Entropy Rate of a Markov Chain

For a stationary Markov chain the entropy rate is given by:

$$H(\chi) = H(\chi)' = \lim H(X_n \mid X_{n-1},\dots,X_1)$$
$$= \lim H(X_n \mid X_{n-1}) = H(X_2 \mid X_1)$$

Where the conditional entropy is computed using the given stationary distribution. Recall that the stationary distribution μ is the solution of the equations:

$$\mu = \sum_i \mu_i P_{ij} \qquad \text{for all j.}$$

We explicitly express the conditional entropy in the following slide.

Conditional Entropy Rate for a SMC

Theorem (*Conditional Entropy rate of a MC*): Let $\{X_i\}$ be a SMC with stationary distribution μ and transition matrix P. Let $X_1 \sim \mu$. Then the entropy rate is:

$$H(\chi) = -\sum_{ij} \mu_i P_{ij} \log P_{ij}$$

Proof:

$$H(\chi) = H(X_2 \mid X_1) = \sum_i \mu_i \Big(\sum_j -P_{ij} \log P_{ij}\Big)$$

Example (*Two state MC*): The entropy rate of the two state Markov chain in the previous example is:

$$H(\chi) = H(X_2 \mid X_1) = \frac{\beta}{\alpha+\beta} H(\alpha) + \frac{\alpha}{\alpha+\beta} H(\beta)$$

If the Markov chain is irreducible and aperiodic, it has unique stationary distribution on the states, and any initial distribution tends to the stationary distribution as n grows.

Example: ER of Random Walk

As an example of stochastic process lets take the example of a random walk on a connected graph. Consider a graph with m nodes with weight $W_{ij} \geq 0$ on the edge joining node i with node j. A particle walk randomly from node to node in this graph.

The random walk is X_m is a sequence of vertices of the graph. Given $X_n = i$, the next vertex j is choosen from among the nodes connected to node i with a probability proportional to the weight of the edge connecting i to j.

Thus,

$$P_{ij} = \frac{W_{ij}}{\sum_k W_{ik}}$$

ER of a Random Walk

In this case the stationary distribution has a surprisingly simple form, which we will guess and verify. The stationary distribution for this MC assigns probability to node i proportional to the total weight of the edges emanating from node i. Let:

$$W_i = \sum_j W_{ij}$$

Be the total weight of edges emanating from node i and let

$$W = \sum_{i,j:j>i} W_{ij}$$

Be the sum of weights of all the edges. Then $\sum_i W_i = 2W$. We now guess that the stationary distribution is:

$$\mu_i = \frac{W_i}{2W}$$

ER of Random Walk

We check that μP=μ:

$$\sum_i \mu_i P_{ij} = \sum_i \frac{W_i}{2W}\frac{W_{ij}}{W_i} = \sum_i \frac{W_{ij}}{2W} = \frac{W_j}{2W} = \mu_j$$

Thus, the stationary probability of state i is proportional to the weight of edges emanating from node i. This stationary distribution has an interesting property of locality: It depends only on the total weight and the weight of edges connected to the node and therefore it does not change if the weights on some other parts of the graph are changed while keeping the total weight constant.

The entropy rate can be computed as follows:

$$H(\chi) = H(X_1 \mid X_2) = -\sum_{ij} \mu_i \sum_j P_{ij} \log P_{ij}$$

ER of Random Walk

$$\begin{aligned}
&= -\sum_i \frac{W_i}{2W}\sum_j \frac{W_{ij}}{W_i}\log\frac{W_{ij}}{W_i} \\
&= -\sum_i\sum_j \frac{W_{ij}}{2W}\log\frac{W_{ij}}{W_i} \\
&= -\sum_i\sum_j \frac{W_{ij}}{2W}\log\frac{W_{ij}}{2W} + \sum_i\sum_j \frac{W_{ij}}{2W}\log\frac{W_i}{2W} \\
&= H\left(\ldots,\frac{W_{ij}}{2W},\ldots\right) - H\left(\ldots,\frac{W_i}{2W},\ldots\right)
\end{aligned}$$

If all the edges have equal weight, , the stationary distribution puts weight $E_i/2E$ on node i, where E_i is the number of edges emanating from node i and E is the total number of edges in the graph. In this case the entropy rate of the random walk is:

$$H(\chi) = \log(2E) - H\left(\frac{E_1}{2E},\frac{E_2}{2E},\ldots,\frac{E_m}{2E}\right)$$

Apparently the entropy rate, which is the average transition entropy, depends only on the entropy of the stationary distribution and the total number of edges

Example

Random walk on a chessboard. Let's king move at random on a 8x8 chessboard. The king has eight moves in the interior, five moves at the edges and three moves at the corners. Using this and the preceding results, the stationary probabilities are, respectively, 8/420, 5/420 and 3/420, and the entropy rate is 0.92log8. The factor of 0.92 is due to edge effects; we would have an entropy rate of log8 on an infinite chessboard. Find the entropy of the other pieces for exercize!

It is easy to see that a stationary random walk on a graph is time reversible; that is, the probability of any sequence of states is the same forward or backward:

$$\Pr\{X_1 = x_1, X_2 = x_2, \ldots X_n = x_n\} = \Pr\{X_n = x_1, X_{n-1} = x_2, \ldots X_1 = x_n\}$$

The converse is also true, that is any time reversible Markov chain can be represented as a random walk on an undirected weighted graph.