Cenni di teoria dell'informazione II

Gregorio D'Agostino

17 Maggio 2021

Insiemi convessi

- Due punti x_1 ed x_2 definiscono un insieme convesso $\mathcal{C}$ costituito dalle loro combinazioni lineari a coefficienti positivi e somma unitaria

$$\forall a > 0, b > 0 : a + b = 1 : x \in \mathcal{C} \Rightarrow x = a \cdot x_1 + b \cdot x_2.$$

Insiemi convessi

- Due punti x_1 ed x_2 definiscono un insieme convesso $\mathcal{C}$ costituito dalle loro combinazioni lineari a coefficienti positivi e somma unitaria

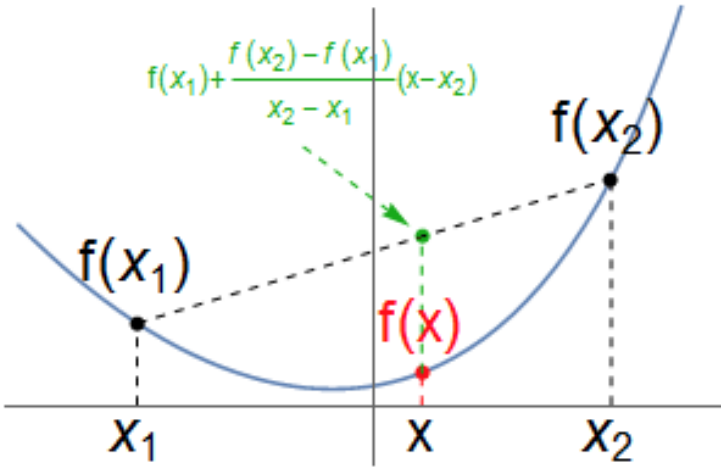
$$\forall a > 0, b > 0 : a + b = 1 : x \in \mathcal{C} \Rightarrow x = a \cdot x_1 + b \cdot x_2.$$

- La definizione si estende a k punti $x_1, x_2, \ldots, x_k$, l'insieme convesso da essi generato $\mathcal{C}$ è definito dalle loro combinazioni lineari a coefficienti positivi e somma unitaria:

$$\forall a_i > 0, \sum_i a_i = 1 : x \in \mathcal{C} \Rightarrow x = \sum_i a_i \cdot x_i.$$

Funzioni convesse

Una funzione si dice convessa se preserva la convessità cioè se il trasformato di qualsiasi combinazione convessa è maggiorato dalla stessa combinazione convessa dei trasformati. Nel caso di due punti:



Esempio di funzione convessa.

$$f\left(a \cdot x_1 + b \cdot x_2\right) \leq a \cdot f(x_1) + b \cdot f(x_2);$$

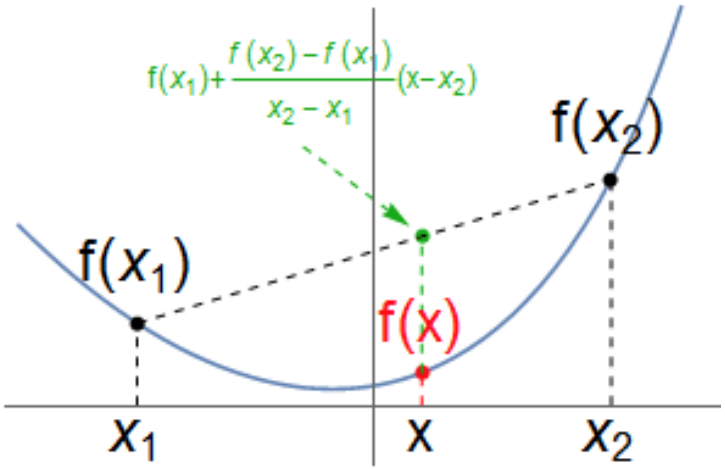
con $a > 0$, $b > 0$ e $a + b = 1$.

- In altri termini, scelto un λ compreso tra zero ed uno, $(0 \leq 1 - \lambda \leq 1)$:

$$f\left((1-\lambda) \cdot x_1 + \lambda \cdot x_2\right) \leq (1-\lambda) \cdot f(x_1) + \lambda \cdot f(x_2).$$

Funzioni convesse

Una funzione si dice convessa se preserva la convessità cioè se il trasformato di qualsiasi combinazione convessa è maggiorato dalla stessa combinazione convessa dei trasformati. Nel caso di due punti:



Esempio di funzione convessa.

$$f\left(a\cdot x_1+b\cdot x_2\right)\leq a\cdot f(x_1)+b\cdot f(x_2);$$

con $a>0$, $b>0$ e $a+b=1$.

- In altri termini, scelto un λ compreso tra zero ed uno, $(0\leq 1-\lambda\leq 1)$:

$$f\left((1-\lambda)\cdot x_1+\lambda\cdot x_2\right)\leq (1-\lambda)\cdot f(x_1)+\lambda\cdot f(x_2).$$

- Le funzioni convesse lisce in un intervallo hanno derivata prima crescente, cioè derivata seconda positiva.

Funzioni strettamente convesse

- Una funzione si dice strettamente convessa se è convessa e l'eguaglianza vale solo per λ nullo o unitario oppure per $x^1 = x^2$.

$$f\left((1-\lambda)\cdot x_1 + \lambda \cdot x_2\right) < (1-\lambda)\cdot f(x_1) + \lambda \cdot f(x_2);$$

per $0 < \lambda < 1$.

Funzioni strettamente convesse

- Una funzione si dice strettamente convessa se è convessa e l'eguaglianza vale solo per λ nullo o unitario oppure per $x^1 = x^2$.

$$f\left((1-\lambda)\cdot x_1 + \lambda \cdot x_2\right) < (1-\lambda)\cdot f(x_1) + \lambda \cdot f(x_2);$$

  per $0 < \lambda < 1$.
- Per λ nullo o unitario vale l'eguaglianza. Per $x^1 = x^2$ l'eguaglianza sussiste per tutti i λ.

Diseguaglianza di Jensen

- Se una funzione convessa definisce una variabile stocastica, allora vale la diseguaglianza:

$$E\left[f(\xi)\right] \geq f\left(E[\xi]\right).$$

Diseguaglianza di Jensen

- Se una funzione convessa definisce una variabile stocastica, allora vale la diseguaglianza:

$$E\left[f(\xi)\right] \geq f\left(E[\xi]\right).$$

- La dimostreremo solo per le variabili discrete:

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

Diseguaglianza di Jensen

- Se una funzione convessa definisce una variabile stocastica, allora vale la diseguaglianza:

$$E\left[f(\xi)\right] \geq f\left(E[\xi]\right).$$

- La dimostreremo solo per le variabili discrete:

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

- Dimostreremo per induzione su M la diseguaglianza tra i termini interni della precedente:

$$\sum_{i=1}^{M} f(x^i) \cdot p_i \geq f\left(\sum_{i=1}^{M} x^i \cdot p_i\right).$$

Diseguaglianza di Jensen - induzione

- Il caso $M = 2$ è la definizione di funzione convessa e quindi è verificato

Diseguaglianza di Jensen - induzione

- Il caso $M = 2$ è la definizione di funzione convessa e quindi è verificato
- Ricorsione su M:

$$S_{M+1} = \sum_{i=1}^{M+1} f(x^i) \cdot p_i = \sum_{i=1}^{M} f(x^i) \cdot p_i + p_{M+1} f(x^{M+1});$$

Diseguaglianza di Jensen - induzione

- Il caso $M = 2$ è la definizione di funzione convessa e quindi è verificato
- Ricorsione su M:

$$S_{M+1} = \sum_{i=1}^{M+1} f(x^i) \cdot p_i = \sum_{i=1}^{M} f(x^i) \cdot p_i + p_{M+1} f(x^{M+1});$$

- Poniamo $\lambda = p_{M+1}$; $1 - \lambda = 1 - p_{M+1} = \sum_{i=1}^{M} p_i$. Per $i = 1, \ldots, M$ definiamo una nuova combinazione lineare normalizzata: $q_i \stackrel{def}{=} p_i/(1 - \lambda)$.

$$S_{M+1} = \sum_{i=1}^{M} f(x^i) \cdot q_i \cdot (1 - \lambda) + \lambda \cdot f(x^{M+1});$$

Diseguaglianza di Jensen - induzione

- Il caso $M = 2$ è la definizione di funzione convessa e quindi è verificato
- Ricorsione su M:

$$S_{M+1} = \sum_{i=1}^{M+1} f(x^i) \cdot p_i = \sum_{i=1}^{M} f(x^i) \cdot p_i + p_{M+1} f(x^{M+1});$$

- Poniamo $\lambda = p_{M+1}$; $1 - \lambda = 1 - p_{M+1} = \sum_{i=1}^{M} p_i$. Per $i = 1, \ldots, M$ definiamo una nuova combinazione lineare normalizzata: $q_i \stackrel{def}{=} p_i/(1 - \lambda)$.

$$S_{M+1} = \sum_{i=1}^{M} f(x^i) \cdot q_i \cdot (1 - \lambda) + \lambda \cdot f(x^{M+1});$$

- Per ipotesi ricorsiva possiamo maggiorare i primi M termini:

$$S_{M+1} \leq (1 - \lambda) f\left(\sum_{i=1}^{M} x^i \cdot q_i\right) + \lambda f(x^{M+1});$$

Diseguaglianza di Jensen - induzione -cont

- Abbiamo ottenuto una combinazione lineare della funzione in due nuovi punti:

$$S_{M+1} \leq (1-\lambda) f\left(\sum_{i=1}^{M} x^i \cdot q_i\right) + \lambda f(x^{M+1});$$

Diseguaglianza di Jensen - induzione -cont

- Abbiamo ottenuto una combinazione lineare della funzione in due nuovi punti:

$$S_{M+1} \leq (1-\lambda) f\left(\sum_{i=1}^{M} x^i \cdot q_i\right) + \lambda f(x^{M+1});$$

- applicando adesso l'ipotesi di convessità:

$$S_{M+1} \leq f\left((1-\lambda)\sum_{i=1}^{M} x^i \cdot q_i + \lambda x^{M+1}\right) = f\left(\sum_{i=1}^{M+1} x^i \cdot p_i\right).$$

Diseguaglianza di Jensen - induzione -cont

- Abbiamo ottenuto una combinazione lineare della funzione in due nuovi punti:

$$S_{M+1} \leq (1-\lambda) f\left(\sum_{i=1}^{M} x^i \cdot q_i\right) + \lambda f(x^{M+1});$$

- applicando adesso l'ipotesi di convessità:

$$S_{M+1} \leq f\left((1-\lambda)\sum_{i=1}^{M} x^i \cdot q_i + \lambda x^{M+1}\right) = f\left(\sum_{i=1}^{M+1} x^i \cdot p_i\right).$$

- Se la funzione è strettamente convessa l'eguaglianza vale solo quando tutti i coefficienti della combinazione lineare sono nulli tranne uno oppure se gli x^i sono tutti uguali.

Una applicazione nota

- Se consideriamo la funzione (strettamente) convessa $f(x) = -log(x)$ ($f''(x) = 1/x^2 > 0$); $x^i = 1/\rho_i$ e per combinazione lineare una distribuzione $a_i = \rho_i$. La diseguaglianza di Jensen diviene:

$$\sum_{i=1}^{M} a_i \cdot f(x^i) = -\sum_{i=1}^{M} \rho_i \cdot log(\rho_i) = \sum_{i=1}^{M} \rho_i \cdot log(1/\rho_i).$$

$$\sum_{i=1}^{M} \rho_i \cdot log(1/\rho_i) \leq log\left(\sum_{i=1}^{M} \rho_i \cdot 1/\rho_i\right) = log(M).$$

Una applicazione nota

- Se consideriamo la funzione (strettamente) convessa $f(x) = -log(x)$ ($f''(x) = 1/x^2 > 0$); $x^i = 1/\rho_i$ e per combinazione lineare una distribuzione $a_i = \rho_i$. La diseguaglianza di Jensen diviene:

$$\sum_{i=1}^{M} a_i \cdot f(x^i) = -\sum_{i=1}^{M} \rho_i \cdot log(\rho_i) = \sum_{i=1}^{M} \rho_i \cdot log(1/\rho_i).$$

$$\sum_{i=1}^{M} \rho_i \cdot log(1/\rho_i) \leq log\left(\sum_{i=1}^{M} \rho_i \cdot 1/\rho_i\right) = log(M).$$

- Abbiamo riottenuto il limite superiore per l'entropia:

$$h \stackrel{def}{=} \sum_{i=1}^{M} log(1/\rho_i) \cdot \rho_i \leq log(M) = h_{max}.$$

Divergenza informazionale

- Date due distribuzioni di probabilità discrete di uguale cardinalità $\mathcal{P} = (p_1, p_2, \ldots, p_M)$, $\mathcal{Q} = (q_1, q_2, \ldots, q_M)$. Ovvero due successioni positive normalizzate:

$$\sum_{i=1}^{M} p_i = \sum_{i=1}^{M} q_i = 1;$$

con $p_i > 0$ e $q_i > 0$.

Divergenza informazionale

- Date due distribuzioni di probabilità discrete di uguale cardinalità $\mathcal{P} = (p_1, p_2, \ldots, p_M)$, $\mathcal{Q} = (q_1, q_2, \ldots, q_M)$. Ovvero due successioni positive normalizzate:

$$\sum_{i=1}^{M} p_i = \sum_{i=1}^{M} q_i = 1;$$

  con $p_i > 0$ e $q_i > 0$.

- La loro Divergenza informazionale $D(\mathcal{P}||\mathcal{Q})$ è definita come segue:

$$D(\mathcal{P}||\mathcal{Q}) \stackrel{def}{=} \sum_{i=1}^{M} p_i \cdot log(p_i/q_i).$$

Divergenza informazionale - Proprietà

- La Divergenza informazionale si annulla quando le due distribuzioni sono uguali.

$$D(\mathcal{P}||\mathcal{Q}) \stackrel{def}{=} \sum_{i=1}^{M} p_i \cdot log(p_i/p_i) = \sum_{i=1}^{M} p_i \cdot log(1) = 0.$$

Divergenza informazionale - Proprietà

- La Divergenza informazionale si annulla quando le due distribuzioni sono uguali.

$$D(\mathcal{P}||\mathcal{Q}) \stackrel{def}{=} \sum_{i=1}^{M} p_i \cdot log(p_i/p_i) = \sum_{i=1}^{M} p_i \cdot log(1) = 0.$$

- La Divergenza informazionale è una grandezza positiva. Per dimostrarlo basta applicare la diseguaglianza di Jensel al caso $f(x) = -log(x)$, $x^i = q_i/p_i$ e $a_i = p_i$

$$D(\mathcal{P}||\mathcal{Q}) = \sum_{i=1}^{M} p_i \cdot log(p_i/q_i) \geq -log\left(\sum_{i=1}^{M} p_i \cdot q_i/p_i\right) = 0.$$

Divergenza informazionale - Proprietà

- La Divergenza informazionale si annulla quando le due distribuzioni sono uguali.

$$D(\mathcal{P}||\mathcal{Q}) \stackrel{def}{=} \sum_{i=1}^{M} p_i \cdot log(p_i/p_i) = \sum_{i=1}^{M} p_i \cdot log(1) = 0.$$

- La Divergenza informazionale è una grandezza positiva. Per dimostrarlo basta applicare la diseguaglianza di Jensel al caso $f(x) = -log(x)$, $x^i = q_i/p_i$ e $a_i = p_i$

$$D(\mathcal{P}||\mathcal{Q}) = \sum_{i=1}^{M} p_i \cdot log(p_i/q_i) \geq -log\left(\sum_{i=1}^{M} p_i \cdot q_i/p_i\right) = 0.$$

- Essendo $-log(x)$ una funzione strettamente convessa, l'eguaglianza si ottiene solo se gli $x^i = q^i/p_i$ sono tutti uguali e quindi $q_i = p_i$.

Misura di indipendenza

- Date due variabili stocastiche discrete ξ ed η possiamo definire la loro distribuzione di probabilità congiunta:

$$\rho_{ij} \stackrel{def}{=} \mathcal{P}(\xi = x^i, \eta = y^j).$$

Misura di indipendenza

- Date due variabili stocastiche discrete ξ ed η possiamo definire la loro distribuzione di probabilità congiunta:

$$\rho_{ij} \stackrel{def}{=} \mathcal{P}(\xi = x^i, \eta = y^j).$$

- Se le variabili stocastiche fossero indipendenti la ρ si fattorizzerebbe:

$$\rho_{ij} = \mathcal{P}(\xi = x^i, \eta = y^j) = \mathcal{P}(\xi = x^i)\mathcal{P}(\eta = y^j) = p_i \cdot q_j;$$

in cui si è posto $q_j \stackrel{def}{=} \mathcal{P}(\eta = y^j)$ e $p_i \stackrel{def}{=} \mathcal{P}(\xi = x^i)$

Misura di indipendenza

- Date due variabili stocastiche discrete ξ ed η possiamo definire la loro distribuzione di probabilità congiunta:

$$\rho_{ij} \stackrel{def}{=} \mathcal{P}(\xi = x^i, \eta = y^j).$$

- Se le variabili stocastiche fossero indipendenti la ρ si fattorizzerebbe:

$$\rho_{ij} = \mathcal{P}(\xi = x^i, \eta = y^j) = \mathcal{P}(\xi = x^i)\mathcal{P}(\eta = y^j) = p_i \cdot q_j;$$

  in cui si è posto $q_j \stackrel{def}{=} \mathcal{P}(\eta = y^j)$ e $p_i \stackrel{def}{=} \mathcal{P}(\xi = x^i)$

- La divergenza delle distribuzioni ρ_{ij} e $p_i \cdot q_j$ fornisce un indice della indipendenza delle variabili. Se le distribuzioni coincidono la divergenza si annulla.

Mutua informazione

- Si definisce Mutua informazione tra due variabili $\mathcal{I}(\xi, \eta)$ la divergenza tra la distribuzione congiunta delle due variabili e quella che avrebbero se fossero indipendenti.

$$\mathcal{I}(\xi \wedge \eta) \stackrel{def}{=} D(\rho_{ij} || p_i q_j).$$

Mutua informazione

- Si definisce Mutua informazione tra due variabili $\mathcal{I}(\xi, \eta)$ la divergenza tra la distribuzione congiunta delle due variabili e quella che avrebbero se fossero indipendenti.

$$\mathcal{I}(\xi \wedge \eta) \stackrel{def}{=} D(\rho_{ij} || p_i q_j).$$

- Esplicitando

$$\mathcal{I}(\xi \wedge \eta) \stackrel{def}{=} \sum_{i=1}^{M} \sum_{J=1}^{M'} \rho_{ij} \cdot log \left(\frac{\rho_{ij}}{p_i \cdot q_i} \right) \geq 0.$$

Mutua informazione

- Si definisce Mutua informazione tra due variabili $\mathcal{I}(\xi, \eta)$ la divergenza tra la distribuzione congiunta delle due variabili e quella che avrebbero se fossero indipendenti.

$$\mathcal{I}(\xi \wedge \eta) \stackrel{def}{=} D(\rho_{ij} || p_i q_j).$$

- Esplicitando

$$\mathcal{I}(\xi \wedge \eta) \stackrel{def}{=} \sum_{i=1}^{M} \sum_{J=1}^{M'} \rho_{ij} \cdot log \left(\frac{\rho_{ij}}{p_i \cdot q_i} \right) \geq 0.$$

- La conoscenza dalla variabile ξ induce una conoscenza parziale sulla variabile η. La mutua informazione quantifica questa grandezza.

Esercizi su Informazione mutua, divergenza ed entropia

- Entropia di una variabile a due valori (A e B). Es $((p(\xi = A), p(\xi = B)) = (1/2, 1/2)$,(0,1), (1/4,3/4) etc).

Esercizi su Informazione mutua, divergenza ed entropia

- Entropia di una variabile a due valori (A e B). Es $((p(\xi = A), p(\xi = B)) = (1/2, 1/2)$,(0,1), (1/4,3/4) etc).
- Divergenza tra due distribuzioni a due valori. Es ((1/2,1/2) e (1/4,3/4))

Esercizi su Informazione mutua, divergenza ed entropia

- Entropia di una variabile a due valori (A e B). Es $((p(\xi = A), p(\xi = B)) = (1/2, 1/2)$,(0,1), (1/4,3/4) etc).
- Divergenza tra due distribuzioni a due valori. Es ((1/2,1/2) e (1/4,3/4))
- Mutua informazione. es $P(\xi, \eta) = \{P(A, A), P(A, B), (B, A), P(B, B)\} = 0, 1/2, 1/2, 0.$
  Calcolare le probabilità "marginali"
  $P(\xi = A) = P(\xi = B) = 1/2.$
  Quale sarebbe la distribuzione congiunta se le variabili stocastiche ξ ed η fossero indipendenti?
  Calcolare la mutua informazione e l'entropia di ξ, η e totale.
  Che c'entra con l'entropia di una variabile uniforme? Come mai la mutua informazione è massima? Vedremo tra poco.

Auto-informazione

- Se due variabili sono identiche cioè assumono sempre gli stessi valori per ogni evento, siamo certi che le variabili sono assolutamente dipendenti. In questo caso la mutua informazione si chiama auto-informazione

$$\mathcal{I}(\xi \wedge \xi) \stackrel{def}{=} \sum_{i=1}^{M} p_i \cdot log\left(\frac{p_i}{p_i \cdot p_i}\right) = \sum_{i=1}^{M} p_i \cdot log\left(\frac{1}{p_i}\right).$$

Auto-informazione

- Se due variabili sono identiche cioè assumono sempre gli stessi valori per ogni evento, siamo certi che le variabili sono assolutamente dipendenti. In questo caso la mutua informazione si chiama auto-informazione

$$\mathcal{I}(\xi \wedge \xi) \stackrel{def}{=} \sum_{i=1}^{M} p_i \cdot log\left(\frac{p_i}{p_i \cdot p_i}\right) = \sum_{i=1}^{M} p_i \cdot log\left(\frac{1}{p_i}\right).$$

- L'autoinformazione coincide con l'entropia della variabile stocastica. Possiamo interpretarlo dicendo che una volta assodato che le due variabili sono identiche l'informazione residua è quella contenuta nella variabile ξ.

Entropia di due variabili

- La definizione è analoga a quella per una variabile:

$$H(\xi,\eta) \stackrel{def}{=} -\sum_{i=1}^{M}\sum_{j=1}^{M'} \mathcal{P}(\xi = x^i, \eta = y^j)\cdot log\left(\mathcal{P}(\xi = x^i, \eta = y^j)\right).$$

Utilizzando le probabilità condizionate :

$$\mathcal{P}(\xi = x^i, \eta = y^j) = \mathcal{P}(\xi = x^i)\cdot\mathcal{P}(\eta = y^j|\xi = x^i);$$

Si può esplicitare la parte di entropia dipendente da ξ:

$$H(\xi,\eta) = -\sum_{i=1}^{M}\sum_{j=1}^{M'} \mathcal{P}(\xi = x^i, \eta = y^j)\cdot log\left(\mathcal{P}(\xi = x^i)\cdot\mathcal{P}(\eta = y^j|\xi = x^i)\right) =$$

$$H(\xi,\eta) = -\sum_{i=1}^{M}\left(\sum_{j=1}^{M'}\mathcal{P}(\xi = x^i, \eta = y^j)\right)\cdot log\left(\mathcal{P}(\xi = x^i)\right) +$$

$$-\sum_{i=1}^{M}\mathcal{P}(\xi = x^i)\left(\sum_{j=1}^{M'}\mathcal{P}(\eta = y^j|\xi = x^i)\cdot log\left(\mathcal{P}(\eta = y^j|\xi = x^i)\right)\right).$$

Entropia di due variabili

- Ricordando che $p_i \stackrel{def}{=} \mathcal{P}(\xi = x^i)$, si può calcolare come distribuzione marginale decomponendola secondo i valori di η:

$$p_i = \sum_{j=1}^{M'} \mathcal{P}(\xi = x^i, \eta = y^j) \stackrel{def}{=} \sum_{j=1}^{M'} \mathcal{P}(\{\xi = x^i\} \cap \{\eta = y^j\}) =$$

$$= \mathcal{P}(\{\xi = x^i\} \cap (\cup_{j=1,M'}\{\eta = y^j\})) = \mathcal{P}(\{\xi = x^i\} \cup \Omega);$$

Quindi il primo termine diviene:

$$-\sum_{i=1}^{M} \left(\sum_{j=1}^{M'} \mathcal{P}(\xi = x^i, \eta = y^j) \right) \cdot log\left(\mathcal{P}(\xi = x^i)\right) =$$

$$= -\sum_{i=1}^{M} p_i \cdot log\left(p_i\right) = H(\xi).$$

Entropia di due variabili

- Definendo le entropie condizionate:

$$H(\eta|\xi = x^i) \stackrel{def}{=} -\sum_{j=1}^{M'} \mathcal{P}(\eta = y^j|\xi = x^i) \cdot log\left(\mathcal{P}(\eta = y^j|\xi = x^i)\right)$$
$$H(\eta|\xi) \stackrel{def}{=} \sum_{i=1}^{M} \mathcal{P}(\xi = x^i) H(\eta|\xi = x^i);$$

Il secondo termine diviene $H(\eta|\xi)$

Entropia di due variabili

- Definendo le entropie condizionate:

$$H(\eta|\xi = x^i) \stackrel{def}{=} -\sum_{j=1}^{M'} \mathcal{P}(\eta = y^j|\xi = x^i) \cdot log\left(\mathcal{P}(\eta = y^j|\xi = x^i)\right)$$
$$H(\eta|\xi) \stackrel{def}{=} \sum_{i=1}^{M} \mathcal{P}(\xi = x^i) H(\eta|\xi = x^i);$$

  Il secondo termine diviene $H(\eta|\xi)$

- Quindi si possono distinguere i due contributi all'entropia :

$$H(\xi, \eta) = H(\xi) + H(\eta|\xi).$$

  L'interpretazione della formula precedente è che l'entropia totale di una coppia di variabili si decompone nell'entropia di una di esse più l'entropia ($H(\eta|\xi)$) dell'altra condizionata dalla prima.

Esercizio su entropia di due variabili

- Es. entropia della coppia ξ ed η con prob. $P(\xi, \eta) = \{P(A, A), P(A, B), (B, A), P(B, B)\} = 1/4, 1/4, 1/4, 1/4.$
  Calcolare le probabilità "marginali":
  $P(\xi = A) = P(\xi = B) = 1/2.$ e identiche per η.
  Verificare che $H(\xi, \eta) = H(\xi) + H(\eta)$
  Quale proprietà della distribuzione congiunta ci dice che variabili stocastiche ξ ed η sono indipendenti?

Esercizio su entropia di due variabili

- Es. entropia della coppia ξ ed η con prob. $P(\xi, \eta) = \{P(A, A), P(A, B), (B, A), P(B, B)\} = 1/4, 1/4, 1/4, 1/4.$
  Calcolare le probabilità "marginali":
  $P(\xi = A) = P(\xi = B) = 1/2.$ e identiche per η.
  Verificare che $H(\xi, \eta) = H(\xi) + H(\eta)$
  Quale proprietà della distribuzione congiunta ci dice che variabili stocastiche ξ ed η sono indipendenti?
- Ripetere nel caso della coppia ξ ed η con prob. $P(\xi, \eta) = \{P(A, A), P(A, B), (B, A), P(B, B)\} = 1/8, 1/4, 1/4, 3/8.$

Entropia di due variabili e mutua informazione

Riprendiamo l'entropia di due variabili:

$$H(\xi,\eta) = H(\xi) + \sum_{i=1}^{M} \mathcal{P}(\xi = x^i) H(\eta|\xi = x^i) = H(\xi) + H(\eta|\xi).$$

L'interpretazione della formula precedente è che l'entropia totale di una coppia di variabili si decompone nell'entropia di una di esse per l'entropia dell'altra condizionata dalla prima.
Riprendiamo la definizione di mutua informazione:

$$\mathcal{I}(\xi \wedge \eta) \stackrel{def}{=} \sum_{i=1}^{M} \sum_{J=1}^{M'} \rho_{ij} \cdot log\left(\frac{\rho_{ij}}{p_i \cdot q_j}\right) =$$

$$= \sum_{i=1}^{M} \sum_{J=1}^{M'} \rho_{ij} \cdot log\left(\frac{P(\xi = x_i|\eta = y_j) q_j}{p_i \cdot q_j}\right);$$

Entropia di due variabili e mutua informazione - cont

$$\mathcal{I}(\xi \wedge \eta) = \sum_{i=1}^{M} \sum_{J=1}^{M'} \rho_{ij} \cdot log \left(\frac{P(\xi = x_i | \eta = y_j)}{p_i} \right);$$

$$\mathcal{I}(\xi \wedge \eta) = \sum_{i=1}^{M} \sum_{J=1}^{M'} \rho_{ij} \cdot log \left(\frac{1}{p_i} \right) + \sum_{i=1}^{M} \sum_{J=1}^{M'} \rho_{ij} \cdot log \left(P(\xi = x_i | \eta = y_j) \right)$$

Il primo addendo è l'entropia di ξ perché $\sum_{J=1}^{M'} \rho_{ij} = p_i$.
Il secondo addendo è l'entropia di ξ condizionata da η (cambiata di segno):

$$\sum_{i=1}^{M} \sum_{J=1}^{M'} q_i P(\xi = x_i | \eta = y_j) \cdot log \left(P(\xi = x_i | \eta = y_j) \right) = H(\xi | \eta);$$

Quindi:

$$\mathcal{I}(\xi \wedge \eta) = H(\xi) - H(\xi | \eta).$$

Legame tra l'entropia condizionata e non condizionata

Abbiamo visto che la mutua informazione è definita positiva, quindi:

$$H(\xi) - H(\xi|\eta) = \mathcal{I}(\xi \wedge \eta) > o.$$

Questo significa che l'entropia di una variabile aleatoria condizionata da un'altra è minore dell'entropia non condizionata:

$$H(\xi) > H(\xi|\eta).$$

Analogamente l'entropia condizionata da un insieme di variabili è minore dell'entropia condizionata da un suo sottinsieme.

$$H(\xi|\eta_1, \eta_2) < H(\xi|\eta_2).$$

Entropia di due variabili stocastiche

In generale (usando $H(\eta|\xi) = H(\eta) - \mathcal{I}(\xi \wedge \eta)$) possiamo scrivere:

$$H(\xi,\eta) = H(\xi) + H(\eta|\xi) = H(\xi) + H(\eta) - \mathcal{I}(\xi \wedge \eta) \leq H(\xi) + H(\eta).$$

L'entropia di una coppia di variabili stocastiche è minore della somma delle loro entropie.

- Se le variabili sono indipendenti l'entropia totale è la somma delle entropie, perché la mutua informazione $\mathcal{I}(\xi \wedge \eta)$ è nulla.

Entropia di due variabili stocastiche

In generale (usando $H(\eta|\xi) = H(\eta) - \mathcal{I}(\xi \wedge \eta)$) possiamo scrivere:

$$H(\xi,\eta) = H(\xi) + H(\eta|\xi) = H(\xi) + H(\eta) - \mathcal{I}(\xi \wedge \eta) \leq H(\xi) + H(\eta).$$

L'entropia di una coppia di variabili stocastiche è minore della somma delle loro entropie.

- Se le variabili sono indipendenti l'entropia totale è la somma delle entropie, perché la mutua informazione $\mathcal{I}(\xi \wedge \eta)$ è nulla.
- Vedremo che se le variabili sono univocamente dipendenti l'entropia condizionata è nulla.

Entropia di due variabili stocastiche

In generale (usando $H(\eta|\xi) = H(\eta) - \mathcal{I}(\xi \wedge \eta)$) possiamo scrivere:

$$H(\xi,\eta) = H(\xi) + H(\eta|\xi) = H(\xi) + H(\eta) - \mathcal{I}(\xi \wedge \eta) \leq H(\xi) + H(\eta).$$

L'entropia di una coppia di variabili stocastiche è minore della somma delle loro entropie.

- Se le variabili sono indipendenti l'entropia totale è la somma delle entropie, perché la mutua informazione $\mathcal{I}(\xi \wedge \eta)$ è nulla.
- Vedremo che se le variabili sono univocamente dipendenti l'entropia condizionata è nulla.
- Quindi la mutua informazione è la differenza tra la somma delle entropie di sue variabili e l'entropia del sistema composto da entrambe:

$$\mathcal{I}(\xi \wedge \eta) = H(\xi) + H(\eta) - H(\xi,\eta).$$

Entropia condizionata da una variabile dipendente

Se una variabile è una funzione invertibile dell'altra $\xi = f(\eta)$ l'entropia condizionata è nulla. Partendo dalla definizione generale:

$$H(\xi|\eta) = \sum_{j=1}^{M'} \mathcal{P}(\eta = y^j) \cdot \left(\sum_{i=1}^{M} \mathcal{P}(\xi = x^i|\eta = y^j) \cdot log \left(\mathcal{P}(\xi = x^i|\eta = y^j) \right) \right).$$

$$H(\xi|\eta) = \sum_{j=1}^{M'} \mathcal{P}(\eta = y^j) \cdot \left(\sum_{i=1}^{M} \mathcal{P}(\xi = f(y^i)|\eta = y^j) \cdot log \left(\mathcal{P}(\xi = f(y^i)|\eta = y^j)) \right) \right).$$

La probabilità condizionata $\mathcal{P}(\xi = f(y^i)|\eta = y^j)$ è diversa da zero solo se $x^i = f(y^i) = f(y^j)$ cioè $f(y^i) = f(y^j)$. Se la funzione f è invertibile $f(y^i) = f(y^j)$ equivale a $y^i = y^j$ ed in questo caso $\mathcal{P}(\xi = f(y^j)|\eta = y^j)$ è uguale ad uno. Il log di uno è zero, quindi tutti i termini sono nulli.

Esercizi su Informazione mutua, divergenza ed entropia

- Entropia di una variabile a due valori (A e B). Es $((p(\xi = A), p(\xi = B)) = (1/2, 1/2)$,(0,1), (1/4,3/4) etc).

Esercizi su Informazione mutua, divergenza ed entropia

- Entropia di una variabile a due valori (A e B). Es $((p(\xi = A), p(\xi = B)) = (1/2, 1/2)$,(0,1), (1/4,3/4) etc).
- Divergenza tra due distribuzioni a due valori. Es ((1/2,1/2) e (1/4,3/4))

Esercizi su Informazione mutua, divergenza ed entropia

- Entropia di una variabile a due valori (A e B). Es $((p(\xi = A), p(\xi = B)) = (1/2, 1/2)$,(0,1), (1/4,3/4) etc).
- Divergenza tra due distribuzioni a due valori. Es ((1/2,1/2) e (1/4,3/4))
- Mutua informazione. es $P(\xi, \eta) = \{P(A, A), P(A, B), (B, A), P(B, B)\} = 0, 1/2, 1/2, 0.$
  Calcolare le probabilità "marginali"
  $P(\xi = A) = P(\xi = B) = 1/2.$
  Quale sarebbe la distribuzione congiunta se le variabili stocastiche ξ ed η fossero indipendenti?
  Calcolare la mutua informazione e l'entropia di ξ, η e totale.
  Che c'entra con l'entropia di una variabile uniforme? Come mai la mutua informazione è massima?

Applicazioni in crittologia

Supponiamo che il messaggio M sia cifrato nel crittogramma C tramite la chiave K: $C \stackrel{def}{=} \mathcal{T}(M, K)$.

La decifrazione si ottiene tramite la trasformazione $M \stackrel{def}{=} \mathcal{D}(C, K)$.

Calcoliamo l'entropia di chiave e messaggio condizionata dalla conoscenza del crittogramma C. Ovvero valutiamo la diversità delle possibili coppie chiave-messaggio fissato il crittogramma:

$$H(K, M|C) = H(K|C) + H(M|K, C) = H(M|C) + H(K|C, M);$$

siccome $H(M|K, C) = 0$ (dato crittogramma e chiave il messaggio è noto e quindi la sua entropia è zero):

$$H(K|C) = H(M|C) + H(K|C, M) \geq H(M|C).$$

L'entropia delle chiavi fissato il crittogramma è maggiore (o uguale) all'entropia del messaggio fissato il crittogramma. Fissato il crittogramma la variabilità del messaggio è minore o uguale a quella della chiave.

Cifrari ideali e perfetti

Un buon cifrario deve produrre crittogrammi che non forniscono informazioni ne' sul messaggio originale, ne' (ancor meno) sulla chiave. Queste considerazioni conducono alle due definizioni di cifrario ideale e perfetto.
Un cifrario si dice ideale se l'informazione mutua tra crittogramma e chiave è nulla:

$$\mathcal{I}(K \wedge C) = 0 = H(K) - H(K|C);$$

ovvero l'entropia (la variabilità) della chiave non dipende dal crittogramma:

$$H(K) = H(K|C).$$

In generale, la conoscenza di un crittogramma aggiunge informazioni sulla possibile chiave, ma se il cifrario è ideale ciò non accade.

Cifrari ideali e perfetti

Un cifrario si dice perfetto se l'informazione mutua tra crittogramma e messaggio è nulla:

$$\mathcal{I}(M \wedge C) = 0 = H(M) - H(M|C);$$

ovvero l'entropia (la variabilità) del messaggio non cambia noto il crittogramma:

$$H(M) = H(M|C).$$

In generale la conoscenza di un crittogramma aggiunge informazioni sul possibile messaggio, ma se la cifratura è perfetta ciò non accade.

Cifrario perfetto

- Se un cifrario è perfetto l'entropia (la variabilità) del messaggio non cambia noto il crittogramma:

$$H(M) = H(M|C).$$

cioè la conoscenza di un crittogramma non aggiunge informazioni sul possibile messaggio.

Cifrario perfetto

- Se un cifrario è perfetto l'entropia (la variabilità) del messaggio non cambia noto il crittogramma:

$$H(M) = H(M|C).$$

  cioè la conoscenza di un crittogramma non aggiunge informazioni sul possibile messaggio.
- In termini probabilistici questo corrisponde all'indipendenza del messaggio dal crittogramma:

$$P(M|C) = P(M).$$

Cifrario perfetto - indipendenza del crittogramma

- La proprietà caratteristica più stringente è l'indipendenza della probabilità di ottenere un crittogramma dal messaggio che cifra:

$$P(M) \cdot P(C|M) = P(M, C) = P(C) \cdot P(M|C) = P(C) \cdot P(M);$$

$$P(M) \cdot P(C|M) = P(C) \cdot P(M);$$

Cifrario perfetto - indipendenza del crittogramma

- La proprietà caratteristica più stringente è l'indipendenza della probabilità di ottenere un crittogramma dal messaggio che cifra:

$$P(M) \cdot P(C|M) = P(M, C) = P(C) \cdot P(M|C) = P(C) \cdot P(M);$$

$$P(M) \cdot P(C|M) = P(C) \cdot P(M);$$

- Quindi $P(C|M) = P(C)$: la probabilità di un crittogramma non dipende dal messaggio che cifra.

Cifrario perfetto - indipendenza del crittogramma

- La proprietà caratteristica più stringente è l'indipendenza della probabilità di ottenere un crittogramma dal messaggio che cifra:

$$P(M) \cdot P(C|M) = P(M, C) = P(C) \cdot P(M|C) = P(C) \cdot P(M);$$

$$P(M) \cdot P(C|M) = P(C) \cdot P(M);$$

- Quindi $P(C|M) = P(C)$: la probabilità di un crittogramma non dipende dal messaggio che cifra.
- La presenza della chiave rende il crittogramma indipendente dal messaggio.

Cifrario perfetto - Numerosità delle chiavi

- Supponiamo $\mathcal{M} = \{M_1, M_2, \ldots, M_{|\mathcal{M}|}\}$sia lo spazio dei messaggi cifrabili e che Le possibili chiavi di cifratura possano appartenere ad uno spazio $\mathcal{K} = \{K_1, K_2, \ldots, K_{|\mathcal{K}|}\}$. I possibili messaggi cifrati (crittogrammi) apparterranno ad un spazio $\mathcal{C} = \{C_1, C_2, \ldots, C_{|\mathcal{C}|}\}$ univocamente determinato dai primi due e dall'algoritmo di cifratura.
  Se la cifratura è perfetta, le cardinalità di questi spazi sono ordinate:
  $$|\mathcal{M}| \leq |\mathcal{C}| \leq |\mathcal{K}|.$$

Cifrario perfetto - Numerosità delle chiavi

- Supponiamo $\mathcal{M} = \{M_1, M_2, \ldots, M_{|\mathcal{M}|}\}$sia lo spazio dei messaggi cifrabili e che Le possibili chiavi di cifratura possano appartenere ad uno spazio $\mathcal{K} = \{K_1, K_2, \ldots, K_{|\mathcal{K}|}\}$. I possibili messaggi cifrati (crittogrammi) apparterranno ad un spazio $\mathcal{C} = \{C_1, C_2, \ldots, C_{|\mathcal{C}|}\}$ univocamente determinato dai primi due e dall'algoritmo di cifratura.
  Se la cifratura è perfetta, le cardinalità di questi spazi sono ordinate:
  $$|\mathcal{M}| \leq |\mathcal{C}| \leq |\mathcal{K}|.$$
- La prima diseguaglianza $|\mathcal{M}| \leq |\mathcal{C}|$ equivale a dire che affinché un cifrario sia perfetto la variabilità dei crittogrammi deve essere più ampia di quella dei testi originali.

Cifrario perfetto - Numerosità delle chiavi

- Supponiamo $\mathcal{M} = \{M_1, M_2, \ldots, M_{|\mathcal{M}|}\}$sia lo spazio dei messaggi cifrabili e che Le possibili chiavi di cifratura possano appartenere ad uno spazio $\mathcal{K} = \{K_1, K_2, \ldots, K_{|\mathcal{K}|}\}$. I possibili messaggi cifrati (crittogrammi) apparterranno ad un spazio $\mathcal{C} = \{C_1, C_2, \ldots, C_{|\mathcal{C}|}\}$ univocamente determinato dai primi due e dall'algoritmo di cifratura.
  Se la cifratura è perfetta, le cardinalità di questi spazi sono ordinate:
  $$|\mathcal{M}| \leq |\mathcal{C}| \leq |\mathcal{K}|.$$
- La prima diseguaglianza $|\mathcal{M}| \leq |\mathcal{C}|$ equivale a dire che affinché un cifrario sia perfetto la variabilità dei crittogrammi deve essere più ampia di quella dei testi originali.
- La seconda diseguaglianza $|\mathcal{C}| \leq |\mathcal{K}|$ equivale a dire che affinché un cifrario sia perfetto la variabilità delle chiavi deve essere più ampia di quella dei crittogrammi.

In un cifrario perfetto le chiavi sono numerose almeno come i crittogrammi e questi almeno quanto i messaggi

- $|\mathcal{K}| \geq |\mathcal{C}|$ Deriva dal fatto che la probabilità che si presenti un crittogramma C è diversa da zero ($P(C|M) = P(C) > 0$) fissato un qualsiasi M; quindi, fissato M esiste almeno una chiave che lo cifra in forma di C.

$$\forall M, C : \exists K : C = \mathcal{T}_K(M).$$

Le chiavi sono numerose almeno come i crittogrammi.

In un cifrario perfetto le chiavi sono numerose almeno come i crittogrammi e questi almeno quanto i messaggi

- $|\mathcal{K}| \geq |\mathcal{C}|$ Deriva dal fatto che la probabilità che si presenti un crittogramma C è diversa da zero ($P(C|M) = P(C) > 0$) fissato un qualsiasi M; quindi, fissato M esiste almeno una chiave che lo cifra in forma di C.

$$\forall M, C : \exists K : C = \mathcal{T}_K(M).$$

Le chiavi sono numerose almeno come i crittogrammi.

- $|\mathcal{C}| \geq |\mathcal{M}|$ Deriva dal fatto che fissata la chiave tutti i messaggi diversi devono essere cifrati in crittogrammi diversi (iniettività); quindi per ogni messaggio c'è almeno un crittogramma.

$$\forall M_i \neq M_j, K : C_i = \mathcal{T}_K(M_i) \neq C_j = \mathcal{T}_{\mathcal{K}}(\mathcal{M}_\rangle).$$

I crittogrammi sono numerosi almeno come i messaggi.

Efficienza di cifratura

- L'efficienza consiste nell'utilizzare il numero minimo di crittogrammi per cifrare tutti i possibili messaggi.

Efficienza di cifratura

- L'efficienza consiste nell'utilizzare il numero minimo di crittogrammi per cifrare tutti i possibili messaggi.
- Una cifratura perfetta per raggiungere la massima efficienza deve produrre una numero di crittogrammi uguale al numero di messaggi: $|\mathcal{C}| = |\mathcal{M}|$.

Efficienza di cifratura

- L'efficienza consiste nell'utilizzare il numero minimo di crittogrammi per cifrare tutti i possibili messaggi.
- Una cifratura perfetta per raggiungere la massima efficienza deve produrre una numero di crittogrammi uguale al numero di messaggi: $|\mathcal{C}| = |\mathcal{M}|$.
- Analogamente affinché la dipendenza della cifratura dalla chiave sia ottimale, chiavi diverse devono produrre crittogrammi diversi. Quindi per ogni crittogramma c'è una chiave: $|\mathcal{C}| = |\mathcal{K}|$.

Efficienza di cifratura

- L'efficienza consiste nell'utilizzare il numero minimo di crittogrammi per cifrare tutti i possibili messaggi.
- Una cifratura perfetta per raggiungere la massima efficienza deve produrre una numero di crittogrammi uguale al numero di messaggi: $|\mathcal{C}| = |\mathcal{M}|$.
- Analogamente affinché la dipendenza della cifratura dalla chiave sia ottimale, chiavi diverse devono produrre crittogrammi diversi. Quindi per ogni crittogramma c'è una chiave: $|\mathcal{C}| = |\mathcal{K}|$.
- Se vogliamo limitare al massimo le risorse per ottenere una cifratura perfetta gli spazi dei messaggi, delle chiavi e dei crittogrammi devono avere tutti la stessa cardinalità $|\mathcal{M}| = |\mathcal{C}| = |\mathcal{K}|$. Siccome la cifratura è sempre iniettiva sia nella chiave che nei messaggi devono esistere relazioni invertibili tra tutte le possibili terne; cioè fissata una coppia di valori il terzo è determinato. $(\mathcal{M}, \mathcal{C}) \Rightarrow \mathcal{K}$, $(\mathcal{M}, \mathcal{K}) \Rightarrow \mathcal{C}$ e $(\mathcal{C}, \mathcal{K}) \Rightarrow \mathcal{M}$

Un esempio di cifrario perfetto: "One time Pad" OTP (Cifratura usa e getta)

- Il messaggio M è una sequenza di bit di lunghezza n. La chiave è un'altra sequenza altrettanto lunga scelta a caso. Il messaggio cifrato si ottiene tramite l'o esclusivo ($\oplus$, XOR o "o aut"):

$$\forall M, K : C = \mathcal{T}_K(M) \stackrel{def}{=} M \oplus K.$$

Un esempio di cifrario perfetto: "One time Pad" OTP (Cifratura usa e getta)

- Il messaggio M è una sequenza di bit di lunghezza n. La chiave è un'altra sequenza altrettanto lunga scelta a caso. Il messaggio cifrato si ottiene tramite l'o esclusivo ($\oplus$, XOR o "o aut"):

$$\forall M, K : C = \mathcal{T}_K(M) \stackrel{def}{=} M \oplus K.$$

- Esempio: $C = \{b_1, b_2, \ldots, b_n\} = \{1, 0, 1, 1, 1, 0, 0\}$ e chiave $K = \{k_1, k_2, \ldots, k_n\} = \{0, 0, 1, 0, 1, 0, 1\}$:

$$\begin{array}{rcrcl} & & M & = & \{1, 0, 1, 1, 1, 0, 0\} \\ & & K & = & \{0, 0, 1, 0, 1, 0, 1\} \\ C & = & M \oplus K & = & \{1, 0, 0, 1, 0, 0, 1\}. \end{array}$$

Un esempio di cifrario perfetto: "One time Pad" OTP o Cifrario di Wernam

- L'o esclusivo è la somma in $\mathbb{Z}_2$ quindi gode di alcune semplici proprietà:

Un esempio di cifrario perfetto: "One time Pad" OTP o Cifrario di Wernam

- L'o esclusivo è la somma in $\mathbb{Z}_2$ quindi gode di alcune semplici proprietà:
  - L'o esclusivo è associativo: $A \oplus (B \oplus C) = (A \oplus B) \oplus C$.

Un esempio di cifrario perfetto: "One time Pad" OTP o Cifrario di Wernam

- L'o esclusivo è la somma in $\mathbb{Z}_2$ quindi gode di alcune semplici proprietà:
  - L'o esclusivo è associativo: $A \oplus (B \oplus C) = (A \oplus B) \oplus C$.
  - La sequenza identicamente nulla è l'elemento neutro: $(\forall S : S \oplus 0 \equiv S)$; $(1 \oplus 0 = 1,\ 0 \oplus 0 = 0)$.

Un esempio di cifrario perfetto: "One time Pad" OTP o Cifrario di Wernam

- L'o esclusivo è la somma in $\mathbb{Z}_2$ quindi gode di alcune semplici proprietà:
  - L'o esclusivo è associativo: $A \oplus (B \oplus C) = (A \oplus B) \oplus C$.
  - La sequenza identicamente nulla è l'elemento neutro: $(\forall S : S \oplus 0 \equiv S)$; $(1 \oplus 0 = 1,\ 0 \oplus 0 = 0)$.
  - L'o esclusivo di qualunque sequenza con se stessa è sempre la sequenza identicamente nulla $(M \oplus M = 0)$: ogni sequenza è autoinversa $(1 \oplus 1 = 0,\ 0 \oplus 0 = 0)$.

Un esempio di cifrario perfetto: "One time Pad" OTP o Cifrario di Wernam

- L'o esclusivo è la somma in $\mathbb{Z}_2$ quindi gode di alcune semplici proprietà:
  - L'o esclusivo è associativo: $A \oplus (B \oplus C) = (A \oplus B) \oplus C$.
  - La sequenza identicamente nulla è l'elemento neutro: $(\forall S : S \oplus 0 \equiv S)$; $(1 \oplus 0 = 1,\ 0 \oplus 0 = 0)$.
  - L'o esclusivo di qualunque sequenza con se stessa è sempre la sequenza identicamente nulla $(M \oplus M = 0)$: ogni sequenza è autoinversa $(1 \oplus 1 = 0,\ 0 \oplus 0 = 0)$.
- Quindi dati C e M la chiave si ricava semplicemente calcolandone l'o esclusivo:

  $$\forall M, C : K = K \oplus (0) = K \oplus (M \oplus M) = (K \oplus M) \oplus (M) = C \oplus M.$$

  avendo usato la proprietà associativa.

Un esempio di cifrario perfetto: "One time Pad" OTP

- La cifratura OTP è perfetta. Per dimostrarlo basta mostrare che $P(M|C)$ non dipende da C.

Un esempio di cifrario perfetto: "One time Pad" OTP

- La cifratura OTP è perfetta. Per dimostrarlo basta mostrare che $P(M|C)$ non dipende da C.
- Essendo K una funzione deterministica di C e M, $P(K|M,C) = 1$ e $P(C|M,K) = 1$ (quando $C = \mathcal{T}_{\mathcal{K}}(\mathcal{M})$).

$$P(M,C) = P(M,C) \cdot P(K|M,C) = P(K,M,C);$$

$$P(K,M,C) = P(K) \cdot P(M|K) \cdot P(C|M;K) = \frac{1}{2^n} \cdot P(M) \cdot 1;$$

$$P(M,C) = \frac{1}{2^n} \cdot P(M);$$

$P(K) = \frac{1}{2^n}$ perché la seguenza della chiave è casuale (variabile aleatoria con distribuzione uniforme) e lunga n ed ogni bit ha probabilità $1/2$.

Un esempio di cifrario perfetto: "One time Pad" OTP

- La cifratura OTP è perfetta. Per dimostrarlo basta mostrare che $P(M|C)$ non dipende da C.
- Essendo K una funzione deterministica di C e M, $P(K|M,C) = 1$ e $P(C|M,K) = 1$ (quando $C = \mathcal{T}_{\mathcal{K}}(\mathcal{M})$).

$$P(M,C) = P(M,C) \cdot P(K|M,C) = P(K,M,C);$$

$$P(K,M,C) = P(K) \cdot P(M|K) \cdot P(C|M;K) = \frac{1}{2^n} \cdot P(M) \cdot 1;$$

$$P(M,C) = \frac{1}{2^n} \cdot P(M);$$

$P(K) = \frac{1}{2^n}$ perché la seguenza della chiave è casuale (variabile aleatoria con distribuzione uniforme) e lunga n ed ogni bit ha probabilità $1/2$.

- Quindi possiamo calcolare la probabilità di ogni crittogramma:

$$P(C) = \sum_M P(M,C) = \sum_M \frac{1}{2^n} \cdot P(M) = \frac{1}{2^n}.$$

Anche crittogrammi, come le chiavi, sono distribuiti uniformemente qualunque sia la distribuzione dei messaggi.

Un esempio di cifrario perfetto: "One time Pad" OTP

- Quindi

$$P(M, C) = \frac{1}{2^n} \cdot P(M) = P(C) \cdot P(M);$$

Un esempio di cifrario perfetto: "One time Pad" OTP

- Quindi
$$P(M, C) = \frac{1}{2^n} \cdot P(M) = P(C) \cdot P(M);$$
- Le variabili aleatorie M e C sono indipendenti. Quindi
$$P(M, C) = P(M|C) \cdot P(C) = P(C) \cdot P(M);$$
- La probabilità dei messaggi non dipende dalla crittazione:
$$P(M|C) = P(M);$$

cioè la cifratura è perfetta.

Decrittazione del cifrario "One time Pad" OTP

- Se un cifrario è perfetto, non significa che sia inviolabile. Se si dispone di una coppia messaggio-crittogramma, nel caso delle OTP, la chiave si ottiene banalmente:

$$\forall M, C : K = C \oplus M.$$

malgrado questo la decrittazione è inutile perché la chiave si usa una volta sola.

Decrittazione del cifrario "One time Pad" OTP

- Se un cifrario è perfetto, non significa che sia inviolabile. Se si dispone di una coppia messaggio-crittogramma, nel caso delle OTP, la chiave si ottiene banalmente:

$$\forall M, C : K = C \oplus M.$$

  malgrado questo la decrittazione è inutile perché la chiave si usa una volta sola.
- La cifratura OTP è anche ideale perché la generazione della chiave è indipendente dal messaggio:

$$P(K|C) = \frac{P(K) \cdot P(C)}{P(C)} = P(K).$$

Decrittazione del cifrario "One time Pad" OTP

- Se un cifrario è perfetto, non significa che sia inviolabile. Se si dispone di una coppia messaggio-crittogramma, nel caso delle OTP, la chiave si ottiene banalmente:

$$\forall M, C : K = C \oplus M.$$

  malgrado questo la decrittazione è inutile perché la chiave si usa una volta sola.
- La cifratura OTP è anche ideale perché la generazione della chiave è indipendente dal messaggio:

$$P(K|C) = \frac{P(K) \cdot P(C)}{P(C)} = P(K).$$

- Per questo motivo la chiave va utilizzata una sola volta.

Messaggio

- L'entropia è un concetto astratto cui si può pervenire da vie diverse: assiomatica, contenuto informativo, auto-informazione. L'auto-informazione è la mutua informazione tra due variabili identicamente distribuite.

Messaggio

- L'entropia è un concetto astratto cui si può pervenire da vie diverse: assiomatica, contenuto informativo, auto-informazione. L'auto-informazione è la mutua informazione tra due variabili identicamente distribuite.
- La divergenza informazionale consente di misurare quanto una distribuzione differisca da un'altra. Ad esempio, note le probabilità dei caratteri (ricavabili dalle loro frequenze relative) si può capire quale sia la lingua e decrittare una codifica sostituzionale.

Messaggio

- L'entropia è un concetto astratto cui si può pervenire da vie diverse: assiomatica, contenuto informativo, auto-informazione. L'auto-informazione è la mutua informazione tra due variabili identicamente distribuite.
- La divergenza informazionale consente di misurare quanto una distribuzione differisca da un'altra. Ad esempio, note le probabilità dei caratteri (ricavabili dalle loro frequenze relative) si può capire quale sia la lingua e decrittare una codifica sostituzionale.
- Analogamente alla probabilità, è possibile definire l'entropia di due variabili. L'entropia dell'insieme di due variabili è pari all'entropia di una di esse più l'entropia dell'altra condizionata dalla prima.

Messaggio

- L'entropia è un concetto astratto cui si può pervenire da vie diverse: assiomatica, contenuto informativo, auto-informazione. L'auto-informazione è la mutua informazione tra due variabili identicamente distribuite.
- La divergenza informazionale consente di misurare quanto una distribuzione differisca da un'altra. Ad esempio, note le probabilità dei caratteri (ricavabili dalle loro frequenze relative) si può capire quale sia la lingua e decrittare una codifica sostituzionale.
- Analogamente alla probabilità, è possibile definire l'entropia di due variabili. L'entropia dell'insieme di due variabili è pari all'entropia di una di esse più l'entropia dell'altra condizionata dalla prima.
- L'entropia dell'insieme di due variabili è sempre inferiore alla somma delle loro entropie.

Messaggio - cont.

- La divergenza informazionale consente di definire la mutua informazione che misura la variabilità condivisa tra due variabili, ovvero il deficit di entropia rispetto al sistema congiunto.

Messaggio - cont.

- La divergenza informazionale consente di definire la mutua informazione che misura la variabilità condivisa tra due variabili, ovvero il deficit di entropia rispetto al sistema congiunto.
- Abbiamo definito il concetto di cifrario ideale e perfetto e visto che lo OTP è un cifrario perfetto.

Messaggio - cont.

- La divergenza informazionale consente di definire la mutua informazione che misura la variabilità condivisa tra due variabili, ovvero il deficit di entropia rispetto al sistema congiunto.
- Abbiamo definito il concetto di cifrario ideale e perfetto e visto che lo OTP è un cifrario perfetto.
- Cifrario perfetto vuol dire inviolabile conoscendo solo crittogrammi, ma può essere facilmente violato conoscendo coppie messaggio-crittogramma.