Probability

"It is unanimously agreed that statistics depends somehow on probability. But, as to what probability is, and how it is connected with statistics, there has seldom been such complete disagreement and breakdown of communication since the Tower of Babel.

Doubtless, much of the disagreement is merely terminological and would disappear under sufficiently sharp analysis."

— Savage, 1954, The foundations of statistics

Richiami di Probabilita'

- E' ora il momento di parlare del concetto di Probabilita' di una osservabile, menzionato piu' volte ma ancora non affrontato.
- Potreste forse aspettarvi che a questo punto inizi una discussione formale, ma quello di cui abbiamo bisogno e' una definizione *operativa,* consistente con una **quantita' fisica**
- Consideriamo una procedura di osservazione sperimentale ben definita, che e' possibile *replicare illimitatamente* ("ensemble" o "collettivo"). Ogni ripetizione da' un valore di una certa osservabile $x \in X$.
- Sia $E \subset X$ un sottoinsieme di possibili risultati di x, chiamato 'evento'.
  Per ogni osservazione dell'ensemble e' ben definito se si verifica E. [NB: stiamo parlando di una definizione operativa: pensiamo a un pannello di strumenti di laboratorio come esempio]
- Definiamo **Probabilita' dell'evento *E***:

$$P(E) = \lim_{N \to \infty} \frac{n(E)}{N}$$

[definizione "frequentista" (Neyman, Pearson…Von Mises 1939)]

- *n(E)* e' il numero di casi in cui si osserva l'evento E in N tentativi
- Il limite non ha il significato matematico, ma quello fisico (analogo alla definizione di velocita' in meccanica)
- <u>Importante</u>: il limite <u>non deve dipendere</u> da **come** si selezionano i N tentativi.
  - Es. successioni di cifre binarie:
    0101010001011110...
    0101010101010101...
- P(E) ***dipende*** dal "collettivo"(o *ensemble*) N - e questo **NON** e' un difetto.

$P(E)$ e' una <u>quantita' fisica misurabile</u> (come massa o velocita') - ha percio' senso parlarne in Fisica, e discuterne le proprieta'. Questa definizione e' compatibile fon quella adottata dalla Meccanica Quantistica quando enuncia le conseguenze osservabili della funzione d'onda (principio di Born).

- NB: Ovviamente P(E) si puo' definire solo per quantita' x osservabili (non μ!)

Qualche Domanda

- » Qual'e' la probabilita' che piova domani ?
- » Qual'e' la probabilita' che una cifra decimale di π sia un '7' ?
- » Agli stati di un sistema 'caotico' deterministico si possono associare delle probabilita' ?
- » Quale differenza c'e' tra numeri 'casuali' e 'pseudocasuali' ?

La Probabilita' di Kolmogorov

- In matematica, si definisce formalmente una "probabilita" sull'insieme S qualunque *funzione di insieme* P, definita su [una σ-algebra* di] sottoinsiemi di S, avente le seguenti proprieta':

$$1.\ \forall A \subset S \quad P(A) \geq 0$$

$$2.\ P(S) = 1$$

$$3.\ \forall \{A_i\}_{i \in \mathcal{N}} : A_i \subset S,\ A_i \cap A_j = \varnothing\ \forall i \neq j \quad P(\cup_i A_i) = \sum_i P(A_i)$$

Posso prendere ***S*** come lo spazio di tutte le possibili osservazioni sperimentali; un suo elemento ("evento elementare") e' una osservazione sperimentale completamente specificata ["intera immagine del pannello strumenti"], mentre A e' un insieme di potenziali osservazioni [es. tutti gli stati in cui lo strumento X segna un certo valore y, indipendentemente dallo stato degli altri)

Si noti che:

- Ogni elemento (evento elementare) corrisponde ad assegnare i valori di *tutte* le quantita' osservabili.
- La intersezione indica la occorrenza simultanea (AND) di due "letture", mentre la UNIONE rappresenta la condizione di OR di due di esse]
- **Sono banalmente verificati per la probabilita' frequentista.** Quindi questa puo' "sfruttare" tutti i teoremi della teoria matematica di Kolmogorov [Proof].
- Naturalmente, anche ALTRI concetti possono verificare gli stessi assiomi. Ma due concetti con le stesse proprietà' formali non sono necessariamente uguali.

✳ Sigma-algebra $\Sigma \subset \mathcal{P}(S)$:

$$1.\ S \in \Sigma$$

$$2.\ \forall A \in \Sigma \implies \bar{A} \in \Sigma$$

$$3.\ \{A_i\}_{i \in \mathcal{N}} : A_i \in \Sigma\, \forall i \implies \cup_i^{\infty} (A_i) \in \Sigma$$

- La "clausola" della σ-algebra e' rilevante solo se la cardinalita' di S e' maggiore del numerabile (cosa di dubbio significato fisico), perche' in tal caso l'insieme delle parti di S crea problemi ("insiemi non-misurabili")

La probabilita' Bayesiana

- La probabilita' frequentista non e' l'unico concetto che soddisfa gli assiomi Kolmogorov.
- Qualcuno parla percio' di "differenti *intepretazioni* della probabilita'", intendendo con quest'ultima parola la definizione di Kolmogorov.
- Noi seguiamo un punto di vista **fisico**, in cui le definizioni devono essere basate su descrizioni operative, e considerate concettualmente distinte anche quando seguono le stesse regole *matematiche.*
  (=e' la matematica a essere una "interpretazione" della realta')
- La "scuola Bayesiana" fa uso di un concetto di "probabilita'" differente da quello von Mises. Per evitare confusione occorre chiamarlo in modo differente: lo indicheremo percio' con suo nome alternativo "*degree of belief*":

> *π(x)* : *"degree of belief"* di un *soggetto*
> nel fatto che la quantita' abbia valore *x*

– Definizione "*soggettiva" di probabilità (Bayes, Ramsey, De Finetti…):*
  – In gran parte costruita sul principio che eventi su cui si hanno informazioni equivalenti sono per definizione 'equiprobabili'
  – Applicabile a qualunque quantita', *anche non osservabile*
  – Useremo un simbolo differente π per evitare confusione con la probabilita' fisica. Lo chiameremo ***d.o.b*** oppure ***credibilita'***
  – Intuitivamente rappresenta l'informazione che il soggetto ha su *x*
  - es. se ho stessa informazione se x1 e x2, allora $\pi(x_1) = \pi(x_2)$
  – **e' stato dimostrato *formalmente essere una "probabilita' di Kolmogorov"*** [Cox] (ma piu' complicato di quella frequentista).
    -> questo non vuol dire che si possa confondere con la p. fisica
  – *Permette una formulazione della* Inferenza statistica basata sulla determinazione di *π per i parametri ignoti*

Non dipende da un ensemble. Ma dipende dal soggetto.

- Riprendiamo Kolmogorov:
    1. $\forall A \subset S \quad P(A) \geq 0$
    2. $P(S) = 1$
    3. $\forall \{A_i\} : A_i \subset S,\ A_i \cap A_j = \varnothing\ \forall i \neq j \quad P(\cup_i A_i) = \sum_i P(A_i)$

e dimostriamo alcuni semplici teoremi sulla probabilita':

- $P(\overline{A}) = 1 - P(A)$

  K3: $P(A) + P(\overline{A}) = P(A \cup \overline{A}) = P(S)$

  K2: $P(A) + P(\overline{A}) = 1 \implies P(\overline{A}) = 1 - P(A)$
- $0 \leq P(A) \leq 1$
    - Dal precedente, applicando K1.
- $P(\varnothing) = 0$
    - Il vuoto e' complementare di S
    - Per converso, $P(A) = 0 \implies A = \varnothing$ **non** e' un teorema. [esempio frequentista ?]
- $A \subset B \implies P(A) \leq P(B)$
    - $B = A \cup (B \backslash A)$; K3 : $P(B) = P(A) + P(B \backslash A) \geq P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
    - $A \cup B = A \cup (B \backslash A) \implies P(A \cup B) = P(A) + P(B \backslash A)$
    - $(B \backslash A) \cup (A \cap B) = B \implies P(B \backslash A) = P(B) - P(A \cap B)$

Si noti che $P(A \cap B)$ ha il significato di **probabilità congiunta** ('AND') mentre $P(A \cup B)$ e' la probabilita' dello OR logico delle due situazioni**.**

Definizione

- Probabilita' *condizionata* (prob. di evento A sotto condizione che si sia verificato l'evento B - avente $P(B) > 0$)

$$P(A|B) \equiv \frac{P(A \cap B)}{P(B)}$$

Ovvero:

$$P(A \cap B) = P(A|B) \cdot P(B)$$

E' importante osservare che la funzione di insieme $P_B(A) \equiv P(A|B)$
e' una probabilita' nel senso di Kolmogorov, definita **sullo stesso spazio S**

Infatti:

K1: $P(A|B) \geq 0$ (perche' $P(A \cap B) \geq 0$ e $P(B) > 0$)

K2: $P_B(S) = P(S|B) = \dfrac{P(S \cap B)}{P(B)} = \dfrac{P(B)}{P(B)} = 1$

K3:

$$P_B(A \cup C) = \frac{P((A \cup C) \cap B)}{P(B)} = \frac{P((A \cap B) \cup (C \cap B))}{P(B)} = \frac{P(A \cap B) + P(C \cap B)}{P(B)} = P_B(A) + P_B(C)$$

[gli stessi passaggi sono validi anche quando l'unione e' numerabile]

- Si noti che si puo' quindi condizionare in catena:

$$P_B(A|C) = \frac{P_B(A \cap C)}{P_B(C)} = \frac{P(A \cap C \cap B)}{P(B)} \cdot \frac{P(B)}{P(B \cap C)} = \frac{P(A \cap (B \cap C))}{P(B \cap C)} = P(A|B \cap C)$$

E quindi:

$$P(A \cap B \cap C) = P(A \cap B|C) \cdot P(C) = P(A|B \cap C) \cdot P(B|C) \cdot P(C)$$

- **Legge della probabilita' totale:**

Sia $\{A_i\}$ una *partizione* (finita o numerabile) di S (per cui $S = \cup_i A_i$)
Allora $\forall B : B = B \cap S = B \cap (\cup_i A_i) = \cup_i (B \cap A_i)$ - da cui:

$$\forall B : P(B) = \sum_i P(B \cap A_i) = \sum_i P(B|A_i)\, P(A_i)$$

Utile quando le probabilita' condizionate sono piu' facili da calcolare di quella complessiva.

💣Osservazione

- P(A|B) si legge "Probabilita' di A **dato** B"
  - Non significa che ci sia una relazione *causale* tra B e A
  - Non significa che ci sia una relazione *temporale* tra B e A
- E' solo la *restrizione* della probabilita' a un ensemble piu' piccolo.
- B *deve* essere un sottoinsieme in uno spazio (o σ-algebra) su cui e' definita la probabilita' P

Esempio:

$$p(x\,|\,\mu) = c \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

non e' corretto, perche' μ e' un *parametro*, non una *osservabile*.

Per indicare una probabilita' dipendente da un parametro, utilizzeremo la notazione $p(x;\mu)$ oppure $p_\mu(x)$.

Si potra' eventualmente avere: $p(x, y\,|\,z, t; \mu, \lambda)$

Definizione

- Definizione: Eventi A, B si dicono ***indipendenti*** se

$$P(A|B) = P(A)$$

o equivalentemente

$$P(A \cap B) = P(A) \cdot P(B)$$

Questa e' la **legge del prodotto di probabilita'.**

Si noti che e' simmetrica: infatti ne segue anche che:

$$P(B|A) = P(B)$$

Si puo' generalizzare:

$$P(A_1 \cap \ldots \cap A_n) = P(A_1 \cap \ldots \cap A_{n-1} | A_n) \cdot P(A_n) =$$

$$= P(A_1 \cap \ldots \cap A_{n-1}) \cdot P(A_n) = \ldots \text{ ricorsivamente...}$$

$$P(A_1 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2) \ldots \cdot P(A_n)$$

- un set di eventi a due a due indipendenti non e' necessariamente un "set indipendente"

- Indipendenti ≠ "Scorrelate"
(non abbiamo parlato ancora di correlazione)

Teorema di Bayes

- Scriviamo la probabilita' della intersezione in due modi diversi:

$$P(B \cap A) = P(B \mid A) \cdot P(A) = P(A \mid B) \cdot P(B)$$

Da cui:

$$P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A)}$$

(Teorema di Bayes)

N.B. diventa banale se A e B sono indipendenti

Il T. di Bayes ***non e'*** cio' che caratterizza la cosiddetta "statistica/inferenza Bayesiana".

Tuttavia, se si decide di fare uso di una probabilita' *soggettiva,* si puo' attribuire una "probabilita" (*d.o.b*) a classi piu' vaste di eventi, e diventa quindi naturale fare un uso *piu' ampio* di *probabilita' condizionate,* e di conseguenza del T. di Bayes.