Lecture #6A: The Asymptotic Equipartition Property
Outline of the lecture

- Law of large numbers
- Typical sequences
Outline of the lecture

- Law of large numbers
- Typical sequences
- Asymptotic Equipartition Property

Outline

1 Law of large numbers
2 Typical sequences
3 Asymptotic Equipartition Property
Weak Law of Large Numbers: Let $X[n]$ be an independent random sequence with mean $\mu_X$ and variance $\sigma^2_X$ defined for $n \geq 1$. Define

$$\hat{\mu}_X[n] \overset{\Delta}{=} (1/n) \sum_{k=1}^{n} X[k] \ \forall \ n \geq 1$$

Then $\hat{\mu}_X[n] \to \mu_X$ in probability as $n \to \infty$. 
Tychebycheff Inequality

- If $X$ is a real-valued random variable with mean $\mu_X$ and variance $\sigma_X^2$.

Let $A$ denote the event $|X - \mu_X| \geq \epsilon$, and $A^c$ the complementary event ($|X - \mu_X| < \epsilon$), then

$$\sigma_X^2 = E[(X - \mu_X)^2|A]P(A) + E[(X - \mu_X)^2|A^c]P(A^c) \geq E[(X - \mu_X)^2|A]P(A)$$
If $X$ is a real-valued random variable with mean $\mu_X$ and variance $\sigma_X^2$.

Let $A$ denote the event $|X - \mu_X| \geq \epsilon$, and $A^c$ the complementary event ($|X - \mu_X| < \epsilon$), then

$$
\sigma_X^2 = E[(X - \mu_X)^2|A]P(A) + E[(X - \mu_X)^2|A^c]P(A^c) \geq E[(X - \mu_X)^2|A]P(A)
$$

Whenever $A$ occurs, $(X - \mu_X)^2 \geq \epsilon^2$, so that

$$
E[(X - \mu_X)^2|A] \geq \epsilon^2
$$

Hence

$$
\sigma_X^2 \geq \epsilon^2 P(A)
$$
Tychebycheff Inequality

- If $X$ is a real-valued random variable with mean $\mu_X$ and variance $\sigma_X^2$.
- Let $A$ denote the event $|X - \mu_X| \geq \epsilon$, and $A^c$ the complementary event ($|X - \mu_X| < \epsilon$), then
  \[
  \sigma_X^2 = E[(X - \mu_X)^2|A]P(A) + E[(X - \mu_X)^2|A^c]P(A^c) \\
  \geq E[(X - \mu_X)^2|A]P(A)
  \]
- Whenever $A$ occurs, $(X - \mu_X)^2 \geq \epsilon^2$, so that
  \[
  E[(X - \mu_X)^2|A] \geq \epsilon^2
  \]
- Hence
  \[
  \sigma_X^2 \geq \epsilon^2 P(A)
  \]
- Alternatively,
  \[
  P(|X - \mu_X| \geq \epsilon) \leq \frac{\sigma_X^2}{\epsilon^2}, \quad \forall \epsilon > 0
  \]

Weak law of Large Numbers

- Sample mean
  \[
  E[\hat{\mu}_X[n]] = \frac{1}{n} \sum_{i=1}^{n} E[X[i]] = \mu_X
  \]
Weak law of Large Numbers

- Sample mean
  \[ E[\hat{\mu}_X[n]] = \frac{1}{n} \sum_{i=1}^{n} E[X[i]] = \mu_X \]

- Sample variance
  \[ Var[\hat{\mu}_X[n]] = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2_X = \frac{\sigma^2_X}{n} \]

Using Tychebycheff Inequality, we get
\[ P(|\hat{\mu}_X[n] - \mu_X| \geq \epsilon) \leq \frac{\sigma^2_X}{n\epsilon^2} \]
Weak law of Large Numbers

- Sample mean
  
  \[ E[\hat{\mu}_X[n]] = (1/n) \sum_{i=1}^{n} E[X[i]] = \mu_X \]

- Sample variance
  
  \[ \text{Var}[\hat{\mu}_X[n]] = (1/n^2) \sum_{i=1}^{n} \sigma_X^2 = \frac{\sigma_X^2}{n} \]

- Using Tychebycheff Inequality, we get
  
  \[ P(|\hat{\mu}_X[n] - \mu_X| \geq \epsilon) \leq \frac{\sigma_X^2}{n\epsilon^2} \]

- Alternatively,
  
  \[ P(|\hat{\mu}_X[n] - \mu_X| \leq \epsilon) \geq 1 - \frac{\sigma_X^2}{n\epsilon^2} \]

Thus

\[ \lim_{n \to \infty} P(|\hat{\mu}_X[n] - \mu_X| \leq \epsilon) = 1 \]
Let $Y$ be the indicator random variable for the event $A$, i.e. $Y = 1$ when $A$ occurs, and 0 otherwise.

$$E[Y] = P(A)$$

Since $Y^2 = Y$, this implies that

$$E[Y^2] = P(A)$$
Weak law of Large Numbers

Let $Y$ be the indicator random variable for the event $A$, i.e. $Y = 1$ when $A$ occurs, and 0 otherwise.

$$E[Y] = P(A)$$

Since $Y^2 = Y$, this implies that

$$E[Y^2] = P(A)$$

Hence

$$\text{Var}[Y] = P(A)[1 - P(A)]$$

Then

$$\hat{\mu}_Y[n] = \frac{n_A}{n}$$

where $n_A$ is the number of times, the event $A$ has occurred.
Weak law of Large Numbers

Then
\[ \hat{\mu}_Y[n] = \frac{n_A}{n} \]

where \( n_A \) is the number of times, the event A has occurred.

From Tychebycheff Inequality, we get
\[
P\left( \left| \frac{nA}{n} - P(A) \right| \geq \epsilon \right) \leq \frac{P(A)[1 - P(A)]}{n\epsilon^2}
\]

Outline

1. Law of large numbers
2. Typical sequences
3. Asymptotic Equipartition Property
Typical sequences

Consider a sequence of \( L = 20 \) bits emitted by a discrete memoryless source (DMS) with

\[
P_U(0) = \frac{3}{4} \quad \text{and} \quad P_U(1) = \frac{1}{4}
\]

Which one of the following is the “real” sequence?

1. 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
2. 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1
3. 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0

Adrish Banerjee
Department of Electrical Engineering
Indian Institute of Technology Kanpur
Kanpur, Uttar Pradesh India

An introduction to Information Theory
Typical sequences

Consider a sequence of $L = 20$ bits emitted by a discrete memoryless source (DMS) with

$$P_U(0) = \frac{3}{4} \text{ and } P_U(1) = \frac{1}{4}$$

Which one of the following is the “real” sequence?

(1) $1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1$

(2) $1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1$

(3) $0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0$

Probability of occurrence of the sequences

(1) $P_{U_1,\cdots,U_{20}}(u_1,\cdots,u_{20}) = \left(\frac{1}{4}\right)^{20}$

(2) $P_{U_1,\cdots,U_{20}}(u_1,\cdots,u_{20}) = \left(\frac{1}{4}\right)^{20}(3)^{14}$

(3) $P_{U_1,\cdots,U_{20}}(u_1,\cdots,u_{20}) = \left(\frac{1}{4}\right)^{20}(3)^{20}$

Let $U$ denote an output sequence of length $L$ emitted a $K$-ary DMS and $P_U(u)$ is the output probability distribution.
Typical sequences

Let $\mathbf{U}$ denote an output sequence of length $L$ emitted a K-ary DMS and $P_U(u)$ is the output probability distribution.

Let $\mathbf{u} = [u_1, u_2, \cdots, u_L]$ denote possible values of $\mathbf{U}$, i.e. $u_j = \{a_1, a_2, \cdots, a_K\}$ for $1 \leq j \leq L$.

Let $n_{a_i}(u)$ denotes the number of occurrence of the letter $a_i$ in the sequence $\mathbf{u}$. Then $\mathbf{u}$ is an $\epsilon$–typical output sequence of length $L$ for this K-ary DMS if

$$(1 - \epsilon)P_U(a_i) \leq \frac{n_{a_i}(u)}{L} \leq (1 + \epsilon)P_U(a_i), \quad 1 \leq i \leq K.$$
Typical sequences

Consider a binary DMS with $P_U(0) = \frac{3}{4}$ and $P_U(1) = \frac{1}{4}$. Let’s choose $\epsilon = \frac{1}{3}$. Then a sequence $u$ of length $L = 20$ is $\epsilon$-typical if and only if both
\[
\frac{2}{3} \cdot \frac{3}{4} \leq \frac{n_0(u)}{20} \leq \frac{4}{3} \cdot \frac{3}{4}
\]
and
\[
\frac{2}{3} \cdot \frac{1}{4} \leq \frac{n_1(u)}{20} \leq \frac{4}{3} \cdot \frac{1}{4}
\]

Equivalently, $u$ is $\epsilon$-typical if and only if both
\[
10 \leq n_0(u) \leq 20
\]
and
\[
4 \leq n_1(u) \leq 6
\]
Typical sequences

Property 1: If $u$ is an $\epsilon$-typical output sequence of length $L$ from a $K$-ary DMS with entropy $H(U)$ in bits, then

$$2^{-(1+\epsilon)LH(U)} \leq P_u(u) \leq 2^{-(1-\epsilon)LH(U)}$$

Proof: From the definition of a DMS, we have

$$P_u(u) = \prod_{j=1}^{L} P_U(u_j) = \prod_{i=1}^{K} [P_U(a_i)]^{n_{z_i}(u)}$$
Typical sequences

- **Property 1:** If \( u \) is an \( \epsilon \)-typical output sequence of length \( L \) from a \( K \)-ary DMS with entropy \( H(U) \) in bits, then

\[
2^{-(1+\epsilon)LH(U)} \leq P_U(u) \leq 2^{-(1-\epsilon)LH(U)}
\]

- **Proof:** From the definition of a DMS, we have

\[
P_U(u) = \prod_{j=1}^{L} P_U(u_j) = \prod_{i=1}^{K} (P_U(a_i))^{n_{ai}(u)}
\]

From the definition of typical sequences, we have

\[
(1 - \epsilon)P_U(a_i) \leq \frac{n_{ai}(u)}{L} \leq (1 + \epsilon)P_U(a_i), \quad 1 \leq i \leq K
\]

Using the right inequality, we get

\[
P_U(u) \geq \prod_{i=1}^{K} [P_U(a_i)]^{(1+\epsilon)LP_U(a_i)}
\]
Typical sequences

- Using the right inequality, we get
  \[ P_U(u) \geq \prod_{i=1}^{K} [P_U(a_i)]^{(1+\epsilon)L_P(a_i)} \]

- Equivalently,
  \[ P_U(u) \geq \prod_{i=1}^{K} 2^{(1+\epsilon)L_P(a_i)} \log_2 P_U(a_i) \]

Simplifying we get,
\[ P_U(u) \geq 2^{(1+\epsilon)L \sum_{i=1}^{K} P_U(a_i) \log_2 P_U(a_i)} \]
Typical sequences

- Using the right inequality, we get
  \[ P_U(u) \geq \prod_{i=1}^{K} [P_U(a_i)]^{(1+\epsilon)L_P U(a_i)} \]

- Equivalently,
  \[ P_U(u) \geq \prod_{i=1}^{K} 2^{(1+\epsilon)L_P U(a_i) \log_2 P_U(a_i)} \]

- Simplifying we get,
  \[ P_U(u) \geq 2^{(1+\epsilon)L \sum_{i=1}^{K} P_U(a_i) \log_2 P_U(a_i)} \]

- Hence
  \[ P_U(u) \geq 2^{-(1+\epsilon)LH(U)} \]

Similar arguments can be used to prove
\[ P_U(u) \leq 2^{-(1-\epsilon)LH(U)} \]
Typical sequences

- **Property 2:** The probability, $1 - P(F)$, that the length $L$ output sequence $U$ from a $K$-ary DMS is $\epsilon$-typical satisfies

$$1 - P(F) > 1 - \frac{K}{Le^2P_{\min}}$$

where $P_{\min}$ is the smallest positive value of $P_U(u)$.

Interested to show that for large $L$, the output sequence $U$ of the DMS is certain to be $\epsilon$–typical.
Typical sequences

- **Property 2**: The probability, $1-P(F)$, that the length $L$ output sequence $U$ from a $K$-ary DMS is $\epsilon$-typical satisfies

$$1 - P(F) > 1 - \frac{K}{L\epsilon^2 P_{\text{min}}}$$

where $P_{\text{min}}$ is the smallest positive value of $P_U(u)$.

- Interested to show that for large $L$, the output sequence $U$ of the DMS is certain to be $\epsilon$–typical.

- We will use Tchebycheff inequality

$$P \left( \left| \frac{nA}{n} - P(A) \right| \geq \epsilon \right) \leq \frac{P(A)[1 - P(A)]}{n\epsilon^2}$$

Let $B_i$ denote the event that $U$ takes on value $u$ such that the condition for $\epsilon$-typical sequence is not satisfied. Then,

$$P(B_i) = P \left( \left| \frac{n_{A_i}(u)}{L} - P_U(a_i) \right| > \epsilon P_U(a_i) \right)$$

$$\leq \frac{P_U(a_i)[1 - P_U(a_i)]}{L[\epsilon P_U(a_i)]^2}$$
Typical sequences

Let $B_i$ denote the event that $U$ takes on value $u$ such that the condition for $\epsilon$-typical sequence is not satisfied. Then,

$$P(B_i) = P\left(\left| \frac{n_i(u)}{L} - P_U(a_i) \right| > \epsilon P_U(a_i) \right) \leq \frac{P_U(a_i)[1 - P_U(a_i)]}{L[\epsilon P_U(a_i)]^2}$$

Simplifying, we have

$$P(B_i) \leq \frac{1 - P_U(a_i)}{Le^2 P_U(a_i)}$$

Let $P_{\min}$ is the minimum non-zero value of $P_U(u)$, we get,

$$P(B_i) < \frac{1}{Le^2 P_{\min}}$$
Typical sequences

- Let $P_{\text{min}}$ is the minimum non-zero value of $P_U(u)$, we get,

$$P(B_i) < \frac{1}{L\epsilon^2 P_{\text{min}}}$$

- Let $F$ be the failure event that $U$ is not $\epsilon$-typical. Since $F$ occurs in at least one of the events, $B_i, 1 \leq i \leq K$, using union bounds we get

$$P(F) \leq \sum_{i=1}^{K} P(B_i) < \frac{K}{L\epsilon^2 P_{\text{min}}}$$

Property 3: The number $M$ of $\epsilon$-typical sequence $u$ from a $K$-ary DMS with entropy $H(U)$ in bits satisfies

$$\left(1 - \frac{K}{L\epsilon^2 P_{\text{min}}}\right) \cdot 2^{(1-\epsilon)LH(U)} < M \leq 2^{(1+\epsilon)LH(U)}$$

where $P_{\text{min}}$ is the smallest positive value of $P_U(u)$. 
Typical sequences

Property 3: The number $M$ of $\epsilon$-typical sequence $u$ from a K-ary DMS with entropy $H(U)$ in bits satisfies

$$
\left(1 - \frac{K}{L\epsilon^2 P_{\text{min}}}\right) \cdot 2^{(1-\epsilon)LH(U)} < M \leq 2^{(1+\epsilon)LH(U)}
$$

where $P_{\text{min}}$ is the smallest positive value of $P_U(u)$.

Proof:

$$
1 = \sum_{u} P_U(u) \geq M \cdot 2^{-(1+\epsilon)LH(U)}
$$

This gives the upper bound

$$
M \leq 2^{(1+\epsilon)LH(U)}
$$
Typical sequences

- Total probability of the \( \epsilon \)-typical sequences is \( 1 - P(F) \), so

\[
1 - P(F) \leq M \cdot 2^{-(1-\epsilon)LH(U)}
\]

- This gives the lower bound

\[
M > \left( 1 - \frac{K}{Le^2P_{\text{min}}} \right) \cdot 2^{(1-\epsilon)LH(U)}
\]
Asymptotic Equipartition Property

Property 3 says that when $L$ is large and $\epsilon$ is small, there are roughly $2^{LH(U)}\epsilon$—typical sequences $u$. 
Asymptotic Equipartition Property

- Property 3 says that when $L$ is large and $\epsilon$ is small, there are roughly $2^{LH(U)}\epsilon$—typical sequences $u$.
- Property 1 says each of these $\epsilon$—typical sequences has probability equal to $2^{-LH(U)}$.
- Property 2 says that the total probability of these $\epsilon$—typical sequences is very nearly 1.
Asymptotic Equipartition Property

- Property 3 says that when $L$ is large and $\epsilon$ is small, there are roughly $2^{L\mathcal{H}(U)}\epsilon$-typical sequences $u$.
- Property 1 says each of these $\epsilon$-typical sequences has probability equal to $2^{-L\mathcal{H}(U)}$.
- Property 2 says that the total probability of these $\epsilon$-typical sequences is very nearly 1.
- These three properties are known as asymptotic equipartition property (AEP) of the output sequence of a DMS.