Definizioni

- X : sorgente di informazione discreta;
- X_k : messaggi prodotti da X ; ogni messaggio è una v.c.d., k è l'indice temporale;
- alfabeto di X : insieme $\{x_1, x_2, ..., x_M\}$ degli M messaggi x_i che la sorgente può produrre;
- $C(x_i)$: codice o sequenza di bit che rappresentano il messaggio x_i ;
- n_i : lunghezza in bit del codice $C(X_i)$;
- $P_{X_k}(x_i)$: probabilità che la sorgente X produca il simbolo x_i ;
- $P_{X_i, X_j}(x_k, x_l)$: probabilità (congiunta) che la sorgente X produca i simboli x_k e x_l rispettivamente negli istanti i e j ;
- sorgente stazionaria: sorgente per cui valga $P_{X_k}(x_i) = cost \; \forall k$; la probabilità del simbolo x_i si pone per definizione $P_X(x_i)$;
- sorgente senza memoria: sorgente per cui valga $P_{X_i, X_j}(x_l, x_m) = P_{X_i}(x_l) P_{X_j}(x_m) \forall i, j, k, l$
- $I(x_i) = \log(\frac{1}{P_X(x_i)})$: informazione del simbolo x_i di una sorgente stazionaria. Nota che la base del logaritmo è 2 (anche in seguito, se non diversamente specificato);
- $H(X) = E[I(x_i)] = \sum_{i=1}^{M} P_X(x_i) \log(\frac{1}{P_X(x_i)})$: entropia di una sorgente stazionaria;
- $\bar{n} = \sum_{i=1}^{M} P_X(x_i) n_i$: costo medio di un codice;
- $\sum_{i=1}^{M} 2^{n_i} \leq 1$: disuguaglianza di Kraft;
- univoca decodificabilità di un codice: ogni codice è univocamente associato ad un messaggio;
- immediata decodificabilità di un codice: nessun codice è l'inizio di un altro (condizione sufficiente all'univoca decodificabilità);
- codice ottimo: un codice univocamente decodificabile C è ottimo se $\forall C' \neq C$ univocamente decodificabile $\bar{n}(C') \geq \bar{n}(C)$;
- entropia condizionata: $H(X|Y = y_j) = \sum_{i=1}^{M} P_{X|Y}(x_i|y_j) \log(\frac{1}{P_{X|Y}(x_i|y_j)})$;
- entropia congiunta: $H(X, Y) = \sum_{i=1}^{M} P_{X,Y}(x_i, y_j) \log(\frac{1}{P_{X,Y}(x_i, y_j)})$;
- entropia di una sorgente con memoria: $H(X) = \lim_{l \to +\infty} H(X_k | X_{k-1}, X_{k-2}, ... X_{k-l})$;
- sorgente di Markov con memoria m : sorgente per cui vale $P(x_k | x_{k-1}, x_{k-2}, ... x_{k-m-1}) = P(x_k | x_{k-1}, x_{k-2}, ... x_{k-m})$;
- entropia di una sorgente di Markov con memoria m :

$H(X)=H(X_k|X_{k-1}, X_{k-2}, \dots X_{k-m})$;

- definizione alternativa dell'entropia: $H_L(X)=\lim_{L\to+\infty}\frac{1}{L}H(X_k, X_{k-1}, \dots, X_{k-L+1})$;
- distribuzione condizionata delle probabilità dei simboli all'uscita di un canale (caso discreto): $P_{Y|X}(y_j|x_i)$;
- densità di probabilità dei simboli all'uscita di un canale (caso continuo) $p_{Y|X}(y_j|x_i)$;
- canale binario simmetrico (BSC): canale con alfabeto d'ingresso e di uscita $\{0,1\}$ e probabilità di errore identica per i due simboli ϵ . Il canale è completamente definito nota la probabilità $P_X(0)=p$ ed il parametro ϵ ;
- canale additivo gaussiano discreto: canale con alfabeto d'ingresso discreto $\{x_1, x_2, \dots x_M\}, x_1, x_2, \dots x_M \in \Re$ ed uscita continua $Y=X+N, N\sim N(\mu, \sigma^2)$;
- canale binario con cancellazione: canale con alfabeto d'ingresso $\{0,1\}$ e di uscita $\{0,1,E\}$, $P_{Y|X}(E|0)=P_{Y|X}(E|1)=\epsilon$, $P_{Y|X}(1|0)=P_{Y|X}(0|1)=0$.

Dimostrazioni

Teorema della massima entropia

Ipotesi

- X sorgente d'informazione discreta, stazionaria, senza memoria;
- $\{x_1, x_2, \dots, x_M\}$ alfabeto di X con M possibili messaggi;

Tesi

$H(X)\le \log M$

Dimostrazione

$$H(X)=\sum_{i=1}^{M} P_X(x_i)\log\left(\frac{1}{P_X(x_i)}\right)$$

$$H(X)=\sum_{i=1}^{M} P_X(x_i)\log\left(\frac{M}{M P_X(x_i)}\right)$$

$$H(X)=\sum_{i=1}^{M} P_X(x_i)\log(M)+\sum_{i=1}^{M} P_X(x_i)\log\left(\frac{1}{M P_X(x_i)}\right)$$

$$H(X)=\log(M)+\sum_{i=1}^{M} P_X(x_i)\log\left(\frac{1}{M P_X(x_i)}\right)$$

$$H(X)-\log(M)=\sum_{i=1}^{M} P_X(x_i)\log\left(\frac{1}{M P_X(x_i)}\right)$$

Per la disuguaglianza del logaritmo, $\ln(x)\le x-1$, si può anche scrivere:

$$H(X)-\log(M)\le\sum_{i=1}^{M} P_X(x_i)\left(\frac{1}{M P_X(x_i)}-1\right)\log(e)$$

$$H(X)-\log(M)\leq\sum_{i=1}^{M}P_X(x_i)(\frac{1}{M\,P_X(x_i)})\log(e)-\sum_{i=1}^{M}P_X(x_i)\log(e)$$

$$H(X)-\log(M)\leq\log(e)-\log(e)$$

$$H(X)-\log(M)\leq 0$$

$H(X)\leq\log(M)$, c.v.d.

La disuguaglianza di Kraft è condizione sufficiente all'esistenza di un codice immediatamente decodificabile

Ipotesi

Le stesse del teorema precedente, inoltre:

- $C(x_i)$ codici che rappresentano x_i ;
- n_i lunghezze dei codici $C(x_i)$;
- $n_1\leq n_2\leq\ldots\leq n_M$;
- vale la disuguaglianza di Kraft.

Tesi

E' sempre possibile costruire un codice immediatamente decodificabile che rispetti la disuguaglianza di Kraft.

Dimostrazione

Si scelgano arbitrariamente gli n_1 bit di $C(x_1)$.

Si scelgano arbitrariamente gli n_2 bit di $C(x_2)$ in modo che non inizino con la sequenza di $C(x_1)$.

Si scelgano arbitrariamente gli n_3 bit di $C(x_3)$ in modo che non inizino né con la sequenza di $C(x_1)$ né con quella di $C(x_2)$.

E' possibile scrivere in questo modo tutti i codici fino a $C(x_M)$ se:

$$numero\,combinazioni\,totali\geq numero\,combinazioni\,già\,usate+1$$

$$2^{n_M}\geq\frac{2^{n_M}}{2^{n_1}}+\frac{2^{n_M}}{2^{n_2}}+\ldots+\frac{2^{n_M}}{2^{n_{M-1}}}+1$$

$$1\geq\frac{1}{2^{n_1}}+\frac{1}{2^{n_2}}+\ldots+\frac{1}{2^{n_{M-1}}}+\frac{1}{2^{n_M}}$$

$$\sum_{i=1}^{M}2^{-n_i}\leq 1$$

ossia se la disuguaglianza di Kraft è rispettata.

La disuguaglianza di Kraft è condizione necessaria all'univoca decodificabilità di un codice

Ipotesi

Le stesse del teorema precedente, inoltre:

- $C(x_i)$ è univocamente decodificabile.

Tesi

$C(x_i)$ rispetta la disuguaglianza di Kraft.

Dimostrazione

Se $C(x_i)$ è univocamente decodificabile si può sempre costruire un altro codice i cui messaggi sono sequenze di N messaggi originali. Si definisce il nuovo alfabeto, i nuovi codice e le loro lunghezze come segue:

$x_{i_1}, x_{i_2}, \dots, x_{i_N}$

$C(x_{i_1}), C(x_{i_2}), \dots, C(x_{i_N})$

$n_{i_1}, n_{i_2}, \dots, n_{i_N}$

Si definisce inoltre la massima lunghezza $n_{max} = max(n_{i_1}, n_{i_2}, \dots, n_{i_N})$. Per il nuovo codice vale la seguente uguaglianza:

$$\left(\sum_{i=0}^{M} 2^{-n_i}\right)^N = \sum_{i_1=0}^{M}\sum_{i_2=0}^{M}\dots\sum_{i_N=0}^{M} 2^{-(n_{i_1}+n_{i_2}+\dots+n_{i_N})}$$

L'uguaglianza è una forma della definizione della potenza di polinomio. La stessa scrittura può assumere anche una terza forma:

$$\sum_{i_1=0}^{M}\sum_{i_2=0}^{M}\dots\sum_{i_N=0}^{M} 2^{-(n_{i_1}+n_{i_2}+\dots+n_{i_N})} = \sum_{n=1}^{N\,n_{max}} A_n 2^{-n}$$

dove A_n è il numero di messaggi con lunghezza totale n, che varia da 1 bit a $N\,n_{max}$ bit. Perchè il nuovo codice sia univocamente decodificabile, deve essere $A_n \le 2^n$, quindi:

$$\sum_{n=1}^{N\,n_{max}} A_n 2^{-n} \le \sum_{n=1}^{N\,n_{max}} 2^n 2^{-n}$$

ossia:

$$\sum_{n=1}^{N\,n_{max}} A_n 2^{-n} \le N\,n_{max}$$

$$\left(\sum_{i=0}^{M} 2^{-n_i}\right)^N \le N\,n_{max}$$

$$\sum_{i=0}^{M} 2^{-n_i} \le (N\,n_{max})^{\frac{1}{N}}$$

L'ultima disuguaglianza, in particolare, deve valere per ogni N. In particolare deve valere per $N \to +\infty$, quindi:

$$\sum_{i=0}^{M} 2^{-n_i} \leq 1$$

che è la disuguaglianza di Kraft, il che dimostra che ogni codice univocamente decodificabile la rispetta.

Primo Teorema di Shannon

Ipotesi

Le stesse del teorema precedente.

Tesi

Se un codice è univocamente decodificabile allora $\bar{n} \geq H(X)$.

Dimostrazione

Per definizione, vale:

$$H(X) - \bar{n} = \sum_{i=1}^{M} P_X(x_i) \log\left(\frac{1}{P_X(x_i)}\right) - \sum_{i=1}^{M} P_X(x_i)\, n_i$$

$$H(X) - \bar{n} = \sum_{i=1}^{M} P_X(x_i) \left[\log\left(\frac{1}{P_X(x_i)}\right) - n_i\right]$$

$$H(X) - \bar{n} = \sum_{i=1}^{M} P_X(x_i) \log\left(\frac{2^{-n_i}}{P_X(x_i)}\right)$$

Utilizzando la disuguaglianza del logaritmo posso scrivere:

$$H(X) - \bar{n} \leq \sum_{i=1}^{M} P_X(x_i) \left(\frac{2^{-n_i}}{P_X(x_i)} - 1\right) \log(e)$$

$$H(X) - \bar{n} \leq \sum_{i=1}^{M} 2^{-n_i} \log(e) - \sum_{i=1}^{M} \frac{1}{P_X(x_i)} \log(e)$$

$$H(X) - \bar{n} \leq \log(e) \left[\sum_{i=1}^{M} 2^{-n_i} - 1\right]$$

Per la disuguaglianza di Kraft $\sum_{i=1}^{M} 2^{-n_i} - 1 \leq 0$, quindi:

$H(X) - \bar{n} \leq 0$ ossia:

$H(X) \leq \bar{n}$ c.v.d.

Ottimalità del codice di Huffmann per sorgenti senza memoria (traccia)

Ipotesi

- X sorgente d'informazione discreta, stazionaria, senza memoria;
- $\{x_1, x_2, \ldots, x_M\}$ alfabeto di X con M possibili messaggi;

- $C(x_i)$ codici che rappresentano x_i prodotti con l'algoritmo di Huffmann;
- n_i lunghezze dei codici $C(x_i)$;

Tesi

$C(x_i)$ è ottimo.

Traccia della dimostrazione

Si dimostra il lemma seguente:

"Tra i codici ottimi ce n'è almeno uno che

- *se* $P_X(x_i)<P_X(x_j)$ *allora* $n_i \geq n_j$;
- *per i due messaggi meno probabili, ad esempio* x_{M-1} *e* x_M *vale* $n_M = n_{M-1}$ *e i rispettivi codici differiscono solo per l'ultimo bit."*

Se C è ottimo e rispetta il lemma, gli ultimi due codici avranno la stessa lunghezza e differiranno per l'ultimo bit, quindi la lunghezza media sarà calcolabile come:

$$\bar{n} = \sum_{i=1}^{M} P_X(x_i)\, n_i = \sum_{i=1}^{M-2} P_X(x_i)\, n_i + n_{M-1}[P_X(x_{M-1}) + P_X(x_M)]$$

Se fosse possibile codificare i messaggi $x_{1,} x_{2,} \ldots x_{M-2}$ in modo ottimo, C sarebbe ottimo.

A questo scopo è possibile costruire una nuova fonte X' che abbia come alfabeto i messaggi $\{x_{1,} x_{2,} \ldots x_{M-2}, x_{M-1} \cup x_M\}$ dove $x_{M-1} \cup x_M$ è un messaggio di probabilità $P_X(x_{M-1}) + P_X(x_M)$ che rappresenta sia x_{M-1} che x_M . X' sarà codificata con un nuovo codice C' identico a C per i primi $M-2$ messaggi e che codifichi $x_{M-1} \cup x_M$ con gli $n_{M-1}-1$ bit in comune in C .

Per come è costruito, C' è ottimo se fosse possibile codificare in modo ottimo i primi $M-3$ messaggi; si può quindi continuare in maniera analoga ricorsivamente fino a provare che C è ottimo. Inoltre la lunghezza media di C' è calcolabile come:

$$\bar{n}' = \bar{n} - P_X(x_{M-1}) - P_X(x_M)$$

Quindi la differenza tra le lunghezze medie:

$$\bar{n} - \bar{n}' = P_X(x_{M-1}) + P_X(x_M)$$

non dipende dai particolari simboli scelti.

Entropia condizionata media

La media dell'entropia condizionata può essere calcolata come:

$$\sum_{j=1}^{N} P_Y(y_j)\, H(X|Y=y1)$$

$$\sum_{j=1}^{N} P_Y(y_j) \sum_{i=1}^{M} P_{X|Y}(x_i|y_j) \log\left(\frac{1}{P_{X|Y}(x_i|y_j)}\right)$$

$$\sum_{i=1}^{M} \sum_{j=1}^{N} P_Y(y_j)\, P_{X|Y}(x_i|y_j) \log\left(\frac{1}{P_{X|Y}(x_i|y_j)}\right)$$

Ma per la definizione di probabilità condizionata si può scrivere

$P(A|B)=P\frac{(A,B)}{P(B)}\rightarrow P(A,B)=P(B)\,P(A|B)$ quindi la scrittura precedente diventa:

$$\sum_{i=1}^{M}\sum_{j=1}^{N} P_{X,Y}(x_i,y_j)\log(\frac{1}{P_{X|Y}(x_i|y_j)})$$

Quest'ultima forma si prende come definizione di entropia condizionata media:

$$H(X|Y)=\sum_{i=1}^{M}\sum_{j=1}^{N} P_{X,Y}(x_i,y_j)\log(\frac{1}{P_{X|Y}(x_i|y_j)}) \quad .$$

L'entropia condizionata è minore o uguale all'entropia semplice

Tesi

$H(X|Y)\leq H(X)$

Dimostrazione

$$H(X|Y)-H(X)=\sum_{i=1}^{M}\sum_{j=1}^{N} P_{X,Y}(x_i,y_j)\log(\frac{1}{P_{X|Y}(x_i|y_j)})-\sum_{i=1}^{M} P_X(x_i)\log(\frac{1}{P_X(x_i)})$$

Per il teorema della somma:

$$H(X|Y)-H(X)=\sum_{i=1}^{M}\sum_{j=1}^{N} P_{X,Y}(x_i,y_j)\log(\frac{1}{P_{X|Y}(x_i|y_j)})-\sum_{i=1}^{M}\sum_{j=1}^{N} P_{X,Y}(x_i,y_j)\log(\frac{1}{P_X(x_i)})$$

$$H(X|Y)-H(X)=\sum_{i=1}^{M}\sum_{j=1}^{N} P_{X,Y}(x_i,y_j)[\log(\frac{1}{P_{X|Y}(x_i|y_j)})+\log(P_X(x_i))]$$

$$H(X|Y)-H(X)=\sum_{i=1}^{M}\sum_{j=1}^{N} P_{X,Y}(x_i,y_j)[\log(\frac{P_X(x_i)}{P_{X|Y}(x_i|y_j)})]$$

$$H(X|Y)-H(X)=\sum_{i=1}^{M}\sum_{j=1}^{N} P_{X,Y}(x_i,y_j)[\log(\frac{P_X(x_i)P_Y(y_j)}{P_{X,Y}(x_i,y_j)})]$$

Per la disuguaglianza del logaritmo si può anche scrivere:

$$H(X|Y)-H(X)\leq\sum_{i=1}^{M}\sum_{j=1}^{N} P_{X,Y}(x_i,y_j)[\frac{P_X(x_i)P_Y(y_j)}{P_{X,Y}(x_i,y_j)}-1]\log e$$

$$H(X|Y)-H(X)\leq\sum_{i=1}^{M}\sum_{j=1}^{N} P_X(x_i)P_Y(y_j)\log e-\sum_{i=1}^{M}\sum_{j=1}^{N} P_{X,Y}(x_i,y_j)\log e$$

$H(X|Y)-H(X)\leq\log e-\log e$

$H(X|Y)-H(X)\leq 0$ quindi

$H(X|Y)\leq H(X)$ c.v.d.

Entropia congiunta di variabili indipendenti

Se $X\perp Y$ allora $P_{X,Y}(x_i,y_j)=P_X(x_i)\,P_Y(y_j)$ quindi

$$H(X,Y)=\sum_{i=1}^{M}\sum_{j=1}^{N} P_X(x_i)P_Y(y_j)\log(\frac{1}{P_X(x_i)P_Y(y_j)})$$

$$H(X,Y)=\sum_{i=1}^{M}\sum_{j=1}^{N}P_X(x_i)P_Y(y_j)\log(\frac{1}{P_X(x_i)})+\sum_{i=1}^{M}\sum_{j=1}^{N}P_X(x_i)P_Y(y_j)\log(\frac{1}{P_Y(y_j)})$$

$$H(X,Y)=\sum_{i=1}^{M}P_X(x_i)\log(\frac{1}{P_X(x_i)})+\sum_{j=1}^{N}P_Y(y_j)\log(\frac{1}{P_Y(y_j)}) \quad \text{ossia}$$

$$H(X,Y)=H(X)+H(Y)$$

Formula alternativa dell'entropia congiunta

Dal momento che $P(A|B)=P\frac{(A,B)}{P(B)}\rightarrow P(A,B)=P(B)\,P(A|B)$ l'espressione dell'entropia congiunta si può riscrivere nella forma:

$$H(X,Y)=\sum_{i=1}^{M}\sum_{j=1}^{N}P_{X,Y}(x_i,y_j)\log(\frac{1}{P_{X,Y}(x_i,y_j)})$$

$$H(X,Y)=\sum_{i=1}^{M}\sum_{j=1}^{N}P_{X,Y}(x_i,y_j)\log(\frac{1}{P_X(x_i)P_{X|Y}(x_i|y_j)})$$

$$H(X,Y)=\sum_{i=1}^{M}\sum_{j=1}^{N}P_{X,Y}(x_i,y_j)\log(\frac{1}{P_X(x_i)})+\sum_{i=1}^{M}\sum_{j=1}^{N}P_{X,Y}(x_i,y_j)\log(\frac{1}{P_{X|Y}(x_i|y_j)})$$

$$H(X,Y)=\sum_{i=1}^{M}P_X(x_i)\log(\frac{1}{P_X(x_i)})+\sum_{i=1}^{M}\sum_{j=1}^{N}P_{X,Y}(x_i,y_j)\log(\frac{1}{P_{X|Y}(x_i|y_j)})$$

$$H(X,Y)=H(X)+H(Y|X)$$

Definizione alternativa dell'entropia per sorgente con memoria

Ipotesi

- X sorgente d'informazione discreta, stazionaria, di Markov con memoria m ;

Tesi

Le due definizioni di entropia sono equivalenti.

Dimostrazione

La definizione alternativa di entropia è

$H_L(X)=\frac{1}{L}H(X_k,X_{k-1},\ldots,X_{k-L+1})$ al tendere di L all'infinito.

Questa definizione può essere scritta nella forma:

$$H_L(X)=\frac{1}{L}H(X_k|X_{k-1},\ldots,X_{k-L+1})+\frac{1}{L}H(X_{k-1},\ldots,X_{k-L+1})$$

$$H_L(X)=\frac{1}{L}H(X_k|X_{k-1},\ldots,X_{k-L+1})+\frac{1}{L}H(X_{k-1}|X_{k-2},\ldots,X_{k-L+1})+\frac{1}{L}H(X_{k-2},\ldots,X_{k-L+1})$$

Si può continuare in questo modo per $L-M$ passi supponendo $M<L-M\rightarrow L>2\mathrm{M}$:

$$H_L(X)=\frac{1}{L}[H(X_k|X_{k-1},\ldots,X_{k-L+1})+\ldots+H(X_{k-L+M}|X_{k-L+M-1},\ldots,X_{k-L+1})]+\frac{1}{L}H(X_{k-L+M-1},\ldots,X_{k-L+1})$$

Inoltre, considerando che i termini della prima sommatoria sono tutti uguali per stazionarietà, si può scrivere:

$$H_L(X)=\frac{L-m}{L}H(X_k|X_{k-1},\dots,X_{k-L+1})+\frac{1}{L}H(X_{k-L+M-1},\dots,X_{k-L+1})$$

Quindi, per $L\to+\infty$:

$$H_L(X)=H(X_k|X_{k-1},\dots,X_{k-L+1})$$

che è la prima definizione di entropia.

Primo teorema di Shannon per sorgenti con memoria

Si suppone di codificare L messaggi consecutivi ($\underline{x}$ vettore dei messaggi) da una fonte X con un codice univocamente decodificabile. Siano i gli M^L messaggi $\underline{x}$ alfabeto di una nuova sorgente, Y , che può quindi essere a sua volta codificata. Le lunghezze $n(y_i)=n(\underline{x})$ dei messaggi devono soddisfare l'uguaglianza di Kraft:

$$\sum_{\underline{x}_i=1}^{M^L}2^{-n(\underline{x})}\le 1 \quad .$$

D'altra parte, è sempre vero che:

$$H_L(X)-\bar{n}=\frac{1}{L}H(\underline{x})-\bar{n}$$

dove $\bar{n}=\frac{1}{L}\sum_{\underline{x}_i=1}^{M^L}P_{\underline{x}}(\underline{x}_i)n(\underline{x}_i)$. La scrittura può essere esplicitata come segue:

$$H_L(X)-\bar{n}=\frac{1}{L}\sum_{\underline{x}_i=1}^{M^L}P_{\underline{x}}(\underline{x}_i)\log\left(\frac{1}{P_{\underline{x}}(\underline{x}_i)}\right)-\frac{1}{L}\sum_{\underline{x}_i=1}^{M^L}P_{\underline{x}}(\underline{x}_i)n(\underline{x}_i)$$

$$H_L(X)-\bar{n}=\frac{1}{L}\sum_{\underline{x}_i=1}^{M^L}P_{\underline{x}}(\underline{x}_i)\left[\log\left(\frac{1}{P_{\underline{x}}(\underline{x}_i)}\right)-n(\underline{x}_i)\right]$$

$$H_L(X)-\bar{n}=\frac{1}{L}\sum_{\underline{x}_i=1}^{M^L}P_{\underline{x}}(\underline{x}_i)\log\left(\frac{2^{-n(\underline{x}_i)}}{P_{\underline{x}}(\underline{x}_i)}\right)$$ ed utilizzando la disuguaglianza del logaritmo:

$$H_L(X)-\bar{n}\le\frac{1}{L}\sum_{\underline{x}_i=1}^{M^L}P_{\underline{x}}(\underline{x}_i)\left(\frac{2^{-n(\underline{x}_i)}}{P_{\underline{x}}(\underline{x}_i)}-1\right)$$ ma per Kraft il secondo membro può al più uguagliare zero, quindi:

$$\bar{n}\le H_L(X) \quad .$$