Recall (a little sloppy, use lecture 12 for recall)

Letter Typical Sequences

- Let $\mathcal{X}$ be a finite alphabet and $\mathcal{X}^n$ be its n-fold extension

- Empirical Frequency ↙ needed to define letter typical sequences

  - $x^n = (x_1, \dots, x_n) \in \mathcal{X}^n$

defined as (a counter) → - $N_{x^n}(a) = \sum_{i=1}^{n} \mathbb{1}_{\{x_i = a\}}$ $a \in \mathcal{X}$

normalized is actual empirical frequency → - $\nu_{x^n}(a) = \frac{1}{n} N_{x^n}(a)$ → $\nu_{x^n}(a)$ is a valid pmf over $\mathcal{X}^n$

Letter Typical Set : $\mathbb{P} \in \mathcal{P}(\mathcal{X})$, $\varepsilon > 0$, $n \in \mathbb{N}$
(can now define w/ notion of empirical frequency)

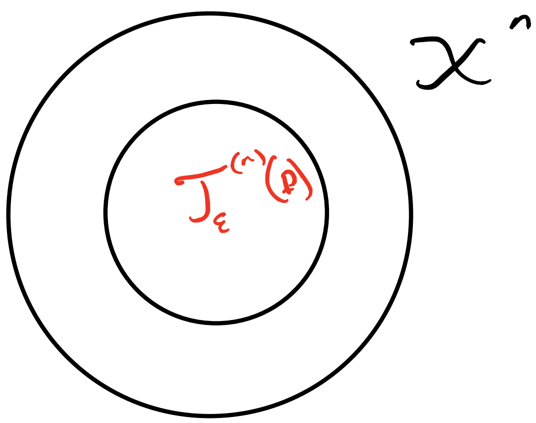
distribution   slackness   length

$$T_\varepsilon^{(n)}(\mathbb{P}) = \{x^n \in \mathcal{X}^n \mid |\nu_{x^n}(a) - p(a)| < \varepsilon\}$$

epsilon letter typical set of length n for $\mathbb{P}$

pmf induced by $\mathbb{P}$: $p(a) \triangleq \mathbb{P}(\{a\})$

Target



(i) $T_\varepsilon^{(n)}(P)$ is "small"

$$\Rightarrow \frac{|T_\varepsilon^{(n)}(P)|}{|\mathcal{X}^n|} \xrightarrow{n\to\infty} 0$$

(ii) $X_1, \dots, X_n$ iid P

$$P^{\otimes n}(X^n \in T_\varepsilon^{(n)}(P)) = P^{\otimes n}(T_\varepsilon^{(n)}(P)) \xrightarrow{n\to\infty} 1$$

i.e want a small but probable subset of entire space.

Properties of letter Typical Sets

① Typical Average Lemma P g $\mathbb{E}_P[g(X)]$

$\forall x^n \in T_\varepsilon^{(n)}(P)$, $\frac{1}{n}\sum_i g(x_i) \approx \mathbb{E}_P[g(X)]$

i.e

$$(1-\varepsilon)\,\mathbb{E}_P[g(X)] \le \frac{1}{n}\sum_i g(x_i) \le (1+\varepsilon)\,\mathbb{E}_P[g(X)]$$

② Fundamental properties of $T_\varepsilon^{(n)}(P)$:

(i) Letter typical sequences are roughly equiprobable

$\forall x^n \in T_\varepsilon^{(n)}(P)$

$$2^{-nH(P)(1+\varepsilon)} \le P^{\otimes n}(\{x^n\}) \le 2^{-nH(P)(1-\varepsilon)}$$

(ii) Letter Typical Set has high prob.

$$\lim_{n\to\infty} \mathbb{P}^{\otimes n}\left(\mathcal{T}_\varepsilon^{(n)}(\mathbb{P})\right) = 1$$

(iii) Cardinality of Letter typical set

for any n

$$(1-\varepsilon)\, 2^{nH(P)(1-\varepsilon)} \leq \left| \mathcal{T}_\varepsilon^{(n)}(\mathbb{P}) \right| \leq 2^{nH(P)(1+\varepsilon)}$$

n large enough

New Material

Refinement of above fundamental properties

Remark: Via an alternative proof technique (Chernoff bound and not LLN) we may obtain the following refined result

$$1 - \delta_\varepsilon(p,n) \leq \mathbb{P}^{\otimes n}\left(\mathcal{T}_\varepsilon^{(n)}(\mathbb{P})\right) \leq 1$$

where

$$\delta_\varepsilon(p,n) = 2|\mathcal{X}|\, e^{-2n\varepsilon^2\mu^2}$$

converges exponentially fast to 1

and

$$\mu := \min_{a \in \mathrm{supp}(\mathbb{P})} \mathbb{P}(\{a\}) > 0$$

Note that

$$\lim_{n\to\infty} \delta_\varepsilon(\mathbb{P}, n) = 0 \quad \forall\, \varepsilon > 0$$

The above refines claim (ii) from properties of typical sets.

On top of that it can be used to get a finite n lower bound on the cardinality of $T_\varepsilon^{(n)}(P)$.

$$(1-\delta_\varepsilon(P,n))\, 2^{-nH(P)(1-\varepsilon)} \leq |T_\varepsilon^{(n)}(P)| \leq 2^{nH(P)(1+\varepsilon)} \quad \forall n \in \mathbb{N}$$

<u>Exercise</u>

Under the assumption that the refined lower bound on the probability of the typical set holds $\left(P^{\otimes n}(T_\varepsilon^{(n)}(P))\right)$, prove the lower bound on $|T_\varepsilon^{(n)}(P)|$.

Jointly Typical Sequences

Extend the notion of typicality to pairs of sequences that are taken from a product space.

Let $\mathcal{X}$ and $\mathcal{Y}$ be finite alphabets and $\mathcal{X}^n$ and $\mathcal{Y}^n$ be their n-fold extension.

For any $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$, define

$$N_{x^n,y^n}(a,b) = \sum_{i=1}^{n} \mathbb{1}_{\{x_i=a,\, y_i=b\}} \qquad (a,b) \in \mathcal{X} \times \mathcal{Y}$$

$\Rightarrow$ Emprical frequency of (x^n, y^n) is defined as

$$\nu_{x^n, y^n}(a,b) := \frac{1}{n} N_{x^n, y^n}(a,b) \qquad (a,b) \in \mathcal{X} \times \mathcal{Y}$$

The empirical frequency is a valid pmf on $\mathcal{X} \times \mathcal{Y}$.
The associated probability measure is called the empirical measure of $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$.

Example: Let $X = \{0, 1\}$, $\mathcal{Y} = \{a, b, c\}$, $n = 10$.

$$x^{10} = 0110001101$$
$$y^{10} = aacabbacac$$

$N_{x^{10}, y^{10}}(0, a) = 3$ $\qquad$ $N_{x^{10}, y^{10}}(1, a) = 2$

$N_{x^{10}, y^{10}}(0, b) = 2$ $\qquad$ $N_{x^{10}, y^{10}}(1, b) = 0$

$N_{x^{10}, y^{10}}(0, c) = 0$ $\qquad$ $N_{x^{10}, y^{10}}(1, c) = 3$

divide by 10 to get $\nu_{x^n, y^n}(a, b)$ for $(a, b) \in \{0, 1\} \times \{a, b, c\}$

Empirical frequency in matrix form is

$$\nu_{x^{10}, y^{10}} = \begin{array}{c|c} x \backslash y & \begin{array}{ccc} a & b & c \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} 3/10 & 2/10 & 0 \\ 2/10 & 0 & 3/10 \end{bmatrix} \end{array}$$

Definition (Joint Letter Typical Set)

Given $n \in \mathbb{N}$, $\varepsilon > 0$ and $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ with pmf p_{XY},

the joint ε-letter typical set is

$$T_\varepsilon^{(n)}(P_{XY}) = \left\{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \left| \nu_{x^n, y^n}(a,b) - p_{XY}(a,b) \right| \leq \varepsilon \, p_{XY}(a,b) \right\}$$

$\forall \ (a,b) \in \mathcal{X} \times \mathcal{Y}$

Properties of letter typical sequences extend to the joint case as follows:

Theorem (Properties of Jointly Typical Sets)

Let $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, $n \in \mathbb{N}$, $\varepsilon > 0$, and consider the jointly typical set $T_\varepsilon^{(n)}(P_{XY})$.

We have

1) If $(x^n, y^n) \in T_\varepsilon^{(n)}(P_{XY})$ then $x^n \in T_\varepsilon^{(n)}(P_X)$ and $y^n \in T_\varepsilon^{(n)}(P_Y)$

2) For each $(x^n, y^n) \in T_\varepsilon^{(n)}(P_{XY})$

(i) $2^{-nH(P_{XY})(1+\varepsilon)} \leq P_{XY}^{\otimes n}(\{(x^n, y^n)\}) \leq 2^{-nH(P_{XY})(1-\varepsilon)}$

(ii) $2^{-nH(P_X)(1+\varepsilon)} \leq P_X^{\otimes n}(\{x^n\}) \leq 2^{-nH(P_X)(1-\varepsilon)}$

(iii) $2^{-nH(P_Y)(1+\varepsilon)} \leq P^{\otimes n}(\{y^n\}) \leq 2^{-nH(P_Y)(1-\varepsilon)}$

3) If $(X_1,Y_1),(X_2,Y_2),\ldots$ are iid according to P_{XY}, then

$$\lim_{n\to\infty} \underbrace{P_{XY}^{\otimes n}\left((X^n,Y^n)\in T_\varepsilon^{(n)}(P_{XY})\right)}_{P_{XY}^{\otimes n}(T_\varepsilon^{(n)}(P_{XY}))} = 1$$

and a refined version is

$$1-\delta_\varepsilon(P_{XY},n) \le P_{XY}^{\otimes n}\left((X^n,Y^n)\in T_\varepsilon^{(n)}(P_{XY})\right) \le 1$$

where

$$\delta_\varepsilon(P_{XY},n) := 2|\mathcal{X}||\mathcal{Y}|e^{-2n\varepsilon^2\mu_{XY}^2}$$

and

$$\mu_{XY} := \min_{(x,y)\in \operatorname{supp}(P_{XY})} P_{XY}(\{(x,y)\}) > 0$$

4) The cardinality of the jointly typical set is

$$(1-\varepsilon)2^{nH(P_{XY})(1-\varepsilon)} \underset{\text{for } n \text{ large enough}}{\le} |T_\varepsilon^{(n)}(P_{XY})| \overset{\text{for any } n}{\le} 2^{nH(P_{XY})(1+\varepsilon)}$$

This can be refined to

$$(1-\delta_\varepsilon(P_{XY},n))2^{nH(P_{XY})(1-\varepsilon)} \le |T_\varepsilon^{(n)}(P_{XY})| \le 2^{nH(P_{XY})(1+\varepsilon)}$$

$$\begin{tikzcd} & P_X \arrow[dr] & \\ P_{XY} \arrow[ur] \arrow[dr] & & P_X \otimes P_Y \\ & P_Y \arrow[ur] & \end{tikzcd}$$

$$(\tilde{X}^n, \tilde{Y}^n) \overset{iid}{\sim} P_X \otimes P_Y$$

$$\Rightarrow \tilde{X}^n \overset{iid}{\sim} P_X$$

$$\tilde{Y}^n \overset{iid}{\sim} P_Y$$

Aside

The last property we need says that joint typicality is a relation that is hard to fabricate/simulate using uncorrelated sequences.

Consider the following setup: Fix $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and let $P_X \in \mathcal{P}(\mathcal{X})$ and $P_Y \in \mathcal{P}(\mathcal{Y})$ be its X- and Y-marginals.

Let $(\tilde{X}, \tilde{Y}) \sim P_X \otimes P_Y$ (i.e $\tilde{X} \perp\!\!\!\perp \tilde{Y}$ w/ the "right" marginals), and consider n iid copies of $(\tilde{X}, \tilde{Y})$ denoted by $(\tilde{X}^n, \tilde{Y}^n)$.

We ask what is the probability that $(\tilde{X}^n, \tilde{Y}^n)$ looks jointly letter typical w.r.t P_{XY}, i.e., $(\tilde{X}^n, \tilde{Y}^n) \in T_\epsilon^{(n)}(P_{XY})$?

Theorem: Under the above setup, the following holds:

$$2^{-n(I(X;Y)+\delta(\varepsilon))} \leq \left(P_X^{\otimes n} \otimes P_Y^{\otimes n}\right)\left((\tilde{X}^n, \tilde{Y}^n) \in T_\varepsilon^{(n)}(P_{XY})\right) \leq 2^{-n(I(X;Y) - \delta(\varepsilon))}$$

$D_{KL}(P_{XY} \| P_X \otimes P_Y)$

where

$$\delta(\varepsilon) \xrightarrow[\varepsilon \to 0]{} 0$$

i.e for VERY dependent events this probability is <u>SMALL</u>

Lemma: Let $\varepsilon > 0$, $X^n \sim P^{\otimes n}$ and $\tilde{P} \ll P$. Then

$$(1-\varepsilon)\, 2^{-n\left(D_{KL}(\tilde{P} \| P) + \varepsilon \log(\mu \cdot \tilde{\mu})\right)} \leq P^{\otimes n}\left(X^n \in T_\varepsilon^{(n)}(\tilde{P})\right) \leq (1+\varepsilon)\, 2^{-n\left(D_{KL}(\tilde{P} \| P) - \varepsilon \log(\mu \cdot \tilde{\mu})\right)}$$

for n sufficiently large (lower bound), where

$$\mu := \min_{a \in \mathrm{supp}(P)} P(\{a\}) \qquad \text{and} \qquad \hat{\mu} := \min_{a \in \mathrm{supp}(\tilde{P})} \tilde{P}(\{a\})$$

- Note that by taking $P = P_X \otimes P_Y$ and $\tilde{P} = P_{XY}$ in the lemma gives the above theorem as a corollary.

Proof: For $\tilde{X}^n \in T_\varepsilon^{(n)}(\tilde{P})$ we have

$$P^{\otimes n}(\tilde{X}^n) = \prod_{i=1}^{n} P(\{\tilde{X}_i\}) = \prod_{a \in supp(\tilde{P})} P(\{a\})^{N_{\tilde{X}^n}(a)}$$

$$\leq \prod_{a \in supp(\tilde{P})} P(\{a\})^{n \cdot \tilde{P}(\{a\})(1-\varepsilon)}$$

$$= 2^{n(1-\varepsilon) \sum_{a \in supp(a)} \tilde{P}(\{a\}) \log \tilde{P}(\{a\})}$$

$$\Rightarrow P^{\otimes n}(X^n \in T_\varepsilon^{(n)}(\tilde{P})) = \sum_{\tilde{x}^n \in T_\varepsilon^{(n)}(\tilde{P})} P^{\otimes n}(\{\tilde{x}^n\})$$

$$\leq |T_\varepsilon^{(n)}(\tilde{P})| \cdot 2^{n(1-\varepsilon) \sum_{a \in supp(\tilde{P})} \tilde{P}(\{a\}) \log P(\{a\})}$$

$$< 2^{n(1+\varepsilon)H(\tilde{P})} \, 2^{n(1-\varepsilon) \sum_{a \in supp(\tilde{P})} \tilde{P}(\{a\}) \log P(\{a\})}$$

$$\leq 2^{-n(D_{KL}(\tilde{P} \| P) - \varepsilon \log_2(\mu \cdot \tilde{\mu}))}$$

A lower bound can be derived similarly