CAPITOLO IX – RACCOLTA STATISTICA DI DATI

9.1 – Variabili aleatorie; definizione e prime proprietà.

Si consideri un esperimento casuale S ed il suo spazio dei campioni Ω , formato da tutti i possibili risultati, ω , di S. Si ricorderà che i sottoinsiemi di Ω , o almeno parte di essi, sono gli *eventi*; tra gli eventi ci sono naturalmente anche quelli formati da un solo elemento, $\{\omega\}$, detti eventi elementari. In Ω sia stata introdotta una funzione di probabilità P.

Diremo *variabile aleatoria* una qualsiasi funzione reale definita sugli elementi di Ω , ossia una qualsiasi legge di associazione tra i campioni $\omega \in \Omega$ ed i numeri reali. Distinguiamo immediatamente tra variabili aleatorie discrete, che possono assumere i loro valori solamente in uno spettro discreto di possibilità, e variabili aleatorie continue, che possono assumere qualsiasi valore in un intervallo, finito o infinito, della retta dei numeri reali. In questo secondo caso i valori della variabile non possono essere finiti.

Di norma si indica una v*ariabile aleatoria* (nel seguito, per brevità, v.a.) con una lettera maiuscola, in genere la X, ed i suoi valori con la medesima lettera al minuscolo, valori che, nel caso discreto, saranno contraddistinti con un indice: dunque, se X è una v.a. discreta che ammette n possibili differenti valori, questi saranno indicati con x_i, dove $i = 1,2,\ldots,n$; se X indica una v.a. continua, il suo generico valore sarà indicato semplicemente con x, compreso in un intervallo $(a,b) \subseteq R$. I valori di una v.a. discreta vengono ordinati dal più piccolo al più grande, in modo che, se $i < j$, si abbia di conseguenza $x_i < x_j$.

Ci serviremo, almeno all'inizio, delle v.a. discrete, certamente più semplici da trattare, per introdurre alcune definizioni che rimarranno valide anche nel caso delle v.a. continue.

La probabilità introdotta nello spazio dei campioni, indicata con $P(E)$ per il generico evento E, viene usata per attribuire una *funzione (o distribuzione)* f *di probabilità* anche ai valori della v.a., e dunque alla v.a. stessa. Diremo, in modo del tutto naturale, che, se x_i indica il generico valore della v.a. discreta X, la sua *probabilità*, indicata come $f(x_i)$, è data dalla probabilità attribuita all'evento $E_i \subseteq \Omega$, costituito dalla totalità dei campioni $\omega \in \Omega$ per i quali $X(\omega) = x_i$: $E_i \equiv \{\omega \in \Omega : X(\omega) = x_i\}$; dunque la probabilità di x_i, $f(x_i)$, sarà:

$$f(x_i) \equiv P(E_i) \equiv P(\{\omega \in \Omega : X(\omega) = x_i\}),$$

o, più brevemente, $f(x_i) \equiv P(X(\omega) = x_i)$.

L'assegnazione della funzione di probabilità alla v.a. X consiste dunque nelle due operazioni seguenti:

- individuazione degli eventi per i quali la v.a. assume il generico valore x_i, già indicati con E_i (e questo per tutti i valori x_i di X);
- calcolo della probabilità di x_i come probabilità di E_i, $P(\{\omega : X(\omega) = x_i\})$.

Possiamo osservare che gli eventi E_i per loro stessa definizione, godono di due notevoli proprietà. La prima di queste si basa sul fatto che la v.a. è definita sull'intero spazio dei campioni, ossia assume un preciso valore per $\forall\, \omega \in \Omega$; ne consegue che la somma logica estesa a tutti questi eventi deve coincidere con Ω :

$$\bigcup_i E_i = \Omega \ .$$

Infatti, se così non fosse, dovrebbe essere $\bigcup_i E_i \subset \Omega$, e dunque dovrebbe esistere (almeno) un campione $\overline{\omega} \in \Omega$ che non appartiene alla somma logica posta a primo membro: tale campione non apparterrebbe a nessuno degli eventi E_i, e di conseguenza in corrispondenza ad esso la v.a. non sarebbe definita, contrariamente alla nostra ipotesi. Inoltre, gli eventi E_i sono a due a due incompatibili, dal momento che ad ogni possibile scelta di $\omega \in \Omega$ rimane associato uno ed un sol valore della v.a.: dunque

$$E_i \wedge E_j = \varnothing \text{ se } i \neq j ,$$

che rappresenta la seconda delle due proprietà preannunciate. Quanto ora detto si esprime con l'affermazione che la famiglia di eventi E_i costituisce una *classe completa di eventi incompatibili*: la completezza è garantita dalla prima proprietà, l'incompatibilità dalla seconda.

Per le v.a. discrete la probabilità $f(x_i)$ del valore x_i viene indicata di norma più brevemente come p_i. Essa deve sempre soddisfare le condizioni:

$$p_i \equiv f(x_i) \geq 0, \qquad \sum_{i=1}^{n} f(x_i) = \sum_{i=1}^{n} p_i = 1,$$

dove la somma va estesa a tutti gli n valori della v.a. discreta (anche se infiniti). La prima condizione risulta nel nostro caso certamente soddisfatta, dal momento che ci si è serviti della probabilità P già introdotta nello spazio dei campioni; la seconda condizione esprime il fatto che ad ogni ripetizione dell'esperimento casuale S che sta alla base di tutto il discorso, la v.a. assume uno (ed uno solo) dei suoi valori. Infatti, dal momento che $\Omega = \bigcup_i E_i$, deve essere

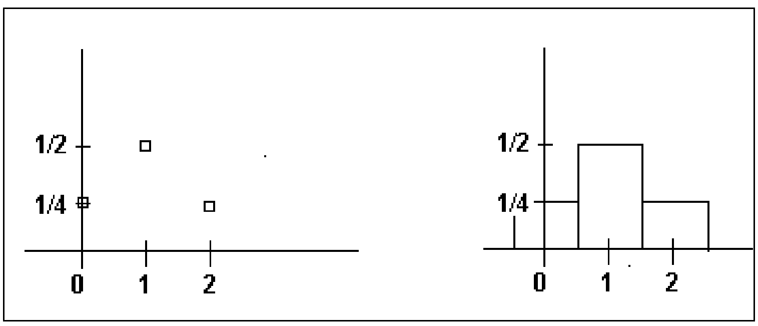
$$P(\Omega) = 1 = P\left(\bigcup_i E_i\right) = \sum_i P(E_i) = \sum_i p_i .$$

Come primo semplice esempio di v.a. discreta X , consideriamo l'esperimento casuale S consistente nel lancio simultaneo di due monete, il cui spazio dei campioni Ω è composto da quattro elementi, $\Omega = \{(TT),(TC),(CT),(CC)\}$. Come valore della v.a. (discreta) consideriamo il numero di teste che si ottengono in ogni possibile risultato dell'esperimento; abbiamo dunque una v.a. i cui valori sono costituiti dai tre numeri reali $0,1,2$; nella notazione proposta scriviamo $n = 3$, $x_1 = 0, x_2 = 1, x_3 = 2$. Si devono ora individuare gli eventi E_i per i quali la v.a. assume il valore x_i; in corrispondenza al primo valore, $x_1 = 0$, otteniamo l'evento $E_1 = \{\omega \in \Omega : X(\omega) = 0\} = \{CC)\}$; in corrispondenza al secondo, $x_2 = 1$, otteniamo l'evento $E_2 = \{\omega \in \Omega : X(\omega) = 1\} = \{CT,TC)\}$, infine, per il terzo valore, $x_3 = 2$, otteniamo l'evento $E_3 = \{\omega \in \Omega : X(\omega) = 2\} = \{TT)\}$. Concludiamo che la probabilità di x_1 è $f(x_1) = p_1 = 1/4$, quella di x_2 è $p_2 = 1/2$, e quella di x_3 è $p_3 = 1/4$. Sono evidentemente verificate le due condizioni richieste. Possiamo costruire la tabella

$$\begin{array}{lll}
x_1 = 0 & E_1 = \{(C,C\} & p_1 = P(E_1) = 1/4 \\
x_2 = 1 & E_2 = \{(T,C),(C,T)\} & p_2 = P(E_2) = 1/2 \\
x_3 = 2 & E_3 = \{(T,T\} & p_3 = P(E_3) = 1/4 .
\end{array}$$

I valori della generica v.a. (discreta) X vengono solitamente riportati in una tabella del tipo di quella riprodotta qui a lato, costruita con i valori ottenuti nell'esempio proposto: gli stessi valori vengono anche indicati in grafico o in istogramma, come illustrato sotto. Si noti come le barre dell'istogramma vengano centrate nel valore x_i al quale si riferiscono.

X=	0	1	2
p=	$1/4$	$1/2$	$1/4$



Per le v.a., sia discrete che continue, e in particolare per queste ultime, si considera anche la *funzione di distribuzione cumulativa* o *funzione di distribuzione* *di probabilità* $F(x_i)$, che rappresenta la probabilità che la v.a. X assuma un valore non superiore ad un prefissato valore x_i; si tratta, in analogia a quanto fatto per la definizione della probabilità, di determinare, per ogni valore x_i, l'evento E_i, costituito ora dalla totalità dei campioni per i quali $X(\omega) \le x_i$, cioè $E_i \equiv \{\omega \in \Omega : X(\omega) \le x_i\}$; dunque ne concludiamo chela probabilità cumulativa di x_i sarà: $F(x_i) \equiv P(E_i) \equiv P(\{\omega \in \Omega : X(\omega) \le x_i\})$, o più semplicemente, $F(x_i) = P(X(\omega) \le x_i)$.

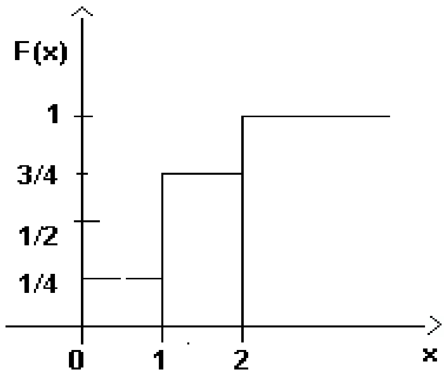
Si vede immediatamente come il valore della probabilità cumulativa si ricolleghi alla probabilità dei singoli valori x_i della v.a. discreta. Abbiamo, ovviamente, la relazione

$$F(x_i) = \sum_{k=1}^{i} f(x_k) = \sum_{x_k \le x_i} f(x_k).$$

Dalla sua stessa definizione discende che la funzione di probabilità cumulativa è una funzione non decrescente con valori non negativi, mai superiore ad uno, valore che raggiunge in corrispondenza al più grande dei valori della v.a.

La tabella seguente riproduce la situazione generata dall'esempio proposto, usato anche per il grafico successivo; abbiamo (si faccia attenzione alle disuguaglianze deboli e forti)

$x < 0$	$0 \le x < 1$	$1 \le x < 2$	$x \ge 2$
$F(x) = 0$	$F(x) = 0.25$	$F(x) = 0.75$	$F(x) = 1$.



Per le v.a. continue non è pensabile di introdurre una funzione di probabilità che misuri la probabilità del verificarsi *esattamente* di un risultato da scegliere in un insieme continuo: tale probabilità sarebbe evidentemente nulla, poiché dovremmo verificare l'eguaglianza delle infinite cifre decimali presenti nel numero che esprime il risultato atteso con le infinite cifre decimali presenti nel numero che esprime il risultato ottenuto. Sarà invece possibile parlare della probabilità che il risultato sia compreso in un intervallo di valori, anche se molto ristretto. Per esempio, se l'esperimento casuale S consiste nel lancio di una freccia su di un bersaglio, la probabilità di colpire un punto che disti *esattamente* 1 cm dal centro è da considerarsi nulla; si può invece valutare la probabilità che la distanza tra il punto colpito ed il centro del bersaglio sia compresa tra 1.0 e 1.5 cm, o tra 1.0 e 1.1 cm, o anche in intervalli più piccoli, comunque di ampiezza finita, non nulla. Tale probabilità dipende sia dal punto iniziale (distanza di 1 cm dal centro) che dall'ampiezza dello intervallo, e cioè dalla precisione richiesta: intendiamo così che la probabilità di colpire una corona circolare centrata nel bersaglio dipende tanto dal raggio minore di tale corona, ossia dalla distanza minima dal bersaglio, quanto dall'ampiezza della corona stessa.

Per le v.a. continue viene introdotta una nuova funzione, anch'essa indicata con $f(x)$, detta *densità di probabilità*, assegnata in modo che il prodotto $f(x)\ \Delta x$ rappresenti la probabilità dell'evento $E(x)$ [1], costituito da tutti e soli i campioni per i quali il valore assunto dalla v.a. risulta compreso nell'intervallo $(x, x+\Delta x)$: $E(x) \equiv \{\omega \in \Omega : x \leq X(\omega) \leq x+\Delta x\}$; quindi

$$f(x)\,\Delta x \equiv P(E(x)) \equiv P((\omega \in \Omega : x \leq X(\omega) \leq x+\Delta x)) \; ,$$

o, analogamente, $f(x)\,\Delta x \equiv P(\{x \leq X(\omega) \leq x+\Delta x\})$. Nel caso in esame scrivere disuguaglianze in senso forte o in senso debole è la stessa cosa, in quanto l'unica differenza tra i due casi sarebbe rappresentata dalla probabilità che si abbia esattamente $X(\omega) = x$ (o analogamente $X(\omega) = x+\Delta x$), evidentemente nulla.

La relazione precedente mostra come la misura della probabilità, $f(x)\,\Delta x$, sia quella dell'area di un rettangolo con base l'intervallo dell'asse reale $(x, x+\Delta x)$, ed altezza $f(x)$, ossia il valore della densità di probabilità, ritenuta costante nell'intervallo. Questa approssimazione della funzione densità con un suo valore è tanto più accettabile quanto più piccolo è l'intervallo Δx. L'intero asse reale, o meglio quella sua porzione, generalmente un intervallo (a,b), nella quale $f(x)0$, può venire immaginata come successione di n intervallini consecutivi, i cui estremi siano indicati come $x_0 = a, x_1, x_2, ..., x_k, ..., x_n = b$, ognuno di ampiezza $\Delta x_k = x_k - x_{k-1}$ (le ampiezze, come si vede, non devono essere necessariamente uguali). In ognuno di essi la funzione densità può ritenersi costante (sempre che gli intervallini siano molto piccoli), scrivendo per esempio $f(x) = f(x_k)$ per $x_{k-1} < x < x_k$ (abbiamo scelto come valore della funzione nel $k-esimo$ intervallo il valore che essa assume nell'estremo superiore dell'intervallo stesso; altre scelte sarebbero ugualmente lecite in quanto comporterebbero differenze non apprezzabili). Allora, la probabilità che la v.a. assuma un valore nel $k-esimo$ intervallo, $I_k = (i_{k-1}, i_k)$, è data dall'area del rettangolo sovrastante l'intervallo:

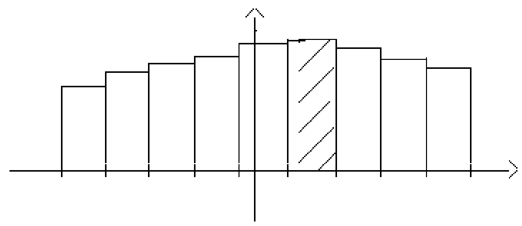
---

[1] Si osservi come, per quanto appena detto, sarebbe più corretto scrivere $E(x, \Delta x)$.

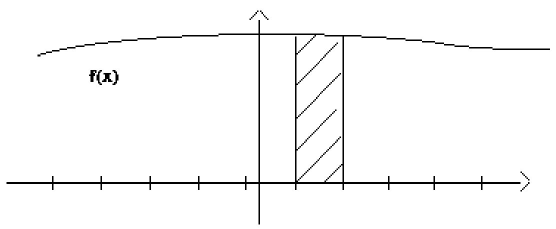
$$P(\{x_{k-1} \le X(\omega) \le x_k\}) = f(x_k)(x_k - x_{k-1}).$$

La funzione densità di probabilità $f(x)$ gode delle seguenti proprietà, analoghe a quelle già mostrate per la probabilità delle v.a. discrete:

- di non essere mai negativa, $f(x) \ge 0$;
- di essere tale che $\sum_k f(x_k)(x_k - x_{k-1}) = 1$.



Quello illustrato in questa sede è un processo di discretizzazione della v.a. continua che introduce un notevole grado di approssimazione, che diventa però sempre minore al crescere del numero degli intervallini, con conseguente diminuzione della loro ampiezza. Contemporaneamente la spezzata alla quale si è ridotto il grafico della funzione densità approssima sempre meglio la curva effettiva. Possiamo sperare che l'approssimazione scompaia quando il numero di intervallini considerati diviene infinitamente elevato, dunque la loro ampiezza sempre minore. I più esperti sanno che in questo caso la sommatoria precedente diviene una somma integrale, e



la seconda proprietà illustrata si riscrive come $\int_{-\infty}^{+\infty} f(x)dx = 1$ (leggi: integrale da meno infinito a più infinito di effe di ics in di ics).

In ogni caso, la conoscenza della funzione di densità di probabilità permette di valutare la probabilità cumulativa, cioè la probabilità che il valore della v.a. non sia maggiore del generico valore a, come

$$F(a) = P(X(\omega) \le a) = \sum_k f(x_k)(x_k - x_{k-1}),$$

dove la somma va estesa ai soli intervalli che si trovano alla sinistra del punto a, cioè a tutti gli intervalli per i quali $x_k \le a$. Di conseguenza valutiamo facilmente anche la probabilità che il valore della v.a. sia compreso nel generico intervallo (a,b), come differenza tra la probabilità che il valore x non sia maggiore di b e la probabilità che x non sia maggiore di a:

$$P(a \le x \le b) = F(b) - F(a).$$

La differenza delle due somme che compaiono a membro destro della relazione precedente si riduce ad un'unica somma, limitata ai rettangoli che hanno base compresa nell'intervallo (a,b).

Le affermazioni precedenti sono in completo accordo con l'esigenza già espressa che la probabilità di ottenere esattamente un generico valore a sia nulla; infatti, la probabilità che sia esattamente $X(\omega) = a$ si può scrivere come $P(X(\omega) = a) = P(a \le X(\omega) \le a)$, e dunque

$$P(X = a) = P(a \le X(\omega) \le a) = F(a) - F(a) = 0.$$

Quanto ora detto a proposito delle v.a. continue dovrebbe avere convinto Chi legge della decisiva importanza della conoscenza della funzione densità di probabilità; il problema sarà proprio quello di determinare, o meglio, di ipotizzare il suo andamento, come vedremo in un paragrafo successivo.

Osserviamo ancora, prima di concludere questo argomento, che:

- se X è una v.a., discreta o continua, e c una costante, anche cX è una v.a.;
- se X_1 e X_2 sono due v.a., sia la somma, $X = X_1 + X_2$, che il prodotto, $X = X_1 \cdot X_2$ delle due sono a loro volta v.a., discrete o continue a seconda che siano discrete o continue le v.a. X_1 ed X_2;
- più in generale, se X è una v.a., una qualsiasi funzione di questa, $Y = \phi(X)$, è a sua volta una v.a., non necessariamente discreta o continua a seconda che sia discreta o continua la v.a. X (si pensi alla variabile aleatoria continua X statura degli elementi di una popolazione, ed alla sua funzione $\phi(X) = INT(X)$, parte intera di X, discreta).

Vediamo degli esempi di quanto detto sopra. Se la variabile aleatoria X conta le teste nel lancio di due monete, la variabile aleatoria $2X$ conta le stesse teste, e ne raddoppia il numero: anche questa è evidentemente una v.a.. Allo stesso modo, è una v.a. anche quella che, oltre a contare le teste, ne calcola successivamente il quadrato, o considera tale numero come argomento di una qualsiasi funzione.

Infine, se X_1 e X_2 sono due v.a. consistenti entrambe nel lancio di un dado, la somma $X = X_1 + X_2$ è anch'essa una v.a. i cui valori sono rappresentati dalla somma dei numeri mostrati da ognuno dei due dadi, cioè dei valori delle singole v.a.

9.2 – Popolazioni di individui e campioni.

Tra i casi più tipici di v.a. continue, e comunque per noi i più interessanti, abbiamo quelli per i quali l'esperimento casuale che ne sta alla base è costituito da singole misure, per esempio effettuate su persone, quali la misura dell'altezza, del peso corporeo, etc. Si faccia attenzione al fatto che nell'esempio proposto la richiesta *casualità* dell'esperimento consiste non nella misura in se stessa (infatti misure ripetute in rapida successione sullo stesso individuo dovrebbero dare sempre lo stesso risultato, per lo meno nel limite dell'errore strumentale), ma *nella scelta dell'individuo da misurare*, in quanto la statura o altra caratteristica sarà differente da individuo a individuo. Si riconosce allora che lo spazio dei campioni è costituito dai singoli *individui*, i quali formano una *popolazione*: la *popolazione* (qui in luogo dello spazio dei campioni) è quindi un insieme di *individui* dotati di proprietà *osservabili* (cioè *misurabili*) comuni, che sono appunto v.a. continue: per esempio la popolazione degli esseri umani, di ciascuno dei quali si può misurare la statura, il peso ecc. Diremo ancora *campione* un qualsiasi sottoinsieme della popolazione, formato da un numero n di individui di questa, con $n \leq N$, se N è il numero di individui della popolazione. Nel concetto di popolazione si ritrova l'ambiguità segnalata a suo tempo per la determinazione dell'insieme universo della teoria degli insiemi. Infatti, parlando di una popolazione di individui, possiamo intendere l'intero genere umano, o i soli individui maschi europei, o i soli individui che presentino una determinata patologia, e così di seguito. Anche in questo caso tuttavia non va esagerata la preoccupazione per i problemi o gli equivoci che ne potrebbero derivare.

Riassumendo, l'esperimento casuale consiste nella scelta preliminare dell'individuo del quale vogliamo misurare una caratteristica osservabile, e tale caratteristica, spesso detta *carattere*, rappresenta una v.a. continua. L'esperimento 'scelta' può dare tanti *risultati* quanti sono gli individui della popolazione, ognuno considerato con una medesima probabilità, $1/\mathrm{N}$ se N sono gli individui presenti nella popolazione (questo vuol dire che non c'è ragione di dare maggior peso alla scelta di un individuo piuttosto che alla scelta di un altro, ossia che gli individui, e dunque i singoli elementi dello spazio dei campioni, sono *equiprobabili*).

9.3 – Speranza matematica e varianza.

Per le distribuzioni di probabilità delle v.a. vengono definiti alcuni parametri dalla cui conoscenza deve essere possibile ricavare le principali caratteristiche delle distribuzioni. In questa sede diamo le definizioni di alcuni di tali parametri considerando v.a. discrete, ma precisando che le definizioni stesse si estenderanno, con le necessarie modifiche formali, alle v.a continue.

Consideriamo dunque una v.a. discreta X, i cui valori x_i, con $= 1, 2, ..., L$, vengano assunti con frequenza relativa p_i: intendiamo dire con questo che su N tentativi, il generico valore x_i si è ripetuto con frequenza n_i, e dunque con frequenza relativa $p_i = n_i / N$. Naturalmente, deve valere $\sum_{i=1}^{L} n_i = N$, da cui $\sum_{i=1}^{L} \frac{n_i}{N} = \sum_{i=1}^{L} p_i = 1$.

Viene introdotto un primo parametro allo scopo di verificare se i valori della v.a. si addensino o meno attorno ad un valore, detto *valore centrale*, che non deve necessariamente essere uno dei valori della v.a.. Questa verifica si può fare in più modi, scegliendo differenti parametri, detti tutti *indicatori di tendenza centrale*. Daremo la precedenza tra di essi alla così detta *speranza matematica* o *valor medio*, definita semplicemente come media aritmetica di tutti i singoli valori, cosa del resto del tutto naturale. Si faccia bene attenzione che tale media va effettuata tenendo conto di tutti i risultati ottenuti negli N tentativi effettuati, e non dei soli valori possibili per la v.a. In altre parole, se per esempio la v.a. potesse assumere solo tre valori, $L = 3$, che indichiamo con $x_1 = 1$, $x_2 = 2$ e $x_3 = 3$, non avrebbe alcun senso procedere al calcolo della loro media $(x_1 + x_2 + x_3)/3 = (1 + 2 + 3)/3 = 2$! Infatti, se nell'effettuazione di $N = 30$ prove, avessimo ottenuto cinque volte il primo, cinque volte il secondo e venti volte il terzo, avremmo dovuto tenere conto di tali frequenze calcolando $\frac{5 \cdot 1 + 5 \cdot 2 + 20 \cdot 3}{30} = 2.5$, e non la media precedente che portava al 2. Dunque, la media sui valori della v.a. è una *media pesata sulle loro frequenze relative*.

La scelta della speranza matematica come indicatore di tendenza centrale non è l'unica possibile, e non è esente da critiche, quale quella, importante, di essere sensibile ai così detti valori *fuori norma*. Infatti una eventuale misura erronea che si differenzi di molto dagli altri valori ottenuti potrebbe spostare, anche notevolmente, il valore medio dal suo valore reale, e cioè quello che si otterrebbe in assenza del dato anomalo, specie se i dati a disposizione sono pochi. In certi casi si preferisce, prima di eseguire il calcolo della media, eliminare il risultato maggiore e quello minore, ma evidentemente nemmeno questa è una garanzia, dal momento che i dati fuori norma potrebbero essere più di uno.

Questo problema potrebbe essere risolto sostituendo alla speranza matematica un altro indicatore di tendenza centrale, per esempio la così detta *mediana*, ossia un numero caratterizzato dal fatto che almeno il 50% dei valori della v.a. stanno alla sua sinistra, ed un numero non inferiore al 50% degli stessi valori stanno alla sua destra. Si badi che anche in questo caso i valori devono venire pesati dalla loro frequenza. Nell'esempio precedente, se considerassimo i soli tre valori $x_1 = 1$, $x_2 = 2$ e $x_3 = 3$, vedremmo che due valori sui tre ammessi non sono maggiori di 2, ed ancora due su tre non sono inferiori a 2: dovremmo allora concludere che tale numero ha alla sua sinistra non meno del 50% dei valori, e lo stesso alla sua destra, e rappresenta quindi la mediana. Al contrario, dovremmo correttamente osservare che su trenta prove, solo dieci non hanno dato esito maggiore di due, e venti non hanno dato esito inferiore a 2: dunque la mediana non sarà 2 ma 3, dal momento che tutti i risultati ottenuti non sono maggiori di 3, e venti su trenta, ossia il 66.67%, non sono inferiori a tre.

Un altro parametro ancora potrebbe essere la *moda*, ossia il valore che si presenta con maggiore frequenza, e dunque con maggiore probabilità. Anche queste due definizioni presentano inconvenienti, tanto da consigliarci, come detto, di scegliere comunque come parametro per la determinazione del valore centrale la media dei valori della v.a., ossia la speranza matematica.

Prima di proseguire, collegandoci alla definizione di mediana, citiamo quella di percentile, con particolare riguardo al venticinquesimo percentile. E' questo un qualsiasi numero che abbia alla sua sinistra non meno del 25% dei valori della v.a., e alla sua destra non meno del 75% di questi. Si vede dunque che la mediana è il cinquantesimo percentile, e che almeno il 50% dei valori di una v.a. sono contenuti tra il venticinquesimo ed il settantacinquesimo percentile. In medicina, assumono spesso grande importanza il decimo ed il novantesimo percentile, prima e dopo dei quali si tende a riconoscere stati patologici.

La semplice determinazione del valore centrale non è comunque sufficiente a permetterci di verificare se i valori della v.a. si addensano o meno attorno ad esso; è evidente infatti che sarà sempre possibile calcolare la media aritmetica, anche se i valori della v.a. risultano, tutti o in parte, significativamente discosti da essa, e dunque indipendentemente dal significato che le si può attribuire. Ad esempio, nel lancio di un dado si può calcolare il valore medio dei risultati possibili, ottenendo 3.5, ma sarebbe ben difficile attribuirgli un qualche significato particolare, meno che mai quello di valore più probabile. Ci occorre dunque un altro parametro, che dia un'idea dell'addensarsi o meno dei valori della v.a. attorno al valore centrale. Un tale parametro sarà necessariamente legato agli *scarti*, e cioè alla differenza di ogni singolo valore dal valore medio; vedremo comunque nel seguito come tali scarti andranno usati.

Per una v.a discreta X che possa assumere n differenti valori, definiamo come *speranza matematica* o *valor medio* la quantità $E(X) = \mu = \sum_{k=1}^{n} p_k x_k$, se p_k indica la probabilità, o la frequenza relativa, del risultato x_k.

Ne consegue che se, in generale, $Y = \Phi(X)$ è una v.a. discreta, funzione biunivoca[2] della v.a X precedente, che ha p_k come probabilità del suo valore x_k, si definisce come valor medio di Y la quantità $E(Y) = E(\Phi(X)) = \sum_{k=1}^{n} \phi(x_k) p_k$. Infatti, il singolo valore y_k della v.a. Y, ottenuto come $y_k = \Phi(x_k)$, ha la medesima probabilità di x_k, dal quale è stato calcolato, e dunque il valor medio di questa nuova v.a. è quello indicato.

Per quanto riguarda la definizione di un parametro che controlli la dispersione dei valori della v.a., come già accennato questo dovrà legarsi allo scarto del singolo valore x_k dal valore centrale, cioè alla differenza $x_k - E(X)$. Anche questo scarto è il valore di una nuova v.a., $Y = X - E(X)$, differenza tra il valore assunto dalla v.a. originale ed il suo valore medio, costante, con valori $y_k = x_k - \mu$. Ci si aspetta che la dispersione dei valori di X sia tanto maggiore quanto maggiori sono gli scarti, ma questa osservazione non deve suggerire di assumere come parametro stimatore della dispersione di X il valore medio di Y. Tale conclusione sarebbe infatti errata, in quanto, per la stessa natura del valore medio, il valore medio degli scarti risulta nullo. Infatti, la probabilità dei valori y_k degli scarti è la medesima, p_k, del corrispondente valore x_k di X, e lo scarto medio è

$$E(Y) = \sum_k p_k y_k = \sum_k p_k (x_k - \mu) = \sum_k p_k x_k - \mu \sum_k p_k = \mu - \mu = 0.$$

La causa di ciò risiede nel fatto che, mediando gli scarti, non si sono distinti scarti positivi da scarti negativi, che nella media si compensano annullandola. Occorre far sì che gli scarti diano contributi che si sommino sempre, eliminando in qualche modo la differenza di segno, cosa che consente sostanzialmente due scelte: o consideriamo il valore assoluto dello scarto, e quindi la sua media

[2] Questa proprietà viene richiesta in questa sede per non appesantire ulteriormente il discorso; si potrebbe vedere come modificare quanto sopra quando tale proprietà venisse a mancare.

$$\sum_k p_k |x_k - \mu|,$$

o consideriamo gli scarti elevati al quadrato, e quindi la media di tali *scarti quadratici*, detta *varianza* ed indicata con σ_X^2 o con $VAR(X)$:

$$\sigma_X^2 = \sum p_k (x_k - \mu)^2 .$$

Le due scelte presentano, al solito, vantaggi e svantaggi. Per esempio la prima fa ricorso al valore assoluto, funzione continua ma non derivabile negli zeri del suo argomento. La seconda, passando dagli scarti dei valori della v.a. $x_k - \mu$ ai loro quadrati, cambia la dimensione fisica di questi: se p.e. i valori x_k sono lunghezze, espresse in centimetri, la varianza sarà espressa in centimetri quadrati, e questo renderà impossibile un collegamento diretto tra speranza matematica e varianza. Si preferisce comunque scegliere questa seconda strada, e dunque adottare quale indicatore della dispersione dei valori della v.a. la varianza, anche a costo di doverne successivamente estrarre la radice.

Dalla definizione di varianza, $\sigma_X^2 = \sum p_k (x_k - \mu)^2$, discende che questa non potrà mai essere negativa, in quanto somma di termini non negativi. Un suo valore elevato indicherà una elevata dispersione dei valori della v.a. rispetto al loro valore medio; una varianza nulla al contrario starebbe ad indicare la coincidenza di tutti i valori della v.a. con il valor medio stesso, e dunque la costanza della v.a. in esame. Si noti come nella definizione di varianza entri il valore medio, il che significa che essa stessa deve essere assunta in coppia con questo, e non altri, indicatori di tendenza centrale: se come parametro di tendenza centrale scegliessimo per esempio la mediana, la definizione di varianza dovrebbe venire modificata per tenere conto di tale scelta (comunque lecita).

Concludiamo questo argomento elencando alcune delle principali proprietà sia del valor medio che della varianza, valevoli per v.a. discrete e continue. Cominciamo con le proprietà del valor medio.

1. Se X è una v.a. costante, nel senso che $X(\omega) = c$ per $\forall\, \omega \in \Omega$, $E(X) = \mu = c$; la dimostrazione è immediata ove si applichi la definizione di valore medio.
2. Sia X una v.a.; cX con $c \in R$ qualsiasi, è a sua volta una v.a.; se $E(X)$ è il valor medio di X, $cE(X)$ è il valor medio della v.a. cX: $E(cX) = cE(X)$; anche la dimostrazione di questa affermazione è immediata.

3. Sia X una v.a. somma di due v.a. X_1 ed X_2; se $E(X_1)$ e $E(X_2)$ indicano rispettivamente i valori medi di X_1 e di X_2, il valore medio di X è semplicemente la somma dei valori medi: $E(X_1 + X_2) = E(X_1) + E(X_2)$; la facile dimostrazione al Lettore.

4. Sia X una v.a. prodotto di due v.a. X_1 e X_2; se $E(X_1)$ e $E(X_2)$ indicano i valori medi rispettivamente di X_1 e di X_2, e se le v.a. X_1 e X_2 sono indipendenti (nel senso che i valori assunti dall'una non hanno alcuna influenza sui valori assunti dall'altra), il valore medio di X è il prodotto dei valori medi delle due variabili a fattore: ossia, e sempre che le v.a. siano <u>stocasticamente indipendenti</u>, $E(X_1 X_2) = E(X_1)E(X_2)$.

Molto spesso, le proprietà 2ª e 3ª si riassumono nella scrittura

$$E(\lambda X_1 + \mu X_2) = \lambda E(X_1) + \mu E(X_2),$$

il che ci dimostra che il calcolo del valor medio è una operazione lineare: il valor medio di una combinazione lineare è la combinazione lineare dei valori medi.

Vediamo ora alcune proprietà della varianza.

1. Dalla definizione di varianza discende che, per qualsiasi v.a. X, la varianza si può ottenere come differenza del valor medio del quadrato di X con il quadrato del valor medio della stessa: $\sigma_X^2 = E(X^2) - E(X)^2$. L'importanza di questo risultato è tale da far sì che qualche volta venga proposto come definizione di varianza, in luogo di quella data in queste pagine. Ne ricaviamo che il valore medio del quadrato di una qualsiasi v.a. deve essere maggiore del quadrato del valore medio, dal momento che la differenza tra questi due rappresenta la varianza, non negativa per sua stessa definizione.

2. Sia X una v.a. (discreta o continua) costante, che ammetta cioè un solo valore: $X(\omega) = c$ per $\forall \omega \in \Omega$: la sua varianza è nulla: $VAR(X) = 0$. Poiché infatti la speranza matematica coincide con c, unico valore assunto dalla v.a., $E(X) = c$, scriviamo direttamente la definizione di varianza, $\sigma_X^2 = \sum_k p_k (c - E(X))^2 = \sum_k p_k (c - c)^2 = 0$.

3. Sia X una v.a., cX è a sua volta una v.a.; se σ_X^2 è la varianza di X, $c^2\sigma_X^2$ è la varianza di cX. La dimostrazione è immediata; infatti, dallo scarto di X, $x_k - E(X)$, si passa allo

scarto di cX, $cx_k - E(cX) = cx_k - cE(X) = c(x_k - E(X))$, e dunque, passando allo scarto elevato al quadrato, abbiamo il fattore c^2.

4. Siano X_1 ed X_2 due v.a. stocasticamente indipendenti (dunque il valor medio del loro prodotto si ottiene come prodotto dei loro valori medi, come in ogni caso il valor medio della loro somma è la somma dei loro valori medi); vale allora $\sigma^2_{X_1+X_2} = \sigma^2_{X_1} + \sigma^2_{X_2}$.

5. Allo stesso modo $\sigma^2_{X_1-X_2} = \sigma^2_{X_1} + \sigma^2_{X_2}$. Si badi a non scrivere la differenza delle varianze in luogo della loro somma: una simile affermazione sarebbe palesemente assurda, in quanto nulla esclude che tale differenza possa assumere anche valori negativi e non potrebbe in alcun modo rappresentare una varianza. Inoltre, si consideri che la varianza rappresenta la dispersione dei valori delle singole v.a., e rappresenta dunque un grado di incertezza; non è pensabile che il passaggio dalle singole v.a. alla loro somma diminuisca questo grado di incertezza: evidentemente esso potrà solamente aumentare.

Riprendiamo infine in esame quella che è stata presentata come prima proprietà della varianza, e ne diamo dimostrazione, sia a titolo di esercizio che, soprattutto, al fine di sottolinearne ancora una volta l'importanza.

Dalla definizione di varianza abbiamo

$$\sigma^2_X = \sum_k p_k (x_k - E(X))^2 = \sum_k p_k \left(x_k^2 - 2E(X)\, x_k + E(X)^2\right) = \sum_k p_k x_k^2 -$$
$$- 2E(X) \sum_k p_k x_k + E(X)^2 \sum_k p_k = E(X^2) - 2E(X)E(X) + E(X)^2 = E(X^2) - E(X)^2,$$

dove si è sfruttato il fatto che $\sum_k p_k x_k^2 = E(X^2)$, che $\sum_k p_k x_k = E(X)$ e che $\sum_k p_k = 1$.

Vediamo un esempio, sempre relativo a tale proprietà. Consideriamo il lancio di due monete, e X sia la v.a. che conta il numero delle teste: essa assume dunque tre valori, $x_1 = 0$, $x_2 = 1$ e $x_3 = 2$, con probabilità (o frequenze relative) $p_1 = 1/4$, $p_2 = 1/2$ e $p_3 = 1/4$. Il valor medio della v.a. X è allora $E(X) = \sum p_i x_l = (1/4) \cdot 0 + (1/2) \cdot 1 + (1/4) \cdot 2 = 1$; la varianza sarà

$$\sigma^2_X = \sum p_i (x_i - E(X))^2 = (1/4)(0-1)^2 + (1/2)(1-1)^2 + (1/4)(2-1)^2 = 1/4 + 1/4 = 1/2.$$

Detta ora $Y = X^2$ la v.a. dei quadrati della precedente, quindi con valori $y_1 = 0$, $y_2 = 1$ e infine $y_3 = 4$, il suo valor medio sarà

$$E(Y) = E\left(X^2\right) = \sum_{i=1}^{3} p_i y_l = (1/4)\cdot 0^2 + (1/2)\cdot 1^2 + (1/4)\cdot 2^2 = 1/2 + 1/4\cdot 4 = 1/2 + 1 = 3/2\,,$$

e dunque possiamo ricalcolare la varianza di X come $\sigma_X^2 = E\left(X^2\right) - E(X)^2 = 3/2 - 1^2 = 1/2$, come doveva essere.

Vediamo ancora un secondo esempio, nel quale si considerano due v.a. uguali, X_1 e X_2, i cui valori sono i numeri delle sei facce di un normale dado da giuoco; sia quindi Y la v.a. differenza delle due, sia cioè $Y = X_1 - X_2$. Le due v.a. sono stocasticamente indipendenti tra loro, dal momento che l'esito del lancio di ciascun dado non influenza l'esito del lancio dell'altro. Come noto, la speranza matematica di ognuna delle due è $E(\mathrm{X}_1) = E(\mathrm{X}_2) = 3.5$, e la varianza di ognuna è

$$VAR(\mathrm{X}_\mathrm{i}) = \frac{(1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2}{6} =$$
$$= \frac{2\left[(2.5)^2 + (1.5)^2 + (0.5)^2\right]}{6} = \frac{17.5}{6} = \frac{105}{36} = \frac{35}{12}$$

con $i = 1{,}2$. Per quanto riguarda la v.a differenza, Y, possiamo affermare che il suo valor medio deve ottenersi come differenza dei valori medi delle due precedenti, e dunque deve essere $E(Y) = E(X_1) - E(X_2) = 0$. Facciamo attenzione a non continuare su questa strada, affermando che la sua varianza si possa ottenere come differenza delle varianze, e dunque si annulli. A parte il fatto che una varianza nulla garantirebbe la certezza di un unico risultato, cosa evidentemente inammissibile nel caso in esame, dobbiamo ricordare che le due varianze andranno sommate e non sottratte.

Calcoliamo la varianza di Y, determinandone i valori con le loro frequenze e le frequenze relative. Costruiamo una tabella mettendo al margine superiore i valori di X_1, ed a margine sinistro i valori di X_2; gli altri elementi della tabella sono ottenuti come differenza tra il valore che intesta la colonna ed il valore che intesta la riga, e sono dunque i valori di Y.
Abbiamo

$X_1 =$	1	2	3	4	5	6
$X_2 = 1$	0	1	2	3	4	5
2	-1	0	1	2	3	4
3	-2	-1	0	1	2	3
4	-3	-2	-1	0	1	2
5	-4	-3	-2	-1	0	1
6	-5	-4	-3	-2	-1	0

.

Si vede come la v.a. differenza possa assumere undici differenti valori, a partire dal minimo $x_1 = -5$, ottenuto quando X_1 vale –1 e X_2 vale 6, al massimo $x_{11} = 5$, ottenuto quando X_1 vale 6 ed X_2 vale 1.

Possiamo così costruire una seconda tabella sulla prima colonna della quale riportiamo, in ordine crescente, gli undici valori y di Y, sulla seconda colonna le loro frequenze n, e sulla terza le corrispondenti frequenze relative. Abbiamo

y	n	p
-5	1	1/36
-4	2	2/36
-3	3	3/36
-2	4	4/36
-1	5	5/36
0	6	6/36
1	5	5/36
2	4	4/36
3	3	3/36
4	2	2/36
5	1	1/36

,

dalla quale è immediata la verifica che $E(Y) = 0$. Di conseguenza, la varianza si può calcolare direttamente come $E(Y^2)$; dal momento che i valori di Y^2 si riducono a solo cinque diversi da zero, sommando le frequenze relative dei risultati positivi di Y con quelle dei risultati negativi (dunque, moltiplicando per due tali frequenze), possiamo scrivere

$$VAR(Y) = E(Y^2) = \sum p_i y_i^2 = \frac{2}{36}\left[(5)^2 + 2(4)^2 + 3(3)^2 + 4(2)^2 + 5(1)^2\right] =$$
$$= \frac{2}{36}[25 + 32 + 27 + 16 + 5] = \frac{210}{36} = \frac{35}{6} = 2\frac{35}{12}.$$

Come si vede, la varianza di Y è esattamente la somma delle varianze di X_1 e X_2 calcolate in precedenza, entrambe uguali a $35/12$.