Recap

- **The AEP:** If \( X_1, X_2, \ldots \) are i.i.d. \( \sim p(x) \), then
  \[
  \frac{1}{n} \log \frac{1}{p(x_1, x_2, \ldots, x_n)} \to H(X) \quad \text{or} \quad p(x_1, x_2, \ldots, x_n) \to 2^{-nH(X)}
  \]

- **The typical set:**
  \[
  \{x^n\} \text{ the set of sequences } (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n
  \]
  that satisfy
  \[
  2^{-n(H(X) - \varepsilon)} \leq p(x_1, x_2, \ldots, x_n) \leq 2^{-n(H(X) + \varepsilon)}
  \]

- Sequences \( X^n \) can be represented using \( nH(X) \) on the average

---

The Need for Entropy Rate

- **Independent, identically distributed (i.i.d.) random variables**
  - The AEP
  - \( nH(X) \) bits suffice on the average to describe \( n \) i.i.d. random variables

- **Dependent random variables?**
  - Entropy rate \( H(\mathcal{X}) \)
  - The entropy \( H(X_1, \ldots, X_n) \) grows asymptotically linearly with \( n \) at the entropy rate
Stochastic Process

- A **stochastic process** is an indexed sequence of random variables with an arbitrary dependence among the random variables.
- The joint probability mass functions is used to characterize the process.

\[ \Pr \{ \{ X_1, X_2, \ldots, X_n \} = (x_1, x_2, \ldots, x_n) \} = p(x_1, x_2, \ldots, x_n), \]
where \( (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \)

Markov Process

- A discrete stochastic process \( X_1, X_2, \ldots \) is said to be a **Markov chain** or a **Markov process** if,

\[
Pr(X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_1 = x_1) = Pr(X_{n+1} = x_{n+1} | X_n = x_n)
\]
- That is, each random variable depends on the one preceding it and is conditionally independent of all the other preceding random variables. And the joint probability mass function is:

\[ p(x_1, x_2, \ldots, x_n) = p(x_1) p(x_2 | x_1) p(x_3 | x_1) \cdots p(x_n | x_{n-1}) \]

Homogeneous Markov Chain

- **homogeneous** = **stationary** = **time-invariant**
- A Markov chain is said to be time-invariant if the conditional probability \( p(X_{n+1} | X_n) \) does not depend on \( n \), i.e.

\[ Pr(X_{n+1} = b | X_n = a) = Pr(X_1 = b | X_0 = a) \]
- Time-invariant also means that the transition probabilities do not change over time.
Irreducible Markov Chain

- If it is possible to go with 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.

Probability Transition Matrix P

- Each element $P_{ij}$ expresses the probability of movement (transition) from one state ($i$) to the next ($j$). That is, $P_{ij} = \Pr\{X_{n+1} = j \mid X_n = i\}$
- A state is the condition or location of an object in the system at a particular time.

Example

- University town → pizza company → lots of phone calls → three kitchens → same set of delivery drivers
- Question:
  - What’s the probability transition matrix?
  - If you start at C, what’s the probability that you’ll be in area B after 2 deliveries? 3 deliveries? 100 deliveries???

From http://ceee.rice.edu/Books/LA/index.html
Example (Cont’)

Properties of the Transition Matrix
- It must be square
- The elements of each row sum to 1
- Each element is between 0 and 1 inclusively
- If all the entries of the transition matrix are between 0 and 1 exclusively, then convergence is guaranteed to take place. Convergence may take place when 0 and 1 are in the transition matrix, but convergence is no longer guaranteed

Invariant
- Many of the long-time properties of Markov chains are connected with the notion of an invariant distribution or measure
- A measure \( \lambda \) is a row vector with non-negative elements
- We say \( \lambda \) is invariant if \( \lambda P = \lambda \)
Two-State Markov Chain Analysis

If the Markov chain has an initial state drawn according to the stationary distribution (or invariant distribution), the resulting process will be stationary. The entropy of the state $X_n$ at time $n$ is $H(X_n) = ???$

Entropy Rate

- The entropy rate of a stochastic process $(X_i)$ is defined by $H(R) = \lim_{n \to \infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n)$
- It describes how the entropy of the sequence grows with $n$, indicating the per symbol entropy of the $n$ random variables.
- A related quantity for entropy rate describes the conditional entropy of the last random variable given the past $H (R | X_1, X_2, \ldots, X_{n-1}) = \lim_{n \to \infty} H (X_n | X_1, X_2, \ldots, X_{n-1})$
- For a stationary stochastic process, $H (R) = H (2R)$

Examples

- Typewriter
- i.i.d. random variables
- Sequence of independent, but not identically distributed random variables
Entropy Rate for a Stationary Markov Chain

\[ H(\mathcal{X}) = H'(\mathcal{X}) = \lim H(X_t | X_{t-1}, \ldots, X_1) \]
\[ = \lim H(X_t | X_{t-1}) = \lim H(X_t | X_{X_{t-1}}) \]
\[ = - \sum_{i,j} \mu_{ij} \log \mu_{ij} \]

Example – two-state Markov chain

Example: Entropy Rate of a Random Walk

- Discussion p66
Entropy Rate of Functions of a Markov Chain

If \( X_1, X_2, \ldots, X_n \) form a Markov chain and \( Y'_i = \phi(X'_i) \),
then \( H(Y'_1 | Y'_2, \ldots, Y'_n) \leq H(\Sigma) \leq H(Y'_1 | Y'_2, \ldots, Y'_n) \).
and \( \lim_{n \to \infty} H(Y'_1 | Y'_2, \ldots, Y'_n) = H(\Sigma) = \lim_{n \to \infty} H(Y'_1 | Y'_2, \ldots, Y'_n) \).