EE 231A
Information Theory
Instructor: Rick Wesel

Handout #6, Solution Set 2
Wednesday April 15, 2020

105 pts

Reading: Chapters 3 and 4 of *Elements of Information Theory*

Lectures 3A: Types of Convergence, the weak law of large numbers, the AEP

1. (8 pts) (*Relative Entropy AEP*).

   (a) Since the $X_1, X_2, \ldots, X_n$ are i.i.d., so are $q(X_1), q(X_2), \ldots, q(X_n)$, and hence we can apply the weak law of large numbers as with the regular AEP to obtain

$$-\frac{1}{n}\log q(X_1, X_2, \ldots, X_n) = \lim -\frac{1}{n}\sum \log q(X_i) \tag{1}$$

$$\rightarrow -E_p(\log q(X)) \quad \text{in probability.} \tag{2}$$

$$\tag{3}$$

   Note that expectation is with respect to p because the X_i are drawn according to p not q. Now express this quantity in the desired form.

$$-E_p(\log q(X)) = -\sum p(x)\log q(x) \tag{4}$$

$$= \sum p(x)\log\frac{p(x)}{q(x)} - \sum p(x)\log p(x) \tag{5}$$

$$= D(\mathbf{p}||\mathbf{q}) + H(\mathbf{p}). \tag{6}$$

   (b) Again, by the weak law of large numbers,

$$-\frac{1}{n}\log\frac{q(X_1, X_2, \ldots, X_n)}{p(X_1, X_2, \ldots, X_n)} = -\frac{1}{n}\sum \log\frac{q(X_i)}{p(X_i)} \tag{7}$$

$$\rightarrow -E_p(\log\frac{q(X)}{p(X)}) \quad \text{in probability.} \tag{8}$$

$$\tag{9}$$

   Expressing in the desired form:

$$-E_p(\log\frac{q(X)}{p(X)}) = -\sum p(x)\log\frac{q(x)}{p(x)} \tag{10}$$

$$= \sum p(x)\log\frac{p(x)}{q(x)} \tag{11}$$

$$= D(\mathbf{p}||\mathbf{q}). \tag{12}$$

   Independent, identically-distributed random variables obey the strong law of large numbers as well. Thus we also have convergence with probability one in the two cases above.

2. (8 pts) *Take it to the limit.*

The sequence pair (x^n, y^n) is drawn i.i.d. according to the p.m.f.

$$p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, y_i)\,. \tag{13}$$

i.e. The pairs are independent of each other but the x_i and y_i within a pair (x_i, y_i) are dependent according to the joint distribution $p(x, y)$.

What is the limit as $n \to \infty$ of

$$\frac{1}{n} \log \frac{p(x^n, y^n)}{p(x^n)p(y^n)}\,? \tag{14}$$

To get full credit you must express your answer in the simplest form and show your argument.

$$\begin{aligned}
\frac{1}{n} \log \frac{p(x^n, y^n)}{p(x^n)p(y^n)} &= \frac{1}{n} \sum_{i=1}^{n} \log \frac{p(x_i, y_i)}{p(x_i)p(y_i)} && (15)\\
&= \frac{1}{n} \sum_{i=1}^{n} \log p(x_i, y_i) - \frac{1}{n} \sum_{i=1}^{n} \log p(x_i) - \frac{1}{n} \sum_{i=1}^{n} \log p(y_i) && (16)\\
&\longrightarrow -H(X, Y) + H(X) + H(Y) && (17)\\
&= I(X; Y) && (18)
\end{aligned}$$

Lectures 3B: Properties of the typical set

3. (20 pts) *The AEP in action.*

(a)

$$H(X) = -\frac{1}{4}\log(\frac{1}{4}) - \frac{3}{4}\log(\frac{3}{4}) \tag{19}$$
$$= -\frac{1}{4}\left(\log(1) - \log(4)\right) - \frac{3}{4}\left(\log(3) - \log(4)\right) \tag{20}$$
$$= (\frac{3}{4} + \frac{1}{4})\log(4) - \frac{3}{4}\log(3) \tag{21}$$
$$= 2 - \frac{3}{4}\log 3 \tag{22}$$

(b) k is the number of ones.

$$-\frac{1}{n}\log p(x_1, x_2, \ldots, x_n) = -\frac{1}{n}\log\left((\frac{3}{4})^k \times (\frac{1}{4})^{n-k}\right) \tag{23}$$
$$= -\frac{1}{n}\left(\log(\frac{3}{4})^k + \log(\frac{1}{4})^{n-k}\right) \tag{24}$$
$$= -\frac{n-k}{n}\log(\frac{1}{4}) - \frac{k}{n}\log(\frac{3}{4}) \tag{25}$$
$$= 2 - \frac{k}{n}\log 3. \tag{26}$$

(c) Recall that membership in the $A_\epsilon^{(n)}$ is equivalent to the condition

$$|-\frac{1}{n}\log p(x_1, x_2, \ldots, x_n) - H(X)| \leq \epsilon \tag{27}$$

We can construct the following table describing every possible $n = 8$ sequence:

k	$\binom{n}{k}$	$p(x_1, x_2, \ldots, x_n)$	$-\frac{1}{n}\log p(x_1, x_2, \ldots, x_n) - H(X)$
0	1	0.00001526	1.1887
1	8	0.00004578	0.9906
2	28	0.00013733	0.7925
3	56	.00041199	0.5944
4	70	0.00123596	0.3962
5	56	0.00370789	0.1981
6	28	0.01112366	0
7	8	0.03337097	-0.1981
8	1	0.10011292	-0.3962

For $\epsilon = 0.2$ The members of the typical set are the sequences with $k \in \{5, 6, 7\}$ so we have

$$\Pr(A_\epsilon^{(n)}) = 56 \times 0.00370789 + 28 \times 0.01112366 + 8 \times 0.03337097 \quad (28)$$
$$= 0.7861 \quad (29)$$

and $|A_\epsilon^{(n)}| = 56 + 28 + 8 = 92$ or 36% of the 2^8 possible sequences.

(d) For $n = 16$ and $\epsilon = 0.2$ we can construct the full 17 row table or realize that only a few consecutive rows will be in $A_\epsilon^{(n)}$. We show an abbreviated table below that includes the sequences in $A_\epsilon^{(n)}$ and the sequences with values of k that are too large or too small by 1 for the sequence to be in $A_\epsilon^{(n)}$.

k	$\binom{n}{k}$	$p(x_1, x_2, \ldots, x_n)$	$-\frac{1}{n}\log p(x_1, x_2, \ldots, x_n) - H(X)$
9	11,440	0.00000458	0.2972
10	8,008	0.00001375	0.1981
11	4,368	0.00004125	0.0991
12	1,820	0.00012374	0
13	560	0.00037121	-0.0991
14	120	0.00111362	-0.1981
15	16	0.00334087	-0.2972

$$\Pr(A_\epsilon^{(n)}) = 8008 \times 0.00001375 + 4368 \times 0.00004125 \quad (30)$$
$$+ 1820 \times 0.00012374 \quad (31)$$
$$+ 560 \times 0.00037121 \quad (32)$$
$$+ 120 \times 0.00111362 \quad (33)$$
$$= 0.8570 \quad (34)$$

and $|A_\epsilon^{(n)}| = 8008 + 4368 + 1820 + 560 + 120 = 14,876$. This is 23% of the 2^{16} possible sequences.

(e) For a fixed ϵ, as n gets larger $\Pr(A_\epsilon^{(n)})$ increases to one. We see this in our example.

(f) We need to check that $|A_\epsilon^{(n)}| \leq 2^{n(H+\epsilon)}$.
H(X) = 0.8113 and $\epsilon = 0.2$ so

$$2^{n(H+\epsilon)} = 2^{n\times 1.0113} \tag{35}$$
$$> 2^n\,. \tag{36}$$

Thus the inequality is trivially satisfied in this case since we must have

$$|A_\epsilon^{(n)}| \leq 2^n \tag{37}$$

Lectures 3C and 3D: AEP data compression, high probability sets vs. typical sets.

4. (12 pts)

The AEP and source coding.

(a) The number of 100-bit binary sequences with three or fewer ones is

$$\binom{100}{0} + \binom{100}{1} + \binom{100}{2} + \binom{100}{3} = 1 + 100 + 4950 + 161700 = 166751\,.$$

The required codeword length is $\lceil \log_2 166751 \rceil = 18$.

(b) Using 18 bits instead of 100 gives about a 5 to one compression. However, we are still not that close to the entropy. $H(0.005) = 0.0454$, so we should be able to get by with 4.54 bits.

(c) The probability that a 100-bit sequence has three or fewer ones is

$$\sum_{i=0}^{3} \binom{100}{i} (0.005)^i (0.995)^{100-i} = 0.60577 + 0.30441 + 0.7572 + 0.01243 = 0.99833$$

Thus the probability that the sequence that is generated cannot be encoded is $1 - 0.99833 = 0.00167$.

Lecture 4A,B,C: Entropy rate, the general AEP,Entropy Rate of stationary processes.

5. (8 pts) *Time's arrow.* By the chain rule for entropy,

$$\begin{aligned} H(X_0|X_{-1},\ldots,X_{-n}) &= H(X_0,X_{-1},\ldots,X_{-n}) - H(X_{-1},\ldots,X_{-n}) && (38)\\ &= H(X_0,X_1,X_2,\ldots,X_n) - H(X_1,X_2,\ldots,X_n) && (39)\\ &= H(X_0|X_1,X_2,\ldots,X_n), && (40) \end{aligned}$$

where (39) follows from stationarity.

6. (8 pts) *Average entropy per element vs. conditional entropy.* By stationarity we have for all $1 \le i \le n$,

$$H(X_n|X^{n-1}) \le H(X_i|X^{i-1}),$$

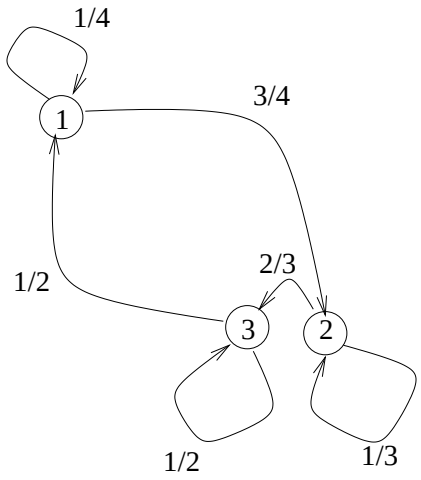
which implies that,

$$\begin{aligned} H(X_n|X^{n-1}) &= \frac{\sum_{i=1}^n H(X_n|X^{n-1})}{n} && (41)\\ &\le \frac{\sum_{i=1}^n H(X_i|X^{i-1})}{n} && (42)\\ &= \frac{H(X_1,X_2,\ldots,X_n)}{n}. && (43) \end{aligned}$$

Lectures 4D and 4E: Stationary Markov chains and the stationary distribution, entropy rate for Markov chains including random walks on a weighted graph.

7. (10 pts) *Adam's Seat Selection.*

   Consider the three-state stationary Markov process governing Adam's choice of a seat for the 231A lecture. Find the stationary probabilities and compute the entropy rate for Adam's seat selection process.



$$\frac{3}{4}\mu_1 = \frac{2}{3}\mu_2 \Longrightarrow \mu_2 = \frac{9}{8}\mu_1 \tag{44}$$

$$\frac{1}{2}\mu_3 = \frac{3}{4}\mu_1 \Longrightarrow \mu_3 = \frac{3}{2}\mu_1 \tag{45}$$

$$\mu_1 + \mu_2 + \mu_3 = 1 \Longrightarrow (1 + \frac{9}{8} + \frac{3}{2})\mu_1 = 1 \tag{46}$$

$$\Longrightarrow \mu_1 = \frac{8}{29} \tag{47}$$

$$\Longrightarrow \mu_2 = \frac{9}{29} \tag{48}$$

$$\Longrightarrow \mu_3 = \frac{12}{29} \tag{49}$$

$$H(\{X\}) = \mu_1 H\left(\frac{1}{4}\right) + \mu_2 H\left(\frac{1}{3}\right) + \mu_3 H\left(\frac{1}{2}\right) \tag{50}$$

$$= 0.9226 \text{ bits.} \tag{51}$$

8. (8 pts) *Random walk on chessboard.* Find the entropy rate of the Markov chain associated with a random walk of a king on the 3×3 chessboard shown below.

1	2	3
4	5	6
7	8	9

Assume that the king must move at each step in the Markov process and that it is equally likely to choose any of the legal chess moves for a king that are available to it. i.e. It can't stay in the same square and it can move to any adjacent square.

The distribution for the king's initial position is the stationary distribution.

The stationary distribution is given by $\mu_i = W_i/2W$, where W_i = number of edges emanating from node i and $2W = \sum_{i=1}^{9} W_i$. By inspection, $W_1 = W_3 = W_7 = W_9 = 3$, $W_2 = W_4 = W_6 = W_8 = 5$, $W_5 = 8$ and $2W = 40$, so $\mu_1 = \mu_3 = \mu_7 = \mu_9 = 3/40$, $\mu_2 = \mu_4 = \mu_6 = \mu_8 = 5/40$ and $\mu_5 = 8/40$. In a random walk the next state is chosen with equal probability among possible choices, so $H(X_2|X_1 = i) = \log 3$ bits for $i = 1, 3, 7, 9$, $H(X_2|X_1 = i) = \log 5$ for $i = 2, 4, 6, 8$ and $H(X_2|X_1 = i) = \log 8$ bits for $i = 5$. Therefore, we can calculate the entropy rate of the king as

$$\mathcal{H} = \sum_{i=1}^{9} \mu_i H(X_2|X_1 = i) \tag{52}$$

$$= 0.3 \log 3 + 0.5 \log 5 + 0.2 \log 8 \tag{53}$$

$$= 2.24 \text{ bits.} \tag{54}$$

9. (8 pts) *Random Walk of a Spider.*

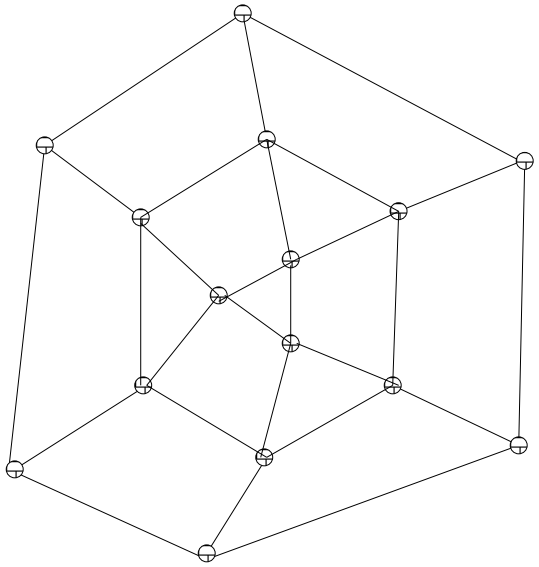


Figure 1: Spider web.

Compute the entropy rate for the random walk of a spider on the web shown in Figure 1.

At each step in the random walk the spider must move to an adjacent node. The spider is equally likely to choose each of the adjacent nodes. Assume that the initial node of the spider is random with the stationary probability mass function.

Solution: There are 6 outer nodes with 3 edges each and 9 inner nodes with 4 edges each. All edges have weight $w_i = 1$.

$$2W = (6 \times 3) + (9 \times 4) = 54 \tag{55}$$

The stationary probabilities are 3/54 for outer nodes and 4/54 for inner nodes.

$H(X_n|X_{n-1})$ is $\log 3$ for outer nodes and $\log 4$ for inner nodes.

$$H(\mathcal{X}) = 6 \times 3/54 \times \log 3 + 9 \times 4/54 \times \log 4 = 1.86 \tag{56}$$

Lecture 1E: Why does entropy involve a logarithm?

10. (15 pts)

(a) For convenience in notation, we will let

$$S_k = \sum_{i=1}^{k} p_i \tag{57}$$

and we will denote $H_2(q, 1-q)$ as $h(q)$. Then we can write the grouping axiom as

$$H_m(p_1, \ldots, p_m) = H_{m-1}(S_2, p_3, \ldots, p_m) + S_2 h\left(\frac{p_2}{S_2}\right). \tag{58}$$

Applying the grouping axiom again, we have

$$\begin{aligned} H_m(p_1, \ldots, p_m) = & \quad H_{m-1}(S_2, p_3, \ldots, p_m) + S_2 h\left(\frac{p_2}{S_2}\right) && (59)\\ = & \quad H_{m-2}(S_3, p_4, \ldots, p_m) + S_3 h\left(\frac{p_3}{S_3}\right) + S_2 h\left(\frac{p_2}{S_2}\right) && (60)\\ & \vdots && (61)\\ = & \quad H_{m-(k-1)}(S_k, p_{k+1}, \ldots, p_m) + \sum_{i=2}^{k} S_i h\left(\frac{p_i}{S_i}\right). && (62) \end{aligned}$$

Now, we apply the same grouping axiom repeatedly to $H_k(p_1/S_k, \ldots, p_k/S_k)$, to obtain

$$\begin{aligned} H_k\left(\frac{p_1}{S_k}, \ldots, \frac{p_k}{S_k}\right) = & \quad H_2\left(\frac{S_{k-1}}{S_k}, \frac{p_k}{S_k}\right) + \sum_{i=2}^{k-1} \frac{S_i}{S_k} h\left(\frac{p_i/S_k}{S_i/S_k}\right) && (63)\\ = & \quad \frac{1}{S_k} \sum_{i=2}^{k} S_i h\left(\frac{p_i}{S_i}\right). && (64) \end{aligned}$$

From (62) and (64), it follows that

$$H_m(p_1, \ldots, p_m) = H_{m-k+1}(S_k, p_{k+1}, \ldots, p_m) + S_k H_k\left(\frac{p_1}{S_k}, \ldots, \frac{p_k}{S_k}\right), \tag{65}$$

which is the extended grouping axiom.

(b) We need to use an axiom that is not explicitly stated in the text, namely that the function H_m is symmetric with respect to its arguments. Using this, we can combine any set of arguments of H_m using the extended grouping axiom.

Consider

$$f(mn) = H_{mn}(\frac{1}{mn}, \frac{1}{mn}, \ldots, \frac{1}{mn}). \tag{66}$$

By repeatedly applying the extended grouping axiom, we have

$$
\begin{aligned}
f(mn) = & \quad H_{mn}(\frac{1}{mn}, \frac{1}{mn}, \ldots, \frac{1}{mn}) && (67)\\
= & \quad H_{mn-(n-1)}(\frac{1}{m}, \frac{1}{mn}, \ldots, \frac{1}{mn}) + \frac{1}{m} H_n(\frac{1}{n}, \ldots, \frac{1}{n}) && (68)\\
= & \quad H_{mn-2(n-1)}(\frac{1}{m}, \frac{1}{m}, \frac{1}{mn}, \ldots, \frac{1}{mn}) + \frac{2}{m} H_n(\frac{1}{n}, \ldots, \frac{1}{n}) && (69)\\
& \vdots && (70)\\
= & \quad H_m(\frac{1}{m}, \ldots . \frac{1}{m}) + H(\frac{1}{n}, \ldots, \frac{1}{n} && (71)\\
= & \quad f(m) + f(n). && (72)
\end{aligned}
$$

We can immediately use this to conclude that $f(m^k) = kf(m)$.

(c) Note that in this problem only, the subscript of H is not the base of the logarithm but rather the number of arguments of H. Another way of stating the original Normalization Axiom is $H_2(\frac{1}{2}, \frac{1}{2}) = \log_2(2)$. The new Normalization Axiom is $H_2(\frac{1}{2}, \frac{1}{2}) = \log_e(2) = \ln(2)$.