MATH 280: HOMEWORK #2

DUE IN LECTURE ON FEB. 5, 2019.

1. Shannon entropy and selective advantage.

One of the themes we will discuss in the work of L. Demetrius is that the Shannon entropy of particular characteristics of a species is related to the selective advantage or disadvantage the species has in different environments. Here is an example we will later treat quantitatively.

Demetrius considers the Shannon entropy of the age at which female members of a species have offspring. The greater this entropy is, the more variable this age is. Suppose, for example, we are discussing seals. High entropy correlates with abundant resources for having offspring and a lack of constraining factors (e.g. threatening polar bears) which might make it more difficult for older less fit members of the group to have offspring successfully.

Demetrius argues that the selective advantage of a species is determined by whether its entropy level matches the environmental conditions. So, a seal species with high entropy in the above sense will do well when there are no polar bears around, but it would have trouble in an environment with a lot of polar bears.

Developing the mathematics behind this idea is one of our main goals. In this problem, the goal is to think qualitatively about how such principles might apply in other domains. In each case, explain your reasoning.

Problem 1. Consider the Shannon entropy of the age at which workers acquire a new job. What would it mean to say that this entropy is very near 0? What sorts of economic conditions might be correlated with having low Shannon entropy?

Problem 2. Consider the Shannon entropy of the age at which a member of a particular political party is first elected to power. What political characteristics of the party would likely be associated to this entropy being large? What characteristics would be associated to small values for this entropy? What would be the Shannon entropy of a regime in which there is a single ruler who rules for life?

2. Shannon entropy and communication rates

Suppose we have N messages, numbered $1, \ldots, N$, we would like to communicate digitally. Message number i has probability p_i of being sent. We have discussed in class the problem of associating to each i a finite string $b(i)$ of 0's and 1's in such a way that $b(i)$ is never an initial sequence of the digits of $b(j)$ if $1 \leq i \neq j \leq N$. Let B be the ordered sequence $\{b(1), \ldots, b(N)\}$ of binary digits used to encode messages. Suppose $b(i)$ has length $\ell(i)$, so that it involves exactly $\ell(i)$ 0's and 1's. The average number of digits per message that will sent when we use B to encode messages is

$$T(B, p_1, \ldots, p_N) = \sum_{i=1}^{N} p_i \cdot \ell(i). \tag{2.1}$$

This is because message i will be sent p_i of the time, and in this case we send $\ell(i)$ digits.

Problem 3. Show that given N and $p_1, \ldots, p_N$, there is a B for which $T(B, p_1, \ldots, p_N)$ is minimal over all possible choices of B. We will call this minimal value $T^{min}(p_1, \ldots, p_N)$. It represents the optimal economy in average bits per message one can achieve by encoding the messages $1, \ldots, N$ into bit strings $b(1), \ldots, b(N)$ as above.

(Hint: The issue here is that there are infinitely many possible choices for B, and not every infinite set of real numbers has a minimal element. Show that one can reduce to the case in which all the p_i are positive, since the i for which $p_i = 0$ have no bearing on $T(B, p_1, \ldots, p_N)$ in (2.1). Then show that you can reduce to considering only those B for which there is some bound on the length $\ell(i)$ of $b(i)$ for all $i = 1, \ldots, N$. In other words, once some $\ell(i)$ becomes sufficiently large, it is impossible that $T(B, p_1, \ldots, p_N)$ is minimal.)

Problem 4. Recall that the Shannon entropy is defined by

$$H(p_1, \ldots, p_N) = -\sum_{i=1}^{N} p_i \cdot \log_2(p_i) \tag{2.2}$$

In class we will discuss when

$$T^{min}(p_1, \ldots, p_N) = H(p_1, \ldots, p_N). \tag{2.3}$$

Show that if all the p_i are rational then $T^{min}(p_1, \ldots, p_N)$ is a rational number. Suppose now that all the $p_i = 1/N$. Use (2.2) to calculate $H(1/N, \ldots, 1/N)$. Then show that if

$$H(1/N, \ldots, 1/N) = T^{min}(1/N, \ldots, 1/N) \tag{2.4}$$

then N is a power of 2. You can use the fact that every positive integer can be factored in a unique way into a product of prime powers.

3. Comments

We'll sketch in class proofs of the following facts:

Theorem 3.1. *If $N = 2^m$ is a power of 2 and each p_i has the form $p_i = 1/2^{m(i)}$, then*

$$T^{min}(p_1, \ldots, p_N) = H(p_1, \ldots, p_N).$$

Notice that the conclusion of this Theorem definitely does not hold for all $(p_1, \ldots, p_N)$ when N is not a power of 2 by Problem 4. Even when $N = 2$, it does not hold for all (p_1, p_2), e.g. when $p_1 = 3/4$ and $p_2 = 1/4$. In this case, one must transmit at least one bit to distinguish between event 1 and event 2, so $T^{min}(3/4, 1/4) = 1$, but $H(3/4, 1/4) < 1$.

One can achieve a transmission rate approaching the Shannon entropy only by sending out long blocks of digits at a time:

Theorem 3.2. *Suppose N is arbitrary and that $(p_1, \ldots, p_N)$ is any probability vector of length N with real components. For each integer $k \geq 1$, consider all possible ordered sequences $q = (i_1, \ldots, i_k)$ of k elements of $\{1, \ldots, N\}$. The probability $p(q)$ of $q = (i_1, \ldots, i_k)$ occurring is $p(q) = p_{i_1} \cdot p_{i_2} \cdots p_{i_k}$ if we assume successive messages are independent of one another. There are N^k such q, and if we number them $q_1, \ldots, q_{N^k}$, one has the optimum transmission rate $T^{min}(q_1, \ldots, q_{N^k})$ for the expected number of bits one will need to send a sequence of k messages. One has*

$$\lim_{k \to \infty} \frac{1}{k} T^{min}(q_1, \ldots, q_{N^k}) = H(p_1, \ldots, p_N) \tag{3.5}$$

The $\frac{1}{k}$ factor on the left makes sense for the following reason. If one had an ideal digital encoding scheme, one would expect that the average amount of information needed to send out a sequence of k messages would be k times the amount needed to send out one message. In fact, it is not hard to show

$$\frac{1}{k} H(q_1, \ldots, q_{N^k}) = H(p_1, \ldots, p_N).$$