How to Quantitate a Markov Chain?
Stochostic project 1

CHIN-NING,CHOU WEI-CHANG,LEE
PROFESSOR RAOUL NORMAND
April 13, 2015

Abstract

In this project,we will study Markov chain in the light of Information Theory ,and then use it to discuss the entropy in the Ehrenfest model. We will use Matlab to simulate the Ehrenfest model to show that the relative entropy is decreasing.

Keywords: Entropy , Markov chain , Kullback Leibler divergence , Ehrenfest model , Thermodynamics

Introduction

In classical thermodynamics, people like to use entropy to evaluate the structural tendency in a system. While the well known Second Law of Thermodynamics claims that in a natural thermodynamic process, the entropy in the participated system will increase, we want to find out whether there also exists some properties in a Markov chain that quantitatively reveal the structural tendency in the system.

In the beginning, we took the definition of entropy that Claude Shannon used in Information theory to construct the entropy in a Markov chain. However, after deducing some corresponding results , we found that such definition does not elegantly exposure the tendency in the system. As a result, we surveyed and tried other models to describe the structure of a Markov chain, rethinking how to interpret Shannon's entropy in a Markov chain.

In this report, we will first apply Shannon's entropy to examine the corresponding results in a Markov chain. Next, some other structural properties will be introduced. In the third part of the report, we will use a simple urn problem which is known as Ehrenst model to demonstrate the properties we used in this project. And we will give some comments on each structural properties and discuss the difference of Shannon's entropy and the entropy in thermodynamic at last.

Entropy ,Relative Entropy and Mutual Information

The entropy of a random variable is a measure of the uncertainty of this random variable.It is a measure of the amount of information required on the average to describe the random variable.
The entropy H(x) of a discrete random variable X is defined by:

$$
\begin{aligned}
H(X) &= -\sum_{x\in X} p(x)\log p(x) \\
&= E_p[\frac{1}{\log p(x)}]
\end{aligned}
$$

And the conditional entropy is given by

$$
\begin{aligned}
H(Y|X) &= \sum_{x\in X} p(x)H(Y|X=x) \\
&= -\sum_{x\in X} p(x)\sum_{y\in Y} p(y|x)\log p(y|x) \\
&= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(y|x)
\end{aligned}
$$

For a pair of discrete random variable X and Y:

$$
\begin{aligned}
H(X,Y) &= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x,y) \\
&= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x)p(y|x) \\
&= H(X) + H(Y|X)
\end{aligned}
$$

We can use the definition above to derive relative entropy,which is a measure of distance between two distributions.The relative entropy or Kullback-lieblar divergence between two pmf p(x) and q(x) is defined as:

$$
\begin{aligned}
D(p||q) &= -\sum_{x\in X} p(x)\log\frac{p(x)}{q(x)} \\
&= E_p[\log\frac{1}{q(x)}] - E_p[\log\frac{1}{p(x)}]
\end{aligned}
$$

The Kullback-Leibler divergence measures the expected number of extra bits required to code samples from P when using Q. It is non-negative but not symmetric :

$$D(p||q) \neq D(q||p) \geq 0$$

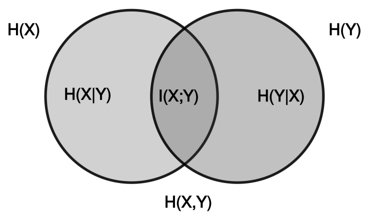
The mutual information I(X;Y) is the relative entropy between the joint distribution and the product distribution p(x).p(y):

$$\begin{aligned} I(X;Y) &= = -\sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\ &= -\sum_{x,y} p(x,y) \log \frac{p(x|y)}{p(x)} \\ &= H(X) - H(X|Y) \end{aligned}$$

The mutual information I(X;Y) is the reduction in the uncertainty of X due to the knowledge of Y.It is easily to derive following property:

$$\begin{aligned} I(X;Y) &= I(Y;X) \\ I(X;X) &= H(X;X) \end{aligned}$$

We can use a Venn diagram to summarize the relationship between the entropy concept we discuss above:



Data processing Inequlity

Under simple Markov assumption,We can show that no clever manipulation of the data can improve the inferences that can be made from the data. $If\ X->Y->Z\ form\ a\ Markov\ chain,\ then\ I(X;Y) \geq I(X;Z)$

$$\begin{aligned} I(X;Y,Z) &= I(X;Z) + I(X;Y|Z) \quad (1) \\ &= I(X;Y) + I(X;Z|Y) \quad (2) \end{aligned}$$

First focus on $I(X;Z|Y)$, we notice that:

$$\begin{aligned}
I(X;Z|Y) &= H(X;Y) + H(X|Z,Y) \\
&= E_P(x,y,z)\log\frac{p(x,z|y)}{p(x|y).p(z|y)} \\
&= E_P(x,y,z)\log\frac{\frac{p(x,y,z)}{p(y)}}{p(x|y).p(z|y)} \\
&= E_P(x,y,z)\log\frac{\frac{p(x,y)p(z|y)}{p(y)}}{p(x|y).p(z|y)} \\
&= E_p(x,y,x)\log 1 &= 0
\end{aligned}$$

From (1) and (2) we can easily get $I(X;Y) \geq I(X;Z)$, Z can be written as a function of Y, thus we cannot increase the information about X by manipulating Y. Note that $I(X;Y|Z) \leq I(X;Y)$ can be derived from (1) and (2),which means the dependence of X and Y is decreased by observing Z.It is because transition matrix is invertible ,so Z->Y->X also forms a Markov chain.Z reveals some information about X and Y.

Entropy rate

Entropy and the Second Law of Thermodynamics

We can model the evolution of a isolated system as a Markov chain with transition matrix obeying the physical laws governing the system.And now we know some basic properties of entropy , what if we try to figure out the statement of the Second Law of Thermodynamics that the entropy of an isolated system is always nondecreasing?
Let μ_n and $\mu_n^{'}$ be two probability distribution at time n on the same space of Markov chain but different initial distribution and μ_{n+1} and $\mu_{n+1}^{'}$be the one at time n+1.The corresponding joint mass function of time n and n+1 can be denoted as p and q.We have:

$$\begin{aligned}
p(x_n, x_{n+1}) &= p(x_n)r(x_{n+1}|x_n) \\
q(x_n, x_{n+1}) &= q(x_n)r(x_{n+1}|x_n)
\end{aligned}$$

where $r(.|.)$ is the transition probability of this Markov chain.By the chain rule for relative entropy ,we can rewrite the relative entropy of p and q as :

$$D(p(x_n,x_{n+1})||q(x_n,x_{n+1})) = D(p(x_n)||q(x_n)) + D(p(x_{n+1}|x_n)||q(x_{n+1}|x_n)) \tag{3}$$

$$= D(p(x_{n+1})||q(x_{n+1})) + D(p(x_n|x_{n+1})||q(x_n|x_{n+1})) \tag{4}$$

Also,

$$D(p(x_{n+1}|x_n)||q(x_{n+1}|x_n)) = D(r(x_{n+1}|x_n)||r(x_{n+1}|x_n)) = 0 \tag{5}$$
$$D(p(x_n|x_{n+1})||q(x_n|x_{n+1}) \geq 0 \tag{6}$$

Substitute (3)(4) with (5)(6) ,We have:

$$D(p(x_n)||q(x_n)) = D(\mu_n||\mu_n^{'}) \geq D(p(x_{n+1})||q(x_{n+1}) = D(\mu_{n+1}||\mu_{n+1}^{'})$$

Distance casued by different initial distribution is decreasing with time n. If we let $\mu_n^{'}$ be any stationary distribution π, $\mu_n^{'}$ will be the same as $\mu_{n+1}^{'}$.

$$D(\mu_n||\pi) \geq D(\mu_{n+1}||\pi)$$

Which implies that any state distribution gets closer and closer to each stationary distribution with time n. $D(\mu_n||\pi)$ is a monotonically non-increasing non-negative sequence and thus have a limit.
The fact that the relative entropy always decreases does not imply that the entropy increases. Markov chain with a non-uniform stationary distribution starts from uniform distribution will decrease since we get maximum entropy when the distribution is uniform.
If the stationary distribution is uniform,we have:

$$D(\mu_n||\pi) = \log|\chi| - H(\mu_n) = \log|\chi| - H(X_n)$$

The monotonic decrease in relative entropy implies a monotonic increase in entropy.Which resonates with the definition of entropy in statistical thermodynamics, the log of number of microstates, where all the states are equally likely in the equilibrium, meaning a uniform distribution.

Simulation on Ehrenfest model

briefly take about Ehrenfest mldel and the simulation

Conclusion