

Appendix to Lecture 4: Entropy Proofs.

1. Gibbs inequality:

To show: $\sum_i p_i \log \frac{p_i}{q_i} \geq 0$.

Let's show $\sum_i p_i \log \frac{q_i}{p_i} \leq 0$. But

$$\sum_i p_i \log \frac{q_i}{p_i} \leq \log\left(\sum_i p_i \cdot \frac{q_i}{p_i}\right) = \log\left(\sum_i q_i\right) = \log 1 = 0.$$

If equality holds, then $\frac{p_i}{q_i} = 1$, all i , so $p_i = q_i$ all i , because \log is strictly convex. Notice, I am being a bit sloppy here: there are special cases where some of the p_i may be zero, for example. These are left to the reader!

2. Entropy satisfies the three basic properties:

a) $-\sum_i p_i \log p_i$ is obviously continuous, since each $p_i \log p_i$ is continuous.

b) $H\left(\left(\frac{1}{k}, \dots, \frac{1}{k}\right)\right) = -\sum_i \frac{1}{k} \log \frac{1}{k} = \log k \leq \log(k+1) = H\left(\left(\frac{1}{k+1}, \dots, \frac{1}{k+1}\right)\right)$.

c) Let $A = \{x_1, \dots, x_k\}$ be the set of possible outcomes, partitioned into $A_j, j = 1, \dots, \ell$. Let $q_j = \sum_{x_i \in A_j} p_i$, where $p_i = p(x_i)$. Then

$$\begin{aligned}
 H(P) &= -\sum p_i \log p_i \\
 &= \sum_{j=1}^{\ell} \sum_{x_i \in A_j} -p_i \log p_i \\
 &= \sum_{j=1}^{\ell} q_j \sum_{x_i \in A_j} -\frac{p_i}{q_j} \log p_i \\
 &= \sum_{j=1}^{\ell} q_j \sum_{x_i \in A_j} -\frac{p_i}{q_j} \left\{ \log \frac{p_i}{q_j} - \frac{p_i}{q_j} \log q_j \right\} \\
 &= \sum_{j=1}^{\ell} q_j H(P_j) + \sum_{j=1}^{\ell} q_j \sum_{x_i \in A_j} -\frac{p_i}{q_j} \log q_j \\
 &= \sum_{j=1}^{\ell} q_j H(P_j) + \sum_{j=1}^{\ell} -q_j \log q_j \\
 &= \sum_{j=1}^{\ell} q_j H(P_j) + H(Q).
 \end{aligned}$$

Here $\sum_{j=1}^{\ell} q_j H(P_j) = H(P|Q)$ is the conditional entropy in terms of the later notation in the main lecture. Recapping: $H(P) = H(Q) + H(P|Q)$.

3. Uniqueness of H :

If H satisfies the three rules just demonstrated, we will show that

$$H((p_1, \dots, p_k)) = - \sum_{i=1}^k c p_i \log p_i,$$

where c is a positive constant. To see this, let's call $F(k) = H((\frac{1}{k}, \dots, \frac{1}{k}))$. We want to show that $F(k) = c \log k$. We are assuming $F(k) \leq F(k+1) \leq F(k+2)$ etc. Use the partition identity to break up a uniform distribution of r^m outcomes into m partitions of r outcomes: Check that this gives

$$F(r^m) = mF(r).$$

Now, let r, s, n be any whole numbers > 1 . We can choose m so that

$$r^m \leq s^n \leq r^{m+1}.$$

Then

$$m \log r \leq n \log s \leq (m+1) \log r,$$

or

$$\frac{m}{n} \leq \frac{\log s}{\log r} \leq \frac{m}{n} + \frac{1}{n}.$$

Similarly,

$$F(r^m) \leq F(s^n) \leq F(r^{m+1}),$$

or,

$$mF(r) \leq nF(s) \leq (m+1)F(r),$$

or,

$$\frac{m}{n} \leq \frac{F(s)}{F(r)} \leq \frac{m}{n} + \frac{1}{n}.$$

Thus,

$$\left| \frac{F(s)}{F(r)} - \frac{\log s}{\log r} \right| \leq \frac{1}{n}.$$

Since n is arbitrary and this last inequality is independent of m , we get that

$$\frac{F(s)}{F(r)} = \frac{\log s}{\log r},$$

or,

$$F(s) = c \log s,$$

where $c > 0$ is a constant independent of s .

For a general distribution $Q = (q_1, \dots, q_k)$ we proceed in two steps. First, assume that all of q_j are rational numbers. Then there is a least common denominator n so that $q_j = \frac{n_j}{n}$, $j = 1, \dots, k$. Now consider $P = (\frac{1}{n}, \dots, \frac{1}{n})$ the uniform distribution of n outcomes. We can bin this into subsets A_1, \dots, A_k , with A_j having n_j equally likely outcomes. The distribution for the binning gives $P(A_j) = \frac{n_j}{n} = q_j$, that is, just the distribution Q . So, using the partition formula and what we just proved about uniform distributions, we get

$$\begin{aligned} c \log n &= H(P) = H(Q) + \sum q_j H(P_j) \\ &= H(Q) + c \sum q_j \log n_j. \end{aligned}$$

Solving, this gives

$$H(Q) = c \log n - \sum q_j \log n_j = -c \sum q_j \log q_j.$$

This proves the result when all q_j are rational. The case of arbitrary q_j 's follows by continuity.