Sorgente di informazione $\mathcal{S}$

Ad ogni istante di tempo (discreto) t_j $\quad (\ldots t_{-2}, t_{-1}, t_0, t_1, t_2, \ldots)$

emette un simbolo dell'alfabeto $\mathcal{X} = \{x_1, x_2, \ldots, x_K\}$

ciascuna ha valore $\in \mathcal{X}$

Abbiamo una successione di v.a. $\quad \ldots X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots$

Tempo	...	t_{-2}	t_{-1}	t_0	t_1	t_2	....
Simboli	...	X_{-2}	X_{-1}	X_0	X_1	X_2	....

$j \qquad j+n$

n

Per descrivere la sorgente $\mathcal{S}$ dovremmo fornire una famiglia infinita di d.p. congiunte (una per ogni sottosequenza) del tipo

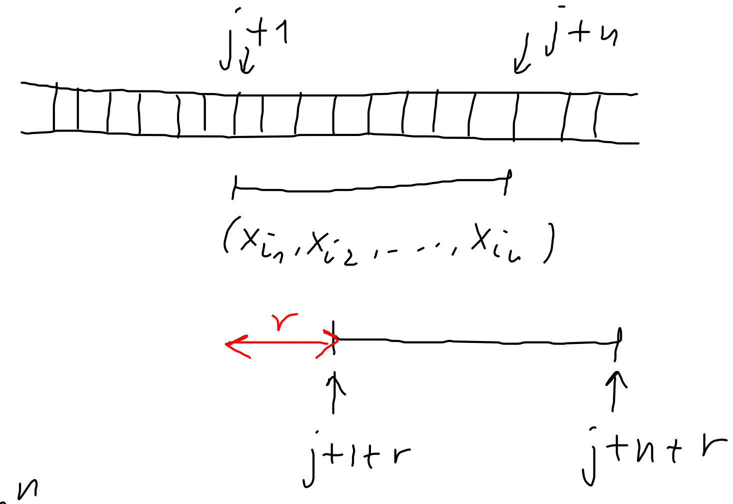
$$\Pr\left[(X_{j+1}, X_{j+2}, \ldots, X_{j+n}) = (x_{i_1}, x_{i_2}, x_{i_3}, \ldots, x_{i_n})\right] \quad \forall j \in \mathbb{Z} \quad \forall n \in \mathbb{N}$$

$$\forall (x_{i_1}, x_{i_2}, \ldots, x_{i_n}) \in \mathcal{X}^n$$

Ci servono ulteriori assunzioni per semplificare la descrizione di $\mathcal{S}$:

- Stazionarietà
- Assenza di memoria

① Stazionarietà



$$\Pr[(X_{j+1}, X_{j+2}, \dots, X_{j+n}) = (x_{i_1}, \dots, x_{i_n})] =$$

$$= \Pr[(X_{j+1+r}, \dots, X_{j+n+r}) = (x_{i_1}, \dots, x_{i_n})]$$

$\forall j \in \mathbb{Z} \quad \forall n \in \mathbb{N} \quad \forall r \in \mathbb{Z} \quad \forall (x_{i_1}, x_{i_2}, \dots, x_{i_n}) \in \mathcal{X}^n$

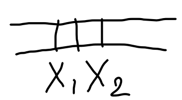
(invarianza rispetto a traslazioni di tempo)

② Assenza di memoria

$$\Pr[(X_{j+1}, \dots, X_{j+n}) = (x_{i_1}, \dots, x_{i_n})] = \prod_{r=1}^{n} \Pr[X_{j+r} = x_{i_r}]$$

in altre parole, $X_{j+1}, \dots, X_{j+r}$ sono <u>indipendenti</u>

Lingua inglese:



$\Pr[X = 'Q'] = 0.12\% = 0.0012 \approx 10^{-3}$

$\Pr[X = 'U'] = 2.73\% = 0.0273 \approx 2 \cdot 10^{-2}$

$\Pr[X_1 = 'Q', X_2 = 'U'] \simeq 10^{-3}$ QU

Se ci fosse assenza di memoria, dovremmo avere:

$$\Pr[X_1 = 'Q', X_2 = 'U'] = \underbrace{\Pr[X_1 = 'Q']}_{\sim 10^{-3}} \underbrace{\Pr[X_2 = 'U']}_{2 \cdot 10^{-2}}$$

$\approx 2 \cdot 10^{-5}$

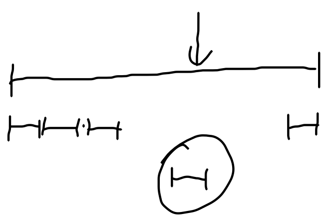
Se vale sia la stazionarietà che l'assenza di memoria, possiamo descrivere la sorgente S con un'unica distribuzione di probabilità:

per assenza di memoria

$$Pr[(X_{j+1}, \dots, X_{j+n}) = (x_{i_1}, \dots, x_{i_n})] = \prod_{r=1}^{n} Pr[X_{j+r} = x_{i_r}]$$

$$\overset{\text{per stazionarietà}}{=} \prod_{r=1}^{n} Pr[X_r = x_{i_r}]$$

Tutte le X_r sono identicamente distribuite



Mi basta una d.p. sull'alfabeto $\mathcal{X} = \{x_1, \dots, x_K\}$ che descrive $Pr[X = x_i]$ per ciascun $i = 1 \dots K$ $\quad (p = (p_1, \dots, p_K))$

Def. Definisco l'entropia della sorgente S come

$$H(S) = H(X) = -\sum_{i=1}^{K} p_i \log p_i$$

("quantità di informazione media per simbolo emessa dalla sorgente S")

Se S è stazionaria ma <u>con</u> memoria, l'entropia può essere definita in due modi:

Def. 1) $$H(S) \overset{def.}{=} \lim_{n \to \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n)$$

t_1

$X_1 \ X_2 \ \dots \ X_n$

$H(X_1, X_2, \dots, X_n)$

Def. 2) $$H(S) \overset{def.}{=} \lim_{n \to \infty} H(X_n | X_1, \dots, X_{n-1})$$

$X_1 \ X_2 \ \dots \ X_{n-1} \ X_n$

Si può dimostrare che le def. 1) e 2) sono equivalenti (vedi testo).

Se abbiamo sia stazionarietà che assenza di memoria:

$$H(X_1, X_2, \dots, X_n) = \sum_{i=1}^{n} H(X_i) \quad \text{(per indipendenza)}$$

$$= n\,H(X) \implies 1/n\, H(X_1, \dots, X_n) = H(X)$$

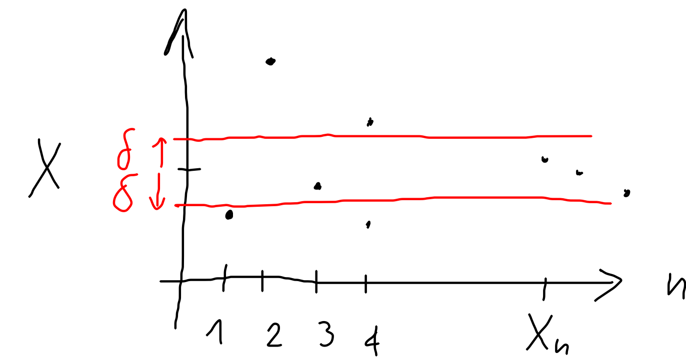
$$H(X_n | X_1, X_2 \dots X_{n-1}) = H(X_n) = H(X) \implies H(X)$$

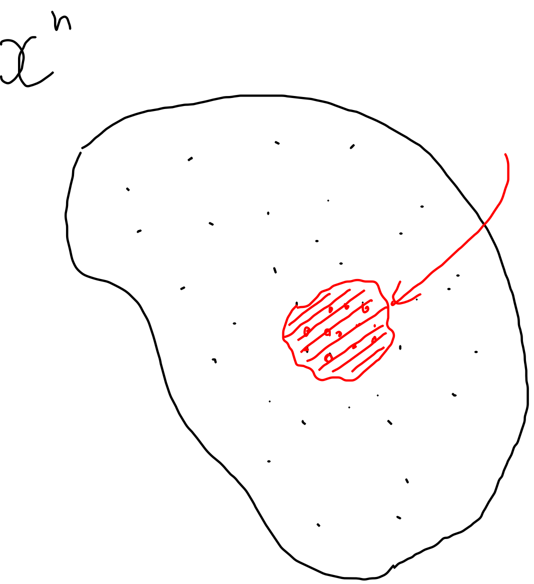
Legge dei grandi numeri.

Sia una successioni di v.a. reali $\{X_n\}$

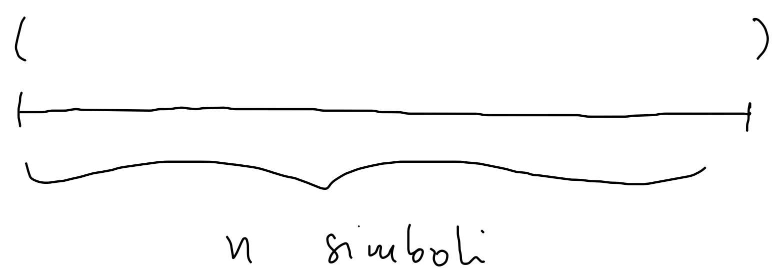
Si ha convergenza in probabilità di $\{X_n\}$ alla v.a. X se

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





Sequenze "tipiche" raccolgono una "grande" probabilità pur essendo in numero "molto basso" rispetto a $|\mathcal{X}^n| = K^n$



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

Def. Una successione $\{X_n\}$ di v.a. reali soddisfa la legge dei grandi numeri se la successione delle medie di $\{X_n\}$ converge in probabilità al suo valore atteso:

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

Succ. delle medie: $Y_1, Y_2, \dots, Y_n, \dots$

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$\Longleftrightarrow$ $\{Y_n\}$ converge in probabilità al suo valore atteso $Y = \mathbb{E}[Y_n]$

(V) Teorema di Chebyshev. Se $\{X_n\}$ è una successione di v.a. <u>indipendenti</u> e per ogni $n \in \mathbb{N}$ ho $Var[X_n] \leq C$ (per qualche costante C) allora $\{X_n\}$ soddisfa la legge dei grandi numeri.

Disuguaglianza di Chebyshev. Se la v.a. X ha varianza $Var[X]$ finita, allora
$$\Pr[\,|X - \mathbb{E}[X]| \geq \delta\,] \leq \frac{Var[X]}{\delta^2}.$$

<u>Dim.</u>
$$\Pr[\,|X - \overbrace{\mathbb{E}X}^{\mu}| \geq \delta\,] = \sum_{x : |x-\mu| \geq \delta} 1 \cdot p(x) \leq \sum_{x : |x-\mu| \geq \delta} \underbrace{\boxed{\frac{(x-\mu)^2}{\delta^2}}}_{\geq 1} \cdot p(x) \leq$$

$$\leq \frac{1}{\delta^2} \underbrace{\sum_x (x-\mu)^2 p(x)}_{= \mathbb{E}_X[(X-\mu)^2] = \mathbb{E}[(X - \mathbb{E}X)^2] = Var[X]} = \frac{Var[X]}{\delta^2}. \qquad \text{QED}$$

Proprietà della varianza:

→ ⊛ $Var[\alpha X+\beta] = \alpha^2 Var[X]$

→ ⊛ Se X_1 e X_2 sono v.a. indipendenti, allora $Var[X_1+X_2] = Var[X_1]+Var[X_2]$

Dim. del teorema di Chebyshev:

Consideriamo la v.a. $Y_n = 1/n \sum_{i=1}^n X_i$

$$Var\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n Var[X_i]$$

$$\Rightarrow Var[Y_n] = \frac{1}{n^2}\sum_{i=1}^n Var[X_i] \le \frac{C\cdot n}{n^2} = \frac{C}{n}.$$

$\mathbb{E}Y_n$

Per la disuguaglianza di Chebyshev applicata a Y_n, ho: $Pr[|Y_n - Y| \ge \delta] \le \frac{C/n}{\delta^2}$

$$= \frac{C}{n\delta^2} \xrightarrow{n\to\infty} 0$$

$\Rightarrow Pr[|Y_n-Y|<\delta] \longrightarrow 1$ per $n\to\infty$