SUNTO DI CIT E QIT

Information Theory Lecture Notes
Gianluca Peri
07-2022

INTRODUZIONE

In questa lezione ci proponiamo di esporre le principali differenze che intercorrono fra la teoria dell'informazione classica e la teoria dell'informazione quantistica. Da ora in avanti ci riferiremo a queste due teorie con gli acronimi CIT (*Classical Information Theory*) e QIT (*Quantum Information Theory*).
Per trattare questo argomento partiremo dalla CIT, e di essa esporremo quelli che riteniamo essere i suoi temi fondamentali; una volta fatto questo passeremo alla trattazione degli stessi temi nel caso quantistico, avendo cura di evidenziare le differenze con il caso classico a mano a mano che esse si paleseranno.

CLASSICAL INFORMATION THEORY

Gli attori principali in teoria dell'informazione sono entità note come *sorgenti* (Σ).

Sorgente
Una sorgente Σ è un'entità che produce sequenzialmente *simboli* x_α facenti parte di un set finito denominato *alfabeto* ($F - \alpha\beta$):

$$F - \alpha\beta = \{x_1, ..., x_F\}$$

Dove l'indice α prende per ovvie ragioni il nome di *indice di alfabeto*.
Si capisce quindi che ad ogni simbolo x_α è associabile una *probabilità di emissione* $p(x_\alpha)$. Assumeremo che tutte le sorgenti abbiano probabilità di emissione *costanti* per ogni simbolo.[1]

Compreso cosa si intenda per *sorgente* e cosa per *simbolo* possiamo adesso definire un altro fondamentale oggetto.

Stringa
Una *stringa* (detta anche "sequenza") è una successione ordinata di simboli. Nello specifico indicheremo con σ_n una stringa di lunghezza n:

$$\sigma_n = (\sigma_{\alpha[1]}, ..., \sigma_{\alpha[n]})$$

dove l'indice fra parentesi quadre per ovvie ragioni prende il nome di *indice di emissione*.

Possiamo dunque capire che, dato che le sorgenti emettono sequenzialmente simboli, possiamo anche dire che le sorgenti emettono stringhe.
Introdotti gli attori principali è adesso importante chiarire cosa ci proponiamo di studiare in teoria dell'informazione.
In teoria dell'informazione il *significato* di un simbolo o di una stringa *non ci interessa affatto!* Il significato è per noi *irrilevante!* Difatti la teoria dell'informazione non vuole essere una teoria dei significati, ma bensì una *teoria delle risorse*. Non una teoria del *comunicato* ma bensì una teoria del *comunicante*. Per intenderci: di un'immagine siamo interessati alla quantità di memoria che essa occupa su un disco rigido e non al luogo che è stato fotografato.

1 Alle volte per riferirsi a sorgenti con probabilità di emissione costanti si perferisce utilizzare il termine *sorgenti i.i.d.* (*indipendent, identically distributed*); dove questi aggettivi si riferiscono ovviamente a come i simboli sono emessi dalla sorgente.

Vogliamo adesso definire una nuova quantità associata ai simboli emessi da una sorgente: la quantità di informazione i che essi trasportano.

Definizione di informazione di un simbolo
Si definisce *informazione di* x_α (*informazione trasmessa da* x_α) come la quantità:

$$i(x_\alpha) \doteq -\log_F p(x_\alpha) \tag{1}$$

Questa è una definizione, non un teorema, e per questo formalmente non necessita di giustificazioni. Tuttavia è utile notare che questa definizione è preferibile ad altre in quanto è particolarmente in accordo con la nostra comprensione intuitiva del concetto di "informazione trasmessa da qualcosa", in quanto:

- L'informazione trasmessa dal simbolo è nulla se la sua probabilità di emissione è 1, e di contro $i(x_\alpha) \to +\infty$ quando $p(x_\alpha) \to 0$. Questo è in accordo con la nostra intuizione, in quanto osservare eventi certi a priori non fornisce alcuna nuova informazione, mentre osservare un evento improbabile fornisce una notevole quantità di informazione. Di mattina osservare il sole sorgere non ci fornisce alcuna informazione, di contro riceveremmo molta informazione se una mattina al posto del sole osservassimo sorgere in cielo una gigantesca mela.
- L'informazione trasmessa da un simbolo è una quantità definita positiva, difatti non avrebbe senso che la ricezione di un simbolo togliesse informazione all'osservatore.
- L'informazione definita con (1) è additiva.[2] L'informazione di una sequenza di simboli è pari alla somma dell'informazione dei singoli simboli, come intuitivamente ci viene naturale richiedere per eventi indipendenti.

  A questo proposito, dato che come già detto possiamo interpretare le stringhe stesse come oggetti emessi dalla sorgente con una certa probabilità $P(\sigma_n)$, e che dunque c'è completa analogia fra stringhe e simboli, possiamo facilmente calcolare l'informazione associata ad una stringa σ_n basandoci sulla definizione di informazione associata ad un simbolo (1):

$$i(\sigma_n) = -\log_F(P(\sigma_n)) = -\log_F\left(\prod_{i=1}^{n} P(x_{\alpha[i]})\right) = \sum_{i=1}^{n} i(x_{\alpha[i]}) \tag{2}$$

- La ragione per cui si è scelto di porre il logaritmo in base F sarà chiara fra poco.

Adesso dobbiamo definire un'altra fondamentale quantità.

Entropia di Shannon
Si consideri una sorgente Σ con alfabeto $F - \alpha\beta = \{x_1, ..., x_F\}$. Sia $p(x_\alpha)$ la probabilità di emissione associata al simbolo dell'alfabeto x_α (per Σ). Allora si *definisce* entropia della sorgente Σ, anche detta *entropia di Shannon*:

$$H(\Sigma) \doteq -\sum_{\alpha=1}^{F} p(x_\alpha) \log_F p(x_\alpha) \tag{3}$$

Questa è la forma canonica di scrittura dell'entropia di Shannon, ma possiamo notare che applicando la definizione di informazione (1) essa può essere scritta come:

$$H(\Sigma) = \sum_{\alpha=1}^{F} p(x_\alpha) i(x_\alpha)$$

Vediamo quindi che l'entropia di Shannon è il valore di aspettazione dell'informazione trasportata da un *singolo simbolo* emesso dalla sorgente.
Per questo motivo alle volte si parla dell'entropia di Shannon come dell'informazione media trasmessa dalla sorgente, o dell'informazione media della sorgente.

2 Additiva per sorgenti come le nostre, aventi probabilità costanti e indipendenti, si dice anche per "sorgenti smemorate".

Notare la similitudine fra l'espressione (3) e la celeberrima entropia alla Boltzmann S:

$$S = -k_B \sum_i p_i \log p_i$$

Questo ci fa capire perché abbiamo scelto di chiamare il valore medio dell'informazione trasportata da un simbolo emesso da Σ *entropia di* Σ.
Possiamo finalmente apprezzare perché si è scelto di definire l'informazione con il logaritmo in base F: fissata la dimensione dell'alfabeto cerchiamo il massimo di $H(\Sigma)$: notiamo che $H(\Sigma)$ è massimo se e solo se il valore dei suoi addendi è massimo dato che sono fra loro indipendenti, dunque vogliamo trovare il valore di p tale che $-p \log_F p$ sia massimo. Ma quindi per trovare il valore di p corrispondente al massimo basta procedere come segue:

$$\frac{d}{dp}\big(-p \log_F p\big) = 0 \;\Rightarrow -\log_F p - p\frac{1}{p} = -\log_F p - 1 = 0 \;\Rightarrow\; \log_F p = -1 \;\Rightarrow\; p = \frac{1}{F}$$

ma dunque il massimo di $H(\Sigma)$ è:

$$\max\big[H(\Sigma)\big] = -\sum_{\alpha=1}^{F} \frac{1}{F} \log_F \frac{1}{F} = \sum_{\alpha=1}^{F} \frac{1}{F} = 1$$

Definire l'informazione con un logaritmo che ha come base la dimensione dell'alfabeto ci permette di *normalizzare l'entropia* della sorgente a 1:

$$0 \leqslant H(\Sigma) \leqslant 1 \tag{4}$$

Con quest'ultima definizione abbiamo finito di delineare il *framework* della CIT; tuttavia non siamo ancora giunti al cuore della teoria, incarnato dai due celebri *teoremi di Shannon*. Il primo tratta i processi di codifica e compressione dell'informazione, ponendo un limite teorico al rate di compressione massimo ottenibile in un dato contesto. Il secondo tratta invece la comunicazione di informazione attraverso canali rumorosi, e stabilisce limiti teorici al tasso di informazione comunicabile attraverso un canale dipendentemente da quanto rumoroso il canale sia. Vogliamo concludere la nostra trattazione della CIT con l'esposizione di questi due teoremi.
Per esporre il primo dobbiamo innanzitutto presentare una carrellata di utili definizioni.

Definizione di schema di codifica/decodifica *(schema C/D)*

- Definiamo **schema di codifica** ($C_{n\to m}$) un processo che trasforma sequenze σ_n in sequenze τ_m.
- Definiamo **rapporto di compressione** di uno schema di codifica $C_{n\to m}$ la quantità

  $$R \doteq \frac{m}{n} < 1$$

- Definiamo **schema di decodifica** ($D_{m\to n}$) un processo che *con una certa probabilità* inverte un certo schema di codifica, e permette di ottenere σ_n a partire da τ_m

  Ed infine definiamo **schema di codifica/decodifica** (C/D) uno schema composto da uno schema di codifica e uno di decodifica. Chiameremo rapporto di compressione dello schema C/D il rapporto di compressione del suo schema di codifica.

Notare che il rapporto di compressione R è tanto più grande quanto più la stringa iniziale viene *espansa* dallo schema di codifica. Ergo un buon algoritmo di compressione corrisponde a bassi valori di R, questa definizione è comoda in quanto ci consente di dire che uno schema di codifica avente rate R mappa stringhe di dimensione n in stringhe di dimensione Rn.
Notare inoltre che per ovvie ragioni in questo contesto lo schema codifica/decodifica viene anche chiamato *schema di compressione/decompressione*.
Notare infine che in base alla nostra definizione di schema di decodifica non è affatto detto che un processo C/D ci permetta di recuperare la nostra stringa iniziale; questo ovviamente non è ideale, vorremmo delle garanzie sul funzionamento del nostro schema, e questo motiva la seguente definizione.

Definizione di reliableness
Uno schema di codifica/decodifica (C/D) è detto *reliable* se rispetta il seguente criterio:

$$\lim_{n\to+\infty} \max_{\sigma_n} \left\{ p\left[D_{m\to n}(C_{n\to m}(\sigma_n)) \neq \sigma_n \right] \right\} = 0$$

Uno schema C/D reliable è dunque uno schema *asintoticamente perfetto*, nel senso che mano a mano che la dimensione di quanto codificato (compresso) aumenta la probabilità di commettere errori tende a zero.
Siamo adesso finalmente pronti per enunciare il primo teorema di Shannon.

Primo teorema di Shannon *(Teorema di codifica senza rumore)*
Per una certa sorgente Σ esiste uno schema C/D *reliable* con rapporto di compressione R *se e solo se*:

$$R > H(\Sigma)$$

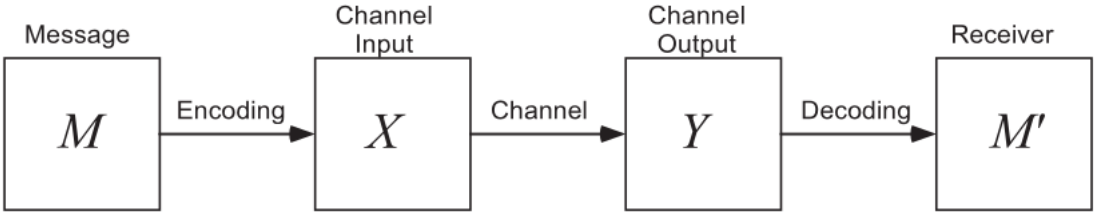
Dunque, come dicevamo, il primo teorema di Shannon pone un limite teorico alla compressione che è sensato applicare all'informazione proveniente da una sorgente. Tanto più l'informazione media dei simboli emessi dalla sorgente è bassa tanto più è possibile comprimere le sue stringhe in maniera reliable. Questo dovrebbe apparire come intuitivamente lampante: data la nostra definizione di informazione una bassa informazione media dei simboli significa una sorgente *altamente prevedibile*, che ammette quindi efficaci schemi di compressione.
Passiamo adesso al secondo teorema di Shannon. Come dicevamo questo ha a che vedere con la trasmissione di informazione attraverso un canale rumoroso (a *rumore scorrelato*).
Consideriamo quindi una sorgente M e un *destinatario* M'; la sorgente emette simboli, i simboli emessi attraversano un canale di comunicazione, ed infine arrivano a M'. Come dicevamo il canale di comunicazione è rumoroso (se non lo fosse potremmo semplicemente ignorare la sua presenza) e dunque l'informazione trasmessa con esso verrà almeno in parte corrotta. Per proteggerci da questi effetti di disturbo occorrerà implementare uno schema C/D che consenta di comunicare bene nonostante il rumore. Si pensi di voler comunicare qualcosa a un nostro amico dall'altra parte di una sala affollata e rumorosa, cosa facciamo per assicurarci che quanto stiamo dicendo venga correttamente recepito dal destinatario? Ci ripetiamo! Comunichiamo la stessa informazione più volte al fine di rendere il messaggio resiliente al rumore. Capiamo quindi che questa volta lo schema C/D dovrà *espandere* le stringhe di Σ anziché tentare di comprimerle il più possibile come nel precedente caso.
In questo contesto definiamo la seguente quantità.

Definizione Rate di Trasmissione
Dato uno schema C/D con $C_{n\to\tau}$ e $D_{\tau\to n}$ definiamo *rate di trasmissione* la quantità:

$$T \doteq \frac{n}{\tau} < 1$$

Dovrà essere minore di uno in quanto, come dicevamo, vogliamo ripetere più volte la medesima informazione per rendere la comunicazione resiliente al rumore.

Come dicevamo stiamo analizzando una situazione dove una sorgente M emette simboli che vengono poi codificati, fatti passare attraverso un canale rumoroso, decodificati e infine ricevuti dal destinatario M'. Viene fuori che per analizzare simili situazioni risulta comodo pensare ad ogni step del processo come ad una differente sorgente.



In questa rappresentazione il canale è modellizzato come una coppia di sorgenti X, Y, rappresentanti rispettivamente l'ingresso e l'uscita del canale. Questo modello ha senso in quanto pensiamo alle quattro sorgenti come *correlate fra loro*. M è correlata a X attraverso il processo di encoding, X è correlata a Y attraverso il canale di comunicazione rumoroso, e infine Y è correlata a M' attraverso il decoding. Diremo che M produce stringhe σ_n che vengono convertite nelle stringhe x_τ che produce X, mentre dopo il canale si hanno le stringhe y_τ di Y che poi vengono decodificate nelle stringhe σ'_n di M'. Notare che questo equivale ad assumere che nel canale di comunicazione i simboli possano unicamente venir *modificati*, e non cancellati.
Per proseguire la nostra analisi abbiamo nuovamente bisogno di una rapida carrellata di definizioni.

Definizione Entropia Congiunta
Si considerino due sorgenti X, Y con alfabeto contenente i simboli x_α, y_α rispettivamente. Allora l'entropia giunta (*joint entropy*) delle due sorgenti è definita come la quantità:

$$H(X,Y) \doteq -\sum_{i,k} p(x_i, y_k) \log p(x_i, y_k)$$

Ovvero: $H(X,Y)$ è l'informazione media trasmessa dalla coppia di sorgenti X, Y. O possiamo semplicemente dire che $H(X,Y)$ è l'informazione media della coppia di sorgenti X, Y.

Definizione Entropia Mutua
Si considerino due sorgenti X, Y. Allora l'entropia mutua è definita come:

$$H(Y:X) \doteq H(Y) + H(X) - H(X,Y)$$

Ovvero: $H(X:Y)$ è l'ammontare di informazione media condivisa fra quanto emesso dalle sorgenti X e Y. O più semplicemente $H(X:Y)$ è l'informazione comune fra X e Y.

Rate di trasmissione achievable
Si consideri una sorgente mittente M (che emette stringhe σ_n) e un destinatario M' che comunicano attraverso un canale rumoroso modellizzato tramite le sorgenti X, Y con uno schema C/D ($n \to \tau \to n$) avente rate di trasmissione T.
Dato questo contesto diremo che il rate di trasmissione $T = n/\tau$ è *achievable* se esiste un protocollo C/D, con quel rate di trasmissione, tale che:

$$\lim_{n \to +\infty} \max_{\sigma_n} \left\{ p\left[D(C(\sigma_n)) \neq \sigma_n \right] \right\} = 0$$

Notare che $p\left[D(C(\sigma_n)) \neq \sigma_n \right]$ è chiamata *probabilità di errore*, ed è la probabilità che alla fine del processo di comunicazione il destinatario M' *non* riceva la stringa σ_n emessa dal mittente M. Notare infine che questa probabilità deve dipendere da qaunto rumore è presente nel canale.

Notare che nella scelta di uno schema C/D grandi valori del rate di trasmissione T sono da preferire, in quanto un grande rate di trasmissione implica il poter comunicare attraverso il canale utilizzando stringhe di lunghezza non troppo maggiore della lunghezza della stringa emessa dal mittente M, il che ci fa risparmiare tempo ed energie rispetto ad un protocollo C/D con rate di trasmissione più basso.

Capacità di un canale rumoroso
Si consideri il medesimo setup della precedente definizione. In questo contesto si definisce *capacità* del canale rumoroso (C) il *massimo* rate di trasmissione T tale da essere achievable in quel canale.

Possiamo finalmente enunciare il secondo teorema di Shannon.

Secondo teorema di Shannon
Dato un canale di comunicazione rumoroso, modellizzato con le sorgenti X, Y, la sua capacità C è pari all'entropia mutua delle sorgenti X, Y.

$$C = H(X : Y)$$

Come il primo anche il secondo teorema di Shannon dovrebbe apparire intuitivamente evidente: l'entropia mutua è una misura di quanto l'informazione delle sorgenti sia correlata, un'alta entropia mutua implica quindi poco rumore e la possibilità di comunicazione efficiente.
Abbiamo concluso la nostra esposizione degli aspetti principali della CIT, passiamo adesso alla QIT. Come dicevamo saremo interessati a mettere in luce le differenze che intercorrono fra le due nelle tematiche fondamentali che abbiamo appena finito di delineare.

QUANTUM INFORMATION THEORY

Vogliamo ripercorrere passo per passo quanto trattato nella precedente sezione. La prima tematica da affrontare dunque è cosa sia una sorgente di informazione quantistica.

Sorgente di informazione quantistica
Una sorgente di informazione quantistica è un'entità che produce sequenzialmente *stati quantistici* ρ_i facenti parte di un set finito

$$\{\rho_1, ..., \rho_n\}$$

Si capisce quindi che ad ogni stato ρ_i è associata una *probabilità di emissione* p_i. Assumeremo che tutte le sorgenti quantistiche abbiano probabilità di emissione *costanti* per ogni stato ρ_i.

Possiamo dunque constatare che la definizione di sorgente quantistica è del tutto analoga alla definizione di sorgente classica, l'unica differenza risiede nel fatto che una sorgente quantistica non emette simboli ma bensì *stati quantistici*.
Risulta adesso *fondamentale* realizzare che una sorgente di informazione quantistica può essere *essa stessa* pensata come uno stato quantistico ρ, che opera su un certo spazio di Hilbert H. Difatti data una generica sorgente quantistica avente alfabeto composto dagli stati $\{\rho_i\}$ con probabilità di emissione $\{p_i\}$ si consideri lo stato quantistico:

$$\rho = \sum_i p_i \rho_i$$

ebbene questo stato ρ rappresenta in tutto e per tutto la sorgente quantistica in esame, in quanto esso è uno degli stati ρ_i con probabilità p_i.
Il primo punto cardine da comprendere è dunque che *le sorgenti quantistiche, che per definizione emettono stati quantistici, sono esse stesse interpretabili come stati quantistici.*
Notare in coda la seguente fondamentale osservazione: l'agente che osserva la sorgente ρ emettere uno stato quantistico a priori non ha idea di quale stato fra i possibili ρ_i la sorgente abbia emesso. Pertanto per l'osservatore a priori ogni stato emesso dalla sorgente è nello stato misto $\rho = \sum_i p_i \rho_i$, che è proprio difatti lo stato che rappresenta la sorgente! Possiamo dunque dire anche che per l'osservatore ogni stato emesso dalla sorgente è a priori identico agli altri (ed è proprio lo stato ρ).
Seguendo il nostro percorso vogliamo adesso definire l'entropia associata ad una sorgente quantistica (ovvero l'entropia associata ad uno stato quantistico). Nella trattazione della CIT abbiamo definito l'entropia definendo prima la quantità d'informazione i trasportata da un simbolo della sorgente e facendo poi vedere che l'entropia di Shannon corrisponde al valore medio dell'informazione emessa dalla sorgente. In un qualche senso quindi capiamo che la definizione dell'informazione i è stata unicamente strumentale alla definizione dell'entropia di Shannon, difatti possiamo notare che nella trattazione delle tematiche più importanti della CIT (si pensi ai due teoremi di Shannon) viene costantemente utilizzata l'entropia di Shannon ma mai l'informazione i.

Ora che siamo in QIT, forti della nostra conoscenza della CIT, non sentiamo quindi la necessità di reintrodurre il concetto di informazione posseduta dagli oggetti emessi da una sorgente. Vogliamo bensì definire l'entropia quantistica direttamente come un'estensione dell'entropia classica di Shannon.
Definiremo quindi l'entropia in QIT, nota come *entropia di Von Neumann*, e poi mostreremo che essa non è altro che una generalizzazione dell'entropia di Shannon al caso quantistico.

Entropia di Von Neumann
Data una sorgente quantistica (stato quantistico) ρ definiamo entropia di Von Neumann $S(\rho)$ la seguente:

$$S(\rho) \doteq -\operatorname{Tr}\left[\rho \log \rho\right]$$

Questa definizione di entropia a prima vista sembra avere poco a che fare con l'entropia di Shannon che siamo abituati ad associare alle sorgenti di informazione. Tuttavia è in realtà facile mostrare che l'entropia di Von Neumann è una generalizzazione dell'entropia di Shannon. Prendiamo in esame una sorgente di informazione quantistica che emetta stati puri mutuamente ortogonali $\{\rho_i = |e_i\rangle \langle e_i|\}$

$$\rho = \sum_{i=1}^{F} p_i |e_i\rangle \langle e_i|$$

Notare che metterci in questa condizione equivale, in pratica, a porci nel caso classico. Difatti classicamente ogni oggetto che può essere emesso da una sorgente è in uno stato puro e distinguibile, dunque mutuamente ortogonale agli altri possibili stati.
Notare inoltre che, nel caso gli stati $\{|e_i\rangle\}$ già non formino una base, possiamo facilmente estenderli ad una base ortonormale dello spazio di Hilbert su cui ρ agisce, difatti gli stati $\{|e_i\rangle\}$ dovendo essere stati fisici sicuramente sono normalizzati, e ai vettori che devono essere aggiunti per formare la base possiamo semplicemente associare probabilità zero, in maniera tale da non modificare ρ con l'estensione.
In questo contesto, ricordando che la traccia di un operatore è la stessa in ogni base, possiamo espandere $S(\rho)$ come segue.

$$S(\rho) = -\operatorname{Tr}\left[\rho \log_F \rho\right] = -\sum_{i=1}^{F} \langle e_i| \left(\sum_{j=1}^{F} p_j |e_j\rangle \langle e_j| \right) \log_F \left(\sum_{k=1}^{F} p_k |e_k\rangle \langle e_k| \right) |e_i\rangle$$

Il termine problematico in questa espressione è:

$$\log_F \left(\sum_{k=1}^{F} p_k |e_k\rangle \langle e_k| \right)$$

difatti in generale per calcolare il logaritmo di una matrice bisogna procedere con un espansione in serie. *Tuttavia* si può dimostrare che per una generica matrice *diagonale* A il logaritmo della stessa non è altro che una matrice diagonale avente come elementi diagonali i logaritmi degli elementi diagonali di A. Fortunatamente

$$\sum_{k=1}^{F} p_k |e_k\rangle \langle e_k|$$

è proprio una matrice diagonale, il che ci autorizza a scrivere:

$$-\sum_{i=1}^{F} \langle e_i| \left(\sum_{j=1}^{F} p_j |e_j\rangle \langle e_j| \right) \log_F \left(\sum_{k=1}^{F} p_k |e_k\rangle \langle e_k| \right) |e_i\rangle =$$

$$= -\sum_{i=1}^{F} \langle e_i| \left(\sum_{j=1}^{F} p_j |e_j\rangle \langle e_j| \right) \left(\sum_{k=1}^{F} \log_F(p_k) |e_k\rangle \langle e_k| \right) |e_k\rangle =$$

$$= -\sum_{i=1}^{F} \langle e_i| \left(\sum_{j,k} p_j \log(p_k) |e_i\rangle \langle e_i|e_j\rangle \langle e_j| \right) |e_i\rangle =$$

per l'ortonormalità:

$$= -\sum_{i=1}^{F} \langle e_i| \left(\sum_{j=1}^{F} p_j \log_F(p_j) |e_j\rangle \langle e_j| \right) |e_i\rangle = -\sum_{i=1}^{F} p_i \log_F p_i$$

Troviamo quindi la medesima struttura dell'entropia di Shannon.
In questa sede è bene notare che gli *stati puri* sono equivalenti a sorgenti che emettono un unico simbolo, e dunque hanno entropia pari a zero. L'osservazione di uno stato puro, noto lo stato puro, non ci fornisce alcuna informazione, esattamente come l'osservazione di una sorgente a singolo simbolo, nota la natura della sorgente stessa, non ci fornisce alcuna informazione. Solo gli *stati misti* hanno entropia di Von Neumann diversa da zero.
Volendo continuare a seguire la nostra *roadmap* vorremmo adesso trattare l'analogo dei due teoremi di Shannon in QIT. Tuttavia si rivela saggio adesso fare una deviazione: fino ad ora abbiamo trovato unicamente forti analogie fra CIT e QIT, la necessità di questa deviazione segna l'inizio delle differenze.
Ritorniamo per un attimo nel contesto della CIT e supponiamo di avere una sorgente mittente M ed una ricevente M' che comunicano attraverso un canale *senza rumore*. In questo contesto ogni simbolo emesso da M arriva inalterato a M', non abbiamo necessita di nessuno schema C/D, e interpretando il canale con la solita coppia di sorgenti X, Y capiamo bene di poter dire che:

$$M \equiv X \equiv Y \equiv M'$$

si capisce quindi che la mutua informazione (entropia mutua) $H(X : Y)$ fra le sorgenti banalmente deve essere tutta l'informazione scambiata $H(X)$, e infatti:

$$H(X : Y) = H(X) + H(Y) - H(X, Y) = H(X) + H(X) - H(X, X) = 2H(X) - H(X) = H(X) \equiv H(Y)$$

è dunque un caso così banale da non necessitare trattazione, che è esattamente il motivo per cui inizialmente non l'abbiamo trattato. *Tuttavia* le cose non sono affatto così semplici in QIT.
Difatti in fisica classica ogni stato è *distinguibile*, mentre in meccanica quantistica solo gli stati ortogonali sono distinguibili! Questo implica che non sia detto che il ricevente sia in grado di estrarre l'informazione inviata dal mittente!
Per comprendere questa cruciale differenza fra CIT e QIT trasliamo il setup appena esposto in un contesto consono alla QIT. In QIT si è soliti antropomorfizzare mittente e ricevente con i due attori Alice e Bob, e noi non ci discosteremo da questa tradizione. Alice possiede una sorgente di informazione classica X con alfabeto $\{x_1, ..., x_n\}$ e rispettive probabilità di emissione $\{p_1, ..., p_n\}$. Alice vuole comunicare con Bob utilizzando un canale quantistico, e per questo motivo codifica la sua sorgente classica in una sorgente quantistica

$$\rho = \sum_{i=1}^{n} p_i \rho_i$$

dove nella codifica ovviamente ρ_1 corrisponde a x_1, ρ_2 a x_2, e così via. La sorgente quantistica ρ emette gli stati quantistici $\{\rho_i\}$, che percorrono un canale di comunicazione privo di rumore, esattamente come nel caso classico, e arrivano a Bob. A questo punto Bob eseguirà delle misure sugli stati quantistici che riceve, possiamo pensare ai risultati di queste misure come generati da una sorgente classica Y.
In QIT il contesto appena esposto è uno dei contesti canonici: abbiamo una certa sorgente classica X la cui informazione viene traslata per intero in uno stato quantistico, che verrà poi comunicato attraverso un canale quantistico. Il destinatario in fondo al canale vorrà poi fare delle misure sugli stati quantistici che riceve per estrarre l'informazione contenuta in essi, traducendola nuovamente in informazione classica. Il processo sotto esame inizia e finisce con sorgenti di informazione classica X, Y con un certo grado di correlazione mutua, i processi quantistici fanno semplicemente da ponte fra queste due sorgenti.
In questo contesto l'entropia di Shannon della sorgente X è nota per ovvie ragioni come *entropia di preparazione*. Nel caso classica avevamo:

$$H(X:Y) = H(X)$$

comunicazione perfetta. Nel caso quantistico come abbiamo già accennato non è così (a causa dell'indistinguibilità degli stati non ortogonali). Nello specifico definiamo **informazione accessibile** il massimo possibile valore di $H(X:Y)$

$$\max H(X:Y) = I_{acc}$$

Questo valore $\max H(X:Y)$ si rivela molto difficile da calcolare esattamente, tuttavia possiamo dire alcune cose su di esso. In primo luogo è evidente che:

$$I_{acc} \leqslant H(X)$$

con l'uguaglianza che vale nel limite classico. Un altro vincolo è posto dal seguente fondamentale teorema.

Holevo Bound
Si considerino due agenti Alice e Bob che comunicano attraverso un canale quantistico privo di rumore nel contesto esposto poco sopra. Allora valida la seguente:

$$H(X:Y) \leqslant S(\rho) - \sum_i p_i S(\rho_i)$$

La quantità $S(\rho) - \sum_i p_i S(\rho_i)$ è nota come **quantità di Holevo**, e solitamente si indica con la lettera χ ($\chi(\rho)$).

Notare che la quantità di Holevo è sempre minore dell'entropia di preparazione $H(X)$

$$H(X:Y) \leqslant \chi(\rho) \leqslant H(X) \tag{5}$$

Dove le uguaglianze sono valide su supporto ortogonale, *id est* nel limite classico.
Notare che la (5) ovviamente implica che:

$$I_{acc} \leqslant \chi(\rho) \leqslant H(X)$$

Questo ci permette di capire che una modalità di comunicazione su canale quantistico strutturata come descritto sopra non potrà mai essere più efficiente di una comunicazione su canale classico privo di rumore. In questo contesto con n *qbits* si possono comunicare al massimo n *bits* di informazione classica.
Dopo questa trattazione siamo finalmente pronti a generalizzare i due teoremi di Shannon al caso quantistico. Seguendo la nostra *roadmap* partiamo dal primo teorema di Shannon, che in QIT prende il nome di *teorema di Schumacher-Josza*. Nel caso classico si mostrava che il miglior rate di compressione R *reliable* per una sorgente X è pari all'entropia della sorgente ($H(X)$). Notare che R è sempre minore di 1. Questo significa che il miglior schema di compressione *reliable* mappa stringhe di lunghezza n in stringhe di lunghezza $Rn = H(X)n$

$$n \to nH(X)$$

Ebbene anticipiamo che una volta traslate le nostre definizioni al caso quantistico troveremo un risultato esattamente analogo!
Nel contesto quantistico abbiamo ovviamente una sorgente di informazione quantistica, ovvero una ρ che opera in un certo spazio H. In questo contesto uno schema C/D con rate di compressione R corrisponde a due operazioni quantistiche C^n e D^n, rispettivamente operazione di compressione e operazione di decompressione, dove l'operazione di compressione porta stati in $H^{\otimes n}$ in stati nello spazio compresso $H^{\otimes nR}$, *ovvero* l'operazione di compressione porta stringhe di lunghezza n di stati emessi dalla sorgente quantistica in stringhe di stati di lunghezza nR. Per concretizzare questo concetto possiamo pensare che l'operazione di compressione permetta di codificare l'informazione contenuta in una stringa di n qbits in una stringa di nR qbits. Adesso per completare il quadro dobbiamo estendere la definizione di *reliability* al caso quantistico, per fare questo ci troviamo costretti ad esporre un'altra rapida carrellata di definizioni.

Fidelity
La *fedeltà* (*fidelity*) di una coppia di stati quantistici ρ, σ è definita come la quantità:

$$F(\rho, \sigma) \doteq Tr\sqrt{\rho^{1/2}\sigma\rho^{1/2}}$$

Notare che $F(\cdot, \cdot)$ è simmetrica nei suoi input. Questa è intesa come una misura di *distanza* fra gli stati ρ, σ. Notare inoltre che nel caso in cui uno dei due stati sia uno *stato puro* si può dimostrare che la fidelity assume una forma molto più semplice:

$$F(|\psi\rangle, \rho) = \sqrt{\langle\psi|\rho|\psi\rangle}$$

alle volte si preferisce addirittura definire la fidelity a partire da quest'ultimo caso particolare, dato che è molto utilizzato.
Si dimostra che

$$0 \leqslant F(\rho, \sigma) \leqslant 1$$

e che $F(\rho, \sigma) = 1$ se $\rho = \sigma$ e infine che $F(\rho, \sigma) = 0$ se ρ e σ sono *ortogonali*, o meglio *hanno supporto su spazi mutuamente ortogonali* (cosi teniamo conto della possibilità di stati misti), e quindi possiamo dire che $F(\rho, \sigma) = 1$ se e solo se ρ e σ sono *perfettamente distinguibili!*

Entanglement fidelity
Si consideri un sistema quantistico Q preparato in uno stato ρ assunto *entangled* con un sistema R esterno, in maniera tale che lo stato complessivo QR sia uno *stato puro*. Lo stato ρ è adesso soggetto ad un operazione $\mathcal{E}$ che lo porta nello stato Q′ entanglato con il sistema R′, quest'ultimo è sempre l'ambiente esterno, è di fatto una copia del sistema R, che non è stato toccato dall'operazione $\mathcal{E}$.
In questo contesto definiamo *entanglement fidelity* $F_E(\rho, \mathcal{E})$ il quadrato della misura di distanza fra i due stati puri QR e Q′R′, ovvero il quadrato della *fidelity* fra essi:

$$F_E(\rho, \mathcal{E}) = F\big(RQ, R'Q'\big)^2 = \langle RQ|\left[\big(\mathbb{I}_R \otimes \mathcal{E}\big)\big(|RQ\rangle\langle RQ|\big)\right]|RQ\rangle$$

Si capisce che, dato che R rimane immutato, questa è una misura di quanto lo stato di Q sia cambiato per effetto dell'operazione E applicata su di esso. Questa definizione è guidata anche dall'intuizione che un *canale quantistico* che preserva bene l'informazione è un canale che preserva bene l'entanglement!
Si capisce che l'informazione di ρ è conservata perfettamente da $\mathcal{E}$ se $F_E(\rho, \mathcal{E}) = 1$, mentre l'informazione viene completamente persa se l'entanglement fidelity è uguale a zero.

Il punto fondamentale che si deve trarre dalle definizioni appena esposte è che *l'entanglement fidelity* è una buona misura di quanto una certa operazione quantistica $\mathcal{E}$ modifichi uno stato quantistico ρ. Grazie a queste conoscenze possiamo finalmente generalizzare il concetto di *reliability* di un processo C/D al caso quantistico.

Definizione di reliability in QIT
Uno schema di compressione/decompressione quantistico, agente sulla stringa di stati quantistici rappresentata dallo stato quantistico $\rho^{\otimes n}$, messo in atto dalle operazioni quantistiche C^n, D^n si definisce *reliable* se

$$\lim_{n\to+\infty} F_E\big(\rho^{\otimes n}, D^n \circ C^n\big) \to 1$$

Possiamo quindi comprendere che la definizione quantistica di *reliableness* è del tutto analoga a quella fornita in CIT.
Adesso siamo finalmente pronti ad enunciare l'analogo quantistico del primo teorema di Shannon.

Teorema di Schumacher-Josza
Si consideri la sorgente quantistica ρ. Allora questa sorgente ammette uno schema di compressione/decompressione quantistico *reliable* con rate di compressione R *se e solo se:*

$$R > S(\rho)$$

dove $S(\rho)$ indica ovviamente *l'entropia di Von Neumann* della sorgente ρ in esame.

Come dicevamo abbiamo una completa analogia con il caso classico.

Tutto quello che ci resta da fare per completare la nostra *roadmap* della teoria dell'informazione è generalizzare il secondo teorema di Shannon al caso quantistico.

Vogliamo quindi parlare di *capacità* di canali quantistici rumorosi (a *rumore scorrelato*). In QIT nel parlare di questo tema assistiamo ad una biforcazione non presente in CIT, difatti posso voler utilizzare il mio canale quantistico sia per comunicare *informazione classica*, ovvero comunicare dei *bit* di informazione, sia per comunicare *informazione quantistica*, ovvero dei *qbit* di informazione. Per un canale quantistico dovranno dunque essere definite *due diversa capacità*: la sua *capacità classica* C e la sua *capacità quantistica* Q.

Definizione informale di capacità di un canale quantistico
In maniera informale possiamo definire la *capacità classica di un canale quantistico* C come:

$$\lim_{n\to+\infty} \frac{\#\,\text{bit comunicati}}{\#\text{usi del canale}} = C \;\; \text{con}: \;\; p_E \to 0$$

Ovvero come il rapporto fra il numero di bit di informazione classica che riesco a comunicare con errore asintoticamente nullo e numero di qbits utilizzati per eseguire la comunicazione. Difatti con il termine **numero di usi di un canale quantistico** si intende il numero di qbits trasmessi sequenzialmente lungo il canale. Aver utilizzato n volte il canale è sinonimo di aver inviato n qbits lungo il canale.
In maniera analoga possiamo definire la *capacità quantistica di un canale quantistico* Q come:

$$\lim_{n\to+\infty} \frac{\#\,\text{qbit comunicati}}{\#\text{usi del canale}} = Q \;\; \text{con}: \;\; p_E \to 0$$

Del tutto analoga alla precedente, solo che adesso inviando stati quantistici vogliamo proprio trasmettere informazione quantistica, qbits di informazione.

Definizione formale di capacità di un canale quantistico
A livello formale modellizziamo un canale quantistico rumoroso come una *mappa CPTP* Φ (*trace preserving*, *completely positive*, *map*). Inoltre modellizziamo anche gli analoghi quantistici dei processi di codifica e decodifica con mappe quantistiche CPTP Φ_C, Φ_D. Chiamando infine k il numero di qbits trasmessi lungo il canale Φ si consideri la seguente quantità:

$$\lim_{\varepsilon\to 0}\lim_{k\to+\infty}\sup\left\{\frac{k}{n}\ :\ \exists\,\Phi_E^{k\to n}\ ,\ \exists\,\Phi_D^{n\to k}\ ,\ \min_{m\in M} F_E\left(|m\rangle\ ,\ \Phi_D^{n\to k}\circ\Phi^{(n)}\circ\Phi_E^{k\to n}\right)>1-\varepsilon\right\} \tag{6}$$

dove k/n è ovviamente il rate di trasmissione T. Adesso: se lo spazio M è composto da stati puri mutuamente ortogonali, ovvero se abbiamo tra le mani una sorgente che invia stati mutuamente ortogonali

$$M = \left\{\,|0\rangle\,,|1\rangle\,\right\}^{\otimes k}$$

allora stiamo evidentemente comunicando informazione classica (gli stati sono distinguibili) e la quantità (6) è la definizione formale di C.
Al contrario se M è composto da stati non ortogonali

$$M = \mathbb{C}^{2^{\otimes k}}$$

stiamo inviando informazione quantistica, e (6) è evidentemente la definizione formale di Q.
Possiamo notare l'evidente analogia fra (6) e la definizione di capacità in CIT: in (6) stiamo cercando il sup del rate di trasmissione *tale da* essere un rate di trasmissione *achievable* (la definizione di achiavableness viene fatta nel contesto quantistico con l'entaglement fidelity F_E, in completa analogia con quanto già visto nella generalizzazione del primo teorema di Shannon).

Notiamo che la capacità di un canale quantistico, al contrario della capacità di un canale classico, *non è additiva*. Ovvero nel caso classico se uso un canale n volte, o se ho n copie del medesimo canale, ottengo una capacità $C^{(n)}$ che, come ci si aspetta, è n volte la capacità per singolo uso $C^{(1)}$.

$$C^{(n)} = nC^{(1)}$$

Nel caso quantistico questo *non* è più sempre vero! E si possono avere situazioni del tipo:[3]

$$C^{(n)} > nC^{(1)}$$

Dove adesso stiamo parlando di capacità classiche su canale quantistico, mentre prima si parlava di capacità classiche su canale classico.
Dunque per parlare in generale della capacità di un canale quantistico, a prescindere da quanti usi se ne faccia, dobbiamo *regolarizzare* la sua capacità, ovvero calcolare il limite del rapporto. Quest'ultima è una quantità costante.

Definizione di capacità regolarizzata
Definiremo *capacità classica regolarizzata di un canale quantistico* la seguente:

$$C_\chi = \lim_{n\to+\infty}\frac{C^{(n)}}{n}$$

dove n è il numero di usi.

Siamo adesso pronti per introdurre il primo pezzo della generalizzazione del secondo teorema di Shannon.

3 La non additività della capacità porta ad una peculiare proprietà dei canali di comunicazione quantistici detta *superattivazione*: supponiamo di avere due canali quantistici, allora anche se la capacità di entrambi i canali è zero non è detto che la capacità totale della coppia di canali sia zero! Detto diversamente: nel mondo della comunicazione quantistica posso avere due linee telefoniche singolarmente rotte, ma che insieme riescono a consentire il passaggio di informazione.

Teorema di Holevo-Schumacher-Westmoreland (1997)
La **capacità classica regolarizzata** è:

$$C_\chi = \max_{\{p_x\}} \chi(\rho)$$

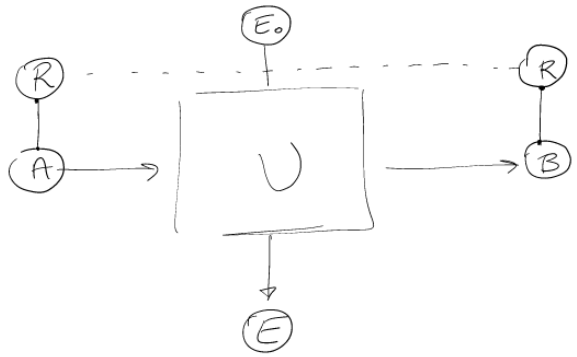
ovvero:

$$C_\chi = \max_{\{p_x\}} \chi(\rho) = \max_{p_x} \left[S\left(\sum_x p_x \rho_x \right) - \sum_x p_x S(\rho_x) \right]$$

Dove il sup scorre su tutte le possibili sorgenti quantistiche $\rho = \sum_x p_x \rho_x$ variando le p_x e *tenendo le* ρ_x *fisse.*
Se invece siamo interessati alla *capacità quantistica (per trasmissione classica) a singolo uso* $C^{(1)}$ allora abbiamo:
Single use capacity
La capacità a singolo uso del canale di comunicazione quantistico Φ (pensato come una *mappa CPTP*) è:

$$C_\chi(\Phi^{\otimes 1}) = C^{(1)} = \max_{\{p_x, \rho_x\}} \chi\big(p_x, \Phi(\rho_x)\big)$$

Mentre per la capacità quantistica Q abbiamo il seguente: possiamo modellizzare la trasmissione di informazione quantistica attraverso un canale quantistico rumoroso come una *mappa quantistica* intesa con il formalismo delle mappe di *interazione sistema-ambiente*.



Dove A è la nostra sorgente, che potrebbe benissimo essere uno stato misto ρ, quindi la purifichiamo con R, considerando il sistema complessivo AR che possiamo sempre pensare come uno stato puro. Pensiamo la mappa Φ agente su A nella seguente maniera:

$$\Phi(\rho) = \mathrm{Tr}_E \left[U(\rho \otimes \rho_E) U^\dagger \right]$$

Dunque consideriamo il sistema complessivo come un sistema isolato, che quindi evolve con dinamica unitaria U. Si noti che per il *teorema di Stinespring* posso di sicuro considerare in questa descrizione lo stato iniziale dell'ambiente (ρ_E) come uno stato puro. Notare che il sistema R rimane immutato, lo possiamo pensare come sottoposto ad un canale ideale $\mathbb{I}d$, ovvero alla matrice identità $\mathbb{I}$. Notare inoltre che quanto ho un sistema puro e lo penso come diviso in due sottosistemi allora i due sottosistemi possono essere stati misti diversi fra loro, ma *avranno eguale entropia.* Questo se ci si pensa è molto intuitivo, e può volendo essere formalmente dimostrato attraverso la *decomposizione di Schmidt*. Possiamo quindi, ad esempio dire che:

$$S(\rho) = S(\rho_R)$$

dato che questi due sistemi insieme formano il sistema AR che per definizione è nello stato puro $|AR\rangle\langle AR|$.

Definizione di entropy exchange
L'entropy exchange $S(\rho, \Phi)$ di una sorgente quantistica ρ che comunica attraverso un canale rumoroso modellizzato da una mappa quantistica Φ (intesa secondo il modello di mappa quantistica come interazione sistema-ambiente) è per definizione:

$$S(\rho, \Phi) \doteq S(E)$$

dove E è lo stato finale dell'ambiente.
Questa quantità è nota come *entropy exchange* in quanto per il teorema di Stinespring lo stato iniziale dell'ambiente lo possiamo sempre prendere come puro, quindi ad entropia nulla, e quindi l'entropia dello stato finale dell'ambiente dev'essere tutta stata fornita dall'interazione Φ della sorgente ρ con l'ambiente. Dunque si capisce che $S(E)$ è proprio l'entropia scambiata dal sistema all'ambiente!

Coherent information
Definiamo *informazione coerente* I_c di una sorgente quantistica ρ che comunica lungo il canale rumoroso Φ la seguente:

$$I_c(\rho, \Phi) \doteq S(\Phi(\rho)) - S(\rho, \Phi)$$

Teorema di Lloyd-Devetok-Shor (2005)
La capacità quantistica di un canale quantistico è:

$$Q(\Phi) = \lim_{n \to +\infty} \frac{1}{n} \max_{\rho^{(n)}} I_c\left(\rho^{(n)}, \Phi^{\otimes n}\right)$$

Dove si può notare che la regolarizzazione della capacità è stata fatta all'interno di quest'ultima formula (notare il limite della frazione). Difatti anche questa capacità quantistica non è additiva!

VANTAGGI DELL'INFORMATICA QUANTISTICA

Siamo stati ovviamente costretti, dati i limiti di tempo, a lasciar fuori dalla nostra *roadmap* numerose tematiche inerenti alla QIT; questo è un peccato, in quanto molti dei temi lasciati fuori hanno il pregio di mettere ben in luce i *vantaggi* dell'informatica quantistica rispetto all'informatica classica. Per completare la nostra trattazione sulle differenze fra CIT e QIT ci sembra quindi doveroso un accenno finale ad alcuni dei principali vantaggi che l'informatica quantistica offre.

- L'informazione quantistica in transito è intrinsecamente sicura grazie al *no cloning theorem* e al fatto che è impossibile distinguere stati quantistici non ortogonali.
- L'informatica quantistica ci consente di sviluppare protocolli di scambio di chiavi private, noti come protocolli QKD (*Quantum Key Distribution*), che consentono di implementare cifrari a chiave privata, gli unici cifrari che possono essere considerati matematicamente inviolabili (vedi ad esempio il *Cifrario di Vernam*).
- In informatica quantistica si ha accesso al *superdense coding*, che consente di trasmettere due bit di informazione classica per ogni qbit manipolato (questo non viola il *teorema di Holevo* in quanto nel superdense coding il setup è differente: si opera su stati entangled condivisi fra mittente e destinatario).
- In alcuni casi gli algoritmi quantistici possono essere utilizzati per eseguire computazioni a velocità molto superiori rispetto a qualunque algoritmo classico noto. L'esempio più famoso è quello *dell'algoritmo di Shor*, in grado di fattorizzare numeri in tempi polinomiali anziché esponenziali come nel caso classico; questo è particolarmente rilevante in quanto in linea di principio consentirebbe di violare i moderni schemi di cifratura asimmetrica, quali ad esempio lo schema RSA, che stanno alla base di ogni sicura comunicazione sulla rete internet.
- Gli stati quantistici possono essere efficacemente trasmessi attraverso protocolli quali il *teletrasporto quantistico*.