

COMP2610/COMP6261 - Information Theory

Lecture 10: Typicality and Approximate Equipartition

Mark Reid and **Aditya Menon**

Research School of Computer Science
The Australian National University



August 20th, 2014

Last time

Markov's inequality

Chebyshev's inequality

Law of large numbers

This time

- Ensembles and sequences
- Typical sets
- Approximation Equipartition (AEP)

- 1 Ensembles and sequences
 - Counting Types of Sequences
- 2 Typical sets
- 3 Asymptotic Equipartition (AEP)
- 4 Wrapping Up

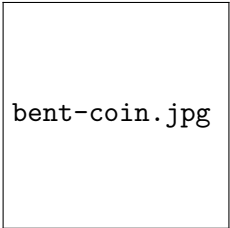
Ensemble

An **ensemble** X is a triple $(x, \mathcal{A}_X, \mathcal{P}_X)$; x is a **random variable** taking **values** in $\mathcal{A}_X = \{a_1, a_2, \dots, a_I\}$ with **probabilities** $\mathcal{P}_X = \{p_1, p_2, \dots, p_I\}$.

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

Ensembles

Example: Bent Coin



bent-coin.jpg

Let X be an **ensemble** with outcomes \mathbf{h} for *heads* with probability 0.9 and \mathbf{t} for *tails* with probability 0.1.

- The **outcome set** is $\mathcal{A}_X = \{\mathbf{h}, \mathbf{t}\}$
- The **probabilities** are $\mathcal{P}_X = \{p_{\mathbf{h}} = 0.9, p_{\mathbf{t}} = 0.1\}$

Extended Ensembles

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

Example (Coin Flips):

hhhhthhththh	→ hh hh th ht ht hh	(6 × 2 outcome blocks)
	→ hhh hth hth thh	(4 × 3 outcome blocks)
	→ 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):


hhhhthhththh	→ hh hh th ht ht hh	(6 × 2 outcome blocks)
	→ hhh hth hth thh	(4 × 3 outcome blocks)
	→ 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 $\mathbf{x} = (x_1, x_2, \dots, x_N)$. The **probability** of \mathbf{x} is defined to be $P(\mathbf{x}) = P(x_1)P(x_2) \dots P(x_N)$.

Extended Ensembles

Example: Bent Coin



bent-coin.jpg

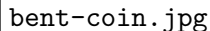
Let X be an ensemble with outcomes $\mathcal{A}_X = \{h, t\}$ with $p_h = 0.9$ and $p_t = 0.1$.

Consider X^4 – i.e., 4 flips of the coin.

$\mathcal{A}_{X^4} = \{hhhh, hhht, hhth, \dots, tttt\}$

Extended Ensembles

Example: Bent Coin



bent-coin.jpg

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

Consider X^4 – i.e., 4 flips of the coin.

$$\mathcal{A}_{X^4} = \{hhhh, hhht, hhth, \dots, tttt\}$$

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

$$P(tttt) = (0.1)^4 = 0.0001$$

$$P(hthh) = (0.9)^3(0.1) \approx 0.0729$$

$$P(htht) = (0.9)^2(0.1)^2 \approx 0.0081.$$

Extended Ensembles

Example: Bent Coin

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

Entropy is additive for independent random variables

Thus,

$$H(X^4) = 4H(X) = 4. (-0.9 \log_2 0.9 - 0.1 \log_2 0.1) = 1.88\text{bits}.$$

More generally,

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

Counting Types of Sequences

In the bent coin example,

$$\begin{aligned}(0.9)^2(0.1)^2 &= P(\text{hhtt}) \\ &= P(\text{htht}) \\ &= P(\text{htth}) \\ &= P(\text{thht}) \\ &= P(\text{thth}) \\ &= P(\text{tthh}).\end{aligned}$$

The **order** of outcomes in the sequence is **irrelevant**

Counting Types of Sequences

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

For a sequence $\mathbf{x} = x_1, x_2, \dots, x_N$, let $n_i = \#$ of times symbol a_i appears in \mathbf{x}

Given the n_i 's, we can compute the probability of seeing \mathbf{x} :

$$\begin{aligned} P(\mathbf{x}) &= P(x_1).P(x_2) \dots P(x_N) \\ &= P(a_1)^{n_1}.P(a_2)^{n_2} \dots P(a_I)^{n_I} \end{aligned}$$

Counting Types of Sequences

Sequence Types

Each unique choice of (n_1, n_2, \dots, n_I) gives a different **type** of sequence

- 4 heads, (3 heads, 1 tail), (2 heads, 2 tails), ...

For a given **type** of sequence how many sequences are there with these symbol counts?

$$\# \text{ of sequences with } n_i \text{ copies of } a_i = \frac{N!}{n_1! n_2! \dots n_I!}$$

Counting Types of Sequences

Example

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

Counting Types of Sequences

Example

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

Each sequence of type $(n_a, n_b, n_c) = (2, 1, 3)$ has length 6 and probability $(0.2)^2(0.3)^1(0.5)^3 = 0.0015$.

Counting Types of Sequences

Example

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

Each sequence of type $(n_a, n_b, n_c) = (2, 1, 3)$ has length 6 and probability $(0.2)^2(0.3)^1(0.5)^3 = 0.0015$.

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

Counting Types of Sequences

Example

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

Each sequence of type $(n_a, n_b, n_c) = (2, 1, 3)$ has length 6 and probability $(0.2)^2(0.3)^1(0.5)^3 = 0.0015$.

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

The probability \mathbf{x} is of type $(2, 1, 3)$ is $(0.0015) \cdot 60 = 0.09$.

- 1 Ensembles and sequences
 - Counting Types of Sequences
- 2 Typical sets
- 3 Asymptotic Equipartition (AEP)
- 4 Wrapping Up

Extended Ensembles

Example

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

$$N = 2$$

\mathbf{x}	$P(\mathbf{x})$
hh	0.5625
ht	0.1875
th	0.1875
tt	0.0625

Extended Ensembles

Example

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

$N = 2$

\mathbf{x}	$P(\mathbf{x})$
hh	0.5625
ht	0.1875
th	0.1875
tt	0.0625

$N = 3$

\mathbf{x}	$P(\mathbf{x})$
hhh	0.4219
hht	0.1406
hth	0.1406
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 ?

$N = 2$

\mathbf{x}	$P(\mathbf{x})$
hh	0.5625
ht	0.1875
th	0.1875
tt	0.0625

$N = 3$

\mathbf{x}	$P(\mathbf{x})$
hhh	0.4219
hht	0.1406
hth	0.1406
thh	0.1406
htt	0.0469
tht	0.0469
tth	0.0469
ttt	0.0156

$N = 4$

\mathbf{x}	$P(\mathbf{x})$	\mathbf{x}	$P(\mathbf{x})$
hhhh	0.3164	thht	0.0352
hhht	0.1055	thth	0.0352
hhtth	0.1055	tthh	0.0352
hthhh	0.1055	httt	0.0117
thhhh	0.1055	thtt	0.0117
httht	0.0352	ttth	0.0117
httth	0.0352	ttth	0.0117
hhtt	0.0352	tttt	0.0039

Observations

As N increases, there is an increasing spread of probabilities

The most likely single sequence will always be the all h's

However, for $N = 4$, the most likely sequence **type** is 3 h's and 1 t

Symbol Frequency in Long Sequences

A natural question to ask is:

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

Symbol Frequency in Long Sequences

A natural question to ask is:

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

Intuitively, we would expect to see

a_i roughly $n_i \approx N \cdot p_i$ times in sequence of length N .

So $P(\mathbf{x}) = P(a_1)^{n_1} P(a_2)^{n_2} \dots P(a_I)^{n_I} \approx p_1^{p_1 N} p_2^{p_2 N} \dots p_I^{p_I N}$

Symbol Frequency in Long Sequences

A natural question to ask is:

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

Intuitively, we would expect to see

a_i roughly $n_i \approx N \cdot p_i$ times in sequence of length N .

So $P(\mathbf{x}) = P(a_1)^{n_1} P(a_2)^{n_2} \dots P(a_I)^{n_I} \approx p_1^{p_1 N} p_2^{p_2 N} \dots p_I^{p_I N}$

So the *information content* $-\log_2 P(\mathbf{x})$ of that sequence is approximately

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

Typical Sets

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

Typical Set

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

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

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

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

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 , \dots , $p_K N$ of a_K .

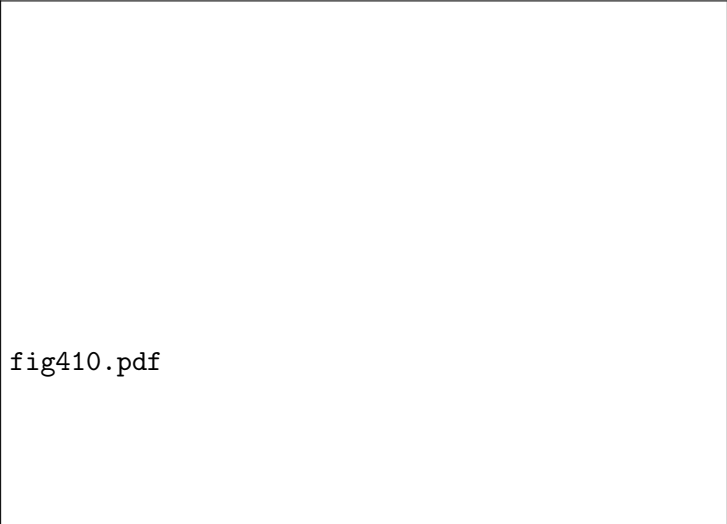


fig410.pdf

Typical Sets

Properties

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

$$2^{-N(H(X)+\beta)} \leq P(\mathbf{x}) \leq 2^{-N(H(X)-\beta)}.$$

Number of sequences in the typical set: For any N, β ,

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

Typical Sets

Proof of Cardinality Bound

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

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

Thus,

$$\begin{aligned} 1 &= \sum_{\mathbf{x}} p(\mathbf{x}) \\ &\geq \sum_{\mathbf{x} \in T_{N\beta}} p(\mathbf{x}) \\ &\geq \sum_{\mathbf{x} \in T_{N\beta}} 2^{-N(H(X)-\beta)} \\ &= 2^{-N(H(X)-\beta)} \cdot |T_{N\beta}|. \end{aligned}$$

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 $p_h = 0.75$, we have

$$-\frac{1}{4} \log_2 P(\text{hhhh}) = 0.4150$$

whereas $H(X) = 0.8113$

The most likely single sequence $\rightarrow \text{hhhh}$

The most likely single sequence **type** $\rightarrow \{\text{hhtt}, \text{htht}, \dots\}$

Typical Sets

Most Likely Sequence

Probability of most likely sequence decays like θ^N

Sequences with $N\theta$ heads contain much more total probability mass

- 1 Ensembles and sequences
 - Counting Types of Sequences
- 2 Typical sets
- 3 Asymptotic Equipartition (AEP)
- 4 Wrapping Up

Asymptotic Equipartition Property

Eventually Informally Equally Divided

Asymptotic Equipartition Property (Informal)

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

For large block sizes “almost all sequences are typical” (i.e., in $T_{N\beta}$)

Probability sequence \mathbf{x} has r 1s for $N = 100$ (left) and $N = 1000$ (right)

Asymptotic Equipartition Property

Formally

Asymptotic Equipartition Property

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

$$-\frac{1}{N} \log_2 P(x_1, \dots, x_N) \rightarrow H(X)$$

Asymptotic Equipartition Property

Formally

Asymptotic Equipartition Property

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

$$-\frac{1}{N} \log_2 P(x_1, \dots, x_N) \rightarrow H(X)$$

Defn: For i.i.d. v_1, v_2, \dots we say $v_N \rightarrow v$ **in probability** if for all $\epsilon > 0$
 $\lim_{N \rightarrow \infty} P(|v_N - v| > \epsilon) = 0$

Asymptotic Equipartition Property

Proof

Since x_1, \dots, x_N are independent,

$$\begin{aligned} -\frac{1}{N} \log p(x_1, \dots, x_N) &= -\frac{1}{N} \log \prod_{n=1}^N p(x_n) \\ &= -\frac{1}{N} \sum_{n=1}^N \log p(x_n). \end{aligned}$$

Let $Y = -\log p(X)$ and $y_n = -\log p(x_n)$. Then, $y_n \sim Y$, and

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

But then by the law of large numbers,

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

Asymptotic Equipartition Property

Comments

For an ensemble with binary outcomes, and low entropy,

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

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

AEP says that for N sufficiently large, we are virtually guaranteed to draw a sequence from this small set

- 1 Ensembles and sequences
 - Counting Types of Sequences
- 2 Typical sets
- 3 Asymptotic Equipartition (AEP)
- 4 Wrapping Up

Summary & Conclusions

- Ensembles and sequences
- Typical sets
- Approximation Equipartition (AEP)