

# Entropy - Seminary 1

Radu Trîmbițaș

October 14, 2013

**Problem 1 (Coin flips)** *A fair coin is flipped until the first head occurs. Let  $X$  denote the number of flips required.*

(a) *Find the entropy  $H(X)$  in bits. The following expressions may be useful:*

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}, \quad \sum_{n=0}^{\infty} nr^n = \frac{r}{(1-r)^2}.$$

(b) *A random variable  $X$  is drawn according to this distribution. Find an "efficient" sequence of yes-no questions of the form, "Is  $X$  contained in the set  $S$ ?" Compare  $H(X)$  to the expected number of questions required to determine  $X$ .*

**Solution.**

(a) The number  $X$  of tosses till the first head appears has the geometric distribution with parameter  $p$  and  $P(X = n) = pq^{n-1}$ . The entropy is

$$\begin{aligned} H(X) &= - \sum_{n=1}^{\infty} pq^{n-1} \log(pq^{n-1}) \\ &= - \log p \sum_{n=1}^{\infty} pq^{n-1} - \log q \sum_{n=1}^{\infty} (n-1)pq^{n-1} \\ &= - \log p \sum_{n=0}^{\infty} pq^n - \log q \sum_{n=0}^{\infty} npq^n \\ &= \frac{-p \log p - q \log q}{p} = \frac{H(p)}{p}. \end{aligned}$$

In our case  $p = q = 1/2$  and  $H(X) = 2$  bits.

(b) Intuitively, it seems clear that the best questions are those that have equally likely chances of receiving a yes or a no answer. Consequently, one possible guess is that the most "efficient" series of questions is: Is  $X = 1$ ? If not, is  $X = 2$ ? If not, is  $X = 3$ ? . . . with a resulting expected

number of questions equal to  $\sum_{n=1}^{\infty} \frac{n}{2^n} = 2$ . This should reinforce the intuition that  $H(X)$  is a measure of the uncertainty of  $X$ . Indeed in this case, the entropy is exactly the same as the average number of questions needed to define  $X$ , and in general  $E(\# \text{ of questions}) \geq H(X)$ . This problem has an interpretation as a source coding problem. Let 0 =no, 1 =yes,  $X$  =Source, and  $Y$  =Encoded Source. Then the set of questions in the above procedure can be written as a collection of  $(X, Y)$  pairs: (1, 1), (2, 01), (3, 001), etc... In fact, this intuitively derived code is the optimal (Huffman) code minimizing the expected number of questions.

■

**Problem 2 (Entropy of functions)** Let  $X$  be a random variable taking on a finite number of values. What is the (general) inequality relationship of  $H(X)$  and  $H(Y)$  if

(a)  $Y = 2^X$

(b)  $Y = \cos X$ .

**Solution.** Let  $y = g(x)$ . Then

$$p(y) = \sum_{x:y=g(x)} p(x).$$

Consider any set of  $x$ 's that map onto a single  $y$ . For this set

$$\sum_{x:y=g(x)} p(x) \log p(x) \leq \sum_{x:y=g(x)} p(x) \log p(y) = p(y) \log p(y).$$

since  $\log$  is a monotone increasing function and  $p(x) \leq \sum_{x:y=g(x)} p(x) = p(y)$ . Extending this argument to the entire range of  $X$  (and  $Y$ ), we obtain

$$\begin{aligned} H(X) &= - \sum_x p(x) \log p(x) \\ &= - \sum_y \sum_{x:y=g(x)} p(x) \log p(x) \\ &\geq - \sum_y p(y) \log p(y) = H(Y). \end{aligned}$$

with equality iff  $g$  is one-to-one with probability one.

$Y = 2^X$  is one-to-one and hence the entropy, which is just a function of the probabilities (and not the values of a random variable) does not change, i.e.,  $H(X) = H(Y)$ .

$Y = \cos(X)$  is not necessarily one-to-one. Hence all that we can say is that  $H(X) \geq H(Y)$ , with equality if cosine is one-to-one on the range of  $X$ . ■

**Problem 3 (Minimum entropy)** What is the minimum value of  $H(p_1, \dots, p_n) = H(\mathbf{p})$  as  $\mathbf{p}$  ranges over the set of  $n$ -dimensional probability vectors? Find all  $\mathbf{p}$ 's which achieve this minimum.

**Solution.** We wish to find all probability vectors  $\mathbf{p} = (p_1, p_2, \dots, p_n)$  which minimize

$$H(\mathbf{p}) = - \sum_i p_i \log p_i$$

Now  $-p_i \log p_i \geq 0$ , with equality iff  $p_i = 0$  or  $1$ . Hence the only possible probability vectors which minimize  $H(\mathbf{p})$  are those with  $p_i = 1$  for some  $i$  and  $p_j = 0, j \neq i$ . There are  $n$  such vectors, i.e.,  $(1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$ , and the minimum value of  $H(\mathbf{p})$  is 0. ■

**Problem 4 (Entropy of functions of a random variable)** Let  $X$  be a discrete random variable. Show that the entropy of a function of  $X$  is less than or equal to the entropy of  $X$  by justifying the following steps:

$$\begin{aligned} H(X|g(X)) &\stackrel{(a)}{=} H(X) + H(g(X)|X) \\ &\stackrel{(b)}{=} H(X) \\ H(X, g(X)) &\stackrel{(c)}{=} H(g(X)) + H(X|g(X)) \\ &\stackrel{(d)}{\geq} H(g(X)). \end{aligned}$$

**Proof.**

- (a)  $H(X, g(X)) = H(X) + H(g(X)|X)$  by the chain rule for entropies.
- (b)  $H(g(X)|X) = 0$  since for any particular value of  $X$ ,  $g(X)$  is fixed, and hence  $H(g(X)|X) = \sum_x p(x)H(g(X)|X = x) = \sum_x 0 = 0$ .
- (c)  $H(X, g(X)) = H(g(X)) + H(X|g(X))$  again by the chain rule.
- (d)  $H(X|g(X)) \geq 0$ , with equality iff  $X$  is a function of  $g(X)$ , i.e.,  $g(\cdot)$  is one-to-one. Hence  $H(X, g(X)) \geq H(g(X))$ .

Combining parts (b) and (d), we obtain  $H(X) \geq H(g(X))$ . ■

**Problem 5 (Drawing with and without replacement)** An urn contains  $r$  red,  $w$  white, and  $b$  black balls. Which has higher entropy, drawing  $k \geq 2$  balls from the urn with replacement or without replacement? Set it up and show why. (There is both a hard way and a relatively simple way to do this.)

**Solution.** Intuitively, it is clear that if the balls are drawn with replacement, the number of possible choices for the  $i$ -th ball is larger, and therefore the conditional entropy is larger. But computing the conditional distributions is slightly involved. It is easier to compute the unconditional entropy.

- With replacement. In this case the conditional distribution of each draw is the same for every draw. Thus

$$X_i = \begin{cases} \text{red} & p_i = \frac{r}{r+w+b} \\ \text{white} & p_i = \frac{w}{r+w+b} \\ \text{black} & p_i = \frac{b}{r+w+b} \end{cases}$$

and therefore

$$\begin{aligned} H(X_i|X_{i-1}, \dots, X_1) &= H(X_i) \\ &= \log(r+w+b) - \frac{r \log r + w \log w + b \log b}{r+w+b} \end{aligned}$$

- Without replacement. The unconditional probability of the  $i$ -th ball being red is still  $r/(r+w+b)$ , etc. Thus the unconditional entropy  $H(X_i)$  is still the same as with replacement. The conditional entropy  $H(X_i|X_{i-1}, \dots, X_1)$  is less than the unconditional entropy, and therefore the entropy of drawing without replacement is lower.

■