

SORSENTI DI INFORMAZIONE

lezioni del corso di

TRASMISSIONE ED ELABORAZIONE NUMERICA
DEI SEGNALI (COMUNICAZIONI ELETTRICHE

AA. 2016-17 (SUN)

Prof. FRANCESCO A. N. PALMERI

PA. LUCERI. 1

Capitolo 1

Sorgenti discrete senza memoria

In questo capitolo analizzeremo le sorgenti tempo-discreto a valori discreti e alfabeto finito (sorgenti discrete). Tale schematizzazione è quella più tipica e costituisce il principale modello di riferimento usato nei moderni sistemi di telecomunicazione. Viene inoltre introdotta l'entropia di sorgente, i limiti teorici per un sistema di compressione senza perdite e alcune tecniche di progetto per codici efficienti.

1.1 Sorgenti discrete

Si assuma che la nostra sorgente S tempo-discreto emetta ad ogni istante di tempo un simbolo appartenente all'alfabeto sorgente

$$\mathcal{A} = \{a_1, \dots, a_n\}, \quad (1.1)$$

ovvero $S(t_i) = a_j$, $j = 1, \dots, n$. Si supponga che la sequenza

$$\dots S(t_{-2}), S(t_{-1}), S(t_0), S(t_1), S(t_2), \dots \quad (1.2)$$

sia denotata per semplicità nel seguente modo

$$\dots S[-2], S[-1], S[0], S[1], S[2], \dots \quad (1.3)$$

dove si è usato il solo argomento intero a indicare gli istanti di tempo (si noti la differenza tra le due funzioni $S(\cdot)$ e $S[\cdot]$). Anche se, come già anticipato, nella maggioranza dei sistemi analizzati la spaziatura tra gli istanti di tempo

è uniforme, tale notazione consente di svincolarsi totalmente dalla associazione diretta alla sequenza degli istanti di tempo. Nel caso di spaziatura uniforme, la sequenza degli interi è da intendersi come sequenza degli istanti di tempo normalizzata.

Si supponga che la sorgente emetta simboli in maniera indipendente, e che tali simboli vengano generati secondo la distribuzione di probabilità

$$\Pi = \{p_1, p_2, \dots, p_n\}. \quad (1.4)$$

La sequenza $\{S[\cdot]\}$ è detta *iid*, ovvero a simboli *indipendenti e identicamente distribuiti* (sorgente discreta senza memoria).

1.1.1 Il Bit-rate

Si supponga che la sorgente emetta simboli a intervalli di tempo regolarmente spaziatu di T secondi. Ad ogni istante di tempo la sorgente quindi sceglie secondo il suo meccanismo aleatorio uno di n simboli. Si definisce *bit-rate*, o *tasso binario*, o *ritmo binario*, o *frequenza di cifra* della sorgente:

$$B_r = \frac{\log_2 n}{T} \text{ bit/sec}. \quad (1.5)$$

La definizione deriva dal fatto che se l'indice corrispondente al simbolo emesso dovesse essere rappresentato con una sequenza binaria, avrebbe bisogno di $\log_2 n$ bit. Ovviamente $\log_2 n$ non è necessariamente un numero intero e una reale associazione ad una stringa binaria richiederebbe $b = \lceil \log_2 n \rceil$ bit, dove il simbolo $\lceil x \rceil$ indica l'intero immediatamente superiore a x . In tale caso utilizzando b bit, non necessariamente tutte le 2^b stringhe binarie sarebbero utilizzate poichè $2^b \geq n$. Pertanto alla definizione di bit-rate non necessariamente corrisponde una associazione ad un codice binario. Inoltre, va menzionato che il bit-rate solo impropriamente misura l'entità di informazione emessa dalla sorgente. Infatti, come vedremo nel seguito, è necessario tenere conto anche della distribuzione di probabilità dei simboli per quantificare rigorosamente l'ammontare di informazione emessa dalla sorgente nell'unità di tempo.

Esempio 0.3 Si consideri di nuovo la sorgente binaria dell'esempio 0.3. I bit siano emessi in maniera indipendente e secondo la distribuzione di probabilità $\Pi = \{0.3, 0.7\}$. Il bit-rate è $B_r = 1/T = 10^6$ bit/sec=1Mbit/sec. Poiché la distribuzione dei due simboli non è uniforme nella sequenza ci saranno in media più "1" che "0". Vedremo in seguito che la definizione corretta di velocità

di informazione, più propriamente che il bit-rate, tiene conto anche della distribuzione di probabilità sui vari simboli. Vedremo anche che quando i simboli sono equiprobabili le due definizioni coincidono.

Esempio 1.1 Si consideri una sorgente che emette simboli appartenenti all'alfabeto $\mathcal{A} = \{1, 2, 3, 4, 5, 6\}$. I simboli siano emessi a intervalli regolari di $T = 0.5$ secondi. La frequenza di simbolo è $R_s = 1/T = 2$ simboli/sec. Il bit-rate è $B_r = \log_2(6/0.5) = 5.17$ bit/sec. Se fosse necessaria una reale associazione ad una sequenza binaria, avremmo bisogno di $b = \lceil \log_2 6 \rceil = 3$ bit, con due simboli inutilizzati. In tale caso la sorgente dopo la associazione (codifica) emetterebbe 6 bit/sec.

1.2 Il concetto di informazione

Un sistema di telecomunicazione è deputato al trasporto dell'informazione tra una o più sorgenti e dei destinatari. E' necessario pertanto capire cosa intendiamo con il termine *informazione*, ovvero tentare una definizione matematica che quantifichi l'informazione generata da una sorgente e quella trasportabile da un canale. Tali concetti sono di fondamentale importanza e costituiscono il cuore della teoria dell'informazione. La teoria mira a quantificare in termini matematici l'entità di ciò che è realmente generato dalla sorgente e in che misura ciò è trasportabile da un canale reale.

Partiamo da un esempio. Consideriamo una sequenza di caratteri dell'alfabeto. E' intuitivo pensare che tale scritto non trasporta alcuna informazione se la sequenza è ripetitiva e prevedibile. Diversamente potremmo dire se la configurazione dei simboli non segue uno schema predeterminato, ma si mostra *innovativa* o imprevedibile. Come casi limite si pensi ad una sequenza periodica o ad una sequenza costante, e ad una sequenza di caratteri scelti completamente a caso. Nella prima situazione tale sorgente è scarsamente informativa, mentre nella seconda vale il viceversa. La definizione di autoinformazione di un evento è basata proprio su questa idea: sorgenti prevedibili, ovvero con distribuzioni dell'alfabeto degeneri, magari con la probabilità concentrata tutta su un sottoinsieme di simboli, generano meno informazione che sorgenti meno prevedibili e più uniformemente distribuite. Le parole *incertezza*, *imprevedibilità*, *informazione* sono da considerare sinonimi nel contesto della Teoria dell'Informazione. Tali idee saranno tradotte in definizioni matematiche nelle definizioni di *autoinformazione* e *entropia di sorgente*.

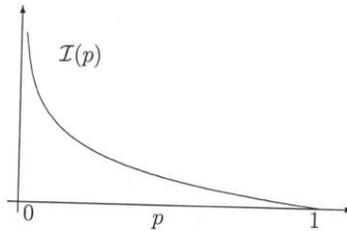


Figura 1.1: Il grafico della autoinformazione \mathcal{I} in funzione di p

1.2.1 Autoinformazione

Consideriamo una sorgente discreta con alfabeto $\mathcal{A} = \{a_1, a_2, \dots, a_n\}$ e distribuzione di probabilità $\Pi = \{p_1, p_2, \dots, p_n\}$. Denotando il simbolo emesso dalla sorgente con S , diremo che l'ammontare di informazione guadagnata dalla osservazione del simbolo a_i , ovvero quando $S = a_i$, è data dalla *autoinformazione* $\mathcal{I}(a_i)$ definita come

$$\mathcal{I}(a_i) = \log\left(\frac{1}{p_i}\right) = -\log p_i. \quad (1.6)$$

\mathcal{I} si misura in *bit* se il logaritmo è in base 2, mentre si misura in *Nat* se il logaritmo è in base naturale.¹ Si noti come la autoinformazione associata al simbolo i -esimo sia solo funzione della probabilità associata a tale simbolo e quindi indipendente dalla etichetta o dal nome ad esso associato. Pertanto adotteremo intercambiabilmente le notazioni $\mathcal{I}(a_i)$ e $\mathcal{I}(p_i)$. Figura 1.1 mostra il grafico della funzione $\mathcal{I}(p) = -\log_2 p$. La definizione, grazie alle proprietà del logaritmo, è particolarmente appropriata poiché soddisfa le seguenti proprietà:

Proprietà 1: $\mathcal{I}(p_i) \geq 0$ poiché $0 \leq p_i \leq 1$. La osservazione di un simbolo trasporta sempre una informazione positiva e mai una perdita di informazione.

¹Anche se piuttosto inusuali, altre basi possono essere utilizzate. Per esempio, se la base del logaritmo è 3, si parla di *trit*, se la base è 4, si parla di *quadrit*, eccetera. Nel seguito di queste note faremo quasi sempre riferimento al bit, che è comunque l'unità minima.



1.2. IL CONCETTO DI INFORMAZIONE

Proprietà 2: $\mathcal{I}(p_i) = 0$ se $p_i = 1$. La osservazione di un evento certo non trasporta alcuna informazione.

Proprietà 3: $\mathcal{I}(p_i) > \mathcal{I}(p_j)$ se $p_i < p_j$. La osservazione di un evento meno probabile trasporta più informazione di un evento più probabile.

Proprietà 4: $\mathcal{I}(p_i, p_j) = \mathcal{I}(p_i) + \mathcal{I}(p_j)$ se a_i e a_j sono i risultati di due osservazioni indipendenti. Dalla osservazione congiunta di due eventi indipendenti si guadagna la somma dell'informazione delle due.

Pertanto, all'insieme dei simboli di un alfabeto sorgente può essere associato l'insieme delle autoinformazioni

$$\{\mathcal{I}(p_1), \mathcal{I}(p_2), \dots, \mathcal{I}(p_n)\}. \tag{1.7}$$

Le prove delle proprietà sono lasciate al lettore per esercizio. Esse si ottengono immediatamente dalla definizione di autoinformazione e dalle proprietà del logaritmo.

Esempio 0.3 Torniamo alla sequenza binaria di di esempio 0.3. Poiché le probabilità associate ai due simboli sono $\Pi = \{0.3, 0.7\}$, le autoinformazioni saranno $\{\mathcal{I}(\frac{1}{2}), \mathcal{I}(\frac{1}{2})\} = \{\log_2 10/3, \log_2 10/7\} = \{1.74, 0.52\}$ misurate in bit. Si noti come lo "0" che è meno probabile sia più informativo. Se i simboli fossero stati equiprobabili, ovvero $\Pi = \{1/2, 1/2\}$, avremmo avuto $\{\mathcal{I}(\frac{1}{2}), \mathcal{I}(\frac{1}{2})\} = \{1, 1\}$, con uguale informazione pari a un bit per entrambi.

Esempio 1.1 Torniamo all'esempio della sorgente che emette simboli dall'alfabeto $\mathcal{A} = \{1, 2, 3, 4, 5, 6\}$, con probabilità uniformi $\Pi = \{1/6, 1/6, \dots, 1/6\}$. Le autoinformazioni sono $\{\mathcal{I}(\frac{1}{6}), \mathcal{I}(\frac{1}{6}), \dots, \mathcal{I}(\frac{1}{6})\} = \{\log_2 6, \log_2 6, \dots, \log_2 6\} = \{2.59, 2.59, \dots, 2.59\}$ bit. Ogni simbolo trasporta un pò più di due bit.

1.2.2 Entropia di sorgente

Se valutiamo la autoinformazione media di una sorgente S , abbiamo

$$\mathcal{H}(S) = E[\mathcal{I}(S)] = \sum_{i=1}^n p_i \mathcal{I}(p_i) = \sum_{i=1}^n p_i \log_2 \frac{1}{p_i} = - \sum_{i=1}^n p_i \log_2 p_i. \tag{1.8}$$

Tale quantità è detta *entropia della sorgente* e rappresenta la informazione mediamente trasportata da ogni simbolo. Si noti che l'entropia dipende solo da come la probabilità è distribuita sui vari simboli. Pertanto adotteremo in

seguito entrambi le notazioni $\mathcal{H}(\Pi)$ e $\mathcal{H}(S)$.² Definendo un vettore (colonna) delle probabilità della sorgente come

$$\boldsymbol{\pi} = (p_1, p_2, \dots, p_n)^T, \quad (1.9)$$

l'entropia può essere calcolata mediante la espressione matriciale

$$\mathcal{H}(\Pi) = -\boldsymbol{\pi}^T (\log_2 \boldsymbol{\pi}), \quad (1.10)$$

dove $(\log_2 \boldsymbol{\pi})$ è il vettore (colonna) dei logaritmi degli elementi di $\boldsymbol{\pi}$.³

L'entropia di sorgente soddisfa le seguenti proprietà:

Proprietà 1: $\mathcal{H}(\Pi) \geq 0$.

Proprietà 2: $\mathcal{H}(\Pi) = 0$, se e solo se $p_k = 1$ per un qualunque k , e tutte le altre probabilità sono nulle. Tale situazione corrisponde ad una sequenza certa e costante del simbolo a_k e pertanto non fornisce alcuna informazione.

Proprietà 3: $\mathcal{H}(\Pi) = \log_2 n$, se e solo se i simboli di sorgente sono equiprobabili ($p_i = 1/n, i = 1, \dots, n$). Tale valore rappresenta il limite superiore per l'entropia della sorgente discreta

$$\mathcal{H}(\Pi) \leq \log_2 n \quad \forall \Pi \quad (1.11)$$

Quindi la distribuzione uniforme è quella a massima entropia.

Proprietà 4: L'entropia è una funzione concava di Π .

²Il termine "entropia" non a caso richiama una analogia con l'entropia termodinamica. Esso fu formalizzato da Shannon nella teoria dell'informazione nella forma qui fornita. Shannon trasse ispirazione dai risultati fondamentali sulla termodinamica di Boltzmann e da una idea di Hartley, che nel 1930 per primo misurò l'informazione di una sorgente come il logaritmo della cardinalità dell'alfabeto. L'entropia "informazionale" qui definita può essere studiata come una definizione matematica autonoma nell'ambito della caratterizzazione delle sorgenti aleatorie, ma esistono degli interessanti collegamenti con la fisica teorica. Le analogie tra termodinamica e informazione sono di straordinario interesse in quanto sembrano fornire una sorprendente visione unificata della fisica e delle comunicazioni (Cover e Thomas, 1991).

³In queste note, quando saranno introdotte delle formule notevoli, cercheremo di fornire sempre la corrispondente espressione matriciale. La sintassi sarà molto simile a quella del software MATLAB allo scopo di consentire allo studente una facile traduzione delle formule.

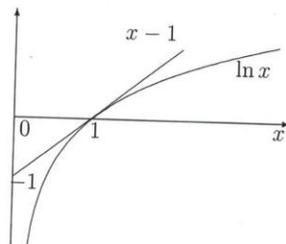


Figura 1.2: La proprietà $\ln x \leq x - 1, x \geq 0$

Prova delle proprietà 1-4: Proprietà 1 e' immediata conseguenza della positività delle autoinformazioni dei vari simboli. Anche Proprietà 2 è immediata se si osserva che $p_i \log_2 \frac{1}{p_i}$ è zero se e solo se $p_i = 0$ o se $p_i = 1$. Per provare Proprietà 3 è necessario usare la proprietà del logaritmo naturale

$$\ln x \leq x - 1, \quad x \geq 0, \quad (1.12)$$

dove l'uguaglianza vale solo se e solo se $x = 1$, punto in cui la retta che costituisce il limite superiore, è tangente al logaritmo. Il grafico di figura 1.2 della funzione logaritmo e della retta mostrano la proprietà. Si considerino quindi due distribuzioni di probabilità $\Pi = \{p_1, p_2, \dots, p_n\}$ e $Q = \{q_1, q_2, \dots, q_n\}$, e la seguente espressione

$$\alpha = \sum_{i=1}^n p_i \log_2 \frac{q_i}{p_i}. \quad (1.13)$$

Passando al logaritmo naturale e usando disuguaglianza (1.12) dimostriamo che $\alpha \leq 0$

$$\begin{aligned} \alpha &= \frac{1}{\ln 2} \sum_{i=1}^n p_i \ln \frac{q_i}{p_i} \\ &\leq \frac{1}{\ln 2} \sum_{i=1}^n p_i \left(\frac{q_i}{p_i} - 1 \right) \\ &= \frac{1}{\ln 2} \sum_{i=1}^n (q_i - p_i) \\ &= \frac{1}{\ln 2} \left(\sum_{i=1}^n q_i - \sum_{i=1}^n p_i \right) = 0 \end{aligned}$$

Pertanto

$$\sum_{i=1}^n p_i \log_2 \frac{q_i}{p_i} \leq 0. \quad (1.14)$$

Ponendo $Q = \{\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\}$, abbiamo

$$\sum_{i=1}^n q_i \log_2 \frac{1}{q_i} = \mathcal{H}(Q) = \log_2 n, \quad (1.15)$$

$$\begin{aligned} \alpha = \sum_{i=1}^n p_i \log_2 \frac{q_i}{p_i} &= \sum_{i=1}^n p_i \log_2 \frac{1}{p_i} - \sum_{i=1}^n p_i \log_2 n \\ &= \sum_{i=1}^n p_i \log_2 \frac{1}{p_i} - \log_2 n \leq 0, \end{aligned}$$

da cui discende che

$$\sum_{i=1}^n p_i \log_2 \frac{1}{p_i} \leq \log_2 n. \quad (1.16)$$

Per provare Proprietà 4, osserviamo innanzitutto come la funzione $x \ln x$ sia una funzione convessa. Infatti $\frac{d^2}{dx^2}(x \ln x) = \frac{1}{x} > 0, \forall x > 0$. Dalla definizione di funzione convessa, per qualunque coppia di valori $0 < p_i < 1$ e $0 < q_i < 1$, e $\lambda \in [0, 1]$, abbiamo che

$$(\lambda p_i + (1 - \lambda)q_i) \ln (\lambda p_i + (1 - \lambda)q_i) \leq \lambda p_i \ln p_i + (1 - \lambda)q_i \ln q_i. \quad (1.17)$$

Cambiando base al logaritmo (ricordiamo che $\log_2 e > 0$) e sommando su i , avendo assunto che $\Pi = \{p_1, \dots, p_n\}$ e $Q = \{q_1, \dots, q_n\}$ sono due possibili distribuzioni per S , abbiamo che

$$\sum_{i=1}^n (\lambda p_i + (1 - \lambda)q_i) \log_2 (\lambda p_i + (1 - \lambda)q_i) \leq \lambda \sum_{i=1}^n p_i \log_2 p_i + (1 - \lambda) \sum_{i=1}^n q_i \log_2 q_i. \quad (1.18)$$

Ovvero

$$\mathcal{H}(\lambda \Pi + (1 - \lambda)Q) \geq \lambda \mathcal{H}(\Pi) + (1 - \lambda)\mathcal{H}(Q), \quad (1.19)$$

che mostra la concavità di \mathcal{H} . \triangle

Esempio 0.3 Torniamo all'esempio della sequenza binaria. Se la distribuzione delle probabilità è $\Pi = \{0.3, 0.7\}$, abbiamo che $\mathcal{H}(\Pi) = 1.74 \cdot 0.3 + 0.52 \cdot 0.7 = 0.88$ bit. Se la distribuzione delle probabilità è uniforme l'entropia diventa: $\mathcal{H}(\Pi) = 1 \cdot 0.5 + 1 \cdot 0.5 = 1$ bit. Si noti come la informazione media sia maggiore nel caso equiprobabile (la entropia di una sorgente binaria sarà più specificamente analizzata nel seguito).

Esempio 1.1 Nell'esempio della sorgente con alfabeto $\mathcal{A} = \{1, 2, 3, 4, 5, 6\}$ e probabilità uniformi, abbiamo che l'entropia è $\mathcal{H}(\Pi) = 6 \cdot \frac{1}{6} \log_2 6 = 2.59$ bit.

1.2.3 Sorgenti a simboli equiprobabili

Un caso tipico di sorgente è quella a simboli aventi una distribuzione di probabilità uniforme. In tale caso l'entropia diventa semplicemente

$$\mathcal{H}(\Pi) = \log_2 n. \quad (1.20)$$

Abbiamo già visto da Proprietà 3 come tale sorgente, ovvero la distribuzione uniforme, sia quella a cui corrisponde la massima entropia. Va notato come questa espressione sia suscettibile di una interpretazione intuitiva: *l'informazione emessa da una sorgente a simboli equiprobabili è pari al numero di bit (non necessariamente un intero) necessario a rappresentare tutti i simboli dell'alfabeto con un codice binario.*

1.2.4 L'entropia di una sorgente binaria

Il caso di una sorgente avente un alfabeto binario, merita una attenzione particolare. L'alfabeto sorgente sia $\mathcal{A} = \{a_1, a_2\}$ e le probabilità corrispondenti siano $\Pi = \{p, (1-p)\}$. L'entropia della sorgente è

$$\mathcal{H}(p) = -p \log_2 p - (1-p) \log_2 (1-p) \quad (1.21)$$

(la notazione $\mathcal{H}(p)$ riflette il fatto che nel caso binario basta un solo parametro a specificare la distribuzione di probabilità). Figura 1.3 mostra l'andamento di \mathcal{H} in funzione di p . Si noti come nei casi in cui tutta la probabilità è concentrata su uno dei simboli, ovvero quando $p = 1$ o $p = 0$, l'entropia è nulla. Ciò è facilmente interpretabile, dato il significato di informazione connesso alla imprevedibilità della sorgente. Si noti inoltre come l'entropia sia massima quando $p = \frac{1}{2}$, che è la situazione di massima incertezza.

1.2.5 ~~La velocità d'informazione~~ ^{il tasso informativo}

La definizione matematica di entropia ci consente di quantificare rigorosamente l'entità di informazione emessa da una sorgente. Per una sorgente che emette simboli a cadenza costante di T secondi, la *velocità media di informazione* della sorgente, detta anche *tasso informativo*, è

$$R = \frac{\mathcal{H}(\Pi)}{T} \text{ bit/sec} \quad (1.22)$$

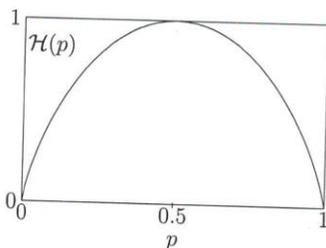


Figura 1.3: L'entropia di una sorgente binaria

Tale definizione è di fondamentale importanza nella teoria dell'informazione. Si noti come tale definizione non necessariamente coincida con quella di bit-rate, eccetto che nel caso in cui i simboli sono equiprobabili:

$$R = \frac{\mathcal{H}(\Pi)}{T} = \frac{\log_2 n}{T} = B_r \text{ bit/sec.} \quad (1.23)$$

Inoltre da Proprietà 3 abbiamo sempre che $R \leq B_r$.⁴

1.3 Sorgenti estese

Nel presentare alcuni dei principali risultati della teoria dell'informazione nel seguito di queste note, sarà spesso utile fare riferimento a *sorgenti estese*, ovvero a sorgenti il cui simbolo è composto da blocchi di simboli emessi in istanti successivi da una sorgente di partenza. Più specificamente si supponga di raggruppare m simboli successivi di una sorgente S

$$\underbrace{\dots, S[k], S[k+1], S[k+2], \dots, S[k+m]}_{S^m [k]}, \underbrace{\dots, S[k+m+1], \dots}_{S^m [k+m]} \quad (1.24)$$

⁴E' opportuno segnalare al lettore che nella letteratura tecnica i termini bit-rate e velocità di informazione (information rate) sono a volte usati in maniera intercambiabile, generando secondo noi un pò di confusione. Qualche autore preferisce infatti usare il termine bit-rate per indicare la velocità di informazione senza usare una terminologia specifica per la quantità qui definita come bit-rate. In queste note manterremo coerenza con le definizioni qui fornite.

1.3. SORGENTI ESTESE

23

Ogni raggruppamento può essere considerato come una nuova sorgente S^m con alfabeto composto dal prodotto cartesiano

$$\mathcal{A}^m = \underbrace{\mathcal{A} \times \mathcal{A} \times \dots \times \mathcal{A}}_m \quad (1.25)$$

La cardinalità del nuovo alfabeto sarà n^m e la sorgente estesa S^m avrà frequenza di simbolo pari a $R_{sm} = \frac{R_s}{m}$ simboli/sec.

Per esempio, in una sorgente con alfabeto binario $\mathcal{A} = \{a_1, a_2\}$, raggruppando i simboli due a due ($m = 2$), abbiamo

$$\mathcal{A}^2 = \{(a_1a_1), (a_1a_2), (a_2a_1), (a_2a_2)\}. \quad (1.26)$$

La sorgente ha cardinalità doppia e frequenza di simbolo pari alla metà.

Nella sorgente composta, ogni elemento dell'alfabeto ha una distribuzione di probabilità data dalle probabilità congiunte degli elementi componenti la m -pla di simboli. In particolare l'entropia della sorgente sarà

$$\mathcal{H}(\Pi^m) = - \sum_{(x_1, x_2, \dots, x_m) \in \mathcal{A}^m} p(x_1, x_2, \dots, x_m) \log_2 p(x_1, x_2, \dots, x_m). \quad (1.27)$$

Una sorgente può in generale esibire una struttura di dipendenza temporale tra i simboli emessi in istanti successivi. Nel caso più semplice (e tipico) di simboli indipendenti, si parla di *sorgente senza memoria*. In tale caso la probabilità congiunta è esprimibile come prodotto delle probabilità marginali. Ad esempio nel caso $\mathcal{A} = \{a_1, a_2\}$, $\Pi = \{p_1, p_2\}$, con $m = 2$, la distribuzione delle probabilità corrispondente ad \mathcal{A}^2 è

$$\Pi^2 = \{p_1^2, p_1p_2, p_2p_1, p_2^2\}, \quad (1.28)$$

La distribuzione relativa ad

$$\mathcal{A}^3 = \{a_1a_1a_1, a_1a_1a_2, a_1a_2a_1, a_1a_2a_2, a_2a_1a_1, a_2a_1a_2, a_2a_2a_1, a_2a_2a_2\}, \quad (1.29)$$

è data da

$$\Pi^3 = \{p_1^3, p_1^2p_2, p_1p_2^2, p_2^3, p_1p_1p_2, p_1p_2p_1, p_1p_2^2, p_1p_2^2, p_1p_2^2, p_2^3\}. \quad (1.30)$$

Per qualunque sorgente a simboli indipendenti vale la seguente

Proposizione: *L'entropia di una sorgente estesa di ordine m , ottenuta da una sorgente a simboli indipendenti e di entropia $\mathcal{H}(\Pi)$ è*

$$\mathcal{H}(\Pi^m) = m\mathcal{H}(\Pi). \quad (1.31)$$

e direttamente dai simboli

Prova: L'indipendenza della sorgente implica che

$$P(x_1, x_2, \dots, x_m) = P(x_1)P(x_2) \cdots P(x_m), \quad (1.32)$$

$\forall (x_1, x_2, \dots, x_m) \in \mathcal{A}^m$. La autoinformazione di ogni simbolo nella sorgente estesa è pertanto

$$\begin{aligned} I(x_1, x_2, \dots, x_m) &= -\log_2 P(x_1, x_2, \dots, x_m) \\ &= -\log_2 P(x_1) - \log_2 P(x_2) - \dots - \log_2 P(x_m) \\ &= -\sum_{l=1}^m \log_2 P(x_l). \end{aligned}$$

L'entropia della sorgente estesa diventa

$$\begin{aligned} \mathcal{H}(\Pi^m) &= - \sum_{(x_1, x_2, \dots, x_m) \in \mathcal{A}^m} P(x_1)P(x_2) \cdots P(x_m) \log_2(P(x_1)P(x_2) \cdots P(x_m)) \\ &= - \sum_{(x_1, x_2, \dots, x_m) \in \mathcal{A}^m} P(x_1)P(x_2) \cdots P(x_m) \sum_{l=1}^m \log_2 P(x_l) \\ &= - \sum_{l=1}^m \sum_{(x_1, x_2, \dots, x_m) \in \mathcal{A}^m} P(x_1)P(x_2) \cdots P(x_m) \log_2 P(x_l) \\ &= - \sum_{l=1}^m \sum_{x_l \in \mathcal{A}} P(x_l) \log_2 P(x_l) \\ &= \sum_{l=1}^m \mathcal{H}(\Pi) = m\mathcal{H}(\Pi). \quad \triangle \end{aligned} \quad (1.33)$$

In particolare, nel caso della sorgente binaria la distribuzione della sorgente originaria è $\Pi = \{p, (1-p)\}$. La sorgente estesa con $m = 2$ ha la distribuzione di probabilità

$$\Pi^2 = \{p^2, p(1-p), (1-p)p, (1-p)^2\}. \quad (1.34)$$

L'entropia è pertanto

$$\mathcal{H}(\Pi^2) = 2\mathcal{H}(p) = 2(-p \log_2 p - (1-p) \log_2 (1-p)). \quad (1.35)$$

Il calcolo generale della distribuzione di probabilità per la sorgente estesa di ordine m a simboli indipendenti è facilmente ottenuta in forma compatta

1.4. SEQUENZE TIPICHE

25

se si organizzano le probabilità in un vettore. Infatti è immediato osservare come il vettore delle probabilità della sorgente estesa si ottenga come

$$\pi^m = \underbrace{\pi \otimes \pi \otimes \dots \otimes \pi}_m, \quad (1.36)$$

dove il simbolo \otimes indica il prodotto di Kronecker definito in Appendice B.

Si noti che la sorgente estesa ha una cadenza che è ridotta di un fattore m , ovvero un simbolo della sorgente estesa viene emesso ogni m simboli della sorgente non estesa. Comunque, anche se cambia la frequenza di simbolo, sia il bit-rate che la velocità di informazione restano invariati

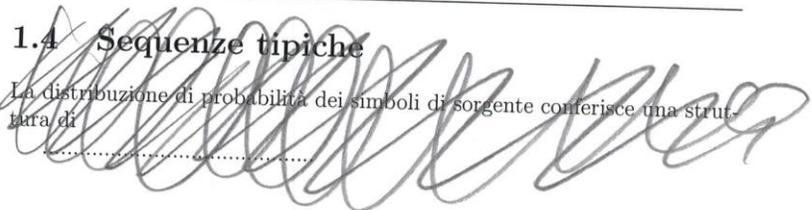
$$B_r^m = \frac{\log_2 n^m}{mT} = \frac{\log_2 n}{T} = B_r, \quad (1.37)$$

$$R_s^m = \frac{\mathcal{H}(\Pi^m)}{mT} = \frac{m\mathcal{H}(\Pi)}{mT} = \frac{\mathcal{H}(\Pi)}{T} = R_s. \quad (1.38)$$

Esempio 1.2 Si consideri una sorgente ternaria con alfabeto sorgente $\mathcal{A} = \{0, 1, *\}$ che emette simboli indipendenti. La distribuzione di probabilità sia $\Pi = \{0.4, 0.4, 0.2\}$. L'entropia della sorgente è pertanto $\mathcal{H}(\Pi) = -0.4 \log_2 0.4 - 0.4 \log_2 0.4 - 0.2 \log_2 0.2 = 1.52$ bit. La entropia di una sorgente estesa con $m = 5$ è $\mathcal{H}(\Pi^5) = 7.61$ bit. Se i simboli della sorgente vengono emessi a cadenza regolare ogni $T = 10$ ms, la frequenza di simbolo della sorgente è $R_s = 100$ simboli/sec. La frequenza di simbolo della sorgente estesa è $R_{s5} = \frac{R_s}{5} = 20$ simboli/sec. La velocità di informazione $R = \frac{1.52}{10^{-2}} = 152$ bit/sec. Si noti come la velocità di informazione non cambi nella sorgente estesa in quanto nonostante l'entropia diventi 5 volte quella originaria, la cadenza $1/T$ si riduce anch'essa di un fattore 5.

1.4 Sequenze tipiche

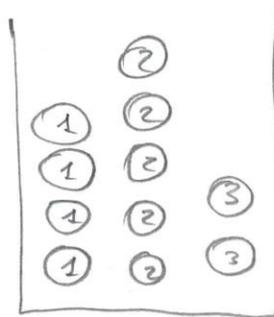
La distribuzione di probabilità dei simboli di sorgente conferisce una struttura di



RICAVIAMO LA FORMULA DELL'ENTROPIA PALTIERRELLY
DA UN ESPERIMENTO:

Esperimento di Wallis

Supponiamo di avere un'urna con N palle.
 le palle sono di 3 tipi: 1, 2 e 3 (magari
 con versi e gialle)



Nella figura
 $\left. \begin{array}{l} N_1 = 4 \\ N_2 = 5 \\ N_3 = 2 \end{array} \right\} N = 11$

Le ne sono N_1 di tipo 1, N_2 di tipo 2 e N_3 di tipo 3.
 Ovviamente $N_1 + N_2 + N_3 = N$.

Supponiamo ora di estrarre le palle una a una e metterle in fila una dopo l'altra

Es. ① ② ① ③ ② ② ③ ① ② ① ②

In quanti modi possiamo disporre le palle?

$$W = \binom{N}{N_1, N_2, N_3} = \frac{N!}{N_1! N_2! N_3!}$$

coeff. multinomiale

La "ricchezza" del campione dipende dalla sua
 molteplicità. Pertanto consideriamo la quantità

$$\frac{\log W}{N}$$

e in particolare il suo limite al crescere di N .

$$\lim_{N \rightarrow \infty} \frac{\log W}{N}$$

Pr. L'VIII. 15

Definiamo i parametri

$$P_1 = \frac{N_1}{N}, \quad P_2 = \frac{N_2}{N}, \quad P_3 = \frac{N_3}{N}$$

$$P_1 + P_2 + P_3 = 1 \quad 0 \leq P_i \leq 1 \quad i=1,2,3$$

e ricordiamo che una utile approssimazione al fattoriale è la formula di Stirling

$$\log n! = n \log n - n + O(\log n)$$

infinitesimo

Quindi

$$\lim_{N \rightarrow \infty} \frac{\log W}{N} = \lim_{N \rightarrow \infty} \frac{\log N! - \sum_{i=1}^3 \log(N P_i)!}{N}$$

$$= \lim_{N \rightarrow \infty} \frac{N \log N - N + O(\log N) - \left[\sum_{i=1}^3 (N P_i \log N P_i - N P_i + O(\log N P_i)) \right]}{N}$$

$$= \lim_{N \rightarrow \infty} \log N - 1 + \frac{O(\log N)}{N} - \left[\sum_{i=1}^3 (P_i \log N P_i - P_i + \frac{O(\log N P_i)}{N}) \right]$$

= 1

$$= \lim_{N \rightarrow \infty} \left(\log N - \sum_{i=1}^3 P_i \log N P_i \right)$$

$$= \lim_{N \rightarrow \infty} \left(\sum_{i=1}^3 P_i \log N - \sum_{i=1}^3 P_i \log N P_i \right) = \sum_{i=1}^3 P_i \log \frac{1}{P_i}$$

= $H(\pi)$

Quindi l'entropia è il limite per $N \rightarrow \infty$ del parametro di completezza del campione. Questa osservazione conferisce all'entropia un senso fisico che va ben oltre l'approccio orientativo delle definizioni riportate in precedenza.

SEQUENZE TIPICHE

PALMERI.16

La struttura di probabilità non uniforme dei simboli di sorgenti conformi alle stringhe una struttura particolare poiché alcuni prodotti di probabilità sempre più piccoli tendono a diventare irrilevanti. Vediamo un esempio:

Si consideri la sorgente binaria con $\mathcal{R} = \{a_1, a_2\}$ a simboli indipendenti e distribuzione di probabilità $\pi = \{0.3, 0.7\}$. Valutiamo le probabilità delle stringhe per ordini crescenti:

$$\begin{aligned} \pi &= \{0.3000 \quad 0.7000\} \\ \mathcal{R}^2 &= \mathcal{R} \times \mathcal{R} = \{a_1^2, a_1 a_2, a_2 a_1, a_2^2\} \\ \pi &= \{0.0900 \quad 0.2100 \quad 0.2100 \quad 0.4900\} \\ \mathcal{R}^3 &= \mathcal{R} \times \mathcal{R} \times \mathcal{R} = \{a_1^3, a_1^2 a_2, \dots, a_2^3\} \\ \pi &= \{0.0270 \quad 0.0630 \quad 0.0630 \quad 0.1470 \quad 0.0630 \quad 0.1470 \quad 0.1470 \quad 0.3430\} \\ \mathcal{R}^4 &= \mathcal{R} \times \mathcal{R} \times \mathcal{R} \times \mathcal{R} = \{a_1^4, a_1^3 a_2, \dots, a_2^4\} \\ \pi &= \{0.0081 \quad 0.0189 \quad 0.0189 \quad 0.0441 \quad 0.0189 \quad 0.0441 \quad 0.0441 \quad 0.1029 \quad 0.0189 \quad 0.0441 \\ &\quad 0.0441 \quad 0.1029 \quad 0.0441 \quad 0.1029 \quad 0.1029 \quad 0.2401\} \\ \mathcal{R}^5 &= \mathcal{R} \times \mathcal{R} \times \mathcal{R} \times \mathcal{R} \times \mathcal{R} = \{a_1^5, a_1^4 a_2, \dots, a_2^5\} \\ \pi &= \{0.0024 \quad 0.0057 \quad 0.0057 \quad 0.0132 \quad 0.0057 \quad 0.0132 \quad 0.0132 \quad 0.0309 \quad 0.0057 \quad 0.0132 \quad 0.0132 \quad 0.0309 \\ &\quad 0.0132 \quad 0.0309 \quad 0.0309 \quad 0.0720 \quad 0.0057 \quad 0.0132 \quad 0.0132 \quad 0.0309 \quad 0.0132 \quad 0.0309 \quad 0.0309 \quad 0.0720 \\ &\quad 0.0132 \quad 0.0309 \quad 0.0309 \quad 0.0720 \quad 0.0309 \quad 0.0720 \quad 0.0720 \quad 0.1681\} \end{aligned}$$

⋮

Si noterà come alcune probabilità vengano diventate molto piccole rispetto alle altre. Questo non sarebbe avvenuto se i simboli fossero stati uniformi ($\pi = \{\frac{1}{2}, \frac{1}{2}\}$). Anche una leggera non uniformità porta ad una struttura molto articolata delle probabilità.

Le stringhe con probabilità non piccola, si denominano
come sequenze tipiche. PALMERI.17

L'intuizione sulla esistenza delle sequenze tipiche è ancora
una volta dovuta alla generalità di Claude Shannon
che formula la proprietà di equipartizione asintotica
(AEP) di cui parleremo dopo.

Esempio (Totocalcio)

Sia $\mathcal{A} = \{1, x, 2\}$ e supponiamo che le partite siano
indipendenti e che abbiano come distribuzione di
probabilità $\pi = \{0.4, 0.4, 0.2\}$

Le sequenze tipiche saranno nella stringa di 13
simboli quelle che hanno una probabilità non
troppo piccola. Ovviamente le sequenze che contengono
molti 2 non sono tipiche.

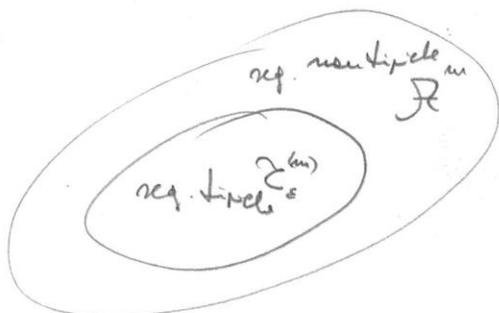
L'entropia è $H(\pi) = 1.5219$ bit. Poiché le
partite sono indipendenti $H(\pi^{13}) = 13 H(\pi)$ ($n=13$)

Ora ci si chiede: quante sono le sequenze tipiche?
La risposta è ovviamente collegata al contenuto
informativo medio della sorgente (entropia) che
dovrebbe fornire la dimensionalità essenziale
dell'insieme tipico. La cosa interessante è che

$$\# \text{seq. tipiche} \approx 2^{n H(\pi)}$$

ovvero l'entropia della sorgente estesa $n H(\pi)$
è quasi approssimativamente il # di seq. tipiche
espresso in bit.

La cosa interessante che emerge dalle proprietà della equipartizione asintotica ^(AEP) è che le sequenze tipiche non differiscono di molto in probabilità e che questa è circa $2^{-nH(\pi)}$. PA. LITIERI. 18



Questo significa che le seq. non tipiche non sono molto "interessanti" perché si verificano raramente.

Nell'esempio del telecalcio abbiamo che dall'AEP le sequenze tipiche sono all'incirca

$$|\Sigma^{(n)}| \approx 2^{13H(\pi)} = 887876$$

contro

$$|\mathcal{R}^{13}| = 3^{13} = 1594323$$

che rappresenta il numero complessivo di sequenze ternarie di lunghezza 13.

(Sfortunatamente le sequenze tipiche sono ancora troppe per poterle giocare tutte!!)

LA PROPRIETA' DI EQUIPARTIZIONE ASINTOTICA AEP₁

(AEP)

Palmerini 19

Teorema (AEP) Se S_1, S_2, \dots sono iid $\sim p(s)$, allora

$$-\frac{1}{n} \log p(S_1, S_2, \dots, S_n) \rightarrow H(S) \quad \text{in probabilità}$$

Prova

Funzioni di v.a. indipendenti, sono indipendenti. Quindi:
 Probabilità $p(S_1, S_2, \dots, S_n) = \prod_{i=1}^n p(S_i)$

$$-\frac{1}{n} \log p(S_1, S_2, \dots, S_n) = -\frac{1}{n} \sum_{i=1}^n \log p(S_i)$$

$\rightarrow -E[\log p(S)]$ in probabilità per la legge debole dei grandi numeri

$$= H(S) \quad \square$$

Legge debole dei grandi numeri:

$$\frac{1}{n} \sum_{i=1}^n Y_i \rightarrow E[Y] \quad \text{in probabilità.}$$

Definizione: l'insieme tipico $\mathcal{Z}_\epsilon^{(n)}$ rispetto a $p(s)$, è l'insieme delle sequenze $(s_1, s_2, \dots, s_n) \in \mathcal{R}^n$ con la proprietà che

$$2^{-n(H(S)+\epsilon)} \leq p(s_1, s_2, \dots, s_n) \leq 2^{-n(H(S)-\epsilon)}$$

Le sequenze sono dette ϵ -tipiche

PROPRIETA' di $\mathcal{Z}_\epsilon^{(n)}$

1. $H(S) - \epsilon \leq -\frac{1}{n} \log p(s_1, s_2, \dots, s_n) \leq H(S) + \epsilon$
2. $P_n \{ \mathcal{Z}_\epsilon^{(n)} \} > 1 - \epsilon$ per n sufficientemente grande
3. $|\mathcal{Z}_\epsilon^{(n)}| \leq 2^{n(H(S)+\epsilon)}$ $|\mathcal{Z}_\epsilon^{(n)}|$ numero degli elementi di $\mathcal{Z}_\epsilon^{(n)}$
4. $|\mathcal{Z}_\epsilon^{(n)}| \geq (1-\epsilon) 2^{n(H(S)-\epsilon)}$ per n sufficientemente grande

PROVE

AEP2

1. Immediato dalla definizione prendendo il \log_2 polinomio

2. Dimende direttamente dal Lemma (AEP).

Traffetti se $-\frac{1}{n} \log_2(s_1^{(n)} \dots s_n^{(n)}) \rightarrow H(s)$ in probabilità, vuol dire che

$$\forall \delta > 0 \exists n_0 : n > n_0 \quad \mathbb{P} \left\{ \left| -\frac{1}{n} \log_2(s_1^{(n)} \dots s_n^{(n)}) - H(s) \right| < \delta \right\} > 1 - \delta$$

Per mettere $\delta = \epsilon$

$$\Rightarrow \mathbb{P} \left\{ \mathcal{Z}_\epsilon^{(n)} \right\} > 1 - \epsilon \quad \square$$

3.

$$1 = \sum_{(s_1, \dots, s_n) \in \mathcal{R}^n} P(s_1^{(n)} \dots s_n^{(n)})$$

$$\underbrace{s[1], s[2], \dots, s[m]}$$

$$\geq \sum_{(s_1, \dots, s_n) \in \mathcal{Z}_\epsilon^{(n)}} P(s_1^{(n)} \dots s_n^{(n)})$$

$$\sum_{s=1}^m$$

$$\geq \sum_{(s_1, \dots, s_n) \in \mathcal{Z}_\epsilon^{(n)}} 2^{-n(H(s)+\epsilon)}$$

$$= 2^{-n(H(s)+\epsilon)} |\mathcal{Z}_\epsilon^{(n)}|$$

$$\Rightarrow |\mathcal{Z}_\epsilon^{(n)}| \leq 2^{n(H(s)+\epsilon)} \quad \square$$

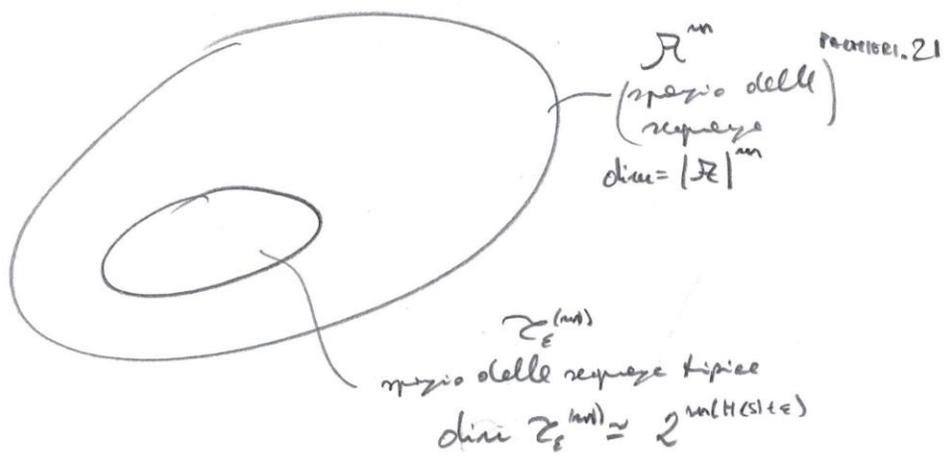
4. Pericoloso per n sufficientemente grande

$$\mathbb{P} \left\{ \mathcal{Z}_\epsilon^{(n)} \right\} > 1 - \epsilon \quad (\text{Proprietà 2.}) \quad \text{e}$$

$$1 - \epsilon < \mathbb{P} \left\{ \mathcal{Z}_\epsilon^{(n)} \right\} \leq \sum_{(s_1, \dots, s_n) \in \mathcal{Z}_\epsilon^{(n)}} P(s_1^{(n)} \dots s_n^{(n)}) \leq \sum_{(s_1, \dots, s_n) \in \mathcal{Z}_\epsilon^{(n)}} 2^{-n(H(s)-\epsilon)}$$

$$= 2^{-n(H(s)-\epsilon)} |\mathcal{Z}_\epsilon^{(n)}|$$

$$|\mathcal{Z}_\epsilon^{(n)}| \geq (1 - \epsilon) 2^{n(H(s)-\epsilon)} \quad \square$$



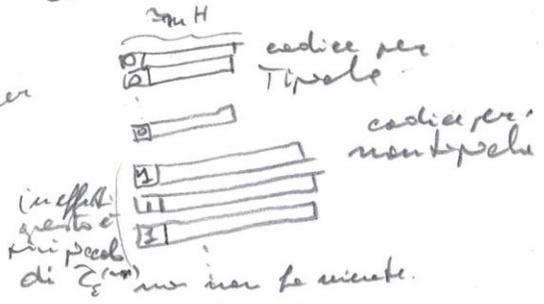
Codifica di sorgente:

Senza invocare il Teorema di Shannon, è possibile inventare una semplice codifica variabilemente ottima. È altrettanto incredibile come sia possibile facilmente trovare un codice variabilemente ottimo usando solo il concetto di "Determinare tutti gli elementi appartenenti a $Z_\epsilon^{(n)}$ ".

Diamo un indice ad ogni s . Poiché ce ne sono un numero $\leq 2^{n(H(S) \pm \epsilon)}$, l'indicizzazione richiede $n(H \pm \epsilon) + 1$ bit (+1 perché potrebbe non essere un intero). Ora per rendere la risposta identificabile includiamo un bit ~~alla fine di ogni~~ ~~stringa~~ ~~stringa~~.

Analogamente possiamo indicizzare con numero di $n \lg |R| + 1$ bit tutte le altre risposte non tipiche.

in effetti le risposte non tipiche non di meno, ma facciamo i conti per eccome



Se $l(s^{(n)})$ la lunghezza del codice per $s^{(n)}$
 Se n è sufficientemente grande $P_r \{ Z_\epsilon^{(n)} \} \geq 1 - \epsilon$

$$E[l(s^{(n)})] = \sum_{s^{(n)}} l(s^{(n)}) p(s^{(n)}) = \sum_{s^{(n)} \in Z_\epsilon^{(n)}} l(s^{(n)}) p(s^{(n)}) + \sum_{s^{(n)} \in Z_\epsilon^{(n)c} } l(s^{(n)}) p(s^{(n)})$$

$$\begin{aligned}
 & \leq \sum_{s^m \in \mathcal{Z}_\epsilon^{(m)}} p(s^m) (n(H(s) + \epsilon) + 2) + \sum_{s^m \in \mathcal{Z}_\epsilon^{(m)c}} (n \log |A| + 2) p(s^m) \quad \text{palmieri.22} \\
 & = P_Z \{ \mathcal{Z}_\epsilon^{(m)} \} (n(H(s) + \epsilon) + 2) + P_Z \{ \mathcal{Z}_\epsilon^{(m)c} \} (n \log |A| + 2) \\
 & \quad \left[1 - P_Z \{ \mathcal{Z}_\epsilon^{(m)} \} \right] \\
 & = \underbrace{P_Z \{ \mathcal{Z}_\epsilon^{(m)} \}}_{\leq 1} [n(H(s) + \epsilon) + 2] + P_Z \{ \mathcal{Z}_\epsilon^{(m)c} \} (n \log |A| + 2) \\
 & \quad + \underbrace{P_Z \{ \mathcal{Z}_\epsilon^{(m)} \} 2 + P_Z \{ \mathcal{Z}_\epsilon^{(m)c} \} 2}_{2} \\
 & \leq n(H(s) + \epsilon) + \epsilon n \log |A| + 2 \\
 & \quad \epsilon' = \epsilon + \epsilon \log |A| + \frac{2}{n} \quad \text{po-essere vero} \\
 & \quad = n(H(s) + \epsilon') \quad \text{relativamente piccolo} \\
 & \quad \quad \quad \text{negliendo } n.
 \end{aligned}$$

per tanto: $\frac{1}{n} E[\ell(s^m)] \rightarrow H(s)$

Teorema

Sia S^m una sequenza iid $\times p(s)$. Sia $\epsilon > 0$.
 Esiste un codice che mappa sequenze S^m di lunghezza m in stringhe binarie tali che la codifica è
 invertibile e che

$$E\left[\frac{1}{n} \ell(S^m)\right] \leq H(S) + \epsilon$$

per n sufficientemente grande.

Quindi la lunghezza media per simbolo po-essere
 resa arbitrariamente vicina all'entropia.

CONVERGENZA DI VARIABILI ALEATORIE

PAG. 23

Dato una sequenza di v.a. X_1, X_2, \dots , diremo che X_1, X_2, X_3, \dots converge a X in probabilità se

$$\forall \varepsilon > 0 \quad \Pr \{ |X_n - X| > \varepsilon \} \rightarrow 0 \text{ per } n \rightarrow \infty$$

ovvero, usando la definizione di limite di una sequenza

$$\forall \varepsilon > 0, \delta > 0 \quad \exists n_0 : n > n_0 \quad \Pr \{ |X_n - X| > \varepsilon \} < \delta$$

Diremo che una sequenza di v.a. X_1, X_2, \dots converge a X in media quadratica se

$$E[(X_n - X)^2] \rightarrow 0, \text{ ovvero}$$

$$\forall \delta > 0 \quad \exists n_0 > 0 : n > n_0 \quad E[(X_n - X)^2] < \delta$$

Una sequenza di v.a. X_1, X_2, \dots converge a X con probabilità uno, o quasi certamente (a.e. almost surely) se

$$\Pr \{ \lim_{n \rightarrow \infty} X_n = X \} = 1$$

Questo significa che tutte le realizzazioni della sequenza tendono a X eccetto qualunqua isolata.

Convergenza di una successione $X_n \rightarrow X$

$$\lim_{n \rightarrow \infty} X_n = X$$

$$\forall \varepsilon > 0 \quad \exists n_0 : n > n_0 \quad |X_n - X| < \varepsilon$$



PALMIERI-24

SORSENTI DISCRETE
CON MEMORIA

Dato una sequenza aleatoria $\{S[k]\}$ con $S[k] \in \mathcal{R} = \{a_1, \dots, a_n\}$

$$\{S\} \rightarrow S[k] \rightarrow$$

PALMIGRANO 25 (PDA)

vogliamo considerare il caso generale di simboli dipendenti.

$$-\infty < k < +\infty$$

$$\underbrace{S[k-m+1] S[k-m] \dots S[k-2] S[k-1] S[k]}_{S^m[-1]} \underbrace{S[k+1] S[k+2] \dots S[k+m-1] S[k+m]}_{S^m[1]}$$

$k = ml + i, \quad i = 1, \dots, m-1$

$$S^m[l] \in \mathcal{R}^m = \underbrace{\mathcal{R} \times \mathcal{R} \times \dots \times \mathcal{R}}_m \quad l = -\infty, \dots, +\infty$$

con $k = ml + i, \quad i = 1, \dots, m-1$

$$P_e(S^m[l]) = P_e\{S[k+l]=s_1, S[k+2+l]=s_2, \dots, S[k+m+l]=s_m\}$$

↑ in genere dipende da l

Se il processo è tale che

$$P_e(S^m[l]) = P(S^m[0]) = p(s_1, s_2, \dots, s_m)$$

il processo è stazionario (in senso stretto)
 si può pertanto scrivere l'entropia congiunta

$$H(S^m) = H(S[1], \dots, S[m]) = \sum_{(s_1, \dots, s_m) \in \mathcal{R}^m} p(s_1, \dots, s_m) \log \frac{1}{p(s_1, \dots, s_m)}$$

Quindi (per un processo stazionario) l'entropia del processo, o entropy rate, o cross entropy è

$$H(S) = \lim_{m \rightarrow \infty} \frac{1}{m} H(S^m) \quad \text{bit/simbolo}$$

Caso particolare, il processo i.i.d., ovvero a simboli indipendenti.

$$H(S^m) = m H(S)$$

$$\Rightarrow H(S) = H(S)$$

Altra quantità di interesse è

(PDZ)

$$H_{\infty}(S) = \lim_{m \rightarrow \infty} H(S[m] | S[m-1], \dots, S[1]) \quad \text{POLYMERI. 26}$$

(che dimostreremo essere non coincidente con $H(S)$)

Poiché $H(S[2] | S[1]) \leq H(S[2])$ (*) con l'ipotesi che $S[2]$ e $S[1]$ sono indipendenti

Iterando si ottiene che

$$\begin{aligned} H(S[m] | S[m-1], \dots, S[1]) &\leq H(S[m] | S[m-1], \dots, S[2]) \\ &= H(S[m-1] | S[m-2], \dots, S[1]) \end{aligned}$$

per stazionarietà

Quindi la sequenza

$H(S[m] | S[m-1], \dots, S[1])$ è non crescente e poiché è positiva, il limite $H_{\infty}(S)$ esiste.

In generale $H_{\infty}(S) \leq H(S[i])$

TEOREMA: Per un processo stazionario $H(S)$ e $H_{\infty}(S)$ esistono e sono coincidenti

$$H(S) = H_{\infty}(S)$$

Prova l'esistenza di $H(S)$ e $H_{\infty}(S)$ è stata già svolta.

Per la regola a catena

$$\frac{H(S[1], \dots, S[m])}{m} = \frac{1}{m} \sum_{i=1}^m H(S[i] | S[i-1], \dots, S[1])$$

(*) Facilmente inferibile, ma dimostrabile immediatamente

$$\begin{aligned} H(S[2] | S[1]) - H(S[2]) &= \sum_{s_1, s_2} p(s_1, s_2) \log_2 \frac{p(s_2)}{p(s_2 | s_1)} && \log_e x = \log_e (x-1) \\ &\leq \log_2 e \sum_{s_1, s_2} p(s_1, s_2) \left(\frac{p(s_2)}{p(s_1 | s_2)} - 1 \right) = 0 && \ln x < x - 1 \end{aligned}$$

~~Teorema della media di Cesàro (da Anellini)~~

(PD3)

Se $a_n \rightarrow a$ e $b_n = \frac{1}{n} \sum_{i=1}^n a_i$, $b_n \rightarrow a$

PALMIERI - 27

Perché sappiamo che $H(S_{t+1} | S_{t-1}, \dots, S_{t-1}) \rightarrow H_{\infty}(S)$
per definizione, il limite a destra tende a $H_{\infty}(S)$
e quello a sinistra a $H(S)$. \square

È possibile estendere l'AEP ai processi dipendenti, infatti
per un processo ergodico e stazionario

$$-\frac{1}{n} \log p(S_{t+1}, \dots, S_{t+n}) \rightarrow H_{\infty}(S) \text{ in probabilità.}$$

Questo significa che a meno di ϵ di $H_{\infty}(S)$ si trova il tipo

di lunghezza n , oppure con probabilità di errore

di ϵ di $H_{\infty}(S)$. Quindi si trova ^{di lunghezza n} ~~caratterizzate~~

da $H_{\infty}(S)$ con n sufficientemente grande,
codificabili con $n H_{\infty}(S) \pm \epsilon$ bit circa.

Perché in generale $H_{\infty}(S) < H(S_{t+1}) \quad \forall i$

La risposta dipende perché non bit che la risposta
indipendente aveva l'entropia marginale.

PROCESSI DI MARKOV

(PDL)

Un processo $\{S[k]\}$ è un Processo di Markov se PA-LT11011.28

$$P_r \{ S[k+1] = s_{k+1} | S[k] = s_k, S[k-1] = s_{k-1}, \dots, S[1] = s_1 \}$$

$$= P_r \{ S[k+1] = s_{k+1} | S[k] = s_k \}$$

$\forall k \in \mathbb{Z}^+ \forall s_1, \dots, s_{k+1} \in \mathcal{S}$ è l'alfabeto \mathcal{S} è discreto (come nel nostro caso)
in parole di CATENA DI MARKOV

Quindi la prob. congiunta per le uscite a catena si scrive

$$P_r \{ S[k] = s_k, S[k-1] = s_{k-1}, \dots, S[1] = s_1 \}$$

$$= P_r \{ S[k] = s_k | S[k-1] = s_{k-1}, \dots, S[1] = s_1 \} P_r \{ S[k-1] = s_{k-1} | S[k-2] = s_{k-2}, \dots, S[1] = s_1 \}$$

$$\dots P_r \{ S[2] = s_2 | S[1] = s_1 \} P_r \{ S[1] = s_1 \}$$

diventa

$$= P_r \{ S[k] = s_k | S[k-1] = s_{k-1} \} P_r \{ S[k-1] = s_{k-1} | S[k-2] = s_{k-2} \} \dots$$

$$\dots P_r \{ S[2] = s_2 | S[1] = s_1 \} P_r \{ S[1] = s_1 \}$$

$$= P_r \{ S[1] = s_1 \} \prod_{i=0}^{k-2} P_r \{ S[k-i] = s_{k-i} | S[k-i-1] = s_{k-i-1} \}$$

DEFINIZIONE

La catena di Markov si dice stazionaria o omogenea

$$P_r \{ S[k+1] = \alpha | S[k] = \beta \} = P_r \{ S[2] = \alpha | S[1] = \beta \}$$

$$\forall k \text{ e } \forall \alpha, \beta \in \mathcal{S}$$

Quindi una catena di Markov omogenea è totalmente caratterizzata da una distribuzione dello stato iniziale $P_r \{ S[1] = s \}$, $s \in \mathcal{S}$, e dalle probabilità condizionate

$$P_r \{ S[k+1] = \alpha | S[k] = \beta \} \quad \forall (\alpha, \beta) \in \mathcal{S}^2$$

gli elementi dell'alfabeto Σ sono anche detti stati della catena.

PDS

PARADISI. 29

Utile dal punto di vista computazionale rappresentare le probabilità dello stato iniziale con un vettore e quelle di transizione dello stato con una matrice

$$\underline{p} = \begin{bmatrix} P\{S[1]=a_1\} \\ P\{S[1]=a_2\} \\ \vdots \\ P\{S[1]=a_n\} \end{bmatrix}$$

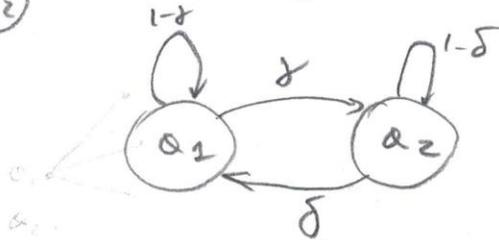
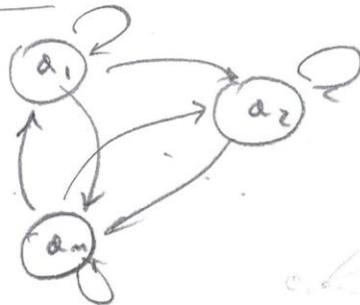
$$\underline{P} = \begin{bmatrix} P\{S[k+2]=a_1 | S[k]=a_1\} & \dots & P\{S[k+2]=a_n | S[k]=a_1\} \\ P\{S[k+1]=a_1 | S[k]=a_1\} & \dots & P\{S[k+1]=a_n | S[k]=a_1\} \\ \vdots & \ddots & \vdots \\ P\{S[k+1]=a_1 | S[k]=a_m\} & \dots & P\{S[k+1]=a_n | S[k]=a_m\} \end{bmatrix}$$

Sono possibili 3 rappresentazioni grafiche

Grafo di stato

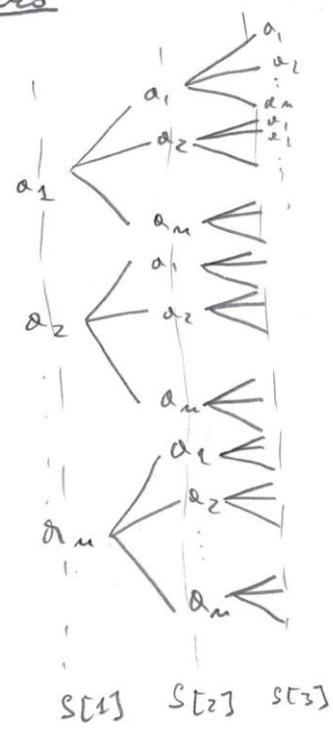
ESEMPIO

$$P = \begin{bmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{bmatrix}$$

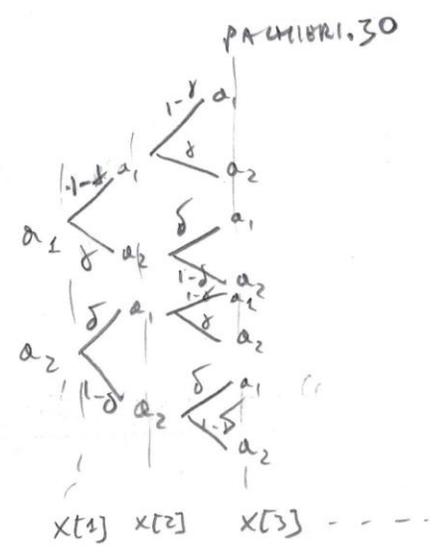


Talbot

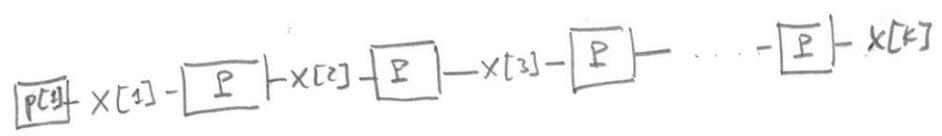
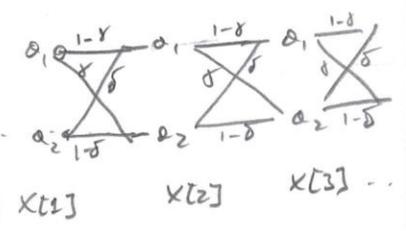
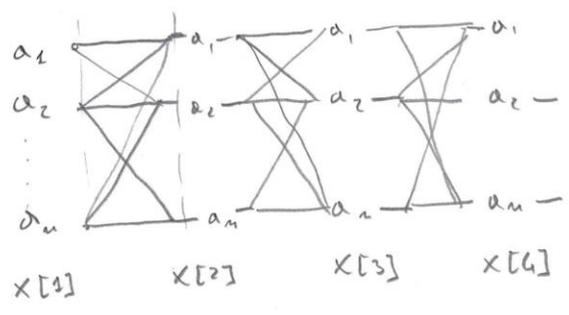
Albers



erweiterte
erweiterte



TRELLIS

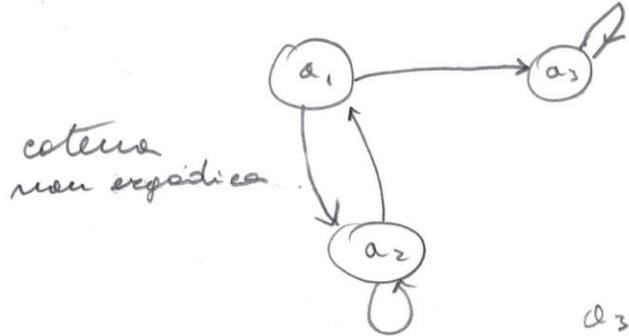


DEFINIZIONE

Una catena di Markov si dice irriducibile se da ogni stato è possibile raggiungere qualunque altro stato in un numero finito di passi.

o ergodica (PDF)
 PUNTO 31

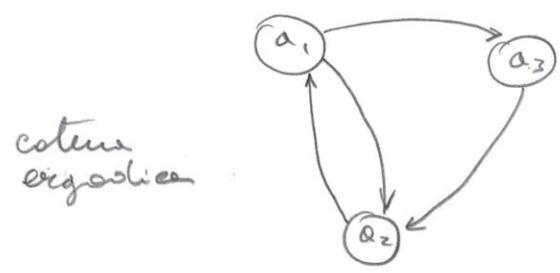
Esempi:



$$P = \begin{bmatrix} 0 & x & x \\ x & x & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

a_3 è uno stato assorbente

⇒ una volta ottenuto non se ne esce più.



$$P = \begin{bmatrix} 0 & x & x \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Per conoscere la distribuzione dello stato al tempo $k+1$,

$$P\{S^{[k+1]} = s_{k+1}\} = \sum_{s_k \in \mathcal{X}} P\{S^{[k+1]} = s_{k+1} | S^{[k]} = s_k\} P\{S^{[k]} = s_k\}$$

in forma matriciale

$$\underline{P}^{[k]} \triangleq \begin{bmatrix} P\{S^{[k]} = a_1\} \\ P\{S^{[k]} = a_2\} \\ \vdots \\ P\{S^{[k]} = a_n\} \end{bmatrix}$$

$$\underline{P}^{[k+1]} = \underline{P}^T \underline{P}^{[k]}$$

$$\underline{P}^{[2]} = \underline{P}^T \underline{P}^{[1]} ; \underline{P}^{[3]} = \underline{P}^T \underline{P}^{[2]} = \underline{P}^{T^2} \underline{P}^{[1]} \dots$$

$$\underline{P}^{[k]} = (\underline{P}^T)^{k-1} \underline{P}^{[1]}$$

nello stesso vettore $\underline{P}^{[k]}$
 $\underline{P}^{[k+1]} = \underline{P}^{[k]} \underline{P}^T$
 $(\dots) (\dots) (\dots)$

Alcune le proprietà "dinamiche" della catena sono determinate dalle proprietà delle potenze di P^T (PDB) P^T 32

DEFINIZIONE

Una catena di Markov si dice regolare, se per qualche k , $(P^T)^k$ contiene solo elementi > 0 .

Questa definizione significa che per qualche k è possibile andare da qualunque stato a qualunque altro stato esattamente in k passi.

Ovviamente catena regolare \Rightarrow catena ergodica

Esempio

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$



catena ergodica

regolare?

$$P^2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad P^{2k} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad k \text{ pari}$$

$$P^3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad P^{2k+1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad k \text{ dispari}$$

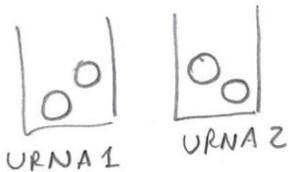
Quindi non è regolare perché per nessun k tutti gli elementi sono > 0 .

Esercizio

URNA DI
EHRENFEST

usato per
studiare fenomeni di diffusione

PD4



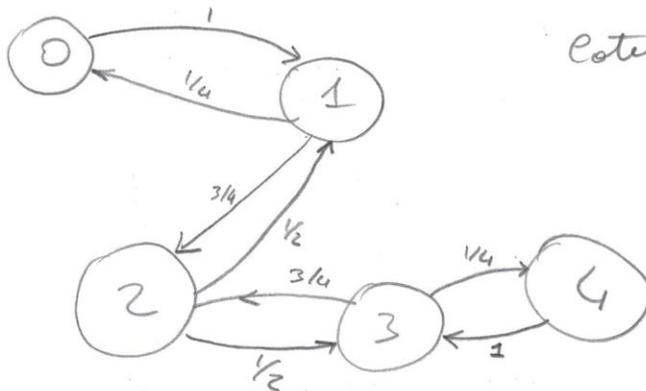
in totale ci sono 4
palle.

PABRIGERIS

Ad ogni passo preleva una palla o caso tra le quattro
e la pone nell'altra urna.

$S[K]$ = # di palle nella prima urna

	$S[K+1]$				
$S[K]$	0	1	2	3	4
0	0	1	0	0	0
1	$1/4$	0	$3/4$	0	0
2	0	$1/2$	0	$1/2$	0
3	0	0	$3/4$	0	$1/4$
4	0	0	0	1	0



catena ergodica

In un numero pari di passi può solo andare
da uno stato ad uno stato che differisce di 2.
Lo stesso per gli indici dispari. Anche per
numeri potenza di $(P^{-1})^k$ ottengo tutti gli elementi > 0 .

Altro esempio

$$P = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}$$



non ergodica, non ricorrente

PD10

PAG. 34

Xolo

Chissà come qualunque catena con una P avente elementi tutti $\neq 0$ è ricorrente. Ci sono sistemi catene con P contenente zero che non ricorrono.

Esempio (Buffer o coda)

$S[K]$ stato del buffer al tempo K

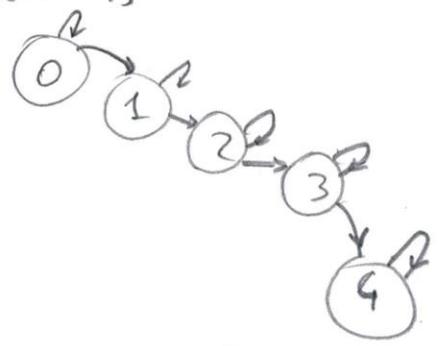
$$S[K] \in \{0, 1, 2, 3, \dots, n\}$$

$n = 4$

$S[K]$ stato iniziale

$$P_{ij} = \begin{cases} P_0 & \{S[K] = 0\} \\ P_1 & \{S[K] = 1\} \\ P_2 & \{S[K] = 2\} \\ P_3 & \{S[K] = n\} \end{cases}$$

	0	1	2	3	4
0	0.2	0.8			
1		0.2	0.8		
2			0.2	0.8	
3				0.2	0.8
4					1



catena non ergodica lo stato 4 è assorbente

$$S[K] = S[K-1] + W[K]$$

$W[K] \in \{0, 1\}$
nuovo arrivo
 $P_{ij}(w) = \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$

Piuttosto il processo di accesso alle casse continue anche delle partenze $w[k] \in \{-1, 0, 1\}$

PD10 bis

PAMIERI.35

$\pi_w = \{0.4, 0.2, 0.4\}$

	0	1	2	3	
0	0.6	0.4			
1	0.4	0.2	0.4		
2		0.4	0.2	0.4	
3			0.4	0.2	0.4
4				0.4	0.6

processo di nascita e morte



catena ergodica e regolare.

ci interessa molto studiare questo processo perché potremo chiederci: qual'è la prob. che la cassa è vuota? Qual'è la prob. che la cassa è vuota?

in generale

$$S[k] = S[k+1] + i[k]$$

n stati
 $\{0, 1, \dots, n\}$

$$i[k] \in \{-1, 0, 1\}$$

$$\pi_i = \left\{ \frac{1-p_0}{2}, p_0, \frac{1-p_0}{2} \right\}$$

$$P = \begin{bmatrix} \frac{p_0+1}{2} & \frac{1-p_0}{2} & & & \\ \frac{1-p_0}{2} & p_0 & \frac{1-p_0}{2} & & \\ & & & \ddots & \\ 0 & \frac{1-p_0}{2} & p_0 & \frac{1-p_0}{2} & \\ & & & \frac{1-p_0}{2} & p_0+1 \end{bmatrix}$$

P simmetrica $p_0 + \frac{1-p_0}{2} = \frac{2p_0+1-p_0}{2}$
 $P = P^T$

PDI

STUDIO DELLA CATENA DI MARKOV

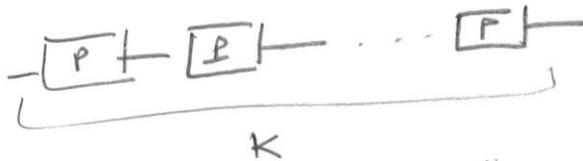
PALMIERI. 36

Primo di tutto notiamo che
 $(P^T)^k = (P^k)^T$ poiché

$$P^T P^T \dots P^T = (P \cdot P \dots P)^T$$

Quindi potremo studiare P^k piuttosto che $(P^T)^k$, come
suggerisce nella letteratura.

Di fatto P^k rappresenta la catena di k step



Domanda: come si comporta P^k per $k \rightarrow \infty$?

TEOREMA

Per una catena di Markov regolare

$$\lim_{k \rightarrow \infty} P^k = W = \begin{bmatrix} \frac{w_1}{w^T} \\ \frac{w_2}{w^T} \\ \vdots \\ \frac{w_n}{w^T} \end{bmatrix}$$

righe
uguali.
 w vettore distribuzione

è come dire
che la catena
di canali
ha un solo
canale con
righe uguali.
Nel caso con
il canale
non
converge
esattamente

Si può collegare la caratteristica di P ai suoi autovettori.
 In effetti se tutti gli autovettori di P abbiano
 lunghezza 1 non per $\lambda = 1$. Se in aggiunta $\lambda = 1$ è un
 autovettore semplice di P , la catena è piacamente
regolare (PD12)
PA 01/01/137

Perché per una catena regolare $P^k \rightarrow W$
 che non dipende da k , dobbiamo avere che
 simbolicamente

$$\underline{p}^{[00]} = \underline{P}^T \underline{p}^{[00]}$$

Andrà $\underline{p}^{[00]}$ può essere trovato come
 soluzione del sistema (*) con il vincolo

$$\underline{u} = \underline{P}^T \underline{u} \quad \sum_{i=1}^n u_i = 1 \quad 0 \leq u_i \leq 1$$

Perché

$$\underline{p}^{[00]} = \lim_{k \rightarrow \infty} \underline{p}^{[k]} = \lim_{k \rightarrow \infty} \underline{P}^{Tk} \underline{p}^{[1]}$$

$$= \underline{W}^T \underline{p}^{[1]} = \begin{bmatrix} w_1 & w_1 & w_1 \\ w_2 & w_2 & w_2 \\ \vdots & \vdots & \vdots \\ w_n & w_n & w_n \end{bmatrix} \begin{bmatrix} p^{[1]}_1 \\ p^{[1]}_2 \\ \vdots \\ p^{[1]}_n \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \underline{w}$$

otteniamo

$$\underline{P}^\infty = \begin{bmatrix} p^{[00]}_1 \\ p^{[00]}_2 \\ \vdots \\ p^{[00]}_n \end{bmatrix}$$

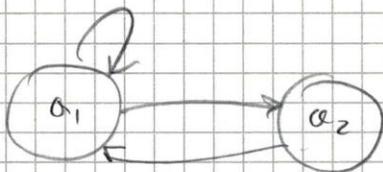
Breynes

81

PACHTER 38

$$P = \begin{bmatrix} 0.1 & 0.9 \\ 1 & 0 \end{bmatrix}$$

negative,
update



$$P[1] = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$$

$$P[2] = \begin{bmatrix} 0.1 & 1 \\ 0.9 & 0 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.55 \\ 0.45 \end{bmatrix}$$

$$P[3] = \begin{bmatrix} 0.1 & 1 \\ 0.9 & 0 \end{bmatrix} \begin{bmatrix} 0.55 \\ 0.45 \end{bmatrix} = \begin{bmatrix} 0.505 \\ 0.495 \end{bmatrix}$$

$$P[4] = \begin{bmatrix} 0.1 & 1 \\ 0.9 & 0 \end{bmatrix} \begin{bmatrix} 0.505 \\ 0.495 \end{bmatrix} = \begin{bmatrix} 0.5545 \\ 0.4455 \end{bmatrix}$$

$$P[5] = \begin{bmatrix} 0.1 & 1 \\ 0.9 & 0 \end{bmatrix} \begin{bmatrix} 0.5545 \\ 0.4455 \end{bmatrix} = \begin{bmatrix} 0.5095 \\ 0.4905 \end{bmatrix}$$

$$p[\infty] = P^T p[0]$$

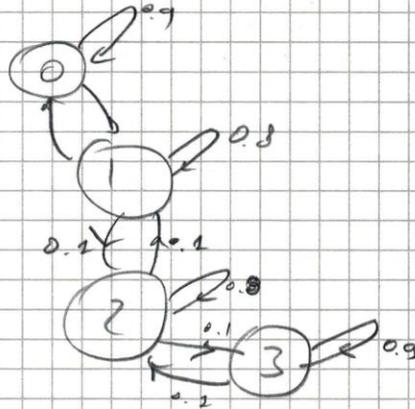
$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0.5 & 1 \\ 0.5 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad \begin{cases} u_1 = \frac{10}{19} \\ u_2 = \frac{9}{19} \end{cases}$$

$$u_1 + u_2 = 1$$

$$P^{\infty} = \begin{bmatrix} \frac{10}{19} & \frac{9}{19} \\ \frac{10}{19} & \frac{9}{19} \end{bmatrix}$$

Gruppo del buffer

$$P = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 \\ 0.1 & 0.8 & 0.1 & 0 \\ 0 & 0.1 & 0.8 & 0.1 \\ 0 & 0 & 0.1 & 0.9 \end{bmatrix}$$



$$u = P^T u$$

$$\begin{pmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 0.9 & 0.1 & 0 & 0 \\ 0.1 & 0.8 & 0.1 & 0 \\ 0 & 0.1 & 0.8 & 0.1 \\ 0 & 0 & 0.1 & 0.9 \end{pmatrix} \begin{pmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \end{pmatrix}$$

$$p[2] = P^T p[1] = \begin{pmatrix} 0.9 & 0.1 & 0 & 0 \\ 0.1 & 0.8 & 0.1 & 0 \\ 0 & 0.1 & 0.8 & 0.1 \\ 0 & 0 & 0.1 & 0.9 \end{pmatrix} \begin{pmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix} = \begin{pmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix}$$

quindi è uniforme $\forall k$.

\Rightarrow Uniforme in tutte le stati $p[\infty]$

Tutte le P che sono stocasticamente irreducibili, ovvero tali che P^T è stocastico, hanno la distribuzione uniforme.

Quelle est une générale simétrico

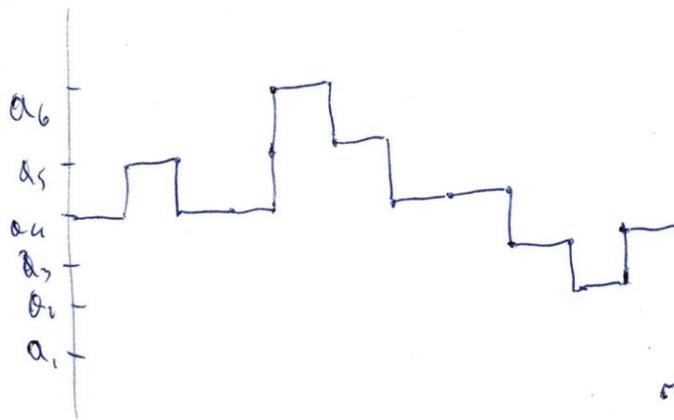
63

$$\begin{bmatrix}
 \frac{p_0+1}{2} & \frac{1-p_0}{2} & & & & \\
 \frac{1-p_0}{2} & p_0 & \frac{1-p_0}{2} & & & \\
 & \frac{1-p_0}{2} & p_0 & \frac{1-p_0}{2} & & \\
 & & \frac{1-p_0}{2} & p_0 & \frac{1-p_0}{2} & \\
 & & & \frac{1-p_0}{2} & p_0 & \frac{1-p_0}{2} \\
 & & & & \frac{1-p_0}{2} & p_0+1 \\
 & & & & & \frac{1-p_0}{2}
 \end{bmatrix}$$

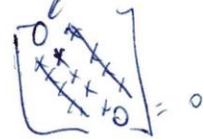
P+LMIR.40

matrice simétrico, qnd
dopionente d'ordico

$$P_{\infty} = \frac{1}{n} e \text{ uniforme.}$$



random walk
sur borniere.
La borniere pas zéro
reflectéte. = 0



ENTROPIA DELLA CATENA DI MARKOV

Per una catena di Markov ^{omogenea (stazionaria)} è facile calcolare l'entropia $H_\infty(S)$

$$H_\infty(S) = \lim_{K \rightarrow \infty} H(S[K] | S[K-1], \dots, S[1])$$

$$= \lim_{K \rightarrow \infty} H(S[K] | S[K-1]) = H(S[K+1] | S[K])$$

per qualunque K

$$H(S[K+1] | S[K]) = \sum_{(s_1, s_2) \in \mathcal{R}^2} p(s_1, s_2) \lg \frac{1}{p(s_1, s_2)}$$

$$= \sum_{(s_1, s_2) \in \mathcal{R}^2} p(s_2 | s_1) p(s_1) \lg \frac{1}{p(s_1, s_2)}$$

$$= \sum_{i, j} P_{ij} p(i) \lg \frac{1}{P_{ij}}$$

Possiamo rendere la distribuzione stazionaria

$$p(i) = [P^{\infty}]_i \quad i = 1, \dots, m$$

$$P_{ij} = P_{ij} \{S[K+1]=j | S[K]=i\}$$

Andrà in forma matriciale.

$$H_\infty(S) = - [P^{\infty}] \left(\underline{P} \odot \lg \underline{P} \right) \underline{e} \quad \text{bit.}$$

Esempio

4.

$$\underline{P} = \begin{bmatrix} 0.3 & 0.7 \\ 0.1 & 0.9 \end{bmatrix}$$

PROBabilità 2

$$\begin{cases} \underline{u} = \underline{P}^T \underline{u} \\ \underline{e}^T \underline{u} = 1 \end{cases} \quad \left\{ \begin{array}{l} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0.3 & 0.1 \\ 0.7 & 0.9 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \\ u_1 + u_2 = 1 \end{array} \right.$$

$$\begin{cases} u_1 = 0.3 u_1 + 0.1 u_2 \\ u_2 = 0.7 u_1 + 0.9 u_2 \\ u_1 + u_2 = 1 \end{cases}$$

$$u_1 = 0.3 u_1 + 0.1 (1 - u_1)$$

$$u_1 - 0.3 u_1 + 0.1 u_2 = 0.1$$

$$0.8 u_1 = 0.1$$

$$\begin{cases} u_1 = \frac{1}{8} \\ u_2 = \frac{7}{8} \end{cases}$$

$$\underline{P}^\infty = \begin{bmatrix} \frac{1}{8} & \frac{7}{8} \\ \frac{1}{8} & \frac{7}{8} \end{bmatrix}$$

$$H_\infty(S) = - \begin{bmatrix} \frac{1}{8} & \frac{7}{8} \end{bmatrix} \begin{bmatrix} 0.3 \lg 0.3 & 0.7 \lg 0.7 \\ 0.1 \lg 0.1 & 0.9 \lg 0.9 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= 0.5205 \text{ bit}$$

l'entropia normale riferita a \underline{P}^∞

$$e^{-} H(S) = \frac{1}{8} \lg 8 + \frac{7}{8} \lg \frac{8}{7} = 0.5436 \text{ bit.}$$

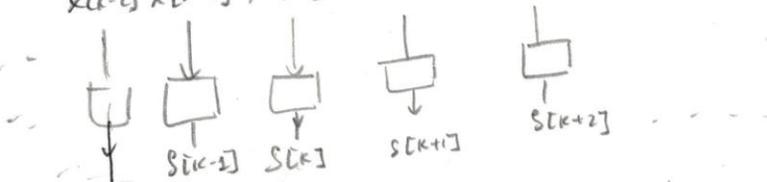
ovviamente $H_\infty(S) < H(S)$

FUNZIONI DI
CATENE DI MARKOV NASCOSTE E CATENE DI MARKOV
 HTM1
 PACHIERI, 63

(Hidden Markov Model) (HMM)

Molto utili nelle applicazioni recenti recenti
 algoritmi che sono funzioni di catene di Markov.
 In altre parole si suppone che ci sia un processo
 di Markov "nascosto" caratterizzato da un
 meccanismo di transizione dallo stato al
 successivo collegato alle osservazioni secondo
 una funzione $s[k] = \phi(x[k])$ (non deterministica
 in parole di frase
 di catene di Markov)

$x[k-2] \ x[k-1] \ x[k] \ x[k+1] \ x[k+2] \dots$



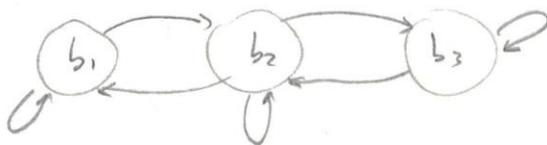
In molti casi infatti la dipendenza temporale fra
 le $s[k]$ non è così semplice ("diretta") come nelle
 catene di Markov, ma è modellabile come un
 doppio meccanismo.

Esempio supponiamo che $\{s[k]\}$ sia il numero di pacchetti
 in arrivo ad un server al tempo k .
 Il numero può dipendere da quello precedente, ma più
 crucialmente dallo stato della rete, Modelliamo un
 tale processo con degli stati

$x[k] \in \{ \text{poco traffico, medio traffico, molto traffico} \}$

$= \{ b_1, b_2, b_3 \}$

$$P_x = \begin{bmatrix} p(b_1|b_1) & p(b_2|b_1) & p(b_3|b_1) \\ p(b_1|b_2) & p(b_2|b_2) & p(b_3|b_2) \\ p(b_1|b_3) & p(b_2|b_3) & p(b_3|b_3) \end{bmatrix}$$



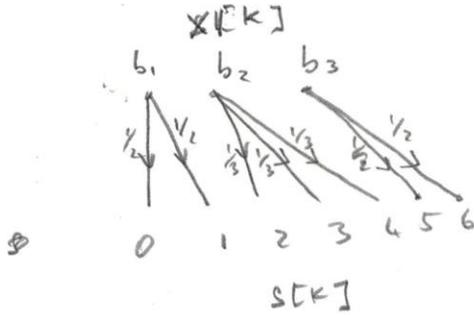
Tre tipi condizionali di traffico possono avere diverse distribuzioni di perdite. Dato che le perdite possono essere

HMM 2

$$S[K] = \{0, 1, 2, 3, 4, 5, 6\}$$

PARTE 44

$$P(S[K]=i | X[K]=b_j) \quad \begin{matrix} i = 0, 1, 2, 3, 4, 5, 6 \\ j = 1, 2, 3 \end{matrix}$$



$$P(S|x) = \left[\begin{array}{c} P_1\{S[K]=0 | X[K]=b_1\} \quad P_2\{S[K]=1 | X[K]=b_1\} \quad \dots \quad P_2\{S[K]=6 | X[K]=b_1\} \\ \hline \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \end{array} \right]$$

