Information Theory and Communication

Stochastic Processes

Ritwik Banerjee

rbanerjee@cs.stonybrook.edu
Entropy Rate

Definition

A **stochastic process** is a sequence of random variable indexed by time. That is, it can be expressed as an indexed sequence of random variables \( \{X_i\} \).

- It is characterized by the joint distribution
  \[
  Pr(X_1, X_2, \ldots, X_n) = (x_1, x_2, \ldots x_n)
  \]
  where \((x_1, x_2, \ldots x_n) \in X^n\) and \(n\) is a positive integer.

- From AEP, we know that on the average, \(nH(X)\) bits are sufficient to describe \(n\) i.i.d. random variables.
- Now we will show that even if the random variables are dependent, under specific situations the entropy \(H(X_1, X_2, \ldots, X_n)\) grows linearly with \(n\).
  - The rate of this linear growth is the entropy \(H(X)\).
  - This is called the **entropy rate** of the process.
  - The entropy rate is also the best achievable data compression.
Stationary Process

Definition

A stochastic process is a **stationary process** if

\[ Pr\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\} = Pr\{X_{1+k} = x_1, X_{2+k} = x_2, \ldots, X_{n+k} = x_n\} \]

for every \( n \) and every “shift” \( k \).

In other words, the joint distribution of any subset of the sequence of random variables is *invariant* w.r.t. shifts in the time index.
Markov Chain

Definition

A discrete stochastic process is called a Markov chain or a Markov process if for \( n = 1, 2, \ldots \),

\[
Pr(x_{n+1}|x_n, x_{n-1}, \ldots x_1) = Pr(x_{n+1}|x_n)
\]

That is, each random variable depends only on the one preceding it and is conditionally independent of all the other preceding random variables.

- For a Markov process, the joint distribution can be expressed as
  \[
p(x_1, x_2, \ldots x_n) = p(x_n|x_{n-1}) \ldots p(x_3|x_2)p(x_2|x_1)p(x_1)
  \]

- A Markov chain is said to be time invariant if \( p(x_{n+1}|x_n) \) is independent of \( n \). That is, \( p(x_2|x_1) = p(x_3|x_2) = \ldots \)

- \( X_n \) is called the state of the process at time \( n \).
A time-invariant Markov process is characterized by two things:

1. the initial state $X_1$, and
2. a probability transition matrix $P$ whose $(i, j)^{th}$ entry is given by $P_{ij} = Pr\{X_{n+1} = j | X_n = i\}$
Types of Markov chains

- If it is possible to go with a positive probability from any state of the Markov chain to any other state in a finite number of steps, then the Markov chain is said to be **irreducible**.

- If the largest common factor of the lengths of different paths from a state to itself is 1, then the Markov chain is said to be **aperiodic**.

- A distribution on the states such that the distribution at time $n + 1$ is the same as the distribution at time $n$ is called a **stationary distribution**.
  - If the initial state of a Markov chain is drawn according to a stationary distribution, then the Markov chain forms a stationary process.

- If a finite-state Markov chain is both irreducible and aperiodic, then the stationary distribution is unique.
  - And no matter what the starting distribution is, the distribution of $X_n$ tends to the stationary distribution as $n \to \infty$. 
Computing Stationary Distributions

The probability transition matrix is given by

\[ P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix} \]

The entropy of the state \( X_n \) at time \( n \) is

\[ H(X_n) = H\left( \frac{\beta}{\alpha + \beta}, \frac{\alpha}{\alpha + \beta} \right) \]
To calculate the rate at which the entropy grows, we will have to take into account the dependency among the $X_i$’s.

But first, we define the **entropy of a stochastic process** \( \{X_i\} \):

\[
H(\mathcal{X}) = \lim_{n \to \infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n)
\]

But this limit may not always exist!

**Example: Typewriter**

Consider the case of a typewriter that has \( m \) equally likely output letters. The typewriter can produce \( m^n \) sequences of length \( n \), all of them equally likely.

The entropy is given by \( H(X_1, \ldots, X_n) = \log m^n \).

The entropy rate (i.e., the entropy of the process) is then given by \( H(\mathcal{X}) = \log m \) bits.
The entropy rate can also be defined as

\[ H'(\mathcal{X}) = \lim_{n \to \infty} H(X_n|X_{n-1}, X_{n-2}, \ldots, X_1) \]

**Theorem**

*For a stationary stochastic process, the two limits defined for \( H(\mathcal{X}) \) and \( H'(\mathcal{X}) \) both exist, and are equal. That is,*

\[ H(\mathcal{X}) = H'(\mathcal{X}) \]