Chapter 29

Entropy, Randomness, and Information

By Sariel Har-Peled, December 3, 2018[1]

> "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..

29.1. Entropy

Definition 29.1.1. The ***entropy*** in bits of a discrete random variable X is given by

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

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

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 $\mathbb{H}(0) = \mathbb{H}(1) = 0$. See Figure 29.1.

The function $\mathbb{H}(p)$ is a concave symmetric around $1/2$ on the interval $[0, 1]$ and achieves its maximum at $1/2$. For a concrete example, consider $\mathbb{H}(3/4) \approx 0.8113$ and $\mathbb{H}(7/8) \approx 0.5436$. Namely, a 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.

We have $\mathbb{H}'(p) = -\lg p + \lg(1-p) = \lg \frac{1-p}{p}$ and $\mathbb{H}''(p) = \frac{p}{1-p} \cdot \left(-\frac{1}{p^2}\right) = -\frac{1}{p(1-p)}$. Thus, $\mathbb{H}''(p) \leq 0$, for all $p \in (0, 1)$, and the $\mathbb{H}(\cdot)$ is concave in this range. Also, $\mathbb{H}'(1/2) = 0$, which implies that $\mathbb{H}(1/2) = 1$ is a maximum of the binary entropy. Namely, a balanced coin has the largest amount of randomness in it.

Example 29.1.2. A random variable X that has probability $1/n$ to be i, for $i = 1, \dots, n$, has entropy $\mathbb{H}(X) = -\sum_{i=1}^{n} \frac{1}{n} \lg \frac{1}{n} = \lg n$.

Note, that the entropy is oblivious to the exact values that the random variable can have, and it is sensitive only to the probability distribution. Thus, a random variables that accepts $-1, +1$ with equal probability has the same entropy (i.e., 1) as a fair coin.

Lemma 29.1.3. *Let X and Y be two independent random variables, and let Z be the random variable (X, Y). Then $\mathbb{H}(Z) = \mathbb{H}(X) + \mathbb{H}(Y)$.*

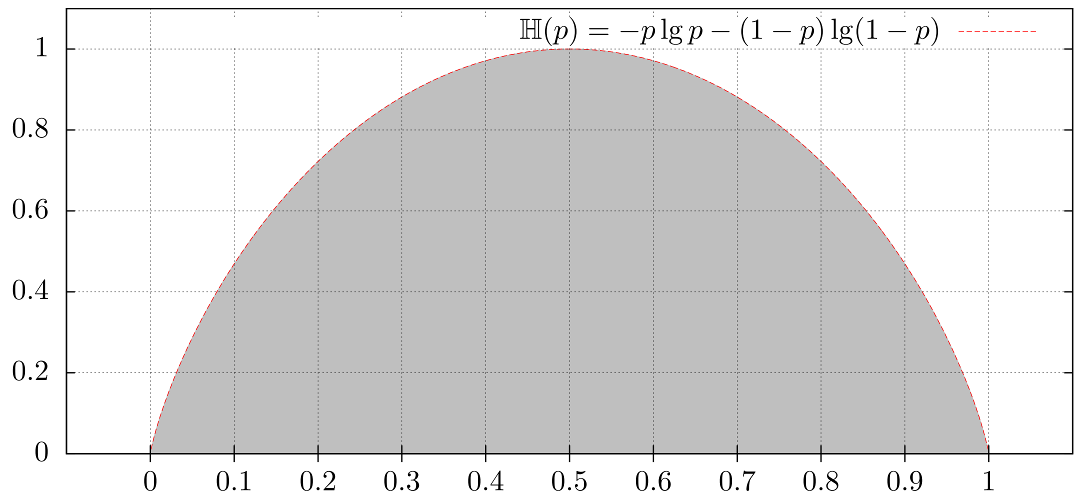

Figure 29.1: The binary entropy function.

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

$$\begin{aligned}
\mathbb{H}(Z) &= \sum_{x,y} \mathbb{P}[(X,Y) = (x,y)] \lg \frac{1}{\mathbb{P}[(X,Y) = (x,y)]} \\
&= \sum_{x,y} \mathbb{P}[X = x] \, \mathbb{P}[Y = y] \lg \frac{1}{\mathbb{P}[X = x] \, \mathbb{P}[Y = y]} \\
&= \sum_{x} \sum_{y} \mathbb{P}[X = x] \, \mathbb{P}[Y = y] \lg \frac{1}{\mathbb{P}[X = x]} \\
&\qquad + \sum_{y} \sum_{x} \mathbb{P}[X = x] \, \mathbb{P}[Y = y] \lg \frac{1}{\mathbb{P}[Y = y]} \\
&= \sum_{x} \mathbb{P}[X = x] \lg \frac{1}{\mathbb{P}[X = x]} + \sum_{y} \mathbb{P}[Y = y] \lg \frac{1}{\mathbb{P}[Y = y]} \\
&= \mathbb{H}(X) + \mathbb{H}(Y).
\end{aligned}$$

■

Lemma 29.1.4. *Suppose that nq is integer in the range $[0, n]$. Then* $\dfrac{2^{n\mathbb{H}(q)}}{n+1} \leq \dbinom{n}{nq} \leq 2^{n\mathbb{H}(q)}$.

Proof: This trivially holds if $q = 0$ or $q = 1$, so assume $0 < q < 1$. We know that

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

As such, since $q^{-nq}(1-q)^{-(1-q)n} = 2^{n(-q\lg q - (1-q)\lg(1-q))} = 2^{n\mathbb{H}(q)}$, we have

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

As for the other direction, let $\mu(k) = \binom{n}{k} q^k (1-q)^{n-k}$. We claim that $\mu(nq) = \binom{n}{nq} q^{nq}(1-q)^{n-nq}$ is the largest term in $\sum_{k=0}^{n} \mu(k) = 1$. Indeed,

$$\Delta_k = \mu(k) - \mu(k+1) = \binom{n}{k} q^k (1-q)^{n-k} \left(1 - \frac{n-k}{k+1} \frac{q}{1-q}\right),$$

and the sign of this quantity is the sign of the last term, which is

$$\operatorname{sign}(\Delta_k) = \operatorname{sign}\left(1 - \frac{(n-k)q}{(k+1)(1-q)}\right) = \operatorname{sign}\left(\frac{(k+1)(1-q) - (n-k)q}{(k+1)(1-q)}\right).$$

Now,

$$(k+1)(1-q) - (n-k)q = k + 1 - kq - q - nq + kq = 1 + k - q - nq.$$

Namely, $\Delta_k \geq 0$ when $k \geq nq + q - 1$, and $\Delta_k < 0$ otherwise. Namely, $\mu(k) < \mu(k+1)$, for $k < nq$, and $\mu(k) \geq \mu(k+1)$ for $k \geq nq$. Namely, $\mu(nq)$ is the largest term in $\sum_{k=0}^{n} \mu(k) = 1$, and as such it is larger than the average. We have $\mu(nq) = \binom{n}{nq} q^{nq}(1-q)^{n-nq} \geq \frac{1}{n+1}$, which implies

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

■

Lemma 29.1.4 can be extended to handle non-integer values of q. This is straightforward, and we omit the easy but tedious details.

Corollary 29.1.5. *We have:*

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

The bounds of Lemma 29.1.4 and Corollary 29.1.5 are loose but sufficient for our purposes. As a sanity check, consider the case when we generate a sequence of n bits using a coin with probability q for head, then by the Chernoff inequality, we will get roughly nq heads in this sequence. As such, the generated sequence Y belongs to $\binom{n}{nq} \approx 2^{n\mathbb{H}(q)}$ possible sequences that have similar probability. As such, $\mathbb{H}(Y) \approx \lg \binom{n}{nq} = n\mathbb{H}(q)$, by Example 29.1.2, a fact that we already know from Lemma 29.1.3.

29.1.1. Extracting randomness

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

Definition 29.1.6. An extraction function Ext takes as input the value of a random variable X and outputs a sequence of bits y, such that $\mathbb{P}\left[Ext(X) = y \;\middle|\; |y| = k \right] = \frac{1}{2^k}$, whenever $\mathbb{P}[|y| = k] > 0$, where $|y|$ denotes the length of y.

As a concrete (easy) example, consider X to be a uniform random integer variable out of $0, \ldots, 7$. All that $Ext(X)$ has to do in this case, is to compute the binary representation of x. However, note that Definition 29.1.6 is somewhat more subtle, as it requires that all extracted sequence of the same length would have the same probability.

Thus, for X a uniform random integer variable in the range $0, \ldots, 11$, the function $Ext(x)$ can output the binary representation for x if $0 \le x \le 7$. However, what do we do if x is between 8 and 11? The idea is to output the binary representation of $x - 8$ as a two bit number. Clearly, Definition 29.1.6 holds for this extraction function, since $\mathbb{P}\Big[Ext(X) = 00 \,\Big|\, |Ext(X)| = 2\Big] = \frac{1}{4}$, as required. This scheme can be of course extracted for any range.

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

Lemma 29.1.7. *Let x/y be a faction, 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 size is equal to $i(x - y)$, but since $y > x$ (as $x/y < 1$), this quantity is negative, as required. ∎

Theorem 29.1.8. *Suppose that the value of a random variable X is chosen uniformly at random from the integers $\{0, \ldots, m-1\}$. Then there is an extraction function for X that outputs on average at least $\lfloor \lg m \rfloor - 1 = \lfloor \mathbb{H}(X) \rfloor - 1$ independent and unbiased bits.*

Proof: We represent m as a sum of unique powers of 2, namely $m = \sum_i a_i 2^i$, where $a_i \in \{0, 1\}$. Thus, we decomposed $\{0, \ldots, m-1\}$ into a disjoint union of blocks that have sizes which are distinct powers of 2. If a number falls inside such a block, we output its relative location in the block, using binary representation of the appropriate length (i.e., k if the block is of size 2^k). One can verify that this is an extraction function, fulfilling Definition 29.1.6.

Now, observe that the claim holds trivially if m is a power of two. Thus, consider the case that m is not a power of 2. If X falls inside a block of size 2^k then the entropy is k. Thus, for the inductive proof, assume that are looking at the largest block in the decomposition, that is $m < 2^{k+1}$, and let $u = \left\lfloor \lg(m - 2^k) \right\rfloor < k$. There must be a block of size u in the decomposition of m. Namely, we have two blocks that we known in the decomposition of m, of sizes 2^k and 2^u. Note, that these two blocks are the largest blocks in the decomposition of m. In particular, $2^k + 2 * 2^u > m$, implying that $2^{u+1} + 2^k - m > 0$.

Let Y be the random variable which is the number of bits output by the extractor algorithm.

By Lemma 29.1.7, since $\frac{m - 2^k}{m} < 1$, we have

$$\frac{m - 2^k}{m} \le \frac{m - 2^k + \left(2^{u+1} + 2^k - m\right)}{m \qquad + \left(2^{u+1} + 2^k - m\right)} = \frac{2^{u+1}}{2^{u+1} + 2^k}.$$

Thus, by induction (we assume the claim holds for all integers smaller than m), we have

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

since $u - k - 1 \leq 0$ as $k > u$. If $u = k - 1$, then $\mathbb{E}[Y] \geq k - \frac{1}{2} \cdot 2 = k - 1$, as required. If $u = k - 2$ then $\mathbb{E}[Y] \geq k - \frac{1}{3} \cdot 3 = k - 1$. Finally, if $u < k - 2$ then

$$\mathbb{E}[Y] \geq k - \frac{2^{u+1}}{2^k}(1 + k - u) = k - \frac{k - u + 1}{2^{k-u-1}} = k - \frac{2 + (k - u - 1)}{2^{k-u-1}} \geq k - 1,$$

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

Bibliography