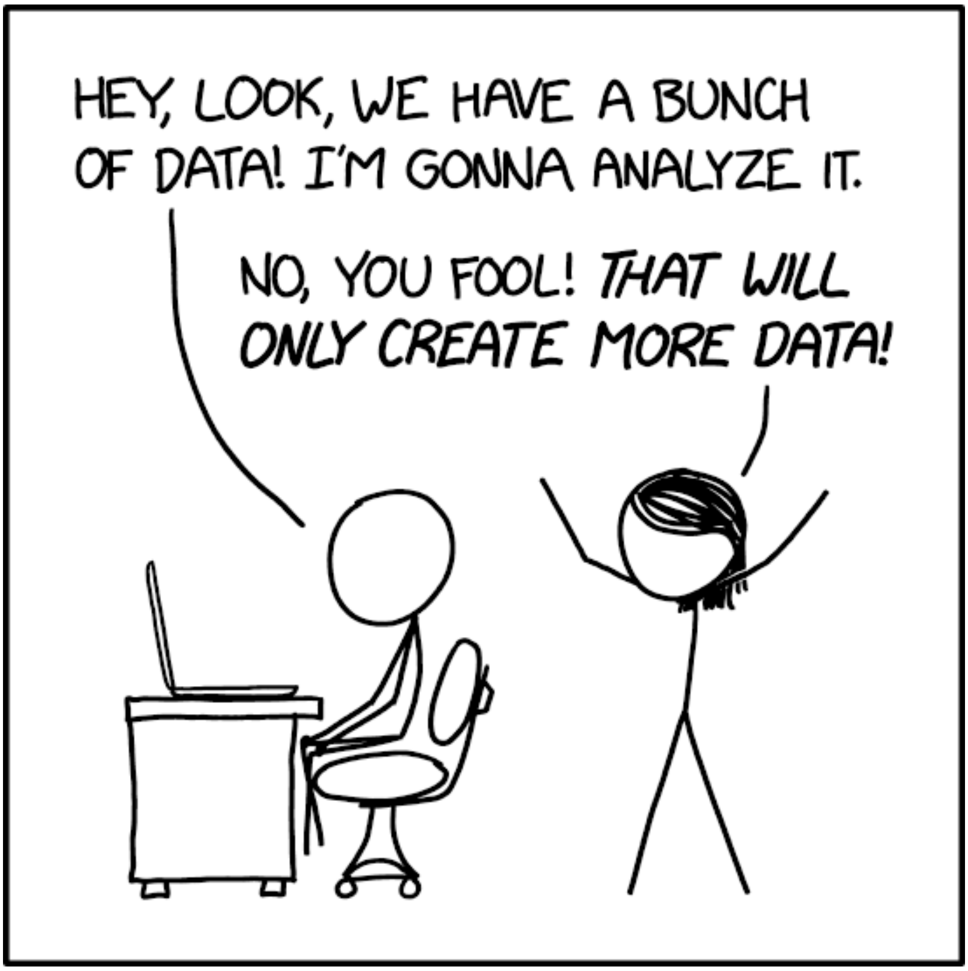
Lezione 0

Il paradosso di "Analizzare i Dati"



Che cosa e' l'Inferenza in Fisica ?

- In generale:
  - Esistono delle *quantita' misurabili* attraverso qualche definizione operativa (**osservabili** x,y...) il cui valore dipende (*in qualche modo*) dai parametri ignoti. Quantita' fisiche (non "variabili")
  - Esistono delle "leggi fisiche" - che determinino i valori e l'evoluzione temporale delle Osservabili. Senza questa assunzione non si fa Fisica. Queste leggi possono essere parametrizzate con dei valori (***parametri***: μ, ν, …) lettere greche - rappresentano *quantita' fisiche* reali , ma inaccessibili all'osservazione diretta
  - Da *osservazione* di osservabile $x \in X$ inferisco *affermazioni* sui parametri ignoti $\mu \in A$.
  - L'output di questo tipo di "analisi dati" **non e'** una certa quantita' di dati ulteriori. Questo tipo di "analisi" e' intrinseco al paradigma scientifico galileiano:

    Teoria -> Predizione -> Osservazione -> *Inferenza* -> Teoria

  - Il caso semplice: supponiamo relazione deterministica:

$$(x, y, \dots) = f(\mu, \nu, \dots)$$

Allora posso fare inferenze deterministiche:

$$(\mu, \nu, \dots) = f^{-1}(x, y, \dots)$$

L'inferenza e' pero' spesso non-deterministica:

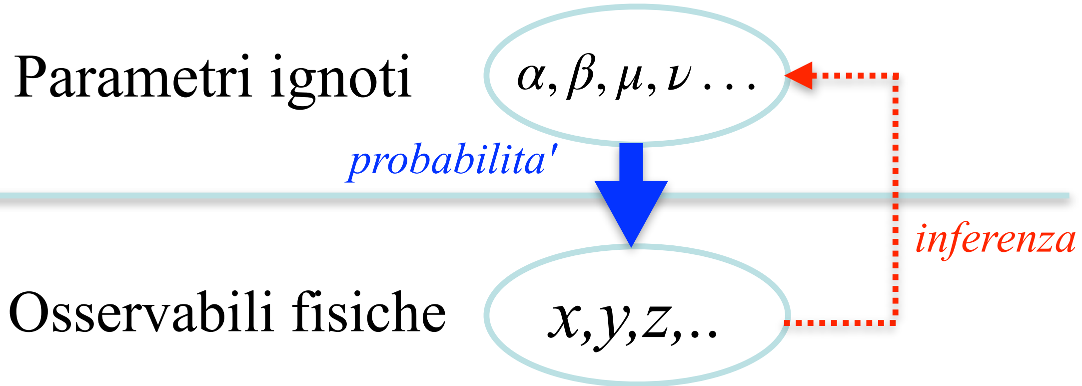
1. f non e' invertibile (***indeterminazione*** = informazione incompleta = ambiguita'= le osservabili non determinano univocamente i parametri
2. f ***non e' una funzione ripetibile/deterministica*** (presenza di "errori"). I parametri non determinano del tutto le osservabili

- In entrambi i casi, si puo' solo fare una inferenza che da' come risultato una regione nello spazio A - cioe' affetta da una ***incertezza.*** Questo si puo' ottenere semplicemente come *immagine inversa* (in senso insiemistico) della osservazione *x*. ("propagazione degli 'errori' " all'indietro da' incertezza su μ).
- Questa incertezza **non** e' pero' *incertezza statistica*.

Che cosa e' la Inferenza Statistica ?

- La ***inferenza statistica*** entra in gioco quando qualche osservabile, non solo non e' determinata, ma ha un comportamento che si puo' descrivere secondo una ***distribuzione di probabilita'*** dipendente dai parametri.
- Si dice che e' una ***osservabile aleatoria/casuale/random.*** Notate che non ho detto *variabile* ma *osservabile,* per sottolineare che ***solo una osservabile puo' avere una distribuzione di probabilita'***.
- Occorre naturalmente definire che cosa si intende con "probabilita"e tra poco lo faremo; ma prima vogliamo discutere ancora un po' il concetto di inferenza, e in che cosa differisce da altre possibili tecniche di "Analisi Dati".
- L'inferenza (anche non statistica) consiste nel cercare di dire qualcosa su μ,ν a partire da osservazioni di x osservabile.
- Solo le *osservabili x* si *misurano* (cioe' si osservano). I *parametri* non si osservano direttamente, quindi non si "misurano" – si possono pero' *stimare* (o *inferire*) a partire dalle misure)
- In statistica, questo si basa sul calcolo di opportune funzioni delle osservabili $s(x, y, \ldots)$ chiamate ***statistiche***.
- Vedremo che una funzione di una variabile aleatoria e' essa stessa una variabile aleatoria (cioe' ha una distribuzione di probabilita')
- La **scienza statistica** e' lo studio del comportamento delle *statistiche* ; la ***inferenza statistica*** e' quel capitolo in cui si discute come usare le statistiche per fare delle inferenze.
- L'inferenza in fisica, se basata su osservabili dotate di una distribuzione di probabilita', e' una inferenza statistica - e ne adotta le metodologie.

L'inferenza (in Statistica) e la Statistica



- <u>Inferenza: determinare i parametri a partire da osservabili **aleatorie**</u>
- Prima di iniziare a parlare di Probabilita' e Statistica, cerchiamo di chiarire che posto occupa l'Inferenza all'interno di tale disciplina, di cui e' solo un capitolo.
- Esempi di questioni statistiche diverse dalla inferenza:

  - Proprieta' delle 'statistiche': <u>***Teoria della Probabilita'***</u>
  - Predizione di *altre* osservabili (es. nel futuro): <u>***Statistica predittiva***</u>
  - Suggerire *azioni* appropriate/ottimali: <u>***Teoria della decisione***</u>

- Naturalmente, la *inferenza* puo' essere una base su cui fare *previsioni*; e le *previsioni* possono aiutare a prendere *decisioni;* tuttavia e' possibile saltare/ignorare passaggi, se la determinazione dei parametri ignoti non e' l'obiettivo primario.
- Idea recente: "***Data Science"****:* fare predizioni *senza* necessita' di una inferenza. Nemmeno necessario conoscere la legge di probabilita' (se esiste): se ne puo' assumere una comoda (*modello*). I parametri eventualmente inferiti non hanno interesse in se'.
- Non ci occupiamo di quanto sopra in questo corso – la statistica ci interessa solo come strumento per ***estrarre dei dati le inferenze*** che ci interessano, allo scopo di mettere in pratica il paradigma galileiano.
- Infine, prima di passare a discutere la probabilita', e per non restare troppo nell'astratto, passero' in rassegna rapidamente alcuni possibili forme che puo' assumere la inferenza statistica.
- Queste corrispondono ad altrettante parti del corso.

Diversi tipi di inferenza statistica

- L'inferenza in se' puo' assumere varie forme (qui elencate in ordine inverso di ricchezza dell'output). L'insieme dei possibili valori dei parametri e' indicato con A.

- ***<u>Inferenza funzionale</u>***:

$$s(x) : x \mapsto f_x(\mu); \quad f : A \to \mathscr{R}$$

La statistica e' una *funzione* dei parametri (dipendente dalle osservazioni) che assegna un qualche genere di “score” a ogni stato possibile della natura, che ne indica in qualche modo la plausibilita'

- ***<u>Stima intervallare:</u>***

$$s(x) : x \mapsto Q_x; \quad Q_x \subset A$$

Determina un sottoinsieme Q dello spazio dei parametri di valori "plausibili" dei parametri

- ***<u>Stima Puntuale</u>***:

$$s(x) : x \mapsto \mu_x \ ; \quad \mu_x \in A$$

- Produce un singolo valore che rappresenta il "miglior valore" per i parametri ignoti.

- ***<u>Test di Ipotesi</u>***

$$s_{\{H_i\}}(x) : x \mapsto i; \quad H_i \subset A \ \forall i \in \{0, \ldots n\}$$

- Dato un set di ipotesi <u>predefinite</u> H_i (“theories”), conclude (decide) quale e' favorita dalle osservazioni. Puo' essere vista anche come "classificazione".

- ***<u>Goodness-Of-Fit (GOF)</u>***

$$s_H(x) : x \mapsto y; \ \ y \in [0,1] \quad (H \subset A)$$

- Valuta quanto bene i dati sono in accordo con una certa ipotesi fissata H

- ***<u>Test di Goodness-Of-Fit</u>***

$$s_H(x) : x \mapsto y; \ \ y \in \{0,1\} \quad (H \subset A)$$

  - Valuta **SE** i dati sono in accordo con una certa ipotesi H oppure no

- **Nel corso tratteremo *<u>tutti</u>* questi tipi di inferenza - in presenza di osservabili aleatorie.**

<u>Nota Bene:</u>

- Molti errori/incomprensioni originano da confusioni/ambiguita' sul tipo di inferenza in gioco.
- Le conclusioni della inferenza riguardo ai parametri non saranno ovviamente sempre "esatte". L'incertezza sulla conclusione dell'inferenza si chiama ***incertezza statistica***, e la **teoria dell'inferenza** consiste nello studio di come quantificarla correttamente.
- E' possibile che la legge di probabilita' della osservabile non sia completamente nota. Questo causa una incertezza ulteriore, che definiremo ***incertezza sistematica***.