Sorgente di informazione $\mathcal{S}$

$$\vec{p} = (p_1, \dots, p_K) \in \mathbb{R}^K \qquad K = |\underbrace{\{x_1, \dots, x_K\}}_{\text{alfabeto } \mathcal{X}}|$$

- Stazionaria
- Senza memoria

Entropia della sorgente: $H(\mathcal{S}) \stackrel{def.}{=} -\sum_{i=1}^{K} p_i \log p_i$

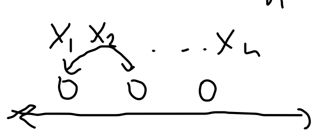
- $Pr[(X_{j+1}, X_{j+2}, \dots, X_{j+n}) = (x_{i_1}, \dots, x_{i_n})] = Pr[(X_{j+r+1}, X_{j+r+2}, \dots, X_{j+r+n}) = (x_{i_1}, \dots, x_{i_n})] \quad \forall j \in \mathbb{Z}, \ \forall r \in \mathbb{Z}$
- $Pr[(X_{j+1}, X_{j+2}, \dots, X_{j+n}) = (x_{i_1}, \dots, x_{i_n})] = \prod_{r=1}^{n} Pr[X_{j+r} = x_{i_r}]$

$$\downarrow$$

$$\prod Pr[X_r = x_{i_r}]$$

Sorgente stazionaria (ma con memoria)

$$H(\mathcal{S}) = \lim_{n \to \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n) \stackrel{\text{Dimostrato nel libro}}{=} \lim_{n \to \infty} H(X_n | X_1, \dots, X_{n-1})$$



n simboli emessi

Nel caso senza memoria:

$$H(X_1, \dots, X_n) = \sum_{r=1}^{n} H(X_r) = -n \sum_{i=1}^{K} p_i \log p_i$$

$$\frac{1}{n} H(\dots) = -\sum_{i=1}^{n} p_i \log p_i$$

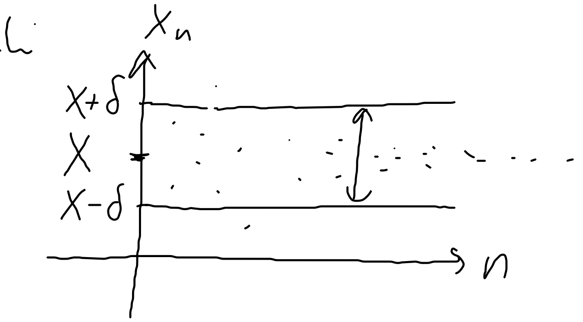
Quali sequenze di simboli sono le più probabili?

Probabilità. Legge dei grandi numeri

Def. Convergenza in probabilità: Sia $\{X_n\}$ ($\in \mathbb{R}$) una successione di v.a. reali

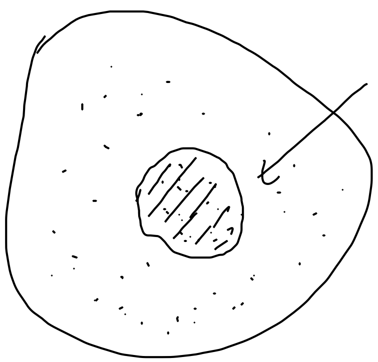
$\{X_n\}$ converge in probabilità alla v.a. X ($\in \mathbb{R}$):

$$\forall \delta > 0 \quad \lim_{n \to \infty} \Pr[\,|X_n - X| < \delta\,] = 1 .$$



(Elevata probabilità che X_n sia nella banda $(X-\delta, X+\delta)$ per ogni n sufficientemente grande)

$\mathcal{K}^n$



Sequenze "<u>tipiche</u>"

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

Ins. di tutte le sequenze di n simboli

Def. La successione $\{X_n\}$ soddisfa la legge debole dei grandi numeri se

$$\lim_{n\to\infty} \Pr\left[\left| \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{Y_n \text{ (media campione)}} - \underbrace{\frac{1}{n}\sum_{i=1}^{n} E[X_i]} \right| < \delta \right] = 1$$

Valore atteso della media campione:

$$E[Y_n] = \frac{1}{n}\sum_i E[X_i]$$

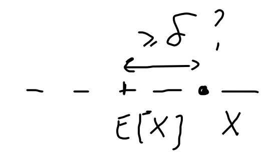
(La successione delle medie di $\{X_n\}$ converge in probabilità al suo valore atteso).

Teorema di Chebyshev: Se $\{X_n\}$ è una successione di v.a. indipendenti, e $\forall n$ $\underset{\text{varianza}}{Var[X_n]} \leq \underset{\text{costante}}{C}$, allora $\{X_n\}$ soddisfa la legge debole dei grandi numeri.

$$Var[X_n] = E[(X_n - E[X_n])^2]$$

(T3.2) Disuguaglianza di Chebyshev. Se la v.a. (reale) X ha varianza $Var[X]$ finita, allora $\forall \delta > 0$

$$Pr[\,|X - E[X]| \geq \delta\,] \leq \frac{Var[X]}{\delta^2}$$



Dim. $Pr[\,|X - E[X]| \geq \delta\,] = \int_{x:|x-EX|\geq\delta} 1 \cdot p(x)\,dx \leq \int_{x:|x-EX|\geq\delta} \underbrace{\frac{(x-EX)^2}{\delta^2}}_{\geq 1} p(x)\,dx \leq \frac{1}{\delta^2} \underbrace{\int_{-\infty}^{+\infty} (x-EX)^2 p(x)\,dx}_{Var[X] = E[(X-EX)^2]}$

$= Var[X] / \delta^2.$

(T3.3)

Dim. del Teorema di Chebyshev. Proprietà della varianza: $Var[\alpha X + \beta] = Var[\alpha X] = \alpha^2 Var[X]$.

Se X_i e X_j sono indipendenti: $Var[X_i + X_j] = Var[X_i] + Var[X_j]$

Consideriamo le v.a. $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$: $Var[Y_n] = Var[\frac{1}{n} \sum_{i=1}^n X_i] = \frac{1}{n^2} \sum_{i=1}^n \underbrace{Var[X_i]}_{\leq C} \leq \frac{Cn}{n^2} = \frac{C}{n}$.

Per Chebyshev applicata a Y_n, $Pr[\,|Y_n - EY_n| < \delta\,] \geq 1 - \frac{Var[Y_n]}{\delta^2} \geq 1 - \frac{C}{n\delta^2} \longrightarrow 1$ per $n \to \infty$. QED

$Pr[\,|Y_n - EY_n| = \delta\,]$

(Staz e assenza di memoria)

Quali sequenze $X^n = (X_1, \dots, X_n)$ di simboli sono più probabili?

Dove si concentra la probabilità $p(X^n)$

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

$p_1 = 0.89$ $\quad p_2 = 0.11$

...001000.1.

1111011110101

Sia $\mathcal{I}(X) = -\log p(X)$ l'autoinformazione di X

(Sappiamo che $E[\mathcal{I}(X)] = E[-\log p(X)] = H(X)$)

Quanto vale $\mathcal{I}(X^n)$?

$$\frac{1}{n}\mathcal{I}(X^n) = -\frac{1}{n}\log p(X^n) = \underbrace{-\frac{1}{n}\sum_{i=1}^{n}\log p(X_i)}_{\text{Media campione delle } \mathcal{I}(X_i) = -\log p(X_i) \in \mathbb{R}}$$

Sono v.a. reali

Per il teorema di Chebyshev,

la media campione delle $\mathcal{I}(X_i)$ converge in prob. alla media dei loro valori attesi:

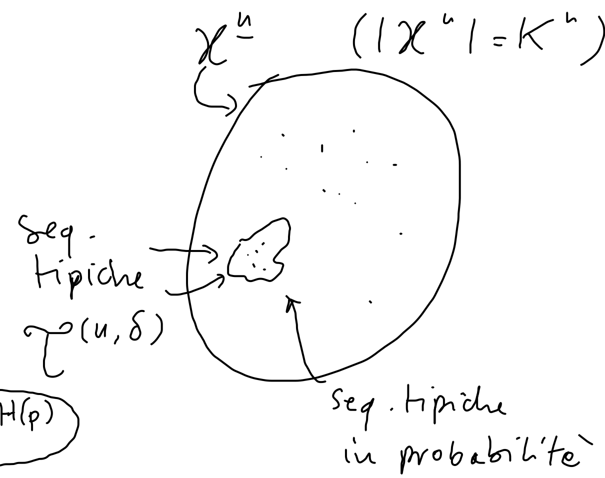
$$\frac{1}{n}\mathcal{I}(X^n) = \frac{1}{n}\sum_{i=1}^{n}[-\log p(X_i)] \xrightarrow{\text{converge in prob.}} \frac{1}{n}\sum_{i=1}^{n}\overbrace{E[-\log p(X_i)]}^{H(X_i)} = \frac{1}{\cancel{n}}\cancel{n}\,H(X).$$

L'autoinformaz. normalizzata converge in prob. all'entropia della sorgente.

$$\iff \quad \forall \delta > 0 \quad \exists \varepsilon_{n,\delta} : \Pr\Big[\underbrace{\Big|\frac{1}{n}\mathcal{I}(X^n) - H(X)\Big| \le \delta}\Big] \ge 1 - \varepsilon$$

Una sequenza $\vec{x} \in \mathcal{X}^n$ che soddisfa questa relazione è detta (n,δ)-tipica (tipica in probabilità):

Def.: $\vec{x} \in \mathcal{X}^n$ è $\underbrace{(n,\delta)}_{\text{lunghezza}}$-tipica in probabilità se $\Big|\frac{1}{n}\mathcal{I}(\vec{x}) - \overbrace{H(\vec{p})}^{=-\sum p_i \log p_i}\Big| \le \delta$.



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

Teo 3.5. Consideriamo $\mathcal{T}^{(n,\delta)}$. Allora:

$p(\vec{x}) \approx 2^{-nH(p)}$

(a) Probabilità di una sequenza tipica: $\forall \vec{x} \in \mathcal{T}^{(n,\delta)}$: $2^{-n[H(\vec{p})+\delta]} \le p(\vec{x}) \le 2^{-n[H(\vec{p})-\delta]}$

(b) Numerosità delle sequenze tipiche: $(1-\varepsilon)\,2^{n[H(\vec{p})-\delta]} \le |\mathcal{T}^{(n,\delta)}| \le 2^{n[H(\vec{p})+\delta]}$

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