Introducción a la Teoría de la Información

Asymptotic Equipartition Property.

Facultad de Ingeniería, UdelaR

2025

Equipartición asintótica

- Consideramos $X_1 \ldots X_n$ variables aleatorias independientes e identicamente distribuidas (*i.i.d.*), $X_i \sim p(x)$, $x \in \mathcal{X}$.
- Definimos $p(x_1 \ldots x_n) = \prod_{i=1}^{n} p(x_i)\,.$
- "Casi toda la probabilidad se concentra en eventos que son casi equiprobables"

$$P\Big\{(x_1 \ldots x_n) : p(x_1 \ldots x_n) = 2^{-n(H \pm \epsilon)}\Big\} \approx 1\,.$$

AEP

Teorema

Sean $X_1 \ldots X_n$ variables aleatorias i.i.d., $X_i \sim p(x)$, cada una de ellas con entropía $H = H(X_i)$. Se cumple

$$-\frac{1}{n} \log p(X_1 \ldots X_n) \to H \quad \text{en probabilidad.}$$

Demostración.

$$\begin{aligned}
-\frac{1}{n} \log p(X_1 \ldots X_n) &= -\frac{1}{n} \sum_{i=1}^{n} \log p(X_i) && \text{(independencia)} \\
&\to -E\left[\log p(X)\right] && \text{(Ley G. Números)} \\
&= H
\end{aligned}$$

$\square$

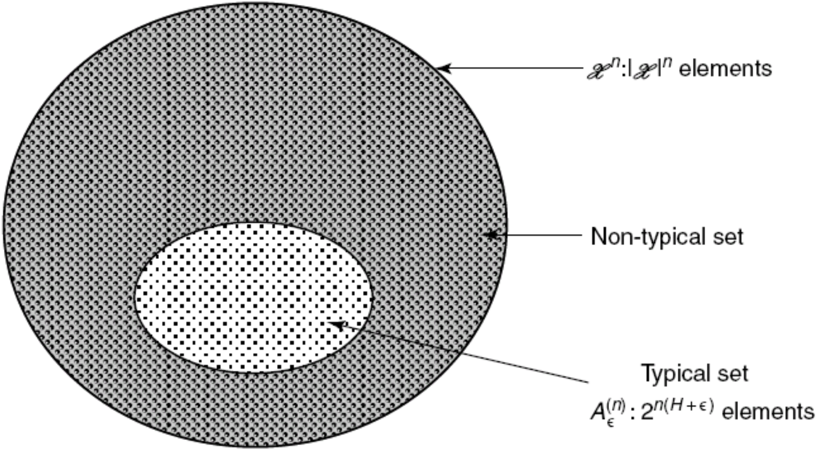
Conjunto Típico de Secuencias

Definición

Sean $X_1 \ldots X_n$ variables aleatorias i.i.d., $X_i \sim p(x)$, $H = H(X_i)$. El *conjunto típico* $A_\epsilon^{(n)}$ con respecto a $p(x)$ es el conjunto de secuencias que satisfacen

$$2^{-n(H+\epsilon)} \leq p(x_1 \ldots x_n) \leq 2^{-n(H-\epsilon)} \, .$$

Conjunto Típico de Secuencias



Propiedades del Conjunto Típico

Teorema

Sean $X_1 \dots X_n$ variables aleatorias i.i.d., $X_i \sim p(x)$, $H = H(X_i)$. El conjunto típico $A_\epsilon^{(n)}$ satisface

1. $(x_1 \dots x_n) \in A_\epsilon^{(n)} \Leftrightarrow H - \epsilon \leq -\frac{1}{n} \log p(x_1 \dots x_n) \leq H + \epsilon$
2. $P\{A_\epsilon^{(n)}\} > 1 - \epsilon$ para n suficientemente grande.
3. $|A_\epsilon^{(n)}| \leq 2^{n(H+\epsilon)}$
4. $|A_\epsilon^{(n)}| \geq (1-\epsilon) 2^{n(H-\epsilon)}$ para n suficientemente grande.

Propiedades del Conjunto Típico

Demostración.

1. $H - \epsilon \leq -\frac{1}{n} \log p(x_1 \dots x_n) \leq H + \epsilon$ es la def. de $A_\epsilon^{(n)}$
2. Como $-\frac{1}{n} \log p(X_1 \dots X_n) \to H$, para todo $\delta > 0$ existe N_0 tal que

$$P\left\{ \left| -\frac{1}{n} \log p(X_1 \dots X_n) - H \right| < \epsilon \right\} > 1 - \delta\,, \quad n > N_0\,.$$

   Tomando $\delta = \epsilon$ obtenemos $P\{A_\epsilon^{(n)}\} > 1 - \epsilon$ para $n > N_0$.

3. $1 \geq \sum_{\mathbf{x} \in A_\epsilon^{(n)}} p(\mathbf{x}) \geq \sum_{\mathbf{x} \in A_\epsilon^{(n)}} 2^{-n(H+\epsilon)} = |A_\epsilon^{(n)}| 2^{-n(H+\epsilon)} \Rightarrow |A_\epsilon^{(n)}| \leq 2^{n(H+\epsilon)}$.
4. $1 - \epsilon \leq \sum_{\mathbf{x} \in A_\epsilon^{(n)}} p(\mathbf{x}) \leq \sum_{\mathbf{x} \in A_\epsilon^{(n)}} 2^{-n(H-\epsilon)} = |A_\epsilon^{(n)}| 2^{-n(H-\epsilon)}$
   $\Rightarrow |A_\epsilon^{(n)}| \geq (1 - \epsilon) 2^{n(H-\epsilon)}$.

$\square$

Propiedades del Conjunto Típico

Demostración.

1. $H - \epsilon \leq -\frac{1}{n}\log p(x_1 \ldots x_n) \leq H + \epsilon$ es la def. de $A_\epsilon^{(n)}$
2. Como $-\frac{1}{n}\log p(X_1 \ldots X_n) \to H$, para todo $\delta > 0$ existe N_0 tal que

$$P\left\{\left|-\frac{1}{n}\log p(X_1 \ldots X_n) - H\right| < \epsilon\right\} > 1 - \delta\,, \quad n > N_0\,.$$

   Tomando $\delta = \epsilon$ obtenemos $P\{A_\epsilon^{(n)}\} > 1 - \epsilon$ para $n > N_0$.
3. $1 \geq \sum_{\mathbf{x}\in A_\epsilon^{(n)}} p(\mathbf{x}) \geq \sum_{\mathbf{x}\in A_\epsilon^{(n)}} 2^{-n(H+\epsilon)} = |A_\epsilon^{(n)}|2^{-n(H+\epsilon)} \Rightarrow |A_\epsilon^{(n)}| \leq 2^{n(H+\epsilon)}$.
4. $1 - \epsilon \leq \sum_{\mathbf{x}\in A_\epsilon^{(n)}} p(\mathbf{x}) \leq \sum_{\mathbf{x}\in A_\epsilon^{(n)}} 2^{-n(H-\epsilon)} = |A_\epsilon^{(n)}|2^{-n(H-\epsilon)}$
   $\Rightarrow |A_\epsilon^{(n)}| \geq (1-\epsilon)2^{n(H-\epsilon)}$.

$\square$

Propiedades del Conjunto Típico

Demostración.

1. $H - \epsilon \leq -\frac{1}{n}\log p(x_1 \ldots x_n) \leq H + \epsilon$ es la def. de $A_\epsilon^{(n)}$
2. Como $-\frac{1}{n}\log p(X_1 \ldots X_n) \to H$, para todo $\delta > 0$ existe N_0 tal que

$$P\left\{\left|-\frac{1}{n}\log p(X_1 \ldots X_n) - H\right| < \epsilon\right\} > 1 - \delta\,, \quad n > N_0\,.$$

   Tomando $\delta = \epsilon$ obtenemos $P\{A_\epsilon^{(n)}\} > 1 - \epsilon$ para $n > N_0$.
3. $1 \geq \sum_{\mathbf{x}\in A_\epsilon^{(n)}} p(\mathbf{x}) \geq \sum_{\mathbf{x}\in A_\epsilon^{(n)}} 2^{-n(H+\epsilon)} = |A_\epsilon^{(n)}|2^{-n(H+\epsilon)} \Rightarrow |A_\epsilon^{(n)}| \leq 2^{n(H+\epsilon)}$.
4. $1 - \epsilon \leq \sum_{\mathbf{x}\in A_\epsilon^{(n)}} p(\mathbf{x}) \leq \sum_{\mathbf{x}\in A_\epsilon^{(n)}} 2^{-n(H-\epsilon)} = |A_\epsilon^{(n)}|2^{-n(H-\epsilon)}$
   $\Rightarrow |A_\epsilon^{(n)}| \geq (1-\epsilon)2^{n(H-\epsilon)}$.

$\square$

Propiedades del Conjunto Típico

Demostración.

1. $H - \epsilon \leq -\frac{1}{n} \log p(x_1 \ldots x_n) \leq H + \epsilon$ es la def. de $A_\epsilon^{(n)}$
2. Como $-\frac{1}{n} \log p(X_1 \ldots X_n) \to H$, para todo $\delta > 0$ existe N_0 tal que

$$P\left\{\left|-\frac{1}{n} \log p(X_1 \ldots X_n) - H\right| < \epsilon\right\} > 1 - \delta\,, \quad n > N_0\,.$$

   Tomando $\delta = \epsilon$ obtenemos $P\{A_\epsilon^{(n)}\} > 1 - \epsilon$ para $n > N_0$.
3. $1 \geq \sum_{\mathbf{x} \in A_\epsilon^{(n)}} p(\mathbf{x}) \geq \sum_{\mathbf{x} \in A_\epsilon^{(n)}} 2^{-n(H+\epsilon)} = |A_\epsilon^{(n)}| 2^{-n(H+\epsilon)} \Rightarrow |A_\epsilon^{(n)}| \leq 2^{n(H+\epsilon)}$.
4. $1 - \epsilon \leq \sum_{\mathbf{x} \in A_\epsilon^{(n)}} p(\mathbf{x}) \leq \sum_{\mathbf{x} \in A_\epsilon^{(n)}} 2^{-n(H-\epsilon)} = |A_\epsilon^{(n)}| 2^{-n(H-\epsilon)}$
   $\Rightarrow |A_\epsilon^{(n)}| \geq (1 - \epsilon) 2^{n(H-\epsilon)}$.

$\square$

Ejemplo

k	$\binom{n}{k}$	$\binom{n}{k} p^k (1-p)^{n-k}$	$-\frac{1}{n}\log p(x^n)$
0	1	1.13E-10	1.321928095
1	25	4.22E-09	1.298529595
2	300	7.60E-08	1.275131095
3	2300	8.74E-07	1.251732595
4	12650	7.21E-06	1.228334095
5	53130	4.54E-05	1.204935595
6	177100	0.000227126	1.181537095
7	480700	0.000924725	1.158138595
8	1081575	0.003120948	1.134740095
9	2042975	0.008842685	1.111341595
10	3268760	0.021222445	1.087943095
11	4457400	0.043409546	1.064544595
12	5200300	0.075966705	1.041146095
13	5200300	0.113950058	1.017747595
14	4457400	0.146507217	0.994349094
15	3268760	0.161157939	0.970950594
16	2042975	0.151085568	0.947552094
17	1081575	0.119979715	0.924153594
18	480700	0.079986477	0.900755094
19	177100	0.044203053	0.877356594
20	53130	0.019891374	0.853958094
21	12650	0.007104062	0.830559594
22	2300	0.001937471	0.807161094
23	300	0.000379071	0.783762594
24	25	4.74E-05	0.760364094
25	1	2.84E-06	0.736965594

$$\begin{aligned} n &= 25 \\ p(1) &= 0{,}6 \\ \epsilon &= 0{,}1 \end{aligned}$$

$$\begin{aligned} h(p) &= 0{,}97095 \\ h(p)-\epsilon &= 0{,}87095 \\ h(p)+\epsilon &= 1{,}07095 \end{aligned}$$

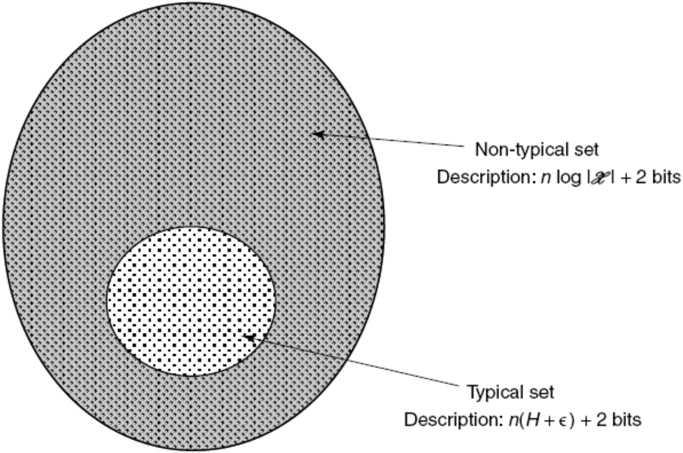
$$\begin{aligned} P\{A_\epsilon^{(n)}\} &= 0{,}9362 \\ |A_\epsilon^{(n)}| &= 26{,}366{,}510 \\ 2^{25} &= 33{,}554{,}432 \end{aligned}$$

Aplicación a Compresión

Denotamos $\mathcal{X}^n$ al conjunto de secuencias de largo n y $x^n = (x_1 \ldots x_n)$.

- Si a cada secuencia de $\mathcal{X}^n$ le asignamos un número en el rango $0 \ldots 2^{n \log |\mathcal{X}|} - 1$, cada secuencia se puede representar como una secuencia binaria de $\lceil n \log |\mathcal{X}| \rceil$ bits.
- Si nos restringimos a $A_\epsilon^{(n)}$, que tiene probabilidad mayor o igual a $1 - \epsilon$, alcanza con $\lceil n(H + \epsilon) \rceil$ bits.

Aplicación a Compresión



Aplicación a Compresión

Teorema

Sean $X_1 \ldots X_n$ variables aleatorias i.i.d., $X_i \sim p(x)$, $H = H(X_i)$. Para todo $\delta > 0$ y todo natural $n > N(\delta)$, existe un código binario para X^n cuyo largo de código, $l(X^n)$, satisface

$$E\left[\frac{1}{n} l(X^n)\right] \leq H + \delta\,.$$

Aplicación a Compresión

Demostración.

Sea $\epsilon = \frac{\delta}{2(1+\log|\mathcal{X}|)}$. Para el código por secuencias típicas, tenemos

$$
\begin{aligned}
E\left[l(X^n)\right] &\leq P\{A_\epsilon^{(n)}\}[n(H+\epsilon)+2] + (1-P\{A_\epsilon^{(n)}\})[n\log|\mathcal{X}|+2] \\
&\leq n(H+\epsilon)+2+\epsilon\Big(n\log|\mathcal{X}|+2\Big), \quad n > N_0(\epsilon) \\
&= n\left(H+\epsilon(1+\log|\mathcal{X}|)+\frac{2(1+\epsilon)}{n}\right) \\
&= n\left(H+\frac{\delta}{2}+\frac{2(1+\epsilon)}{n}\right) \\
&\leq n\left(H+\frac{\delta}{2}+\frac{\delta}{2}\right), \quad n > \text{máx}\left\{N_0(\epsilon), \frac{4(1+\epsilon)}{\delta}\right\} \\
&= n(H+\delta)
\end{aligned}
$$

$\square$

$A_\epsilon^{(n)}$ vs. Otros Conjuntos de Alta Probabilidad

Teorema

Sean $X_1 \ldots X_n$ variables aleatorias i.i.d. con entropía $H = H(X_i)$. Sea también, para cada natural n, $B_\delta^{(n)}$ un subconjunto de $\mathcal{X}^n$ que satisface

$$P\{B_\delta^{(n)}\} \geq 1 - \delta\,, \quad \delta < 1\,.$$

Entonces, para todo $\delta' > 0$ se cumple

$$\frac{1}{n} \log |B_\delta^{(n)}| > H - \delta'\,,$$

si n es suficientemente grande.

$A_\epsilon^{(n)}$ vs. Otros Conjuntos de Alta Probabilidad

Demostración.

Sea $\epsilon = \text{mín}\left\{\frac{\delta'}{2}, \frac{1-\delta}{2}\right\}$.

$$\begin{aligned}
B_\delta^{(n)} \cap A_\epsilon^{(n)} &= \left(B_\delta^{(n)\,c} \cup A_\epsilon^{(n)\,c}\right)^c \\
P\{B_\delta^{(n)} \cap A_\epsilon^{(n)}\} &= 1 - P\left\{B_\delta^{(n)\,c} \cup A_\epsilon^{(n)\,c}\right\} \\
&\geq 1 - P\{B_\delta^{(n)\,c}\} - P\{A_\epsilon^{(n)\,c}\} \\
\textstyle\sum_{B_\delta^{(n)} \cap A_\epsilon^{(n)}} 2^{-n(H-\epsilon)} &\geq 1 - \delta - \epsilon\,, \quad n > N(\epsilon) \\
|B_\delta^{(n)}| 2^{-n(H-\epsilon)} &\geq 1 - \delta - \epsilon \\
|B_\delta^{(n)}| &\geq (1 - \delta - \epsilon) 2^{n(H-\epsilon)} \\
\tfrac{1}{n} \log |B_\delta^{(n)}| &\geq \tfrac{1}{n} \log(1 - \delta - \epsilon) + H - \epsilon\,, \quad \epsilon < 1 - \delta \\
&\geq \tfrac{1}{n} \log(\tfrac{1-\delta}{2}) + H - \tfrac{\delta'}{2}\,, \quad \epsilon \leq \tfrac{1-\delta}{2}, \epsilon \leq \tfrac{\delta'}{2}\,, n > N(\epsilon)\,. \\
&> -\tfrac{\delta'}{2} + H - \tfrac{\delta'}{2}\,, \quad n > \text{máx}\{N(\epsilon), N'\}\,.
\end{aligned}$$

□

Ejemplo

k	$\binom{n}{k}$	$\binom{n}{k} p^k (1-p)^{n-k}$	$-\frac{1}{n} \log p(x^n)$
0	1	1.13E-10	1.321928095
1	25	4.22E-09	1.298529595
2	300	7.60E-08	1.275131095
3	2300	8.74E-07	1.251732595
4	12650	7.21E-06	1.228334095
5	53130	4.54E-05	1.204935595
6	177100	0.000227126	1.181537095
7	480700	0.000924725	1.158138595
8	1081575	0.003120948	1.134740095
9	2042975	0.008842685	1.111341595
10	3268760	0.021222445	1.087943095
11	4457400	0.043409546	1.064544595
12	5200300	0.075966705	1.041146095
13	5200300	0.113950058	1.017747595
14	4457400	0.146507217	0.994349094
15	3268760	0.161157939	0.970950594
16	2042975	0.151085568	0.947552094
17	1081575	0.119979715	0.924153594
18	480700	0.079986477	0.900755094
19	177100	0.044203053	0.877356594
20	53130	0.019891374	0.853958094
21	12650	0.007104062	0.830559594
22	2300	0.001937471	0.807161094
23	300	0.000379071	0.783762594
24	25	4.74E-05	0.760364094
25	1	2.84E-06	0.736965594

$B_\epsilon^{(n)}$ menor conjunto con $P\{B_\epsilon^{(n)}\} \geq 1 - \epsilon = 0{,}9$

$$\begin{aligned} P\{\#1 \geq 12\} &= 0{,}9222 \\ P\{\#1 \geq 13\} &= 0{,}84623 \end{aligned}$$

$$\begin{aligned} P\{A_\epsilon^{(n)}\} &= 0{,}9362 \\ |A_\epsilon^{(n)}| &= 26{,}366{,}510 \\ |B_\epsilon^{(n)}| &= 20{,}457{,}889 \end{aligned}$$

AEP para Procesos Estocásticos

Teorema

Si $\{X_i\}$ es un proceso estacionario y ergódico (por ejemplo Markov irreducible y aperiódico estacionario)

$$-\frac{1}{n}\log p(X_1 \dots X_n) \to H(\mathcal{X})$$

con probabilidad 1.