Appunti di statistica

1. Concetti generali

La probabilità è una quantificazione del grado di aspettativa nei confronti di un evento. Considereremo la probabilità un concetto elementare. La quantificazione della probabilità è un processo frutto di un atteggiamento soggettivo o culturale. Esso si basa difatti su un modello della realtà.

In ambiente scientifico è il successo dell'assunzione che si fa sulla probabilità di un evento a decretarne l'accettazione.

Esiste una sola verifica della probabilità di un evento, la misurazione della frequenza con cui questo avviene, possibile solo nel caso in cui l'evento è ripetibile ed essa stessa soggetta ad errore di natura statistica.

Almeno per la scienza occidentale, a partire dal Rinascimento, si afferma il determinismo, l'idea che una descrizione completa delle cause che lo precedono permetta una previsione esatta dell'evento. Così la fisica del secolo XIX attribuiva alla incompletezza della descrizione che si può dare di un sistema le limitazioni nel prevedere il risultato di una misura effetuata su di esso. La meccanica quantistica ha poi introdotto nella descrizione nella descrizione della dinamica una indeterminazione *non eliminabile*.

Sia x una variabile casuale, cioè una variabile il cui valore non è noto a priori ma può assumere valori in un insieme $\mathcal{A}$ con probabilità in un sottoinsieme dei reali positivi $\mathcal{P}$. Per semplicità possiamo considerare i due casi:

- $\mathcal{A}$ è numerabile, a_i sono i suoi elementi e p_i le corrispondenti probabilità, con $\sum_i p_i = 1$.
- $\mathcal{A}$ è un intervallo I dei reali e $P(x)$ ne denota la densità di probabilità, con $\int_I P(x)dx = 1$.

Corrispondentemente la media statistica di una funzione $f(x)$ è definita dalle relazioni

$$\langle f \rangle = \begin{cases} \sum_{x\in\mathcal{A}} f(x)P(x) \\ \int_{\mathcal{A}} f(x)P(x)dx \end{cases}$$

esercizio 1 Calcolare la media del quadrato del risultato del lancio di un dado a sei facce ideale.

Consideriamo due variabili casuali: x con valori in $\mathcal{A}$ e y con valori in $\mathcal{B}$.

Definisco le variabili x ed y indipendenti se

$$P(a,b) \equiv P(x=a, y=b) = P(x=a)P(y=b)$$

dove ho indicato con $P(x=a, y=b)$ la probabilità che congiuntamente si verifichi che $x=a$ e che $y=b$, detta appunto probabilità congiunta.

esempio: Il modello ideale di un dado a sei facce attribuisce uguale probabilità $1/6$ a tutte le facce. Nel lancio di due dadi siano x il valore ottenuto dal primo dado e y il valore ottenuto dal secondo. L'assunzione che si fa in condizioni normali per un simile esperimento è che le variabili x ed y siano indipendenti, ovvero che per ogni valore assunto da x, y sia $P(x,y) = P(x)P(y) = 1/36$.

esercizio 2 calcolare la probabilità che il risultato del lancio di due dadi ideali ed indipendenti abbia somma 5.

In generale, possiamo definire la probabilità condizionale che $x=a$ quando $y=b$, che denoteremo $P(a,\underline{b})$ attraverso la relazione

$$P(a,b) = P(y=b)P(a,\underline{b}) = P(x=a)P(\underline{a},b) \tag{1}$$

Assunzioni sulla probabilità condizionale insieme alla regola di somma

$$P(x = a) = \sum_{b_i} P(a, b_i) = \sum_{b_i} P(a, \underline{b_i}) P(y = b_i)$$

permettono di risolvere problemi associati a variabili non osservate.
esempio: Si hanno due scatole: la scatola 1 contiene 1 palla bianca (b) e due nere (n), la scatola 2 contiene 2 palle bianche ed una nera.

1) Si scelga una scatola tra le due con eguale probabilità, qual è la probabilità di estrarre una palla bianca? Sia x la variabile casuale associata al colore della palla, y la variabile associata al numero della scatola, allora

$$P(x = b) = P(y = 1)P(b, \underline{1}) + P(y = 2)P(b, \underline{2}) = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{2}{3}$$

quindi la probabilità è $1/2$.
2) Avendo pescato la palla bianca, qual è la probabilità di averla presa dalla scatola 1? Dalla relazione (1)

$$P(\underline{b}, 1) = \frac{P(b, \underline{1})P(y = 1)}{P(x = b)} = \frac{(1/3) \cdot (1/2)}{1/2} = \frac{1}{3}$$

esercizio 3 Siano date due scatole. La scatola 1 contiene una palla bianca, la scatola 2 contiene una palla bianca ed una nera. In una estrazione casuale viene estratta una palla bianca. Qual è la probabilità che sia stata estratta dalla scatola 1? Qual è la probabilità che estraendo un'altra palla dalla stessa scatola questa sia ancora bianca?

2. Entropia

In statistica si definisce l'entropia di un insieme statistico

$$S = -\sum_{x \in \mathcal{A}} P(x) \ln P(x) \tag{2}$$

la base del logaritmo, non è essenziale, alterando la definizione solo per un fattore di scala, di seguito si userà sempre il logaritmo naturale.

Notiamo le seguenti proprietà dell'entropia:

- l'entropia è una funzione positiva ed è nulla solo se la variabile casuale può assumere un solo valore (è predeterminata).
- l'entropia assume un valore massimo per una distribuzione costante della probabilità su tutti i valori che la variabile casuale può assumere; nel caso che questi siano in numero N, l'entropia è massima se $P(x) = 1/N$ per ogni valore di x.
- l'entropia dell'insieme statistico congiunto XY di X e Y è uguale alla somma delle entropie $S(X)$ e $S(y)$ se le distribuzioni statistiche sono indipendenti.
- l'entropia ha una proprietà di composizione: dividiamo l'insieme $\mathcal{A}$ dei valori cha la variabile casuale x può assumere in gruppi disgiunti g_i †. La probabilità che un valore di x appartenga al gruppo g_i è $P(x \in g_i = \sum_{a \in g_i} P(a)$. Allora possiamo esprimere l'entropia dell'insieme $\mathcal{P}$ sulla base dell'entropia di ciascuna delle sue sottoparti:

$$S = -\sum_{a \in \mathcal{A}} P(a) \ln P(a) = \sum_{g_i} P(x \in g_i) \ln P(x \in g_i) + \sum_{g_i} P(x \in g_i) S(G_i)$$

† due insiemi $\mathcal{A}$ e $\mathcal{B}$ si dicono disgiunti se nessun elemento di $\mathcal{A}$ è elemento di $\mathcal{B}$.

esempio:

Sia dato un insieme statistico in cui la variabile casuale x può assumere 4 valori arbitrari con probabilità $(1/2, 1/6, 1/6, 1/6)$. L'entropia del sistema vale pertanto $S = -1/2 \ln 1/2 - 3/6 \ln 1/6 = 1/2 \ln 12$. Suddividiamo l'insieme nei due insiemi $(x, (a_1), P(x = a_1)$ e $x, (a_2, a_3, a_4), P(x = a_i | x \in (a_2, a_3, a_4) = 1/3$. L'entropia del primo è nulla, l'entropia del secondo vale $\ln 3$. Sulla base della proprietà di decomposizione possiamo calcolare l'entropia nella forma

$$S = -\frac{1}{2} \ln \frac{1}{2} - \frac{1}{2} \ln \frac{1}{2} - \frac{1}{2} \ln 3 = \frac{1}{2} \ln 12$$

esercizio 4 Calcolare l'entropia dell'insieme statistico di probabilità $P_i = (1/2)^i$.

3. Quando le probabilità degli eventi elementari sono uguali.

L'insieme statistico più comune, dal punto di vista modellistico è l'insieme statistico di probabilità uniforme $P_i = 1/N$ dove N è il numero dei valori che la variabile casuale può assumere.

E' una ammissione di ignoranza sulla complessità della realtà. Per esempio assumiamo fino a prova contraria che nel lancio di una moneta ciascuna delle facce risulti con uguale probabilità.

A volte, alcuni di questi eventi sono equivalenti sotto il profilo dell'osservatore. In questo caso è il numero totale degli eventi che verificano una data condizione, che ne determina la probabilità.

3.1. Permutazioni

Si chiama permutazione di un insieme ordinato un riordinamento degli elementi dell'insieme. Il numero delle permutazioni di un insieme di N elementi è detto fattoriale di N ed indicato in genere con $N!$. Esso ha la proprietà $N! = N * (N-1)!$.

Costruiamo per esempio le sei permutazioni dell'insieme di tre elementi numerati 1,2,3:
[123][132][213][231][312][321].

Possiamo costruire tutte le possibili permutazioni di quattro elementi numerati da 1 a 4 a partire dalle permutazioni di tre elementi. E' sufficiente considere che l'elemento 4 potrà trovarsi in 4 diverse posizioni:
[4???][?4??][??4?][???4]

dove con la notazione ??? è stata indicata una qualunque delle permutazioni degli elementi 1,2,3. Il numero di permutazioni di 4 elementi è pertanto $4 \cdot 6 = 24$.

esercizio 5 Si estraggano da una scatola palline numerate da 1 a 10 rimettendole nell'urna dopo averle pescate. Calcolare la probabilità che il risultato dell'estrazione sia 111, la probabilità che sia 123 e la probabilità che escano 1, 2 e 3 indipendentemente dal loro ordine.

Si chiamano disposizioni di n elementi presi tra N, indicate con D_N^n i sottoinsiemi ordinati che si possono ottenere selezionando n elementi da un insieme di N.

Per esempio, da un insieme di 3 elementi, numerati 1,2,3 possiamo ottenere le disposizioni [12][21][13][31][23][32].

E' facile comprendere che le disposizioni sono legate alle permutazioni:

$$\mathrm{D}_N^n = \frac{N!}{(N-n)!}$$

A volte non si è interessati all'ordinamento interno delle disposizioni; si chiamano combinazioni le disposizioni senza ordinamento interno. Il numero delle combinazioni, di solito indicato con la notazione $\binom{N}{n}$ è uguale al numero delle disposizioni diviso per il numero di modi in cui è possibile riordinare internamente una disposizione:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

esercizio 6 qual è la probabilità che i cinque numeri estratti fra cento siano 1,2,3,4,5?

esercizio 7 qual è la probabilità che di cinque numeri estratti fra cento tre siano 1,2,3?

esercizio 8 in quanti modi si possono disporre n oggetti distinti in N scatole, in modo che ogni scatola non contenga più di un oggetto?

esercizio 9 in quanti modi si possono disporre n oggetti indistinguibili in N scatole, senza limitazioni?

4. Soluzioni

1. la media statistica del quadrato delle facce di un dado a sei facce è data da

$$\frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2)=\frac{91}{6}$$

2. la probabilità che la somma di due dadi a sei facce ideali ed indipendenti sia 5 è data da:

$$P(1,4)+P(2,3)+P(3,2)+P(4,1)=4\cdot\frac{1}{36}$$

3. siano b ed n rispettivamente i risultati: il colore della palla è bianco e il colore della palla è nero. Allora: $P(b,\underline{1})P(1)=P(\underline{b},1)P(b)$, pertanto la probabilità che la palla bianca sia stata estratta dalla scatola 1 è:

$$P(1,\underline{b})=\frac{P(b,\underline{1})P(1)}{P(b)}=\frac{1\cdot 1/2}{3/4}=\frac{2}{3}$$

La probabilità che in una ulteriore estrazione si ripresenti la palla bianca sarà pertanto:

$$P(bb,\underline{b})=P(\underline{b},1)P(b,\underline{1})+P(\underline{b},2)P(b,\underline{2})=\frac{2}{3}+\frac{1}{3}\frac{1}{2}=\frac{5}{6}$$

dove ho indicato con bb la ripetizione dell'evento b.

4. notiamo innanzitutto che, come necessario $\sum_i P(i)=\sum_i 1/2^i=1$. Possiamo fattorizzare l'insieme delle probabilità P_i in

$$P_i=\frac{1}{2^i}=\begin{cases}\frac{1}{2} & i=1\\ \frac{1}{2}\frac{1}{2^{i-1}} & i>1\end{cases}$$

Per la proprietà di decomposizione dell'entropia, $S(P_i)=S(\frac{1}{2},\frac{1}{2})+\frac{1}{2}S(1)+\frac{1}{2}S(P_i)$. Da cui $S(P_i)=2S(\frac{1}{2},\frac{1}{2})=2\ln 2$.

5. la probabilità che esca tre volte 1 è data, in assenza di correlazione fra estrazioni successive da:

$$P(1,1,1)=P(1)P(1)P(1)=\frac{1}{1000}$$

Uguale è la probabilità dell'estrazione 1 2 3. La probabilità che escano 1, 2 e 3 indipendentemente dal loro ordine è invece data da:

$$P(\overline{1,2,3})=P(1,2,3)\cdot 3!=\frac{6}{1000}$$

dove ho indicato con $\overline{1,2,3}$ una qualunque permutazione di $1,2,3$.

6. la probabilità che i cinque numeri estratti dalla serie 1:100 siano 1,2,3,4,5 è data da:

$$P(\overline{1,2,3,4,5}) = \frac{1}{\binom{100}{5}} = \frac{5!95!}{100!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5}{100 \cdot 99 \cdot 98 \cdot 97 \cdot 96}$$

7. la probabilità che dei cinque numeri estratti nelle condizioni di cui sopra, tre siano 1,2,3 è data dalla probabilità che 1,2,3 siano i primi estratti moltiplicato per in numero di modi in cui è possibile permutare i 5 numeri estratti:

$$P(\overline{1,2,3,X,Y}) = P(1,2,3) \cdot 5! = \frac{5! \cdot 3!}{100!}$$

8. in tanti quanti sono i modi di prendere n oggetti indistinguibili (scatole) tra N, pertanto in $\binom{N}{n}$ modi.

9. possiamo rappresentare le N scatole con $N-1$ tratti verticali. Per esempio la notazione $oo|o||o$ significa 2 oggetti nella prima scatola, 1 nella seconda, nessuno nella terza, 1 nella quarta. Il numero di disposizioni degli n oggetti nelle N scatole è pertanto uguale al modo di permutare $N-1+n$ oggetti tenendo conto che permutazioni interne delle $N-1$ barre o degli n oggetti non cambiano il significato della rappresentazione, ovvero $\frac{(N+n-1)!}{(N_1)!n!}$