Stationary Information Sources

Aakash Ghosh

May 22, 2023

Timeline

We are going to look at:

1. Defining the space
2. Defining Ergodicity

Setting up our space

- For an alphabet A we define the space Ω such that if $x \in \Omega$ then

$$x = \{\ldots x_{-2}, x_{-1}, x_0, x_1, x_2, hdots\}$$

Setting up our space

- For an alphabet A we define the space Ω such that if $x \in \Omega$ then

$$x = \{ \ldots x_{-2}, x_{-1}, x_0, x_1, x_2, hdots \}$$

- We define functions X_i such that $X_i(x) = x_i \forall i \in \mathbb{Z}$

Setting up our space

- For an alphabet A we define the space Ω such that if $x \in \Omega$ then

$$x = \{ ... x_{-2}, x_{-1}, x_0, x_1, x_2, hdots \}$$

- We define functions X_i such that $X_i(x) = x_i \forall i \in \mathbb{Z}$
- We take the sample space $\mathcal{F}$ to be the smallest set such that all X_i are measurable.
- We define $[x_{i_1} x_{i_2} ... x_{i_n}] = \{ y | y_{i_j} = x_{i_j} \forall 1 \leq j \leq n \}$
- we define a probability measure μ on such as space.

Shift transformation

- We want x to be some kind of information stream from the same source.

Shift transformation

- We want x to be some kind of information stream from the same source.
- μ shouldn't depend on the point from where we are parsing our stream

Shift transformation

- We want x to be some kind of information stream from the same source.
- μ shouldn't depend on the point from where we are parsing our streami.e. μ is shift invariant.

Shift transformation

- We want x to be some kind of information stream from the same source.
- μ shouldn't depend on the point from where we are parsing our streami.e. μ is shift invariant.
- We define T as the shift transformation. $Tx = y$ if $x_i = y_{i-1} \forall \in \mathbb{Z}$

Shift transformation

- We want x to be some kind of information stream from the same source.
- μ shouldn't depend on the point from where we are parsing our streami.e. μ is shift invariant.
- We define T as the shift transformation. $Tx = y$ if $x_i = y_{i-1} \forall \in \mathbb{Z}$
- We assume $\mu T^{-1} = \mu$(i.e. E and $T(E)$ have the same measure)

Shift transformation

- We want x to be some kind of information stream from the same source.
- μ shouldn't depend on the point from where we are parsing our streami.e. μ is shift invariant.
- We define T as the shift transformation. $Tx = y$ if $x_i = y_{i-1} \forall \in \mathbb{Z}$
- We assume $\mu T^{-1} = \mu$(i.e. E and $T(E)$ have the same measure)

Relation between EIS and SIS

- From the shift invariance, it follows that random vectors $\{X_0, X_1 \dots X_n\}$ and $\{X_i, X_{i+1}, X_{i+2} \dots X_{i+n}\}$ have the same distribution.
- We can therefore define an EIS on A^n defined by

$$\mu_n(\{x_1, x_2 \dots x_n\}) = \mu([x_1, x_2, x_3 \dots x_n])$$

Rate of information

- We therefore have an average rate of information gain in the first n steps as

$$-\frac{1}{n}\log \mu([x_1, x_2, x_3 \dots x_n])$$

- Expectation of our random vector is given by

$$\frac{1}{n}H([X_1X_2...X_n])$$

Topological Nature of SIS

- We use discrete topology on individual bits.
- We define:
$$d(x, y) = \sum 2^{-|i|} d(x_i, y_i)$$
- If μ, ν are measures, then so is $t\mu + (1 - t)\nu, 0 \leq t \leq 1$ i.e. the space of probability measures are convex and compact.

Ergodicity

- We say f is T invariant if $f = f \circ T$ a.e.
- The indicator function 1_E is invariant iff $\mu(E \Delta T(E)) = 0$
- Define:

$$\mathcal{I}_{T,\mu} = \{E | \mu(E \Delta T(E)) = 0\}$$

  All such sets are called invariant sets.
- If for $E \in \mathcal{I}$, $\mu E = 0$ or 1, then T is μ ergodic.
- A random variable is T invariant iff it is I measurable.

Time average and space average

- We deine:
$$A_n f(\omega) = \frac{1}{n}\left(f(\omega) + f(T^1\omega) + f(T^2\omega) \dots + f(T^{n-1}\omega)\right)$$
$A_n f(\omega)$ is the time average of ω under T
- Similarly, $E[f] = \int_\omega f(\omega)\mu(d\omega)$ is the space average of f.

Convergence of time average and space average

For an L^1 random variable f, $\lim_{n\to\infty} A_n f(\omega)$ exists and $\lim_{n\to\infty} A_n f(\omega) = E[f]$ almost surely and in L^1.

Convergence of time average and space average

For an L^1 random variable f, $\lim_{n\to\infty} A_n f(\omega)$ exists and $\lim_{n\to\infty} A_n f(\omega) = E[f]$ almost surely and in L^1.
Further, if μ is ergodic then $E[f|\mathcal{I}](\omega) = E[f]$ almost surely, and if f is L^p then convergence is in L^p as well.

References