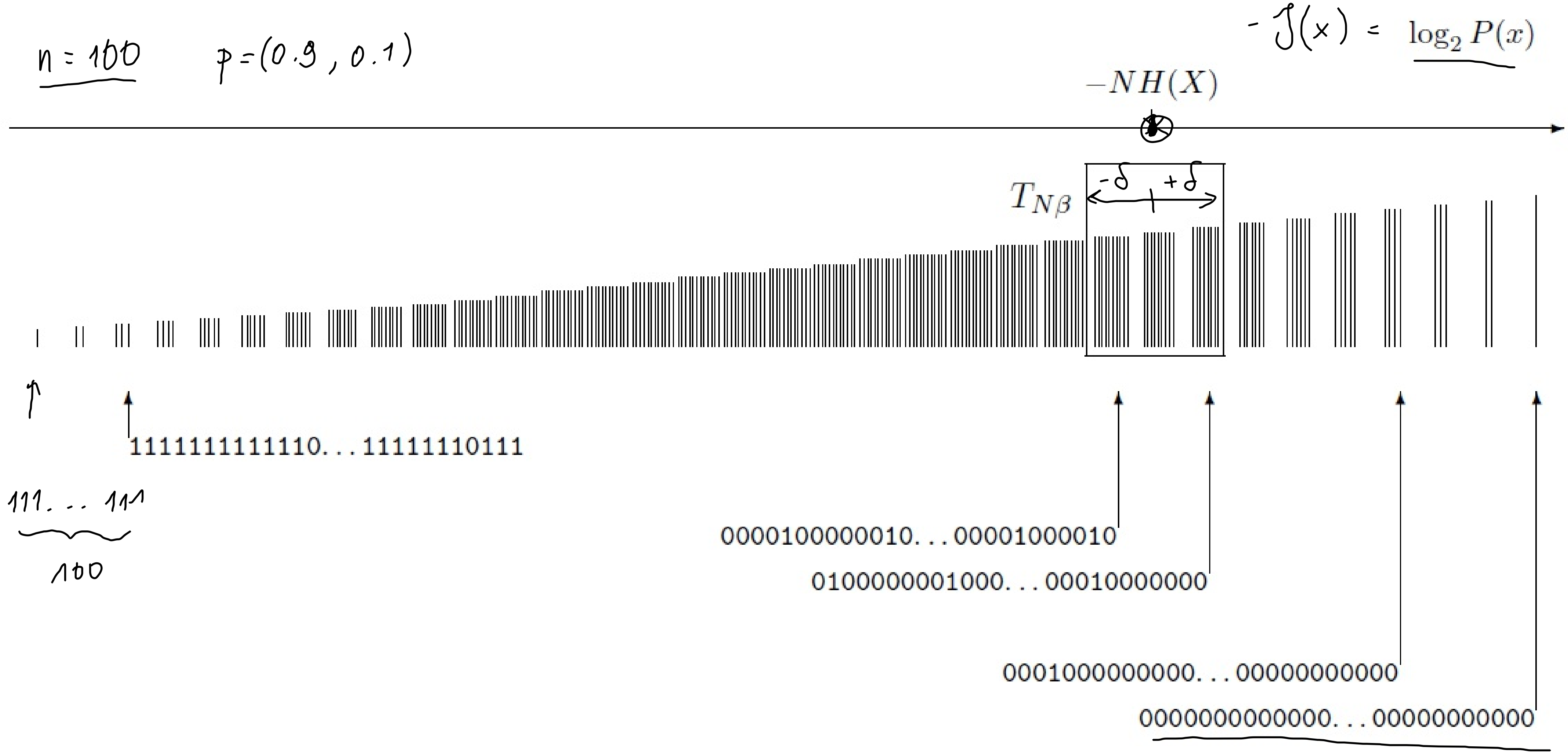
$\mathcal{X} = \{0, 1\}$

$p_X = \underline{(1/2, 1/2)}$ $\qquad q_X = (3/4, 1/4)$

Sequenze di n simboli ; $\underline{n = 3}$ $\qquad \mathcal{X}^n$

X^n	$p(X^n)$	$q(X^n)$		$I(X^n) = -\log_2 q(X^n)$
000	1/8	27/64		$\log_2 64/27$
001	1/8	9/64	54/64	$\log_2 64/9$
010	1/8	9/64	27/64	
100	1/8	9/64		⋮
011	1/8	3/64		⋮
101	1/8	3/64		⋮
110	1/8	3/64		⋮
111	1/8	1/64	64/64	$\log_2 64$



n = 100
p = (0.9, 0.1)
$-J(x) = \log_2 P(x)$
$-NH(X)$
$T_{N\beta}$
$-\delta$
$+\delta$
1111111111110...11111110111
111...111
100
0000100000010...00001000010
0100000001000...00010000000
0001000000000...00000000000
0000000000000...00000000000

Autoinformazione della sequenza X^n : $\mathcal{I}(X^n) = -\log p(X^n)$

Teorema di Chebyshev : Se $\{Y_n\} = (Y_1, Y_2, \dots, Y_k, \dots)$ è una successione di v.a. reali e indipendenti (identicamente distribuite) e la cui varianza $Var[Y_n]$ è limitata da una costante ($\exists C > 0 : Var[Y_n] \leq C$ per ogni n) allora $\{Y_n\}$ soddisfa la legge (debole) dei grandi numeri, ovvero

$$Y_n \xrightarrow{prob.} \mathbb{E}[Y_n] \qquad \left(\lim_{n \to \infty} Pr\left[\left| Y_n - \mathbb{E}[Y_n] \right| < \delta \right] = 1 \quad \forall \delta > 0 \right)$$

$$\mathbb{E}_X[\mathcal{I}(X)] = \mathbb{E}_X[-\log p(X)] = H(X) = -\sum_{i=1}^{K} p_i \log p_i \qquad (\text{dove } p_i = Pr[X = x_i])$$

sorgente stazionaria e senza memoria

Consideriamo $\frac{1}{n}\mathcal{I}(X^n)$. Ho $\boxed{\frac{1}{n}\mathcal{I}(X^n)} = -\frac{1}{n}\log p(X^n) = -\frac{1}{n}\log\left(\prod_{j=1}^{n} p(X_j)\right) =$

Y_n

$X^n = (X_1, X_2, \dots, X_j, \dots, X_n)$

$$= \underbrace{-\frac{1}{n}\sum_{j=1}^{n} \log p(X_j)}.$$

Media campione di $\mathcal{I}(X)$ sugli n simboli

Scelgo $\underbrace{Y_n := \frac{1}{n} \mathcal{I}(X^n)}$. Inoltre, $\mathbb{E}[Y_n] = \mathbb{E}\left[\frac{1}{n}\mathcal{I}(X^n)\right] = \mathbb{E}\left[-\frac{1}{n}\sum_{j=1}^{n} \log p(X_j)\right]$

Autoinformazione normalizzata

$$= \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}\left[-\log p(X_j)\right]$$

le X_j sono identicamente distrib. (stazionarietà)

$$= \frac{1}{n}\sum_{j=1}^{n} \underbrace{\mathbb{E}\left[-\log p(X)\right]}_{H(X)} = \frac{1}{n}\cdot n \cdot H(X) = H(X).$$

Ora invoco il teorema di Chebyshev con $Y_n = \frac{1}{n}\mathcal{I}(X^n)$ (e quindi $\mathbb{E}[Y_n] = H(X) = H(p_X)$)

(lecito perché $0 \le \frac{1}{n}\mathcal{I}(X^n) \le \log \frac{1}{p_{min}}$ dove $p_{min} = \min_{x\in\mathcal{X}} p(x) > 0$ senza di perdita generalità

e quindi $\mathrm{Var}[\frac{1}{n}\mathcal{I}(X^n)]$ è limitata da una costante)

Quindi $Y_n \xrightarrow{prob} \mathbb{E}[Y_n]$

ovvero $\frac{1}{n}\mathcal{I}(X^n) \xrightarrow{prob.} H(p_X)$

In altre parole, per ogni n, $\forall \delta > 0$, $\Pr\left[\boxed{\left|\frac{1}{n}\mathcal{I}(X^n) - H(p_X)\right| < \delta}\right] = 1 - \varepsilon$

le seq. che soddisfano questo sono dette tipiche

dove $\varepsilon = \varepsilon(n,\delta)$ tende a zero per $n \to \infty$.

Fissata una lunghezza n delle sequenze,

fissato $\delta > 0$,

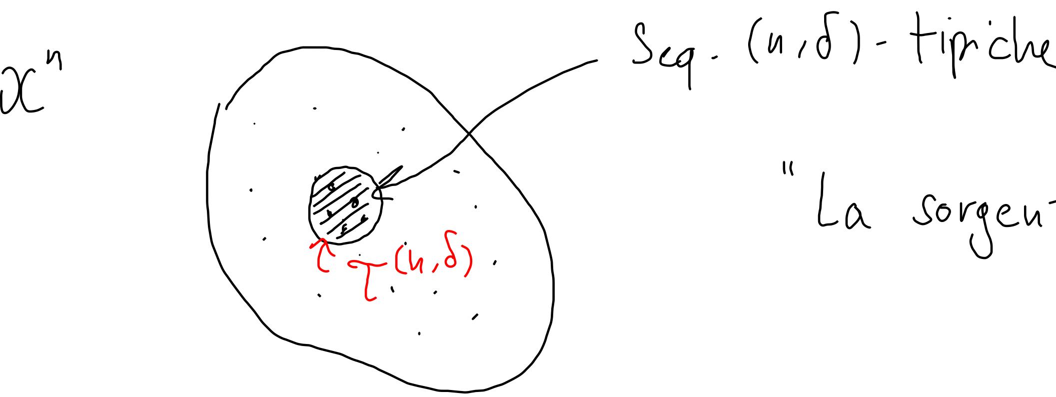
chiamiamo una sequenza $X^n = (X_1, \dots, X_n)$ (n,δ)-tipica se

$$\boxed{\left| \frac{1}{n} \underbrace{\mathcal{I}(X^n)}_{-\log p(X^n)} - H(p_X) \right| \leq \delta}$$

(la sua autoinformazione normalizzata è δ-prossima all'entropia della sorgente)

Per costruzione,

$$\Pr[X^n \text{ sia } (n,\delta)\text{-tipica}] = 1 - \varepsilon \quad \text{con } \varepsilon = \varepsilon(n,\delta) \xrightarrow{n \to \infty} 0$$



L'insieme delle seq. (n,δ)-tipiche è denotato $\mathcal{T}^{(n,\delta)}$

"La sorgente genera quasi esclusivamente sequenze tipiche"

Teorema (3.5 nel libro).

(a) Probabilità di una sequenza tipica: Sia $x \in \mathcal{T}^{(n,\delta)}$. Allora

$\rightarrow$ $2^{-n(H(p_X)+\delta)} \leq p(x) \leq 2^{-n(H(p_X)-\delta)}$ $\qquad (\ p(x) \approx 2^{-nH(p_X)}\ )$

(b) Numerosità delle sequenze tipiche: $(1-\varepsilon)2^{n(H(p_X)-\delta)} \leq |\mathcal{T}^{(n,\delta)}| \leq 2^{n(H(p_X)+\delta)}$

$(\ |\mathcal{T}^{(n,\delta)}| \approx 2^{nH(p_X)}\ )$

Dim. (per brevità, scrivo $H = H(p_X)$)

(a) Dalla definizione di tipicità:

$$-\delta \leq -\tfrac{1}{n}\log p(x) - H \leq \delta$$

$$\delta \geq \tfrac{1}{n}\log p(x) + H \geq -\delta$$

$$\delta - H \geq \tfrac{1}{n}\log p(x) \geq -\delta - H$$

$\alpha \mapsto 2^{\alpha}$

Moltiplico per n

$$n(\delta - H) \geq \log p(x) \geq -n(H+\delta)$$

$$2^{n(\delta-H)} \geq p(x) \geq 2^{-n(H+\delta)} \implies 2^{-n(H+\delta)} \leq p(x) \leq 2^{-n(H-\delta)}$$

Limitaz. superiore

(b) $\underline{1} \geq \sum_{x \in \mathcal{T}^{(n,\delta)}} p(x) \geq$

$$\geq |\mathcal{T}^{(n,\delta)}| \cdot \min_{x \in \mathcal{T}^{(n,\delta)}} p(x) \overset{(a)}{\geq} \underline{|\mathcal{T}^{(n,\delta)}| \cdot 2^{-n(H+\delta)}}$$

Quindi, confrontando primo e ultimo termine,

$$|\mathcal{T}^{(n,\delta)}| \leq \underline{2^{+n(H+\delta)}}$$

Limitaz. inferiore: $|\mathcal{T}^{(n,\delta)}| \cdot \max_{x \in \mathcal{T}^{(n,\delta)}} p(x) \geq \sum_{x \in \mathcal{T}^{(n,\delta)}} p(x) = 1 - \varepsilon$

$$\overset{(a)}{\wedge\!\!|}$$

$$|\mathcal{T}^{(n,\delta)}| \cdot 2^{-n(H-\delta)}$$

Quindi, confrontando, $|\mathcal{T}^{(n,\delta)}| \geq (1-\varepsilon) \cdot 2^{+n(H-\delta)}$. QED.

$\mathcal{X} = \{x_1, \dots x_K\}$

$$K = 2^{\log_2 K}$$

$$\frac{|\mathcal{T}^{(n,\delta)}|}{|\mathcal{X}^n|} \leq \frac{2^{n(H+\delta)}}{K^n} = \frac{2^{n(H+\delta)}}{2^{n \cdot \log K}} =$$

$$= 2^{n[H+\delta-\log K]}$$

$$= 2^{-n[\underbrace{\log K - H}_{>0} - \delta]}$$

Ma l'entropia $H \leq \log K$

(ed è $= \log K$ solo se p_X è uniforme)

Se p_X non è uniforme, ho $H < \log K$

Se δ è tale che $\log K > H + \delta$, allora $\log K - H - \delta > 0$

e quindi $2^{-n(\underbrace{\dots}_{>0})} \xrightarrow[n \to \infty]{} 0$; $\mathcal{T}^{(n,\delta)}$ rappresenta una frazione asintoticamente infinitesimale delle possibili sequenze