Causal Networks

Giuseppe Magazzù

2021 - 2022

Contents

1 The Potential Outcome Framework **1**
1.1 Potential Outcome 1

2 Flow of Associations and Causal Graphs **3**
2.1 Bayesian Networks 4
2.2 Causal Networks 4
2.3 D-Separation 6

3 Causal Models **8**
3.1 Intervention and do-Operator 8
3.2 Modularity and Adjustment Formula 9
3.3 Backdoor Adjustment 10

4 Structural Causal Models **11**

5 Randomized Experiments **12**

6 Nonparametric Identification **13**

7 Estimation **14**

8 Unobserved Confounding **15**

9 Instrumental Variables **16**

10 Causal Discovery from Observational Data **17**

11 Transfer Learning and Transportability **18**

12 Counterfactuals **19**

Chapter 1

The Potential Outcome Framework

Fundamental Problem of Causal Inference:

- Average Treatment Effects and Missing Data Interpretation
- Ignorability - Exchangeability
- Conditional Exchangeability - Uncounfoundedness
- Positivity - Overlap - Common Support and Extrapolation
- No interference, Consistency, SUTVA

1.1 Potential Outcome

- X **Treatment** variabile aleatoria
- Y **Outcome** variabile aleatoria
- Z **Covariate** insieme di variabili aleatorie

Il **potential outcome** $Y(x)$ denota quale sarebbe l'**outcome** quando $X = x$, ovvero se il treatment che è stato scelto è x. Una volta che viene osservato il **potential outcome** Y(x) questo assume valore Y chiamato **outcome**.

Una **popolazione** consiste di molti individui o unità. Ogni individuo (unità) è associato a uno o più variabili Z **covariate**.

Denotiamo X, Y, Z dell'individuo i-esimo come X_i, Y_i, Z_i.

$Y(x)$ è una variabile aleatoria
$Y_i(x)$ non è trattata come una variabile aleatoria poiché specifica per l'individuo

Individual Treatment Effect (ITE)

Per verificare se c'è una relazione causale tra X e Y si può calcolare la seguente differenza $\tau_i \triangleq Y_i(1) - Y_i(0)$.

Se il potential outcome è uguale in entrambi i casi ($\tau_i = 0$) allora non c'è relazione causale. Nel caso contrario si può notare che diversi treatment portano a diversi outcome e quindi c'è una relazione causale.

Non possiamo osservare tutti i **potential outcome** poiché l'osservarne uno influenzerebbe gli altri. I **potential outcome** che non possono essere osservati vengono chiamati **counterfactual**, mentre quello che osserviamo è il **factual**.

$$
\begin{array}{lllll}
Y(1) = ? & \Rightarrow & Y = 0, X = 1 & \Rightarrow & Y(1) = ? \quad \text{counterfactual} \\
Y(0) = ? & \Rightarrow & Y = 0, X = 0 & \Rightarrow & Y(0) = 1 \quad \text{factual}
\end{array}
$$

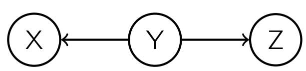
Average Treatment Effect (ATE)

$\tau \triangleq \mathbb{E}[Y_i(1) - Y_i(0)]$

Chapter 2

Flow of Associations and Causal Graphs

In un grafo diretto indichiamo con $pa(X)$ i **genitori** del nodo X e $ch(Y)$ i **figli** del nodo Y.



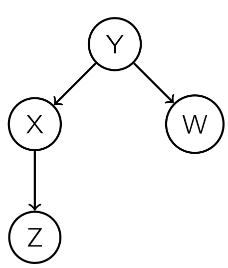
$$pa(Y) = \emptyset \quad ch(Y) = \{X, Z\}$$
$$pa(X) = Y \quad ch(X) = \emptyset$$

Un **cammino** è una qualsiasi sequenza di nodi adiacenti, indipendentemente dalla direzione degli archi che li collega.

Un **cammino diretto** è un qualsiasi **cammino** tra due nodi in cui tutti gli archi che li collega hanno la stessa direzione.

$de(Y)$ è l'insieme dei nodi **discendenti** del nodo Y, ovvero tutti i nodi che possono essere raggiunti da Y.

$an(Z)$ è l'insieme dei nodi **antenati** del nodo Z, ovvero tutti i nodi che da Z possono essere raggiunti.



$$an(Z) = \{X, Y\}$$
$$an(Y) = \emptyset$$

$$de(Y) = \{X, Z, W\}$$
$$de(W) = \emptyset$$

2.1 Bayesian Networks

In una rete bayesiana vogliamo modellare la distribuzione di probabilità $P(X_1, \dots, X_n)$. Tramite la **Chain Rule** possiamo riscriverla nel seguente modo:

$$\begin{aligned} P(X_1, \dots, X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \cdots P(X_n|X_1, \dots, X_{n-1}) \\ &= P(X_1) \prod_{i=2}^{n} P(X_i|X_1, \dots, X_{i-1}) \end{aligned}$$

Local Markov Assumption: Un nodo X è indipendente da tutti i suoi non-discendenti date l'evidenze di tutti i nodi genitori $pa(X)$.

Data questa assunzione possiamo applicare la fattorizzazione della probabilità P.

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i|pa(X_i))$$

In questo modo possiamo ridurre il numero di parametri della rete.

Una distribuzione di probabilità P si dice che sia Markov se ogni nodo X rispetta la Local Markov Assumption.

L'assunzione di Markov non ci da informazioni riguardo relazioni di dipendenza tra i nodi. Quindi estendiamo quest'assunzione.

Minimality Assumption

- Dato $pa(X)$, un nodo X è indipendente da tutti i suoi non-discendenti.
- I nodi adiacenti sono dipendenti.

Dato un DAG G, se P è Markov allora sappiamo:

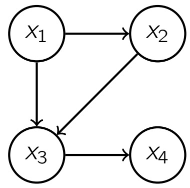
- P soddisfa un insieme di indipendenze specificate dalla struttura di G.
- se P soddisfa pure la Minimality Assumption, allora l'insieme di indipendenze è minimale, ovvero P non soddisfa altre indipendenze in G. Questo equivale a dire che tutti i nodi adiacenti sono dipendenti.

2.2 Causal Networks

Una variabile X si dice **causa** di una variabile Y se Y può cambiare in risposta a un cambiamento di X.

Causal Edge Assumption: Ogni variabile associata a un nodo è causata dalle variabili dei nodi genitori.

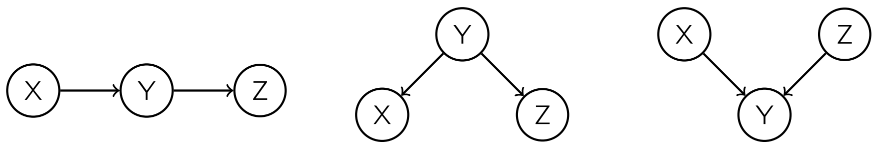
Un **Grafo Causale** è un DAG in cui è soddisfatta la proprietà **Causal Edge Assumption**.



X_1 causa direttamente X_2 e X_3
X_2 causa direttamente X_3
X_3 causa direttamente X_4

X_1 causa indirettamente X_4

Per capire la differenza tra association flow e causation flow utilizziamo dei grafi elementari.



(a) Chain (b) Fork (c) Collider

Il flusso di associazione esprime che due nodi del grafo sono associati o meno.

Vogliamo sapere se due nodi sono (statisticamente) dipendenti o indipendenti.

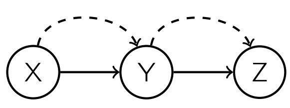
Un-connected Nodes

Due nodi sconnessi non sono associati. Questo può essere dimostrato applicando la fattorizzazione.

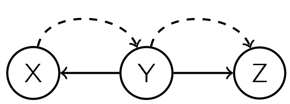
$$\begin{aligned} P(X,Y) &= P(X|Y)P(Y) && \text{(probabilità composta)} \\ &= P(X|pa(X))P(Y|pa(Y)) && \text{(fattorizzazione)} \\ &= P(X)P(Y) && \text{indipendenza} \end{aligned}$$

Due nodi connessi adiacenti sono dipendenti per la **local edge assumption**.

Chain e Fork



X causa Y che causa quindi Z



Y causa X e Z

Condizionando Y in entrambi i casi, il flusso di associazione da X a Z viene bloccato in Y. Per la **Local Markov Assumption** $X \perp\!\!\!\perp Z|Y$.

Proof. Data la Chain $X \to Y \to Z$, $P(X, Z|Y) \stackrel{?}{=} P(X|Y)P(Z|Y)$

$$\begin{aligned} P(X, Y, Z) &= P(X)P(Y|X)P(Z|X, Y) && \text{(probabilità composta)} \\ &= P(X)P(Y|X)P(Z|X) && \text{(fattorizzazione)} \end{aligned}$$

$$P(X, Z|Y) = \frac{P(X, Y, Z)}{P(Y)} = \frac{P(X)P(Y|X)P(Z|X)}{P(Y)} = P(X|Y)P(Z|Y)$$

□

Proof. Data la Fork $X \leftarrow Y \to Z$, $P(X, Z|Y) \stackrel{?}{=} P(X|Y)P(Z|Y)$

$$\begin{aligned} P(X, Y, Z) &= P(Y)P(X|Y)P(Z|X, Y) && \text{(probabilità composta)} \\ &= P(Y)P(X|Y)P(Z|Y) && \text{(fattorizzazione)} \end{aligned}$$

$$P(X, Z|Y) = \frac{P(X, Y, Z)}{P(Y)} = \frac{P(Y)P(X|Y)P(Z|Y)}{P(Y)} = P(X|Y)P(Z|Y)$$

□

Collider

Proof. Data il Collider $X \to Y \leftarrow Z$, $P(X, Z) \stackrel{?}{=} P(X)P(Z)$

$$\begin{aligned} P(X, Y) &= \sum_y P(X, Z|Y = y) && \text{(marginalizzazione)} \\ &= \sum_y P(X)P(Z)P(Y = y|X, Z) && \text{(probabilità composta)} \\ &= P(X)P(Z)\cancel{\sum_y P(Y = y|X, Z)} && \text{(somma unitaria)} \\ &= P(X)P(Z) \end{aligned}$$

□

2.3 D-Separation

Il flusso di associazione è simmetrico, mentre il flusso di causazione non è simmetrico e scorre nei **direct path**.

Un percorso p si dice bloccato da un insieme di nodi S sse

- p contiene una chain $A \to B \to C$ o una fork $A \leftarrow B \to C$ e $B \in S$
- p contiene un collider $A \to B \leftarrow C$ e $B \notin S, de(B) \not\subseteq S$

Se l'insieme S blocca ogni percorso tra due nodi X e Y, allora X e Y si dicono **d-separati** da S, e quindi indipendenti dato S.

ESEMPI...

La d-separazione implica l'indipendenza condizionale.

- Due variabili X e Y sono d-separate nel grafo G quando condizionate dall'insieme S ($X \perp\!\!\!\perp_G Y|S$).
- Due variabili X e Y sono indipendenti nella distribuzione p quando condizionate dall'insieme S ($X \perp\!\!\!\perp_p Y|S$).

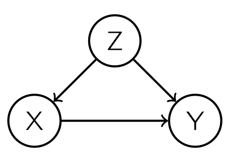
Global Markov Assumption

Dato una distribuzione P che è Markov rispetto al grafo G, se le variabili X e Y sono d-separate in G dato l'insieme condizionale S, allora X e Y sono indipendenti in P condizionalmente a S ($X \perp\!\!\!\perp_G Y|S \Rightarrow X \perp\!\!\!\perp_p Y|S$).

Global Markov Assumption $\Leftrightarrow$ Local Markov Assumption $\Leftrightarrow$ Bayesian Factorization

L'associazione è la causazione scorrono nei grafi diretti. Nei grafi causali (DAG + Local Edge Assumption) la causazione scorre nei percorsi diretti.

- Causal Association: associazione nei percorsi diretti
- Confounding Association: nel percorso è presente una associazione non causale che rende l'associazione di non causazione



- X $\leftarrow$ Z $\rightarrow$ Y   Confounding path
- X $\rightarrow$ Y, Z $\rightarrow$ Y   Causal path

L'associazione è simmetrica, la causazione è non simmetrica. La causazione è un sottoinsieme dell'associazione.

- Local Markov
- Minimality
- Causal Edge

Chapter 3

Causal Models

In un **randomized control experiment** tutti i fattori $X_1, ..., X_n$ che influenzano l'outcome Y sono statici o variati randomicamente tranne uno X_i. In questo modo un cambiamento nell'outcome Y sarà dovuto solo da quel fattore X_i.

Spesso non è possibile effettuare gli esperimenti per cause pratiche o etiche, quindi si eseguono degli studi osservazionali in un cui si raccolgono i dati rispetto ad alterarli.

Il problema degli studi osservazionali è distinguere la causazione dalla correlazione.

3.1 Intervention and do-Operator

Intervening	**Conditioning**
Cambiamo il sistema fissando $X = x$ e osserviamo i cambiamenti su tutta la popolazione.	Non cambiamo nulla e consideriamo un sottoinsieme della popolazione $X = x$.

Denotiamo l'intervention con il do-operator $do(X = x)$.

$$P(Y(x) = y) = P(Y = y|do(X = x)) = P(y|do(x))$$

$P(Y\|X = x) = P(y\|x)$	observational distribution
$P(Y\|do(X = x)) = P(y\|do(x))$	interventional distribution

Un'espressione con do è detta interventional expression, mentre senza do è detta observational expression.

Un'interventional expression che può essere ridotta a una observational expression è detta **identificabile**.

Una stima è detta causale se contiene il do-operator, statistica altrimenti.

pre-intervention distribution	**post-intervention distribution**
$P(Y\|do(x), Z = z)$	$P(Y\|x, Z = z)$

3.2 Modularity and Adjustment Formula

Il **causal mechanism** è un meccanismo che genera X_i come distribuzione condizionale di X_i dati i genitori (cause) $pa(X_i)$, ovvero $P(X_i|pa(X_i))$.

Assunzione: Le intervention sono locali

Un intervention su una variabile X_i cambia solo il suo causal mechanism, non cambia il causal mechanism che genera un'altra variabile X_j.

Modularity - Independence Mechanism - Invariance

Se interveniamo su un insieme di variabili S fissando valori costanti, allora per ogni variabile $X_i \in \{X_1, ..., X_n\}$ possiamo affermare che:

1. Se $X_i \in S$, allora $P(X_i = x|pa(X_i)) = 1$ se x è il valore assegnato dall'intervention $do(X_i = x)$, 0 altrimenti
2. Se $X_i \notin S$, allora $P(X_i = x|pa(X_i))$ rimane inalterato

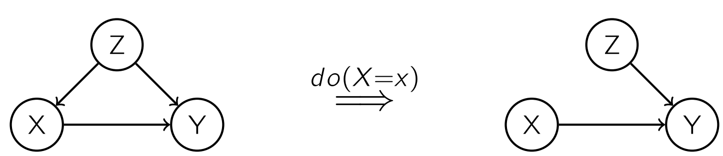
Data una variabile $X_i \in S$, un valore x di X_i è **consistente** con l'intervention su X_i se x è uguale al valore che è stato assegnato nell'intervention $do(X_i = x)$.

Una volta impostato il valore di X_i non importa più il valore dei nodi genitori $pa(X_i)$, poiché al variare di $pa(X_i)$ non varierà il valore di X_i.

Questo possiamo rappresentarlo attraverso un grafo (detto manipulated graph o **post-intervention graph**) in cui si rimuoviamo gli archi entranti in X_i; se non ce ne sono, vuol dire che l'intervention non ha conseguenze sulla distribuzione post-intervention $p(y|do(x)) = p(y|x)$.

Adjustment Formula

Dato il seguente grafo causale e il suo corrispettivo grafo manipolato a seguito dell'intervention sulla variabile X.



pre-intervention graph

post-intervention graph

Possiamo osservare sotto la condizione di **modularity assumption** che:

1. $P(Y = y|do(X = x)) = Pm(Y = y|X = x)$
2. $P(Z|do(X = x)) = P_m(Z|x) = P(Z)$
3. $P(Y|do(X = x), Z) = P_m(Y|x, Z) = P(Y|x, Z)$

La probabilità marginale $P(Z)$ e la probabilità condizionale $P(Y|x, Z)$ sono invarianti all'intervention. Questo perché il causal mechanism di X non influenza il causal mechanism delle altre variabili.

Date le seguenti uguaglianze:

1. $P_m(Z = z|X = x) = P_m(Z = z) = P(Z = z)$
2. $P_m(Y = y|X = x, Z = z) = P(Y = y|X = x, Z = z)$

Possiamo scrivere la seguente interventional expression:

$$\begin{aligned}
P(Y = y|do(X = x)) &= P_m(Y = y|X = x) \\
&= \sum_z P_m(Y = y|X = x, Z = z) P_m(Z = z|X = x) \\
&= \sum_z P_m(Y = y|X = x, Z = z) P_m(Z = z) \\
&= \sum_z P(Y = y|X = x, Z = z) P(Z = z)
\end{aligned}$$

Il causal mechanism di Y, $P(Y = y|X = x, Z = z)$ viene aggiustato, pesato da Z (adjusting/controlling for Z).

Quest'espressione prende il nome di Adjustment Formula e ci permette di rispondere a una interventional query solo usando dati osservazionali.

Causal Effect Rule

Generalizzazione della adjustment formula.

Dato un grafo G in cui un insieme di variabili $pa(X)$ sono i genitori di X

$$\begin{aligned}
P(Y = y|do(X = x)) &= \sum_u P(Y = y|X = x, pa(X) = u) P(pa(X) = u) \\
&= \sum_u \frac{P(Y = y, X = x, pa(X) = u)}{P(X = x|pa(X) = u)}
\end{aligned}$$

3.3 Backdoor Adjustment

Chapter 4

Structural Causal Models

Chapter 5

Randomized Experiments

Chapter 6

Nonparametric Identification

Chapter 7

Estimation

Chapter 8

Unobserved Confounding

Chapter 9

Instrumental Variables

Chapter 10

Causal Discovery from Observational Data

Chapter 11

Transfer Learning and Transportability

Chapter 12

Counterfactuals