23.2.8. *Let x/y be a fraction, such that $x/y < 1$. Then, for any i , we have $x/y < (x+i)/(y+i)$.*

Proof: We need to prove that $x(y+i) - (x+i)y < 0$. The left side is equal to $i(x-y)$, but since $y > x$ (as $x/y < 1$), this quantity is negative, as required. ■

23.2.4.5 A uniform variable extractor...

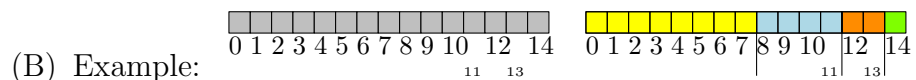
- Theorem 23.2.9.** (A) X : random variable chosen uniformly at random from $\{0, \dots, m-1\}$.
 (B) Then there is an extraction function for X :
 (A) outputs on average at least

$$\lfloor \lg m \rfloor - 1 = \lfloor \mathbb{H}(X) \rfloor - 1$$

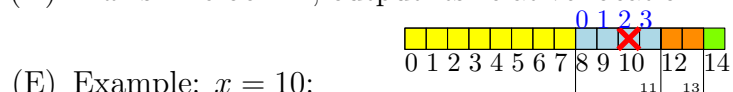
independent and unbiased bits.

23.2.4.6 Proof

- (A) m : A sum of unique powers of 2, namely $m = \sum_i a_i 2^i$, where $a_i \in \{0, 1\}$.



- (C) decomposed $\{0, \dots, m-1\}$ into disjoint union of blocks sizes are powers of 2.
 (D) If x is in block 2^k , output its relative location in the block in binary representation.



then falls into block 2^2 ...
 x relative location is 2. Output 2 written using two bits,
 Output: "10".

23.2.4.7 Proof continued

- (A) Valid extractor...
- (B) Theorem holds if m is a power of two. Only one block.
- (C) m not a power of 2...
- (D) X falls in block of size 2^k : then output k complete random bits..
... entropy is k .
- (E) Let $2^k < m < 2^{k+1}$ biggest block.
- (F) $u = \lfloor \lg(m - 2^k) \rfloor < k$.
There must be a block of size u in the decomposition of m .
- (G) two blocks in decomposition of m : sizes 2^k and 2^u .
- (H) Largest two blocks...
- (I) $2^k + 2 * 2^u > m \implies 2^{u+1} + 2^k - m > 0$.
- (J) Y : random variable = number of bits output by extractor.

23.2.4.8 Proof continued

- (A) By lemma, since $\frac{m-2^k}{m} < 1$:

$$\frac{m - 2^k}{m} \leq \frac{m - 2^k + (2^{u+1} + 2^k - m)}{m + (2^{u+1} + 2^k - m)} = \frac{2^{u+1}}{2^{u+1} + 2^k}.$$

- (B) By induction (assumed holds for all numbers smaller than m):

$$\begin{aligned} \mathbf{E}[Y] &\geq \frac{2^k}{m}k + \frac{m - 2^k}{m} \left(\underbrace{\lfloor \lg(m - 2^k) \rfloor}_u - 1 \right) \\ &= \frac{2^k}{m}k + \frac{m - 2^k}{m} (\underbrace{k - k}_{=0} + u - 1) \\ &= k + \frac{m - 2^k}{m} (u - k - 1) \end{aligned}$$

23.2.4.9 Proof continued..

- (A) We have:

$$\begin{aligned} \mathbf{E}[Y] &\geq k + \frac{m - 2^k}{m} (u - k - 1) \\ &\geq k + \frac{2^{u+1}}{2^{u+1} + 2^k} (u - k - 1) \\ &= k - \frac{2^{u+1}}{2^{u+1} + 2^k} (1 + k - u), \end{aligned}$$

since $u - k - 1 \leq 0$ as $k > u$.

- (B) If $u = k - 1$, then $\mathbf{E}[Y] \geq k - \frac{1}{2} \cdot 2 = k - 1$, as required.
- (C) If $u = k - 2$ then $\mathbf{E}[Y] \geq k - \frac{1}{3} \cdot 3 = k - 1$.

23.2.4.10 Proof continued.....

(A) $\mathbf{E}[Y] \geq k - \frac{2^{u+1}}{2^{u+1}+2^k}(1+k-u)$.

And $u - k - 1 \leq 0$ as $k > u$.

(B) If $u < k - 2$ then

$$\begin{aligned}\mathbf{E}[Y] &\geq k - \frac{2^{u+1}}{2^k}(1+k-u) \\ &= k - \frac{k-u+1}{2^{k-u-1}} \\ &= k - \frac{2+(k-u-1)}{2^{k-u-1}} \\ &\geq k-1,\end{aligned}$$

since $(2+i)/2^i \leq 1$ for $i \geq 2$.