Single-shot Thermo, lecture 7

2.12.2014

- For exercise sheet 5, you have time until next week.
- Quick repetition last time:

$$p \xrightarrow{\varepsilon\text{-noisy}} q \iff \exists q' : D(q,q') \le \varepsilon,\ p \xrightarrow{\text{noisy}} q'.$$

Smooth entropies:

$$H_\infty^\varepsilon(p) := \max_{p' : D(p,p') \le \varepsilon} H_\infty(p'),$$

$$H_0^\varepsilon(p) := \min_{p' : D(p,p') \le \varepsilon} H_0(p').$$

Minimum I with $S_I \xrightarrow{\varepsilon\text{-noisy}} p$ is $I_\infty^\varepsilon(p) := \log d_p - H_\infty^\varepsilon(p)$;

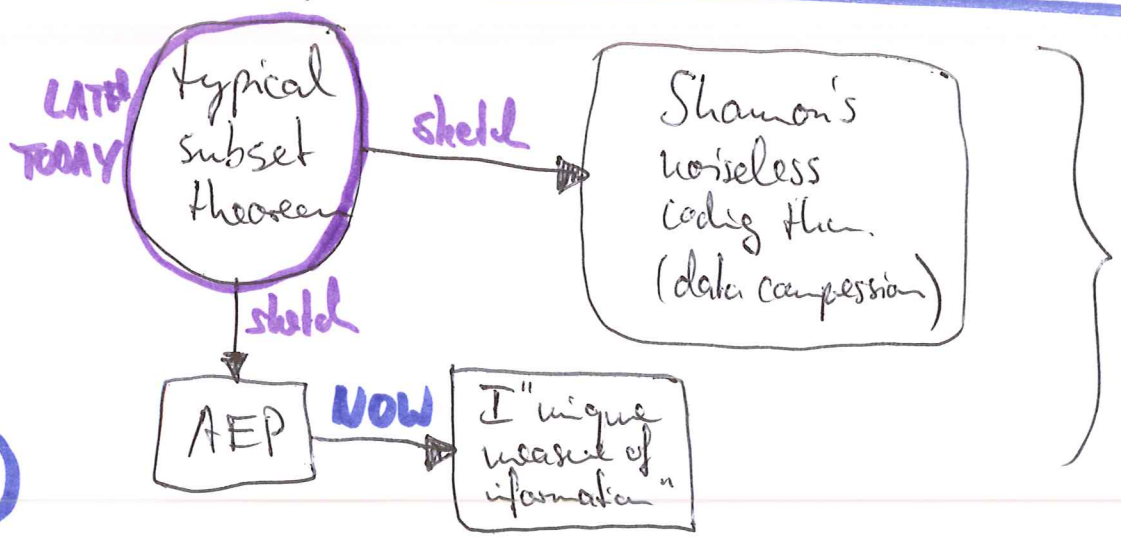
maximum I with $p \xrightarrow{\varepsilon\text{-noisy}} S_I$ is at least $I_0^\varepsilon(p)$.

Asymptotic equipartition property (AEP) (proof sketch today):

$$\lim_{n\to\infty} \frac{1}{n} H_{0/\infty}^\varepsilon(p^{\otimes n}) = H(p) \Rightarrow \lim_{n\to\infty} \frac{1}{n} I_{0/\infty}^\varepsilon(p^{\otimes n}) = I(p).$$

$\forall \varepsilon \in (0,1)$.

2.10. State conversion in the thermodynamic limit: recovering Shannon and von Neumann entropy



Use in addition:

If $I_0^{\varepsilon/2}(p) \geq I_\infty^{\varepsilon/2}(q)$ then $p \xrightarrow{\varepsilon\text{-noisy}} q$. (homework)

Theorem ("unique measure of information"):
p, q prob. vectors, not maximally mixed, $0 < \varepsilon < 1$. For every $n \in \mathbb{N}$, let m_n be the largest integer with

$$p^{\otimes n} \xrightarrow{\varepsilon\text{-noisy}} q^{\otimes m_n}$$

Then $\lim_{n\to\infty} \frac{m_n}{n} = \frac{I(p)}{I(q)}$.

Converse + interpretation: see last lecture.

Proof: If $\frac{m}{n} \leq \frac{\frac{1}{n} I_0^{\varepsilon/2}(p^{\otimes n})}{\frac{1}{m} I_\infty^{\varepsilon/2}(q^{\otimes m})}$ then $p^{\otimes n} \xrightarrow{\varepsilon\text{-noisy}} q^{\otimes m}$.

But this is impossible for $m = m_n + 1$

$$\Rightarrow \frac{m_n+1}{n} > \frac{\frac{1}{n} I_0^{\varepsilon/2}(p^{\otimes n})}{\frac{1}{m_n+1} I_\infty^{\varepsilon/2}(q^{\otimes(m_n+1)})}.$$

Easy to see: $m_n \xrightarrow{n\to\infty} \infty$ so

$$\liminf_{n\to\infty} \frac{m_n}{n} \geq \liminf_{n\to\infty} \frac{\frac{1}{n} I_0^{\varepsilon/2}(p^{\otimes n})}{\frac{1}{m_n+1} I_\infty^{\varepsilon/2}(q^{\otimes(m_n+1)})} = \frac{\lim_{n\to\infty} \frac{1}{n} I_0^{\varepsilon/2}(p^{\otimes n})}{\lim_{m\to\infty} \frac{1}{m} I_\infty^{\varepsilon}(q^{\otimes m})}$$

$$= \frac{I(p)}{I(q)}.$$

We have $p^{\otimes n} \xrightarrow{\text{noisy}} q_n$ with $D(q^{\otimes m_n}, q_n) \leq \varepsilon$

Choose $\delta > 0$ such that $\varepsilon + \delta < 1$.

Homework: I_∞^δ is a nonuniformity monotone $\Rightarrow$

$$I_\infty^\delta(p^{\otimes n}) \geq I_\infty^\delta(q_n) = I_\infty(q_n') \text{ with } D(q_n, q_n') \leq \delta.$$

$$\Rightarrow D(q^{\otimes m_n}, q_n') \leq D(q^{\otimes m_n}, q_n) + D(q_n, q_n') \leq \varepsilon + \delta$$

$$\Rightarrow I_\infty^{\delta+\varepsilon}(q^{\otimes m_n}) = \min_{q' : D(q^{\otimes m_n}, q') \leq \delta+\varepsilon} I_\infty(q')$$

$$\leq I_\infty(q_n') \leq I_\infty^\delta(p^{\otimes n}).$$

$$\Rightarrow \frac{m_n}{n} \leq \frac{\frac{1}{n} I_\infty^\delta(p^{\otimes n})}{\frac{1}{m_n} I_\infty^{\varepsilon+\delta}(q^{\otimes m_n})}$$

$$\Rightarrow \limsup_{n\to\infty} \frac{m_n}{n} \leq \frac{\lim_{n\to\infty} \frac{1}{n} I_\infty^\delta(p^{\otimes n})}{\lim_{m\to\infty} \frac{1}{m} I_\infty^{\varepsilon+\delta}(q^{\otimes m_n})} = \frac{I(p)}{I(q)}. \quad \square$$

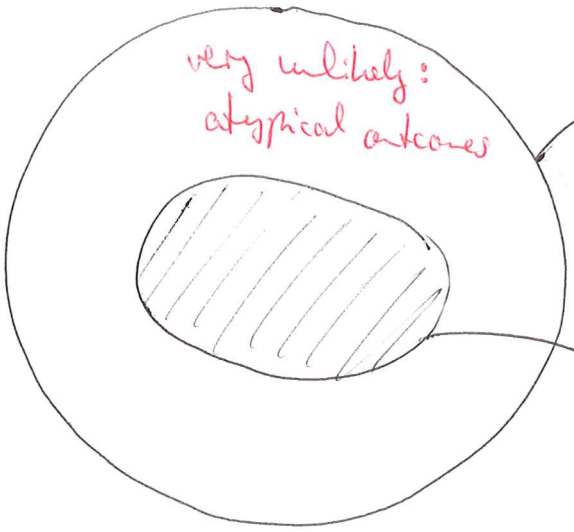
2.11. Typical subsets and data compression

Literature: Cover, Thomas, Elements of Information Theory, Wiley 2006

Consider independent, identically distributed ("i.i.d.") random variables $X_1, \dots, X_n$, taking values in an alphabet $\mathcal{X}$.

Ex.: n coin tosses. $\mathcal{X} = \{0, 1\}$ (0 ↑ heads, 1 ↑ tails), $p(X_1, \dots, X_n) = p(X_1) \cdot \dots \cdot p(X_n)$.

(earlier notation: $p \Rightarrow (p(0), p(1))$; $p(x_1, \dots, x_n) \Rightarrow p^{\otimes n}(x_1, \dots, x_n)$)



all $|\mathcal{X}|^n$ possible outcome sequences

"typical" set $A_\varepsilon^{(n)}$:
$2^{n(H+\varepsilon)}$ elements,
$H = H(p)$ Shannon entropy
and $p(A_\varepsilon^{(n)}) \geq 1-\varepsilon$
small, but highly probable.

leave on blackboard!

Def. Given a sequence of rndm variables $X_1, X_2, \dots$ we say that the sequence converges to X *in probability* if $\forall \varepsilon > 0$, $\Pr\{|X_n - X| > \varepsilon\} \xrightarrow{n\to\infty} 0$.

Theorem: If $X_1, X_2, \dots$ are i.i.d. $\sim p(x)$, then

$$-\frac{1}{n} \log p(X_1, X_2, \dots, X_n) \to H(X) \text{ in probability}$$

Proof: By the weak law of large numbers,

$$-\frac{1}{n} \log p(X_1, \dots, X_n) = -\frac{1}{n} \sum_{i=1}^{n} \underbrace{\log p(X_i)}_{\text{also i.i.d.}}$$

$$\longrightarrow -\mathbb{E} \log p(X) \text{ in probability}$$

$$= -\sum_{x \in \mathcal{X}} p(x) \log p(x) = H(X).$$

□

"Almost all outcomes are almost equally surprising"

E.g. biased coin, $p = (\frac{2}{3}, \frac{1}{3}) \Rightarrow H(p) \approx 0.92$ (< 1 bit, $\log = \log_2$)

$\Rightarrow$ If n large, then with high probability

$$p(x_1, \dots, x_n) \approx 2^{-n \cdot H(p)} \approx 2^{0.92 \cdot n}$$

Def.: The typical (sub)set $A_\varepsilon^{(n)}$ with respect to $p(x)$ is the set of sequences $(x_1, \dots, x_n) \in \mathcal{X}^n$ with

$$2^{-n(H(X)+\varepsilon)} \leq p(x_1, \dots, x_n) \leq 2^{-n(H(X)-\varepsilon)}.$$

Theorem:

(i) $(x_1, \dots, x_n) \in A_\varepsilon^{(n)} \iff H(X) - \varepsilon \leq -\frac{1}{n} \log p(x_1, \dots, x_n) \leq H(X) + \varepsilon$

(ii) $\lim_{n \to \infty} \Pr\{A_\varepsilon^{(n)}\} = 1$

(iii) $|A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$

(iv) $\forall \delta > 0$: $|A_\varepsilon^{(n)}| \geq (1-\delta) 2^{n(H(X)-\varepsilon)}$ if n is large enough.

Proof: (i) by definition, (ii) by $-\frac{1}{n} \log p(x_1, \dots, x_n) \to H(X)$ in probability

(iii) $1 \geq \Pr(A_\varepsilon^{(n)}) = \sum_{x \in A_\varepsilon^{(n)}} \underbrace{p(x)}_{\geq 2^{-n(H(X)+\varepsilon)}} \geq |A_\varepsilon^{(n)}| \cdot 2^{-n(H(X)+\varepsilon)}$

(iv) $|A_\varepsilon^{(n)}| \cdot 2^{-n(H(X)-\varepsilon)} \geq \sum_{x \in A_\varepsilon^{(n)}} p(x)$

$= \Pr\left(A_\varepsilon^{(n)}\right) \geq 1-\delta$ if n is large enough

$\Rightarrow |A_\varepsilon^{(n)}| \geq 2^{n(H(X)-\varepsilon)} (1-\delta)$. □

Consequence: data compression

- encode every $x \in A_\varepsilon^{(n)}$ into a (binary) codeword

$C_x = 0 \underbrace{\cdots\cdots}_{\lceil n(H(X)+\varepsilon)\rceil}$

$\lceil n(H(X)+\varepsilon)\rceil$ bits, giving index of x in all of $A_\varepsilon^{(n)}$

and every $x \in \mathcal{X}^n \setminus A_\varepsilon^{(n)}$ as

$C_x = 1 \underbrace{\cdots\cdots}_{\lceil n \log_2 |\mathcal{X}|\rceil}$

$\lceil n \log_2 |\mathcal{X}| \rceil$ bits, encoding x

Expected codeword length:

$$\mathbb{E}(\ell(C_x)) = \sum_{x \in A_\varepsilon^{(n)}} p(x)\,\ell(C_x) + \sum_{x \notin A_\varepsilon^{(n)}} p(x)\,\ell(C_x)$$

$$\leq \sum_{x \in A_\varepsilon^{(n)}} p(x)\left[n(H(X)+\varepsilon)+2\right]$$

$$+ \sum_{x \notin A_\varepsilon^{(n)}} p(x)\left[n \log_2 |\mathcal{X}| + 2\right]$$

$\Pr\left(|A_\varepsilon^{(n)}| \le 1\right)$ and $\Pr\left((A_\varepsilon^{(n)})^c\right) \le \delta$ if n large enough

$$\le n(H(X)+\varepsilon) + 2 + \delta n \log_2 |\mathcal{X}| + 2\delta$$

$$\Rightarrow \mathbb{E}\left(\frac{1}{n}\ell(C_x)\right) \le H(X) + \varepsilon + \delta \log_2 |\mathcal{X}|$$

$$= H(X) + \varepsilon' \text{ for } n \text{ large enough}$$

and ε' can be made as small as one wishes.

Easy to show: no better compression rate is possible unless error probability $\to 1$. Best rate of compression: $H(X)$.

Sketch: How to get the AEP from this

We can take $\varepsilon_n \to 0$ slowly with n such that for $A^{(n)} := A_\varepsilon^{(n)}$

$$p^{\otimes n}(A^{(n)}) \ge 1-\varepsilon_n$$

$$(1-\varepsilon_n)\, 2^{n(H-\varepsilon_n)} \le |A^{(n)}| \le 2^{n(H+\varepsilon_n)}$$

Set $q^{(n)}(x) := \begin{cases} p^{\otimes n}(x) / p^{\otimes n}(A^{(n)}) & \text{if } x \in A^{(n)} \\ 0 & \text{otherwise} \end{cases}$

$$\Rightarrow D(p^{\otimes n}, q^{(n)}) = \frac{1}{2}\left(1 - p^{\otimes n}(A^{(n)})\right) \le \frac{\varepsilon_n}{2} \xrightarrow{n\to\infty} 0$$

n large enough $\Rightarrow H_0^\varepsilon(p^{\otimes n}) \le H_0(q^{(n)}) = \log |A^{(n)}|$

$$\le n(H + \varepsilon_n).$$

$$\Rightarrow \limsup_{n\to\infty} \frac{1}{n} H_0^{\varepsilon}(p^{\otimes n}) \leq \limsup_{n\to\infty} \left(H(\quad) + \varepsilon_n\right) = H.$$

For the converse: show that any distr. $q^{(n)}$ supported on a subset $\mathcal{Q}^{(n)}$ of size $|\mathcal{Q}^{(n)}| \leq 2^{n(H(p)-\delta)}$ has $D(q^{(n)}, p^{\otimes n}) > \varepsilon$ if n is large enough.

Problem 22: discuss briefly that lowering of 2nd energy level is **not** counted.

- raise only [redacted] invest $I_\infty(p)$ and get p with 100% security
- then lower 2nd level:

$$\Rightarrow \langle W \rangle_{\text{total}} = I(p)$$

but only on **average**!