

Premessa.



Con riferimento al gioco del lotto, fissiamo l'attenzione sul numero che sarà estratto per primo sulla ruota di Palermo sabato prossimo. Allo stato della nostra attuale informazione, quel numero, che ancora non conosciamo, è uno degli interi compresi tra 1 e 90.

Siamo quindi di fronte ad una partizione di novanta casi elementari ω_h : "Il primo estratto sulla ruota di Palermo sabato prossimo è h " ($h = 1, 2, \dots, 90$).

Diremo anche che è *aleatorio* (nel senso di non ancora conosciuto) il numero "primo estratto ..." il cui valore è da ritenersi *ben determinato sin da questo momento ma non noto per carenza di informazione*.

Si possono costruire esempi a volontà. E' aleatorio il numero dei pezzi che saranno dichiarati accettabili dopo il

collaudo che effettueremo su n pezzi prodotti da una macchina, è aleatorio il numero di “*teste*” che usciranno in n lanci successivi di una moneta, il numero di errori tipografici nel testo di cui ci accingiamo a leggere le bozze

Abbiamo sin qui indicato esempi di numeri aleatori che conducono alla considerazione di partizioni finite. Ogni evento della corrispondente partizione fa riferimento ad uno dei possibili, distinti, valori o, come diremo nel seguito, ad una possibile *determinazione* del numero aleatorio.

E' agevole, peraltro, pensare a numeri aleatori le cui possibili determinazioni sono un'infinità (numerabile o continua).

E' numerabile, ad esempio, l'insieme dei possibili valori del “numero razionale scelto a caso nell'intervallo $(0, 1)$ ”. E si potrà pensare ad un esperimento concettuale che permetta di parlare di

“scelta a caso” di un numero razionale.

Conviene considerare numerabile l'insieme delle possibili determinazioni del numero, aleatorio, di “arrivi” di un certo fenomeno entro un assegnato intervallo di tempo. Potrà trattarsi, concretamente, del numero di chiamate telefoniche ad un centralino nella prossima ora, o del numero di arresti di un'attrezzatura per guasto o altro nelle prossime ventiquattro ore, del numero di sinistri che saranno denunciati ad una compagnia assicuratrice nel prossimo esercizio In questi esempi infatti, non è possibile fissare a priori, in condizioni generali, un numero massimo di arrivi, sicchè appare opportuno accogliere l'ipotesi che le possibili determinazioni del numero di “arrivi” siano lo zero e gli interi positivi.

Si perviene poi alla considerazione di un numero aleatorio le cui determinazioni riempiono un insieme che ha la potenza del

continuo pensando alle possibili posizioni di un punto su un'asse. Ad esempio: fissato su un bersaglio piano un punto da chiamarsi "centro", si fissi l'attenzione sulla distanza da esso del punto che sarà colpito al prossimo colpo. Tale distanza è misurata da un numero aleatorio le cui possibili determinazioni riempiono un intervallo di numeri reali non negativi.

La "ulteriore durata in vita" di un individuo o di un'attrezzatura offre un altro esempio di numero aleatorio le cui determinazioni possibili sono quelle di un intervallo di numeri reali che sarà teoricamente opportuno considerare superiormente illimitato.

Interessa a questo punto sottolineare due aspetti del discorso precedente, che sono:

- 1) parlando di un numero aleatorio occorre pensare che trattasi di un numero compiutamente determinato ma

non ancora a noi noto per carenza di informazione, si riferisca essa ad eventi futuri o passati;

- 2) la considerazione di un numero aleatorio ci ha condotti spontaneamente alla considerazione di una partizione di casi elementari possibili.

In generale, per quel che riguarda la partizione, indicato - come faremo nel seguito - con X o altra lettera maiuscola un numero aleatorio e con le corrispondenti lettere minuscole le possibili determinazioni distinte di X , gli eventi aleatori possibili sono del tipo $\{X = x\}$ con $x \in A$, essendo A un generico insieme di numeri reali. In particolare: $\{X = x_h\}$, $h \in N$ e $\{X = x\}$, $x \in I$, ove N è l'insieme finito di interi o l'insieme dei numeri naturali (e ciò quando la partizione è finita, rispettivamente numerabile) ed I un insieme di numeri reali avente la potenza del continuo (quando tale è la potenza dell'insieme delle determinazioni di X).

Per quel che concerne la carenza di informazione va osservato che, in generale, non ci troviamo di fatto in uno stato di completa ignoranza sull'andamento del fenomeno, a cui il numero aleatorio è legato. Possediamo invece una certa informazione che, in quanto incompleta, non ci consente di precisare il valore del numero aleatorio, cioè la determinazione che esso ha assunto o assumerà. Essa però, oltre a consentirci eventualmente di individuare l'insieme delle possibili determinazioni del numero aleatorio, riuscirà sufficiente per permetterci di distribuire la probabilità sugli eventi della partizione da esso generata. Ci serviremo, allo scopo, di valutazioni coerenti formulate sulla base di opinioni convenienti che conducano, quando possibile, a modelli di distribuzioni semplici o operativamente agili.

Tipicamente, però, la costruzione della distribuzione

avviene attraverso valutazioni sintetiche via via più numerose. Intendiamo dire con ciò che, di norma, noi saremo in grado di incominciare con il riassumere le nostre opinioni e informazioni in un numero certo che ci dica l'ordine di grandezza che ci attendiamo per il numero aleatorio in considerazione e, poi, in altri numeri certi che ci indichino l'ordine di grandezza dei possibili scarti (o di loro convenienti funzioni) attorno a quella prima valutazione riassuntiva.

Questo processo si arresta, di norma e a seconda delle esigenze dei singoli problemi, ad un modesto numero di valutazioni sintetiche che non permettono in generale di costruire l'intera distribuzione ma che sono sufficienti per la trattazione del problema in esame. Tali questioni saranno trattate a suo luogo ove vedremo come tali valutazioni sintetiche si possano altresì ricavare una volta assegnata la distribuzione di

probabilità sulla partizione. Riprendiamo ora le esemplificazioni sopra indicate.

Con riferimento al numero aleatorio “primo estratto ... “ appare ragionevole accogliere l’ipotesi di “simmetria” e dare quindi una distribuzione uniforme sui novanta casi elementari, assegnando cioè il valore $1/90$ ad ogni caso elementare ω_h ($h = 1, 2, \dots, 90$).

Con riferimento alla distanza aleatoria dal centro del bersaglio del punto che sarà colpito nel prossimo colpo (e, del pari, per “la ulteriore durata in vita ... ”) appare naturale attribuire probabilità nulla al caso elementare $\{X = x\}$, qualunque sia x nell’intervallo I di possibili valori e assegnare probabilità positive a *intervalli di valori* x contenuti in I , introducendo - nel modo che sarà chiarito nel primo capitolo - una densità della distribuzione.

Capitolo I°:

Le variabili aleatorie

unidimensionali.

I.1 La variabile aleatoria come funzione P -misurabile.

Come detto in premessa, la considerazione di un numero aleatorio conduce spontaneamente alla considerazione di una partizione: quella degli eventi $\{X = x\}$, ove x percorre l'insieme delle possibili determinazioni, e che le informazioni che possediamo consentono, almeno teoricamente, di assegnare una distribuzione di probabilità su quella partizione.

Più in generale, ora, diremo che stiamo considerando un numero aleatorio ogni volta che agli eventi di una partizione, sulla quale è stata assegnata una distribuzione di probabilità, siano associate le determinazioni numeriche di una grandezza, dimensionale o no.

Con ciò un numero aleatorio X appare come una funzione, che indicheremo nel seguito con il simbolo $x(\omega)$, definita sui "punti" ω di una partizione Ω .

Il codominio di tale funzione è un insieme di numeri reali che può essere finito, numerabile, continuo. Sulla partizione Ω , d'altra parte, è stata assegnata una distribuzione di probabilità. Se vogliamo che tale distribuzione induca, tramite la funzione $x(\omega)$, una distribuzione di probabilità sull'insieme dei possibili valori del numero aleatorio [codominio di $x(\omega)$] occorre che la funzione stessa sia opportuna e cioè *misurabile*, nel senso che ora preciseremo.

Dire che su una partizione Ω è stata assegnata una distribuzione di probabilità significa dire, lo ricordiamo, che è stata assegnata una funzione (misura) P non negativa, normalizzata, completamente additiva, sugli eventi di una famiglia boreliana $\{B\}$, di Ω (sui casi elementari se la partizione è discreta).

Su Ω sia definita ora una funzione $x(\omega)$. Fissato in Ω un

evento E (insieme di “punti” ω) rimarrà individuato un insieme J_E di numeri reali, codominio della $x(\omega)$ al variare di ω in E . L’insieme J_E è detto anche *immagine* di E [generata da $x(\omega)$] nell’insieme dei numeri reali.

Fissato, viceversa, un insieme J di numeri reali rimane individuato in Ω un insieme E_J di punti ω in corrispondenza dei quali riesce $x(\omega) \in J$, cioè si verifica l’evento $E_J = \{x(\omega) \in J\}$ (la determinazione di X è un numero di J). L’insieme E_J dicesi anche *controimmagine* di J in Ω .

Interessa a questo punto pervenire alla valutazione di probabilità di eventi del tipo $\{x(\omega) \in J\}$. Ma per quali insiemi J ?

Se le determinazioni possibili del numero aleatorio $X = x(\omega)$ sono in numero discreto, gli J che interesano si riducono, manifestamente, a punti dell’asse reale o a loro riunioni finite o al più numerabili.

In condizioni generali basterà limitarsi alla considerazione di intervalli (degeneri o no, limitati o no, aperti o chiusi, superiormente chiusi ...). Se, infatti, sono valutabili le probabilità delle controimmagini E_I di intervalli I , lo sono quelle di eventi Ω controimmagini di boreliani dell'asse reale, poiché esse appartengono alla famiglia completamente additiva contenente gli eventi E_I .

Basterà anzi, come vedremo, considerare particolari intervalli: quelli inferiormente limitati e superiormente chiusi. Siano J un tale intervallo e x il suo estremo superiore. Consideriamo l'evento $\{x(\omega) \in J\} = \{x(\omega) \leq x\}$, ovvero l'evento: "il numero aleatorio $X = x(\omega)$ assume una determinazione non superiore ad x ". Valutare la probabilità di tale evento significa valutare la probabilità dell'evento E_J , controimmagine in Ω di J .

Occorre dunque che E_J appartenga alla famiglia di eventi

per i quali rimane assegnata la probabilità P in Ω , ovvero che sia un evento che diremo P -misurabile. Sono tali gli eventi della famiglia $\{B\}$ dei boreliani o riunioni di essi con eventi di probabilità nulla (ricordiamo che, in generale, non ogni evento di Ω riesce misurabile se si vuole che la misura P sia completamente additiva!). In conclusione: la funzione $x(\omega)$, che descrive il numero aleatorio in esame, deve godere della seguente proprietà:

Comunque si fissi un numero reale x , l'insieme dei punti ω di Ω in corrispondenza dei quali riesce $\{x(\omega) \leq x\}$ dev'essere P -misurabile, ovvero dev'essere possibile assegnare la probabilità all'evento $\{x(\omega) \leq x\}$.

Una funzione $x(\omega)$ che goda di tale proprietà è detta anche una *funzione P -misurabile*.

Raccogliendo le nozioni sin qui esposte possiamo

finalmente pervenire alla seguente definizione, in via assiomatica, di numero aleatorio:

Un numero aleatorio X è una funzione reale, P -misurabile, definita sui punti di una partizione sulla quale sia stata assegnata una distribuzione, P , di probabilità.

Nel seguito in luogo di *numero aleatorio* diremo *variabile aleatoria*, e ciò per adeguarci alla terminologia ormai imposta ovunque da tempo. In realtà però, nella nozione di *variabile aleatoria* non vi è nulla di ... *variabile*. Trattasi, lo ripetiamo ancora, di un numero (o grandezza dimensionale) compiutamente determinato ma non ancora conosciuto per mancanza di informazione.

I.2 Funzione di ripartizione di una variabile aleatoria.

Nel precedente paragrafo è stato detto che la considerazione

di una variabile aleatoria X conduce in sostanza, fissato arbitrariamente un numero reale x , alla considerazione dell'evento che indicheremo nel seguito con $\{X \leq x\}$: "La variabile aleatoria X assume una determinazione non superiore a x ". E' per tali eventi che dobbiamo saper valutare la probabilità.

Diamo ora la seguente definizione:

Dicesi *funzione di ripartizione della v.a. X* la funzione di variabile reale, $F(x)$ definita dalla:

$$F(x) = \text{Prob } \{X \leq x\}.$$

La funzione di ripartizione è ovviamente limitata:

$$0 \leq F(x) \leq 1.$$

Essa riesce poi monotona non decrescente. Siano infatti x_1 e x_2 due numeri reali qualunque con $x_2 > x_1$. E':

$$F(x_2) = \text{Prob } \{X \leq x_2\} \geq \text{Prob } \{X \leq x_1\} = F(x_1),$$

poiché l'evento $\{X \leq x_1\}$ implica l'evento $\{X \leq x_2\}$ (e la P è

completamente additiva).

E' chiaro inoltre, da quanto precede, che:

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

La $F(x)$ può riuscire continua o no. In quest'ultimo caso o punti di discontinuità possono essere al più un'infinità numerabile e le discontinuità sono di prima specie (salti). Ciò è conseguenza del fatto che $F(x)$ è monotona e limitata. Inoltre negli eventuali punti di discontinuità la $F(x)$ è continua a destra. Tale proprietà, ci limitiamo ad affermarlo, è conseguenza della completa additività della probabilità P (cioè della continuità della misura P).

E' agevole verificare a questo punto che assegnata la $F(x)$, assegnata cioè la probabilità ad ogni evento $\{X \leq x\}$, rimane assegnata la probabilità ad ogni evento del tipo $\{X < x\}$, $\{X = x\}$, $\{X > x\}$, $\{X \geq x\}$, $\{a < X < b\}$, $\{a \leq X < b\}$, $\{a < X \leq b\}$, $\{a$

$\leq X \leq b\}$ ovvero, come detto al precedente paragrafo, riamne assegnata la probabilità per ogni intervallo dell'asse reale (e quindi per la famiglia boreliana dell'asse, generata dagli intervalli).

In particolare: la probabilità dell'evento $\{X = x_0\}$ è la differenza tra il limite destro e quello sinistro della $F(x)$ nel punto x_0 :

$$\text{Pr ob}\{X = x_0\} = \lim_{x \rightarrow x_0^+} F(x) - \lim_{x \rightarrow x_0^-} F(x).$$

Dalla

$$\text{Pr ob}\{X \leq a\} + \text{Pr ob}\{a < X \leq b\} = \text{Pr ob}\{X \leq b\}$$

segue la

$$\text{Pr ob}\{a < X \leq b\} = F(b) - F(a).$$

Si dirà che una distribuzione è *limitata* se le possibili determinazioni di X cadono tutte entro un intervallo limitato.

Torna utile interpretare la distribuzione di probabilità di una

v.a. alla stregua della distribuzione di una massa (unitaria) su una retta. La $F(x)$, allora, misura la “massa depositata alla sinistra del punto di ascissa x ” (compresa quella eventualmente concentrata in x stesso). L’incremento $F(b)-F(a)$ misura la massa contenuta nell’intervallo superiormente chiuso $(a, b]$.

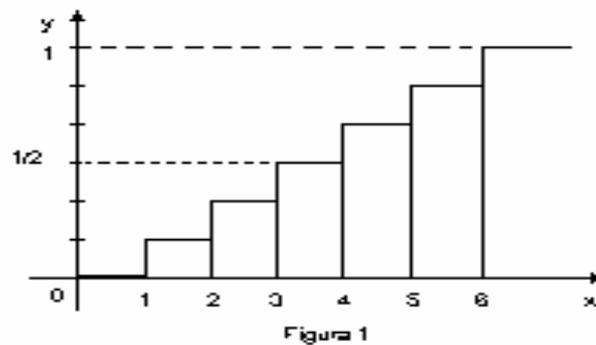
Se le determinazioni della v.a. X sono in numero discreto, la distribuzione concentra la massa in un numero discreto di punti assegnando alla determinazione x_h la probabilità p_h ($h = 1, 2, \dots$; $\sum p_h = 1$). Si parla anche di v.a. *discreta*.

E’ questo, ad esempio, il caso della v.a. “punto realizzabile col lancio di un dado”. Le possibili determinazioni sono i sei numeri interi $1, 2, \dots, 6$.

Per ragioni di “simmetria” assumiamo la equiprobabilità dei sei casi elementari possibili; la distribuzione concentra allora masse pari a $1/6$ in ognuno dei 6 punti di ascissa $x = 1, x = 2, \dots$,

$x = 6$.

La corrispondente funzione di ripartizione, il cui diagramma è indicato in *figura 1*, è costante a tratti, continua a destra.



quando le determinazioni di una v.a. X riempiono un insieme continuo di numeri reali possono presentarsi i seguenti casi:

- a) *La distribuzione concentra ancora la massa in un numero finito (o numerabile) di punti x_h ($h = 1, 2, \dots$). Agli intervalli dell'asse reale che non contengono alcun punto x_h viene allora attribuita probabilità nulla. La funzione di*

ripartizione $F(x)$ è ancora una funzione costante a tratti con “salti” di ampiezza p_h nei punti x_h . Una situazione siffatta si incontra tipicamente ogni volta che si “discretizza” una v.a. i cui valori possibili riempiono un continuo di numeri reali. Si pensi, ad esempio, alla v.a. X “statura di un individuo scelto a caso in una collettività”. Le sue determinazioni possibili formano un insieme di numeri reali positivi. Se, come d’ordinario, si conviene di misurare la statura approssimata al centimetro, la variabile diventa discreta. Dire allora che la statura è 178 cm significa in realtà che la statura di quell’individuo è compresa tra 177,5 e 178,5 cm. L’insieme delle possibili determinazioni viene cioè diviso in intervalli di ampiezza pari a 1 cm a partire da una statura minima quale, ad esempio, 150 cm e sino ad una statura massima quale, ad esempio, 220 cm. La funzione di

ripartizione concentra la massa nei 71 punti $x_1=150$, $x_2=151$, ..., $x_{71}=220$. La massa concentrata in $x_1=150$ è la probabilità che la statura non superi 150,5 cm, quella in $x_2=151$ è la probabilità che la statura sia compresa nell'intervallo (150,5 - 151,5), ecc. In effetti tutto avviene come se la v.a. X ammettesse solamente un numero finito di possibili determinazioni.

b) *La distribuzione non concentra masse in alcun punto: è "diffusa".* La $F(x)$ è continua ed esiste una funzione, $f(x)$, tale che:

$$F(x) = \int_{-\infty}^x f(t)dt .$$

La $f(x)$ non è negativa [$F(x)$ monotona non decrescente] e normalizzata, nel senso che

$$\int_{-\infty}^{+\infty} f(x)dx = \lim_{x \rightarrow +\infty} F(x) = 1 .$$

Nei punti di continuità della $f(x)$ risulta

$$\frac{dF(x)}{dx} = f(x).$$

La funzione $f(x)$ è detta anche *densità* della distribuzione. Si intende l'appropriatezza del termine osservando che la massa contenuta nell'intervallo $(x, x+\Delta x)$, cioè l'incremento $F(x+\Delta x) - F(x)$, vale, nelle nostre ipotesi e se x è un punto di discontinuità della densità, $f(x)\Delta x$ a meno di infinitesimi di ordine superiore al primo rispetto Δx .

La $F(x)$ monotona, continua, è derivabile quasi ovunque (salvo cioè un insieme di misura nulla) ed è l'integrale della propria derivata. Tale situazione si presenta quando la $F(x)$ è *assolutamente continua*. Non ricorderemo qui la definizione di funzione assolutamente continua preferendo dire anche nel seguito che nel caso che stiamo esaminando la v.a. X è continua e la sua distribuzione di probabilità è dotata di densità.

- c) *La $F(x)$ è continua, e (per la monotonia) derivabile quasi ovunque ma con derivata nulla (e non è quindi rappresentabile come integrale della propria derivata). La distribuzione non concentra la massa in alcun punto [perché la $F(x)$ è continua] ma la diffonde su un insieme di misura nulla (non esiste quindi, finita, una densità). E' un caso limite, interessante da un punto di vista teorico, ma del quale non avremo occasione di occuparci nel seguito.*
- d) Il caso più generale riguarda una distribuzione che è una “mistura” (combinazione lineare convessa) dei modelli descritti in a), b), c). La più generale funzione di ripartizione è cioè una funzione del tipo

$$F(x) = c_1 F_1(x) + c_2 F_2(x) + c_3 F_3(x),$$

con $c_i \geq 0$, $\sum c_i = 1$, e le F_i funzioni di ripartizione

appartenenti ai tre tipi considerati in a), b), c).

I.3 Valori caratteristici di una distribuzione.

La funzione di ripartizione descrive compiutamente una v.a. Frequentemente però, e specie nelle applicazioni, interessa (o basta o non è possibile altro che) conoscere alcuni valori caratteristici o sintetici della distribuzione che ne forniscano un'immagine riassuntiva sufficiente per gli scopi prefissi.

Si pensi ad esempio di dover considerare un guadagno aleatorio X , frutto dell'investimento di un capitale in un'attività economica i cui profitti non sono "certi" (com'è sempre realistico pensare). In un primo momento non interesserà considerare la distribuzione di X (e certo non sarà agevole, in condizioni generali, indicare la funzione di ripartizione). Interesserà però almeno la considerazione di un valore certo che ci consenta di orientarci sull'ordine di grandezza che ci attendiamo per il guadagno aleatorio in questione. Si ripensi, allo

scopo, all'interpretazione in termini di scommessa della definizione di probabilità nella concezione soggettiva. Valutare pari a p la probabilità di un certo evento E significa essere disposti a scambiare l'importo certo p con il guadagno aleatorio che assume il valore 1 se si verifica E e 0 nel caso opposto.

Orbene diremo *speranza matematica del guadagno aleatorio X l'importo certo m , che indicheremo anche col simbolo $E(X)$, che siamo disposti a scambiare con il guadagno aleatorio X , nel senso cioè che siamo disposti a ricevere (pagare) m per lasciare (richiedere) al competitore il possesso del guadagno che si realizzerà (o che si è realizzato ma che non conosciamo).*

La valutazione di tale importo certo assume sinteticamente la valutazione della distribuzione condensando in un unico valore le opinioni e informazioni sulla grandezza aleatoria X .

La considerazione della sola speranza matematica di X non è però sufficiente nella larga maggioranza dei casi. Si pensi, infatti, di dover decidere la scelta dell'investimento del capitale in due attività che assicurano guadagni aleatori le cui speranze matematiche giudichiamo uguali. Giudicheremo, in tal caso, indifferente investire nell'una o nell'altra attività il nostro capitale ? . Certamente no, se sappiamo che a parità di speranza matematica le determinazioni possibili (positive e negative) di un guadagno aleatorio sono molto maggiori in valore assoluto di quelle dell'altro sì che a fronte di eventuali grossi guadagni positivi possano verificarsi perdite molto elevate.

A parità di speranza matematica del guadagno, un investimento, appare allora più "rischioso" dell'altro e la decisione circa la scelta dell'investimento sarà influenzata dalla valutazione della rischiosità, ottenuta attraverso una misura della

dispersione della distribuzione attorno alla già stimata speranza matematica. Allo scopo potremo considerare una nuova v.a.: quella le cui determinazioni sono i quadrati degli scarti del guadagno aleatorio dalla sua speranza matematica. Valutata la speranza matematica di tale nuova variabile, ci saremo procurati un nuovo valore caratteristico, detto *varianza*, che ci darà indicazioni sulla dispersione della distribuzione. Tanto maggiore è la varianza tanto più dispersa è la distribuzione attorno alla speranza matematica.

Preciseremo successivamente tali concetti. Per il momento, con le considerazioni precedenti, ci siamo posti nella situazione di chi, trovandosi di fronte alla distribuzione di una massa comunque ripartita su un filo rigido (di massa trascurabile), si preoccupa anzi tutto di determinare il baricentro per passare successivamente alla ricerca del momento di inerzia attorno ad

esso, riassumendo nella conoscenza di queste grandezze quella della distribuzione nella sua interezza. Potremmo pensare di proseguire nella ricerca di altri valori sintetici valutando le speranze matematiche di convenienti funzioni della v.a. X onde pervenire ad una conoscenza sempre più estesa della distribuzione.

Diremo, viceversa, a questo punto come si costruiscono tali valori caratteristici, *nota la distribuzione*. Il problema di inferire sulla distribuzione attraverso la conoscenza di valori sintetici sarà ripreso successivamente.

I.4 Speranza matematica di una variabile aleatoria.

Sia X una v.a. e sia $F(x)$ la sua funzione di ripartizione.

Dicesi *speranza matematica* (o *valor medio*) di X il numero certo, il cui significato è stato dichiarato nel paragrafo

precedente, definito dalla

$$E(X) = \sum_h x_h p_h \quad [1.1]$$

se la distribuzione è discreta e la serie a secondo membro converge assolutamente (e quindi incondizionatamente) o dalla

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx \quad [1.2]$$

se la distribuzione è dotata di densità ed esiste, finito, l'integrale

$$\int_{-\infty}^{+\infty} |x|f(x)dx$$

Si osservi che la serie a secondo membro della [1.1], rispettivamente l'integrale della [1.2], potrebbero esistere finiti e non verificarsi l'assoluta sommabilità. In tali casi diremo che non esiste la speranza matematica di X . Manifestamente la speranza matematica di X esiste per ogni distribuzione limitata (in particolare quando le determinazioni di X sono in numero finito).

Le definizioni [1.1] e [1.2] riguardano due tipi particolari di distribuzioni (discreta e assolutamente continua).

Nel caso generale (e quindi anche nei due ora indicati) la definizione è data in termini dell'integrale di Stieltjes:

$$E(X) = \int_{-\infty}^{+\infty} x dF(x) \quad [1.3]$$

(ferma restando la condizione di assoluta sommabilità).

Dalle definizioni testè date appare che il numero $E(X)$ è interpretabile (come già suggerito) alla stregua di ascissa del baricentro di una distribuzione lineare di masse (massa totale unitaria).

Quando riesca $E(X) = 0$ la v.a. dicesi *equa*. Si osservi che è

$$E(aX+b) = aE(X)+b, \quad [1.4]$$

sicchè, in particolare, la v.a. $X-E(X)$ è *equa*.

Interessa ora considerare i valori medi di altre v.a., convenienti funzioni della v.a. X (in particolare, potenze intere

positive).

Sia $g(x)$ una funzione reale della variabile reale x , assolutamente sommabile rispetto alla funzione di ripartizione $F(x)$ della v.a. X , sicchè esista, finito, l'integrale

$$\int_{-\infty}^{+\infty} |g(x)| dF(x).$$

Diremo allora speranza matematica della v.a. $Y = g(X)$ il numero, che indicheremo con $E[g(X)]$, definito dalla:

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x) dF(x) \quad [1.5].$$

Va qui osservato che secondo la [1.3] la speranza matematica di $Y = g(X)$ è data dall'integrale

$$\int_{-\infty}^{+\infty} y dG(y)$$

ove $G(y)$ è la funzione di ripartizione della Y la quale, come si intuisce, può essere costruita nota la $F(x)$ e data la $g(x)$.

Notiamo che la [1.5] dice che il valor medio di una v.a. Y

funzione della v.a. X , si può calcolare operando direttamente con la distribuzione della Y .

Torna utile per il seguito il seguente risultato che generalizza la [1.4]. Siano $Y_1 = g_1(X)$ e $Y_2 = g_2(X)$ due v.a. funzioni della v.a. X ed esistano $E(Y_1)$, $E(Y_2)$.

Dalla

$$\int_{-\infty}^{+\infty} |g_1(x) + g_2(x)| dF(x) \leq \int_{-\infty}^{+\infty} |g_1(x)| dF(x) + \int_{-\infty}^{+\infty} |g_2(x)| dF(x)$$

segue l'esistenza di $E(Y_1 + Y_2)$ e risulta anzi, dalla [1.5], che:

$$E(Y_1 + Y_2) = E(Y_1) + E(Y_2) \quad [1.6].$$

La [1.6] stabilisce, in un caso particolare, la *proprietà additiva* della speranza matematica. Su tale proprietà ritorneremo nel successivo Capitolo.

I.5 Momenti di una variabile aleatoria.

Considereremo ora particolari funzioni della v.a. X : le

potenze X^n e le potenze del suo valore assoluto $|X|^n$, con n intero positivo.

Le speranze matematiche, quando esistano finite,

$$m_n = E(X^n) = \int_{-\infty}^{+\infty} x^n dF(x) ,$$

$$\beta_n = E(|X|^n) = \int_{-\infty}^{+\infty} |x|^n dF(x) ,$$

diconsi rispettivamente *momento di ordine n* , e *momento assoluto di ordine n* dalla v.a. X (o della sua distribuzione).

Dalle definizioni scende che: se esiste il momento m_n esistono il momento assoluto β_n e tutti i momenti m_k e β_k di ordine k inferiore a n . In particolare è $m_0=1$ e $m_1=E(X)$.

L'importanza della considerazione dei momenti segue dal seguente teorema che ci limitiamo solo ad enunciare. Esistano i momenti m_n di ordine n comunque elevato della v.a. X e la serie di potenze

$$\sum \frac{m_n}{n!} z^n$$

abbia raggio di convergenza positivo. La successione $\{m_n\}$ dei momenti individua allora la funzione di ripartizione, $F(x)$, di X .

Altre questioni nelle quali torna utile la considerazione dei momenti riguardano la “forma” della distribuzione. Una distribuzione si dice *simmetrica* se esiste un numero a tale che, per ogni x ,

$$F(a - x) = 1 - F(a + x) - \text{Prob}\{X = a + x\}$$

o nel caso di una distribuzione dotata di densità,

$$f(a - x) = f(a + x) .$$

Il numero a è detto, in tali condizioni, centro di simmetria.

Manifestamente se una distribuzione è simmetrica ed esiste la speranza matematica, questa è il centro di simmetria.

Definiamo a questo punto i momenti di ordine n attorno ad un valore c , qualunque, come i numeri $E[(X-c)^n]$. in particolare

diconsi *momenti centrali* quelli valutati attorno a $c = E(X) = m_1$.

Il momento centrale di ordine n sarà indicato nel seguito con μ_n ,

sicchè $\mu_n = E[(X-m_1)^n]$.

Se esistono i momenti, la simmetria di una distribuzione è caratterizzata dal fatto che i momenti centrali di ordine dispari (quanti esistono) sono nulli. Per misurare la deviazione dall'andamento simmetrico di una distribuzione si fa ricorso, se esiste il momento centrale terzo, al cosiddetto *coefficiente di asimmetria*,

$$\gamma = \frac{\mu_3}{(\mu_2)^{3/2}} .$$

A seconda del segno di γ si parla di asimmetria positiva o negativa.

I.6 Varianza. - Disuguaglianza di Bienaymè-Cebicev.

Particolarmente significativa è la considerazione del momento centrale secondo detto anche *varianza* di X ,

$$\text{var}(X) = \mu_2 = E[(X - m_1)^2] .$$

La determinazione positiva della radice quadrata di μ_2 è detta anche *scarto quadratico medio* (brevemente s.q.m.) e indicata usualmente con σ (e per tale motivo denoteremo talvolta, nel seguito, con $\sigma^2(X)$ o con σ_x^2 la varianza di X).

Nell'analogia meccanica, precedentemente indicata a proposito della speranza matematica, il significato di μ_2 è quello di momento di inerzia attorno al baricentro. In altri termini, σ è la distanza dal baricentro dei punti (simmetrici rispetto ad esso) nei quali andrebbe concentrata la massa unitaria (metà sull'uno e metà sull'altro, ad esempio) per avere una distribuzione concentrata che presenti lo stesso momento di inerzia (rispetto al baricentro) della distribuzione originale.

E' agevole intendere che il secondo momento della variabile X attorno ad un valore c è minimo quando esso è calcolato attorno al valor medio m_1 . La verifica formale è immediata. E' infatti:

$$(X - c)^2 = [(X - m_1) + (m_1 - c)]^2$$

e di qui, in forza della [1.6],

$$\begin{aligned} E[(X - c)^2] &= E[(X - m_1)^2] + 2(m_1 - c)E(X - m_1) + (m_1 - c)^2 = \\ &= E[(X - m_1)^2] + (m_1 - c)^2, \end{aligned}$$

sicchè

$$\text{var}(X) \leq E[(X - c)^2]$$

e il minimo di $E[(x - c)^2]$ si realizza proprio in corrispondenza di $c = m_1$.

Torna utile per il seguito tenere presente la:

$$\text{var}(X) = E(X)^2 - E[(X)]^2. \quad [1.7]$$

Si osservi poi che sussiste la

$$\text{var}(aX + b) = a^2 \text{var}(X) ,$$

come facilmente si prova.

Tra le variabili $Y = aX + b$, funzioni lineari della X , particolarmente interessante è la

$$Y = \frac{X - m_1}{\sigma} \quad [\text{ove } m_1 = \mathbf{E}(X) \text{ e } \sigma = \sigma(X)]$$

che dicesi *variabile ridotta* (associata alla X). Essa è una variabile equa [$\mathbf{E}(Y) = 0$] la cui varianza è unitaria; inoltre è una variabile adimensionale (mentre X potrebbe essere misurata in cm, kg, ..., a seconda del suo significato).

Il valor medio m_1 di una variabile aleatoria X fornisce una prima immagine riassuntiva, *posizionale*, della distribuzione: la varianza (o lo s.q.m.) ha invece le caratteristiche di un parametro di *dispersione* nel senso che la conoscenza della varianza permette di intendere se la distribuzione sia più o meno concentrata attorno al valor medio m .

Celebre in proposito è la *disuguaglianza di Bienaymè-Cebicev*. Essa dice che, noti la speranza matematica m_1 e lo s.q.m. σ , fissato comunque un numero reale positivo t , risulta

$$\text{Pr ob}\{|X - m_1| \geq t\sigma\} \leq \frac{1}{t^2}$$

o, anche, posto $\varepsilon = t\sigma$:

$$\text{Pr ob}\{|X - m_1| \geq \varepsilon\} \leq \frac{\text{var}(X)}{\varepsilon^2} . \quad [1.8]$$

si intende subito il sussistere di tale disuguaglianza pensando in termini dell'analogia meccanica, La massa esterna all'intervallo di ampiezza $2t\sigma$ centrato nel baricentro non può superare $1/t^2$ perché - altrimenti - il momento di inerzia della distribuzione sarebbe maggiore di $\sigma^2 = \text{var}(X)$.

La prova della [1.8] per via analitica si consegue del resto facilmente. Sia ε un numero positivo, arbitrariamente scelto. E':

$$\text{Pr ob}\{|X - m_1| \geq \varepsilon\} = \int_{|x-m_1| \geq \varepsilon} dF(x) .$$

Per x variabile nell'insieme di integrazione è

$$\frac{|x - m_1|}{\varepsilon} \geq 1 ,$$

sicchè:

$$\begin{aligned} \int_{|x-m_1| \geq \varepsilon} dF(x) &\leq \frac{1}{\varepsilon^2} \int_{|x-m_1| \geq \varepsilon} (x - m_1)^2 dF(x) \leq \\ &\leq \frac{1}{\varepsilon^2} \int_{-\infty}^{+\infty} (x - m_1)^2 dF(x) = \frac{\text{var}(X)}{\varepsilon^2}. \end{aligned}$$

Si osservi che nella maggiorazione si può considerare, più

in generale, la potenza h -esima del rapporto $\frac{|x - m_1|}{\varepsilon}$. Quando

esista, finito, il momento centrale h -esimo si perviene allora alla

seguinte generalizzazione della disuguaglianza di Bienaymè-

Cebicev:

$$\text{Pr ob}\{|X - m_1| \geq \varepsilon\} \leq \frac{1}{\varepsilon^h} E(|X - m_1|^h) . \quad [1.9]$$

La disuguaglianza di Bienaymè-Cebicev è stata stabilita

senza ipotizzare per la distribuzione altro che l'esistenza del

valor medio e della varianza. Per colpa di tale generalità, la disuguaglianza è generalmente poco significativa quando la si debba impiegare nei casi particolari.

I.7 Indici posizionali. - Mediana, frattili, moda.

Si è detto che il valor medio è un indice posizionale. Altri indici di questo tipo sono la mediana, i frattili, la moda.

Dicesi *mediana*, della distribuzione un'ascissa, $x_{1/2}$, che divide l'asse dei numeri reali in due semirette su ciascuna delle quali rimane depositata la metà della massa della distribuzione della variabile X . Una tale definizione si adatta propriamente al caso di una distribuzione continua ove, allora, $x_{1/2}$ è soluzione della $F(x) = 1/2$.

Può avvenire - considerando in particolare funzioni di ripartizione costanti a tratti - che esistano più valori x

soddisfacenti la $F(x) = 1/2$; ognuno di essi è detto *mediana*.

Può accadere altresì che non esista alcuna ascissa $x_{1/2}$ tale che $F(x_{1/2}) = 1 - F(x_{1/2}) = 1/2$. Per ovviare a tale inconveniente e ponendosi nelle condizioni più generali, diremo *mediana* ogni valore $x_{1/2}$ che soddisfa contemporaneamente le:

$$\text{Pr ob}\{X \leq x\} \geq \frac{1}{2}, \quad \text{Pr ob}\{X \geq x\} \geq \frac{1}{2}.$$

La mediana è caratterizzata da una proprietà analoga a quella del valor medio: il primo momento assoluto $E(|X - c|)$ è minimo se c è mediana.

E' infatti:

$$E(|X - c|) = \int_{-\infty}^{+\infty} |x - c| dF(x) = - \int_{-\infty}^c (x - c) dF(x) + \int_c^{+\infty} (x - c) dF(x).$$

Facendo comparire $x_{1/2}$ negli integrali

$$\int_{-\infty}^c = \int_{-\infty}^{x_{1/2}} + \int_{x_{1/2}}^c \quad \text{e} \quad \int_c^{+\infty} = \int_c^{x_{1/2}} + \int_{x_{1/2}}^{+\infty}$$

e ponendo, sotto il segno di integrale, $x - c = x - x_{1/2} + x_{1/2} - c$ si

ottiene facilmente:

$$\begin{aligned} E(|X - c|) &= E(|X - x_{1/2}|) + 2 \int_c^{x_{1/2}} (x - c) dF(x) + (x_{1/2} - c) \left[\int_{x_{1/2}}^{+\infty} dF(x) - \int_{-\infty}^{x_{1/2}} dF(x) \right] = \\ &= E(|X - x_{1/2}|) + 2 \int_c^{x_{1/2}} (x - c) dF(x) \end{aligned}$$

(per la definizione di mediana). L'integrale $\int_c^{x_{1/2}} (x - c) dF(x)$ è non

negativo per c maggiore, minore o uguale a $x_{1/2}$; donde la conclusione.

In modo analogo a quello indicato per la mediana si definiscono i *frattili*. Fissato comunque un numero ν compreso tra 0 e 1, il numero (o i numeri) x_ν , soluzione della $F(x) = \nu$ dicesi, *frattile* di ordine ν . In particolare sono considerati i *quartili*, $\nu = 1/4, 1/2, 3/4$ (per $\nu = 1/2$ ritroviamo la mediana); i *decili*, $\nu = 1/10, \dots, 9/10$ e i *centili*, $\nu = 1/100, \dots, 99/100$.

Ben definita solo per le distribuzioni assolutamente continue e discrete è la *moda*. Nel primo caso essa è la

determinazione di X cui compete la massima densità (quando tale massimo è unico; altrimenti ogni punto massimo relativo è una moda e la distribuzione dicesi *plurimodale*).

Nel secondo caso, se p_h sono i “salti” della funzione di ripartizione $F(x)$ in corrispondenza alle determinazioni x_h della v.a. X , un punto x_h è una moda se risulta $p_{h-1} < p_h$ e $p_h > p_{h+1}$.

In *figura 2* è indicata la densità di una distribuzione

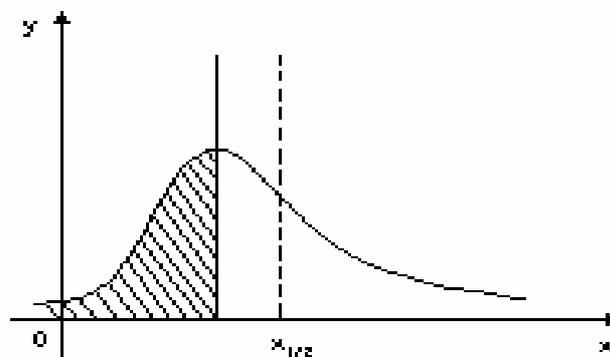


Figura 2

unimodale; la densità presenta una “coda” lunga a destra della moda, breve a sinistra; è asimmetrica e il coefficiente di asimmetria, γ , è positivo (si scambiano le conclusioni circa

l'ampiezza delle code se $\gamma < 0$).

Chiudiamo queste notizie indicando ancora altri valori tipici impiegati per esaminare la maggior o minor dispersione della distribuzione.

Limitiamo le nostre considerazioni allo *scarto semplice medio* che è il primo momento assoluto attorno alla mediana (con il che esso diviene minimo) e al *campo di variazione* che misura l'ampiezza del minimo intervallo, se esiste limitato, entro il quale cadono tutte le determinazioni della variabile aleatoria.

Come indice di dispersione è inoltre usata la *semidifferenza interquartile*, $(x_{3/4} - x_{1/4})/2$.

Capitolo II°:

Le variabili aleatorie

multiple.

II.1 Distribuzioni bidimensionali congiunte.

Nella premessa abbiamo considerato la variabile X distanza dal centro del punto colpito in un bersaglio piano. Il punto colpito, fissato un sistema cartesiano di riferimento, è individuato dalle sue due coordinate, sicchè si presenta spontanea anche la considerazione di una coppia, ordinata, di variabili aleatorie X, Y (ascissa e ordinata del punto).

In tal caso si parla, indifferentemente, di coppia di variabili o di *variabile aleatoria doppia*, o bidimensionale, o di vettore aleatorio a due componenti (unidimensionali) ed è immediato intendere come si possa accedere all'estensione di tale nozione per pervenire a quella di variabile aleatoria n -dimensionale (X_1, X_2, \dots, X_n) (o vettore aleatorio a n componenti). Soffermeremo, però, nel seguito, la nostra attenzione sul caso di variabili doppie, solo accennando alle estensioni.

Interesserà anzitutto introdurre la nozione di funzione di ripartizione estendendo in modo opportuno la nozione data precedentemente per una variabile unidimensionale.

Dicesi, precisamente, *funzione di ripartizione congiunta* della variabile doppia (X, Y) e si indica con $F(x, y)$, la funzione definita dalla:

$$F(x, y) = \text{Prob}\{X \leq x \cap Y \leq y\} .$$

La $F(x, y)$ è definita sull'intero piano; come funzione delle sue variabili, separatamente, è monotona non decrescente.

E'

$$\sum_{h,k} [g(x_h, y_k)] p_{hk}$$

e

$$\lim_{x \rightarrow -\infty} F(x, y) = \lim_{y \rightarrow -\infty} F(x, y) = 0 .$$

La $F(x, y)$ misura la "massa" contenute nel quadrante delimitato superiormente dalle semirette uscenti dal punto (x, y) ,

parallele agli assi e orientate come i semiassi negativi di x, y .

La massa contenuta nel rettangolo che ha il lati paralleli agli assi e vertici $P_{11}(x_1, y_1), P_{12}(x_1, y_2), P_{21}(x_2, y_1), P_{22}(x_2, y_2)$, (con $x_1 < x_2$ e $y_1 < y_2$) è data allora da:

$$F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) .$$

Si pensi, infatti, con riferimento alla *figura 3*, che per valutare la massa depositata sul rettangolo $P_{11} P_{12} P_{21} P_{22}$ occorre sottrarre dalla massa contenuta nel quadrante di vertice P_{22} , quella contenuta nei quadranti di vertici P_{12}, P_{21} e restituire quella del

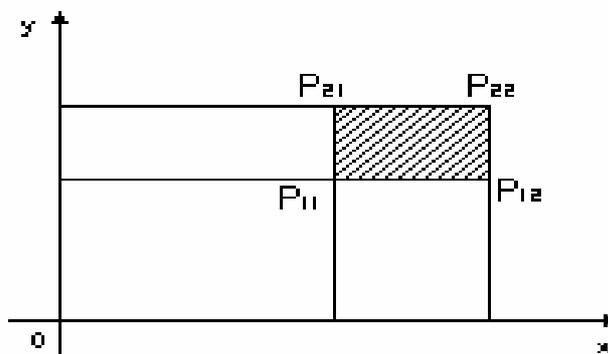


Figura 3

quadrante di vertice P_{11} che in tal modo è stata tolta due volte.

Valutata la massa che la distribuzione assegna ai rettangoli del piano, è facile valutare quella ripartita sui plurirettangoli (riunioni di rettangoli disgiunti) e pensare di “prolungare” tale valutazione sugli insiemi della famiglia boreliana generata dai rettangoli del piano.

Come nel caso unidimensionale limiteremo sostanzialmente le nostre considerazioni a due tipi particolari di distribuzioni congiunte: a) discrete e, b) dotate di densità.

a) La distribuzione è di tipo discreto se concentra masse p_{hk} nei punti (x_h, y_k) di un insieme al più numerabile di punti del piano. Va inteso che i numeri x_h sono le possibili determinazioni di una v.a. X e y_k quelle di una v.a. Y e

$$p_{hk} = \text{Pr ob}\{X = x_h \cap Y = y_k\} .$$

Dev'essere, ovviamente,

$$\sum_h \sum_k p_{hk} = 1 .$$

Si pensi, per semplificare, alla coppia di v.a.: X “punto realizzato col lancio di un dado”, Y “punto realizzato col lancio di un secondo dado (o col secondo lancio di un medesimo dado)”. Le determinazioni possibili sono le trentasei coppie ordinate (h,k) con $h, k = 1, 2, \dots, 6$. Per ragioni di “simmetria” e di indipendenza tra i due lanci, assumiamo $p_{hk} = 1/36$.

Per una distribuzione di tipo discreto è:

$$F(x, y) = \sum_{x_h \leq x} \sum_{y_k \leq y} p_{hk} .$$

La $F(x,y)$ è una funzione costante “a pezzi”, con discontinuità lungo intervalli o semirette uscenti dai punti (x_h, y_k) ove sono concentrate le masse.

In *figura 4* è indicato il valore della $F(x,y)$ nel caso del

citato esempio dei punti realizzabili con due lanci.

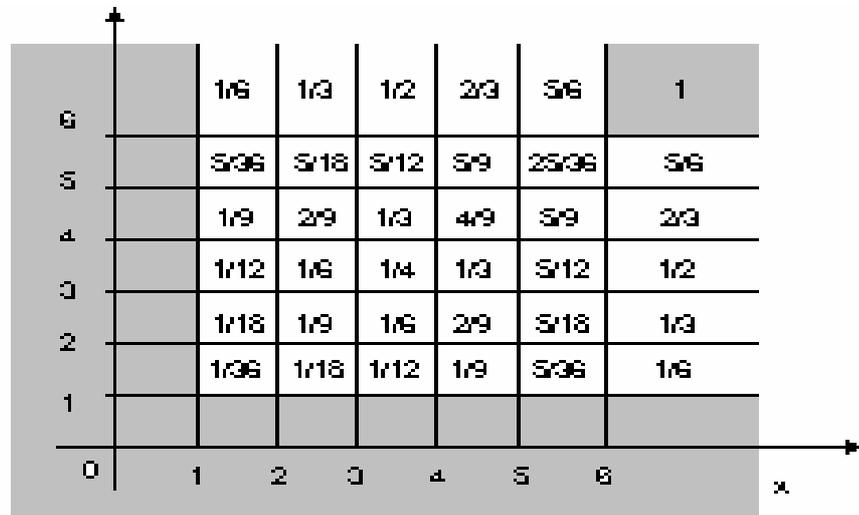


Figura 4

b) Una distribuzione si dice dotata di densità se esiste una

funzione $f(x, y)$ non negativa, normalizzata $\left(\int_{-\infty-\infty}^{+\infty+\infty} f(x, y) dx dy = 1 \right)$,

tale che per la funzione di ripartizione $F(x, y)$ risulti:

$$F(x, y) = \int_{-\infty-\infty}^x \int_{-\infty-\infty}^y f(\xi, \eta) d\xi d\eta .$$

La $F(x, y)$ è allora una funzione continua nella due variabili.

Nei punti di continuità della densità $f(x, y)$ è:

$$\frac{\delta^2 F(x, y)}{\delta x \delta y} = f(x, y) .$$

Osservando ora che, nell'ipotesi di continuità delle derivate seconde miste della $F(x, y)$, si ha:

$$\lim_{\Delta x \rightarrow 0, \Delta y \rightarrow 0} \frac{1}{\Delta x \Delta y} [F(x + \Delta x, y + \Delta y) - F(x, y + \Delta y) - F(x + \Delta x, y) + F(x, y)] = \frac{\delta^2 F(x, y)}{\delta x \delta y},$$

si intende che sussiste, trascurando infinitesimi di ordine superiore al secondo (rispetto alla distanza $\sqrt{\Delta x^2 + \Delta y^2}$, la:

$$\text{Prob}\{x < X \leq x + \Delta x \cap y < Y \leq y + \Delta y\} = f(x, y) \Delta x \Delta y,$$

che porge il significato della densità.

Per passare poi alla considerazione di casi più generali, occorre pensare che una distribuzione bidimensionale può presentare una varietà di comportamento maggiore di quelle unidimensionali.

Si rifletta (e soccorre anche qui l'immagine meccanica di distribuzione di masse sul piano) che la distribuzione può concentrare masse non solo su punti isolati ma altresì su "curve"

(varietà unidimensionali), sicchè il caso più generale si presenta come una mistura di distribuzioni delle quali una concentra masse su punti isolati, una seconda distribuisce masse su varietà unidimensionali con densità lineare, e una terza distribuisce masse su insiemi bidimensionali con densità superficiale.

II.2 Distribuzioni marginali e distribuzioni subordinate.

La funzione di ripartizione $F(x, y)$ della v.a. doppia (X, Y) è la funzione di ripartizione di una distribuzione di probabilità che è stata detta, lo ricordiamo, congiunta.

Sempre con riferimento alla coppia di quelle v.a., dicesi *distribuzione marginale* della v.a. X la distribuzione unidimensionale la cui funzione di ripartizione è data dalla

$$F_1(x) = \lim_{y \rightarrow +\infty} F(x, y) = \text{Prob}\{X \leq x \cap Y < +\infty\} ,$$

sicchè, pensando alla distribuzione di masse del piano, $F_1(x)$

misura la massa contenuta nel semipiano alla sinistra della retta (parallela all'asse y) luogo dei punti di ascissa x e quella eventualmente concentrata su tale retta.

Analogamente, la

$$F_2(y) = \lim_{x \rightarrow +\infty} F(x, y) = \text{Pr ob}\{X < +\infty \cap Y \leq y\}$$

dicesi funzione di ripartizione della distribuzione marginale della Y .

Il significato della nozione e l'appropriatezza del termine si intendono in modo particolarmente agevole se si fa riferimento alle distribuzioni discrete.

Notiamo, allo scopo, che in generale una distribuzione discreta è descritta dalla matrice delle probabilità p_{hk} attribuite alle determinazioni (x_h, y_k) . Si veda in proposito lo schema di

figura 5.

Y	y_1	y_2	\dots	y_j	\dots	y_m
-----	-------	-------	---------	-------	---------	-------

<u>X</u>							
x_1	p_{11}	p_{12}	\dots	p_{1j}	\dots	p_{1m}	$p_{1.}$
x_2	p_{21}	p_{22}	\dots	p_{2j}	\dots	p_{2m}	$p_{2.}$
\cdot	\cdot	\cdot	\dots	\cdot	\dots	\cdot	\cdot
\cdot	\cdot	\cdot	\dots	\cdot	\dots	\cdot	\cdot
x_i	p_{i1}	p_{i2}	\dots	p_{ij}	\dots	p_{im}	$p_{i.}$
\cdot	\cdot	\cdot	\dots	\cdot	\dots	\cdot	\cdot
\cdot	\cdot	\cdot	\dots	\cdot	\dots	\cdot	\cdot
x_n	p_{n1}	p_{n2}	\dots	p_{nj}	\dots	p_{nm}	$p_{n.}$
	$p_{.1}$	$p_{.2}$	\dots	$p_{.j}$	\dots	$p_{.m}$	

Figura 5

La tabella, beninteso, potrebbe estendersi all'infinito nelle due direzioni.

La somma della riga i -esima della matrice riportata in figura 5 è un numero, che denotiamo con $p_{i.}$, il cui significato è quello di probabilità dell'evento $\{X = x_i\}$.

Infatti:

$$\begin{aligned}
 p_{i.} &= p_{i1} + p_{i2} + \dots + p_{im} = \Pr ob\{X = x_i \cap Y = y_1\} + \dots + \Pr ob\{X = x_i \cap Y = y_m\} = \\
 &= \Pr ob\{X = x_i \cap Y < +\infty\} = \Pr ob\{X = x_i\}
 \end{aligned}$$

Le $p_{i.}$, fornisco, al variare di i , la distribuzione marginale di probabilità della X .

Analogo il significato delle $p_{.j}$ ottenute sommando gli elementi della colonna j -esima della matrice.

Le $p_{.j} = Prob\{X < +\infty \cap Y = y_j\}$ descrivono, al variare di j , la distribuzione marginale di probabilità della Y . Si noti che le $p_{.i}$ e $p_{.j}$ sono segnate “ai margini” della matrice di *figura 5*.

Nel caso delle distribuzioni dotate di densità è

$$F_1(x) = \int_{-\infty}^x d\xi \int_{-\infty}^{+\infty} f(\xi, \eta) d\eta ,$$

la cui densità

$$f_1(x) = \int_{-\infty}^{+\infty} f(x, \eta) d\eta$$

è detta densità della distribuzione marginale della X .

Analogamente

$$f_2(y) = \int_{-\infty}^{+\infty} f(\xi, y) d\xi$$

è detta densità della distribuzione marginale della Y .

Un'altra nozione di distribuzione interessa ora considerare ed è quella di *distribuzione secondaria (o condizionata)* ad un'ipotesi.

Con riferimento ad una variabile doppia (X, Y) di tipo discreto e ricordando la nozione di probabilità subordinata, è

$$\Pr ob\{X = x_i | Y = y_j\} = \frac{\Pr ob\{X = x_i \cap Y = y_j\}}{\Pr ob\{Y = y_j\}} = \frac{p_{ij}}{p_{.j}}$$

sicchè risulta $\sum_i \Pr ob\{X = x_i | Y = y_j\} = 1$. Le $\frac{p_{ij}}{p_{.j}}$, fissa

rimanendo l'ipotesi $Y = y_j$, forniscono dunque, al variare di i , una distribuzione di probabilità. Tale distribuzione è detta distribuzione (della variabile X) subordinata all'ipotesi $\{Y = y_j\}$.

Si indica con

$$F(x|y_j) = \Pr ob\{X \leq x | Y = y_j\}$$

la sua funzione di ripartizione.

Nel caso delle distribuzioni dotate di densità cominciamo

col definire la distribuzione della X subordinata all'ipotesi $y < Y$

$\leq y + k$ supposta di probabilità non nulla, attraverso la:

$$\begin{aligned} \text{Pr ob}\{X \leq x | y < Y \leq y + k\} &= \frac{\text{Pr ob}\{X \leq x \cap y < Y \leq y + k\}}{\text{Pr ob}\{y < Y \leq y + k\}} = \\ &= \frac{\int_{-\infty}^x d\xi \int_y^{y+k} f(\xi, \eta) d\eta}{\int_{-\infty}^x d\xi \int_y^{y+k} f(\xi, \eta) d\eta} = \frac{\int_{-\infty}^x f(\xi, \theta) d\xi}{\int_{-\infty}^x f(\xi, \theta) d\xi}. \end{aligned}$$

L'ultima uguaglianza sussiste in forza del teorema della

media integrale $\left(\int_y^{y+k} f(\xi, \eta) = k \cdot f(\xi, \theta) \right)$ [con θ opportuno in $(y, y + k)$].

Facendo ora tendere k allo zero, supposte soddisfatte le opportune condizioni di continuità della $f(x, y)$ che consentono il passaggio al limite sotto il segno di integrale e se la densità della distribuzione marginale $f_2(y)$ è non nulla nel punto y , si perviene alla:

$$F(x|y) = \lim_{k \rightarrow 0} \text{Pr ob} \{ X \leq x | y < Y \leq y + k \} = \frac{\int_{-\infty}^x f(\xi, y) d\xi}{f_2(y)} .$$

La $F(x|y)$ è, nelle nostre ipotesi, una funzione di ripartizione continua; la sua densità è

$$f(x|y) = \frac{f(x, y)}{f_2(y)} .$$

Analogo il significato della funzione $F(y|x)$ di ripartizione della Y subordinatamente all'ipotesi $\{X = x\}$ e della densità

$$f(y|x) = \frac{f(x, y)}{f_1(x)} .$$

Esemplificheremo i concetti ora esposti di distribuzione marginale e subordinata con riferimento al caso di una v.a. doppia (X, Y) discreta. Fissiamo, allo scopo, l'attenzione su una collettività di individui (ad esempio coscritti alla leva in una data sessione) e consideriamo per ogni individuo due caratteristiche somatiche: statura (misurata in cm) e peso (misurato in kg). Si

pensi di sorteggiare “a caso” un individuo da quella collettività (si intende con ciò riferirsi ad un meccanismo di scelta che attribuisce ad ognuno degli individui la medesima probabilità di essere sorteggiato). Sono aleatorie, prima del sorteggio, la statura X il peso Y di quell’individuo. Siano x_1, x_2, \dots, x_n le possibili determinazioni della statura e y_1, y_2, \dots, y_m quelle del peso. Anche in forza di precedenti osservazioni su collettività “analoghe”, si sia costruita la distribuzione di probabilità congiunta $F(x, y)$ descritta da una tabella quale quella di *figura 5*. Allora p_{hk} è la probabilità che la statura X dell’individuo scelto a caso misuri x_h cm e il peso Y sia di y_k kg. La $p_{h.}$ è la probabilità che un individuo scelto a caso abbia statura pari a x_h cm; la $\frac{p_{hk}}{p_{.k}}$ è la probabilità che sia x_h cm la statura di un individuo il cui peso è y_k kg.

II.3 Valori caratteristici di una distrib. bidimensionale.

Sia (X, Y) una v.a. doppia e sia $F(x,y)$ la funzione di ripartizione della distribuzione congiunta di X, Y . Consideriamo poi la v.a. $Z = g(X, Y)$, essendo $g(x, y)$ una funzione reale, misurabile, delle variabili reali x, y .

Manifestamente la distribuzione della v.a. Z è deducibile dalla distribuzione congiunta delle X, Y per tramite della funzione g (che trasforma insiemi del piano in insiemi sull'asse reale). In vista di valutare la speranza matematica di Z dovremmo pertanto costruire dapprima la detta distribuzione; il che, lo si intende agevolmente, può comportare notevoli difficoltà, [dipendenti, in genere, dalla natura della funzione $g(x, y)$]. Ciò però non è necessario perchè quella speranza matematica può essere costruita direttamente sulla distribuzione congiunta di X, Y .

Precisamente: la distribuzione sia discreta e la serie

$$\sum_{h,k} [g(x_h, y_k)] p_{hk}$$

risulti convergente.

Allora è, qui *per definizione*:

$$E(Z) = E[g(X, Y)] = \sum_{h,k} g(x_h, y_k) p_{hk} \quad [2.1].$$

Analogamente, se la $F(x, y)$ è dotata di densità ed esiste finito l'integrale

$$\int_{-\infty-\infty}^{+\infty+\infty} |g(x, y)| f(x, y) dx dy,$$

la speranza matematica di Z è data, per definizione, da :

$$E(Z) = E[g(X, Y)] = \int_{-\infty-\infty}^{+\infty+\infty} g(x, y) f(x, y) dx dy . [2.2]$$

Ripetiamo che le [2.1], [2.2] sono definizioni ma che è possibile provare che i numeri calcolati a secondo membro delle [2.1], [2.2] sono, di fatto, la speranza matematica della v.a. Z

quale si troverebbe ragionando sulla distribuzione unidimensionale della stessa.

Ciò posto, cominciamo col provare che sussiste la già annunciata *proprietà additiva della speranza matematica*.

Precisamente, posto $g(x, y) = x + y$, troviamo, nell'ipotesi - ad esempio - di una distribuzione dotata di densità:

$$\begin{aligned} E[X + Y] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + y) f(x, y) dx dy = \int_{-\infty}^{+\infty} x f(x, y) dx dy + \int_{-\infty}^{+\infty} y f(x, y) dx dy = \\ &= \int_{-\infty}^{+\infty} x \left[\int_{-\infty}^{+\infty} f(x, y) dy \right] dx + \int_{-\infty}^{+\infty} y \left[\int_{-\infty}^{+\infty} f(x, y) dx \right] dy = \\ &= \int_{-\infty}^{+\infty} x f_1(x) dx + \int_{-\infty}^{+\infty} y f_2(y) dy = E(X) + E(Y). \end{aligned}$$

Analoga la distribuzione nel caso di una distribuzione discreta.

Consideriamo ancora la v.a. $Z = X + Y$ e, nell'ipotesi che la distribuzione congiunta di X, Y sia dotata di densità, ricerchiamone la funzione di ripartizione, cioè la funzione $F(z)$

definita dalla:

$$F(z) = \text{Prob}\{X + Y \leq z\}.$$

Si tratta di ricercare la massa che la distribuzione congiunta assegna al semipiano $x + y \leq z$ (vedi figura 6).

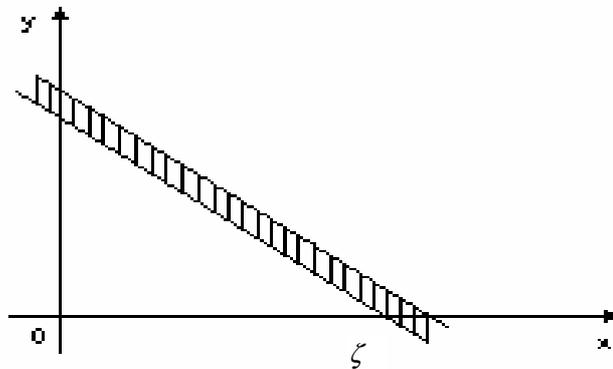


Figura 6

Sia $f(x, y)$ la densità della distribuzione congiunta. La massa contenuta nella striscia tra le rette $x + y = \zeta$ e $x + y = \zeta + d\zeta$ è, a meno di infinitesimi di ordine superiore al primo rispetto a $d\zeta$ uguale a

$$d\zeta \int_{-\infty}^{+\infty} f(x, \zeta - x) dx$$

e ne viene la:

$$F(z) = \int_{-\infty}^z d\xi \int_{-\infty}^{+\infty} f(x, \xi - x) dx . \quad [2.3]$$

La densità della distribuzione unidimensionale di Z è poi la:

$$f(z) = \int_{-\infty}^{+\infty} f(x, z - x) dx . \quad [2.4]$$

Notiamo ora che, determinata la densità [2.4], la speranza matematica di $Z = X + Y$ andrebbe calcolata secondo la:

$$E(Z) = \int_{-\infty}^{+\infty} z f(z) dz = \int_{-\infty}^{+\infty} z \left[\int_{-\infty}^{+\infty} f(x, z - x) dx \right] dz . \quad [2.5]$$

La speranza matematica [2.5] è quella già calcolata secondo la definizione di $E(X + Y)$.

Abbandonando ora il caso particolare $g(x, y) = x + y$ rivolgeremo la nostra attenzione al calcolo delle speranze matematiche di altre variabili aleatorie $g(X, Y)$.

Interessano, in particolare, le speranze matematiche delle

v.a. $Z = g(X, Y) = X^h \times Y^k$, prodotti di potenze intere dell v.a. X e Y .

La quantità $E(X^h \times Y^k)$ è detta anche *momento di ordine* $h+k$ (della distribuzione doppia) e verrà indicata col simbolo m_{hk} .

Particolarmente significativa riesce la considerazione dei (due) momenti primi m_{10} , m_{01} e dei (tre) momenti secondi m_{20} , m_{11} , m_{02} .

I momenti primi

$$m_{10} = E(X), \quad m_{01} = E(Y)$$

forniscono le speranze matematiche (o valori medi) di X , rispettivamente Y .

Le indicheremo nel seguito con m_x , rispettivamente m_y .

Come abbiamo già constatato, m_x , m_y sono speranze matematiche di variabili unidimensionali valutate sulle rispettive distribuzioni marginali.

Del pari su quelle distribuzioni sono valutati i momenti secondi

$$m_{20} = \mathbf{E}(X^2) \quad e \quad m_{02} = \mathbf{E}(Y^2)$$

e i momenti centrali secondi (o varianze):

$$\mu_{20} = \text{var}(X) = \mathbf{E}[(X - m_x)^2]$$

$$\mu_{02} = \text{var}(Y) = \mathbf{E}[(Y - m_y)^2].$$

Il momento secondo centrale misto

$$\mu_{11} = \mathbf{E}[(X - m_x)(Y - m_y)],$$

è detto anche *covarianza* delle v.a. X, Y e si indica con il simbolo $\text{cov}(X, Y)$.

In termini dell'interpretazione meccanica, i numeri m_x, m_y forniscono le coordinate del pbaricentro della distribuzione di massa sul piano, i momenti secondi sono momenti di inerzia rispetto ad assi paralleli agli assi coordinati, la covarianza è il momento centrifugo rispetto al baricentro.

Giova ancora annotare la:

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y) \quad [2.6]$$

che fornisce la varianza della somma di due v.a. La [2.6] si

ricava immediatamente dalla:

$$\begin{aligned} \text{var}(X + Y) &= \mathbf{E}\{[(X + Y) - (m_x + m_y)]^2\} = \\ &= \mathbf{E}[(X - m_x)^2] + \mathbf{E}[(Y - m_y)^2] + 2\mathbf{E}[(X - m_x)(Y - m_y)]. \end{aligned}$$

Dicesi funzione caratteristica della v.a. doppia (X, Y) la funzione delle variabili reali u, v ,

$$\varphi(u, v) = \mathbf{E}(e^{i(uX+vY)}).$$

E subito visto che:

a) $\varphi(0,0) = 1$,

b) $|\varphi(u, v)| = |\mathbf{E}(e^{i(uX+vY)})| \leq |\mathbf{E}(e^{i(uX+vY)})| = 1$,

c) $\varphi(-u, -v) = \overline{\varphi(u, v)}$.

Anche nel caso bidimensionale, quando esistano i momenti di ordine $h + k$ esistono le derivate parziali:

$$\frac{\delta^{h+k} \varphi(u, v)}{\delta^h u \delta^k v} = i^{h+k} E(X^h Y^k e^{i(uX+vY)}) ,$$

sicchè

$$m_{hk} = E(X^h Y^k) = \frac{1}{i^{h+k}} \left(\frac{\delta^{h+k} \varphi(u, v)}{\delta^h u \delta^k v} \right)_{0,0} .$$

La $\varphi(u, 0) = \varphi_1(u)$ è poi la funzione caratteristica della distribuzione marginale della X , mentre la $\varphi(0, v) = \varphi_2(v)$ è quella della distribuzione marginale della Y .

Sussiste, ci limitiamo ad affermarlo, l'analogo del teorema di inversione di Lèvy. In altri termini, la conoscenza della $\varphi(u, v)$ permette di risalire a quella della funzione di ripartizione $F(x, y)$ della distribuzione congiunta.

II.4 Variabili aleatorie indipendenti.

Due variabili aleatorie X, Y diconsi *indipendenti* se ogni evento che faccia riferimento alle determinazioni numeriche

dell'una è indipendente stocasticamente da qualunque evento che riguardi le determinazioni dell'altra.

Si riguarderanno, ad esempio, indipendenti le v.a. X “punto realizzato col lancio di un primo dado”, Y “punto realizzato col lancio di un secondo dado”, quando non appaiano ragioni che ci portino a valutare le probabilità delle determinazioni di Y visto X , in modo diverso da quello seguito ignorando il risultato dell'altro lancio (e analogamente scambiando Y con X).

L'ascissa X e l'ordinata Y del punto colpito nel tiro di un bersaglio potranno essere giudicate indipendenti in una prima approssimazione. Si potrà pensare, però, che per colpa di difetti nel congegno di tiro o di atteggiamento del tiratore, i colpi “alti” si accompagnino più volentieri con i colpi “a destra” o simili altre deviazioni dall'indipendenza.

Le v.a. X , Y “ulteriori durate in vita degli individui di una

coppia” difficilmente potranno essere giudicate indipendenti se si tratta di individui apparentati (coniugi, fratelli, fratelli siamesi, ...), ragionevolmente invece, se si tratta di individui appartenenti a collettività diverse.

Siano ora $F(x,y)$, $F_1(x)$, $F_2(y)$ la funzione di ripartizione della distribuzione congiunta, rispettivamente quelle delle distribuzioni marginali di X e di Y .

Condizione necessaria e sufficiente perché X , Y siano indipendenti è che risulti, per ogni x e y ,

$$F(x, y) = F_1(x) \times F_2(y) .$$

Ciò comporta nel caso discreto le

$$p_{ij} = p_i \times p_j, \quad i, j = 1, 2, \dots,$$

e nel caso delle distribuzioni dotate di densità la

$$f(x, y) = f_1(x) \times f_2(y) .$$

Notiamo subito che se X e Y sono indipendenti risulta:

$$E(XY) = E(X)E(Y).$$

Nel caso di distribuzioni discrete si ha, infatti,

$$E(XY) = \sum_{h,k} x_h y_k p_{hk} = \sum_{h,k} x_h y_k p_h \cdot p_k = \sum_h x_h p_h \cdot \sum_k y_k p_k = E(X)E(Y).$$

Analoga la dimostrazione nel caso delle distribuzioni dotate di densità.

Dopo di ciò, in termini della funzione caratteristica, la condizione di indipendenza stocastica tra X e Y è dichiarata dalla

$$\varphi(u, v) = \varphi_1(u) \times \varphi_2(v), \quad [2.7]$$

come segue dalla

$$E(e^{i(uX+vY)}) = E(e^{iuX} \cdot e^{ivY}) = E(e^{iuX}) \cdot E(e^{ivY}).$$

Si badi però che la condizione $E(XY) = E(X)E(Y)$ necessaria per l'indipendenza, potrebbe essere soddisfatta anche in condizioni di dipendenza stocastica.

Notando che si ha:

$$\begin{aligned} \text{cov}(X, Y) &= \mathbf{E}[(X - m_x)(Y - m_y)] = \\ &= \mathbf{E}(XY - m_x Y - m_y X + m_x m_y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y), \end{aligned}$$

la condizione che la speranza matematica del prodotto sia uguale al prodotto delle speranze matematiche comporta l'annullarsi della covarianza.

Le due variabili X , Y diconsi in tale situazione, *non correlate* o anche *ortogonali* (sicchè, in particolare, due v.a. indipendenti sono non correlate mentre non è vero il viceversa).

La condizione [2.7] per le funzioni caratteristiche, è invece, ci limitiamo ad affermarlo, condizione necessaria e sufficiente per l'indipendenza.

Osserviamo qui quanto segue, a proposito della somma, Z , di due v.a. X , Y indipendenti. Dette $\varphi_Z(u)$ la funzione caratteristica di $Z = X + Y$ e $\varphi_X(u)$ e $\varphi_Y(u)$ quelle degli addendi, è:

$$\varphi_Z(u) = E[e^{iu(X+Y)}] = E[e^{iuX} \cdot e^{iuY}] = E(e^{iuX}) \cdot E(e^{iuY}) = \varphi_X(u) \cdot \varphi_Y(u).$$

Dunque: la funzione caratteristica della somma di due v.a. indipendenti è il prodotto delle loro funzioni caratteristiche.

Per quanto riguarda poi la funzione di ripartizione di Z si osservi che ora le [2.3], [2.4] divengono rispettivamente:

$$F(z) = \int_{-\infty}^z d\xi \int_{-\infty}^{+\infty} f_1(x) f_2(\xi - x) dx, \quad [2.3']$$

$$f(z) = \int_{-\infty}^{+\infty} f_1(x) f_2(z - x) dx. \quad [2.4']$$

Pertanto, quando la distribuzione congiunta di X e Y indipendenti è dotata di densità, la densità della somma è il *prodotto convulotorio* delle densità degli addendi; con tale nome si designa infatti l'integrale a secondo membro della [2.4'].

Osserviamo infine che, nel caso di indipendenza o non correlazione la [2.6] si riduce alla:

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) \quad [2.6']$$

che ci dice che allora la varianza è additiva.

II.5 Correlazione. - Regressione. - Regressione lineare.

Abbandoneremo ora il caso particolare di indipendenza tra le v.a. X, Y (caso piuttosto limite) per studiare la correlazione che, in generale, intercede tra due v.a. . Abbiamo già detto che quando risulti $cov(X, Y) = 0$ le due v.a. diconsi non correlate o ortogonali; nel caso opposto la correlazione tra esse dicesi positiva o negativa a seconda del segno della covarianza. Onde misurare, poi, l'entità di tipi particolari di correlazione tra X e Y possono introdursi vari indici.

Soffermeremo la nostra attenzione sull'*indice di correlazione lineare (di Bravais)* definito nel modo seguente:

$$\rho = \frac{cov(X, Y)}{\sigma(X)\sigma(Y)} .$$

E', intanto, evidente che ρ è un "indice" cioè una grandezza

adimensionale, essendo, la covarianza di X, Y misurata nelle stesse unità del prodotto $\sigma(X)\sigma(Y)$ e gli s.q.m. secondo le unità di misura di X , rispettivamente Y .

In generale risulta:

$$|\rho| \leq 1 . \quad [2.8]$$

Consideriamo, infatti, la v.a. $\lambda X + \mu Y$ con λ, μ parametri reali, supponendo per semplicità formale $E(X) = E(Y) = 0$.

Valutiamone la varianza.

E':

$$\begin{aligned} \text{var}(\lambda X + \mu Y) &= E[(\lambda X + \mu Y)^2] = E[(\lambda X)^2] + 2E[\lambda X \mu Y] + E[(\mu Y)^2] = \\ &= \lambda^2 \text{var}(X) + 2\lambda\mu \text{cov}(X, Y) + \mu^2 \text{var}(Y). \end{aligned}$$

La varianza di una v.a. è non negativa, sicchè la forma quadratica nelle variabili reali λ, μ ,

$$\lambda^2 \text{var}(X) + 2\lambda\mu \text{cov}(X, Y) + \mu^2 \text{var}(Y), \quad [2.9]$$

è definita positiva o semidefinita positiva.

Ciò implica, com'è ben noto, la

$$|\text{cov}(X, Y)| \leq \sigma(X)\sigma(Y)$$

ovvero la [2.8].

Osserviamo ancora che la [2.9] è demidefinita positiva allora e allora soltanto che $\rho = \pm 1$. In quella condizione è $E[(\lambda X + \mu Y)^2] = 0$ oltre che per $\lambda = \mu = 0$ per tutte le coppie di numeri λ, μ il cui rapporto è una costante (per tutti i punti del piano λ, μ appartenenti ad una retta passante per l'origine).

Sicchè allora e allora soltanto che $\rho = \pm 1$ risulta nulla la varianza della v.a. $kX - Y$ (avendo posto $k = -\lambda/\mu$) e avremo quindi, a meno di un evento di P -misura nulla,

$$\text{Prob}\{kX - Y \neq 0\} = 0.$$

La v.a. $kX - Y$ coincide, a meno di quell'evento con la sua speranza matematica che, per ipotesi, è nulla.

In altri termini: allora e allora soltanto che ρ assume uno dei

suoi valori estremi, ± 1 , la dipendenza stocastica tra X e Y è addirittura una dipendenza funzionale di tipo particolare (lineare): $Y = kX$.

Il significato dell'indice di correlazione, ρ , appare anche evidenziato dalle considerazioni che ora svolgeremo e che riguardano la nozione di regressione (propriamente di *regressione lineare*) tra due v.a.

con riferimento alla coppia (X, Y) si fissi una determinazione, diciamola y della v.a. Y , e si determini, in corrispondenza, la speranza matematica $E(X/y)$ della v.a. X subordinatamente all'ipotesi $Y = y$.

Al variare di y , il valor medio $E(X/y)$ descrive un insieme di punti che viene denominato *curva di regressione di X su Y* .

Quanto precede lasci intendere che trattasi della curva del piano x, y di equazioni parametriche:

$$\begin{cases} x = E(X|y) = m_1(y), \\ y = y. \end{cases}$$

Se la distribuzione è discreta, trattasi - ovviamente - di un insieme di punti isolati che, per estensione di linguaggio, viene ancora chiamato “curva” di regressione. Chiaramente poi se X e Y sono indipendenti, la curva di regressione di X su Y è semplicemente la retta parallela all’asse y , di equazione $x = m_x$ passante per il baricentro della distribuzione.

Analogo il significato di *curva di regressione di Y su X* ; essa è il luogo di equazioni parametriche:

$$\begin{cases} x = x, \\ y = E(Y|x) = m_2(x) \end{cases}$$

e nel caso di indipendenza, trattasi della retta di equazione $y = m_y$ parallela all’asse x .

Per intendere su un esempio il significato delle curve di regressione si ripensi alle v.a. $X =$ statura, $Y =$ peso di un

individuo scelto a caso in una collettività. Le osservazioni compiute sulla collettività abbiano posto in evidenza un insieme di possibili coppie ordinate (x_h, y_k) quali quelle che appaiono nella *figura 7*.

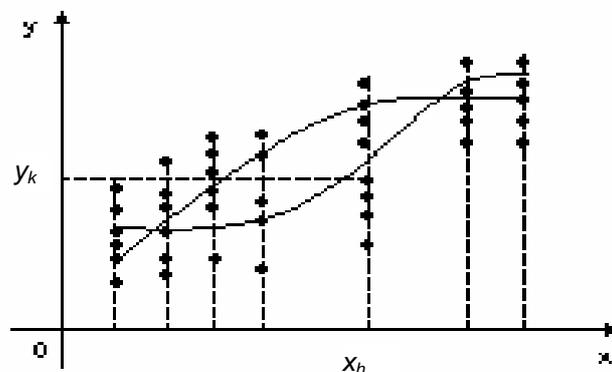


Figura 7

La coppia (x_h, y_k) si sia presentata n_{hk} volte sulle n osservazioni (sicchè n_{hk}/n è la frequenza osservata di quella coppia di determinazioni di X, Y).

Assumendo pari a $1/n$ la probabilità di sorteggiare un determinato individuo dalla collettività (adottando cioè una scelta a caso) valuteremo uguale a n_{hk}/n la probabilità che

quell'individuo abbia statura uguale a x_h cm e peso uguale a y_k kg. La curva di regressione della statura sul peso fornisce ora la statura media in corrispondenza ad ogni singola determinazione del peso (una statura tipica, per così dire, rapportata al peso). Analogamente il significato dell'altra curva di regressione. Occorre soffermare l'attenzione sulla asimmetria delle due nozioni: Una cosa è la statura tipica per pesi assegnati, altra il peso tipico per assegnate stature (le curve di regressione non coincidono se non casualmente).

Le curve di regressione godono di una proprietà particolarmente significativa, che enunceremo con riferimento alla curva di regressione di X su Y . E':

$$E\{[X - m_1(Y)]^2\} = \min_{g(Y)} E\{[X - g(Y)]^2\},$$

ossia la deviazione quadratica media di X da una funzione $g(Y)$ della v.a. Y è minima quando $g(Y)$ è uguale a $m_1(Y)$.

Invero, nell'ipotesi di distribuzione dotata di densità, risulta:

$$\begin{aligned} E[(X - g(Y))^2] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [x - g(x)]^2 f(x, y) dx dy = \\ &= \int_{-\infty}^{+\infty} f_2(y) \left| \int_{-\infty}^{+\infty} [x - g(y)]^2 f(x|y) dx \right| dy \end{aligned}$$

e l'integrale $\int_{-\infty}^{+\infty} [x - g(y)]^2 f(x|y) dx$ è minimo quando $g(y) = m_I(y)$,

poiché fornisce - allora - la varianza di X nella distribuzione subordinata all'ipotesi $Y = y$ (si ricordi allo scopo la proprietà del minimo della varianza di una v.a.; vedi paragrafo I.7).

Le curve di regressione, in generale distinte, si scostano, di norma, dall'andamento rettilineo. Nelle applicazioni pratiche ove questo scostamento è usualmente poco sensibile (o addirittura nullo) si preferisce considerare, in luogo delle curve di regressione, delle *rette di regressione* che le approssimano nel senso che sarà tosto chiarito.

Precisamente: dicesi *retta di regressione lineare di X su Y* quella delle rette di equazione $x = \alpha y + \beta$ in corrispondenza alla quale riesce minima la deviazione quadratica media $E\{[X - (\alpha Y + \beta)]^2\}$, delle distanze misurate sulle parallele all'asse x , tra le determinazioni di X e le corrispondenti determinazioni di $\alpha Y + \beta$. Si rifletta, per intendere il significato della scelta approssimazione, che ad ogni determinazione y della v.a. Y corrisponde una distribuzione, subordinata, della X e si ricordi la proprietà della curva di regressione di X su Y .

Ora è:

$$E\{[X - (\alpha Y + \beta)]^2\} = E[X^2 - 2X(\alpha Y + \beta) + (\alpha^2 Y^2 + 2\alpha\beta Y + \beta^2)] = \\ E(X^2) - 2\alpha E(XY) - 2\beta E(X) + \alpha^2 E(Y^2) + 2\alpha\beta E(Y) + \beta^2$$

e la soluzione del sistema di equazioni ottenuto annullando la derivate parziali rispetto α, β è:

$$\alpha = \rho \frac{\sigma_x}{\sigma_y},$$
$$\beta = m_x - \rho \frac{\sigma_x}{\sigma_y} m_y,$$

sicchè la cercata retta di regressione lineare di X su Y è la retta di equazione

$$x - m_x = \rho \frac{\sigma_x}{\sigma_y} (y - m_y) .$$

In modo analogo si trova che la:

$$y - m_y = \rho \frac{\sigma_y}{\sigma_x} (x - m_x)$$

è l'equazione della retta di regressione lineare di Y su X .

Si osservi che se $\rho=0$ (non correlazione tra X e Y) le rette di regressione sono parallele agli assi coordinati. Se $\rho = \pm 1$ le due rette coincidono (sappiamo già che in quei casi la dipendenza stocastica tra X e Y è un legame di dipendenza lineare).

In ogni caso le rette di regressione passano per il baricentro

della distribuzione. Se $\rho \neq 0$, l'inclinazione sull'asse delle ascisse è positiva o negativa a seconda che sia $\rho > 0$ (correlazione positiva) o $\rho < 0$ (correlazione negativa).

II.6 Ellisse di inerzia. - Ellisse di concentrazione.

Si è visto che le rette di regressione godono della proprietà di rendere minima la speranza matematica del quadrato degli scarti tra una v.a. e una funzione lineare dell'altra. Gli scarti sono misurati in entrambi i casi sulle parallele ad un asse coordinato.

Consideriamo ora una generica retta r del piano (x, y) e indichiamo con Δ_r la “distanza aleatoria dei punti immagine delle possibili coppie di determinazioni della v.a. doppia (X, Y) da r (misurata, quindi, sulla direzione ortogonale alla r)”. Per ogni retta r si calcoli la speranza matematica, $E(\Delta_r^2)$, del quadrato

della distanza.

La retta r che rende minima tale speranza matematica è detta *retta di regressione ortogonale*. Manifestamente tale retta deve passare per il baricentro (m_x, m_y) della distribuzione. (Si riflette sul fatto che $E(\Delta_r^2)$ misura il momento di inerzia della distribuzione rispetto alla retta r).

Detto θ l'angolo che la generica retta r per il baricentro forma col verso positivo dell'asse x , l'equazione della r è:

$$(x - m_x)\sin\theta - (y - m_y)\cos\theta = 0$$

e risulta

$$\begin{aligned} E(D_r^2) &= E\left\{[(X - m_x)\sin\theta - (Y - m_y)\cos\theta]^2\right\} = \\ &= \text{var}(X)\sin^2\theta - 2\rho\sigma(X)\sigma(Y)\sin\theta\cos\theta + \text{var}(Y)\cos^2\theta = \\ &= \sigma_x^2\sin^2\theta - 2\rho\sigma_x\sigma_y\sin\theta\cos\theta + \sigma_y^2\cos^2\theta. \end{aligned}$$

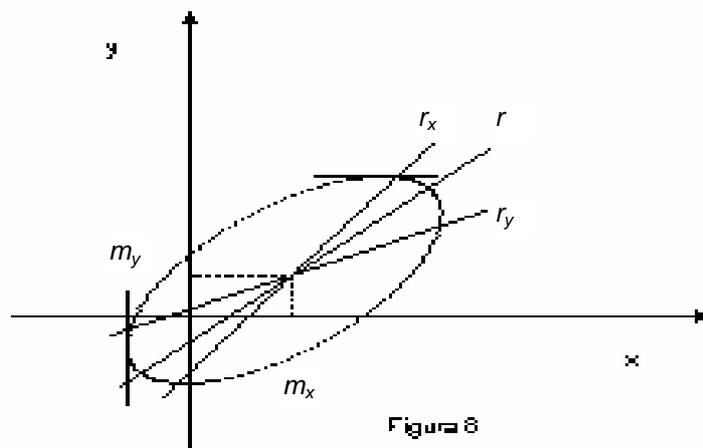
Al variare ora di θ si pensi di staccare, a partire dal baricentro e sui due versi di ogni direzione r , un segmento di lunghezza pari al reciproco di $[E(\Delta_r^2)]^{1/2}$. Il luogo dei punti in tal

modo ottenuto è un'ellisse, detta *ellisse di inerzia*, di equazione

$$\frac{(x - m_x)^2}{\sigma_x^2} - 2\rho \frac{(x - m_x)(y - m_y)}{\sigma_x \sigma_y} + \frac{(y - m_y)^2}{\sigma_y^2} = \frac{1}{\sigma_x^2 \sigma_y^2}$$

La retta di regressione ortogonale [univocamente determinata, salvo il caso $\rho = 0$ e $\sigma^2(X) = \sigma^2(Y)$] è asse maggiore dell'ellisse.

Le rette di regressione lineare sono diametri, ciascuno coniugato con un'asse coordinato. Si veda la *figura 8* ove con r_x è indicata la retta di regressione di X su Y , con r_y la retta di regressione di Y su X e con \bar{r} la retta di regressione ortogonale.



La famiglia di ellissi omotetiche

$$\frac{(x - m_x)^2}{\sigma_x^2} - 2\rho \frac{(x - m_x)(y - m_y)}{\sigma_x \sigma_y} + \frac{(y - m_y)^2}{\sigma_y^2} = c^2, \quad [2.10]$$

cui appartiene per $c = \frac{1}{\sigma_x \sigma_y}$ l'ellisse di inerzia, è per più aspetti

interessante. Si osservi intanto che per $\rho = 0$ (non correlazione)

e per $\sigma_x = \sigma_y$ trattasi di circonferenze (ogni diametro è asse maggiore).

Per $\rho^2 = 1$ le ellissi degenerano su una retta doppia (sappiamo già il significato della condizione: la massa è tutta raccolta su una retta).

Sia $|\rho| < 1$; l'ellisse della famiglia [2.10], con $c^2 = 4(1 - \rho^2)$ gode di una particolare proprietà: è l'ellisse entro la quale si può pensare concentrata in modo uniforme la massa unitaria mantenendo invariati i momenti primi e secondi della distribuzione originaria della v.a. doppia (X, Y) . L'ellisse in

questione è detta *ellisse di concentrazione*.

Bibliografia.

- J. Bass Elements de calcul des probabilitès, Masson,
1962.
- B. De Finetti Teoria della probabilità, Einaudi, 1963.
- M. Cugiani Metodi dell'analisi numerica, Utet, 1967.
- H. Cramèr Mathematical methods of statistics, Princeton
University press, 1946.
- T. S. Ferguson Mathematical statistics, Academic press,
1967.
- B. V. Gnedenko The theory of probability, Chelsea publishing
company, 1962.
- W. Feller An introduction to probability theory and its
applications, vol II, J. Wiley, 1966.
- L. J. Savage The foundations of statistics, J. Wiley, 1954.
- F. Spitzer Principles of random walk, Van Nostrand co.
inc., 1964.

- S. S. Wilks *Mathematical statistics*, J. Wiley, 1962.

Indice.

- Premessa.....pag. 2

- Capitolo I°: Le variabili aleatorie unidimensionali.

I.1 La variabile aleatoria come funzione P-misurabile.....

11

I.2 Funzione di ripartizione di una variabile aleatoria.....

16

I.3 Valori caratteristici di una distribuzione..... 26

I.4 Speranza matematica di una variabile aleatoria..... 30

I.5 Momenti di una variabile aleatoria.....

35

I.6 Varianza. - Disuguaglianza di Bienaymè-Cebicev..... 38

I.7 Indici posizionali. - Mediana, frattili, moda.....

43

- Capitolo II°: Le variabili aleatorie multiple.

II.1 Distribuzioni bidimensionali congiunte.....

49

	II.2 Distribuzioni marginali e distribuzioni subordinate.....	
56		
	II.3 Valori caratteristici di una distribuzione bidimensionale..	64
	II.4 Variabili aleatorie indipendenti.....	
72		
	II.5 Correlazione. - Regressione. - Regressione lineare.....	78
	II.6 Ellisse di inerzia. - Ellisse di concentrazione.....	88
	- Bibliografia.....	93