3장. AEP

* Random Variable의 수렴.

- Law of large numbers
  - $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{n\to\infty} E[X]$
  - X의 산술평균은 Expectation 으로 수렴.
- 수렴의 종류 $(Y_n \to Y)$

① Almost sure convergence 강한 수렴

- $P(\{w : \lim_{n\to\infty} Y_n(w) = Y(w)\}) = 1$

② Mean square convergence 약한 수렴

- $\lim_{n\to\infty} E[|Y_n - Y|^2] = 0$

√ ③ Convergence in probability 주로 사용

- $\lim_{n\to\infty} P(\{w : |Y_n(w) - Y(w)| > \varepsilon\}) = 0$
- 차이가 ε보다 클 확률이 0. ε 범위 내로 수렴.
- $\forall \varepsilon > 0$ 에 대해 만족.

④ Convergence in distribution (cdf)

- 모든 y에 대해 continuous F,

  $\lim_{n\to\infty} F_n(y) \longrightarrow F_Y(y)$
- Y라는 distribution 으로 수렴.

- Weak law of large numbers
  - $X_1, X_2, \cdots$ iid, $S_n = \frac{X_1 + \cdots + X_n}{n}$ 에 대해 (X의 mean μ, variance σ^2)
  - $E[S_n] = \mu$, $Var[S_n] = 0$ in $n \to \infty$

* Asymptotic Equipartition Property

— n개의 iid rvs $X_1, \cdots, X_n \sim p(X_1, X_2, \cdots) = \prod_i p(X_i) = p(X^n)$

- n이 충분히 크면, $P(X_i)$가 의미 없이, 특정 sequence 나올 확률이 동일해진다 ㅎㅎㅎ

- $\underbrace{\frac{1}{n} \log \frac{1}{p(X_1, X_2, \cdots, X_n)}}_{}$ 는 $H(X)$로 수렴. $n \to \infty$면 : sequence의 정보량이 비슷해짐.

sequence

— $X_1, X_2, \cdots X_n$은 iid with $p(x)$.

- $-\frac{1}{n} \log p(X_1, X_2, \cdots X_n) \longrightarrow H(X)$ in probability.

pf) X_i가 iid이므로 $\log(X_i)$도 각 independent.

$$
\begin{aligned}
-\frac{1}{n} \log p(X_1, X_2, \cdots, X_n) &= -\frac{1}{n} \sum_{i=1}^{n} \log p(X_i) \\
&= -\sum_{i=1}^{n} \frac{\log p(X_i)}{n} \\
&\to -E[\log p(X)] \quad \text{weak of large numbers.} \\
&= H(X) \ \square
\end{aligned}
$$

- 의미 : $n \to \infty$면 모든 sequence의 정보량은 상수($H(X)$)로 수렴. 똑같은 정보량을 가짐

= Equipartition.

* Typical Set

— $p(x)$에 대해 $A_\varepsilon^{(n)}$은 다음의 조건을 만족하는 $(x_1, \cdots, x_n)$ set.

- $2^{-n(H(X)+\varepsilon)} \le p(x_1, x_2, \cdots, x_n) \le 2^{-n(H(X)-\varepsilon)}$

— 특징 (임의의 ε에 대해)

1. $(x_1, x_2, \cdots, x_n) \in A_\varepsilon^{(n)}$ 이면, $H(X)-\varepsilon \le -\frac{1}{n}\log p(x_1, \cdots, x_n) \le H(X)+\varepsilon$.
2. $\Pr\{A_\varepsilon^{(n)}\} > 1-\varepsilon$: n이 충분히 크면, typical set 확률이 대부분임.
3. 개수 : $|A_\varepsilon^{(n)}| \le 2^{n(H(X)+\varepsilon)}$: 크기는 전체에 비해 꽤 작음.
4. : $|A_\varepsilon^{(n)}| \ge (1-\varepsilon) 2^{n(H(X)-\varepsilon)}$

by ②, $\Pr\{|-\frac{1}{n}\log(x_1, \cdots, x_n) - H(X)| < \varepsilon\} > 1-\delta$. 임의의 δ에 대해 n이 존재.

③ proof : $1 = \sum_x p(x) \ge \sum_{x \in A_\varepsilon^{(n)}} p(x) \ge \sum_{x \in A_\varepsilon^{(n)}} 2^{-n(H(X)+\varepsilon)} = 2^{-n(H(X)+\varepsilon)} \cdot |A_\varepsilon^{(n)}|$ □.

④ ∵ ②에 의해 n이 크면 $Pr\{A_\varepsilon^{(n)}\} > 1-\varepsilon$

∴ $1-\varepsilon < Pr\{A_\varepsilon^{(n)}\} = \sum_{A_\varepsilon} p(x)$

$\le \sum 2^{-n(H(X)-\varepsilon)} = 2^{-n(H(X)-\varepsilon)} \cdot |A_\varepsilon^{(n)}|$

∴ $|A_\varepsilon^{(n)}| \ge (1-\varepsilon)\, 2^{n(H(X)-\varepsilon)}$

⇒ typical set의 size는 $2^{n(H(X))}$의 $\pm\varepsilon$ 정도.

ex) Binary RV : $p(0)=p$, $p(1)=1-p$ 에서

- typical : np개의 0, $n(1-p)$개의 1로 구성
- typical set 확률 = $p^{np} \cdot (1-p)^{n(1-p)}$

$= 2^{n(p\log p + (1-p)\log(1-p))} = 2^{-nH(X)}$

- typical set 개수 = $\binom{n}{np} = \frac{n!}{np!\,(n-np)!}$

$n! \approx n^n e^{-n}\sqrt{2\pi n}$ 을 이용.

$$\frac{n!}{np!\; n(1-p)!} \approx \frac{n^n e^{-n}}{(np)^{np} e^{-np} \cdot (n-np)^{(n-np)} \cdot e^{-(n-np)}}$$

$$= \frac{1}{p^{np}(1-p)^{n(1-p)}} = 2^{nH(X)}$$

Binary RV에 대해

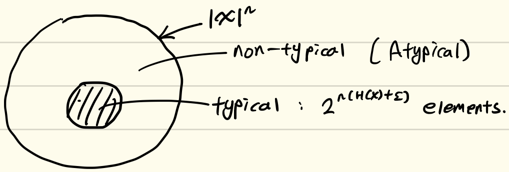
⇒ typ 확률 $= 2^{-nH(X)}$

typ 개수 $= 2^{n(H(X))}$

* Data Compression by AEP

— 전체 vs. typical set

•



- 경우의 수 : non-typical >> typical
- 확률 : non-typical < typical

— 압축

- $x \to$ encoder $\to C(x)$. 압축 performance는 $C(x)$의 평균길이.
- Source code

typical set : $n(H(X)+\varepsilon)+1$ bits 필요.

Atypical set : $n\log|\mathcal{X}|+1$ bits 필요

$+1$: typ, atyp 구분 bit.

0 1

— Expect length : $l(x^n)$

· $E[l(x^n)] = \sum_{x^n} p(x^n)\, l(x^n)$

압축 performance.

각 codeword 길이
- typ : $n(H(X)+\varepsilon)+1 \quad +1$
- atyp : $n(\log|\mathcal{X}|)+1 \quad +1$

· $$E[l(X^n)] = \sum_{x^n} p(x^n)\, l(x^n) = \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n)\, l(x^n) + \sum_{A_\varepsilon^c} p(x^n)\, l(x^n)$$

$$\le \sum_{typ} p(x^n)\,(n(H+\varepsilon)+2) + \sum_{Atyp} p(x^n)\,(n\log|\mathcal{X}|+2)$$

$$= \underbrace{Pr\{A_\varepsilon^{(n)}\}}_{\le 1}(n(H+\varepsilon)+2) + \underbrace{Pr\{A_\varepsilon^{(n)c}\}}_{\le \varepsilon}(n\log|\mathcal{X}|+2)$$

↳ $Pr\{A_\varepsilon^{(n)}\} \ge 1-\varepsilon$

$$\le n(H+\varepsilon)+2+\underbrace{\varepsilon n\log|\mathcal{X}|+2\varepsilon}_{\varepsilon\text{는 매우 작음.}} = n(H+\varepsilon')$$

— 그러므로 $\varepsilon > 0$ 에 대해 n이 충분히 크면

· $E\left[\frac{1}{n} l(X^n)\right] \le H(X)+\varepsilon$ 압축 한계.

— typical set : 확률 합이 $\ge 1-\varepsilon$가 되는 최소 set.

· $Pr(A_\varepsilon^{(n)}) \ge 1-\varepsilon$

· $|A_\varepsilon^{(n)}| \le 2^{n(H(X)+\varepsilon)}$