Cenni di teoria dell'informazione

Gregorio D'Agostino

11 Maggio 2021

Crittoanilisi: Analisi probabilistica delle sequenze

Metodo delle frequenze - Legge dei grandi numeri

Entropia

L'approccio probabilistico

- Una delle domande fondamentali quando si è di fronte ad un flusso o anche ad un insieme di dati non ordinato riguarda la loro regolarità.

L'approccio probabilistico

- Una delle domande fondamentali quando si è di fronte ad un flusso o anche ad un insieme di dati non ordinato riguarda la loro regolarità.
- L'attaccante si chiede se può estrarre elementi conoscitivi dalla sequenza dei dati.

L'approccio probabilistico

- Una delle domande fondamentali quando si è di fronte ad un flusso o anche ad un insieme di dati non ordinato riguarda la loro regolarità.
- L'attaccante si chiede se può estrarre elementi conoscitivi dalla sequenza dei dati.
- Il difensore si chiede se stia rendendo pubblica qualche parte dell'informazione che intende custodire o trasmettere.

L'approccio probabilistico

- Una delle domande fondamentali quando si è di fronte ad un flusso o anche ad un insieme di dati non ordinato riguarda la loro regolarità.
- L'attaccante si chiede se può estrarre elementi conoscitivi dalla sequenza dei dati.
- Il difensore si chiede se stia rendendo pubblica qualche parte dell'informazione che intende custodire o trasmettere.
- In entrambi i casi l'approccio probabilistico consente di valutare quantitativamente il problema.

Attesa probabilistica

- Data una variabile stocastica discreta η che può assumere i valori $x^1, x^2, \ldots, x^M$, si definisce valore d'attesa (o semplicemente attesa) di una funzione di η la sua somma sui valori x^i pesata con le probabilità degli stessi:

$$E[f(\eta)] \stackrel{def}{=} \sum_{i=1}^{M} f(x^i) \cdot p_i.$$

Se la cardinalità M è finita la somma esiste sempre; se la variabile stocastica assume un insieme di valori numerabile, la somma si estende all'infinito e non sempre converge. Le funzioni limitate (funzioni di prova) sono sempre sommabili.

Attesa probabilistica

- Data una variabile stocastica discreta η che può assumere i valori $x^1, x^2, \ldots, x^M$, si definisce valore d'attesa (o semplicemente attesa) di una funzione di η la sua somma sui valori x^i pesata con le probabilità degli stessi:

$$E[f(\eta)] \stackrel{def}{=} \sum_{i=1}^{M} f(x^i) \cdot p_i.$$

  Se la cardinalità M è finita la somma esiste sempre; se la variabile stocastica assume un insieme di valori numerabile, la somma si estende all'infinito e non sempre converge. Le funzioni limitate (funzioni di prova) sono sempre sommabili.

- *L'attesa viene anche denominata "aspettazione" o spesso impropriamente "speranza". A volte si sente dire la "speranza di vita di un paziente è ..." Ogni paziente spera di vivere a lungo ed in buona salute, la sua attesa di vita non ha nulla a che fare con i suoi desideri. In inglese il termine utilizzato è expectation che indica ciò che razionalmente si ritiene che accada.*

Proprietà dell'attesa

- Linearità:

$$E[a \cdot f(\xi) + b \cdot g(\eta) + c] = a \cdot E[f(\xi)] + b \cdot E[f(\eta)] + c;$$

in cui a,b e c sono costanti, non variabili aleatorie come ξ ed η.

Proprietà dell'attesa

- Linearità:

$$E[a \cdot f(\xi) + b \cdot g(\eta) + c] = a \cdot E[f(\xi)] + b \cdot E[f(\eta)] + c;$$

in cui a,b e c sono costanti, non variabili aleatorie come ξ ed η.

- Preserva il segno dell'argomento

$$f > 0 \Rightarrow E[f(\xi)] > 0.$$

Proprietà dell'attesa

- Linearità:

$$E[a \cdot f(\xi) + b \cdot g(\eta) + c] = a \cdot E[f(\xi)] + b \cdot E[f(\eta)] + c;$$

in cui a,b e c sono costanti, non variabili aleatorie come ξ ed η.

- Preserva il segno dell'argomento

$$f > 0 \Rightarrow E[f(\xi)] > 0.$$

- Se due variabili sono indipendenti le attese possono essere calcolate separatamente:

$$E_{\xi\eta}[f(\xi)g(\eta)] = \sum_{i,j} \rho_{ij}^{\xi\eta} f(x^i)g(y^j) = \sum_{i,j} \rho_i^{\xi} \rho_j^{\eta} f(x^i)g(y^j);$$

$$E_{\xi\eta}[f(\xi)g(\eta)] = \left(\sum_i \rho_j^{\eta} f(x^i)\right)\left(\sum_j \rho_j^{\eta} g(y^j)\right) = E_{\xi}[f(\xi)]E_{\eta}[g(\eta)]$$

Descrittori sintetici

- Una delle grandezze più utilizzate per quantificare l'intensità di una variabile stocastica η è il suo valore di attesa che si indica con μ. Quando esiste si calcola:

$$\mu \stackrel{def}{=} E[\eta] = \sum_{i=1}^{M} x^i \cdot p_i;$$

in cui x^i sono i valori assumibili dalla η; se M è finito o p_i converge abbastanza velocemente, il valore d'attesa esiste. Esso esprime sinteticamente l'estensione tipica di una variabile stocastica.

- Similmente (quando esiste) si definisce la varianza:

$$Var[\eta] \stackrel{def}{=} \sigma^2 \stackrel{def}{=} E[(\eta - \mu)^2] = \sum_{i=1}^{M} (x^i - \mu)^2 \cdot p_i.$$

Questo è un indice sintetico della dispersione della variabile intorno al suo valore di attesa.

La frequenza dei caratteri

- Una delle caratteristiche stocastiche di una sorgente è la frequenza dei caratteri (sia in senso stretto, quando sono lettere, sia in senso gereralizzato come sequenza di 8 bit (numeri tra 0 e 255).

La frequenza dei caratteri

- Una delle caratteristiche stocastiche di una sorgente è la frequenza dei caratteri (sia in senso stretto, quando sono lettere, sia in senso gereralizzato come sequenza di 8 bit (numeri tra 0 e 255).
- Un attaccante che osservi un lungo flusso di caratteri generato da una sorgente può innanzi tutto calcolare le frequenze dei caratteri ed osservarne la eventuale stabilità.

La frequenza dei caratteri

- Una delle caratteristiche stocastiche di una sorgente è la frequenza dei caratteri (sia in senso stretto, quando sono lettere, sia in senso gereralizzato come sequenza di 8 bit (numeri tra 0 e 255).
- Un attaccante che osservi un lungo flusso di caratteri generato da una sorgente può innanzi tutto calcolare le frequenze dei caratteri ed osservarne la eventuale stabilità.
- Questa operazione si chiama analisi delle frequenze e come vedremo consente di decrittare tutti i cifrari sostituzionali.

La frequenza dei caratteri

- Una delle caratteristiche stocastiche di una sorgente è la frequenza dei caratteri (sia in senso stretto, quando sono lettere, sia in senso gereralizzato come sequenza di 8 bit (numeri tra 0 e 255).
- Un attaccante che osservi un lungo flusso di caratteri generato da una sorgente può innanzi tutto calcolare le frequenze dei caratteri ed osservarne la eventuale stabilità.
- Questa operazione si chiama analisi delle frequenze e come vedremo consente di decrittare tutti i cifrari sostituzionali.
- Calcolando la frequenza di un carattere nel crittogramma e comparandola con quelle dei caratteri nella lingua in cui è scritto il messaggio, possiamo dedurre quale carattere sostituisca.

La frequenza dei caratteri

- Una delle caratteristiche stocastiche di una sorgente è la frequenza dei caratteri (sia in senso stretto, quando sono lettere, sia in senso gereralizzato come sequenza di 8 bit (numeri tra 0 e 255).
- Un attaccante che osservi un lungo flusso di caratteri generato da una sorgente può innanzi tutto calcolare le frequenze dei caratteri ed osservarne la eventuale stabilità.
- Questa operazione si chiama analisi delle frequenze e come vedremo consente di decrittare tutti i cifrari sostituzionali.
- Calcolando la frequenza di un carattere nel crittogramma e comparandola con quelle dei caratteri nella lingua in cui è scritto il messaggio, possiamo dedurre quale carattere sostituisca.
- Come sempre daremo un fondamento matematico a queste osservazioni generiche.

Le leggi dei grandi numeri

- Dato un evento (composito) A con probabilità p in molti casi è possibile costruire N repliche identiche ed indipendenti del sistema per osservare quante volte si presenta l'evento desiderato. Alternativamente possiamo pensare ad una variabile stocastica (χ) detta indicatore di A che vale uno se si verifica l'evento A e zero quando non si verifica. La funzione ξ associata alla variabile aleatoria è:

$$\forall e \in \Omega : \begin{cases} \chi(A) : \xi(e) = 1 \Leftrightarrow e \in A, \\ \chi(A) : \xi(e) = 0 \Leftrightarrow e \notin A. \end{cases}$$

Questa situazione corrisponde ad una schema di prove reiterate in cui il sistema è posto nelle stesse condizioni N volte e non vi è alcun legame tra le varie repliche. Ovvero vi sono N variabili indipendenti ξ_i identicamente distribuite con distribuzione dicotomica.

Le leggi dei grandi numeri -cont

- La frequenza ν dell'evento si definisce come il numero di volte in cui si presenta l'evento A che corrisponde alla somma delle variabili ξ_i:

$$\nu = \sum_{i=1}^{N} \xi_i.$$

Le leggi dei grandi numeri -cont

- La frequenza ν dell'evento si definisce come il numero di volte in cui si presenta l'evento A che corrisponde alla somma delle variabili ξ_i:

$$\nu = \sum_{i=1}^{N} \xi_i.$$

- La legge dei grandi numeri (in forma debole) ci assicura che la frequenza relativa ν/N tende (debolmente) alla probabilità p:

$$\frac{\nu}{N} \xrightarrow{\mathcal{P}} p.$$

Le leggi dei grandi numeri -cont

- La frequenza ν dell'evento si definisce come il numero di volte in cui si presenta l'evento A che corrisponde alla somma delle variabili ξ_i:

$$\nu = \sum_{i=1}^{N} \xi_i.$$

- La legge dei grandi numeri (in forma debole) ci assicura che la frequenza relativa ν/N tende (debolmente) alla probabilità p:

$$\frac{\nu}{N} \xrightarrow{\mathcal{P}} p.$$

- Il simbolo $\xrightarrow{\mathcal{P}}$ indica la "convergenza debole". $\eta_n \xrightarrow{\mathcal{P}} \eta$ si legge η_n converge in probabilità ad η e formalmente significa:

$$\forall \delta > 0, \epsilon > 0 : \exists N_{\delta,\epsilon} : \forall N > N_{\delta,\epsilon} \mathcal{P}\left(|\eta_n - \eta| > \delta\right) < \epsilon.$$

La diseguaglianza di Chebyshev

- Data una variabile stocastica la sua varianza consente di dare una stima degli scarti della variabile dal suo valore di attesa:

$$\mathcal{P}\left(|\xi-\mu|>\delta\right)=E\left[\chi\left(|\xi-\mu|>\delta\right)\right]=E\left[\chi\left(\frac{|\xi-\mu|^2}{\delta^2}>1\right)\right]$$

$$\mathcal{P}\left(|\xi-\mu|>\delta\right)\leq E\left[\frac{|\xi-\mu|^2}{\delta^2}\chi\left(\frac{|\xi-\mu|^2}{\delta^2}>1\right)\right]\leq E\left[\frac{|\xi-\mu|^2}{\delta^2}\right].$$

$$\mathcal{P}\left(|\xi-\mu|>\delta\right)\leq E\left[\frac{|\xi-\mu|^2}{\delta^2}\right]=Var[\xi]/\delta^2.$$

La diseguaglianza di Chebyshev

- Data una variabile stocastica la sua varianza consente di dare una stima degli scarti della variabile dal suo valore di attesa:

$$\mathcal{P}\left(|\xi-\mu|>\delta\right)=E\left[\chi\left(|\xi-\mu|>\delta\right)\right]=E\left[\chi\left(\frac{|\xi-\mu|^2}{\delta^2}>1\right)\right]$$

$$\mathcal{P}\left(|\xi-\mu|>\delta\right)\leq E\left[\frac{|\xi-\mu|^2}{\delta^2}\chi\left(\frac{|\xi-\mu|^2}{\delta^2}>1\right)\right]\leq E\left[\frac{|\xi-\mu|^2}{\delta^2}\right].$$

$$\mathcal{P}\left(|\xi-\mu|>\delta\right)\leq E\left[\frac{|\xi-\mu|^2}{\delta^2}\right]=Var[\xi]/\delta^2.$$

- La varianza fornisce una limitazione alla probabilità di scartare dal valore di attesa.

Legge dei grandi numeri

- Consideriamo la frequenza relativa cioè la variabile $\frac{\nu}{N} = \frac{1}{N}\sum_{i=1}^{N} \xi_i$. ed applichiamo la diseguaglianza di Ch.

Legge dei grandi numeri

- Consideriamo la frequenza relativa cioè la variabile $\frac{\nu}{N} = \frac{1}{N}\sum_{i=1}^{N} \xi_i$. ed applichiamo la diseguaglianza di Ch.
- Calcoliamo l'attesa di $\frac{\nu}{N}$:

$$E\left[\frac{1}{N}\sum_{i=1}^{N}\xi_i\right] = \frac{1}{N}\sum_{i=1}^{N} E\left[\xi_i\right] = \frac{1}{N}\sum_{i=1}^{N}\left[p\cdot 1 + (1-p)\cdot 0\right] = p.$$

Legge dei grandi numeri

- Consideriamo la frequenza relativa cioè la variabile $\frac{\nu}{N} = \frac{1}{N}\sum_{i=1}^{N} \xi_i$. ed applichiamo la diseguaglianza di Ch.
- Calcoliamo l'attesa di $\frac{\nu}{N}$:

$$E\left[\frac{1}{N}\sum_{i=1}^{N}\xi_i\right] = \frac{1}{N}\sum_{i=1}^{N} E\left[\xi_i\right] = \frac{1}{N}\sum_{i=1}^{N}\left[p\cdot 1 + (1-p)\cdot 0\right] = p.$$

- Calcoliamo la devianza $E[\xi^2]$: [Si noti che $\xi_i^2 = \xi_i$]

$$E\left[\left(\frac{1}{N}\sum_{i=1}^{N}\xi_i\right)^2\right] = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} E\left[\xi_i\cdot\xi_j\right] =$$

$$= \frac{1}{N^2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}{}^{(i)} E\left[\xi_i\xi_j\right] + \sum_{j=1}^{N} E\left[\xi_i\right] + \right) = \frac{1}{N^2}(N(N-1)p^2+Np) =$$

$$= \frac{1}{N^2}(N(N-1)p^2 - Np) = \frac{1}{N}((N-1)p^2 + p).$$

Legge dei grandi numeri -cont

- Calcoliamo la varianza
  $Var[\xi] = E[(\xi - \mu)^2] = E[\xi^2] + \mu^2 - 2E[\xi]\mu = E[\xi^2] - \mu^2.$

Legge dei grandi numeri -cont

- Calcoliamo la varianza
  $Var[\xi] = E[(\xi - \mu)^2] = E[\xi^2] + \mu^2 - 2E[\xi]\mu = E[\xi^2] - \mu^2.$
- Nel caso della frequenza relativa:

$$Var(\xi) = E[(\nu/N)^2] - p^2 = \frac{1}{N}((N-1)p^2 + p) - p = \frac{p(1-p)}{N}$$

Legge dei grandi numeri -cont

- Calcoliamo la varianza
  $Var[\xi] = E[(\xi - \mu)^2] = E[\xi^2] + \mu^2 - 2E[\xi]\mu = E[\xi^2] - \mu^2.$
- Nel caso della frequenza relativa:

$$Var(\xi) = E[(\nu/N)^2] - p^2 = \frac{1}{N}((N-1)p^2 + p) - p = \frac{p(1-p)}{N}$$

- Applicando la diseguaglianza di Chebyshev alla frequenza relativa si ottiene:

$$\mathcal{P}\left(\left|\frac{\nu}{N} - p\right| > \delta\right) \leq \frac{1}{\delta^2} Var\left(\frac{\nu}{N}\right) = \frac{p(1-p)}{N \cdot \delta^2} < \epsilon$$

Legge dei grandi numeri -cont

- Calcoliamo la varianza
  $Var[\xi] = E[(\xi - \mu)^2] = E[\xi^2] + \mu^2 - 2E[\xi]\mu = E[\xi^2] - \mu^2.$
- Nel caso della frequenza relativa:

$$Var(\xi) = E[(\nu/N)^2] - p^2 = \frac{1}{N}((N-1)p^2 + p) - p = \frac{p(1-p)}{N}$$

- Applicando la diseguaglianza di Chebyshev alla frequenza relativa si ottiene:

$$\mathcal{P}\left(\left|\frac{\nu}{N} - p\right| > \delta\right) \leq \frac{1}{\delta^2} Var\left(\frac{\nu}{N}\right) = \frac{p(1-p)}{N \cdot \delta^2} < \epsilon$$

- Che è vera per tutti gli $N > N_{\epsilon,\delta}$

$$N_{\delta,\epsilon} = \frac{p(1-p)}{\epsilon \cdot \delta^2}.$$

Osservazioni utili

- In ogni linguaggio le lettere dell'alfabeto vengono utilizzate con una certa frequenza. Ogni lettera ha una frequenza diversa.

Osservazioni utili

- In ogni linguaggio le lettere dell'alfabeto vengono utilizzate con una certa frequenza. Ogni lettera ha una frequenza diversa.
- Nella lingua parlata o scritta l'ipotesi di indipendenza dei caratteri non è valida, ma la legge dei grandi numeri sussiste.

Osservazioni utili

- In ogni linguaggio le lettere dell'alfabeto vengono utilizzate con una certa frequenza. Ogni lettera ha una frequenza diversa.
- Nella lingua parlata o scritta l'ipotesi di indipendenza dei caratteri non è valida, ma la legge dei grandi numeri sussiste.
- Questa proprietà suggerisce un metodo per la crittanalisi dei testi codificati con cifrature sostituzionali:
  Si calcolano le frequenze dei caratteri cifrati e si identificano con le lettere dell'alfabeto che posseggono la stessa frequenza.
  Più i testi sono lunghi e meno è possibile commettere degli errori nelle attribuzioni. Ovviamente questo metodo è velocissimo perché scala linearmente con la lunghezza del messaggio.

Osservazioni utili

- In ogni linguaggio le lettere dell'alfabeto vengono utilizzate con una certa frequenza. Ogni lettera ha una frequenza diversa.
- Nella lingua parlata o scritta l'ipotesi di indipendenza dei caratteri non è valida, ma la legge dei grandi numeri sussiste.
- Questa proprietà suggerisce un metodo per la crittanalisi dei testi codificati con cifrature sostituzionali:
  Si calcolano le frequenze dei caratteri cifrati e si identificano con le lettere dell'alfabeto che posseggono la stessa frequenza.
  Più i testi sono lunghi e meno è possibile commettere degli errori nelle attribuzioni. Ovviamente questo metodo è velocissimo perché scala linearmente con la lunghezza del messaggio.
- In generale: più lungo è il messaggio cifrato, più facile è la decrittazione.

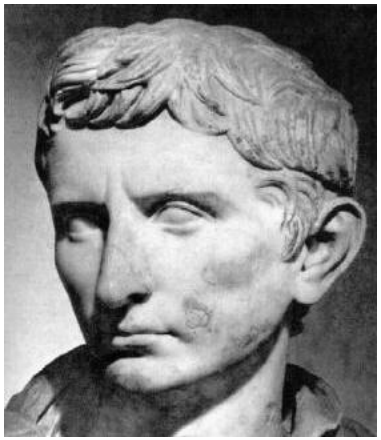
Osservazioni utili

- In ogni linguaggio le lettere dell'alfabeto vengono utilizzate con una certa frequenza. Ogni lettera ha una frequenza diversa.
- Nella lingua parlata o scritta l'ipotesi di indipendenza dei caratteri non è valida, ma la legge dei grandi numeri sussiste.
- Questa proprietà suggerisce un metodo per la crittanalisi dei testi codificati con cifrature sostituzionali:
  Si calcolano le frequenze dei caratteri cifrati e si identificano con le lettere dell'alfabeto che posseggono la stessa frequenza.
  Più i testi sono lunghi e meno è possibile commettere degli errori nelle attribuzioni. Ovviamente questo metodo è velocissimo perché scala linearmente con la lunghezza del messaggio.
- In generale: più lungo è il messaggio cifrato, più facile è la decrittazione.
- Esercizio: cifrare e decrittare un testo qualsiasi con una cifratura sostituzionale mediante il metodo delle frequenze.

La cifratura di Ottaviano Augusto

Livia Drusilla Claudia; Roma, 30 gennaio 58 a.C. ?

Roma, 28 settembre 29.



Gaius Iulius Caesar Octavianus Augustus

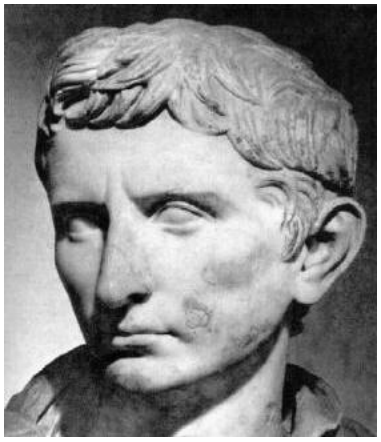
Roma, 23 settembre 63 a.C. ? Nola, 19 agosto 14 d.C

- Ottaviano era nipote naturale e figlio adottivo di Giulio Cesare e fu imperatore molto a lungo durante e dopo la guerra civile.

La cifratura di Ottaviano Augusto

Livia Drusilla Claudia; Roma, 30 gennaio 58 a.C. ?

Roma, 28 settembre 29.



Gaius Iulius Caesar Octavianus Augustus

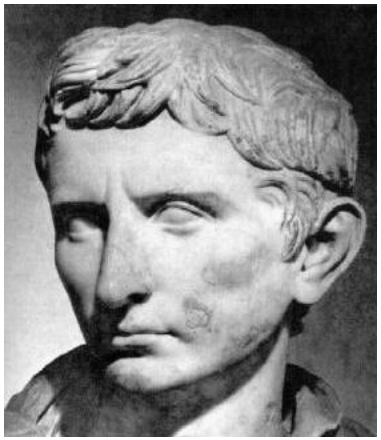
Roma, 23 settembre 63 a.C. ? Nola, 19 agosto 14 d.C

- Ottaviano era nipote naturale e figlio adottivo di Giulio Cesare e fu imperatore molto a lungo durante e dopo la guerra civile.
- Nel comunicare con la moglie Livia Drusilla sulle questioni dell'Urbe usavano una cifratura formidabile. Scrivevano in greco ed utilizzavano il primo libro dell'iliade (di cui avevano due copie identiche) per convertire le lettere in numeri.

La cifratura di Ottaviano Augusto

Livia Drusilla Claudia; Roma, 30 gennaio 58 a.C. ?

Roma, 28 settembre 29.



Gaius Iulius Caesar Octavianus Augustus

Roma, 23 settembre 63 a.C. ? Nola, 19 agosto 14 d.C

- Ottaviano era nipote naturale e figlio adottivo di Giulio Cesare e fu imperatore molto a lungo durante e dopo la guerra civile.
- Nel comunicare con la moglie Livia Drusilla sulle questioni dell'Urbe usavano una cifratura formidabile. Scrivevano in greco ed utilizzavano il primo libro dell'iliade (di cui avevano due copie identiche) per convertire le lettere in numeri.
- Ogni lettera era numerata sequenzialmente per cui ad ogni carattere corrispondevano molti numeri che, quindi, avevano frequenze confrontabili.

La cifratura polialfabetica

- La cifratura di Augusto è un esempio di cifratura polialfabetica in cui la corrispondenza sostituzionale è polidroma (uno a tanti).

La cifratura polialfabetica

- La cifratura di Augusto è un esempio di cifratura polialfabetica in cui la corrispondenza sostituzionale è polidroma (uno a tanti).
- Le cifrature polialfabetice bilanciate sono fatte in modo che tutti i caratteri usati abbiano frequenze simili; quindi la tecnica di analisi delle frequenze fallisce miseramente.

La cifratura polialfabetica

- La cifratura di Augusto è un esempio di cifratura polialfabetica in cui la corrispondenza sostituzionale è polidroma (uno a tanti).
- Le cifrature polialfabetice bilanciate sono fatte in modo che tutti i caratteri usati abbiano frequenze simili; quindi la tecnica di analisi delle frequenze fallisce miseramente.
- Ovviamente questo rafforzamento della cifratura ha un costo di ridondanza: si è costretti a ricorrere ad un alfabeto con più caratteri con un conseguente allungamento del testo cifrato rispetto al testo in chiaro.

La cifratura polialfabetica

- La cifratura di Augusto è un esempio di cifratura polialfabetica in cui la corrispondenza sostituzionale è polidroma (uno a tanti).
- Le cifrature polialfabetice bilanciate sono fatte in modo che tutti i caratteri usati abbiano frequenze simili; quindi la tecnica di analisi delle frequenze fallisce miseramente.
- Ovviamente questo rafforzamento della cifratura ha un costo di ridondanza: si è costretti a ricorrere ad un alfabeto con più caratteri con un conseguente allungamento del testo cifrato rispetto al testo in chiaro.
- I messaggi di Augusto restarono enigmatici per molti anni, fino a che non fu scoperta una copia dell'Iliade della moglie in cui sopra le lettere erano scritti dei numeri sequenziali. Sostituendo ai numeri dei messaggi le lettere corrispondenti si decifrarono tutti i messaggi scambiati dai coniugi.

La cifratura polialfabetica - cont

- La cifratura polialfabetica previene la decrittazione basata sulle frequenze delle lettere.

La cifratura polialfabetica - cont

- La cifratura polialfabetica previene la decrittazione basata sulle frequenze delle lettere.
- Se il testo è abbastanza lungo si possono calcolare le frequenze dei digrammi (sequenze di due lettere). La codifica polialfabetica non protegge contro questa tecnica crittoanalitica. Ma la cifratura di Augusto è robusta anche rispetto a questa analisi.

La cifratura polialfabetica - cont

- La cifratura polialfabetica previene la decrittazione basata sulle frequenze delle lettere.
- Se il testo è abbastanza lungo si possono calcolare le frequenze dei digrammi (sequenze di due lettere). La codifica polialfabetica non protegge contro questa tecnica crittoanalitica. Ma la cifratura di Augusto è robusta anche rispetto a questa analisi.
- In questo caso gruppi di caratteri codificati (numeri) diversi vengono aggregati tra loro in modo da ottenere frequenze simili a quelle delle lettere che si suppone codifichino e rispettino le frequenze dei digrammi.

La cifratura polialfabetica - cont

- La cifratura polialfabetica previene la decrittazione basata sulle frequenze delle lettere.
- Se il testo è abbastanza lungo si possono calcolare le frequenze dei digrammi (sequenze di due lettere). La codifica polialfabetica non protegge contro questa tecnica crittoanalitica. Ma la cifratura di Augusto è robusta anche rispetto a questa analisi.
- In questo caso gruppi di caratteri codificati (numeri) diversi vengono aggregati tra loro in modo da ottenere frequenze simili a quelle delle lettere che si suppone codifichino e rispettino le frequenze dei digrammi.
- La tecnica può essere estesa ai poligrammi aumentando i vincoli per l'identificazione dei caratteri alfabetici.

Misurare l'informazione

- In generale osservando un flusso di dati si può tentare di svelare il contenuto originale. Un problema connesso a questo è la misura dell'informazione estraibile da (o contenuta in) un flusso di dati.

Misurare l'informazione

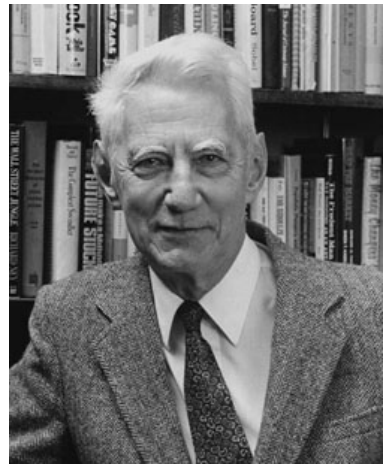
- In generale osservando un flusso di dati si può tentare di svelare il contenuto originale. Un problema connesso a questo è la misura dell'informazione estraibile da (o contenuta in) un flusso di dati.
- Il concetto di Entropia informazionale nasce proprio dall'esigenza di quantificare il contenuto informativo di sequenze astratte di segnali.

Misurare l'informazione

- In generale osservando un flusso di dati si può tentare di svelare il contenuto originale. Un problema connesso a questo è la misura dell'informazione estraibile da (o contenuta in) un flusso di dati.
- Il concetto di Entropia informazionale nasce proprio dall'esigenza di quantificare il contenuto informativo di sequenze astratte di segnali.
- Affinché il concetto risponda bene alle nostre esigenze deve dare un contenuto informativo nullo alle sequenze di numeri casuali.

Concetto di Entropia

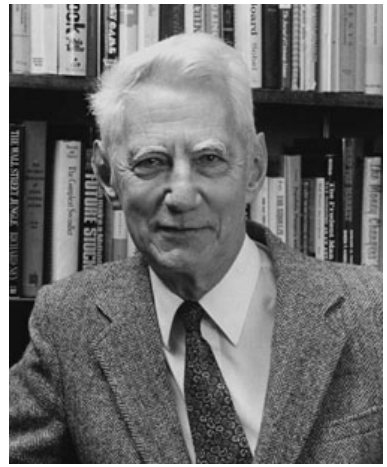
Il termine Entropia assume significati molteplici. Noi ci occuperemo dell'entropia informazionale ed utilizzeremo l'approccio classico che si deve a Claude Shannon. Per i suoi contributi negli anni dopo la seconda guerra mondiale, viene considerato il "padre della teoria dell'informazione".



Claude Shannon 30 aprile 1916, Petoskey, Michigan, USA; 24 febbraio 2001, Medford, Massachusetts, USA. "Il padre della teoria dell'inf

Concetto di Entropia

Il termine Entropia assume significati molteplici. Noi ci occuperemo dell'entropia informazionale ed utilizzeremo l'approccio classico che si deve a Claude Shannon. Per i suoi contributi negli anni dopo la seconda guerra mondiale, viene considerato il "padre della teoria dell'informazione".

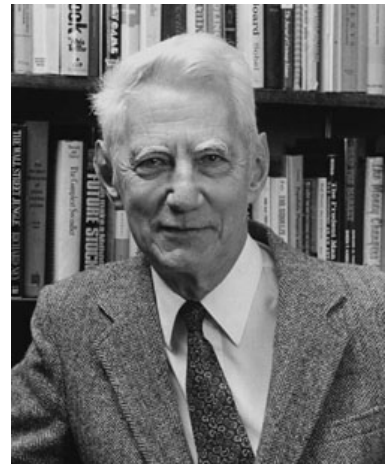


Claude Shannon 30 aprile 1916, Petoskey, Michigan, USA; 24 febbraio 2001, Medford, Massachusetts, USA. "Il padre della teoria dell'inf

- Il termine entropia nasce in fisica ad opera di Rudolf Clausius (a metà del 1800) come funzione di stato che misura il livello di disordine dei sistemi termodinamici e contribuisce all'equilibrio termodinamico.

Concetto di Entropia

Il termine Entropia assume significati molteplici. Noi ci occuperemo dell'entropia informazionale ed utilizzeremo l'approccio classico che si deve a Claude Shannon. Per i suoi contributi negli anni dopo la seconda guerra mondiale, viene considerato il "padre della teoria dell'informazione".



Claude Shannon 30 aprile 1916, Petoskey, Michigan, USA; 24 febbraio 2001, Medford, Massachusetts, USA. "Il padre della teoria dell'inf

- Il termine entropia nasce in fisica ad opera di Rudolf Clausius (a metà del 1800) come funzione di stato che misura il livello di disordine dei sistemi termodinamici e contribuisce all'equilibrio termodinamico.
- L'etimologia del termine viede da ἐν en, "dentro", e τροπή tropé "movimento" e significa "cambiamento interno". Un cambiamento del sistema non legato alla sua energia.

Entropia in fisica

- In meccanica statistica ed in teoria dei campi (in particolare nella formulazione di Gibbs) si lega indissolubilmente l'entropia termodinamica al numero degli stati microscopici che caratterizzano una stato macroscopico. Precisamente l'entropia è il logaritmo del numero di tali stati.

Entropia in fisica

- In meccanica statistica ed in teoria dei campi (in particolare nella formulazione di Gibbs) si lega indissolubilmente l'entropia termodinamica al numero degli stati microscopici che caratterizzano una stato macroscopico. Precisamente l'entropia è il logaritmo del numero di tali stati.
- L'entropia di Shannon è l'applicazione naturale dei concetti fondamentali della meccanica statistica ai sistemi informazionali.

Misurare il contenuto informativo

- Una immagine, un testo, dei segni sulle pareti di una grotta possono contenere informazione. In alcuni casi il messaggio contenuto è semplice in altri casi meno chiaro. Il problema è come distinguere segni privi di significato da altri con un contenuto.

Misurare il contenuto informativo

- Una immagine, un testo, dei segni sulle pareti di una grotta possono contenere informazione. In alcuni casi il messaggio contenuto è semplice in altri casi meno chiaro. Il problema è come distinguere segni privi di significato da altri con un contenuto.
- Andando oltre gli aspetti qualitativi, il problema primario consiste nel quantificare (misurare) il contenuto informativo di una produzione umana. Scrivere $E = mc^2$ in piccoli caratteri o a caratteri cubitali non altera il contenuti informativo dell'equazione. Quindi il problema non è misurare la grandezza, l'eleganza o il numero dei caratteri.

Misurare il contenuto informativo

- Una immagine, un testo, dei segni sulle pareti di una grotta possono contenere informazione. In alcuni casi il messaggio contenuto è semplice in altri casi meno chiaro. Il problema è come distinguere segni privi di significato da altri con un contenuto.
- Andando oltre gli aspetti qualitativi, il problema primario consiste nel quantificare (misurare) il contenuto informativo di una produzione umana. Scrivere $E = mc^2$ in piccoli caratteri o a caratteri cubitali non altera il contenuti informativo dell'equazione. Quindi il problema non è misurare la grandezza, l'eleganza o il numero dei caratteri.
- Misurare la rarità di un evento è il punto di partenza per comprendere il legame tra informazione e disordine.

Contenuto informativo di un evento

- Una funzione che misuri il contenuto informativo di un evento deve dipendere dalla sua probabilità:

$$\mathcal{I}(A) \stackrel{def}{=} f(p(A)).$$

Contenuto informativo di un evento

- Una funzione che misuri il contenuto informativo di un evento deve dipendere dalla sua probabilità:

$$\mathcal{I}(A) \stackrel{def}{=} f(p(A)).$$

- Ma non può essere $p(A)$ perché più è piccola la probabilità e più sappiamo sul sistema. Quindi f deve essere una funzione monotona decrescente:

$$\forall x < y : f(x) > f(y).$$

Contenuto informativo di un evento

- Una funzione che misuri il contenuto informativo di un evento deve dipendere dalla sua probabilità:

$$\mathcal{I}(A) \stackrel{def}{=} f(p(A)).$$

- Ma non può essere $p(A)$ perché più è piccola la probabilità e più sappiamo sul sistema. Quindi f deve essere una funzione monotona decrescente:

$$\forall x < y : f(x) > f(y).$$

- Per precisare la definizione si impongono dei vincoli "naturali" sulla funzione:

Contenuto informativo di un evento

- Una funzione che misuri il contenuto informativo di un evento deve dipendere dalla sua probabilità:

$$\mathcal{I}(A) \stackrel{def}{=} f(p(A)).$$

- Ma non può essere $p(A)$ perché più è piccola la probabilità e più sappiamo sul sistema. Quindi f deve essere una funzione monotona decrescente:

$$\forall x < y : f(x) > f(y).$$

- Per precisare la definizione si impongono dei vincoli "naturali" sulla funzione:
- Innanzitutto si chiede che sia una funzione positiva perché un contenuto informativo negativo è di difficile interpretazione.

Contenuto informativo di un evento

- Una funzione che misuri il contenuto informativo di un evento deve dipendere dalla sua probabilità:

$$\mathcal{I}(A) \stackrel{def}{=} f(p(A)).$$

- Ma non può essere $p(A)$ perché più è piccola la probabilità e più sappiamo sul sistema. Quindi f deve essere una funzione monotona decrescente:

$$\forall x < y : f(x) > f(y).$$

- Per precisare la definizione si impongono dei vincoli "naturali" sulla funzione:
- Innanzitutto si chiede che sia una funzione positiva perché un contenuto informativo negativo è di difficile interpretazione.
- Deve annullarsi quando la probabilità dell'evento che si verifica è unitaria perché l'evento è "quasi sicuro" e non traiamo alcuna informazione aggiuntiva:

$$f(1) = 0.$$

Ipotesi sul legame contenuto informativo di un evento e probabilità

- La funzione f deve divergere per $p(A)$ tendente a zero perché se si verifica un evento quasi impossibile l'informazione cresce indefinitamente.

$$\lim_{x \to 0^+} f(x) = \infty.$$

Ipotesi sul legame contenuto informativo di un evento e probabilità

- La funzione f deve divergere per $p(A)$ tendente a zero perché se si verifica un evento quasi impossibile l'informazione cresce indefinitamente.

$$\lim_{x \to 0^+} f(x) = \infty.$$

- Se si verificano due eventi indipendenti l'informazione acquisita deve essere la somma delle informazioni legate ai singoli eventi:

$$\mathcal{I}(A \cap B) = \mathcal{I}(A) + \mathcal{I}(B);$$

che, essendo $\mathcal{P}(A \cap B) = \mathcal{P}(A)\mathcal{P}(B)$, equivale alla proprietà di fattorizzazione della f:

$$f(x \cdot y) = f(x) + f(y).$$

Ipotesi sul legame contenuto informativo di un evento e probabilità

- La funzione f deve divergere per $p(A)$ tendente a zero perché se si verifica un evento quasi impossibile l'informazione cresce indefinitamente.

$$\lim_{x \to 0^+} f(x) = \infty.$$

- Se si verificano due eventi indipendenti l'informazione acquisita deve essere la somma delle informazioni legate ai singoli eventi:

$$\mathcal{I}(A \cap B) = \mathcal{I}(A) + \mathcal{I}(B);$$

che, essendo $\mathcal{P}(A \cap B) = \mathcal{P}(A)\mathcal{P}(B)$, equivale alla proprietà di fattorizzazione della f:

$$f(x \cdot y) = f(x) + f(y).$$

- Si impone che il contenuto informativo sia una funzione continua della probabilità.

Ipotesi sul legame contenuto informativo di un evento e probabilità

- Infine si fissa una normalizzazione (arbitrariamente):

$$\mathcal{I}(A) = 1 \Leftrightarrow p(A) = 1/e;$$

che equivale a:

$$f(1/e) = 1.$$

Ipotesi sul legame contenuto informativo di un evento e probabilità

- Infine si fissa una normalizzazione (arbitrariamente):

$$\mathcal{I}(A) = 1 \Leftrightarrow p(A) = 1/e;$$

che equivale a:

$$f(1/e) = 1.$$

- Con queste condizioni la funzione f è definita univocamente:

$$\mathcal{I}(A) \stackrel{def}{=} -log(p(A)) = log\left(\frac{1}{p(A)}\right).$$

Altre normalizzazioni porterebbero al logaritmo in altre basi.

Unicità della definizione di Shannon

- Per dimostrarlo basta costruire per una qualsiasi $f(x)$, la funzione $g(x) = f(x) + log(x)$ e verificare che è identicamente nulla. Poniamo $x_1 = 1/e$, valgono le proprietà:

Unicità della definizione di Shannon

- Per dimostrarlo basta costruire per una qualsiasi $f(x)$, la funzione $g(x) = f(x) + log(x)$ e verificare che è identicamente nulla. Poniamo $x_1 = 1/e$, valgono le proprietà:
- $$g(x_1) = f(x_1) + log(1/e) = 1 - 1 = 0;$$

Unicità della definizione di Shannon

- Per dimostrarlo basta costruire per una qualsiasi $f(x)$, la funzione $g(x) = f(x) + log(x)$ e verificare che è identicamente nulla. Poniamo $x_1 = 1/e$, valgono le proprietà:
- $$g(x_1) = f(x_1) + log(1/e) = 1 - 1 = 0;$$
- La g si annulla per tutte le potenze di x_1:

$$g((x_1)^m) = f((x_1)^m) + log(x_1^m) = mf(x_1) + m \cdot log(x_1) = m - m = 0;$$

Unicità della definizione di Shannon

- Per dimostrarlo basta costruire per una qualsiasi $f(x)$, la funzione $g(x) = f(x) + log(x)$ e verificare che è identicamente nulla. Poniamo $x_1 = 1/e$, valgono le proprietà:
- $$g(x_1) = f(x_1) + log(1/e) = 1 - 1 = 0;$$
- La g si annulla per tutte le potenze di x_1:

$$g((x_1)^m) = f((x_1)^m) + log(x_1^m) = mf(x_1) + m \cdot log(x_1) = m - m = 0;$$

- La g si annulla per tutte le radici di x_1:

$$g((x_1)^{1/n}) = n \cdot (1/n) g((x_1)^{1/n}) = (1/n) f\left(\left((x_1)^{1/n}\right)^n\right) + n log(x_1^{1/n})$$

$$g((x_1)^{1/n}) = f(x_1) + n(1/n) \cdot log(x_1) = g(x_1) = 0;$$

Unicità della definizione di Shannon

- Per dimostrarlo basta costruire per una qualsiasi $f(x)$, la funzione $g(x) = f(x) + log(x)$ e verificare che è identicamente nulla. Poniamo $x_1 = 1/e$, valgono le proprietà:
- $$g(x_1) = f(x_1) + log(1/e) = 1 - 1 = 0;$$
- La g si annulla per tutte le potenze di x_1:

$$g((x_1)^m) = f((x_1)^m) + log(x_1^m) = mf(x_1) + m \cdot log(x_1) = m - m = 0;$$

- La g si annulla per tutte le radici di x_1:

$$g((x_1)^{1/n}) = n \cdot (1/n) g((x_1)^{1/n}) = (1/n) f\left(\left((x_1)^{1/n}\right)^n\right) + n log(x_1^{1/n})$$

$$g((x_1)^{1/n}) = f(x_1) + n(1/n) \cdot log(x_1) = g(x_1) = 0;$$

- La g si annulla per tutte le potenze razionali di x_1:

$$\forall m, n : g((x_1)^{m/n}) = (m/n) \cdot g((x_1)) = 0;$$

Unicità della definizione di Shannon

- Per dimostrarlo basta costruire per una qualsiasi $f(x)$, la funzione $g(x) = f(x) + log(x)$ e verificare che è identicamente nulla. Poniamo $x_1 = 1/e$, valgono le proprietà:
- $$g(x_1) = f(x_1) + log(1/e) = 1 - 1 = 0;$$
- La g si annulla per tutte le potenze di x_1:

$$g((x_1)^m) = f((x_1)^m) + log(x_1^m) = mf(x_1) + m \cdot log(x_1) = m - m = 0;$$

- La g si annulla per tutte le radici di x_1:

$$g((x_1)^{1/n}) = n \cdot (1/n) g((x_1)^{1/n}) = (1/n) f\left(\left((x_1)^{1/n}\right)^n\right) + n log(x_1^{1/n})$$

$$g((x_1)^{1/n}) = f(x_1) + n(1/n) \cdot log(x_1) = g(x_1) = 0;$$

- La g si annulla per tutte le potenze razionali di x_1:

$$\forall m, n : g((x_1)^{m/n}) = (m/n) \cdot g((x_1)) = 0;$$

- I numeri $X_1^{m/n}$ sono densi nell'intervallo (0,1] e quindi, per la continuità, la funzione g è nulla in tutto l'intervallo $(0, 1]$.

Entropia di Shannon

- La definizione di contenuto informativo ci consente di definire la Entropia di una sorgente

Entropia di Shannon

- La definizione di contenuto informativo ci consente di definire la Entropia di una sorgente
- Una sorgente è un sistema che genera una sequenza di variabili stocastiche discrete. Dal punto di vista probabilistico genera una catena stocastica: $\xi = \{\xi_1, \xi_2, \ldots, \xi_i, \ldots\}$

Entropia di Shannon

- La definizione di contenuto informativo ci consente di definire la Entropia di una sorgente
- Una sorgente è un sistema che genera una sequenza di variabili stocastiche discrete. Dal punto di vista probabilistico genera una catena stocastica: $\xi = \{\xi_1, \xi_2, \ldots, \xi_i, \ldots\}$
- L'entropia della sorgente è il valore d'attesa dell'informazione della catena stocastica:

$$h(s) \stackrel{def}{=} E[\mathcal{I}(\xi)] = E[\mathcal{I}(\xi)].$$

Entropia di Shannon

- La definizione di contenuto informativo ci consente di definire la Entropia di una sorgente
- Una sorgente è un sistema che genera una sequenza di variabili stocastiche discrete. Dal punto di vista probabilistico genera una catena stocastica: $\xi = \{\xi_1, \xi_2, \ldots, \xi_i, \ldots\}$
- L'entropia della sorgente è il valore d'attesa dell'informazione della catena stocastica:

$$h(s) \stackrel{def}{=} E[\mathcal{I}(\xi)] = E[\mathcal{I}(\xi)].$$

- Vedremo in seguito il caso generale, vediamo adesso il caso di una catena stazionaria senza memoria.

Entropia di una sorgente senza memoria

- Una sorgente si dice senza memoria quando la sequenza di variabili stocastiche da essa generata è costituita da variabili stocastiche identiche ed indipendenti tra loro.

Entropia di una sorgente senza memoria

- Una sorgente si dice senza memoria quando la sequenza di variabili stocastiche da essa generata è costituita da variabili stocastiche identiche ed indipendenti tra loro.
- L'entropia in questo caso si definisce (o calcola) come attesa dell'informazione a ciascuna singola variabile $\xi = \xi^i$ della catena stocastica:

$$h \stackrel{def}{=} E[\mathcal{I}(\xi)] = -\sum_{i=1}^{M} \rho_i \cdot log(\rho_i).$$

Proprietà dell'entropia

- I valori a probablità nulla non contribuiscono all'entropia:

$$\lim_{\rho_i \to 0} \rho_i \cdot log(\rho_i) = 0.$$

Proprietà dell'entropia

- I valori a probablità nulla non contribuiscono all'entropia:

$$\lim_{\rho_i \to 0} \rho_i \cdot log(\rho_i) = 0.$$

- L'entropia di una sorgente che emette un numero finito di caratteri è limitata superiormente. Essendo il numero di valori assumibili finito, è possibile calcolare il limite superiore dell'entropia analiticamente. Bisogna massimizzare h con il vincolo imposto dalla normalizzazione:

$$\begin{cases} h &= -\sum_{i=1,M} \rho_i \cdot log(\rho_i) \\ 1 &= \sum_{i=1,M} \rho_i. \end{cases}$$

Proprietà dell'entropia

- I valori a probablità nulla non contribuiscono all'entropia:

$$\lim_{\rho_i \to 0} \rho_i \cdot log(\rho_i) = 0.$$

- L'entropia di una sorgente che emette un numero finito di caratteri è limitata superiormente. Essendo il numero di valori assumibili finito, è possibile calcolare il limite superiore dell'entropia analiticamente. Bisogna massimizzare h con il vincolo imposto dalla normalizzazione:

$$\begin{cases} h &= -\sum_{i=1,M} \rho_i \cdot log(\rho_i) \\ 1 &= \sum_{i=1,M} \rho_i. \end{cases}$$

- Derivando rispetto a ρ_i, con il metodo dei moltiplicatori di Lagrange:

$$-\rho_i \cdot \frac{1}{\rho_i} - log(\rho_i) = \lambda$$

cioè i ρ_i sono tutti uguali e dunque $\rho_i = \frac{1}{M}$.

Massima entropia di una sorgente

- Corrisponde al caso uniforme in cui tutte le ρ_i sono pari a $1/M$:

$$h_{max} = -\sum_{i=1}^{M} \rho_i \cdot log(\rho_i) = -\sum_{i=1}^{M} \frac{1}{M} \cdot log(1/M) = log(M).$$

L'entropia massima corrisponde al logaritmo del numero degli stati.

Massima entropia di una sorgente

- Corrisponde al caso uniforme in cui tutte le ρ_i sono pari a $1/M$:

$$h_{max} = -\sum_{i=1}^{M} \rho_i \cdot log(\rho_i) = -\sum_{i=1}^{M} \frac{1}{M} \cdot log(1/M) = log(M).$$

  L'entropia massima corrisponde al logaritmo del numero degli stati.

- In generale data un qualsiasi valore dell'entropia esiste sempre il numero di stati equivalente, detto indice di Nei:

$$h = log(N_{stati});$$

  che corrisponde al numero minimo di stati con cui si può ottenere la stessa entropia:

$$N_{stati} = e^h.$$

Massima entropia di una sorgente

- Corrisponde al caso uniforme in cui tutte le ρ_i sono pari a $1/M$:

$$h_{max} = -\sum_{i=1}^{M} \rho_i \cdot log(\rho_i) = -\sum_{i=1}^{M} \frac{1}{M} \cdot log(1/M) = log(M).$$

  L'entropia massima corrisponde al logaritmo del numero degli stati.

- In generale data un qualsiasi valore dell'entropia esiste sempre il numero di stati equivalente, detto indice di Nei:

$$h = log(N_{stati});$$

  che corrisponde al numero minimo di stati con cui si può ottenere la stessa entropia:

$$N_{stati} = e^h.$$

- Il valore minimo dell'entropia, cioè zero, si ottiene quando la probabilità è concentrata in un solo stato. In questo caso la sorgente deve emette un solo carattere.

Messaggio

- Reiterando il rilevamento dello stato di un sistema è possibile stimare la frequenza relativa di ogni stato. Questo vale anche quando si osserva una sorgente cifrata di caratteri.

Messaggio

- Reiterando il rilevamento dello stato di un sistema è possibile stimare la frequenza relativa di ogni stato. Questo vale anche quando si osserva una sorgente cifrata di caratteri.
- La legge dei grandi numeri ci assicura che la frequenza relativa di uno stato (o di un carattere) tende alla sua probabilità.

Messaggio

- Reiterando il rilevamento dello stato di un sistema è possibile stimare la frequenza relativa di ogni stato. Questo vale anche quando si osserva una sorgente cifrata di caratteri.
- La legge dei grandi numeri ci assicura che la frequenza relativa di uno stato (o di un carattere) tende alla sua probabilità.
- L'analisi delle frequenze dei caratteri consente di decrittare messaggi codificati con qualsiasi criterio sostituzionale (non polialfabetico) purché siano abbastanza lunghi.

Messaggio

- Reiterando il rilevamento dello stato di un sistema è possibile stimare la frequenza relativa di ogni stato. Questo vale anche quando si osserva una sorgente cifrata di caratteri.
- La legge dei grandi numeri ci assicura che la frequenza relativa di uno stato (o di un carattere) tende alla sua probabilità.
- L'analisi delle frequenze dei caratteri consente di decrittare messaggi codificati con qualsiasi criterio sostituzionale (non polialfabetico) purché siano abbastanza lunghi.
- La crittografia di Augusto non è decrittabile con il metodo delle frequenze.

Messaggio

- Reiterando il rilevamento dello stato di un sistema è possibile stimare la frequenza relativa di ogni stato. Questo vale anche quando si osserva una sorgente cifrata di caratteri.
- La legge dei grandi numeri ci assicura che la frequenza relativa di uno stato (o di un carattere) tende alla sua probabilità.
- L'analisi delle frequenze dei caratteri consente di decrittare messaggi codificati con qualsiasi criterio sostituzionale (non polialfabetico) purché siano abbastanza lunghi.
- La crittografia di Augusto non è decrittabile con il metodo delle frequenze.
- In generale: più è lungo un messaggio e più è facile scoprire la chiave.

Messaggio

- Reiterando il rilevamento dello stato di un sistema è possibile stimare la frequenza relativa di ogni stato. Questo vale anche quando si osserva una sorgente cifrata di caratteri.
- La legge dei grandi numeri ci assicura che la frequenza relativa di uno stato (o di un carattere) tende alla sua probabilità.
- L'analisi delle frequenze dei caratteri consente di decrittare messaggi codificati con qualsiasi criterio sostituzionale (non polialfabetico) purché siano abbastanza lunghi.
- La crittografia di Augusto non è decrittabile con il metodo delle frequenze.
- In generale: più è lungo un messaggio e più è facile scoprire la chiave.
- Esiste un indicatore classico dell'informazione contenuta in una sequenza stocastica (ovvero nella sua sorgente) che viene chiamato entropia informazionale o di Shannon.