

Homework no. 3

1. *Entropy of a sum (EIT 2.18).* Let X and Y be random variables that take on values x_1, x_2, \dots, x_r and y_1, y_2, \dots, y_s , respectively. Let $Z = X + Y$.

- Show that $H(Z|X) = H(Y|X)$. Argue that if X, Y are independent, then $H(Y) \leq H(Z)$ and $H(X) \leq H(Z)$. Thus the addition of *independent* random variables adds uncertainty.
- Give an example (of necessarily dependent random variables) in which $H(X) > H(Z)$ and $H(Y) > H(Z)$.
- Under what conditions does $H(Z) = H(X) + H(Y)$?

2. *Strong vs. Weak Typicality.* In class we defined the typical set $\mathcal{A}_\epsilon^{(n)}$ as the set of sequences \mathbf{x}^n such that

$$\left| -\frac{1}{n} \log \mathbb{P}(\mathbf{x}^n) - H(X) \right| \leq \epsilon.$$

These sequences are called *weakly typical*. A sequence \mathbf{x}^n is strongly typical if it has roughly the right number of occurrences of each of the source alphabet symbols, for example,

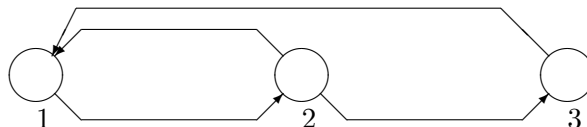
$$\left| \frac{\sum_i 1(x_i = a)}{n} - \mathbb{P}(X = a) \right| \leq \delta$$

for every $a \in \mathcal{X}$, where $1(\cdot)$ is the indicator function equal to 1 if its argument is true, 0 otherwise (it is also known as the Kronecker delta).

- Show that every strongly typical sequence is weakly typical (that is, fix δ , there is an $\epsilon(\delta)$ such that every δ -strongly typical sequence is also $\epsilon(\delta)$ -weakly typical, and $\epsilon(\delta) \rightarrow 0$ as $\delta \rightarrow 0$).
 - Show that a weakly typical sequence need not be strongly typical (construct a simple example, recall the best friend of the statistician).
3. *Entropy rate of constrained sequences.* In magnetic recording, the mechanism of recording and reading the bits imposes constraints on the sequences of bits that can be recorded. For example, to ensure proper synchronization, it is often necessary to limit the length of runs of 0's between two 1's. Also to reduce intersymbol interference, it may be necessary to require at least one 0 between any two 1's. We will consider a simple example of such a constraint.

Suppose that we are required to have at least one 0 and at most two 0's between any pair of 1's in a sequences. Thus, sequences like 101001 and 0101001 are valid sequences, but 0110010 and 0000101 are not. We wish to calculate the number of valid sequences of length n .

- Show that the set of constrained sequences is the same as the set of allowed paths on the following state diagram:



- (b) Let $X_i(n)$ be the number of valid paths of length n ending at state i . Argue that $\mathbf{X}(n) = [X_1(n) \ X_2(n) \ X_3(n)]^T$ satisfies the following recursion:

$$\begin{bmatrix} X_1(n) \\ X_2(n) \\ X_3(n) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1(n-1) \\ X_2(n-1) \\ X_3(n-1) \end{bmatrix}, \quad (1)$$

with initial conditions $\mathbf{X}(1) = [1 \ 1 \ 0]^T$.

- (c) Let

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \quad (2)$$

Then we have by induction

$$\mathbf{X}(n) = A\mathbf{X}(n-1) = A^2\mathbf{X}(n-2) = \dots = A^{n-1}\mathbf{X}(1). \quad (3)$$

Using the eigenvalue decomposition of A for the case of distinct eigenvalues, we can write $A = U^{-1}\Lambda U$, where Λ is a matrix of eigenvalues. Then $A^{n-1} = U^{-1}\Lambda^{n-1}U$. Show that we can write

$$\mathbf{X}(n) = \lambda_1^{n-1}\mathbf{Y}_1 + \lambda_2^{n-1}\mathbf{Y}_2 + \lambda_3^{n-1}\mathbf{Y}_3, \quad (4)$$

where $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3$ do not depend on n . For large n , this sum is dominated by the largest term. Therefore argue that for $i = 1, 2, 3$, we have

$$\frac{1}{n} \log X_i(n) \rightarrow \log \lambda, \quad (5)$$

where λ is the largest (positive) eigenvalue. Thus the number of sequences of length n grows as λ^n for large n . Calculate λ for the matrix A above.

- (d) We will now take a different approach. Consider a Markov chain whose state diagram is the one given in part (a), but with arbitrary transition probabilities. Therefore the probability transition matrix of this Markov chain is

$$P = \begin{bmatrix} 0 & \alpha & 1 \\ 1 & 0 & 0 \\ 0 & 1-\alpha & 0 \end{bmatrix}. \quad (6)$$

Show that the stationary distribution of this Markov chain is

$$\mu = \left[\frac{1}{3-\alpha}, \frac{1}{3-\alpha}, \frac{1-\alpha}{3-\alpha} \right]^T. \quad (7)$$

- (e) Maximize the entropy rate of the Markov chain over choices of α . What is the maximum entropy rate of the chain?
- (f) Compare the maximum entropy rate in part (e) with $\log \lambda$ in part (c). Why are the two answers the same?