COMP2610 / COMP6261 - Information Theory

Lecture 10: Typicality and Asymptotic Equipartition Property

Australian
National
University

21 August, 2018

Last time

Markov'

Chebys

Law of large numbers

Law of Large Numbers

Theorem

Let $X_1, \ldots, X_n$ be a sequence of iid random variables, with

$$\mathbb{E}[X_i] = \mu$$

and $\mathbb{V}[X_i$

Then, for a

$$\lim_{n\to\infty} p(|\bar{X}_n - \mu| <$$

This is also called $\bar{X}_n \to \mu$ **in probability**.

Definition: For random variables $v_1, v_2, \ldots$, we say $v_n \to v$ **in probability** if for all $\beta > 0$ $\lim_{n\to\infty} P(|v_n - v| > \beta) = 0$.

β is fixed (not shrinking like $\frac{1}{n}$). Not max/min. Reduction in variability.

This time

- Ensembles and sequences
- Typ
- Asymptotic Equipartition Property (AEP)

Ensembles

Ensem

An **ense** taking **values** i $p_2, \ldots, p_I\}$.

We will call $\mathcal{A}_X$ the alphabet of the ensemble

Ensembles

Example: Bent Coin

Let X be an ensemble with outcomes $\mathtt{h}$ for *heads* and $\mathtt{t}$ for *tails*.

- $\mathcal{A}_X = \{\mathtt{h}, \mathtt{t}\}$
- The proba

Extended Ensembles

We can also consider blocks of outcomes, which will be useful to describe sequences:

Example (Coin Flips):

`hhhh` blocks)

blocks)

$\rightarrow$ `hhhh thht hthh` (3 4 outcome blocks)

Extended Ensembles

We can also consider blocks of outcomes, which will be useful to describe sequences:

Example (Coin Flips):

`hhhh` blocks)

blocks)

$\rightarrow$ `hhhh thht hthh` (3 4 outcome blocks)

Extended Ensemble

Let X be a single ensemble. The **extended ensemble** of blocks of size N is denoted X^N. Outcomes from X^N are denoted $\mathbf{x} = (x_1, x_2, \ldots, x_N)$. The **probability** of $\mathbf{x}$ is defined to be $P(\mathbf{x}) = P(x_1)P(x_2)\ldots P(x_N)$.

Extended Ensembles

Example: Bent Coin

Let X be an ensemble with outcomes $\mathcal{A}_X = \{\text{h}, \text{t}\}$ with $p_\text{h} = 0.9$ and $p_\text{t} = 0.1$.

Extended Ensembles

Example: Bent Coin

Let X be an ensemble with outcomes $\mathcal{A}_X = \{\mathtt{h}, \mathtt{t}\}$ with $p_\mathtt{h} = 0.9$ and $p_\mathtt{t} = 0.1$.

$$P(\mathtt{hhhh}) = (0.9)^4 \approx 0.6561$$

$$P(\mathtt{tttt}) = (0.1)^4 = 0.0001$$

$$P(\mathtt{hthh}) = 0.9 \cdot 0.1 \cdot 0.9 \cdot 0.9 = (0.9)\ (0.1) \approx 0.0729$$

$$P(\mathtt{htht}) = 0.9 \cdot 0.1 \cdot 0.9 \cdot 0.1 = (0.9)^2(0.1)^2 \approx 0.0081.$$

Extended Ensembles

Example: Bent Coin

- Entropy of extended ensembles

We can view X^4 as comprising 4 independent random variables, based on the ense

Entropy i

Thus,

$$H(X^4) = 4H(X) = 4.(-0.9\log_2 0.9 \qquad \text{s.}$$

More generally,

$$H(X^N) = NH(X).$$

Counting Types of Sequences

Criteria for dividing 2^N sequences into **types**

In the bent coin example,

$$= P$$
$$= P$$
$$= P$$

The order of outcomes in the sequence is irrelevant

Counting Types of Sequences

Let X be an ensemble with alphabet $\mathcal{A}_X = \{a_1, \ldots, a_I\}$.

Let $p(x = a_i) = p_i$.

For a sequ

let n_i = # o

Given the n_i's, we can compute the probability of s

$$\begin{aligned} P(\mathbf{x}) &= P(x_1) \cdot P(x_2) \ldots \\ &= P(a_1)^{n_1} \cdot P(a_2)^{n_2} \cdot \ldots \cdot \quad (\quad_I)^{\prime} \\ &= p_1^{n_1} \cdot p_2^{n_2} \ldots p_I^{n_I} \end{aligned}$$

Sufficient statistics: $\{n_1, n_2, \ldots, n_I\}$. Use it as a criteria of partitioning.

Counting Types of Sequences

Sequence Types

Each unique choice of $(n_1, n_2, \ldots, n_l)$ gives a different type of sequence

- 4 heads, (3 heads, 1 tail), (2 heads, 2 tails), ...
- Sequences in each type are equiprobable

For a give
symbol c

# of sequences with n_i copies of $\quad \frac{\quad !}{\qquad\qquad}$

$$\binom{N}{n_1}\binom{N-n_1}{n_2}\binom{N-n_1-n_2}{n_3}\cdots$$

$$= \frac{N!}{n_1!(N-n_1)!} \cdot \frac{(N-n_1)!}{n_2!(N-n_1-n_2)!} \cdot \frac{(N-n_1-n_2)!}{n_3!(N-n_1-n_2-n_3)!} \cdots$$

Counting Types of Sequences

Example

Probability of types

Let $\mathcal{A} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ with $P(\mathtt{a}) = 0.2$, $P(\mathtt{b}) = 0.3$, $P(\mathtt{c}) = 0.5$.

Counting Types of Sequences

Example

Probability of types

Let $\mathcal{A} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ with $P(\mathtt{a}) = 0.2$, $P(\mathtt{b}) = 0.3$, $P(\mathtt{c}) = 0.5$.

Each sequence with 2 a's, 3 b's and 5 c's has probability $(0.2)^2(0.3)^3(0.5)^5$.

Counting Types of Sequences

Example

Probability of types

Let $\mathcal{A} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ with $P(\mathtt{a}) = 0.2$, $P(\mathtt{b}) = 0.3$, $P(\mathtt{c}) = 0.5$.

Each sequence with 2 a's, 1 b, and 3 c's has probability
$(0.2)^2(0.3)^1(0.5)^3 = 0.012$.

There are $\frac{6!}{2!1!3!} = 60$ such sequences.

Counting Types of Sequences

Example

Let $\mathcal{A} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ with $P(\mathtt{a}) = 0.2$, $P(\mathtt{b}) = 0.3$, $P(\mathtt{c}) = 0.5$.

Each sequence has probability $(0.2)^2(0.3)^1(0.5)^3 = 0.0075$.

There are $\frac{6!}{2!1!3!} = 60$ such sequences.

The probability $\mathbf{x}$ is of type $(2, 1, 3)$ is $(0.0$

Study probabilities at the level of types (most likely, average/typical)

Extended Ensembles

Example

With $p_h = 0.75$, what are the probabilities for X^N?

$N = 2$

x	
hh	
ht	
th	0.1875
tt	0.0625

Extended Ensembles

Example

With $p_h = 0.75$, what are the probabilities for X^N?

$N = 2$		$N = 3$	
x			
hh			
ht			
th	0.1875	hth	0.1406
tt	0.0625	thh	0.1406
		htt	0.0469
		tht	0.0469
		tth	0.0469
		ttt	0.0156

Extended Ensembles

Example

With $p_h = 0.75$, what are the probabilities for X^N?

$N = 2$		$N = 3$		$N = 4$			
x						$P(\mathbf{x})$	
hh						0.0352	
ht						0.0352	
th	0.1875	hth	0.1406	h		.0352	
tt	0.0625	thh	0.1406	h		.0117	
		htt	0.0469	t		.0117	
		tht	0.0469	h		.0117	
		tth	0.0469	htth	0.0352	ttth	0.0117
		ttt	0.0156	hhtt	0.0352	tttt	0.0039

Observations

As N increases, there is an increasing spread of probabilities

The most li

Howeve $\texttt{h}$ $\texttt{t}$

Not surprising because $3 = N \cdot p_{\texttt{h}}$, pretty m

Symbol Frequency in Long Sequences

To judge if a sequence is typical/average, a natural question to ask is:

How often does each symbol appear in a sequence **x** from X^N?

Intuitively, in a sequence of length N, let a_i appear for n_i times.
Then **in expectation**

Note p_i

$$P(\mathbf{x}) = P(a_1)^{n_1} P(a_2)^{n_2} \dots P(a_l$$

Symbol Frequency in Long Sequences

To judge if a sequence is typical/average, a natural question to ask is:

How often does each symbol appear in a sequence $\mathbf{x}$ from X^N?

Intuitively, in a sequence of length N, let a_i appear for n_i times.
Then **in expectation**

Note p_i

$$P(\mathbf{x}) = P(a_1)^{n_1} P(a_2)^{n_2} \dots P(a_I$$

So the *information content* $-\log_2 P(\mathbf{x})$ of th y

$$-p_1 N \log_2 p_1 - \dots - p_I N \log_2 p_I = -N \sum_{i=1}^{I} p_i \log_2 p_i = NH(X)$$

Typical Sets

We want to consider elements $\mathbf{x}$ that have $-\log_2 P(\mathbf{x})$ "close" to $NH(X)$

Typical Set

For "close

$$\beta \qquad 2$$

$$= \left\{ \mathbf{x} : \left| -\frac{1}{N} \log_2 P(\mathbf{x} \right. \right.$$

Union of types

Typical Sets

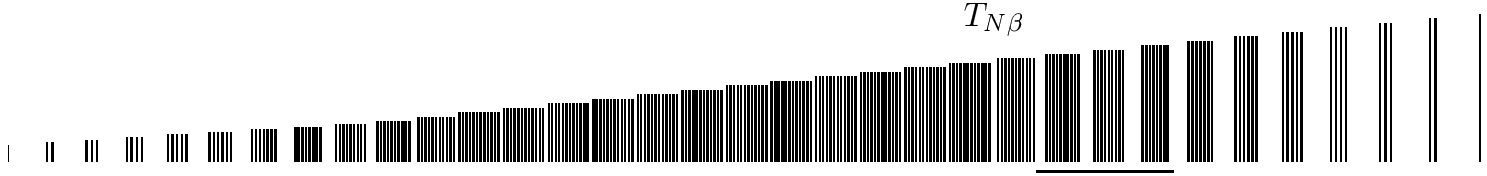
We want to consider elements $\mathbf{x}$ that have $-\log_2 P(\mathbf{x})$ "close" to $NH(X)$

Typical Set

For "closeness" $\beta > 0$ the typical set $T_{N\beta}$ for X^N is

$$T_{N\beta} \stackrel{\text{def}}{=} \{\mathbf{x} : \left| -\frac{1}{N}\log_2 P(\mathbf{x}) - H(X) \right| < \beta\}$$

Union of types



What when $\beta = 0$ (and replace $<$ by $\leq$)?

Criterion based on information content. Other criterion (KL divergence)?

Typical Sets

The name “typical” is used since $\mathbf{x} \in T_{N\beta}$ will have roughly $p_1 N$ occurrences of symbol a_1, $p_2 N$ of a_2, . . ., $p_K N$ of a_K.

$\mathbf{x}$	$\log_2(P(\mathbf{x}))$
...1.....1.....1....1.1.....1.....1.....1.....1.....11...	−50.1
.....1.....1.....1.....1.....1.....1.....1.....1....	−37.3
.....1...1..1...1...11..1.1.....11.....1..1.1..1..1.....1.	65.9
	56.4
	53.2
	43.7
	46.8
.....1..1..1.....111.....1.....1.....1.1...1...1.....1	−56.4
.....1.....1.....1.....1.....1.....1.....1.....1.....	
.....1.....1.....1.....1.....1..1.1.1.1..1....	
1.....1.....1.....1.....1.....1.....	
.....11.1.....1.....1.....1.....	
.1.....1...1.1.....1.....11.....1.1...1.....	
.....1...1..1.....1..11.1.1.1...1.....1.....1.....1..1.....	−59.5
.....11.1.....1....1..1.....1.....1.....1.....1.....	−46.8
.....	−15.2
11111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111	−332.1

Randomly drawn sequences for $P(1) = 0.1$. Note: $H(X) \approx 0.47$

Typical Sets

Properties

Typical sequences are nearly equiprobable: Every $\mathbf{x} \in T_{N\beta}$ has

Variatio

Number of sequences in the typical set: For any

$$|T_{N\beta}| \leq 2^{N(H(X}$$

Typical Sets

Proof of Cardinality Bound

For every $\mathbf{x} \in T_{N\beta}$,

$$p(\mathbf{x}) \geq 2^{-N(H(X)-\beta)}.$$

Thus,

$$1 \geq \sum_{\mathbf{x} \in T_{N\beta}} 2^{-N(H(}$$

$$\geq \sum_{\mathbf{x} \in T_{N\beta}} 2^{-N(H(X)+\beta)}$$

$$= 2^{-N(H(X)-\beta)} \cdot |T_{N\beta}|.$$

Thus

$$|T_{N\beta}| \leq 2^{N(H(X)+\beta)}$$

Typical Sets

Most Likely Sequence

The most likely sequence may not belong to the typical set

e.g. with

$=$

whereas $H(X) = 0.8113$

The most likely single sequence $\rightarrow$ `hhhh`

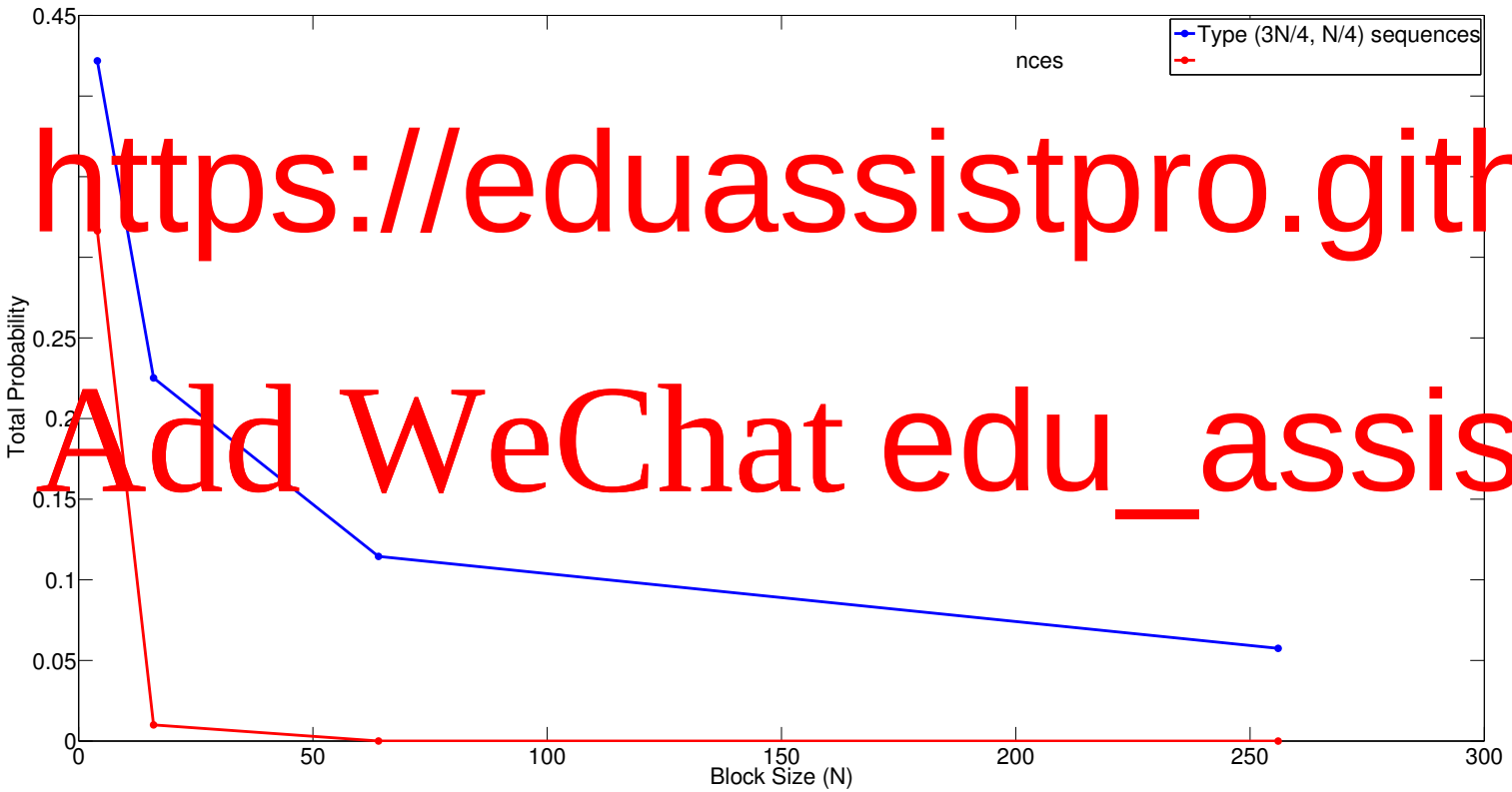
The most likely single sequence type $\rightarrow \{\texttt{hhht}, \texttt{hthh}, \ldots\}$

Typical Sets

Most Likely Sequence

Probability of most likely sequence decays like $(p_{\mathrm{h}})^N$ ($p_{\mathrm{h}} = 0.75$)

Sequences with $N \cdot p_{\mathrm{h}}$ heads contain much more total probability mass



Blue curve corresponds to typical set with $\beta = 0$. What if $\beta > 0$?

Asymptotic Equipartition Property

Eventually Equally Divided

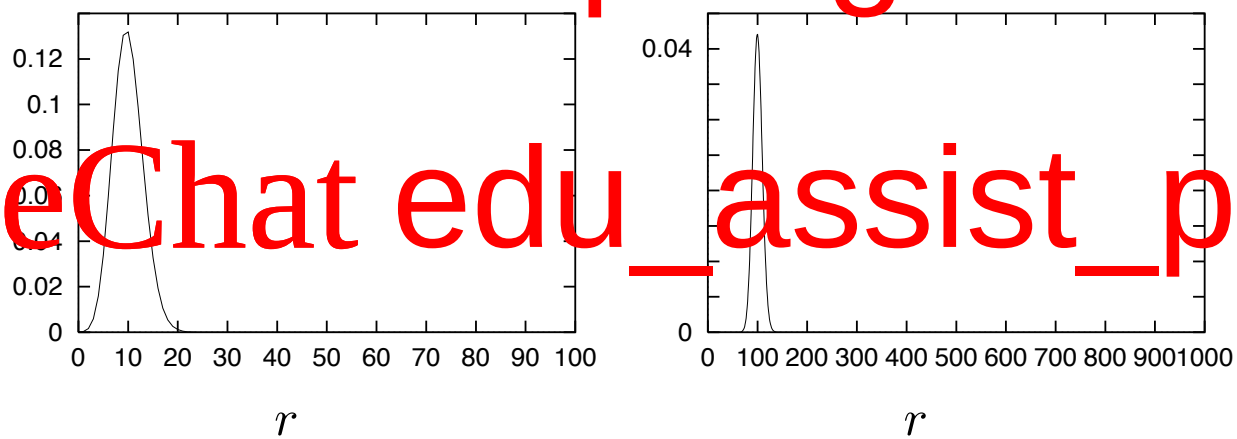
Informally

Asymptotic Equipartition Property (Informal)

As $N \to \infty$, $\log_2 P(x_1, \ldots, x_N)$ is close to $-NH(X)$ with high probability.

For la $T_{N\beta}$)

$$n(r)P(\mathbf{x}) = \binom{N}{r} p_1^r (1-p_1)^{N-r}$$



Probability sequence $\mathbf{x}$ has r heads for $N = 100$ (left) and $N = 1000$ (right). Here $P(X = \text{head}) = 0.1$.

Asymptotic Equipartition Property

Formally

Asymptotic Equipartition Property

If $x_1, x_2, \ldots$ are i.i.d. with distribution P then, in probability

$$\frac{1}{}$$

In precise!

$$(\forall \beta > 0) \lim_{N \to \infty} p \left| -\frac{}{N} \log_2 P(x_1, \ldots, \right. \qquad 1.$$

Exactly the probability of $\mathbf{x} \in T_{N\beta}$.

Asymptotic Equipartition Property

Formally

Asymptotic Equipartition Property

If $x_1, x_2, \ldots$ are i.i.d. with distribution P then, **in probability**,

$$\frac{1}{}$$

In precise!

$$(\forall \beta > 0) \lim_{N \to \infty} p \left| -\frac{}{N} \log_2 P(x_1, \ldots, \right. \qquad 1.$$

Exactly the probability of $\mathbf{x} \in T_{N\beta}$.

Recall definition: for random variables $v_1, v_2, \ldots$, we say $v_N \to v$ **in probability** if for all $\beta > 0$ $\lim_{N\to\infty} P(|v_N - v| > \beta) = 0$

Here v_N corresponds to $-\frac{1}{N} \log_2 P(x_1, \ldots, x_N)$.

Asymptotic Equipartition Property

Comments

Why is it surprising/significant?

For an ens

i.e. the typical set is a small fraction of all possible sequences

AEP says that for N sufficiently large we are $\forall i$ a sequence from this small set

Significance in information theory

Asymptotic Equipartition Property

Proof

Since $x_1, \ldots, x_N$ are independent,

$$-\frac{1}{N}\log p(x_1, \ldots, x_N) = -\frac{1}{n}\log \prod_{n=1}^{N} p(x_n)$$

Let $Y = -\log p(X)$ and $y_n = -\log p(x_n)$. T

$$\mathbb{E}[Y] = H(X)$$

But then by the law of large numbers,

$$(\forall \beta > 0) \lim_{N\to\infty} p\left(\left|\frac{1}{N}\sum_{n=1}^{N} y_n - H(X)\right| > \beta\right) = 0.$$

Summary & Conclusions

- Ensembles and sequences

- Typ

- Asymptotic Equipartition Property (AEP)

Next: Source Coding.