

**PARTE PRIMA**

**PROBABILITÀ**



## CAPITOLO 1

### LA PROBABILITÀ

Tutti i fenomeni che ci circondano hanno una strana caratteristica. Da un lato essi sono differenti l'uno dall'altro, dall'altro lato hanno sempre qualcosa che li accomuna. Probabilmente ciò deriva dalla particolare prospettiva nella quale noi li osserviamo. Per esempio un gruppo di cavie per uno sperimentatore rappresenta una successione di copie di uno stesso ceppo, ma egli è interessato agli aspetti comuni e spera che quelli individuali siano il meno rilevanti possibile, proprio perché vuole “replicare” su tali cavie un certo esperimento. È possibile che un osservatore esterno osservi gli abitanti di questo mondo con lo stesso spirito di quel ricercatore, ignorando le differenze tra gli individui (di cui noi magari andiamo molto orgogliosi) proprio perché è interessato agli aspetti comuni che sottostanno alla variabilità.

Anche il più semplice degli esperimenti, se si scende ad un grado di osservazione opportunamente preciso, ha un grado di variabilità nel risultato che induce aleatorietà. Proprio per questo non si può parlare in assoluto di uno stesso fenomeno replicato, perché il fenomeno o è un altro “simile” o è ripetuto in un istante successivo. È l'osservatore che giudica quei due fenomeni come due repliche di qualcosa di comune che non è né l'uno né l'altro, ma solo un modello teorico (o idea) che egli ha in mente.

Il più semplice dei fenomeni che possono essere rappresentati attraverso delle leggi o modelli stocastici è il risultato del *lancio di una moneta*. Ammettendo di avere una moneta equa, ossia che non abbia una faccia più pesante dell'altra (si pensi all'astrazione che si fa in fisica sul comportamento dei gas o dei gravi), e che sia impossibile che la moneta resti in equilibrio sul

bordo o che si perda, i due possibili esiti dell'esperimento (o *evento*) sono *testa* (*T*) e *croce* (*C*). È risaputo che per quanta cura si metta nel collocare la moneta sulla mano e nell'imprimerle una certa forza, il risultato è imprevedibile, per quanto le condizioni iniziali siano determinate esattamente. Dall'altro lato è pure esperienza comune che in queste condizioni, ripetendo nel tempo l'esperimento, i risultati *nel complesso* mostrano una regolarità che non può essere trascurata; in particolare non si verificano mai casi in cui vi è una eccessiva predominanza di un risultato di tipo *T* o *C* rispetto ad un altro (anzi, quando ciò si verifica, ad essere messo in discussione non è il modello teorico, bensì l'equità della moneta). Pertanto, se il *singolo* risultato è imprevedibile, è "meno imprevedibile" il complesso di un insieme di risultati basato su un numero più ampio di *repliche*.

## 1. Probabilità

Dopo avere definito i possibili eventi o gli esiti possibili di un esperimento aleatorio, è immediato chiedersi qual è la *probabilità* che ha ciascuno dei possibili eventi di verificarsi, infatti il concetto di probabilità è usato correntemente.

Nel diciassettesimo secolo, quando cominciarono a diffondersi i giochi di carte, si cercò di formalizzare la probabilità come il *numero di casi favorevoli sul numero di casi possibili, giudicati questi tutti equiprobabili*. Ora, a parte la tautologia di fondare la probabilità sull'equiprobabilità, come ci si deve regolare quando i casi sono manifestamente non equiprobabili?

Successivamente con lo sviluppo delle scienze osservative si definì la probabilità di un esito in modo *statistico*, come limite del numero di esperimenti che hanno condotto a quel risultato rapportato al numero *sufficientemente elevato* di esperimenti provati. Anche qui però non mancano le difficoltà teoriche e pratiche. Primo, si fa riferimento alla citata ripetibilità di un esperimento, con tutte le implicazioni di arbitrarietà connesse, lasciando così non risolti i problemi (la maggior parte) in cui l'esito è incerto ma l'esperimento non è affatto ripetibile (per es., l'esito di una operazione chirurgica, che è basato non solo su statistiche di operazioni effettuate su altri, ma anche sullo stato del paziente). Inoltre dal punto di vista pratico quanti devono essere gli esperimenti prima di potere affermare la convergenza al limite come praticamente raggiunta? Ed inoltre, potendo per effetto del caso i casi favorevoli essere molto più (o meno) numerosi di quanto la probabilità non "prescriva", la valutazione di probabilità non è certa e quindi si deve parlare di una *probabilità della probabilità*?

Nel secolo appena trascorso si è andata affermando, non senza forti contrasti ed opposizioni, la concezione *soggettivista*, di cui uno dei più grandi

rappresentanti è l'italiano Bruno De Finetti [8], concezione che rinuncia esplicitamente alla possibilità di esprimere su qualsiasi evento una probabilità “oggettiva”, cioè tale che sia un valore comune per chiunque la esprima; in realtà la probabilità, quantificando una incertezza dovuta a carenza di conoscenza, dipende dallo stato di informazione del soggetto che la esprime. Anche a questa concezione non sono mancate critiche sia sulle pretese implicazioni filosofiche connesse con questa scelta (l'aleatorietà è solo una nostra carenza di conoscenza o è una caratteristica intrinseca del fenomeno?), sia sulla reale utilità che essa comporta.

1.1 Quale che sia la concezione della probabilità scelta e in qualunque modo una probabilità possa essere formulata, vi sono alcune regole che devono essere rispettate, affinché si possa applicare la probabilità ad un insieme di possibili realizzazioni dell'esperimento aleatorio, o *eventi*, che costituiscono lo *spazio campionario*:  $\mathcal{A} = \{A_1, A_2, \dots, A_n\}$  (in questo momento assumiamo che  $n$  sia finito). Anzi è possibile definire la probabilità attraverso un'impostazione *assiomatica*, che ignora del tutto la scelta delle probabilità assegnate ad i vari eventi, concentrandosi esclusivamente sulle *regole* che qualunque assegnazione deve rispettare.

Poiché la probabilità è una *misura* su cui faremo alcune operazioni algebriche, quali:

- data la probabilità che un evento  $A$  accada,  $\mathcal{P}(A)$ , determinare la probabilità che *non* accada,  $\mathcal{P}(\bar{A})$ ,
- data le probabilità che accadano i due eventi  $A_1$  e  $A_2$ ,  $\mathcal{P}(A_1)$  e  $\mathcal{P}(A_2)$ , determinare la probabilità che ne accada *almeno uno* di essi  $\mathcal{P}(A_1 \cup A_2)$ ,
- oppure che accadano *entrambi*,  $\mathcal{P}(A_1 \cap A_2)$ ,

dal primo spazio si deve definire un altro spazio, cui appartengano non solo tutti gli eventi che appartengono ad  $\mathcal{A}$ , ma anche tutte le possibili *unioni* ed *intersezioni* fra essi e relativi *eventi complementari* (o negazioni  $\bar{A}_i$ ), compreso l'evento certo  $\Omega = \{A_1 \cup A_2 \cup \dots \cup A_n\}$ . Cioè definiamo uno *spazio*  $\mathcal{B}$  dove valgono le seguenti proprietà:

1. l'evento  $\Omega \in \mathcal{B}$ ;
2. se  $A_i \in \mathcal{B}$ , allora anche  $\bar{A}_i \in \mathcal{B}$ ;
3. se  $A_i \in \mathcal{B}$  e  $A_j \in \mathcal{B}$ , allora anche  $(A_i \cup A_j) \in \mathcal{B}$ .

A questo punto possiamo definire un'*algebra* che, effettuando le tre operazioni di negazione, unione ed intersezione su qualsiasi insieme di elementi appartenenti a  $\mathcal{B}$ , ci garantisca di ottenere sempre elementi ancora all'interno di  $\mathcal{B}$ .

Ricordando le leggi di De Morgan:

- $\overline{A_i \cup A_j} = \bar{A}_i \cap \bar{A}_j$ ,
- $\overline{A_i \cap A_j} = \bar{A}_i \cup \bar{A}_j$ ,

si può dimostrare che vale anche la:

4. se  $A_i \in \mathbf{B}$  e  $A_j \in \mathbf{B}$ , allora anche  $(A_i \cap A_j) \in \mathbf{B}$ .

**Esercizio 1.1.** Si dimostri la proprietà 4.

Su uno spazio  $\mathbf{B}$  di questo genere si può definire una particolare misura che viene detta *probabilità*, avente le seguenti proprietà:

*Assiomi della probabilità:*

- I. la probabilità è una misura *non negativa*:  $\mathcal{P}(A) \geq 0$ ;
- II. la misura della probabilità è *normalizzata all'unità*:  $\mathcal{P}(\Omega) = 1$ ;
- III. la probabilità dell'unione di eventi incompatibili (mutuamente escludentisi) è *additiva*. Ossia, indicando con  $\emptyset$  l'insieme vuoto, si ha:

se  $A_i \cap A_j = \emptyset$  per ogni  $(i \neq j)$ , allora  $\mathcal{P}(A_i \cup A_j) = \mathcal{P}(A_i) + \mathcal{P}(A_j)$ .

**Esercizio 1.2.** Si dimostri che  $\mathcal{P}(\bar{A}) = 1 - \mathcal{P}(A)$ .

**Esercizio 1.3.** Si dimostri che se  $A_i \cap A_j = \emptyset$  per ogni  $i \neq j$ , allora se:

$$\mathcal{P}\left(\bigcup_{i=1}^m A_i\right) = \sum_{i=1}^m \mathcal{P}(A_i), \text{ con } m < n, \text{ ed in particolare } \mathcal{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathcal{P}(A_i) = 1.$$

Se  $n \rightarrow \infty$ , l'algebra deve diventare una  $\sigma$ -algebra, cioè la terza proprietà sarà:

se  $A_1, A_2, \dots, A_i, \dots \in \mathbf{B}$ , allora  $(A_1 \cup A_2 \cup \dots \cup A_i \cup \dots) \in \mathbf{B}$ ,

ed il terzo assioma, ancora per eventi a due a due incompatibili, diventa:

$$\mathcal{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathcal{P}(A_i).$$

In questo caso si parla di *completa additività*.

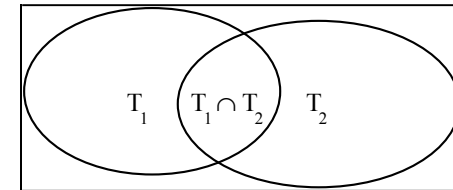
1.2 Date le precedenti definizioni, possiamo dire che si può parlare di assegnazione di una misura di probabilità ad un insieme di eventi, quando è definita la terna

$$\{\mathbf{A}, \mathbf{B}, \mathcal{P}\}.$$

In questo modo siamo in condizioni di assegnare una probabilità, non solo agli eventi che compongono  $\mathbf{A}$ , ma anche a tutti quelli che ne possono derivare per negazione, unione ed intersezione di essi.

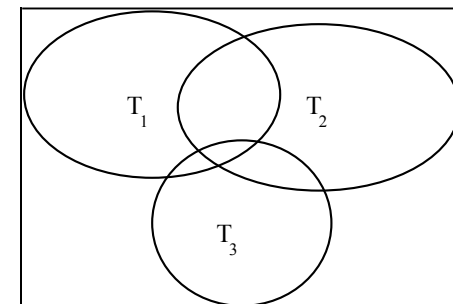
Nel caso in cui il nostro insieme  $A$  è costituito da elementi tra loro *incompatibili* ed *esaustivi*, cioè che costituiscono una *partizione* dello spazio campionario, le cose sono abbastanza semplici. Infatti in questo caso la probabilità dell'unione di qualunque coppia, o terna, o quaterna ... di eventi è data dalla somma delle probabilità dei singoli eventi; mentre la probabilità di una qualunque intersezione è nulla. Se invece gli eventi non costituiscono una partizione, si può operare in due modi: determinare qual è la probabilità delle intersezioni coinvolte, o costruire lo spazio degli *eventi elementari*.

Per esempio, sia dato il gioco a testa e croce con due lanci; si vuole determinare la probabilità di ottenere *almeno una testa*: cioè la probabilità dell'unione dei due eventi  $T_1$  e  $T_2$ . Supponendo che  $\mathcal{P}(T_1) = \mathcal{P}(T_2) = 0.5$ , non è vero che  $\mathcal{P}(T_1 \cap T_2) = 0$ , cioè è del tutto ammissibile che esca testa al primo e al secondo lancio, anzi si ammetta che tale probabilità valga 0.25. In questo caso la  $\mathcal{P}(T_1 \cup T_2) = \mathcal{P}(T_1) + \mathcal{P}(T_2) - \mathcal{P}(T_1 \cap T_2)$ . Di ciò ci si può rendere conto attraverso i diagrammi di Venn, che associano la probabilità di un evento all'area racchiusa entro la relativa "patata".



Infatti con  $\mathcal{P}(T_1) + \mathcal{P}(T_2)$  abbiamo sommato due volte l'area relativa a  $T_1 \cap T_2$  e quindi essa va eliminata. Però più il numero degli elementi non incompatibili che si uniscono è elevato, più la situazione si complica. Per esempio, se abbiamo tre eventi non incompatibili si ha:

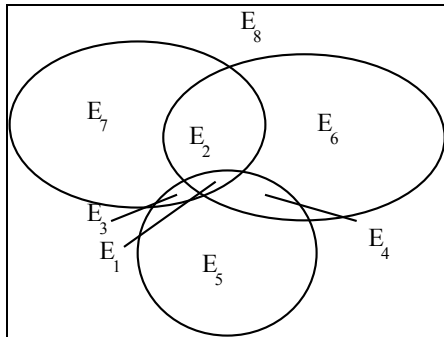
$$\begin{aligned} \mathcal{P}(T_1 \cup T_2 \cup T_3) &= \\ &= \mathcal{P}(T_1) + \mathcal{P}(T_2) + \mathcal{P}(T_3) + \\ &- \mathcal{P}(T_1 \cap T_2) - \mathcal{P}(T_1 \cap T_3) - \mathcal{P}(T_2 \cap T_3) + \\ &+ \mathcal{P}(T_1 \cap T_2 \cap T_3) \end{aligned}$$



Infatti come si vede occorre sottrarre le probabilità delle intersezioni a due a due, che nella somma della prima riga sono state contate due volte; ma in questo modo l'intersezione a tre, che nella prima riga era stata contata tre volte, nella seconda viene sottratta tre volte e quindi bisogna aggiungerla nuovamente con la terza riga. È ovvio che una generalizzazione di questa procedura porta ben presto ad un calcolo impraticabile (sottrarre le probabilità delle intersezioni a quattro a quattro, sommare quelle a cinque a cinque, ...).

1.3 Un metodo alternativo consiste nella determinazione degli *eventi elementari*, cioè la costruzione di un nuovo spazio di eventi, ottenuto da quello precedente, in cui questi siano *incompatibili* ed *esaustivi*. Nella figura si vede che ciò si ottiene considerando gli eventi:

$$\begin{aligned} E_1 &= T_1 \cap T_2 \cap T_3 & E_2 &= T_1 \cap T_2 \cap C_3 & E_3 &= T_1 \cap C_2 \cap T_3 & E_4 &= C_1 \cap T_2 \cap T_3 \\ E_5 &= C_1 \cap C_2 \cap T_3 & E_6 &= C_1 \cap T_2 \cap C_3 & E_7 &= T_1 \cap C_2 \cap C_3 & E_8 &= C_1 \cap C_2 \cap C_3 \end{aligned}$$



che sono in numero di  $2^3$ . Infatti occorre considerare l'*insieme delle parti*, costituito dal numero di modi di presentarsi di ciascun evento  $m$  (in questo caso pari a 2:  $T$  e  $C$ ) elevato al numero di eventi  $n$ .

Tale numero si ottiene sommando quanti eventi elementari si ottengono con nessuna  $C$ , con una  $C$ , con due  $C$ , ... con  $n$   $C$ . Tali eventi sono il numero di *combinazioni* di  $n$  oggetti a zero a zero, a uno a uno, ... , a  $n$  a  $n$ .

Tali numeri, è noto dal calcolo combinatorio, sono i coefficienti binomiali  ${}_k C_n = N! / [(N - k)! k!]$ , la cui somma è pari allo sviluppo del binomio  $(1 + 1)^n$ :

$$\begin{aligned} \binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{n} &= \binom{n}{0} 1^n 1^0 + \binom{n}{1} 1^{n-1} 1^1 + \dots + \binom{n}{n} 1^0 1^n = \\ &= (1 + 1)^n = 2^n. \end{aligned}$$



## 2. Coerenza

Abbiamo analizzato a fondo la proprietà 3.: se  $A_i \cap A_j = \emptyset$  per ogni  $i \neq j$ , allora  $\mathcal{P}(A_i \cup A_j) = \mathcal{P}(A_i) + \mathcal{P}(A_j)$ .

Per quanto riguarda la proprietà 1.:  $\mathcal{P}(A) \geq 0$ , essa può essere accettata senza difficoltà anche intuitivamente, osservando che essa si ritrova in tutte le concezioni probabilità citate nel paragrafo precedente.

È la proprietà 2.:  $\mathcal{P}(\Omega) = 1$  che necessita di qualche ulteriore commento.

2.1 Secondo la concezione classica della probabilità se  $\mathcal{A} = \{A_1, A_2, \dots, A_k\}$  è un insieme di eventi incompatibili e  $\mathcal{P}(A_i) = n_i / n$  (numero di casi favorevoli  $n_i$  sul numero di casi possibili  $n$  se equiprobabili, con  $\sum_i n_i = n$ ), si ricava:

$$\sum_i \mathcal{P}(A_i) = \sum_i n_i / n = 1.$$

Altrettanto secondo la concezione frequentista, se  $\mathcal{P}(A_i) = f_i$ , essendo  $f$  il limite del rapporto  $n_i / n$  realmente osservato su  $n$  prove.

In queste condizioni se la quota di ingresso per una scommessa viene pagata proporzionalmente a tale probabilità, si ha una scommessa *equa*, cioè non vi è una vincita certa. Infatti, se  $S$  è l'importo per partecipare ad una scommessa avente come vincita una quantità unitaria (per esempio 1 euro) su  $n$  prove ripetute ci *aspettiamo* che  $m$  di queste diano esito positivo, ed  $n - m$  negativo, ottenendo alla fine delle  $n$  scommesse un ricavo pari a  $m$ , mentre il costo dell'operazione cioè delle  $n$  scommesse ripetute sarà stato pari ad  $nS$ . Da qui, se si vuole che alla fine non vi sia né una perdita, né una vincita, si dovrà porre  $nS = m$  e quindi  $S$  dovrà essere fissato pari ad  $m / n$ .

Si noti bene che ciò non vuol dire che non vi sarà un vincitore ed un perdente, poiché non è affatto detto che su un numero di prove *limitato* il numero di successi sarà pari proprio a  $m$ .

**Esercizio 1.4.** Si generalizzi la proprietà appena esposta ad uno spazio  $\mathcal{A}$  con  $k$  eventi incompatibili.

2.2 L'impostazione soggettivista invece, ribalta la prospettiva, fissando come condizione la coerenza e ricavando l'additività come risultato. Infatti si parte dall'equazione di coerenza costruita sommando i costi di partecipare a ciascuna delle scommesse ed eguagliando tale somma all'importo della vincita:  $S_1 + S_2 + \dots + S_n = V$ .

Se così non fosse infatti saremmo nel caso detto "scommessa olandese" in cui è possibile ottenere una combinazione di scommesse che porta ad una vincita o ad una perdita certa. Per esempio nella partita a testa e croce se scommettere su  $T$  costa 50 Centesimi e su  $C$  40 Cent. con vincita pari a 1 €, scommettendo sia su  $T$  che su  $C$  si ha una vincita certa di 10 Cent., qualunque

sia l'esito del lancio. Se invece scommettere su  $T$  costa 50 Cent. e su  $C$  60 Cent. con vincita pari a 1 €, scommettendo sia su  $T$  che su  $C$  si ha una perdita certa di 10 Cent. Per costringerci ad assumere una posizione coerente si potrebbe ipotizzare di formulare dapprima la scommessa e poi essere costretti a prendere la posizione del banco o della punta: in questo caso per non andare incontro ad una perdita certa dobbiamo formulare una scommessa coerente. Si noti che nella realtà tutte le scommesse non coerenti, formulate per esempio da allibratori, non consentono di assumere la posizione del banco.

La  $\mathcal{P}(A_i)$  è qui per definizione  $S_i / V$ , da cui  $\sum_i \mathcal{P}(A_i) = 1$ .

### 3. Indipendenza

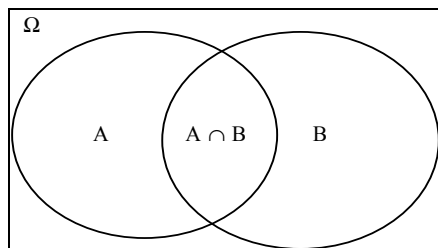
Occorre definire che cosa si intende per un evento  $A$  subordinato ad un altro  $B$ . Ammettiamo di avere espresso la  $\mathcal{P}(A)$  e successivamente siamo venuti a conoscenza che si è verificato l'evento  $B$ . Occorre esprimere la valutazione di probabilità su  $A$  alla luce della nuova conoscenza su  $B$ . Indichiamo questa nuova probabilità come  $\mathcal{P}(A|B)$ . In particolare si potrebbe verificare il caso che  $\mathcal{P}(A) = \mathcal{P}(A|B)$ , in questo caso gli eventi  $A$  e  $B$  si dicono *indipendenti*. Altrimenti i due eventi si dicono *dipendenti*.

Per esempio,  $A$  sia la vittoria della squadra  $X$  sulla squadra  $Y$  in un certo incontro,  $B$  sia il fatto che ad un certo punto del gioco, la squadra  $X$  è in vantaggio; naturalmente la valutazione di probabilità sul fatto che  $X$  alla fine vinca sarà maggiore rispetto all'inizio della partita. Oppure  $A$  come prima e  $B$  che ad un certo punto piove; se nessuna delle due squadre è favorita da questo evento, si può supporre che la  $\mathcal{P}(A)$  resti inalterata.

Come quarto assioma della probabilità si ha:

$$\text{IV. } \mathcal{P}(A|B) = \mathcal{P}(A \cap B) / \mathcal{P}(B).$$

Possiamo giustificare questo assioma considerando che il verificarsi di  $B$  ha ristretto tutto lo spazio  $\Omega$  a quella porzione contenuta in  $B$  e quindi ciò che di  $A$  può verificarsi è la sua parte che interseca  $B$ , mentre la parte rimanente non può verificarsi più. D'altra parte ora la massa di probabilità unitaria non è più quella riferita al vecchio  $\Omega$ , bensì si è ristretta a  $B$ .



Quindi la condizione di indipendenza è caratterizzata da:

$$\mathcal{P}(A \cap B) = \mathcal{P}(A) \mathcal{P}(B).$$

**Esercizio 1.5.** Si dimostri che se  $\mathcal{P}(A) = \mathcal{P}(A | B)$  allora:

$$a) \mathcal{P}(A) = \mathcal{P}(A | \bar{B}) \text{ e } b) \mathcal{P}(\bar{A}) = \mathcal{P}(\bar{A} | \bar{B}).$$

*Suggerimento:* Si consideri che: a)  $(A \cap B) \cup (A \cap \bar{B}) = A$  e che  
b)  $\mathcal{P}(\bar{A} \cap \bar{B}) = 1 - \mathcal{P}(A \cup B)$ .

3.1 Estendendo la condizione di indipendenza a insiemi di più di due eventi, si può vedere come l'indipendenza *non* gode della proprietà transitiva.

**Esercizio 1.6.** Si consideri uno spazio partizionato in quattro eventi equiprobabili:  $E_1, E_2, E_3, E_4$ :

a) dati i tre eventi:  $A = E_1 \cup E_2, B = E_1 \cup E_3$  e  $C = E_2 \cup E_4$ ,  
si verifichi che se  $\mathcal{P}(A) = \mathcal{P}(A | B)$  e  $\mathcal{P}(A) = \mathcal{P}(A | C)$ ,  
non è vero che  $\mathcal{P}(B) = \mathcal{P}(B | C)$

b) dati i tre eventi:  $A = E_1 \cup E_2, B = E_1 \cup E_3$  e  $C = E_1 \cup E_4$ ,  
allora:  $\mathcal{P}(A \cap B) = \mathcal{P}(A \cap C) = \mathcal{P}(B \cap C) = 1/4$ ,  
ma:  $\mathcal{P}(A \cap B \cap C) = \mathcal{P}(E_1) = 1/4 \neq \mathcal{P}(A) \mathcal{P}(B) \mathcal{P}(C) = 1/8$ .

D'altronde dato uno spazio partizionato in otto eventi equiprobabili  $E_1, E_2, \dots, E_8$  e dati i tre eventi:  
 $A = E_1 \cup E_2 \cup E_3 \cup E_4, B = E_1 \cup E_5 \cup E_6 \cup E_7$  e  $C = E_1 \cup E_2 \cup E_7 \cup E_8$ ,  
si ha:

c)  $\mathcal{P}(A \cap B \cap C) = \mathcal{P}(E_1) = 1/8 = \mathcal{P}(A) \mathcal{P}(B) \mathcal{P}(C)$ , ma:  
 $\mathcal{P}(A \cap B) = \mathcal{P}(E_1) = 1/8 \neq \mathcal{P}(A) \mathcal{P}(B) = 1/4$ .

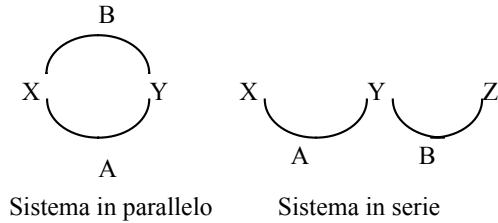
Quindi affinché tre eventi possano dirsi indipendenti, occorre non solo che siano indipendenti a due a due, ma anche a tre a tre e viceversa. Questa proprietà si estende facilmente a insiemi costituiti da  $n$  eventi.

È comune considerare la probabilità di un'intersezione di eventi come il prodotto delle loro probabilità. Occorre però precisare che tale prodotto si può effettuare solo se gli eventi sono indipendenti. Per esempio nel gioco a testa e croce, lanciando due volte la stessa moneta o due monete eguali, si suppone che il risultato di un lancio non influenzi il risultato dell'altro. Ora ciò può essere assunto dalla conoscenza del fenomeno fisico così come può essere dettato dall'esperienza empirica condotta su un gran numero di prove ripetute.

Tornando all'esempio del paragrafo 1.3, se possiamo ammettere che i tre lanci sono a due a due tutti indipendenti, allora, ammettendo che le probabilità di testa e di croce per ciascuno dei tre lanci siano tutte uguali e pari a  $1/2$ , tutti gli eventi da  $E_1$  ad  $E_8$  hanno la stessa probabilità pari a  $1/8$ .

#### 4. Sistemi in serie e in parallelo

Si definisce sistema *in parallelo* un sistema composto da un certo numero di componenti elementari, di cui basta che ne funzioni almeno uno, affinché il sistema funzioni. Per esempio per raggiungere  $Y$  da  $X$  vi siano due strade alternative e su ciascuna vi sia un ponte; basta che solo uno di essi sia funzionante. Si definisce sistema *in serie* un sistema composto da un certo numero di componenti elementari che devono funzionare tutti contemporaneamente affinché il sistema funzioni. Per esempio, se devo raggiungere dal punto  $X$  il punto  $Z$  e vi sono lungo il percorso due ponti  $A$  e  $B$  uno dietro l'altro, occorre che essi siano entrambi percorribili.



4.1 Qual è la probabilità che il sistema in parallelo funzioni? Se abbiamo  $\mathcal{P}(A)$  e  $\mathcal{P}(B)$  la probabilità di funzionamento di  $A$  e di  $B$ , bisogna determinare qual è lo spazio degli eventi partizionato:

$$\begin{array}{ll}
 A \cap B & A \cap \bar{B} \\
 \bar{A} \cap B & \bar{A} \cap \bar{B}.
 \end{array}$$

Quindi l'unico evento che non porta ad un sistema funzionante è il quarto: quello in cui *tutti* i componenti non sono funzionanti.

Evidentemente se i componenti fossero in numero superiore, come sappiamo, il partizionamento diventerebbe sempre più pesante. Però a ben vedere è inutile eseguirlo, in quanto possiamo calcolare la probabilità dell'unico caso in cui il sistema *non* funziona e quindi ricavare la probabilità di funzionamento come complemento all'unità.

Se i due eventi sono indipendenti si ha  $\mathcal{P}(\bar{A} \cap \bar{B}) = \mathcal{P}(\bar{A}) \mathcal{P}(\bar{B})$ . E quindi  $\mathcal{P}(S) = 1 - \mathcal{P}(\bar{A}) \mathcal{P}(\bar{B})$ . Generalizzando a più componenti indipendenti, si ha:

$$\mathcal{P}(S) = 1 - \prod_{i=1}^n [1 - \mathcal{P}(A_i)].$$

Qual è la probabilità che il sistema in serie funzioni? In questo caso, poiché devono funzionare tutti i componenti, sempre nel caso di eventi indipendenti, si ha:

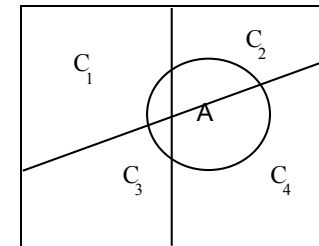
$$\mathcal{P}(S) = \prod_{i=1}^n \mathcal{P}(A_i).$$

## 5. Formula di Bayes

Spesso il problema che si pone è il seguente: si è osservato il risultato di un esperimento che può dipendere da differenti cause; qual è la probabilità che a generare quel risultato sia stata questa o quella causa?

Formalmente: si abbia lo spazio degli eventi partizionato secondo quelle che chiameremo cause:  $C_1, C_2, \dots, C_n$  (dunque ammettiamo che a generare quel risultato possa essere stata *una ed una sola* di quelle cause); su tale spazio si sovrapponga lo spazio dei risultati, per esempio risultato positivo  $A$  o negativo  $\bar{A}$ ; si desidera conoscere la probabilità delle cause, *dopo* aver osservato quale risultato è stato prodotto.

Per esempio una serie di pezzi provengano da un certo numero di macchine che possono produrre pezzi buoni o difettosi; si estragga un pezzo a caso e, avendo visto che esso è difettoso, si voglia determinare la probabilità che esso sia stato prodotto da una o dall'altra macchina. Per risolvere questo problema devono essere note quelle che chiameremo le probabilità *a priori*  $\mathcal{P}(C_i)$ , cioè nell'esempio le probabilità che ciascun pezzo sia prodotto da questa o quella macchina *senza tenere conto* dell'esito verificatosi, ossia prima di aver osservato se il pezzo estratto sia buono o difettoso. Inoltre devono essere note le *verosimiglianze*, ossia le probabilità che ciascun effetto sia prodotto dalle varie cause  $\mathcal{P}(A | C_i)$ , nell'esempio la difettosità delle varie macchine. Ciò che si cerca sono le probabilità *a posteriori*  $\mathcal{P}(C_i | A)$



5.1 Per determinare tali probabilità, partiamo dalla definizione di probabilità condizionata:

$$\mathcal{P}(C_i | A) \mathcal{P}(A) = \mathcal{P}(C_i \cap A) = \mathcal{P}(A \cap C_i) = \mathcal{P}(C_i) \mathcal{P}(A | C_i)$$

da cui:

$$\mathcal{P}(C_i | A) = \frac{\mathcal{P}(C_i) \mathcal{P}(A | C_i)}{\mathcal{P}(A)}.$$

Per ricavare  $\mathcal{P}(A)$  possiamo considerare che:

$$A = A \cap \Omega = A \cap (C_1 \cup C_2 \cup \dots \cup C_n) = (A \cap C_1) \cup (A \cap C_2) \cup \dots \cup (A \cap C_n)$$

ed essendo gli eventi  $(A \cap C_i)$  incompatibili tra loro a causa dell'incompatibilità delle  $C_i$ , si ha:

$$\mathcal{P}(A) = \mathcal{P}(C_1) \mathcal{P}(A|C_1) + \mathcal{P}(C_2) \mathcal{P}(A|C_2) + \dots + \mathcal{P}(C_n) \mathcal{P}(A|C_n) ,$$

da cui in definitiva:

$$\mathcal{P}(C_i|A) = \frac{\mathcal{P}(C_i) \mathcal{P}(A|C_i)}{\sum_{j=1}^n \mathcal{P}(C_j) \mathcal{P}(A|C_j)} .$$

Vi sono alcuni particolari che vogliamo sottolineare:

- così come la  $\sum_i \mathcal{P}(C_i) = 1$ , anche la  $\sum_i \mathcal{P}(C_i|A) = 1$  ; infatti se prima era tutto lo spazio  $\Omega$  partizionato tra le varie cause  $C_i$ , ora è solo il sotto-spazio  $A$ ;
- poiché la massa unitaria di probabilità è suddivisa in un egual numero  $n$  di parti sia con le probabilità *a priori* che *a posteriori*, ogni  $\mathcal{P}(C_i|A)$  può essere in generale minore o maggiore o uguale della relativa  $\mathcal{P}(C_i)$ .

Non si cada nell'errore di pensare che, per fissato  $i$ ,  $\mathcal{P}(C_i|A)$  dev'essere minore del rispettivo  $\mathcal{P}(C_i)$ , come un'osservazione superficiale della figura precedente può suggerire. Infatti  $\mathcal{P}(C_i)$  è la porzione di  $C_i$  in  $\Omega$ , mentre  $\mathcal{P}(C_i|A)$  è la porzione di  $C_i$  in  $A$ .

**Esercizio 1.7.** Si tracci coi diagrammi di Venn uno spazio in cui  $\mathcal{P}(C_i) = \mathcal{P}(C_i|A)$  per ogni  $i$ .

**Esercizio 1.8.** Si tracci come prima uno spazio in cui  $\mathcal{P}(C_1|A) = 1$  .

5.2 Per applicare la formula di Bayes al caso in cui le cause non sono tra loro incompatibili ed esaustive, occorre dapprima partizionare lo spazio delle cause. Per esempio, in un sito ambientale in certi giorni vengono scaricate le acque depurate provenienti da due fabbriche. Poiché le due fabbriche possono anche scaricare nella stessa giornata, le cause non sono incompatibili. Durante un prelievo si è accertata una condizione di inquinamento dovuto al malfunzionamento di uno dei due depuratori. Ammettiamo di conoscere qual è il numero di giorni in un anno in cui ciascuna fabbrica scarica e, rapportandolo al numero di giorni di lavorazione, otteniamo la probabilità che ciascuna fabbrica scarichi nel sito  $\mathcal{P}(C_i)$ ; ammettiamo di conoscere anche la probabilità di malfunzionamento di ogni depuratore, ossia la probabilità che – ammesso che una fabbrica abbia scaricato – il depuratore non abbia operato correttamente  $\mathcal{P}(A|C_i)$ . Partizioniamo lo spazio delle cause, scomponendole nelle quattro cause elementari, e troviamo le nuove verosimiglianze da associarvi, assumendo l'indipendenza tra le cause e tra i malfunzionamenti:

	<i>Eventi possibili</i>	<i>Probabilità a priori</i>	<i>Verosimiglianze</i>
1	scaricano entrambe	$\mathcal{P}(C_1 \cap C_2) = \mathcal{P}(C_1) \mathcal{P}(C_2)$	$1 - [\mathcal{P}(\bar{A} C_1) \mathcal{P}(\bar{A} C_2)]$
2	scarica solo la prima	$\mathcal{P}(C_1 \cap \bar{C}_2) = \mathcal{P}(C_1) \mathcal{P}(\bar{C}_2)$	$\mathcal{P}(A C_1)$
3	scarica solo la seconda	$\mathcal{P}(\bar{C}_1 \cap C_2) = \mathcal{P}(\bar{C}_1) \mathcal{P}(C_2)$	$\mathcal{P}(A C_2)$
4	non scarica nessuna	$\mathcal{P}(\bar{C}_1 \cap \bar{C}_2) = \mathcal{P}(\bar{C}_1) \mathcal{P}(\bar{C}_2)$	0

# PROBLEMI PER IL CAPITOLO 1

1. In una partita a testa e croce con una moneta equa, qual è la probabilità che esca la faccia assegnata almeno una volta entro i primi cinque lanci?  
*Sugg.:* Si determini la probabilità che la faccia assegnata *non* esca entro i primi cinque lanci.

$$\text{Ris.}: 1 - (1/2)^5.$$

Ed in una partita a dadi?

$$\text{Ris.}: 1 - (5/6)^5.$$

- Quanti lanci devo effettuare nei due casi per essere ragionevolmente sicuro ( $Prob > 0.99$ ) di ottenere almeno una volta la faccia assegnata?

$$\text{Ris.}: 1 - (1/2)^n > 0.99, \text{ da cui } n = 7;$$

$$\text{Ris.}: 1 - (5/6)^n > 0.99, \text{ da cui } n = 26.$$

2. Un sistema è costituito da quattro elementi che funzionano indipendentemente. Se la probabilità di guasto dei quattro elementi è costante e pari a 0.01, qual è la probabilità di guasto del sistema, ammesso che esso sia fuori servizio se anche solo uno degli elementi è guasto?

$$\text{Ris.}: 1 - 0.99^4.$$

- E qual è, ammesso che il sistema sia fuori servizio solo se tutti e quattro gli elementi sono guasti?

$$\text{Ris.}: 0.01^4.$$

- E qual è, ammesso che sia fuori servizio se sono guasti almeno due dei quattro elementi?

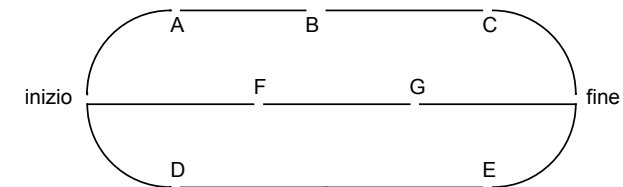
*Sugg.:* Si partizioni lo spazio degli eventi.

$$\text{Ris.}: 0.01^4 + 4 \cdot 0.99 \cdot 0.01^3 + 6 \cdot 0.99^2 \cdot 0.01^2 = .000592.$$

- Nei primi due casi quanto dev'essere la probabilità di guasto di ciascun elemento  $p$  per ottenere una probabilità di guasto del sistema di 0.01?

$$\text{Ris.}: 1 - (1 - p)^4 = 0.01; p^4 = 0.01.$$

3. Per andare dall'**inizio** alla **fine** del percorso si può scegliere una delle tre strade: nella prima occorre che funzionino (consentano il passaggio) gli elementi **A**, **B** e **C**; nella seconda **F** e **G**; nella terza **D** ed **E**.



Ammesso che il fatto che un elemento sia percorribile è indipendente da un altro e che le probabilità di funzionamento siano:

$P(A)$	$P(B)$	$P(C)$	$P(D)$	$P(E)$	$P(F)$	$P(G)$
0.99	0.98	0.99	0.95	0.99	0.95	0.95

determinare qual è la probabilità di poter andare dall'inizio alla fine.

*Sugg.:* Determinare qual è la probabilità che ognuna delle 3 strade sia percorribile.

$$\text{Risp.: } 1 - (1 - 0.99 \cdot 0.98 \cdot 0.99) (1 - 0.95 \cdot 0.95) (1 - 0.95 \cdot 0.99) .$$

4. Un prodotto semilavorato, prima di essere immesso nel ciclo produttivo, viene sottoposto a tre test; il test *A* ha probabilità 0.95 di scoprire se il pezzo è difettoso, il test *B* 0.92 ed il test *C* 0.9. Qual è la probabilità di immettere nel ciclo produttivo un pezzo difettoso?

$$\text{Risp.: } (1 - 0.95) \cdot (1 - 0.92) \cdot (1 - 0.9) = 0.0004 .$$

- Il test *A* ha anche una probabilità di 0.01 di identificare il pezzo come difettoso, quando non lo è; il test *B* di 0.02 ed il test *C* di 0.05. Qual è la probabilità che un pezzo buono venga scartato?

$$\text{Risp.: } 1 - (1 - 0.01) \cdot (1 - 0.02) \cdot (1 - 0.05) = 0.07831 .$$

- Se il 99% dei pezzi è buono, qual è la probabilità di corretta diagnosi?

$$\text{Risp.: } 0.01 \cdot (1 - 0.0004) + 0.99 \cdot (1 - 0.07831) = 0.9224691 .$$

5. Avendo 3 macchine e 3 pezzi da lavorare, qual è la probabilità che, distribuendo a caso i pezzi alle macchine, ogni pezzo vada su una macchina diversa?

*Sugg.:* Dato il primo pezzo assegnato in modo qualsiasi, si determini la probabilità di assegnare il secondo ad un'altra macchina ed il terzo alla rimanente.

$$\text{Risp.: } 2/3 \cdot 1/3 .$$

6. Tre macchine producono pezzi di identica caratteristica, che confluiscono in un unico magazzino. La macchina *A* produce 10 pezzi per ora, la *B* 40 pezzi e la *C* 50. La prima produce con difettosità pari a 0.01, la seconda pari a 0.05 e la terza 0.1. Qual è la probabilità di ottenere un pezzo difettoso, prendendone uno a caso?

$$\text{Risp.: } 0.071 .$$

- Se dal magazzino si preleva un pezzo ed esso risulta difettoso, qual è la probabilità che provenga dalla prima, dalla seconda o dalla terza macchina?

*Sugg.:* Applicare la formula di Bayes con *a priori*:  $\mathcal{P}(A) = 10/100$ ;  $\mathcal{P}(B) = 40/100$ ;  $\mathcal{P}(C) = 50/100$  e verosimiglianze:  $\mathcal{P}(\text{difett.} | A) = 0.01$ ;  $\mathcal{P}(\text{difett.} | B) = 0.05$ ;  $\mathcal{P}(\text{difett.} | C) = 0.1$

$$\text{Risp.: } \mathcal{P}(A | \text{difett.}) = 0.014; \mathcal{P}(B | \text{difett.}) = 0.282; \mathcal{P}(C | \text{difett.}) = 0.704 .$$

7. Si supponga di partecipare al seguente gioco: possiamo puntare su tre carte, di cui solo una nasconde un premio. Dopo avere scelto la carta il conduttore del gioco scopre delle altre due una che certamente non è la vincente. A questo punto abbiamo la possibilità di permanere nella scelta o di cambiare carta. Determinare la strategia migliore.

*Sugg.:* Si partizioni lo spazio degli eventi per ognuna delle due strategie.

$$\text{Risp.: } \text{La probabilità di successo cambiando è di } 2/3, \text{ restando è } 1/3 .$$



## CAPITOLO 2

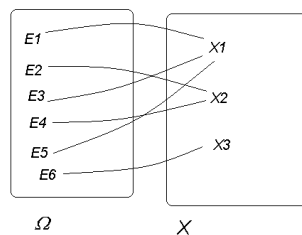
### VARIABILI CASUALI (GENERALITÀ)

Quando lo spazio degli eventi è opportunamente partizionato, si può far corrispondere a ciascun risultato un numero. La maggior parte dei giochi di carte, di dadi, la roulette, si basano su questa operazione, e lo studio del calcolo delle probabilità che affronteremo sarà basato su tali assegnazioni.

#### 1. Variabili casuali discrete

Iniziamo con l'esempio più semplice. Possiamo formalizzare l'esito del gioco a testa e croce, assegnando ad esso una *variabile*  $X$ . Per esempio, dato che gli esiti sono solo due, la variabile dovrà essere definita solo in due punti, diciamo 0 e 1, e quindi assegniamo  $T$  a 0 e  $C$  a 1 (oppure naturalmente  $T$  a 1 e  $C$  a 0). In tal caso quando l'esito è noto, è noto anche il risultato della variabile, ma finché tale esito non è noto, non si sa quale valore tale variabile assumerà; pertanto chiameremo una variabile siffatta *variabile casuale* (v.c.) o *aleatoria* per sottolineare il fatto che il valore che essa assume dipende da un fenomeno aleatorio. Passando dal gioco della moneta a quello del dado, gli esiti possibili sono sei. Ora è possibile associare ad ogni faccia un numero diverso, per esempio da 1 a 6, e quindi la v.c. sarà definita in sei punti distinti, ma in realtà il gioco potrebbe essere legato all'apparire di una faccia con numero pari o con numero dispari di punti, o all'apparire o non della faccia con cinque punti; allora in questo caso la v.c. sarebbe ancora definita solo in due punti.

La cosa essenziale è la seguente: dato un risultato, dev'essere definito univocamente il valore della variabile casuale, e quindi la legge che lega esito e v.c. dev'essere *suriettiva*: ad ogni esito deve corrispondere *uno ed un solo* valore della v.c. Per esempio una legge del tipo: *i punti dispari valgono  $X=0$  ed i punti minori di tre valgono  $X=1$* , non definisce una v.c. poiché il punto uno è incluso in entrambe le definizioni, mentre il quattro ed il sei non sono contemplati. Invece: *i punti dispari valgono  $X=1$ , i punti due e quattro valgono  $X=2$  e il sei vale  $X=3$*  è una v.c. definita su tre valori: 1, 2 e 3.



Una v.c.  $X$  quindi si basa su una partizione dello spazio degli eventi. Infatti, data la particolare legge suriettiva con la quale abbiamo ottenuto la v.c., qualunque evento si verifichi il valore della v.c. è uno solo, inoltre non si possono realizzare eventi che non abbiano un valore di v.c.

Ma si badi che la legge di formazione di tale v.c., come già sottolineato, non è unica. Infatti nel caso precedente si potrebbe essere interessati per esempio al numero di  $T$  ottenute indipendentemente dall'ordine di arrivo. E quindi  $E_1$  dà luogo ad  $X=3$ ;  $E_2, E_3$  e  $E_4$  danno luogo a  $X=2$ ;  $E_5, E_6$  e  $E_7$  danno luogo a  $X=1$ ;  $E_8$  a  $X=0$ . Ovviamente, poiché gli eventi  $E$  sono incompatibili, unendoli in modo da assegnare ogni  $E_i$  ad un solo gruppo, i risultati saranno anch'essi incompatibili. Questo quindi è il modo più diretto per creare v.c.: da una partizione dello spazio campionario.

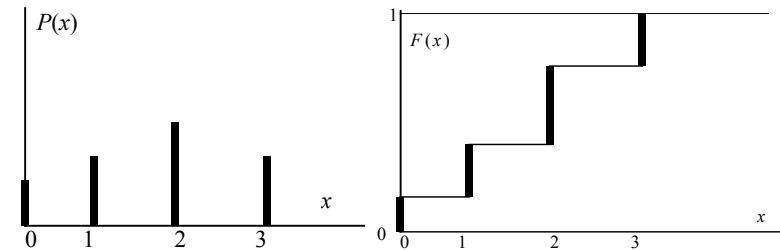
Naturalmente, poiché la v.c. costituisce una partizione dello spazio degli eventi, su essa può essere assegnata una misura di probabilità. La notazione di questo sarà ovviamente del tipo:  $\mathcal{P}(X=0)$ ; questo significherà, per esempio per la particolare v.c. legata al gioco a testa e croce, la probabilità che si realizzi l'evento connesso col valore 0 della v.c.

È possibile in realtà fare riferimento non ad un valore specifico del risultato, come in questo caso 0, ma ad un valore generico diciamo  $x$  che indicheremo con la lettera minuscola:  $\mathcal{P}(X=x)$ ; dipendentemente dai vari valori che si assegneranno ad  $x$  si otterranno i vari valori di probabilità connessi. In realtà, poiché questo sarà il caso più comune durante il nostro studio la notazione precedente  $\mathcal{P}(X=x)$  sarà abbreviata in  $\mathcal{P}(x)$ . Occorre però ricordare sempre che ciò sta per: *la probabilità che la v.c.  $X$  sia uguale ad un particolare valore da specificare  $x$* .

La legge suriettiva che lega lo spazio degli eventi alla v.c. fa sì che:

- a) i particolari valori che la v.c. può assumere sono disgiunti e quindi le relative probabilità sono sommabili e
- b) la somma di tali probabilità è 1.

Per esempio lanciando tre monete la v.c.  $X$  numero delle teste può assumere valori  $x = 0, 1, 2, 3$ ; lanciandone  $n$  la v.c.  $X$  può assumere valori  $x = 0, 1, 2, \dots, n$ . Tale situazione può essere rappresentata su un diagramma cartesiano, riportando in ascissa i valori di  $x$  ed in ordinata i valori delle  $P(x)$ .



Il primo grafico rappresenta la *distribuzione di probabilità* (d.p.), il secondo la *funzione di ripartizione* o *cumulativa* (f.r.). In quest'ultimo caso, invece di considerare la probabilità di un punto, si considera la probabilità riferita ad un intervallo illimitato inferiormente e limitato superiormente:

$$F(x) = P(X \leq x).$$

Si badi bene che la f.r. è definita su *tutto l'asse reale*, indipendentemente dall'intervallo e dal dominio di definizione della v.c. Nella letteratura anglosassone questa funzione è chiamata *cumulative distribution function* (c.d.f.), o semplicemente *distribution function*, con ovvi possibili disguidi per il lettore italiano. Se la v.c. è discreta la c.d.f. si presenta come una funzione a gradini, i cui salti avvengono in corrispondenza dei punti ove la v.c. pone una massa non nulla e sono di entità pari alla probabilità di quei punti.

Possiamo notare che la  $F(x)$  gode delle seguenti proprietà:

1.  $\lim_{x \rightarrow \infty} F(x) = 1$
2.  $\lim_{x \rightarrow -\infty} F(x) = 0$
3.  $F(x + \Delta x) - F(x) \geq 0$ .

La prima proprietà deriva dal secondo assioma della probabilità:  $P(\Omega) = 1$ ; infatti rappresenta tutta la massa di probabilità, diffusa o concentrata, che è unitaria poiché la v.c. è una suddivisione *esaustiva* dello spazio campionario.

La seconda proprietà è evidente da sé.

La terza, comporta che la  $F(x)$  sia una funzione non decrescente; infatti la  $P(x)$  è non negativa.

## 2. Variabili casuali continue

Le v.c. che abbiamo descritto finora sono *discrete* in quanto i valori che esse possono assumere si possono mettere in corrispondenza biunivoca coi numeri interi (nei casi finora esaminati con un insieme limitato di essi), cioè i possibili risultati possono essere enumerati; quindi la v.c. è discreta e la massa di probabilità sarà distribuita su un insieme numerabile (eventualmente anche illimitato, ma con la potenza del numerabile) di punti. Invece vi può essere il caso in cui l'esito del risultato sia diverso.

Come esempio prendiamo il gioco delle bocce. Il risultato del tiro, ammesso che esso non sia nullo, è determinato dalla distanza della boccia dal pallino; tale distanza è una variabile continua (in corrispondenza biunivoca con l'asse dei numeri reali, o di una sua porzione) e quindi la v.c. che noi determiniamo è data proprio da tale distanza e quindi è una v.c. *continua*. (In realtà anche in questo caso la rappresentazione potrebbe essere una v.c. discreta, per esempio valutando se il risultato è più vicino (0) o più lontano (1) dal pallino del lancio precedente, oppure valutando la distanza espressa secondo un'assegnata precisione, diciamo del centimetro: in questo senso la rilevazione o *misurazione* di una misura continua può essere sempre considerata come discreta). In questo caso la probabilità che un lancio finisca *esattamente* ad una distanza assegnata dal pallino è praticamente nulla. Eppure, quando lanciamo una boccia, *uno* dei valori di  $X$  si realizzerà, anche se a parità di tutti gli altri valori che erano possibili prima del lancio aveva una probabilità assolutamente trascurabile di verificarsi. In questo caso la massa di probabilità si dice che è *diffusa* su un certo insieme di definizione.

Formalizzando il discorso, invece di chiedersi qual è la  $\mathcal{P}(X=x)$ , ci si chieda: qual è la  $\mathcal{P}(X \leq x)$ , la probabilità che la v.c.  $X$  sia minore di un valore assegnato  $x$ , ossia si costruisca per prima la *c.d.f.*

Una nuova notazione ci permette di evitare la necessità di distinguere i due casi continuo e discreto.

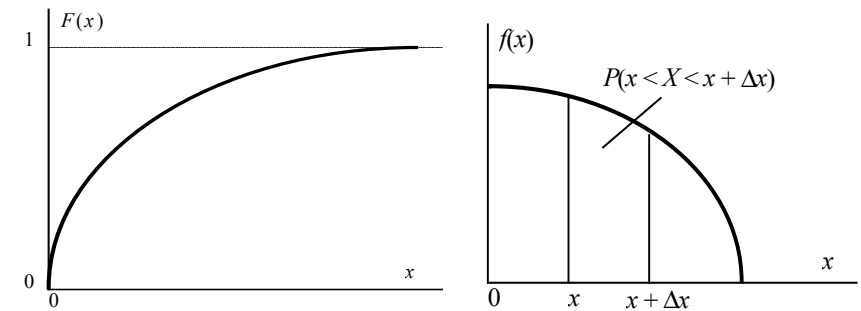
Se la v.c. è continua la  $\mathcal{F}(x)$  risulta una funzione che, pur mantenendo le proprietà già descritte, risulta crescente con continuità ed assume il senso formale di una funzione integranda, essendo derivabile in tutti i punti:

$$\mathcal{F}(x) = \int_{-\infty}^x f(\xi) d\xi \quad \text{e quindi} \quad f(x) dx = d\mathcal{F}(x)$$

si può ricavare la sua derivata  $f(x)$  che prende il nome di *funzione di densità di probabilità* (f.d.). Pertanto sul grafico della f.d. le probabilità sono rappresentate come aree, in quanto corrispondono ad un'operazione di integrazione definita.

La costruzione della f.d. può quindi ottenersi chiedendosi non qual è la  $\mathcal{P}(X=x)$ , ma qual è la  $\mathcal{P}(x < X \leq x + \Delta x)$ ; ossia la probabilità di cadere non in

un punto ma in un certo intervallo tra  $x$  e  $x + \Delta x$ . Diminuendo sempre più il valore di  $\Delta x$  fino a renderlo infinitesimo  $dx$ , si ottiene l'area di un rettangolo pari a  $\mathcal{P}(x < X \leq x + dx)$  avente base infinitesima  $dx$  ed altezza pari alla quantità  $f(x)$ .



Si noti bene che mentre la funzione di distribuzione di probabilità è un insieme di valori di probabilità attribuiti ai vari valori che la v.c. può assumere, e non ai valori che la v.c. *non* può assumere (per es., la probabilità che su 10 lanci di una moneta escano 5 teste ma non 4,5 teste); la densità di probabilità non esprime una probabilità finché tale funzione non è moltiplicata per il differenziale  $dx$ , che rende tale probabilità infinitesima.

Anche nel caso discreto la  $\mathcal{P}(x)$  si può ricavare formalmente come  $d\mathcal{F}(x)$ , dando però a questo differenziale un significato particolare: esso è ovviamente nullo nei punti  $x$  ove la v.c. non concentra massa di probabilità e quindi non è definita, mentre è pari al salto  $\mathcal{P}(x) = \mathcal{F}(x + \Delta x) - \mathcal{F}(x - \Delta x)$  nei punti  $x$  ove la v.c. è definita. Essa può essere facilmente calcolata sommando tutte le probabilità relative ai valori di  $X < x$  ove la v.c. è definita. Si noti che, mentre per v.c. continue la  $\mathcal{F}(x)$  è continua a destra e a sinistra, per v.c. discrete essa è *continua a destra*, poiché il salto viene effettuato proprio in corrispondenza del valore di  $x$  e non subito dopo, come prescrive la scrittura  $\mathcal{F}(x) = \mathcal{P}(X \leq x)$ .

Pertanto la funzione di ripartizione consente una rappresentazione unitaria sia nel caso di v.c. discrete che continue.

### 3. Variabili casuali multiple

Si possono prendere in considerazione contemporaneamente più variabili casuali. Per esempio nel caso discreto si possono lanciare due monete o due dadi. In questo caso la v.c. è costituita da una coppia di valori (l'esito del primo lancio e del secondo) e si parla di v.c. *doppia* o più in generale *multipla*.

Se si è in presenza di v.c. multiple occorre definire come la massa di probabilità si distribuisce su queste variabili.

3.1 Nel caso di due v.c. entrambe discrete i punti dove è concentrata massa di probabilità sono quelli ove le due v.c. sono *congiuntamente* definite.

Nel caso del lancio di due dadi la d.p. *congiunta* è riportata nella matrice sottostante. Nell'ultima colonna sono riportate le *distribuzioni di probabilità marginali*, ossia le probabilità riferite solo alla prima v.c., indipendentemente dal valore dell'altra. Similmente nell'ultima riga è riportata la probabilità marginale della seconda:

$$\mathcal{P}(X_1=x_1) = \sum_{x_2} \mathcal{P}(X_1=x_1 \cap X_2=x_2) \quad \text{e} \quad \mathcal{P}(X_2=x_2) = \sum_{x_1} \mathcal{P}(X_1=x_1 \cap X_2=x_2)$$

Primo lancio	Secondo lancio						
	1	2	3	4	5	6	
1	$\mathcal{P}(X_1=1 \cap X_2=1)$	$\mathcal{P}(X_1=1 \cap X_2=2)$	$\mathcal{P}(X_1=1 \cap X_2=3)$	$\mathcal{P}(X_1=1 \cap X_2=4)$	$\mathcal{P}(X_1=1 \cap X_2=5)$	$\mathcal{P}(X_1=1 \cap X_2=6)$	$\mathcal{P}(X_1=1)$
2	$\mathcal{P}(X_1=2 \cap X_2=1)$	$\mathcal{P}(X_1=2 \cap X_2=2)$	$\mathcal{P}(X_1=2 \cap X_2=3)$	$\mathcal{P}(X_1=2 \cap X_2=4)$	$\mathcal{P}(X_1=2 \cap X_2=5)$	$\mathcal{P}(X_1=2 \cap X_2=6)$	$\mathcal{P}(X_1=2)$
3	$\mathcal{P}(X_1=3 \cap X_2=1)$	$\mathcal{P}(X_1=3 \cap X_2=2)$	$\mathcal{P}(X_1=3 \cap X_2=3)$	$\mathcal{P}(X_1=3 \cap X_2=4)$	$\mathcal{P}(X_1=3 \cap X_2=5)$	$\mathcal{P}(X_1=3 \cap X_2=6)$	$\mathcal{P}(X_1=3)$
4	$\mathcal{P}(X_1=4 \cap X_2=1)$	$\mathcal{P}(X_1=4 \cap X_2=2)$	$\mathcal{P}(X_1=4 \cap X_2=3)$	$\mathcal{P}(X_1=4 \cap X_2=4)$	$\mathcal{P}(X_1=4 \cap X_2=5)$	$\mathcal{P}(X_1=4 \cap X_2=6)$	$\mathcal{P}(X_1=4)$
5	$\mathcal{P}(X_1=5 \cap X_2=1)$	$\mathcal{P}(X_1=5 \cap X_2=2)$	$\mathcal{P}(X_1=5 \cap X_2=3)$	$\mathcal{P}(X_1=5 \cap X_2=4)$	$\mathcal{P}(X_1=5 \cap X_2=5)$	$\mathcal{P}(X_1=5 \cap X_2=6)$	$\mathcal{P}(X_1=5)$
6	$\mathcal{P}(X_1=6 \cap X_2=1)$	$\mathcal{P}(X_1=6 \cap X_2=2)$	$\mathcal{P}(X_1=6 \cap X_2=3)$	$\mathcal{P}(X_1=6 \cap X_2=4)$	$\mathcal{P}(X_1=6 \cap X_2=5)$	$\mathcal{P}(X_1=6 \cap X_2=6)$	$\mathcal{P}(X_1=6)$
	$\mathcal{P}(X_2=1)$	$\mathcal{P}(X_2=2)$	$\mathcal{P}(X_2=3)$	$\mathcal{P}(X_2=4)$	$\mathcal{P}(X_2=5)$	$\mathcal{P}(X_2=6)$	1

Si definisce distribuzione di probabilità condizionata della  $X_1$  rispetto alla  $X_2$  quella che si ottiene dividendo la d.p. congiunta per la d.p. marginale:

$$\mathcal{P}(X_1 = x_1 | X_2 = x_2) = \mathcal{P}(X_1 = x_1 \cap X_2 = x_2) / \mathcal{P}(X_2 = x_2) .$$

Per definizione di indipendenza, se tale probabilità è pari a  $\mathcal{P}(X_1 = x_1)$  e ciò per tutti i valori  $x_1$  e  $x_2$ , allora le due v.c. si dicono *indipendenti*. Ossia si deve verificare che:

$$\mathcal{P}(X_1 = x_1 \cap X_2 = x_2) = \mathcal{P}(X_1 = x_1) \mathcal{P}(X_2 = x_2) .$$

3.2 Nel caso di due v.c. *continue*,  $X_1$  e  $X_2$ , si definisce la f.r. *congiunta* e la derivata seconda mista rispetto alle due variabili  $x_1$  e  $x_2$ , detta f.d. *congiunta*:

$$\mathcal{H}(x_1, x_2) = \mathcal{P}(X_1 \leq x_1 \cap X_2 \leq x_2) \quad \text{e} \quad f(x_1, x_2) = d^2 \mathcal{H}(x_1, x_2) / dx_1 dx_2 .$$

Quest'ultima, moltiplicata per i rispettivi differenziali, esprime la probabilità infinitesima che  $(x_1 < X_1 \leq x_1 + \Delta x_1)$  insieme a  $(x_2 < X_2 \leq x_2 + \Delta x_2)$ .

Per esempio, rilevando la posizione della boccia secondo la coppia di coordinate cartesiane (la distanza dal pallino secondo il lato lungo e secondo il lato corto del campo) o radiali con centro il pallino, ci si chiede che essa cada in una prefissata posizione espressa dalle due coordinate.

Si definiscono le funzioni di densità marginali come:

$$f(x_1) = \int f(x_1, x_2) dx_2 \quad \text{e} \quad f(x_2) = \int f(x_1, x_2) dx_1 ,$$

ove gli integrali precedenti si intendono definiti ed estesi rispettivamente su tutto il campo di definizione della  $x_2$  e della  $x_1$  e corrispondono quindi alle f.d. relative alle singole v.c.  $x_1$  o  $x_2$ .

Anche qui si definiscono le *funzioni di densità condizionate*:

$$f(x_1 | x_2) = f(x_1, x_2) / f(x_2) \quad \text{e} \quad f(x_2 | x_1) = f(x_1, x_2) / f(x_1).$$

Se  $f(x_1, x_2) = f(x_1)f(x_2)$  per tutti i valori di  $x_1$  e  $x_2$ , le due v.c. si dicono *indipendenti*.

#### 4. Distribuzioni di frequenze

Quando si è in presenza di una v.c. che rappresenta l'esito di un esperimento ripetibile è possibile ottenere una serie di risultati  $(x_1, \dots, x_n)$  ognuno dei quali è la realizzazione di tale v.c. Se tali esperimenti sono condotti *indipendentemente* l'uno dall'altro lasciando le condizioni al contorno di tale esperimento sempre costanti, si può pensare a tale  $n$ -upla indifferentemente come  $n$  realizzazioni della stessa v.c.  $X$  o come una realizzazione di  $n$  copie della stessa v.c.  $(X_1, \dots, X_n)$  da ognuna delle quali viene estratta una realizzazione. Dal punto di vista sostanziale, nel caso non si sia interessati all'ordine di estrazione delle prove, le due modalità sono equivalenti; dal punto di vista formale talvolta sarà comodo considerare questa o quella rappresentazione. Per adesso considereremo la prima.

4.1 Se la v.c. è discreta e concentra massa in pochi punti, diciamo  $m$ , mentre il numero di realizzazioni è superiore ( $n > m$ ), certamente alcuni dei valori della v.c. saranno ottenuti più volte. Tale numero esprime la *frequenza* di ciascun valore della v.c. in quell'esperimento. Per esempio, lanciando una moneta 10 volte si abbiano 6 teste e 4 croci: in questo caso, se a  $T$  è associato il valore  $X=1$  e a  $C$  il valore  $X=0$ , la frequenza  $n$  di  $x=1$  è pari a 6 e quella di  $x=0$  è 4. La *distribuzione di frequenza* (d.f.) è quindi:

$x$	$n$	$f$
0	4	0.4
1	6	0.6
	10	1.0

Nella terza colonna abbiamo riportato la *distribuzione di frequenza relativa*,  $f_i = n_i / n$ , ottenuta semplicemente dividendo ciascuna frequenza, che sarà detta quindi *assoluta*, per il totale. Possiamo costruire allora la funzione di ripartizione empirica assoluta e relativa accumulando le rispettive frequenze dall'alto. L'ultima casella di tali colonne sarà rispettivamente  $n$  e 1. In generale possiamo rappresentare una distribuzione di frequenza assoluta e relativa ed accumulata come la seguente tabella:

$x_i$	$n_i$	$f_i$	$h_i$	$F_i$
$x_1$	$n_1$	$f_1$	$n_1$	$f_1$
...	...	...	...	...
$x_i$	$n_i$	$f_i$	$n_1 + n_2 + \dots + n_i$	$f_1 + f_2 + \dots + f_i$
...	...	...	...	...
$x_m$	$n_m$	$f_m$	$n$	1
	$n$	1		

4.2 Se invece la v.c. è continua, certamente gli  $n$  valori ottenuti saranno distinti, a meno che non si proceda ad una *discretizzazione* della variabile, ossia alla formulazione di intervalli non sovrappoventisi, dividendo l'intervallo ove la v.c. è definita in  $m$  classi.

$X$	$n_i$	$f_i$	$h_i$	$F_i$
$x_0 \mid x_1$	$n_1$	$f_1$	$n_1$	$f_1$
...	...	...	...	...
$x_{i-1} \mid x_i$	$n_i$	$f_i$	$n_1 + n_2 + \dots + n_i$	$f_1 + f_2 + \dots + f_i$
...	...	...	...	...
$x_{m-1} \mid x_m$	$n_m$	$f_m$	$n$	1
	$n$	1		

ove con  $\mid$  indichiamo che la classe è chiusa a destra (per es., il valore  $x_1$  fa parte della prima classe e non della seconda).

Naturalmente se la v.c. non è limitata inferiormente o superiormente la prima o l'ultima classe possono essere semirette:  $x_0 = -\infty$  e  $x_m = \infty$ . In questo caso le frequenze vanno attribuite alla classe oppure si potrà trovare un valore di comodo cui attribuirle, come il valore centrale di ciascuna classe (resta però il problema delle classi non limitate). In ogni caso si può notare che dal punto di vista formale la d.f. relativa e la d.p. godono delle stesse proprietà; esse sono infatti una misura non negativa e normalizzata all'unità.

Per le v.c. discrete le frequenze insistono su punti particolari, per le v.c. continue su intervalli. Il parallelismo si estende anche ad altre caratteristiche. Si possono costruire le distribuzioni di frequenza cumulate. Anche per la d.f. si può procedere ad una rappresentazione grafica, che sarà del tutto analoga a quella della d.p.: ad aste per d.f., a gradini per le cumulate (v. **Problema 2**).

Nel caso in cui per ogni unità si rilevino più caratteristiche, e quindi si è in presenza di v.c. multiple, occorre registrare il numero di osservazioni che cadono nell'intersezione delle modalità per le varie variabili. Per esempio, per una v.c. doppia si ottiene la distribuzione di frequenza bivariata che si rappresenta come una tabella a doppia entrata, proprio come una v.c. bivariata, ma al posto delle probabilità si troveranno le frequenze assolute.

Relativizzando tali frequenze sui totali marginali di riga, di colonna, o sul totale generale, si ottengono le distribuzioni di frequenza condizionate.



## PROBLEMI PER IL CAPITOLO 2

- Da un magazzino sono state prelevate a caso 100 casse contenenti ognuna 10 prodotti; per ciascuna scatola sono stati contati quanti pezzi vi erano aventi una certa caratteristica. Dai dati così ottenuti si costruisca la distribuzione di frequenza e la funzione di ripartizione empirica assolute e relative.

```

1 1 1 2 0 0 3 2 1 2 2 0 0 1 1 2 2 1 0 1
1 2 2 1 1 2 2 1 0 1 1 1 1 0 3 1 3 3 1 1
1 0 0 2 0 1 1 0 0 1 0 2 1 0 0 1 2 1 2 0
0 0 0 3 0 1 0 1 1 2 1 1 1 0 4 0 0 1 2 0
0 1 2 4 0 1 2 1 1 3 1 0 1 1 2 0 0 0 1 2

```

*Risp.:*

$x$	$n_i$	$f_i$	$h_i$	$F_i$
0	31	0.31	31	0.31
1	41	0.41	72	0.72
2	20	0.20	92	0.92
3	6	0.06	98	0.98
4	2	0.02	100	1.00
Totale	100	1.00		

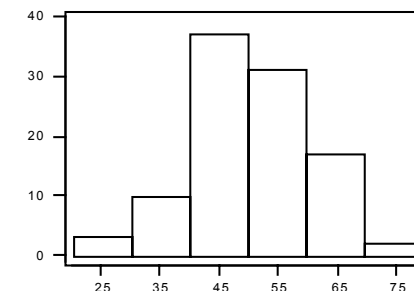
- Si rappresentino graficamente le quattro distribuzioni così ottenute.
- Su 100 pezzi è stata rilevata una certa caratteristica; costruire la distribuzione di frequenza e la funzione di ripartizione empirica assolute e relative.

```

31 50 62 43 63 38 57 53 68 52 66 47 74 54 59 54 41 42 46 67
63 51 55 49 43 57 43 47 58 69 57 57 41 57 64 40 51 44 38 41
34 72 56 57 49 47 30 49 47 56 51 41 39 53 40 66 50 36 47 58
47 49 26 28 53 53 46 37 51 46 32 53 43 64 23 66 33 50 64 46
41 61 49 45 40 58 63 59 49 41 49 41 68 65 52 61 43 42 41 56

```

- Sugg.:* Essendo una distribuzione di frequenza di variabile continua, per determinare la distribuzione di frequenza, essa va prima raggruppata in classi. La rappresentazione grafica più opportuna, in questo caso, è quella di un istogramma come quello rappresentato qui sotto, in cui ad ogni classe si associa un rettangolo avente base proporzionale all'ampiezza della classe ed area proporzionale alla frequenza associata a quella classe.



- Si rappresentino graficamente le quattro distribuzioni così ottenute.

*Sugg.:* Per la distribuzione cumulativa è preferibile invece tracciare la distribuzione basata sui valori originari. (Si ricordi che si ottiene una funzione a gradini).

3. Si supponga che per produrre un pezzo si possa usare la macchina  $A$  o la macchina  $B$ . La prima ha una difettosità pari a  $p_1$  e la seconda  $p_2$ . Supponendo che la scelta della macchina sia casuale con eguale probabilità, si costruisca la v.c. doppia, costituita dal tipo di macchina e dal tipo di pezzo.

- Si dica se le due v.c. in generale sono dipendenti e si determini la condizione per la quale sono indipendenti.

*Risp.:* Le due v.c. sono indipendenti per  $p_1 = p_2$

- Qual è la probabilità di ottenere un pezzo difettoso?

*Risp.:*  $\frac{1}{2} (p_1 + p_2)$

4. Dalle coppie di dati ( $X$  e  $Y$ ) sottostanti si costruisca la distribuzione di frequenza doppia, assoluta e relativa.

$X$	$Y$	$X$	$Y$	$X$	$Y$	$X$	$Y$	$X$	$Y$	$X$	$Y$	$X$	$Y$	$X$	$Y$	$X$	$Y$	$X$	$Y$
4	0	3	1	2	2	2	0	2	2	2	0	3	1	2	1	3	2	0	1
1	1	2	1	1	2	2	0	2	3	4	1	4	0	3	1	4	1	3	0
2	3	0	0	3	1	3	0	2	1	4	1	2	2	4	1	4	1	3	1
3	1	1	2	0	1	4	1	3	1	3	1	2	2	1	1	3	0	4	0
3	0	2	1	4	0	4	1	2	0	3	0	2	0	3	1	4	0	2	0

## CAPITOLO 3

### MOMENTI E MEDIE

Spesso è utile sintetizzare la v.c. secondo alcune quantità che esprimono questa o quella caratteristica che si intende mettere in luce. Ciò viene fatto attraverso i momenti e le medie.

#### 1. *Momenti*

Il primo di tali *momenti* è il *valor medio teorico* o *valore atteso* (v.a.) o *expectation*. Esso si definisce per le v.c. discrete come:

$$E(X) = \sum_x x \mathcal{P}(x),$$

mentre per le v.c. continue si dovrà calcolare l'integrale:

$$E(X) = \int_x x f(x) dx,$$

ove gli estremi della sommatoria o dell'integrale vanno estesi su tutto il campo di definizione della  $X$ .

Ricordando che gli estremi del campo di definizione della f.r. sono pari a  $-\infty$  e  $\infty$ , è possibile usare la notazione unificata:

$$E(X) = \int_{-\infty}^{\infty} x d\mathcal{F}(x).$$

Tali tipi di integrali d'ora in poi li indicheremo semplicemente come  $\int x d\mathcal{F}(x)$  intendendoli sempre come integrali definiti calcolati su  $\mathfrak{R}$ . Questo è un tipo di integrale particolare poiché come si vede fa uso della quantità  $d\mathcal{F}(x)$ .

Come già detto ciò comporta che per le v.c. discrete nei punti  $x$  ove la  $X$  non è definita si attribuisce a  $d\mathcal{H}(x)$  valore nullo, mentre nei punti  $x$  ove la  $X$  è definita si attribuisce peso pari a  $\mathcal{P}(x)$ . Per le v.c. continue, poiché  $d\mathcal{H}(x) = f(x) dx$ , l'integrale si calcola usualmente, ma entro l'intervallo dove la v.c. è definita. Quindi la notazione  $d\mathcal{H}(x)$  ci solleva dal considerare sia se la v.c. è discreta o continua, sia l'intervallo ove essa è definita.

1.1 Ma che cosa rappresenta il valor medio teorico? Tale formula ha un analogo nella comune media aritmetica effettuata su  $n$  osservazioni: se essa viene applicata, anziché alla d.p., alla d.f., si ottiene infatti:

$$M = \sum_i x_i f_i = x_1 f_1 + x_2 f_2 + \dots + x_m f_m = (x_1 n_1 + x_2 n_2 + \dots + x_m n_m) / n = \sum_i x_i f_i$$

ossia, sostituendo ad ogni valore  $x_i$  questo valore  $M$ , la somma ponderata non cambia:

$$M = \sum_i x_i f_i = \sum_i M f_i = M \sum_i f_i = M.$$

In modo finora molto intuitivo, facendo riferimento alla probabilità come limite delle frequenze osservate, si può ammettere che effettuando un numero sufficientemente elevato di prove, il valore  $f_i$  tende a  $\mathcal{P}(x_i)$  e quindi  $M$  tende a  $\mathcal{E}(X)$ . In realtà, come la media aritmetica rappresenta il *baricentro* di un insieme di osservazioni, così il v.a. rappresenta il baricentro della v.c.

Chiariamo meglio questo concetto discutendo le due proprietà del v.a.

Definita la v.c.  $X' = X - \mathcal{E}(X)$ , si ha:

1.  $\mathcal{E}[X'] = \mathcal{E}[X - \mathcal{E}(X)] = 0$ ;
2.  $\mathcal{E}[X']^2 = \mathcal{E}[(X - c)^2] = \min_c$  per  $c = \mathcal{E}(X)$ .

Infatti:

$$\begin{aligned} \int [x - \mathcal{E}(X)] d\mathcal{H}(x) &= \int x d\mathcal{H}(x) - \int \mathcal{E}(X) d\mathcal{H}(x) = \int x d\mathcal{H}(x) - \mathcal{E}(X) \int d\mathcal{H}(x) \\ &= \mathcal{E}(X) - \mathcal{E}(X) = 0. \end{aligned}$$

Si noti che:  $\int \mathcal{E}(X) d\mathcal{H}(x) = \mathcal{E}(X) \int d\mathcal{H}(x)$ ,

poiché  $\mathcal{E}(X)$  è una costante rispetto ad  $X$ , in quanto in  $\mathcal{E}(X)$  l'operatore integrale è già stato effettuato; in sostanza, se ad ogni valore  $x$  si sostituisce un unico valore  $c$  e tale valore  $c$  è proprio il v.a., resta inalterata la somma integrale dei prodotti  $x d\mathcal{H}(x)$ . D'altro lato il fatto che  $\int d\mathcal{H}(x) = 1$  deriva dalla seconda proprietà delle v.c. che prevede che tutta la massa di probabilità sia relativizzata ad 1.

1.2 La prima proprietà introduce una nuova classe di momenti: i *momenti centrati*, ossia i momenti calcolati non sulla v.c. originaria, ma su una sua trasformazione ottenuta scartando ogni valore  $x$  da una prefissata costante  $c$ .

Per motivi che saranno subito chiari, di solito tale costante è il v.a.

Per commentare la seconda proprietà, si può fare riferimento alla *funzione di perdita*. Si ammetta di dover indovinare quale sarà il prossimo valore estratto dalla v.c.  $X$  e che pagheremo una penale tanto più elevata quanto più la nostra previsione  $c$  sarà lontana, secondo la metrica quadratica, dal valore estratto. Se tale previsione viene effettuata su  $n$  valori, alla fine il costo sopportato sarà  $\sum_i (x_i - c)^2 f_i$ , che potrà essere minimizzato scegliendo opportunamente  $c$ .

$$\text{Ossia:} \quad \frac{\partial}{\partial c} \sum_i (x_i - c)^2 f_i = -2 \sum_i (x_i - c) f_i = 0$$

$$\text{da cui:} \quad \sum_i x_i f_i = \sum_i c f_i = c \sum_i f_i = c = M.$$

Effettuando la derivata seconda ci si rende conto che essa è sempre positiva e quindi ciò che otteniamo è effettivamente un minimo. Analoghi passaggi si effettuano se si fa riferimento non ad  $n$  osservazioni ma alla v.c.  $X$ . In questo caso si vorrà minimizzare il v.a. della funzione di perdita  $E(X - c)^2$ :

$$\frac{\partial}{\partial c} E(X - c)^2 = -2 E(X - c) = 0$$

$$\text{da cui:} \quad c = E(X).$$

La scelta di assumere il v.a. come quantità da minimizzare si giustifica, in senso frequentista, considerando il v.a. come il limite della media aritmetica quando  $n$  diverge, in quanto si suppone che le frequenze relative tendano a stabilizzarsi intorno alla probabilità. Alternativamente, in senso soggettivista, si può dire che il valore atteso (da qui il nome) si ottiene sostituendo nella formula della media aritmetica il valore della probabilità a quello della frequenza relativa. Si noti però che tale valore atteso *non* necessariamente è uno dei valori che realmente possono verificarsi.

Per esempio,  $x$  sia la v.c. associata al gioco a testa e croce con  $P(0) = 1 - \varphi$  e  $P(1) = \varphi$ . In questo caso:

$$E(X) = \sum_x x P(x) = 0 \times (1 - \varphi) + 1 \times \varphi = \varphi.$$

Essendo  $\varphi$  in generale compreso tra zero e uno,  $E(X)$  non è uno dei valori ammissibili di  $X$ .

La seconda proprietà introduce la classe più ampia di momenti di ordine superiore al primo. La formula generale di questi momenti è:

$$E[(X - c)^p] = {}_c\mu_p,$$

ove  $p$  costituisce il *grado* e  $c$  l'*origine* del momento. Ove  $p$  dovesse essere un reale occorre che l'argomento della potenza sia positivo. In questo caso si fa riferimento ai *momenti assoluti*:  $E(|X - c|)^p = {}_c\mu_p'$ . Dato il fatto che i momenti centrati sono i più frequentemente usati, per economia di scrittura essi vengono indicati senza l'indicazione dell'origine:  $E[X - E(X)]^p = \mu_p$ .

Ancora per brevità, il v.a. viene indicato semplicemente con  $\mu$ .

Per quanto riguarda la quantità  $E[X - E(X)]^2 = E[X^2]$ , essa è nota come *varianza* e si indica  $var(X)$  (talvolta con  $\sigma^2$ ). Se il v.a. rappresenta il baricentro della v.c., ossia un parametro di posizione, la varianza è il principale parametro di *dispersione*. Infatti se la v.c.  $x$  assumesse solo un valore, allora anche il v.a. sarebbe uguale a quel valore e la varianza sarebbe zero. Più dispersi sono i valori che la v.c. può assumere, più grande è la varianza. Si noti però che mentre l'unità di misura di  $E(X)$  è pari a quella dei valori  $x$ , la  $var(X)$  è espressa nei quadrati dell'unità di misura dei valori  $x$ .

## 2. Medie algebriche

Accanto ai momenti vi sono altre quantità sintetiche, dette *medie*, che hanno la caratteristica di essere espresse nella stessa unità di misura della  $X$ .

Infatti queste medie vengono ottenute estraendo dal relativo momento la radice di ordine pari a quello del momento stesso:

$$M_p(X) = [E(X^p)]^{1/p} = \mu_p^{1/p}.$$

Naturalmente la media del primo ordine o *media aritmetica* è uguale al momento primo. Non si faccia confusione su ciò che comunemente si intende per media: la media aritmetica applicata a dati osservati; anche le altre medie potenziate sono medie e possono essere applicate a osservazioni o a v.c.

Alcune delle altre medie sono degne di commento. Per esempio la media del secondo ordine o *media quadratica*. Quando tale media è basata sul momento centrato prende il nome di *scarto quadratico medio* o *deviazione standard*, spesso indicato con  $\sigma$ .

Altre medie importanti sono la *media geometrica*, ottenuta per  $p \rightarrow 0$

$$\lim_{p \rightarrow 0} M_p(X) = \lim_{p \rightarrow 0} \exp \left\{ \frac{1}{p} \log[E(X^p)] \right\} = \exp \left\{ \lim_{p \rightarrow 0} \frac{E[X^p \log(X)]}{E(X^p)} \right\}.$$

Da cui:

$$M_0(X) = \exp\{E[\log(X)]\}.$$

Vi è inoltre da citare la *media armonica* per  $p = -1$ .

Tutte queste medie sono utilizzabili per vari problemi; esse possono essere ricavate come particolari somme che restano inalterate sostituendo ai valori osservati un valore costante (vedi **Problemi N.2**).

Valgono due proprietà: *a*) le medie algebriche sono sempre comprese nell'intervallo di definizione della  $X$ , e *b*) se  $x > 0$ , esse sono crescenti con  $p$ .

Quindi:

1.  $\min(X) \leq M_p(X) \leq \max(X)$ .
2.  $M_{-1}(|X|) \leq M_0(|X|) \leq M_1(|X|) \leq M_2(|X|) \leq M_3(|X|)$ .

### 3. Relazioni tra momenti

È possibile per i momenti di ordine  $p$  intero ottenere delle relazioni che legano i momenti centrati a quelli di origine zero:

$$\begin{aligned}\mu_p &= E(X^p) = E[X - E(X)]^p = \\ &= E[X^p - p X^{p-1} E(X) + \dots + (-1)^i \binom{p}{i} X^{p-i} E^i(X) + \dots + (-1)^p E^p(X)] \\ &= E[X^p] - p E[X^{p-1}] E(X) + \dots + (-1)^i \binom{p}{i} E[X^{p-i}] E^i(X) + \dots + (-1)^p E^p(X) \\ &= {}_0\mu_p - p {}_0\mu_{p-1} {}_0\mu_1 + \binom{p}{2} {}_0\mu_{p-2} {}_0\mu_1^2 + \dots + (-1)^i \binom{p}{i} {}_0\mu_{p-i} {}_0\mu_1^i + \dots + (-1)^p {}_0\mu_1^p\end{aligned}$$

In particolare si ha:

$$\begin{aligned}\mu_2 &= {}_0\mu_2 - {}_0\mu_1^2 = E(X^2) - E^2(X) = \text{var}(X) \\ \mu_3 &= {}_0\mu_3 - 3 {}_0\mu_2 {}_0\mu_1 + 2 {}_0\mu_1^3 \\ \mu_4 &= {}_0\mu_4 - 4 {}_0\mu_3 {}_0\mu_1 + 6 {}_0\mu_2 {}_0\mu_1^2 - 3 {}_0\mu_1^4.\end{aligned}$$

### 4. Quantili e percentili

Le medie ed i momenti che abbiamo trattato finora richiedono che le v.c. siano di tipo quantitativo, che consentano cioè di effettuare i calcoli algebrici previsti, quali le somme. Occorre cioè che siano grandezze misurate su scale di intervallo, ossia grandezze dove sia fissata un'unità di misura (il tempo, la temperatura in gradi centigradi), o su scale di rapporti, ove vi sia anche uno zero assoluto (altezza, peso, temperatura in gradi assoluti, quantità monetarie).

4.1 Altri tipi di medie sono i *quantili* che, al contrario delle medie algebriche, non richiedono calcoli algebrici sui valori della v.c. Fissato un valore di probabilità, diciamo  $\alpha$ , il quantile  $x_\alpha$  stacca sulla f.d. o sulla f.p. alla sua sinistra un'area *almeno* pari ad  $\alpha$ , ossia:  $\mathcal{P}(X \leq x_\alpha) = \mathcal{F}(x_\alpha) \geq \alpha$ .

L'eguaglianza può essere raggiunta per v.c. continue, mentre per v.c. discrete occorre prendere il più piccolo valore di  $x$  che realizza:  $\mathcal{F}(x_\alpha) \geq \alpha$ .

Per  $\alpha = 1/2$  si ottiene la *mediana*, per  $\alpha = 1/4$  e  $3/4$  il primo ed il terzo *quartile*. Analogamente i *percentili* sono tali da staccare a sinistra un'area pari a  $100\alpha\%$ . Il 50° percentile è la mediana, il 25° il primo quartile. Occorre solo che la v.c. sia *ordinabile*, ossia si possa stabilire che un valore preceda o segua un altro valore secondo una scala che appunto chiamiamo *ordinale*. Per esempio i giudizi di valore sono delle grandezze qualitative ma ordinabili.

I quantili possono essere calcolati ovviamente anche per v.c. quantitative.

4.2 Un altro indice di posizione è la *moda* o *valore prevalente*. Esso è il valore della v.c. cui corrisponde, se discreta, il valore di probabilità massimo nella d.p. o, se continua, il valore di densità massima nella f.d.

Affinché si possa determinare la moda, la v.c. può essere anche qualitativa *non ordinabile*, ma chiaramente essa può essere determinata anche per v.c. ordinabili o quantitative. Le f.d. o le d.p. possono presentare anche più massimi, in tal caso si parla di mode relative ed assolute; se invece presentano un solo massimo si parla di distribuzioni *unimodali*.

## 5. Trasformazioni lineari e non lineari e standardizzazione

5.1 Definiamo *trasformazione lineare* di  $X$  la v.a.  $Y$

$$Y = aX + b$$

con  $a$  e  $b$  costanti. Si ha:

$$E(Y) = \int (aX + b) dF(x) = a \int X dF(x) + b \int dF(x) = aE(X) + b$$

e

$$var(Y) = \int [aX + b - aE(X) - b]^2 dF(x) = a^2 \int [X - E(X)]^2 dF(x) = a^2 var(X)$$

5.2 Si definisce una particolare trasformazione lineare, la *standardizzazione*:

$$X^* = X / \sqrt{var(X)} = [X - E(X)] / \sqrt{var(X)}.$$

Tale trasformazione gode della stessa proprietà della  $X^*$  (vedi par. 1.1) per quanto riguarda il v.a.:

$$E(X^*) = 0;$$

ed inoltre:

$$var(X^*) = E\{[X - E(X)] / \sqrt{var(X)}\}^2 = E[X - E(X)]^2 / var(X) = 1.$$

Questa trasformazione è molto importante perché ci mette in condizione di poter confrontare v.c. che hanno *forma distribuzionale* differente, anche se hanno una diversa posizione e dispersione dei valori. Tale trasformazione rende anche la v.c. ed i relativi momenti *privi di unità di misura* (numeri puri).

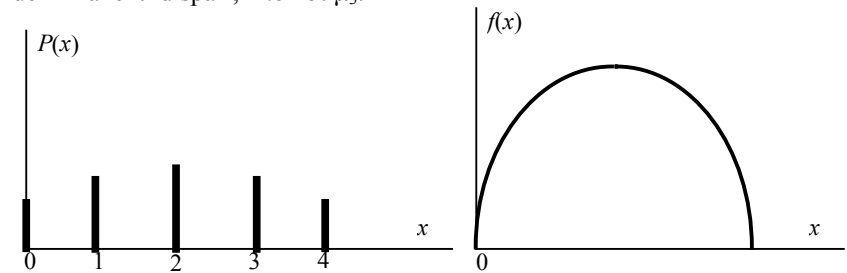
Per forma distribuzionale possiamo per ora intendere anche solo l'aspetto grafico che presentano le d.p. o le f.d. In particolare vi sono distribuzioni *simmetriche* ossia che rispetto ad un centro (il v.a.) presentano probabilità pari per *scarti*  $X - c$  eguali in valore assoluto, ma opposti:

$$\text{se } X_1 - c = -(X_2 - c) \quad \text{allora} \quad P(X_1 - c) = P(X_2 - c).$$

In questo caso qualunque momento di grado dispari risulta nullo. Infatti per ogni scarto positivo ne corrisponde uno negativo di eguale entità e ponderato in modo identico. Naturalmente non si può prendere il momento



primo centrato, che è *sempre* nullo, ma si potrà fare riferimento al più piccolo dei rimanenti dispari, il *terzo*:  $\mu_3$ .



Ora se la distribuzione è simmetrica, allora  $\mu_3 = 0$ ; altrimenti  $\mu_3$  sarà positivo se si ha una coda più lunga a destra o negativo se la coda è a sinistra. Tuttavia poiché tale momento risente dell'unità di misura e dell'ordine medio di grandezza dei valori  $x$  non è facile effettuare confronti sull'asimmetria di diverse v.c. o osservazioni. Se invece si confrontano i momenti terzi delle v.c. standardizzate, tali problemi non sussistono più.

Abbiamo così definito un *coefficiente di asimmetria* che è un numero puro

$$\beta_1 = \mu_3'' = [\mathcal{E}(X - c)^3] / \sqrt{[\text{var}(X)]^3}$$

Un altro coefficiente che descrive la forma di una distribuzione su cui però rimandiamo la discussione è il *coefficiente di curtosi*:

$$\beta_2 = \mu_4'' = [\mathcal{E}(X - c)^4] / \sqrt{[\text{var}(X)]^4}.$$

5.3 Se la v.c.  $Y$  è una funzione non lineare di un'altra v.c.  $X$ ,  $Y = h(X)$ , si possono ricavare i valori approssimati di  $\mathcal{E}(X)$  e  $\text{var}(X)$ . Dalla scomposizione in serie di Taylor, arrestata al primo ordine, si ha:

$$Y \approx h[\mathcal{E}(X)] + h'[\mathcal{E}(X)] [X - \mathcal{E}(X)]$$

$$\text{da cui: } \mathcal{E}(Y) \approx h[\mathcal{E}(X)] \quad \text{e} \quad \text{var}(Y) \approx \{h'[\mathcal{E}(X)]\}^2 \text{var}(X)$$

ove  $h[\mathcal{E}(X)]$  e  $h'[\mathcal{E}(X)]$  sono i valori della funzione  $h(\cdot)$  e della sua derivata calcolati in  $\mathcal{E}(X)$ .

## 6. Momenti di variabili casuali multiple

Si definiscono i *momenti misti* (centrati) di una v.c. doppia come:

$$\mu_{p,q} = \mathcal{E}(X_1^p, X_2^q) = \iint [X_1 - \mathcal{E}(X_1)]^p [X_2 - \mathcal{E}(X_2)]^q d^2 \mathcal{P}(x_1, x_2).$$

Nel caso in cui uno dei due gradi, per esempio  $q$ , è zero, si ottiene il momento semplice dell'altra variabile:

$$\mu_{p,0} = \mathcal{E}(X_1^p X_2^0) = \mu_p.$$

$$\begin{aligned} \text{Infatti: } \iint [X_1 - E(X_1)]^p d^2 \mathcal{H}(x_1, x_2) &= \\ &= \int [X_1 - E(X_1)]^p \int d^2 \mathcal{H}(x_1, x_2) = \int [X_1 - E(X_1)]^p d \mathcal{H}(x_1) \end{aligned}$$

**Esercizio 3.1.** Si generalizzi l'espressione per ottenere i momenti di v.c. multiple

Il più comune di tali momenti è il momento misto di secondo ordine detto *covarianza* che, espresso in funzione dei momenti di origine zero, è:

$$\mu_{1,1} = \text{cov}(X_1, X_2) = E(X_1' X_2) = E(X_1 X_2) - E(X_1) E(X_2) .$$

Se le due v.c. sono indipendenti allora  $\text{cov}(X_1, X_2) = 0$ .

Infatti in tal caso:

$$\begin{aligned} \text{cov}(X_1, X_2) &= \iint [X_1 - E(X_1)] [X_2 - E(X_2)] d^2 \mathcal{H}(x_1, x_2) = \\ &= \int [X_1 - E(X_1)] d \mathcal{H}(x_1) \int [X_2 - E(X_2)] d \mathcal{H}(x_2) = 0. \end{aligned}$$

Si badi che in generale non è vero il contrario, ossia la  $\text{cov}(X_1, X_2) = 0$  non implica l'indipendenza. Se  $\text{cov}(X_1, X_2) = 0$  usualmente si parla di *indipendenza in senso debole* e quindi l'indipendenza in senso assoluto si definisce come *indipendenza in senso forte* proprio per distinguerla da quella in senso debole.

## 7. La disuguaglianza di Cauchy-Schwartz e la correlazione

Possiamo allora dimostrare la *disuguaglianza di Cauchy-Schwartz*:

$$E^2(X Y) \leq E(X^2) E(Y^2) .$$

Infatti: data la quantità arbitraria  $k$ , poiché

$$E(X + kY)^2 = k^2 E(Y^2) + 2k E(X Y) + E(X^2) \geq 0,$$

per ogni  $k$ , allora il discriminante della precedente equazione di secondo grado in  $k$  dovrà essere non positivo, poiché non ci potranno essere zeri reali distinti:

$$\Delta = 4 E^2(XY) - 4 E(X^2) E(Y^2) \leq 0.$$

Se le v.c.  $X$  e  $Y$  sono scartate dai propri v.a., si ha:

$$\text{cov}^2(X, Y) \leq \text{var}(X) \text{var}(Y) .$$

Il caso di eguaglianza si raggiunge se  $Y|X$  è una v.c. degenere, ossia priva di variabilità (cioè, assegnato il valore  $x$ ,  $y$  è univocamente determinato), ed in particolare tale che vi sia un legame *lineare* deterministico tra le due v.c.:

$$Y = \alpha + \beta X.$$

(Ovviamente identiche considerazioni valgono per  $X|Y$ ). In questo caso è possibile fissare un  $k$  tale che  $X + kY = 0$  (per es.:  $\beta = -1/k$  e  $\alpha = 0$ ).

7.1 Si definisce il *coefficiente di correlazione lineare*:

$$\rho = \text{cov}(X, Y) / \sqrt{\text{var}(X) \text{var}(Y)}.$$

Esso è caratterizzato dal fatto di essere un numero puro, ossia è espresso in un'unità di misura adimensionale. Per la disuguaglianza di Cauchy-Schwartz si ha  $\rho^2 \leq 1$  e quindi:

$$-1 \leq \rho \leq 1.$$

Allora nel caso di perfetta dipendenza lineare ( $Y = \alpha + \beta X$ ),  $\rho$  raggiunge i limiti  $-1$  o  $1$ ; nel caso di indipendenza in senso debole esso vale  $0$ , in quanto  $\text{cov}(X, Y)$  è nulla. Pertanto l'indipendenza in senso debole si chiama anche *indipendenza lineare*.

## 8. Vettore di medie e matrice di varianza e covarianza

Una v.c. multipla può essere riguardata come un vettore di v.c. singole:

$$\mathbf{X} = (X_1, X_2, \dots, X_n)^T$$

che rappresentiamo come un vettore colonna e che scriviamo per economia di spazio come il trasposto (indicato con l'apice  $T$ ) di un vettore riga.

Esso è caratterizzato dai primi due momenti: il *vettore di medie* costituito dal vettore colonna dei v.a. delle v.c. unidimensionali:

$$\mathbb{E}(\mathbf{X}) = [\mathbb{E}(X_1), \mathbb{E}(X_2), \dots, \mathbb{E}(X_n)]^T$$

e dalla *matrice di varianza e covarianza* (o semplicemente matrice di covarianza), matrice che prevede sulla diagonale principale le varianze e negli elementi extradiagonali le covarianze:

$$\text{var}(\mathbf{X}) = \begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & \dots & \text{cov}(X_1, X_n) \\ \text{cov}(X_1, X_2) & \text{var}(X_2) & \dots & \text{cov}(X_2, X_n) \\ \dots & \dots & \dots & \dots \\ \text{cov}(X_1, X_n) & \text{cov}(X_2, X_n) & \dots & \text{var}(X_n) \end{bmatrix}$$

Essa risulta simmetrica poiché:

$$\text{cov}(X_i, X_j) = \text{cov}(X_j, X_i).$$

Essa inoltre risulta semi-definita positiva. Infatti in termini matriciali essa si può esprimere come

$$\text{var}(\mathbf{X}) = \mathbb{E}(\mathbf{X}'\mathbf{X}^T),$$

ove con l'apice intendiamo il vettore scartato dal proprio v.a.:  $\mathbf{X}' = \mathbf{X} - \mathbb{E}(\mathbf{X})$ .

Si noti che il prodotto  $\mathbf{X}'\mathbf{X}^T$  dà una matrice di dimensioni  $n \times n$ . Premoltiplicando e postmoltiplicando  $\text{var}(\mathbf{X})$  per un vettore  $\mathbf{a}$  arbitrario non nullo, sia ha:

$$\mathbf{a}^T \text{var}(\mathbf{X}) \mathbf{a} = \mathbf{a}^T \mathbb{E}(\mathbf{X} \mathbf{X}^T) \mathbf{a} = \mathbb{E}(\mathbf{X}^T \mathbf{a} \mathbf{a}^T \mathbf{X}) = \mathbb{E}(Y^T Y) = \text{var}(Y) \geq 0$$

ottenendo così una *forma quadratica* certamente non negativa, poiché è la varianza di  $Y$ , una v.c. unidimensionale  $Y$ , combinazione lineare di  $\mathbf{X}$  ( $Y = \mathbf{X}^T \mathbf{a}$ ), avente v.a. nullo. Come è noto, ciò è sufficiente a affermare che la forma quadratica all'interno è appunto semi-definita positiva.

## 9. Combinazioni lineari e non lineari di variabili casuali

Spesso la v.c. di interesse è la somma di più v. c. senza riguardo all'ordine di arrivo o di realizzazione. Per esempio, la vita di un sistema, costituito da elementi che entrano in funzione ad uno ad uno man mano che i precedenti si vanno guastando, è data dalla somma delle vite di ciascun componente; la v.c. numero di successi su  $n$  prove è dato dalla somma di v.c. di tipo 0 (per l'insuccesso) o 1 (per il successo), (sommando tanti 1 quanti sono i successi, e tanti 0 quanti sono gli insuccessi, otteniamo il numero di successi).

9.1 Più in generale si abbia una *combinazione lineare*:

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n = \sum_i a_i X_i$$

in cui i coefficienti  $a_i$  sono arbitrari (purché non tutti nulli). In particolare: se essi sono tutti 1, si ha la somma; se essi sono tutti pari a  $1/n$  si ha la media aritmetica. Ricaviamo i primi due momenti della v.c.  $Y$ .

Si osservi che, applicando l'operatore  $\mathbb{E}$  alla v.c.  $Y$ , si ha un integrale nelle  $n$  dimensioni  $x_1, x_2, \dots, x_n$ :

$$\begin{aligned} \mathbb{E}(Y) &= \int_{x_1} \int_{x_2} \dots \int_{x_n} [a_1 X_1 + a_2 X_2 + \dots + a_n X_n] d^n \mathcal{F}(x_1, x_2, \dots, x_n) = \\ &= \int_{x_1} a_1 X_1 \int_{x_2} \dots \int_{x_n} d^n \mathcal{F}(x_1, x_2, \dots, x_n) + \dots + \int_{x_n} a_n X_n \int_{x_1} \dots \int_{x_{n-1}} d^n \mathcal{F}(x_1, x_2, \dots, x_n) \end{aligned}$$

ma ciascuno degli integrali multipli ad  $n-1$  dimensioni che non include v.c.  $X_i$ , per definizione di funzione marginale, risulta  $d\mathcal{F}(x_i)$  (nella  $i$ -esima v.c. non integrata); per cui si ha:

$$\begin{aligned} \mathbb{E}(Y) &= \int_{x_1} a_1 X_1 d\mathcal{F}(x_1) + \int_{x_2} a_2 X_2 d\mathcal{F}(x_2) + \dots + \int_{x_n} a_n X_n d\mathcal{F}(x_n) = \\ &= a_1 \mathbb{E}(X_1) + a_2 \mathbb{E}(X_2) + \dots + a_n \mathbb{E}(X_n) \end{aligned}$$

Da cui:

$$\mathbb{E}(Y) = \mathbb{E}(\sum_i a_i X_i) = \sum_i a_i \mathbb{E}(X_i)$$

Analogamente:

$$\begin{aligned} \text{var}(Y) &= \mathbb{E}(\sum_i a_i X_i)^2 - \mathbb{E}^2(\sum_i a_i X_i) = \\ &= \sum_i a_i^2 \mathbb{E}[X_i]^2 + 2 \sum_{i < j} a_i a_j \mathbb{E}(X_i X_j) - \sum_i a_i^2 \mathbb{E}^2(X_i) - 2 \sum_{i < j} a_i a_j \mathbb{E}(X_i) \mathbb{E}(X_j) \end{aligned}$$

E quindi:

$$\text{var}(Y) = \sum_i a_i^2 \text{var}(X_i) + 2 \sum_{i < j} a_i a_j \text{cov}(X_i, X_j)$$

Nel caso di indipendenza lineare tra tutte le coppie di v.c., risulta:

$$\text{var}(Y) = \sum_i a_i^2 \text{var}(X_i) .$$

9.2 Se la v.c.  $Z$  è funzione non lineare di due v.c.  $X$  e  $Y$ ,  $Z = h(X, Y)$ , i valori approssimati di  $E(Z)$  e  $\text{var}(Z)$ , dalla scomposizione in serie di Taylor arrestata al I ordine, risultano:

$$Z \approx h[E(X), E(Y)] + h'_x[E(X), E(Y)] [X - E(X)] + h'_y[E(X), E(Y)] [Y - E(Y)]$$

$$\text{sono:} \quad E(Z) \approx h[E(X), E(Y)]$$

$$\begin{aligned} \text{var}(Z) \approx & \{h'_x[E(X), E(Y)]\}^2 \text{var}(X) + \{h'_y[E(X), E(Y)]\}^2 \text{var}(Y) + \\ & + 2 h'_x[E(X), E(Y)] h'_y[E(X), E(Y)] [Y - E(Y)] \text{cov}(X, Y) \end{aligned}$$

ove  $h'_x()$  e  $h'_y()$  sono le derivate prime rispetto a  $x$  e  $y$ .

## 10. Funzione caratteristica, generatrice dei momenti e generatrice dei cumulanti

10.1 La *funzione caratteristica* (f.c.) è una trasformazione della d.p. o della f.d.:

$$C_X(t) = E(e^{itX}) = \int e^{itx} d\mathcal{F}(x)$$

ove la variabile  $t$  è un parametro non aleatorio e  $i = \sqrt{-1}$ . Come si vede tale trasformazione dà una funzione non più in  $x$ , variabile che è stata integrata, bensì nella nuova variabile  $t$ . Tale funzione gode di alcune proprietà:

- La f.c. *esiste sempre*, nel senso che l'integrale è sempre limitato. Infatti:  $|\int e^{itx} d\mathcal{F}(x)| \leq \int |e^{itx}| d\mathcal{F}(x) = \int d\mathcal{F}(x) = 1$ , essendo  $|e^{itx}| = \sqrt{[\cos^2(tx) + \sin^2(tx)]} = 1$ .
- La f.c. è in *corrispondenza biunivoca* con la f.r. e quindi con la f.d. o la d.p.; pertanto conoscere la f.c. significa determinare univocamente la v.c.
- La f.c. di una *trasformazione lineare*  $Y = aX + b$  è:

$$C_Y(t) = E\{e^{(aX+b)it}\} = e^{ibt} E\{e^{iatX}\} = e^{ibt} C_X(at)$$

- La *somma*  $Y$  di v.c.  $X_i$  *indipendenti*,  $Y = \sum_i X_i$ , è una v.c. la cui f.c. è data dal prodotto delle f.c. delle v.c. che compongono la somma:

$$C_Y(t) = E\{e^{itY}\} = \int e^{ity} d\mathcal{F}(y) = \int \dots \int \exp[it \sum_i x_i] d^n \mathcal{F}(x_1, x_2, \dots, x_n)$$

e avendosi per l'indipendenza

$$d^n \mathcal{F}(x_1, x_2, \dots, x_n) = d\mathcal{F}(x_1) d\mathcal{F}(x_2) \dots d\mathcal{F}(x_n) = \prod_j d\mathcal{F}(x_j)$$

si ha:

$$C_Y(t) = \int \dots \int \prod_j \exp[it \sum_j x_j] d\mathcal{F}(x_j) = \prod_j \int \exp[it x_j] d\mathcal{F}(x_j) = \prod_j C_{X_j}(t)$$

- La f.c. di una v.c. *simmetrica intorno a 0*, ossia se  $d\mathcal{H}(x) = d\mathcal{H}(-x)$ , è reale:

$$\begin{aligned} C_X(t) &= \int_{-\infty}^0 e^{itx} d\mathcal{H}(x) + \int_0^{\infty} e^{itx} d\mathcal{H}(x) = \int_0^{\infty} e^{-itx} d\mathcal{H}(-x) + \int_0^{\infty} e^{itx} d\mathcal{H}(x) = \\ &= \int_0^{\infty} \{[\cos(tx) - i \sin(tx)] + [\cos(tx) + i \sin(tx)]\} d\mathcal{H}(x) = 2 \int_0^{\infty} \cos(tx) d\mathcal{H}(x) \end{aligned}$$

10.2 La *funzione generatrice dei momenti* (f.g.m.) è molto simile alla f.c.

$$\mathcal{M}_X(t) = \mathcal{E}(e^{tX}) = \int e^{tx} d\mathcal{H}(x)$$

ed essa si differenzia da quella solo per essere una funzione reale e non complessa. Ciò però comporta che non è garantita la sua esistenza.

Le altre proprietà della f.g.m., qualora esista, sono le stesse della f.c. Poiché le v.c. che tratteremo saranno dotate tutte di f.g.m., faremo sempre riferimento a questa quando dovremo studiare le v.c., mentre ci limiteremo a usare le f.c. quando dovremo ricavare qualche proprietà su v.c. generiche (che potrebbero anche non essere dotate di f.g.m.). La f.g.m. prende questo nome perché da essa si ricavano facilmente i momenti di origine zero e di grado  $k$  intero (se essi esistono finiti) della v.c., come la derivata di ordine  $k$  calcolata nel punto  $t=0$ . È chiaro quindi che il calcolo dei momenti in questo modo è più agevole, riducendosi a derivate successive, anziché ad integrali o sommatorie. Sviluppando in serie di Taylor la f.g.m., si ha:

$$\mathcal{M}_X(t) = \mathcal{E}\left\{\sum_{k=0}^{\infty} \frac{1}{k!} (tX)^k\right\} = \left\{\sum_{k=0}^{\infty} \frac{1}{k!} \mathcal{E}(tX)^k\right\} = \left\{\sum_{k=0}^{\infty} \frac{1}{k!} t^k {}_0\mu_k\right\},$$

che derivata una, due, tre volte, ..., calcolata nel punto  $t=0$ , dà:

$$\mathcal{M}_X'(0) = {}_0\mu_1; \mathcal{M}_X''(0) = {}_0\mu_2; \mathcal{M}_X'''(0) = {}_0\mu_3; \dots$$

Da questi sappiamo calcolare i momenti centrali.

10.3 Un'altra funzione connessa con la f.c. è la *funzione generatrice dei cumulanti* (f.g.c.). Essa è definita come

$$\mathcal{K}_X(t) = \log[\mathcal{M}_X(t)].$$

Vediamo a cosa corrispondono le prime quattro derivate:

$$\mathcal{K}_X'(t) = \mathcal{M}_X'(t) / \mathcal{M}_X(t)$$

$$\mathcal{K}_X''(t) = \mathcal{M}_X''(t) / \mathcal{M}_X(t) - [\mathcal{M}_X'(t) / \mathcal{M}_X(t)]^2$$

$$\begin{aligned} \mathcal{K}_X'''(t) &= \mathcal{M}_X'''(t) / \mathcal{M}_X(t) - \mathcal{M}_X''(t) \mathcal{M}_X'(t) / [\mathcal{M}_X(t)]^2 \\ &\quad - 2 [\mathcal{M}_X'(t) / \mathcal{M}_X(t)] \{ \mathcal{M}_X''(t) / \mathcal{M}_X(t) - [\mathcal{M}_X'(t) / \mathcal{M}_X(t)]^2 \} \end{aligned}$$

$$\begin{aligned} \mathcal{K}_X''''(t) &= \mathcal{M}_X''''(t) / \mathcal{M}_X(t) - \mathcal{M}_X'''(t) \mathcal{M}_X'(t) / [\mathcal{M}_X(t)]^3 - 3 \{ \mathcal{M}_X''(t) \mathcal{M}_X'(t) / [\mathcal{M}_X(t)]^2 \\ &\quad + [\mathcal{M}_X'(t)]^2 / [\mathcal{M}_X(t)]^2 - 2 \mathcal{M}_X''(t) [\mathcal{M}_X'(t)]^2 / [\mathcal{M}_X(t)]^3 \} + \\ &\quad + 2 \{ 3 \mathcal{M}_X''(t) [\mathcal{M}_X'(t)]^2 / [\mathcal{M}_X(t)]^3 - [\mathcal{M}_X''(t)]^4 / [\mathcal{M}_X(t)]^4 \} \end{aligned}$$

che, calcolate nel punto  $t=0$ , danno i primi quattro *cumulanti*:

$$\kappa_1 = {}_0\mu_1 = E(X)$$

$$\kappa_2 = {}_0\mu_2 - {}_0\mu_1^2 = \text{var}(X)$$

$$\kappa_3 = {}_0\mu_3 - 3 {}_0\mu_2 {}_0\mu_1 + 2 {}_0\mu_1^3 = \mu_3$$

$$\begin{aligned}\kappa_4 &= {}_0\mu_4 - 4 {}_0\mu_3 {}_0\mu_1 + 6 {}_0\mu_2 {}_0\mu_1^2 - 3 {}_0\mu_1^4 - 3 {}_0\mu_2^2 + 6 {}_0\mu_2 {}_0\mu_1^2 \\ &= \mu_4 - 3 ({}_0\mu_2 - {}_0\mu_1^2)^2 \\ &= \mu_4 - 3 \mu_2^2\end{aligned}$$

Vediamo che per calcolare i momenti centrati i cumulanti sono più comodi.

10.4 Una questione nota come “il problema dei cumulanti” riguarda la possibilità di ricostruire univocamente la f.r., nota la successione dei momenti. Ossia: date due v.c. aventi la stessa successione di momenti, esse sono uguali? La risposta è: certamente sì, se la successione dei coefficienti  ${}_0\mu_k / k!$  nell’espansione in serie è convergente.

## 11. Distribuzione del minimo e del massimo

Date  $n$  v.c.  $\{X_1, X_2, \dots, X_n\}$  aventi f.r. comune  $F_X(x)$ . Si definiscano le v.c.

$$\min = \min[X_1, X_2, \dots, X_n] \quad \text{e} \quad \max = \max[X_1, X_2, \dots, X_n]$$

ossia il valore più piccolo e quello più grande che si può ottenere in una generica estrazione  $\{X_1, X_2, \dots, X_n\}$  dalla  $n$ -upla di v.c.

Ovviamente fin quando la  $n$ -upla non è estratta,  $\min$  e  $\max$  sono v.c.

Per ottenere la f.r. di  $\max$ , si consideri che ciò equivale a ricercare la probabilità che tutte le v.c.  $X_i$  assumano valore inferiore o uguale ad un certo valore  $y$ . Questa è una v.c. *continua*, definita in  $\mathfrak{R}$ , qualunque sia la natura delle  $X_i$ : affermare che  $\max \leq y$  equivale a dire che  $\{X_1 \leq y \cap \dots \cap X_n \leq y\}$ .

Supponendo l’indipendenza delle v.c., si ha la f.r. del *massimo*:

$$F_{\max}(y) = \text{Prob}\{X_1 \leq y, \dots, X_n \leq y\} = \prod_{i=1}^n \text{Prob}\{X_i \leq y\} = \prod_{i=1}^n F_X(y) = [F_X(y)]^n$$

da cui, se la v.c.  $X$  è continua, derivando rispetto a  $y$ , possiamo ricavare la f.d.:

$$f_{\max}(y) = n f_X(y) [F_X(y)]^{n-1}.$$

La f.r. del *minimo* invece corrisponde a ricercare la probabilità che tutte le v.c.  $X_i$  assumano valore superiore a  $\min = y$ .

Ancora sotto l’ipotesi di indipendenza, si ha:

$$1 - F_{\min}(y) = \text{Prob}\{X_1 > y, \dots, X_n > y\} = \prod_{i=1}^n \text{Prob}\{X_i > y\} = [1 - F_X(y)]^n$$

la cui f.d. risulta:

$$f_{\min}(y) = n f_X(y) [1 - F_X(y)]^{n-1}$$

### PROBLEMI PER IL CAPITOLO 3

- Sui dati del Capitolo 2 (Problemi 1) si calcolino i primi quattro momenti empirici centrati.
- Si determini la mediana ed il primo ed il terzo quartile.
- Si standardizzi la variabile e si determinino i coefficienti di asimmetria e curtosi.

Risp.:

$x_i$	$n_i$	$x_i n_i$	$x_i^2 n_i$	$x_i^3 n_i$	$x_i^4 n_i$
0	31	0	0	0	0
1	41	41	41	41	41
2	20	40	80	160	320
3	6	18	54	162	486
4	2	8	32	128	512
<b>Totale</b>	<b>100</b>	<b>107</b>	<b>207</b>	<b>491</b>	<b>1359</b>

Da cui:

${}_0\mu_1$	${}_0\mu_2$	${}_0\mu_3$	${}_0\mu_4$
1.07	2.07	4.91	13.59
$\mu_1$	$\mu_2$	$\mu_3$	$\mu_4$
0	0.9251	0.7154	2.8625

- Lo stesso si faccia sui dati del Capitolo 2 (Problemi 2)
- Un'auto percorre 15 Km a 50 Km/h, 30 Km a 60 Km/h e 65 Km a 85 Km/h. Determinare la velocità media.

*Sugg.:* La velocità media è la somma dei percorsi diviso la somma dei tempi, occorre quindi calcolare la media armonica con pesi gli spazi percorsi

$$\text{Risp.: } 110 / (15 / 50 + 30 / 60 + 65 / 85) = 70.3 .$$

- Abbiamo investito un capitale di 1.000 euro con tassi di interesse composto per 3 anni al 10%, per 5 anni al 7.5% e per 2 anni all'8%. Determinare il tasso medio.

*Sugg.:* Occorre eguagliare il montante  $M$  alla fine dell'investimento del capitale  $C$  reinvestito ogni anno con gli interessi variabili maturati al montante che si otterrebbe con interesse costante

$$M = C (1 + i_1)^{n_1} (1 + i_2)^{n_2} \dots (1 + i_k)^{n_k} = C (1 + i^*)^{n_1 + n_2 + \dots + n_k}$$

La risposta si ottiene quindi applicando la media geometrica ai coefficienti di interesse  $(1 + i)$

$$\text{Risp.: } (1.1^3 * 1.075^5 * 1.08^2)^{1/10} - 1 = 8.3448\% .$$

- Cinque amici decidono di mettere insieme i propri soldi ed andare a mangiare insieme. Tre di loro hanno 10 euro, uno ha 15 euro e l'ultimo ha 20 euro. Quanto potrà spendere ciascuno di essi?

*Sugg.:* Essendo in presenza di quantità additive, si applichi la media aritmetica



$$\text{Risp.: } (3 * 10 + 1 * 15 + 1 * 20) / 5 = 13 .$$

- Abbiamo 10 lastre di metallo quadrate di fissato spessore, 5 hanno lato 10 cm, 3 lato 20 cm e 2 lato 15 cm. Fondendo il materiale e ricavando 10 lastre eguali e di spessore pari al precedente, quale sarà la lunghezza del lato?

*Sugg.:* Occorre sommare le superfici per ottenere la superficie totale. Usare la media quadratica

$$\text{Risp.: } \sqrt{(10^2 * 5 + 20^2 * 3 + 15^2 * 2) / 10} = 14.6629 .$$

- E se al posto delle lastre ci fossero cubi, quale sarebbe il lato del cubo medio?

*Sugg.:* Per ottenere il volume totale sommare i volumi: usare la media cubica.

$$\text{Risp.: } [(10^3 * 5 + 20^3 * 3 + 15^3 * 2) / 10]^{1/3} = 15.2906 .$$

3. Sui dati del Capitolo 2 (Problemi 4) si calcolino i totali marginali e le distribuzioni condizionate per  $X$  e per  $Y$ .

- Calcolare il vettore delle medie e la matrice di varianza e covarianza empirica.

$$\text{Risp.: } (2.580, 0.880), \begin{pmatrix} 1.2436 & -0.2104 \\ -0.2104 & 0.6256 \end{pmatrix} .$$

- Si calcoli la media e la varianza della v.c.  $Z = 2 * X - 3 * Y$ .

4. Data la v.c.  $X$ , simmetrica e con v.a. nullo, e la v.c.  $Y = X^2$ , si determini se le due variabili sono indipendenti in senso forte ed in senso debole.

*Risp.:* Le due v.c., a meno di casi particolari, sono dipendenti in senso forte, poiché  $f(y | x) = f(x, y) / f(x)$  in generale è diversa da  $f(y)$ ; ma sono indipendenti in senso debole, dato che:  $cov(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y) = E(X \cdot Y) - E(X^3) = 0$ .

5. Data la v.c. indicatrice, che assume solo i valori  $X = 1$ , con probabilità  $\phi$ , e  $X = 0$ , con probabilità  $\psi = 1 - \phi$ . Si calcolino i primi quattro momenti centrati.

- Data la v.c. uniforme, che ha d.p. pari a  $f(x) = k$  per  $0 \leq X \leq 1$  e  $f(x) = 0$  altrove. Si verifichi che  $f(x)$  è una f.d. determinando il valore di  $k$  opportuno; si calcolino quindi i primi quattro momenti centrati.



## CAPITOLO 4

### VARIABILI CASUALI DISCRETE

Iniziamo lo studio sistematico di alcune v.c. che saranno utili. In questo capitolo saranno trattate le v.c. discrete, nei successivi quelle continue.

#### 1. La variabile casuale binomiale

Nel Capitolo 2 abbiamo introdotto la v.c. più semplice, che ha solo due possibili valori 0 e 1, che idealmente possiamo associare all'insuccesso o al successo di un particolare esperimento aleatorio. Tale variabile è caratterizzata da un solo *parametro* che esprime la probabilità del successo  $\mathcal{P}(1) = \varphi$ ; la probabilità quindi dell'insuccesso è  $\mathcal{P}(0) = 1 - \varphi$ , che possiamo indicare con  $\psi$ , anche se certamente questo non è un parametro libero. Tale v.c. è detta *bernoulliana* (dal nome del matematico svizzero Bernoulli) e quindi è del tutto specificata una volta fissato  $\varphi$ .

Si supponga di realizzare  $n$  prove *indipendenti* che si possono considerare repliche di questa v.c., oppure al solito considerare  $n$  v.c. tutte copie della stessa. In questo caso diciamo che le v.c. sono *identicamente ed indipendentemente distribuite* (i.i.d.). La v.c. alla quale siamo interessati è il numero di successi  $X$  sulle  $n$  prove. Ovviamente questa  $X$  è una v.c. discreta che può assumere i valori interi tra 0 e  $n$ :  $x = 0, 1, \dots, n$ .

Per determinare la probabilità di ciascuna determinazione  $x$  si possono seguire due ragionamenti: uno induttivo ed uno basato sulla funzione caratteristica.

1.1 Per induzione. Si costruisca lo spazio dei possibili esiti con  $n$  pari a 2.

Tale spazio è dato dai quattro risultati:  $T_1T_2$ ,  $T_1C_2$ ,  $C_1T_2$ ,  $C_1C_2$ , che hanno probabilità:

$$\mathcal{P}(T_1T_2) = \varphi^2, \quad \mathcal{P}(T_1C_2) = \varphi\psi, \quad \mathcal{P}(C_1T_2) = \psi\varphi, \quad \mathcal{P}(C_1C_2) = \psi^2;$$

da cui la d.p. è:

$x$	$\mathcal{P}(x)$
0	$\psi^2$
1	$2\varphi\psi$
2	$\varphi^2$
	1

Portando il numero delle prove a tre si ottiene in base alle stesse considerazioni la seguente distribuzione:

$x$	$\mathcal{P}(x)$
0	$\psi^3$
1	$3\varphi\psi^2$
2	$3\varphi^2\psi$
3	$\varphi^3$
	1

Effettuando calcoli di questo genere si verifichi sempre che: a) i valori probabilità che si ottengono sono sempre positivi; b) la loro somma dia uno.

Generalizzando per  $n$  prove, si ottiene una v.c. il cui termine generico  $x$  ha probabilità:

$$\mathcal{P}(x) = \binom{n}{x} \varphi^x \psi^{n-x}.$$

Si noti anche qui che la somma delle probabilità:

$$\sum_{x=0}^n \mathcal{P}(x) = \sum_{x=0}^n \binom{n}{x} \varphi^x \psi^{n-x} = (\varphi + \psi)^n = 1^n = 1.$$

dà uno, essendo la sommatoria lo sviluppo del binomio  $(\varphi + \psi)^n$ .

1.2 Estendendo il ragionamento al caso in cui ogni esperimento abbia più di due possibili esiti,  $A_1, A_2, \dots, A_k$ , ciascuno con probabilità  $\varphi_1, \varphi_2, \dots, \varphi_k$  (con  $\varphi_1 + \varphi_2 + \dots + \varphi_k = 1$ ) nel caso di  $n$  prove indipendenti la probabilità di ottenere  $x_1$  realizzazioni di tipo  $A_1$ ,  $x_2$  di tipo  $A_2$ ,  $\dots$ ,  $x_k$  di tipo  $A_k$  (con  $x_1 + x_2 + \dots + x_k = n$ ) è data dalla distribuzione *multinomiale*:

$$\mathcal{P}(x_1, x_2, \dots, x_k) = \binom{n}{x_1 x_2 \dots x_k} \varphi_1^{x_1} \varphi_2^{x_2} \dots \varphi_k^{x_k} = \frac{n!}{x_1! x_2! \dots x_k!} \varphi_1^{x_1} \varphi_2^{x_2} \dots \varphi_k^{x_k}$$

## 2. La Binomiale come somma di Bernoulliane

Un modo più rigoroso di introdurre la v.c. *binomiale* è considerare questa variabile come somma di v.c. *bernoulliane* i.i.d.. Come abbiamo già avuto modo di notare infatti il numero  $X$  di successi su  $n$  prove si può pensare come la somma di  $n$  v.c., pari a 1 se quella prova è stata un successo e 0 se è stata un insuccesso.

Pertanto, se le prove sono indipendenti, la v.c. somma di *bernoulliane* avrà f.c. pari alla produttoria delle f.c. delle *bernoulliane*. Ora calcolare la f.c. della *bernoulliana* è facile, infatti essa risulta:

$$C_X(t) = \mathbb{E}(e^{itX}) = \sum_{x=0}^1 e^{itx} P(x) = \psi e^{it0} + \varphi e^{it1} = \varphi e^{it} + \psi$$

Da cui è possibile calcolare la f.c. della v.c. risultante dalla somma di  $n$  di queste  $X_i$  indipendenti:

$$C_{\sum_i X_i}(t) = \prod_i C_{X_i}(t) = (\varphi e^{it} + \psi)^n.$$

Se ora calcoliamo la f.c. della *binomiale*, otteniamo:

$$C_Y(t) = \mathbb{E}(e^{itY}) = \sum_{y=0}^n e^{ity} P(y) = \sum_{y=0}^n \binom{n}{y} (\varphi e^{it})^y \psi^{n-y} = (\varphi e^{it} + \psi)^n.$$

Pertanto le v.c.  $Y$  e  $\sum_i X_i$  sono uguali avendo la stessa f.c., grazie alla biunivocità tra le funzioni caratteristiche e le distribuzioni di probabilità.

Una immediata conseguenza della precedente relazione è la seguente: la somma di  $k$  v. c. *binomiali* indipendenti con lo stesso parametro  $\varphi$  e l'altro parametro anche differente  $n_1, n_2, \dots, n_k$  è una v.c. *binomiale* con parametro  $\varphi$  ed  $n = n_1 + n_2 + \dots + n_k$ . In simboli:

$$Y = X_1 + X_2 + \dots + X_k, \text{ con } X_i \propto \text{Bin}(\varphi, n_i) \text{ indep.} \Rightarrow Y \propto \text{Bin}(\varphi, \sum_i n_i).$$

ove il simbolo  $\propto$  significa “si distribuisce come”, ossia “ha f.d.” o “ha d.p.”

Per esempio, si estraggano con reimmissione  $n_1$  palline dalla prima urna,  $n_2$  dalla seconda,  $\dots$ ,  $n_k$  dalla  $k$ -esima, urne tutte di pari composizione tra palline bianche e nere; la v.c. *binomiale* risponde alla domanda: “come si distribuisce la v.c. costituita dal numero totale di palline di un certo colore estratte?”

2.1. Dalla f.g.c. della *binomiale*,  $\mathcal{K}_X(t) = n \log(\varphi e^t + \psi)$ , ricaviamo i cumulanti:

$$\mathcal{K}'(t) = n \varphi e^t (\varphi e^t + \psi)^{-1}$$

$$\mathcal{K}''(t) = n \varphi \psi e^t (\varphi e^t + \psi)^{-2}$$

$$\mathcal{K}'''(t) = n \varphi \psi e^t (\psi - \varphi e^t) (\varphi e^t + \psi)^{-3}$$

$$\mathcal{K}''''(t) = n \varphi \psi e^t [\psi^2 - 4\varphi \psi e^t + \varphi^2 e^{2t}] (\varphi e^t + \psi)^{-4}$$

Da cui:

$$\kappa_1 = E(X) = n\varphi;$$

$$\kappa_2 = \mu_2 = \text{var}(X) = n\varphi\psi;$$

$$\kappa_3 = \mu_3 = n\varphi\psi (1 - 2\varphi);$$

$$\kappa_4 = \mu_4 - 3 \mu_2^2 = n\varphi\psi (1 - 6\varphi + 6\varphi^2).$$

I primi due momenti potevano essere ricavati direttamente, ricordando che la *binomiale* è la somma di  $n$  *bernoulliane* i.i.d. Inoltre si vede che per  $\varphi = 1/2$  il momento terzo è nullo ed infatti la distribuzione risulta simmetrica.

Inoltre, calcolando gli indici di asimmetria e di curtosi, si vede che al divergere di  $n$  i due indici tendono rispettivamente a:

$$\lim_{n \rightarrow \infty} \beta_1 = \lim_{n \rightarrow \infty} \kappa_3 \kappa_2^{-3/2} = \lim_{n \rightarrow \infty} n\varphi\psi (1 - 2\varphi) (n\varphi\psi)^{-3/2} = 0$$

$$\lim_{n \rightarrow \infty} \beta_2 = \lim_{n \rightarrow \infty} \kappa_4 \kappa_2^{-2} + 3 = \lim_{n \rightarrow \infty} n\varphi\psi (1 - 6\varphi + 6\varphi^2) (n\varphi\psi)^{-2} + 3 = 3$$

In seguito saremo in grado di commentare più ampiamente questo risultato.

### 3. La convergenza in legge

Spesso è necessario considerare una *successione* di v.c. governata da un *parametro* che chiameremo  $n$ .

Serviamoci di un esempio per chiarire questo concetto.

Supponiamo di effettuare da un'urna contenente  $V$  palline bianche e  $N - V$  nere, un'estrazione; naturalmente indicando con  $X$  l'evento pallina bianca, si ha  $\mathcal{P}(x) = V/N$ .

Possiamo fare un altro esperimento: estraiamo da quell'urna *due* palline; in questo caso la v.c.  $X$  numero di palline bianche potrà risultare pari a 0 o 1 o 2. Se le due estrazioni vengono effettuate in modo *indipendente*, per esempio reimbussolando ogni pallina una volta estratta prima di estrarre la successiva, come sappiamo, la v.c.  $X$  è una *binomiale*.

Ci si può chiedere: se il numero di estrazioni effettuate in queste condizioni cresce fino a raggiungere valori "molto grandi" che cosa succede alla nostra v.c.? Si noti che in questo caso la cosa ha non solo un interesse teorico, ma anche pratico, tenendo conto della laboriosità del calcolo del numero combinatorio  ${}_k C_n$  quando  $n$  cresce e  $k$  è vicino a  $n/2$ .

In altre parole abbiamo creato una successione di v.c. tra loro simili, in quanto sono tutte *binomiali*, ma diverse per il fatto che il parametro  $n$  va crescendo secondo i numeri naturali.

Più formalmente possiamo definire il concetto di *convergenza in legge o distribuzione*. Questo non è un limite matematico bensì stocastico. Se la *successione delle f.r.*  $\{F_{X_n}(x)\}$ , corrispondenti alla successione delle v.c.  $\{X_n\}$ , ha come limite matematico una f.r.  $F_Y(x)$ ; la v.c.  $Y$  che corrisponde a tale f.r. sarà la v.c. a cui converge in legge la successione delle v.c.

$$\begin{aligned} \{X_1, \quad X_2, \quad \dots \quad X_n, \quad \dots\} &\xrightarrow{L} Y \\ \{F_{X_1}(x), \quad F_{X_2}(x), \quad \dots \quad F_{X_n}(x), \quad \dots\} &\rightarrow F_Y(x) \\ \{C_{X_1}(t), \quad C_{X_2}(t), \quad \dots \quad C_{X_n}(t), \quad \dots\} &\rightarrow C_Y(t) \end{aligned}$$

Spesso il limite matematico  $F_Y(x)$  non è facilmente calcolabile, mentre lo è il limite  $C_Y(t)$  della successione delle f.c.  $\{C_{X_n}(t)\}$ . Allora ci viene incontro la proprietà di biunivocità tra f.c. e f.r. che – come si dimostra – si mantiene anche per i limiti di successioni. Possiamo quindi determinare la v.c. limite  $Y$  come quella corrispondente alla f.c.  $C_Y(t)$ .

#### 4. La variabile casuale Poisson

Quando si costruiscono queste successioni di v.c. spesso bisogna però determinare delle restrizioni. Nel caso prima ipotizzato per esempio è chiaro che la v.c. che si ottiene avrà media e varianza infinite al divergere di  $n$  e quindi non sarà maneggiabile. Si può far sì che il v.a. si mantenga costante:

$$E(X) = n\varphi = \mu,$$

facendo tendere a zero  $\varphi$  con la stessa velocità con cui  $n$  tende ad infinito.

Effettuando il limite della f.p. sotto questa condizione (essendoci anche qui corrispondenza biunivoca tra f.r. e f.p., la v.c. limite che si determina sarà la stessa), si ottiene:

$$\begin{aligned} \mathcal{P}(x) &= \lim_{n \rightarrow \infty} \binom{n}{x} \varphi^x \psi^{n-x} = \lim_{n \rightarrow \infty} \frac{n(n-1) \dots (n-x+1)}{x!} \frac{\mu^x}{n^x} \left(1 - \frac{\mu}{n}\right)^{n-x} \\ &= \frac{\mu^x}{x!} \lim_{n \rightarrow \infty} \frac{n}{n} \frac{n-1}{n} \dots \frac{n-x+1}{n} \lim_{n \rightarrow \infty} \left(1 - \frac{\mu}{n}\right)^n \lim_{n \rightarrow \infty} \left(1 - \frac{\mu}{n}\right)^{-x}. \end{aligned}$$

Il primo ed il terzo limite valgono 1, mentre il secondo è un limite notevole.

Si ha quindi la d.p. della v.c. *Poisson*:

$$\mathcal{P}(x) = \frac{\mu^x}{x!} e^{-\mu}.$$

Questa, per come è stata costruita, è una v.c. discreta con  $x = 0, 1, 2, \dots$ . Si noti la relazione recursiva per il calcolo delle probabilità:

$$\mathcal{P}(x) = \mathcal{P}(x-1) \mu / x.$$

Ricaviamo la f.g.m. e la f.g.c.:

$$\mathcal{M}(t) = \sum_{x=0}^{\infty} \frac{\mu^x}{x!} e^{-\mu} e^{tx} = e^{-\mu} \exp\{\mu e^t\}, \quad \mathcal{K}(t) = \log[\mathcal{M}(t)] = -\mu + \mu e^t$$

Pertanto media, varianza, momento terzo sono tutti pari a  $\mu$ .

Ciò era ottenibile portando al limite per  $n \rightarrow \infty$ , con  $n\phi = \mu$ , i tre momenti.

Ricavando gli indici di asimmetria e curtosi  $\beta_1$  e  $\beta_2$ , si vede che per  $\mu \rightarrow \infty$ :

$$\lim_{\mu \rightarrow \infty} \beta_1 = \lim_{\mu \rightarrow \infty} \mu_3 / \sigma^3 = \lim_{\mu \rightarrow \infty} \kappa_3 / \sigma^3 = \lim_{\mu \rightarrow \infty} \mu / \mu^{3/2} = 0$$

$$\lim_{\mu \rightarrow \infty} \beta_2 = \lim_{\mu \rightarrow \infty} \mu_4 / \sigma^4 = \lim_{\mu \rightarrow \infty} \kappa_4 / \sigma^4 + 3 = \lim_{\mu \rightarrow \infty} \mu / \mu^2 + 3 = 3.$$

4.1 Dalla f.g.m. possiamo ricavare un'altra importante proprietà: la somma di  $k$  v.c. indipendenti distribuite secondo *Poisson* ognuna con parametro  $\mu_i$ , è una v.c. Poisson con  $\mu$  la somma dei  $\mu_i$ . In simboli:

$$Y = X_1 + X_2 + \dots + X_k, \text{ con } X_i \propto \text{Pois}(\mu_i) \text{ indep.} \Rightarrow Y \propto \text{Pois}(\sum_i \mu_i).$$

4.2 La v.c. *Poisson* trova applicazione quando si è in presenza di eventi con probabilità molto bassa. In particolare si può usare la *Poisson* come approssimazione della *binomiale* quando  $n$  è molto grande e  $\phi$  (o  $\psi$ ) molto piccolo. In genere si ritiene che l'approssimazione sia buona per  $n > 30$  e  $n\phi < 5$ . In questo caso, come si può capire facilmente, si evita il calcolo del numero combinatorio che diventa sempre più pesante all'aumentare di  $n$ .

## 5. La variabile casuale Binomiale negativa

Nell'introdurre la v.c. *binomiale* abbiamo fatto riferimento al seguente schema: si ripete un certo esperimento nelle stesse condizioni ed in modo indipendente un certo numero di volte.

Qui ci poniamo nella stessa situazione (i.i.d.) ma fissiamo il numero di successi, mentre teniamo come v.c. il numero di prove da effettuare per raggiungere quel numero di successi. Per ottenere ciò, le prime  $n-1$  prove dovranno presentare  $k-1$  successi in una successione qualsiasi, mentre l' $n$ -esima prova dovrà essere un successo.

A causa dell'indipendenza delle prove possiamo scrivere:

$$\mathcal{P}(n) = \phi \text{ Bin}_{n-1}(k-1) = \binom{n-1}{k-1} \phi^{k-1} \psi^{n-k} \phi = \binom{n-1}{k-1} \phi^k \psi^{n-k} = \frac{(n-1)!}{(n-k)!(k-1)!} \phi^k \psi^{n-k}$$

Riparametrizzando con  $y = n - k$  si ha:

$$\mathcal{P}(y) = \frac{(k+y-1)!}{y!(k-1)!} \phi^k \psi^y.$$



Mentre prima  $n$  era definito sugli interi a partire da  $k$  (infatti, per avere  $k$  successi, almeno  $k$  prove si devono fare), ora  $y$  è definito sui naturali:  $y = 0, 1, \dots$ , ossia il numero di prove da effettuare *oltre*  $k$  per ottenere il  $k$ -esimo successo. Questa è la d.p. della v.c. detta *binomiale negativa*.

5.1 Il termine “*negativa*” nasce dall’esponente che si trova nello sviluppo in serie qui sotto riportato.

Per verificare che  $\sum_{y=0}^{\infty} \mathcal{P}(y) = 1$  si consideri la serie:

$$\varphi^{-k} = \sum_{y=0}^{\infty} \frac{(-1)^y}{y!} k(k+1) \dots (k+y-1) (\varphi-1)^y.$$

Ricaviamo la f.g.m. e la f.g.c.:

$$\mathcal{M}_Y(t) = \sum_{y=0}^{\infty} \binom{k+y-1}{k-1} \varphi^k [1 - (\varphi) e^t]^y = \varphi^k [1 - (1 - \varphi) e^t]^{-k}.$$

$$\mathcal{K}_Y(t) = k \log(\varphi) - k \log[1 - (1 - \varphi) e^t];$$

le cui derivate sono:

$$\mathcal{K}_Y'(t) = \frac{k e^t (1 - \varphi)}{1 - (1 - \varphi) e^t};$$

$$\mathcal{K}_Y''(t) = \frac{k e^t (1 - \varphi) [1 - (1 - \varphi) e^t] - k (1 - \varphi) e^t [- (1 - \varphi) e^t]}{[1 - (1 - \varphi) e^t]^2}.$$

Da cui:

$$\mathbb{E}(Y) = k\psi / \varphi \quad \text{e} \quad \text{var}(Y) = k\psi / \varphi^2.$$

Come si vede all’aumentare di  $\varphi$  la media del *ritardo* e la dispersione della v.c. sono sempre più basse; mentre al diminuire di  $\varphi$  (e all’aumentare di  $\psi$ ) la media e la varianza aumentano, cioè non solo è più lunga l’attesa media, ma il fenomeno diventa sempre più disperso.

## 6. Riepilogo sulle variabili casuali bernoulliane

Le tre v.c. che abbiamo fin qui esaminato provengono dallo schema bernoulliano. È interessante osservare le relazioni tra media e varianza:

	$\mathbb{E}(X)$		$\text{var}(X)$
<i>Binomiale</i>	$n\varphi$	$>$	$n\varphi\psi$
<i>Poisson</i>	$\mu$	$=$	$\mu$
<i>Bin. negativa</i>	$k\psi / \varphi$	$<$	$k\psi / \varphi^2$

## 7. La variabile casuale ipergeometrica

Supponiamo di avere un'urna di composizione nota, diciamo  $M$  palline bianche e  $N - M$  nere, e di effettuare un'estrazione di  $n$  palline a caso ma *senza reimbussolamento* o, che è lo stesso, *in blocco*. Vediamo subito che questi due schemi si equivalgono, infatti siamo interessati al numero di palline bianche  $v$  presenti nel campione estratto e ciò è indifferente rispetto al fatto che le palline siano estratte ad una ad una o tutte insieme.

Cosa cambia rispetto allo schema bernoulliano? Che le prove successive *non* sono indipendenti, infatti la probabilità di estrarre una pallina bianca va cambiando a secondo di quante palline bianche si vanno estraendo.

Per poter formulare una v.c. è conveniente, come sappiamo, partizionare lo spazio campionario: in questo caso possiamo far riferimento a tutti i possibili risultati. Essi sono pari al numero combinatorio  ${}_nC_N = N! / [(N - n)! n!]$  che esprime il numero di tutte le distinte *combinazioni* di  $N$  oggetti ad  $n$  ad  $n$ , indipendentemente dall'ordine, che appunto a noi qui non interessa, in conseguenza della v.c. che stiamo studiando. Tali combinazioni le giudichiamo *equiprobabili*, in quanto ognuna delle palline, delle coppie, delle terne, ... , dei gruppi di  $n$  palline, ha la stessa probabilità *a priori* di essere estratta. Fra questi possibili esiti, quali presentano esattamente  $v$  palline bianche? Limitiamo la nostra attenzione al sottogruppo costituito da  $M$  palline bianche nell'urna; di queste,  $v$  dovranno venire a far parte del campione e quindi il modo di scegliere tali palline da estrarre è pari a  ${}_vC_M$ , mentre per ciascuna di tali scelte vi sono  ${}_{n-v}C_{N-M}$  possibili modi di estrarre le  $n - v$  nere dal gruppo delle  $N - M$  dell'urna.

In conclusione la probabilità cercata è:

$$\mathcal{P}(v) = \frac{\binom{M}{v} \binom{N-M}{n-v}}{\binom{N}{n}}.$$

Per accertarsi che questa è una d.p. occorre sommare i suoi valori su tutto il campo di definizione della v.c., campo che è più facile descrivere per esclusione: tutti i valori  $v$  che danno senso ai combinatori del numeratore. Formalmente:

$$\min(v) = \max(0, n - N + M); \quad \max(v) = \min(M, n).$$

7.1 Per ricavare i primi due momenti non possiamo fare riferimento alla f.g.m. in quanto la sua espressione è troppo complicata.

Per ricavare la media occorre considerare che anche questa è una somma di v.c. *bernoulliane*, ma non i.i.d. Ricordando che la *bernoulliana* ha v.a. pari alla probabilità di successo, la v.c. relativa alla prima estrazione ha v.a. pari a

$$\mathbb{E}(X_1) = \mathbb{P}(X_1) = M/N = \phi,$$

la seconda:

$$\begin{aligned}\mathbb{E}(X_2) &= \mathbb{P}(X_2) = \mathbb{P}(X_2 \cap X_1) + \mathbb{P}(X_2 \cap \bar{X}_1) = \mathbb{P}(X_2 | X_1) \mathbb{P}(X_1) + \mathbb{P}(X_2 | \bar{X}_1) \mathbb{P}(\bar{X}_1) \\ &= \frac{M-1}{N-1} \frac{M}{N} + \frac{M}{N-1} \frac{N-M}{N} = \frac{M^2 - M + N M - M^2}{N(N-1)} = \frac{M}{N}.\end{aligned}$$

Quindi la probabilità *non condizionata* della seconda estrazione è uguale alla prima. Per induzione si può dimostrare che anche le successive hanno v.a. costante e pari a  $M/N$ .

Analogo ragionamento porta a calcolare il momento secondo *non condizionata* della  $i$ -esima estrazione, che risulta pari a quello della prima, e quindi la varianza risulta pari a  $\text{var}(X_i) = \phi(1-\phi) = \phi\psi$ .

Il v.a. di  $V$  si può ottenere quindi sommando i valori attesi di  $X_1, \dots, X_n$ :

$$\mathbb{E}(V) = \mathbb{E}(X_1) + \dots + \mathbb{E}(X_n) = n M/N.$$

Se si pensa ad  $M/N$  come al parametro  $\phi$  della v.c. *bernoulliana*, si vede che *binomiale* ed *ipergeometrica* hanno lo stesso v.a.

Per calcolare la  $\text{var}(V)$  si deve far riferimento sempre alla somma di *bernoulliane* non i.i.d., considerandone la covarianza:

$$\begin{aligned}\text{cov}(X_1, X_2) &= \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1) \mathbb{E}(X_2) = \mathbb{P}(X_1 \cap X_2) - \mathbb{P}(X_1) \mathbb{P}(X_2) = \\ &= \frac{M-1}{N-1} \frac{M}{N} - \frac{M^2}{N^2} = -\frac{M(N-M)}{N^2(N-1)} = -\frac{\phi\psi}{(N-1)}\end{aligned}$$

Anche qui per induzione si può dimostrare che tutte le altre coppie  $i$  e  $j$  hanno la stessa covarianza non condizionata. Pertanto possiamo scrivere:

$$\text{var}(V) = \sum_{i=1}^n \text{var}(X_i) + 2 \sum_{i \neq j}^n \text{cov}(X_i, X_j) = n\phi\psi - 2 \frac{n(n-1)}{2} \frac{\phi\psi}{(N-1)}$$

Da cui:

$$\text{var}(V) = n\phi\psi \frac{N-n}{N-1}.$$

Si può osservare che la varianza dell'*ipergeometrica* è inferiore a quella della *binomiale*, per il fattore correttivo  $(N-n)/(N-1)$ . Ovviamente questo fattore correttivo è uguale ad 1 per  $n=1$  (estrarre un'unità con o senza reimmissione è lo stesso) o per  $N \rightarrow \infty$  (infatti reimbussolare, o no, altera trascurabilmente la composizione di un'urna con un numero grandissimo di palline). Infatti in quest'ultimo caso la distribuzione *ipergeometrica* tende in legge alla distribuzione *binomiale*. Non diamo di ciò una dimostrazione formale, essendo quella intuitiva già sufficiente.

## 8. Distribuzione delle medie di posizione

Data l' $n$ -upla di v.c.  $\{X_1, X_2, \dots, X_n\}$  ed una generica estrazione da essa  $\{x_1, x_2, \dots, x_n\}$ , che, una volta ordinata in ordine crescente, scriviamo come  $\{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$  e dato un valore  $k$  compreso tra 1 ed  $n$ , si ha la *media di posizione* (o *statistica d'ordine*), data dal valore  $x_{(k)}$ . Ovviamente fin tanto che l'estrazione non è stata effettuata, questo valore è una v.c. e lo indicheremo con  $X_{(k)}$ . Si noti che per  $k = 1$  si ha il *minimo*, mentre per  $k = n$  si ha il *massimo*.

Tali medie *non algebriche*, calcolate su un insieme di  $n$  osservazioni, corrispondono ai quantili delle v.c.

Si desidera ora determinare la f.d. di una generica  $X_{(k)}$ , esprimente la probabilità che la  $k$ -esima osservazione, su  $n$  estratte, assuma un certo valore compreso tra  $y$  e  $y + dy$ . Essa, sfruttando la distribuzione *trinomiale* (*multinomiale* per tre possibili risultati), dà:

$$f_{X_{(k)}}(y) dy = \text{Prob}\{X_{(k)} \leq y\} =$$

$$\text{Prob}\{(k-1) \text{ delle } X_i \leq y \cap \text{una } X_i \text{ cada tra } (y \text{ e } y+dy) \cap (n-k) \text{ delle } X_i > y\} =$$

$$= \frac{n!}{(k-1)! 1! (n-k)!} [F_X(y)]^{k-1} f(y) [1 - F_X(y)]^{n-k} dy.$$

Analogamente per calcolare la f.d. congiunta di  $X_{(l)}$  e  $X_{(k)}$  (con  $l < k < n$ ):

$$f_{X_{(l)}, X_{(k)}}(y, z) dy dz = \text{Prob}\{X_{(l)} \leq y \cap X_{(k)} \leq z\} =$$

$$= \frac{n!}{(l-1)! 1! (k-l-1)! 1! (n-k)!} [F_X(y)]^{l-1} f_X(y) [F_X(z) - F_X(y)]^{k-l-1} f_X(z) [1 - F_X(z)]^{n-k} dy dz$$

facilmente generalizzabile per più medie di posizione.

8.1 Definiamo ora: a) la *mediana* campionaria, come la statistica d'ordine centrale se  $n$  è dispari e la media delle due statistiche d'ordine centrali se  $n$  è pari, b) il *campo di variazione* (*range*) campionario, come  $X_{(n)} - X_{(1)}$  e c) il *valore centrale* (*midrange*), come  $[X_{(n)} + X_{(1)}] / 2$ . Le relative f.d. si ottengono dalla formula precedente per opportune trasformazioni ed integrazioni.

Per esempio per ottenere la f.d. del campo di variazione e del valore centrale, ponendo  $l = 1$  e  $k = n$ , e si eseguano le trasformazioni:

$$R = X_{(n)} - X_{(1)} \quad \text{e} \quad T = (X_{(1)} + X_{(n)}) / 2$$

da cui:

$$X_{(n)} = T + R / 2 \quad \text{e} \quad X_{(1)} = T - R / 2$$

avente *Jacobiano* pari a  $-1$ . Da ciò si ottiene:

$$f_{R, T}(r, t) = n(n-1) f(t-r/2) [F_X(t+r/2) - F_X(t-r/2)]^{n-2} f(t+r/2)$$

che, integrata una volta su  $r$  e una volta su  $t$ , danno quanto richiesto.

#### PROBLEMI PER IL CAPITOLO 4

1. Una macchina produce pezzi con difettosità pari al 5% e li inscatola a gruppi di 10. Calcolare la probabilità che in una scatola vi sia più di un pezzo difettoso.

*Risp.:* 0.086138 .

- Se si effettua un controllo basato sull'esame di tre scatole e l'intero lotto viene scartato se più di una scatola ha più di un pezzo difettoso, qual è la probabilità che il lotto non superi il controllo?

*Risp.:* 0.02098 .

- Se si effettua un controllo basato sull'esame di tre scatole e l'intero lotto viene scartato se complessivamente si trovano più di 3 pezzi difettosi, qual è la probabilità che il lotto non superi il controllo?

*Sugg.:* Considerare il numero di pezzi difettosi come la somma di quelli trovati nelle tre scatole.

*Risp.:* 0.0608 .

- Un controllo viene effettuato scartando l'intero lotto se, esaminando 3 scatole da 5 pezzi, viene trovato anche solo un pezzo difettoso. Fissare la difettosità del processo in modo tale da limitare all'1% la probabilità di vedersi scartare il lotto.

*Risp.:* 0.00067 .

2. Da un magazzino composto da migliaia di pezzi, si estraggono 50 pezzi. Determinare qual è la probabilità di trovare pezzi difettosi se la difettosità della produzione è pari all'1%.

*Sugg.:* Usare l'approssimazione data dalla *Poisson*.

*Risp.:* 0.3935 .

- Qual è la probabilità di trovare più di due pezzi difettosi?
  - Vi sarebbe differenza se i pezzi fossero raggruppati in scatole da 10 pezzi e si desiderasse sapere la probabilità di ottenere più di due pezzi difettosi complessivamente? Perché?
  - Ed invece se i pezzi fossero raggruppati in scatole da 25, e si desiderasse sapere la probabilità che in ognuna delle due scatole non vi è più di un pezzo difettoso?
3. Un controllo viene effettuato scartando l'intera produzione se viene trovato anche solo un pezzo difettoso su cinquanta. Determinare quanto dev'essere la difettosità per limitare all'1% il rischio di tale eventualità.

*Risp.:* 0.0002 .

- Qual è la probabilità in questo caso che capitino più di due pezzi difettosi?

4. Il magazzino ci fornisce a caso pezzi di tipo  $A$  e di tipo  $B$ , con eguale probabilità. Qual è la probabilità che al decimo pezzo richiesto ancora non sia arrivato il terzo pezzo di tipo  $A$ ?

*Sugg.: Usare la binomiale negativa.*

- Determinare quanti pezzi occorre prelevare affinché sia minore dell'1% la probabilità di non avere almeno 3 pezzi di tipo  $A$ .

5. A tutti è noto il gioco del lotto. Da un'urna contenente 90 palline numerate, si estraggono senza reimmissione 5 palline. Ciò viene ripetuto per 10 diverse urne dette "ruote". Si supponga di giocare una "terno secco", ossia di scommettere sull'uscita di tre numeri. Quanto dovrebbe essere pagata la scommessa in caso di vincita se il gioco fosse equo?

*Sugg.: Usare l'ipergeometrica con  $N = 90$ ,  $M = 3$ ;  $n = 5$ ;  $v = 3$ .*

- Se si giocasse una quaterna su tutte le ruote, quale sarebbe la probabilità di fare almeno un ambo su almeno una ruota?

*Sugg.: Si calcoli la probabilità di fare ambo, terna o quaterna su una ruota e poi si applichi la regola della somma delle probabilità per eventi *compatibili* ed *indipendenti*.*

6. Da un mazzo di 52 carte ne abbiamo ricevuto 5. Qui troviamo che vi sono due Assi. Qual è la probabilità, chiedendo altre tre carte:

- di ricevere un altro Asso, facendo "tris"?
- di ricevere altri due Assi, facendo "poker"?
- di ricevere una coppia di altre carte, facendo "doppia coppia"?
- di ricevere un tris di altre carte, facendo "full" non d'Assi?
- di fare "full" d'Assi?
- di migliorare in qualunque modo il nostro gioco in uno dei modi precedenti?

## CAPITOLO 5

### VARIABILI CASUALI CONTINUE: I

Iniziamo ora lo studio delle v.c. continue. Come abbiamo già chiarito, la v.c. continua è tale perché i valori che può assumere sono in corrispondenza biunivoca con un segmento o una retta o una semiretta. Tipicamente le sue realizzazioni sono valori quali: la resa di una reazione, la lunghezza di un pezzo prodotto, il tempo entro cui si verifica un arrivo o un guasto.

#### 1. La variabile casuale uniforme

La più semplice delle v.c. continue è la v.c. *uniforme*. Essa è definita su un segmento reale  $[a, b]$  limitato ove la f.d. si mantiene costante.

A fini computazionali, non ha importanza se tale intervallo è chiuso o aperto, in quanto l'area complessiva sottesa alla curva non si modifica se non infinitesimamente se gli estremi  $a$  e  $b$  sono inclusi o no.

Per ricavare tale costante si ricordi che l'area sottesa alla curva dev'essere:

$$\int_a^b k \, dx = k(b-a) = 1.$$

pertanto  $k = 1/(b-a)$ . I momenti si ricavano anche senza l'uso della f.g.m.:

$$\mathbb{E}(X^m) = \frac{1}{b-a} \int_a^b x^m \, dx = \frac{b^{m+1} - a^{m+1}}{(m+1)(b-a)}.$$

È facile da qui ricavare media e varianza, per il caso più usuale in cui  $b = 1$  ed  $a = 0$ , v.c. che d'ora in poi indicheremo brevemente con  $U$ :

$$\mathbb{E}(U) = 1/2 \quad \text{e} \quad \text{var}(U) = 1/3 - 1/4 = 1/12.$$

1.1 Questa v.c. gode di un'importante proprietà: la f.r. di *qualunque* v.c. è una  $U$ .  
 Infatti, data  $y = F(x)$ , essendo la  $F()$  una funzione monotona non decrescente, si ha:  $\mathcal{P}(Y \leq y) = \mathcal{P}(X \leq x)$ , ossia  $F(y) = F(x)$  che, opportunamente differenziato, dà  $dF(y) = dF(x)$ , cioè  $f(y) dy = f(x) dx$ . Pertanto dev'essere:  $f(y) = 1$ .

## 2. Tasso di guasto e funzione di affidabilità

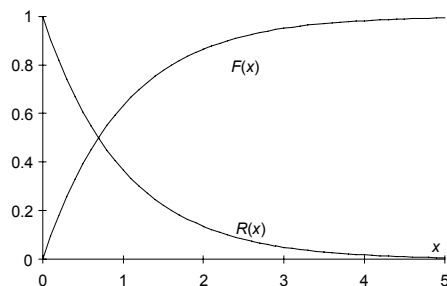
Prima di andare avanti con lo studio delle v.c. è opportuno definire due nuove funzioni: la *funzione di affidabilità* (f.a.), definita come:

$$R(x) = 1 - F(x)$$

e la funzione *tasso di guasto* (t.g.) o *tasso di arrivo*, definita come:

$$\lambda(x) = f(x) / R(x).$$

La prima rappresenta la probabilità che  $X > x$ .



Usualmente le v.c. su cui calcoliamo la f.a. esprimono il tempo al quale si verifica un fenomeno, quale il guasto di un componente. In questi casi se la f.r. esprime la probabilità che il fenomeno si verifichi *entro* il tempo  $x$ , la f.a. esprime la probabilità che il fenomeno *non* si verifichi entro il tempo  $x$ .

Pertanto in questi casi l'affidabilità misura la probabilità che al trascorrere del tempo il sistema sia ancora funzionante (non si sia verificato il guasto).

Allora il t.g., opportunamente moltiplicato per il differenziale, esprime la probabilità che il fenomeno si verifichi al tempo  $x$  – o meglio nell'intervallo infinitesimale compreso tra  $x$  e  $x + dx$  – *ammesso che non si sia verificato prima*:

$$\begin{aligned} \lambda(x) dx &= \mathcal{P}(x \leq X \leq x + dx \mid X > x) = \mathcal{P}(x \leq X \leq x + dx \cap X > x) / \mathcal{P}(X > x) \\ &= \mathcal{P}(x < X \leq x + dx) / \mathcal{P}(X > x) = f(x) dx / R(x) \end{aligned}$$

ricordando il significato della funzione di densità e di affidabilità, da cui si ha:

$$\lambda(x) = f(x) / R(x).$$



### 3. La variabile casuale esponenziale

Ricerchiamo quella v.c. che ha t.g. costante nel tempo.

Dobbiamo risolvere l'equazione differenziale:  $\lambda(x) = f(x) / R(x) = \text{cost.}$

$$\lambda(t) = f(x) / R(x) = \frac{dF(x)/dx}{[1 - F(x)]} = -d \log[1 - F(x)]/dx = \lambda;$$

da cui, separando i differenziali e integrando, si ha:

$$\log[1 - F(x)] = -\lambda x + k; \quad 1 - F(x) = e^{-\lambda x} e^k; \quad F(x) = 1 - e^{-\lambda x} e^k$$

(ove la costante di integrazione  $k$  risulta pari a zero, imponendo il campo di definizione  $x > 0$ , per cui si ha  $\lim_{x \rightarrow 0} F(x) = 0$ ). Quindi:

$$R(x) = 1 - F(x) = \exp\{-\lambda x\}; \quad F(x) = 1 - \exp\{-\lambda x\}; \quad f(x) = \lambda \exp\{-\lambda x\}.$$

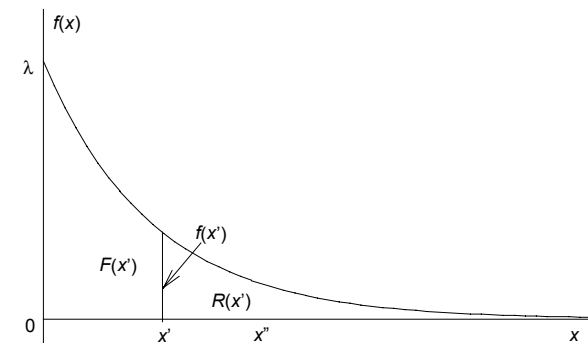
3.1 Abbiamo così ricavato la f.d. della v.c. *esponenziale*, definita per  $x \in \mathbb{R}^+$ , che dipende da un solo parametro  $\lambda$ , che chiameremo *parametro di scala*, perché esso può essere eliminato con la trasformazione  $y = \lambda x$ .

Infatti:  $f(y) dy = f(x) dx; \quad f(y) = f(x) dx / dy;$

ove:  $dx / dy = 1 / \lambda,$

da cui:  $f(y) = \lambda \exp\{-y\} / \lambda = \exp\{-y\}.$

Questa v.c. è per costruzione quella che ha t.g. costante nel tempo.



Vediamo che questa caratteristica comporta che un fenomeno il cui verificarsi nel tempo è regolato da questa v.c. possiede una notevole proprietà: il fenomeno si dice *senza memoria*. Infatti si desidera conoscere qual è l'affidabilità di una v.c. al tempo  $x''$  ammesso che si sia accertato che al tempo  $x'$  esso è ancora funzionante:

$$\begin{aligned}\mathcal{P}(X > x'' \mid X > x') &= \mathcal{P}(X > x'' \cap X > x') / \mathcal{P}(X > x') = \mathcal{P}(X > x'') / \mathcal{P}(X > x') \\ &= \mathcal{R}(x'') / \mathcal{R}(x') = \exp\{-\lambda(x'')\} / \exp\{-\lambda(x')\} = \exp\{-\lambda(x'' - x')\}.\end{aligned}$$

Dall'espressione precedente si può notare che tale affidabilità si riduce a quella calcolata spostando l'origine dei tempi da 0 a  $x'$ . Quindi, assicurarsi che il sistema funziona al tempo  $x'$  equivale a sostituirlo in quell'istante con uno nuovo. Si badi bene però che ciò non vuol dire certo che il sistema non si guasterà mai! Significa solo che la probabilità che l'evento si verifichi non aumenta col trascorrere del suo funzionamento: non ha *usura*; in realtà, man mano che il sistema è *esposto* al fenomeno di guasto, aumenta la probabilità di guasto solo perché è più lungo l'intervallo di esposizione. Un sistema che ha funzionato una quantità indefinita di tempo, se è certo che è ancora funzionante, ed un sistema nuovo hanno pari affidabilità.

3.2 Si supponga di avere un materiale radioattivo composto da  $n$  atomi e che la distribuzione di probabilità del tempo al quale ogni singolo atomo decade, emettendo una particella, sia regolata da una v.c. *esponenziale* con parametro  $\lambda$ . Se  $P$  è la probabilità che un atomo decada durante un certo intervallo di tempo, in ogni intervallo quante particelle ci si aspetta che vengano emesse? All'inizio vi sono  $n$  atomi e quindi all'inizio ci si aspetta che vengano emesse  $nP$  particelle. Dopo un certo tempo gli atomi ancora non decaduti saranno  $n - m$ . È vero che gli atomi che non sono finora decaduti "riazzerano" la propria affidabilità, ma ora sono in numero minore rispetto all'inizio; e quindi il numero atteso di emissioni è  $(n - m)P$ . Di ciò si ha verifica nel fatto che un pezzo di materiale radioattivo emette all'inizio una quantità maggiore di particelle e via via un numero sempre minore: non perché diminuisca la probabilità di decadere di ogni singolo atomo, ma perché diminuiscono gli atomi esposti al fenomeno. Se si ammette che in ogni istante il numero di particelle emesse è pari al numero atteso, moltiplicando per  $n$  la  $\mathcal{R}(x)$ , si ottiene il numero di atomi che si prevede non saranno ancora decaduti al tempo  $x$ .

Più grande è  $\lambda$ , più alta è la *velocità di decadimento*.

3.3 La f.g.m. e la f.g.c. della v.c. *esponenziale* sono:

$$\mathcal{M}_X(t) = \lambda \int_0^{\infty} e^{-\lambda x} e^{tx} dx = \frac{\lambda}{\lambda - t} \quad \text{e} \quad \mathcal{K}_X(t) = \log(\lambda) - \log(\lambda - t),$$

Da cui ricaviamo:

$$\mathcal{E}(X) = 1 / \lambda \quad \text{e} \quad \text{var}(X) = 1 / \lambda^2.$$

È possibile anche calcolare il tempo di dimezzamento, corrispondente all'istante nel quale ci si aspetta che il numero di atomi non decaduti siano pari alla metà di quelli che erano al tempo di inizio osservazione:

$$\mathcal{R}(x_{1/2}) = \mathcal{F}(x_{1/2}) = \exp\{-\lambda x_{1/2}\} = 1/2, \text{ da cui: } x_{1/2} = \log(2) / \lambda = \log(2) \mathcal{E}(X).$$

Come si vede il v.a. è più alto della mediana, cioè  $x_{1/2} = \log(2) \mathcal{E}(X)$ , ossia la metà degli elementi di una popolazione durano meno del 69.31% della vita attesa; e ancora: il 63.21% =  $\mathcal{F}[\mathcal{E}(X)]$  degli elementi durano meno della vita attesa. È a causa della forte asimmetria positiva che gli elementi, che durano di più della vita mediana, possono anche durare molto di più, innalzando così la vita media stessa.

#### 4. La variabile casuale Weibull

La v.c. *Weibull* è caratterizzata dal fatto di avere t.g. esponenziale:

$$\lambda(x) = f(x) / \mathcal{R}(x) = -\partial \log[\mathcal{R}(x)] / \partial x = \lambda (x - \xi)^{\alpha-1}$$

$$\log[\mathcal{R}(x)] = -\lambda \int (x - \xi)^{\alpha-1} dx = -(\lambda/\alpha) (x - \xi)^\alpha$$

$$\mathcal{R}(x) = \exp\{-\lambda (x - \xi)^\alpha / \alpha\}; \quad \mathcal{F}(x) = 1 - \exp\{-\lambda (x - \xi)^\alpha / \alpha\}.$$

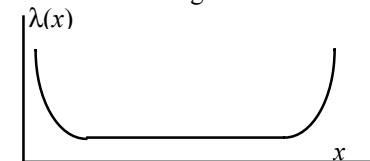
Da cui:

$$f(x) = \lambda \exp\{-\lambda (x - \xi)^\alpha / \alpha\} (x - \xi)^{\alpha-1}.$$

Tuttavia attraverso la trasformazione  $y = \lambda (x - \xi)^\alpha / \alpha$ , essa si riconduce alla v.c. *esponenziale* avente  $\lambda = 1$ . Pertanto per calcolare per esempio la f.r. si può fare uso della trasformazione inversa:

$$x = (\alpha y / \lambda)^{1/\alpha} + \xi.$$

Questa v.c. spesso è usata per descrivere dei fenomeni che hanno un t.g. variabile nel tempo. Per esempio vi sono fenomeni, quali per esempio dei prodotti molto “giovani”, che possono presentare un t.g. *decescente*; infatti possono essere caratterizzati da una elevata “mortalità infantile”. Altri prodotti “maturi” hanno t.g. pressoché *costante*, in quanto non sono molto soggetti ad usura. Altri ancora, che chiameremo “obsoleti”, hanno spesso t.g. *crescente*, in quanto l'usura ha sempre più un ruolo determinante nella vita di quel prodotto. Pertanto si suole descrivere il t.g. della vita del prodotto come una funzione che ha una forma tipica a “vasca da bagno”.



## 5. La variabile casuale Gamma

Come si può osservare nella v.c. esponenziale:

$$d \log[1 - \mathcal{F}(x)] / dx = -f(x) / [1 - \mathcal{F}(x)] = k \Leftrightarrow d \log[f(x)] / dx = f'(x) / f(x) = k.$$

Infatti:

$$\frac{d^2 \log[1 - \mathcal{F}(x)]}{dx^2} = - \frac{f'(x) [1 - \mathcal{F}(x)] + f^2(x)}{[1 - \mathcal{F}(x)]^2} = 0;$$

da cui:

$$\frac{f(x)}{1 - \mathcal{F}(x)} = \frac{f'(x)}{f(x)} = k.$$

Possiamo ora ricavare la v.c. con f.d. che soddisfa l'equazione:

$$d \log[f(z)] / dz = -z^m; \quad \log[f(z)] = -z^{m+1} / (m+1) + c;$$

da cui:

$$f(z) = k \exp\{-z^{m+1} / (m+1)\};$$

e facendo uso della trasformazione:  $\lambda x = z^{m+1} / (m+1)$ , e indicando con  $\alpha = 1/(m+1)$ , si ricava:

$$f(x) = k' \lambda^\alpha e^{-\lambda x} x^{\alpha-1}.$$

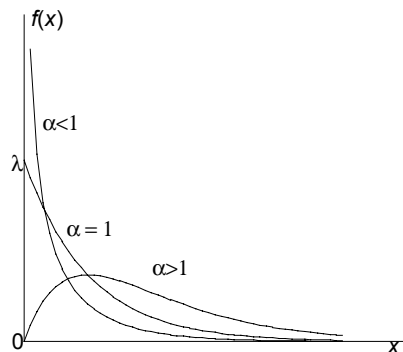
Il fattore di normalizzazione è l'integrale euleriano del II tipo:

$$\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx,$$

che, calcolando che  $\Gamma(1) = 1$ , gode della proprietà  $\Gamma(\alpha) = (\alpha-1) \Gamma(\alpha-1)$  e che per  $\alpha$  intero si riduce a  $\Gamma(n) = (n-1)!$

5.1 Pertanto la v.c. *Gamma* è una v.c. continua definita per  $x \geq 0$ , con f.d.:

$$f(x) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha e^{-\lambda x} x^{\alpha-1}.$$



Questa v.c. dipende da due parametri:  $\lambda$  *parametro di scala*, che può essere eliminato con la trasformazione  $y = \lambda x$ :

$$f(y) = f(x) dx / dy = \frac{1}{\Gamma(\alpha)} \lambda e^{-\lambda x} (\lambda x)^{\alpha-1} / \lambda = \frac{1}{\Gamma(\alpha)} e^{-y} (y)^{\alpha-1}$$

e il *parametro di forma*  $\alpha > 0$ . Infatti al variare di questo parametro si osservano diverse forme della f.d.: in particolare, per  $0 < \alpha < 1$ , si ha un asintoto verticale per  $x=0$  e poi la funzione è sempre asintoticamente decrescente; per  $\alpha = 1$ , si ha la funzione esponenziale che dal valore  $\lambda$  in corrispondenza di  $x=0$ , decresce asintoticamente; mentre per  $\alpha > 1$  si ha  $f(0) = 0$  e poi la funzione cresce fino ad un massimo per poi decrescere ancora una volta esponenzialmente.

I percentili della v.c.  $Y \propto \text{Gamma}(\lambda = 1/2, \alpha = n/2)$  sono tabulati nel relativo **Prontuario**.

Per ottenere i percentili della v.c.  $X$  con qualunque valore di  $\lambda$ , si ricorra alla trasformazione lineare  $y = 2\lambda x$ .

5.2 La f.g.m. e la f.g.c. sono rispettivamente:

$$\mathcal{M}_X(t) = \int_0^{\infty} \frac{1}{\Gamma(\alpha)} \lambda^{\alpha} e^{-\lambda x} x^{\alpha-1} e^{tx} dx = \int_0^{\infty} \frac{1}{\Gamma(\alpha)} \lambda^{\alpha} e^{-(\lambda-t)x} x^{\alpha-1} dx = \left( \frac{\lambda}{\lambda-t} \right)^{\alpha}$$

$$\mathcal{K}_X(t) = \alpha [\log(\lambda) - \log(\lambda-t)]$$

Da qui ricaviamo facilmente i primi due momenti:

$$\mathcal{E}(X) = \alpha / \lambda \quad \text{e} \quad \text{var}(X) = \alpha / \lambda^2.$$

Inoltre la f.g.m. ci mette in condizione di dimostrare la più importante proprietà della v.c. *Gamma*: la v.c. costituita dalla somma di  $k$  v.c. indipendenti  $x_i$  distribuite secondo una *Gamma* con stesso  $\lambda$  e parametri di forma  $\alpha_1, \dots, \alpha_k$ , si distribuisce secondo una *Gamma* di parametri  $\lambda$  e  $\alpha = \alpha_1 + \dots + \alpha_k$ . Formalmente cioè, data:

$$Y = X_1 + \dots + X_k \text{ con } X_i \propto \text{Gamma}(\lambda, \alpha_i) \text{ indep., si ha: } Y \propto \text{Gamma}(\lambda, \sum_i \alpha_i).$$

Per la dimostrazione basta osservare che:

$$\mathcal{M}_Y(t) = \prod_{i=1}^n \mathcal{M}_{X_i}(t) = \left( \frac{\lambda}{\lambda-t} \right)^{\alpha} \text{ ove } \alpha = \sum_{i=1}^n \alpha_i.$$

In particolare, visto che l'esponenziale è una particolare *Gamma* con  $\alpha = 1$ , sommando  $n$  v.c. *esponenziali* i.i.d. di parametro  $\lambda$  si ha una *Gamma*( $\lambda, n$ ).

## 6. Processi in serie, in parallelo e in stand-by

Si supponga di avere un sistema composto da  $n$  elementi *indipendenti* aventi durata distribuita per tutti secondo la v.c. *esponenziale* di parametro  $\lambda$ .

Immaginiamo tre modalità di funzionamento del sistema.

*Sistema in serie.* In questo caso il sistema si guasta appena va fuori servizio il primo elemento; siamo quindi interessati alla distribuzione del *minimo* (*min*):

$$F_{\min}(y) = 1 - [1 - F_X(y)]^n \quad \text{e} \quad f_{\min}(y) = n f_X(y) [1 - F_X(y)]^{n-1}$$

(si ricordi che nei sistemi in serie avevamo calcolato l'affidabilità come prodotto delle affidabilità). Per v.c. *esponenziali* possiamo ricavare:

$$F_{\min}(y) = 1 - e^{-\lambda y n} \quad \text{e} \quad f_{\min}(y) = n \lambda e^{-\lambda y} e^{-(n-1)\lambda y} = n \lambda e^{-n\lambda y}$$

che risulta ancora una *esponenziale* di parametro (t.g.)  $n\lambda$ .

*Sistema in parallelo.* Il sistema si guasta appena va fuori servizio l'ultimo elemento; siamo quindi interessati alla distribuzione del *massimo* (*Max*):

$$F_{\max}(y) = [F_X(y)]^n \quad \text{e} \quad f_{\max}(y) = n f_X(y) [F_X(y)]^{n-1}$$

(si ricordi che nei sistemi in parallelo avevamo calcolato la probabilità di guasto come prodotto delle probabilità di guasto); per v.c. *esponenziali* si ha:

$$F_{\max}(y) = [1 - e^{-\lambda y}]^n \quad \text{e} \quad f_{\max}(y) = n \lambda e^{-\lambda y} [1 - e^{-\lambda y}]^{n-1}$$

Si può dimostrare che il v.a. e la f.g.m. di questa v.c. risultano rispettivamente

$$\mathcal{E}(\max) = \frac{1}{\lambda} \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} \frac{1}{k} \quad \mathcal{M}_{\max}(t) = 1 + \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} \frac{t}{\lambda k - t}$$

*Sistema in stand-by.* Si supponga invece che gli elementi funzionino uno alla volta: appena si guasta il primo, entra in funzione il secondo e così via. In questo caso la vita del sistema è data dalla *somma* delle vite dei singoli elementi, che per v.c. *esponenziali* – come abbiamo visto – si distribuisce come una v.c. *Gamma* di parametri  $\lambda$  e  $n$ .

6.1 Come è intuibile, i tre tipi di sistemi – a parità di  $n$  e  $\lambda$  – sono in ordine crescente di affidabilità, come si può osservare anche dalla tabella sotto riportata per  $\lambda = 1$  dei v.a., delle varianze e dei coefficienti di variazione (*CV*).

Numero di elementi	Sistema in serie			Sistema in parallelo			Sistema in stand-by		
	$\mathcal{E}(X)$	$V(X)$	$CV$	$\mathcal{E}(X)$	$V(X)$	$CV$	$\mathcal{E}(X)$	$V(X)$	$CV$
1	1.00	1.00	1.00	1.00	1.00	1.00	1	1	1.00
2	0.50	0.25	1.00	1.50	1.25	0.75	2	2	0.71
3	0.33	0.11	1.00	1.83	1.36	0.64	3	3	0.58
4	0.25	0.06	1.00	2.08	1.42	0.57	4	4	0.50
5	0.20	0.04	1.00	2.28	1.46	0.53	5	5	0.45

## 7. Il processo di Poisson

Si supponga ora di avere nel sistema stand-by un problema diverso: si desidera determinare qual è la probabilità di un certo numero di arrivi (guasti) in un fissato intervallo di tempo  $(0, T)$ , in modo da poter dimensionare opportunamente un sistema costituito da più elementi indipendenti. Si supponga che l'elemento guastatosi non venga *sostituito* da uno nuovo (cosa che lo riporterebbe al suo istante 0), ma venga semplicemente *ripristinato* (cosa che lo riporta ad uno stato pari a quello che aveva prima di guastarsi).

7.1 La d.p. della v.c. in questo caso si può determinare per induzione. La probabilità che non si verifichi alcun arrivo è pari all'affidabilità nel punto  $T$ :  $\mathcal{P}_T(0) = \mathcal{R}(T)$ . Per calcolare la probabilità di un arrivo, pensiamo ad un punto generico  $t$  nell'intervallo  $(0, T)$  ove tale arrivo si verifichi

$$\begin{array}{ccc} A = X_1 \geq t & B = t \leq X_1 \leq t + dt & C = X_2 > T \mid X_2 > t \\ \hline 0 & T & T \end{array}$$

Si deve verificare contemporaneamente che: 1) fino al tempo  $t$  non si realizzi alcun arrivo (ossia, il primo arrivo non si verifichi prima di  $t$ :  $X_1 \geq t$ ), 2) al tempo  $t$  si abbia l'arrivo, ammesso che esso non si sia verificato prima ( $t \leq X_1 \leq t + dt \mid X_1 \geq t$ ), ed infine che: 3) non si verifichi alcun altro arrivo tra  $t$  e  $T$  (ossia, il secondo arrivo si verifichi dopo  $T$ :  $X_2 > T \mid X_2 > t$ ). Il primo evento è  $A$  ed ha probabilità  $\mathcal{R}(t)$ ; il secondo evento è  $B|A$  ed ha probabilità  $\lambda(t) dt$ ; infine il terzo evento è  $C$  ed ha probabilità  $\mathcal{R}(T) / \mathcal{R}(t)$ , che indica l'affidabilità condizionata di un elemento che ha già lavorato fino a  $t$  e viene ora *ripristinato*. La probabilità dell'intersezione dei tre eventi è data dal prodotto di queste tre probabilità, perché gli eventi risultano indipendenti per costruzione. In particolare, la seconda è la probabilità di un evento  $B$  condizionato al primo  $A$ :  $\mathcal{P}(A \cap B) = \mathcal{P}(A) \mathcal{P}(B|A)$ ; mentre la terza riguarda un altro elemento, la cui vita si suppone sia indipendente da quella del primo. Ora, poiché il punto  $t$  è un punto generico, per ottenere la probabilità desiderata si devono sommare le probabilità ottenute al variare di  $t$  (qui si può effettuare la somma poiché gli eventi nel tempo sono incompatibili: l'arrivo, se è unico, può verificarsi solo in uno dei punti  $t$  compresi tra 0 e  $T$ ). Ricordando che  $\lambda(t) dt = -d\log[\mathcal{R}(T)]$  e  $\mathcal{R}(0) = 1$ , si ha:

$$\mathcal{P}_T(1) = \int_0^T \mathcal{R}(t) \lambda(t) \mathcal{R}(T)/\mathcal{R}(t) dt = -\mathcal{R}(T) \int_0^T d\log[\mathcal{R}(t)] = -\mathcal{R}(T) \log[\mathcal{R}(T)].$$

Per calcolare la probabilità di due arrivi indipendenti, si ammetta che entro  $t$  si sia verificato un evento, mentre il secondo evento si verifichi al tempo  $t$ :

$$\mathcal{P}_T(2) = \int_0^T \mathcal{P}_t(2) \lambda(t) \mathcal{R}(T)/\mathcal{R}(t) dt = \mathcal{R}(T) \int_0^T \log[\mathcal{R}(t)] d\log[\mathcal{R}(t)] = \frac{1}{2} \mathcal{R}(T) \{\log[\mathcal{R}(T)]\}^2.$$

Iterando per  $k > 2$  e sostituendo  $-\log[\mathcal{R}(T)] = \int_0^T \lambda(t) dt = \Lambda(T)$ , si perviene a:

$$\mathcal{P}_T(k) = \int_0^T \mathcal{P}_{k-1}(t) \lambda(t) \mathcal{R}(T) / \mathcal{R}(t) dt = \mathcal{R}(T) \{-\log[\mathcal{R}(T)]\}^k / k!$$

$$\mathcal{P}_T(k) = e^{-\Lambda(T)} [\Lambda(T)]^k / k!$$

che risulta la d.p. di *Poisson*.

Questo processo prende il nome di *Processo di Poisson non omogeneo*.

7.2 Da qui, per arrivi governati v.c. *esponenziali*, aventi cioè t.g. costante nel tempo:  $\lambda(t) = \lambda$  e  $\Lambda(T) = \lambda T$ , si ha il *Processo di Poisson omogeneo*:

$$\mathcal{P}_T(k) = \frac{1}{k!} (\lambda T)^k e^{-\lambda T}.$$

In questo caso, a causa dell'omogeneità del processo che risulta senza memoria o usura – come sappiamo – i due sistemi sostituzione e ripristino sono equivalenti, infatti risulta:  $\mathcal{R}(T) / \mathcal{R}(t) = \mathcal{R}(T-t)$ .

Come si vede, ponendo  $\lambda T = \mu$ , questa v.c. coincide con una *Poisson*. Pertanto la *Poisson*, oltre a costituire una v.c. limite della *binomiale*, rappresenta anche la v.c. che regola il numero di arrivi nel caso di un processo che presenta le seguenti caratteristiche:

1. i numeri degli arrivi in intervalli disgiunti in  $(0, T)$  sono indipendenti (*indipendenza*);
2. intervalli di eguale lunghezza hanno stessa probabilità di contenere un egual numero di eventi (*omogeneità*).

7.3 Si può ricavare il processo di *Poisson* facendo uso della formula di Bayes. Ricordando che la v.c. *Gamma* rappresenta la d.p. della vita di  $n$  elementi i.i.d. a t.g. costante,  $f(t | n) dt$  (ossia che in un tempo  $t$  vi siano  $n-1$  guasti e l'ultimo continui a funzionare), scambiando il ruolo di parametro e di v.c. tra  $T$  e  $n$ , si può ricavare la probabilità di  $n-1$  guasti in un intervallo  $(0, T)$ :

$$\mathcal{P}(n-1 | T) = \frac{f(T | n) \mathcal{P}(n-1)}{\sum_{n=0}^{\infty} f(T | n) \mathcal{P}(n-1)}.$$

Ora il problema sta nel determinare la probabilità *a priori*  $\mathcal{P}(n-1)$ . In mancanza di informazioni, si può dare una probabilità uniforme a tutti i valori di  $n$  secondo quello che viene detto l'assioma dell'ignoranza. Pertanto il valore di  $\mathcal{P}(n-1)$ , ridotto ad una costante, si elide a numeratore e denominatore e quindi occorre solo trovare la costante di normalizzazione  $c$ .

Talvolta suscita perplessità l'uso di una distribuzione *a priori* “degenere” del tipo di quelle usate qui. Infatti notiamo che, dovendo distribuire la massa di probabilità su un numero infinito (anche se numerabile) di punti, quali quelli su cui è definita la *Poisson*, la probabilità che si assegna ad ogni punto è infinitesima. Questo tuttavia non è un problema né dal punto di vista formale, né tanto meno dal punto di vista computazionale, perché come abbiamo visto tali *a priori* si elidono.



Ricavando la costante di normalizzazione, che risulta pari a  $c = 1/\lambda$ , si ha:

$$\mathcal{P}(n-1 | T) = c \frac{1}{\Gamma(n)} \lambda^n e^{-\lambda T} T^{n-1} = \frac{1}{(n-1)!} e^{-\lambda T} (\lambda T)^{n-1};$$

da cui: 
$$\mathcal{P}(n | T) = \frac{1}{n!} e^{-\mu} \mu^n.$$

## 8. La variabile casuale Gumbel

Dal modello esponenziale deriva la v.c. di *Gumbel* attraverso una semplice trasformazione logaritmica. Data una v.c. *X* esponenziale di parametro  $\lambda$ , la

$$Z = a \pm b \log(\lambda X)$$

ammettendo un coefficiente  $b$  sempre positivo, assumerà diverse espressioni, a seconda che il segno nell'espressione sia positivo o negativo. Tale v.c. prende il nome di variabile “dei valori estremi”: “dei massimi” per segno negativo e “dei minimi” per segno positivo.

8.1 Nel caso di *Gumbel* “dei massimi” occorre tenere conto del cambiamento di segno nella trasformazione, scambiando c.d.f. e f.a., e quindi avremo:

$$\mathcal{H}(z) = \exp\{-e^{-(z-a)/b}\}$$

(si verifichi che per  $z \rightarrow +\infty$  la  $\mathcal{H}(z) \rightarrow 1$ , mentre per  $z \rightarrow -\infty$  la  $\mathcal{H}(z) \rightarrow 0$ ).

Questa v.c. può essere ricavata come distribuzione asintotica, per  $n \rightarrow \infty$ , del massimo di  $n$  v.c. i.i.d. aventi t.g. un generico  $\Lambda(t)$ , per cui:  $\mathcal{R}(t) = \exp\{-\Lambda(t)\}$

$$Z = \text{Max}\{X_1, X_2, \dots, X_n\}$$

Poiché siamo interessati al superamento di una “soglia” di  $z$  molto elevata,  $z^*$ , si può ammettere che il t.g. cresca con  $z^*$ , secondo una  $\mathcal{H}(z^*)$ , e che l'affidabilità diminuisca proporzionalmente a  $n$ , per cui scriviamo

$$\mathcal{R}_X(z^*) = [1 - \mathcal{F}_X(z^*)] = [\exp\{-\mathcal{H}(z^*)\}] = 1/n$$

Allora avremo

$$\mathcal{R}_X(z) / \mathcal{R}_X(z^*) = \exp\{\mathcal{H}(z^*) - \mathcal{H}(z)\} = [1 - \mathcal{F}_X(z)] / [1/n]$$

da cui

$$\mathcal{F}_X(z) = 1 - [1/n] \exp\{\mathcal{H}(z^*) - \mathcal{H}(z)\}$$

Data l'indipendenza, la c.d.f. di  $Z$  si può scrivere come  $\mathcal{F}_Z(z) = [\mathcal{F}_X(z)]^n$ , da cui

$$\mathcal{F}_Z(z) = [\mathcal{F}_X(z)]^n = \left[1 - \frac{1}{n} \exp\{\mathcal{H}(z^*) - \mathcal{H}(z)\}\right]^n$$

Passando al limite per  $n \rightarrow \infty$  si ha:

$$\mathcal{F}_Z(z) = \exp\{-e^{[\mathcal{H}(z^*) - \mathcal{H}(z)]}\}$$

Approssimando in serie  $\mathcal{H}(z)$  intorno al punto di interesse  $z^*$  ed arrestando lo sviluppo al I ordine, si ha:

$$\mathcal{H}(z) \cong \mathcal{H}(z^*) + \mathcal{H}'(z^*)(z^* - z)$$

da cui

$$F_Z(z) = \exp\{-e^{[\mathcal{H}'(z^*)(z - z^*)]}\}$$

che, ponendo  $\mathcal{H}'(z^*) = 1/b$  e  $z^* = a$ , restituisce immediatamente la c.d.f. della *Gumbel* “dei massimi”.

8.2 Questa v.c. può essere interpretata anche come la probabilità di non avere arrivi di *magnitudo* superiore ad una soglia fissata, entro un certo assegnato intervallo temporale.

Supponiamo che il verificarsi di un certo fenomeno segua il *Processo di Poisson omogeneo* e che la magnitudo  $Z$  di tale fenomeno sia una v.a. anch'essa con legge di tipo esponenziale  $\mathcal{F}(z) = 1 - e^{-\gamma z}$ , con  $\gamma$  positivo e  $z \in \mathfrak{R}^+$  (ossia, quando il fenomeno si verifica, si ha una probabilità pari a  $e^{-\gamma z}$  che esso non superi la magnitudo  $z$ ). In queste condizioni, ammettendo che il considerare solo fenomeni di magnitudo superiore a  $z$  fissato conduca ad un *Processo di Poisson omogeneo* in cui sia stato solo modificato il tasso di arrivo in modo proporzionale alla probabilità  $e^{-\gamma z}$ , si ottiene il nuovo tasso di arrivo  $\lambda e^{-\gamma z}$ :

$$\mathcal{P}_{T,z}(k) = (\lambda e^{-\gamma z} T)^k \exp\{-\lambda e^{-\gamma z} T\} / k!$$

Da qui si può ottenere la probabilità di non ottenere superamenti della soglia  $z$  entro il tempo  $T$ , calcolando la probabilità precedente in  $k = 0$ :

$$\mathcal{P}_{T,z}(0) = \exp\{-\lambda T e^{-\gamma z}\}$$

che si riconduce alla *Gumbel* ponendo  $\gamma = 1/b$  e  $\log(\lambda T) = a/b$ .

8.2. Nel caso in cui segno della trasformazione  $Z = a \pm b \log(\lambda X)$  sia positivo, la c.d.f. risulta

$$\mathcal{F}(z) = 1 - \exp\{-e^{(z-a)/b}\}$$

(si verifichi ancora che per  $z \rightarrow +\infty$  la  $\mathcal{F}(z) \rightarrow 1$ , mentre per  $z \rightarrow -\infty$  la  $\mathcal{F}(z) \rightarrow 0$ ). Come si osserva facilmente in questo caso la c.d.f. è quella di una *Weibull*,  $\mathcal{F}(x) = 1 - \exp\{-\lambda(x-\xi)^\alpha/\alpha\}$ , ponendo  $\log(\lambda/\alpha) = -a/b$ ,  $\alpha = 1/b$  e  $z = \log(x-\xi)$ .

## PROBLEMI PER IL CAPITOLO 5

1. Si supponga di avere a disposizione un elemento, come una lampadina, avente tasso di guasto costante e v.a. pari a 30 ore. Fino a che istante l'affidabilità si mantiene superiore a 0.9?

*Sugg.:* La vita dell'elemento è una v.c. *esponenziale* con tasso di guasto 1/30.

*Risp.:* fino a 3.16 ore.

- Si determini la probabilità che la durata più breve tra 10 lampadine del tipo precedente superi le 10 ore.

$$\text{Risp.: } \mathcal{R}(y) = e^{-n\lambda y} = e^{-10 \cdot 0.033 \cdot 10} = e^{-3.33} = 0.0357$$

2. Si supponga ora di avere un sistema costituito da 3 elementi del tipo precedente e tali che uno entra in funzione quando il precedente si è guastato. Fino a che istante l'affidabilità si mantiene superiore a 0.9?

*Sugg.:* La v.c. "vita del sistema" è la somma delle vite dei tre elementi.

*Risp.:* La vita del sistema è una v.c. *Gamma* con  $\lambda = 1/30$  ed  $\alpha = 3$ . Dai prontuari si legge il valore di  $y = 2.2$  e da qui il valore della vita pari a  $2.2 \cdot 30 / 2 = 33.06$  ore.

- Si supponga ora che si voglia un'affidabilità superiore a 0.9 per almeno 50 ore. Da quanti elementi dev'essere composto il sistema?

*Sugg.:* La v.c. numero di arrivi segue il processo di *Poisson* con v.a.  $\mu = \lambda T = 50/30$ .

Si determini di tale v.c. il punto della f.r. pari o superiore a 0.9: un numero di arrivi superiore si avrà con una probabilità inferiore a 0.1.

*Risp.:* Il numero massimo di guasti che può verificarsi con prob. 0.9 è 3, quindi occorre dotarsi di 4 elementi.

3. In un processo produttivo in una giornata il numero medio di pezzi di un certo tipo richiesti al magazzino è pari a 5. Se la distribuzione del numero di richieste segue il processo di *Poisson*, di quanti pezzi dev'essere rifornito il magazzino all'inizio della settimana per limitare all'1% la probabilità di restare senza pezzi?

4. Si determini il v.a. e la varianza della v.c. *Weibull*:  $f(x) = \lambda \exp\{-\lambda x^\alpha/\alpha\} x^{\alpha-1}$ .

$$\text{Risp.: } E(X) = \lambda \int_0^\infty x^\alpha \exp\{-\lambda x^\alpha/\alpha\} dx = \left(\frac{\alpha}{\lambda}\right)^{1/\alpha} \int_0^\infty y^{1/\alpha} e^{-y} dy = \left(\frac{\alpha}{\lambda}\right)^{1/\alpha} \Gamma\left(\frac{1}{\alpha} + 1\right),$$

$$E(X^2) = \lambda \int_0^\infty x^{\alpha+1} \exp\{-\lambda x^\alpha/\alpha\} dx = \left(\frac{\alpha}{\lambda}\right)^{2/\alpha} \int_0^\infty y^{2/\alpha} e^{-y} dy = \left(\frac{\alpha}{\lambda}\right)^{2/\alpha} \Gamma\left(\frac{2}{\alpha} + 1\right),$$

$$\text{var}(X) = \left(\frac{\alpha}{\lambda}\right)^{2/\alpha} \left[ \Gamma\left(\frac{2}{\alpha} + 1\right) - \Gamma^2\left(\frac{1}{\alpha} + 1\right) \right].$$

- La vita di un processo è una v.c. *Weibull* con  $\xi = 0$ ,  $\lambda = 0.1$  ed  $\alpha = 2$ . Si determini il punto che ha affidabilità pari a 0.9.

*Sugg.:* Si determini il punto che ha affidabilità 0.9 sull'esponenziale e quindi con la nota trasformazione ci si riconduca al punto sulla *Weibull*.

*Risp.:* 1.451623

- *Determinazione della scorta ottima*

Supponiamo che ora si desideri determinare la scorta ottima non in base ad un rischio fissato, ma in base a funzioni di costo assegnate da minimizzare. Si abbia quindi un costo di scorta che supponiamo incrementi in modo lineare con la dimensione  $n$  di essa secondo un coefficiente  $c$ , il *costo deterministico*. D'altro lato esiste un *costo aleatorio*, dato da una penale  $C$ , che pagheremo solo se la scorta non dovesse essere sufficiente, ossia se dovessimo avere un numero di richieste superiori a  $n$ . È opportuno valutare questo costo aleatorio moltiplicandolo per la probabilità di pagarlo  $P$ : ciò si può giustificare *a lungo andare* con la considerazione di “ammortare” le volte che si andrà incontro al pagamento della penale accantonando una quantità fissa per ogni prova. In questo caso la  $P$  è pari a  $\mathcal{P}(X > n) = 1 - \mathcal{F}(n)$  e quindi il costo aleatorio sarà valutato come  $C[1 - \mathcal{F}(n)]$ . Il costo totale ovviamente sarà dato dalla somma dei due costi: si tratta di trovare il valore di  $n$  che minimizza questa somma.

Un esempio numerico: sia  $c = 10$ ,  $C = 1000$  e d.p. una *Poisson* con v.a.  $\mu = 1$ .

Dimensione scorta	Costo deterministico	$\mathcal{P}(n)$	Costo aleatorio	Costo totale
0	0	0.3679	632.121	632.121
1	10	0.3679	264.241	274.241
2	20	0.1839	80.301	100.301
3	30	0.0613	18.988	48.988
4	40	0.0153	3.660	43.660
5	50	0.0031	0.594	50.594

Come si può osservare, il minimo della funzione di costo totale si ottiene in corrispondenza di  $n = 4$  e risulta pari a 43.66.

Analiticamente il problema può essere risolto in modo esplicito. Si tratta di arrestarsi al valore di  $n$  per cui un incremento ad  $n + 1$  provoca un incremento del costo totale, ossia quando  $C_T(n + 1) > C_T(n)$ , ossia

$$\begin{aligned}\Delta C_T(n) &= \{c(n + 1) + C[1 - \mathcal{F}(n + 1)]\} - \{cn + C[1 - \mathcal{F}(n)]\} = \\ &= c - C[\mathcal{F}(n + 1) - \mathcal{F}(n)] = c - C[\mathcal{P}(n + 1)] > 0\end{aligned}$$

cioè quando  $\mathcal{P}(n + 1) < c / C$ , la probabilità di ottenere una richiesta pari a  $n$  è inferiore al rapporto tra il costo di scorta unitario  $c$  ed il costo di penale  $C$ . In questo caso l'incremento di costo apportato da un elemento di scorta in più non giustifica la riduzione nel rischio di pagare la penale che questo apporta. È evidente quindi che, immaginando una d.p. che diminuisce asintoticamente, aumentando il rapporto  $c / C$  il valore ottimale di  $n$  tende ad arretrare, mentre all'aumentare di tale rapporto tale valore aumenta. Questa regola va comunque verificata numericamente (in quanto si ottiene solo un minimo relativo in caso di d.p. unimodale), poiché vi sono casi in cui il minimo assoluto si ha per  $n = 0$ , ossia non conviene fare scorte (se  $c$  è troppo grande o  $C$  troppo piccolo). Come si vede dalla colonna  $\mathcal{P}(n)$ , che riporta i valori della d.p. della *Poisson* con  $\mu = 1$ , il punto per cui  $\mathcal{P}(n) < c / C = 0.01$  è per  $n \geq 5$ ; quindi ci si deve arrestare a  $n = 4$ .

## CAPITOLO 6

### VARIABILI CASUALI CONTINUE: II

In questo capitolo continuiamo lo studio delle v.c. continue ed in particolare introdurremo la v.c. normale nonché il Teorema limite centrale.

#### 1. La variabile casuale normale o gaussiana

Ricaviamo la v.c. avente f.d. che soddisfa l'equazione:

$$d \log[f(x)] / dx = - (x - \mu) / \sigma^2,$$

imponendo così che la costante a denominatore sia positiva. Da ciò si ha:

$$\log[f(x)] = - \frac{1}{2} (x - \mu)^2 / \sigma^2 + \text{cost} ;$$

da cui, operando la trasformazione lineare  $z = (x - \mu) / \sigma$ , sia ha:

$$f(x) = k \exp\left\{-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right\} = k' \exp\{-\frac{1}{2} z^2\}$$

Come si vede, questa è una particolare *Gamma* con parametro di forma pari a  $\frac{1}{2}$ , ma poiché la v.c. è elevata a potenza positiva, è ammessa anche la metà negativa dei numeri reali. Pertanto la costante di integrazione si può calcolare tenendo conto che l'area è raddoppiata rispetto alla *Gamma*.

Alternativamente si può considerare l'integrale definito doppio:

$$A^2 = \int_{-\infty}^{\infty} \exp\{-\frac{1}{2} z^2\} dz \int_{-\infty}^{\infty} \exp\{-\frac{1}{2} y^2\} dy = \int_0^{2\pi} d\varphi \int_0^{\infty} \rho \exp\{-\frac{1}{2} \rho^2\} d\rho = 2\pi$$

avendo fatto uso della trasformazione polare:  $z = \rho \cos(\varphi)$  e  $y = \rho \sin(\varphi)$ , avente *Jacobiano* pari a  $\rho$ .

Le costanti di normalizzazione sono  $k' = 1 / \sqrt{2\pi}$  e  $k = 1 / b \sqrt{2\pi}$ .

1.1 La forma generale della v.c. *normale* o *gaussiana* ha quindi f.d.:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

Questa è una funzione che tende asintoticamente a zero per  $x \rightarrow \pm\infty$ , simmetrica rispetto all'asse  $x = \mu$ ; essa presenta il tipico aspetto campanulare, con un massimo in corrispondenza di  $x = \mu$  e due flessi in corrispondenza dei punti  $x_1 = \mu - \sigma$  e  $x_2 = \mu + \sigma$ : all'esterno di tali flessi la concavità è rivolta verso l'alto, all'interno verso il basso.

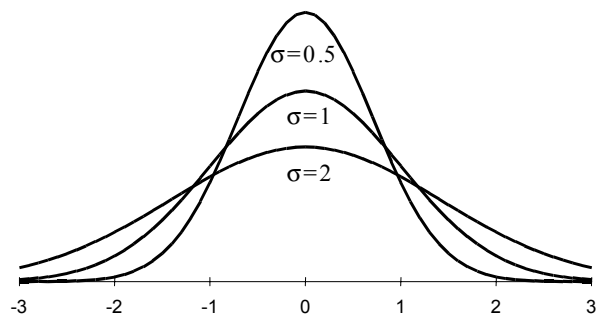
Con opportuna trasformazione lineare si possono eliminare i due parametri: uno di *posizione*  $\mu$  ed uno di *scala*  $\sigma$ :

$$Z = \frac{X - \mu}{\sigma}; \quad \text{da cui: } X = \sigma Z + \mu$$

che portano alla f.d. che si ottiene anche ponendo  $\mu = 0$  e  $\sigma = 1$ :

$$f(z) = f(x) dx / dz = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2} z^2\}.$$

Funzione di densità della normale con  $\mu = 0$



Come si vede dalla figura, all'aumentare del parametro di scala, aumenta la dispersione o variabilità della curva. È facilmente intuibile invece cosa succede al variare di  $\mu$ : varia il centro della curva da 0 al nuovo  $\mu$ .

1.2 Ricaviamo della  $Z$  la f.g.m.:

$$\begin{aligned} \mathcal{M}_Z(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-\frac{1}{2} z^2 + zt\} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-\frac{1}{2} z^2 + zt - \frac{1}{2} t^2\} dz \exp\{\frac{1}{2} t^2\} \\ &= \exp\{\frac{1}{2} t^2\} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-\frac{1}{2} (z-t)^2\} dz = \exp\{\frac{1}{2} t^2\}; \end{aligned}$$

la f.c.:

$$C_Z(t) = \mathcal{M}_Z(it) = \exp\{-\frac{1}{2} t^2\}.$$

(si noti che tale funzione, essendo simmetrica intorno allo zero, ha la parte immaginaria identicamente nulla) ed infine la f.g.c.:

$$\mathcal{K}_Z(t) = \log[\mathcal{M}_Z(t)] = \frac{1}{2} t^2.$$

Grazie alla proprietà:

$$\mathcal{M}_{a+bZ}(t) = \mathbb{E}\{\exp[(a + bZ)t]\} = \mathbb{E}\{\exp[(bZ)t]\} e^{at} = e^{at} \mathcal{M}_Z(bt),$$

si ricava la forma generale della  $X$  della f.g.m. e della f.g.c.:

$$\mathcal{M}_X(t) = \mathcal{M}_{a+bZ}(t) = e^{at} \exp\{\frac{1}{2}(bt)^2\}$$

$$\mathcal{K}_X(t) = \log[\mathcal{M}_X(t)] = at + \frac{1}{2}(bt)^2.$$

1.3 Da ciò si ricava:

$$\mathbb{E}(X) = \mu \quad \text{var}(X) = \sigma^2 \quad \mu_3 = 0 \quad \mu_4 - 3\mu_2^2 = 0.$$

Da queste relazioni vediamo alcune proprietà salienti della *normale*:

- il v.a. ed il parametro di posizione coincidono: questo è il motivo perché spesso tale momento primo è indicato semplicemente con la lettera  $\mu$ ;
- la deviazione standard e il parametro di scala coincidono: questo è il motivo perché spesso tale momento è indicato con la lettera  $\sigma$ ;
- quindi la trasformazione  $z = (x - \mu) / \sigma$  risulta essere la standardizzazione, pertanto la v.c.  $Z$  che si ottiene ponendo  $\mu = 0$  e  $\sigma = 1$  è la v.c. *normale standardizzata*  $\mathcal{N}(0, 1)$ ;
- la distribuzione è simmetrica e quindi  $\mu_3 = 0$ ;
- l'indice di curtosi  $\beta_2 = \mu_4 / \mu_2^2 = 3$ : data la straordinaria importanza che la *normale* assume nel calcolo delle probabilità e nella statistica, tale indice viene assunto, insieme all'indice di asimmetria, come indice di divergenza dalla normalità.

Abbiamo scelto questa strada, forse tortuosa, per arrivare alla espressione più comunemente usata della *gaussiana*; speriamo però che ciò costringa a riflettere sul fatto che le due quantità  $\mu$  e  $\sigma$  sono i *parametri* della distribuzione, che solo accidentalmente coincidono coi *momenti*. In tutte le altre distribuzioni che abbiamo incontrato e che incontreremo, ciò non si verifica mai (con l'eccezione della *Poisson*).

La standardizzazione qui è utile in quanto riconduce qualunque v.c. normale a quella con v.a. nullo e varianza unitaria. Ciò consente anche di poter usufruire di un solo prontuario, che consente di determinare la f.r., poiché il calcolo dell'integrale non è fattibile se non grazie ad appropriate approssimazioni numeriche.

Riportiamo nei Prontuari la f.r. della  $\mathcal{N}(0, 1)$ ,  $\Phi(z)$ , al variare della seconda cifra decimale di  $z$ .

## 2. Somme e combinazioni lineari di normali

Si abbia una v.c. combinazione lineare di v.c. *normali* indipendenti:

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$$

con  $X_i \propto iid \mathcal{N}(\mu_i, \sigma_i)$ . Questa v.c. è distribuita secondo una *normale* con:

$$\mathbb{E}(Y) = a_1 \mu_1 + a_2 \mu_2 + \dots + a_n \mu_n$$

e

$$\text{var}(Y) = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \dots + a_n^2 \sigma_n^2$$

$$Y \propto \mathcal{N}(\sum_i a_i \mu_i, \sum_i a_i^2 \sigma_i^2).$$

Per dimostrare questo teorema basta fare riferimento alle f.g.m. delle  $X_i$ :

$$\mathcal{M}_Y(t) = \prod_i \mathcal{M}_{X_i}(a_i t) = \prod_i \exp\{\mu_i a_i t\} \exp\{\sigma_i^2 (a_i t)^2 / 2\} = \exp\{\mu t\} \exp\{\sigma^2 t^2 / 2\}.$$

Da cui si vede che la f.g.m. alla quale si perviene è quella di una  $\mathcal{N}(\mu, \sigma^2)$ , avendo indicato con  $\mu = \sum_i a_i \mu_i$  e con  $\sigma^2 = \sum_i a_i^2 \sigma_i^2$ .

## 3. Il teorema limite centrale

Esponiamo ora il più importante teorema del calcolo delle probabilità, teorema che sarà spesso invocato durante lo studio della statistica: il *teorema limite centrale* (t.l.c.). Tale teorema fa riferimento ad una successione di v.c. usando lo stesso concetto di *convergenza in legge* che già abbiamo usato per introdurre la v.c. *Poisson*.

In particolare si consideri una successione di v.c. *indipendenti ed egualmente distribuite* (i.i.d.):

$$\{\mathbf{X}\} = \{X_1, X_2, \dots, X_n, \dots\},$$

aventi  $\mathbb{E}[X_i] = \mathbb{E}[X]$  e  $\text{var}[X_i] = \text{var}[X]$ , entrambe finite.

Si noti bene che non si fa alcun riferimento ad una *specificata forma distribuzionale*, purché essa sia comune per tutte. Come vedremo si pongono delle condizioni generalmente soddisfatte.

Si consideri dalla successione  $\{\mathbf{X}\}$ , la successione delle *somme parziali*:

$$\{\mathbf{S}\} = \{S_1, S_2, \dots, S_n, \dots\},$$

ove

$$S_1 = X_1, \quad S_2 = X_1 + X_2, \quad \dots, \quad S_n = X_1 + \dots + X_n, \quad \dots$$

Si osservi che il v.a. e la varianza di ciascuna somma sono pari a  $\mathbb{E}[S_n] = n \mathbb{E}[X]$  e, grazie all'indipendenza delle  $X_i$  tra loro,  $\text{var}[S_n] = n \text{var}[X]$ .



Da questa si può costruire la successione delle *somme parziali standardizzate* che risulta:

$$\{\mathbf{S}^*\} = \{S_1^*, S_2^*, \dots, S_n^*, \dots\},$$

ove

$$S_n^* = \frac{(X_1 + \dots + X_n) - n \mathbb{E}[X]}{\sqrt{n \operatorname{var}[X]}}.$$

Il t.l.c. afferma: *la successione  $\{\mathbf{S}^*\}$  ha come distribuzione limite la normale standardizzata* (ossia converge in legge alla  $\mathcal{N}(0, 1)$ ).

3.1 È questo il caso che più frequentemente riscontreremo in statistica e perciò a questo caso ci limiteremo, anche se esistono versioni del teorema che pongono delle condizioni meno restrittive, come per es. una forma funzionale non eguale tra le v.c. Per dimostrare il teorema, possiamo fare uso di una proprietà delle f.c. che abbiamo già citato e che riguarda i limiti di successioni: per determinare il limite della  $\{\mathbf{S}^*\}$ , dobbiamo determinare il limite della successione costituita dalle f.r.  $\{\mathbf{F}(\mathbf{S}^*)\}$  corrispondenti; poiché questo limite non è fattibile, possiamo determinare il limite delle corrispondenti f.c.  $\{C_{S_n^*}(t)\}$ : la v.c. che corrisponde al limite di questa successione è la v.c. a cui converge in legge la  $\{\mathbf{S}^*\}$ . Determiniamo quindi l'espressione della f.c. della generica:

$$S_n^* = \frac{(X_1 + \dots + X_n) - n \mathbb{E}[X]}{\sqrt{n \operatorname{var}[X]}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mathbb{E}[X]}{\sqrt{\operatorname{var}[X]}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i$$

ove con  $Z_i$  indichiamo la v.c. standardizzata:  $Z_i = [X_i - \mathbb{E}(X)] / \sqrt{\operatorname{var}(Z)}$ , avente  $\mathbb{E}(Z_i) = 0$  e  $\mathbb{E}(Z_i^2) = 1$ . La f.c. di  $S_n^*$  si può esprimere come:

$$C_{S_n^*}(t) = \prod_{i=1}^n C_{Z_i}(t/\sqrt{n}) = \{\mathbb{E}[\exp\{itZ_i/\sqrt{n}\}]\}^n;$$

infatti, grazie all'indipendenza delle v.c.  $X_i$ , la f.c. della somma delle  $Z_i$  è la produttoria delle f.c. di ogni singola  $Z_i$ , che per il fatto di essere identicamente distribuite hanno stessa f.c. Inoltre, come abbiamo già visto, il fattore  $1/\sqrt{n}$  che moltiplica  $Z_i$  può essere considerato come fattore che moltiplica la variabile  $t$ . Tale espressione, sviluppata in serie di Taylor, sotto la condizione che il momento centrato di ordine  $k$  sia tale che:  $\mu_k / (n^{k/2} k!) = o(n)$ , cioè sia un infinitesimo di ordine superiore a  $n$  per ogni  $k > 2$ , è:

$$\begin{aligned} C_{S_n^*}(t) &= \{\mathbb{E}[1 + itZ_i/\sqrt{n} + \frac{1}{2}(itZ_i/\sqrt{n})^2 + o(n)]\}^n \\ &= [1 + it\mathbb{E}(Z_i)/\sqrt{n} - \frac{1}{2}t^2\mathbb{E}(Z_i^2)/n + o(n)]^n = [1 - \frac{1}{2}t^2/n + o(n)]^n. \end{aligned}$$

Possiamo ora effettuare il limite per  $n \rightarrow \infty$  della  $C_{S_n^*}(t)$ , che risulta:

$$\lim_{n \rightarrow \infty} C_{S_n^*}(t) = \lim_{n \rightarrow \infty} [1 - \frac{1}{2}t^2/n + o(n)]^n = \exp\{-\frac{1}{2}t^2\}.$$

Come si vede, abbiamo dimostrato che la successione  $\{C_{S_n^*}(t)\}$  ha come limite la f.c.  $\exp\{-\frac{1}{2}t^2\}$  che corrisponde a quella della normale standardizzata. Pertanto si può dire in base a quanto esposto sulle successioni di v.c., di f.r. e di f.c. che la  $\{\mathbf{S}^*\}$  converge in legge alla *normale standardizzata*.

3.2 Cerchiamo di capire quali sono le implicazioni di questo teorema.

Si osservi come si faccia riferimento ad un insieme di tante (teoricamente infinite) v.c. che si sommano l'una all'altra, ma il cui effetto globale non ha una varianza infinita a causa della standardizzazione, quindi nel complesso ogni v.c. avrà un'influenza infinitesima. Pertanto possiamo pensare tali v.c. come tante piccolissime cause che si sovrappongono additivamente ed indipendentemente l'una all'altra, dando luogo ad un effetto che invece ha una sua variabilità non trascurabile.

Questo è proprio il modo come il calcolo delle probabilità formalizza *l'errore accidentale* o *caso*: un fenomeno che, rispetto ad una intensità media attesa, presenta ad ogni replica una deviazione verso l'alto o verso il basso, senza prediligere questo o quel segno; con deviazioni di piccola entità più frequenti e con deviazioni di maggiore entità, ma meno frequenti.

Sono innumerevoli gli esempi in tutti i campi sperimentali di fenomeni che si comportano secondo questo modello. Per esempio: la misura condotta con un elevato grado di precisione di una distanza tra due punti, il peso o l'altezza di una popolazione selezionata di animali o vegetali, la resistenza di una popolazione di provini tutti realizzati con lo stesso standard produttivo.

Ovunque un fenomeno si possa pensare come costituito da un "centro" deterministico su cui si sovrappone *additivamente* una serie di infinite cause di peso infinitesimo, allora la popolazione costituita dalle possibili rilevazioni di quel fenomeno si distribuirà secondo la legge degli errori accidentali, ossia *normalmente*.

#### 4. La Normale come limite della Binomiale

L'applicazione più immediata del t.l.c. riguarda la possibilità di usare sotto certe condizioni la v.c. *normale* come v.c. limite della *binomiale*. Questa v.c. è stata introdotta come somma di v.c. *bernoulliane* i.i.d. e quindi quando il parametro  $n$  di tale distribuzione diverge, siamo automaticamente nelle condizioni del t.l.c. Tuttavia nelle applicazioni pratiche in cui il valore di  $n$  può essere grande ma non certo infinito, è opportuno stabilire delle regole in cui l'approssimazione della normale può essere adottata.

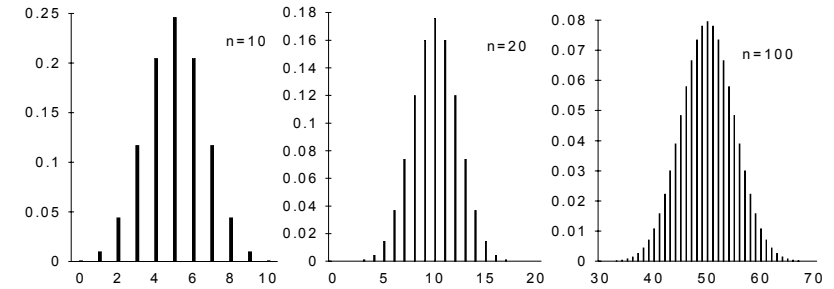
In base a tali considerazioni in genere si ammette che l'approssimazione è già buona per  $n > 30$  e  $\min(n\phi, n\psi) > 10$ . Come si vede questo caso, se si vuole, è complementare a quello della convergenza alla *Poisson*, che si verifica quando  $\phi$  è piccolo.

4.1 Tuttavia resta qui da chiarire un aspetto che può passare inosservato.

Come è possibile che una v.c. continua come la *normale* approssimi bene una v.c. discreta come la *binomiale*?

In realtà è proprio il passaggio al limite congiunto alla standardizzazione che fa compiere alla v.c. discreta un “salto” nel continuo. Infatti si badi bene che se non si effettuasse l’operazione di standardizzazione, si avrebbe media e varianza che divergerebbero, rendendo la v.c. non più maneggiabile, invece questa operazione costringe media e varianza a rimanere costanti.

Cosicché, quando al divergere di  $n$  aumentano il numero di punti ove la v.c. discreta è definita, tali punti anziché disperdersi sempre più, vanno “infittendosi” sempre più, fino a determinare al limite una v.c. continua.



4.2 Per migliorare questa approssimazione, si può effettuare quella che viene detta *correzione per la continuità*.

Si tratta di considerare ogni valore di probabilità che insiste su un punto ove la v.c. discreta è definita come una massa da “spalmare”, da diffondere su un opportuno intervallo.

In particolare se la v.c. discreta è definita come la *binomiale* sugli interi  $x = 0, 1, \dots, n$ , si potrà pensare la  $\text{Prob}(X=x) = \text{Prob}(x-0.5 \leq X \leq x+0.5)$ . Più in generale se si desidera la probabilità che  $X$  sia compresa tra due estremi, essa sarà pari a:

$$\text{Prob}(x_1 \leq X \leq x_2) = \text{Prob}(x_1 - 0.5 \leq X \leq x_2 + 0.5).$$

Ciò chiaramente dà la possibilità di distinguere il calcolo di questa probabilità da quella in cui gli estremi dell’intervallo non sono compresi:

$$\text{Prob}(x_1 < X < x_2) = \text{Prob}(x_1 + 0.5 \leq X \leq x_2 - 0.5).$$

4.3 Si noti che tutte le v.c. che godono della proprietà dell’additività, come la *Gamma* e la *Poisson*, opportunamente standardizzate e al divergere di un opportuno parametro, tendono alla v.c. *normale*, in quanto possono sempre essere pensate come somme di v.c. *iid*. Per riconoscere che una v.c.  $X_{(n)}$  gode della proprietà dell’additività, basta osservare se la f.c. si presenta come:

$$C_{X_{(n)}}(t) = [C_{X_{(1)}}(t)]^n$$

## 5. La variabile casuale lognormale

Abbiamo qui l'occasione di accennare ad un'altra distribuzione strettamente connessa alla normale: la distribuzione *lognormale*.

Formalmente essa è definita come la v.c.  $Y$  che attraverso la trasformazione logaritmica si riduce ad una v.c. *normale*:

$$X = \log(Y) \propto \mathcal{N}(\mu, \sigma)$$

Ovviamente  $Y$  è una v.c. continua definita sul semiasse positivo. Non effettuiamo lo studio completo di questa variabile, che così facilmente può trasformarsi nella *normale*. Tuttavia si noti come v.a. e deviazione standard siano legate proporzionalmente.

Infatti si ha:

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} \frac{e^x}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx = \exp\{\mu + \sigma^2/2\} \\ E(Y^2) &= \int_{-\infty}^{\infty} \frac{e^{2x}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx = \exp\{2\mu + 2\sigma^2\} \\ var(Y) &= \exp\{2\mu + 2\sigma^2\} - \exp\{2\mu + \sigma^2\} = E^2(Y) [\exp\{\sigma^2\} - 1]. \end{aligned}$$

5.1 In realtà tale variabile si presta a modellare bene un fenomeno simile a quello commentato per il t.l.c. in cui però le v.c. si cumulano *moltiplicativamente*:

$$Y = X_1 \cdot \dots \cdot X_2 \cdot \dots \cdot X_n \dots$$

ed in effetti in questo caso la trasformazione logaritmica restituisce le condizioni additive del t.l.c.:

$$\log(Y) = \log(X_1) + \log(X_2) + \dots + \log(X_n) \dots$$

Sono tanti i fenomeni casuali in cui la variabilità (deviazione standard) è proporzionale al livello medio, per esempio: la crescita mensile del peso di un animale, l'energia dissipata durante una lavorazione meccanica. In questi casi per *linearizzare* il fenomeno si ricorra alla trasformazione logaritmica.

## PROBLEMI PER IL CAPITOLO 6

1. La lunghezza dei pezzi prodotti da una macchina ha distribuzione *normale* con v.a. 100 mm e varianza 4 mm<sup>2</sup>. Supponendo che verranno scartati tutti quelli che hanno lunghezza inferiore a 96 o superiore a 104, calcolare la difettosità del processo produttivo.

*Sugg.:* Si calcoli il complemento a 1 di:  $\Phi\left(\frac{104-100}{2}\right) - \Phi\left(\frac{96-100}{2}\right)$ .

Per quest'ultimo termine si ricorra alla relazione  $\Phi(-z) = 1 - \Phi(z)$ .

*Risp.:* 0.0456.

- Supponendo che invece si desideri una difettosità pari all'1%, quali specifiche in termini di varianza possiamo soddisfare?

*Sugg.:* L'area al di là delle specifiche  $1 - \Phi(z_2) + \Phi(z_1)$  deve essere pari a 0.01. Se le due aree vengono poste simmetricamente intorno alla media, si ha  $\Phi(z_2) = 0.995$ , e  $z_1 = -z_2$ ; da cui si ricava  $x_2 = \sigma z_2 - \mu$  e  $x_1 = \sigma z_1 - \mu$ .

*Risp.:*  $x_2 = 105.152$  e  $x_1 = 94.848$ .

- Si supponga ora che con le prime specifiche ( $104 \div 96$ ), si desideri una difettosità dell'1%. Che varianza deve avere il processo produttivo?

*Sugg.:* Supponendo una ripartizione simmetrica della difettosità nelle due code della distribuzione, da uno dei due percentili  $\Phi(z_2) = 0.995$ , o  $z_1 = -z_2$ , si ricavi  $\sigma = (104-100)/z_2$  o equivalentemente  $\sigma = (100-96)/z_1$ .

*Risp.:*  $\sigma^2 = 2.415$ .

2. Ci si accerti che usando intervalli non simmetrici, l'area coperta dalla distribuzione *normale*, a parità di lunghezza dell'intervallo, è minore. Per esempio con v.a. 100 mm e varianza 4 mm<sup>2</sup>, si calcoli l'area compresa tra le specifiche  $97 \div 105$  o  $95 \div 103$ .

3. Si supponga che una procedura di controllo sia tale che, se la media aritmetica di 9 pezzi estratti a caso non è compresa tra 98 e 102 mm, il lotto viene scartato. Determinare la probabilità che il lotto venga scartato se la produzione è distribuita *normalmente* con v.a. pari a 100 mm e varianza 4 mm<sup>2</sup>.

*Sugg.:* Si determini la distribuzione della media aritmetica di 9 v.c. *normali* indipendenti.

*Risp.:* 0.26%.

- E se il v.a. è alterato di 1 mm in più o in meno, qual è la probabilità che il lotto venga scartato?

*Risp.:* 6.68%.

4. Un pezzo è costituito da 3 elementi indipendenti la lunghezza di ciascuno dei quali è una v.c. *normale* con v.a. e varianza rispettivamente: 100 mm, 2 mm<sup>2</sup>; 50 mm, 1 mm<sup>2</sup>; 50 mm, 1.5 mm<sup>2</sup>. Si determini la distribuzione della lunghezza del pezzo ed il suo v.a. e varianza.

- Si determini v.a. e varianza della v.c.  $Y = 3X_1 - 5X_2 + 3X_3$ , sapendo che  $X_1, X_2$  e  $X_3$  sono indipendenti ed hanno rispettivamente v.a. e varianza 100 mm, 2 mm<sup>2</sup>; 50 mm, 1 mm<sup>2</sup>; 50 mm, 1.5 mm<sup>2</sup>.

5. Una macchina confeziona dei pacchi di pasta. Si sa che il v.a. di tali pacchi è pari a 1000 gr e la varianza è pari a 25 gr<sup>2</sup>, non è nota però la forma distribuzionale, ma si sa che è unimodale e non molto asimmetrica. Se si esaminano 100 pacchi, qual è la probabilità che la loro media aritmetica si discosti dal v.a. di più di 1 gr?

*Sugg.:* Si usi l'approssimazione *normale* per la distribuzione della media aritmetica.

*Risp.:* 4.56% .

- Se ora si usano 2 gruppi di 50 pacchi, qual è la probabilità che entrambe le medie si discostino dal v.a. di oltre 1 gr?

*Risp.:*  $(1 - 0.8427)^2 = 2.47\%$  .

- Quanti pacchi devo controllare affinché la probabilità che la media di tali pacchi si discosti di oltre 1 gr dal v.a. sia solo dell'1%?

*Sugg.:*  $\Phi(z_1) = \Phi\left(\frac{999 - 1000}{5/\sqrt{n}}\right) = 0.005$ .

*Risp.:*  $n \geq 166$ .

6. Un magazzino ha in egual numero pezzi di tipo *A* e di tipo *B*. Se si estraggono 30 pezzi, qual è la probabilità di avere un numero di pezzi di tipo *A* inferiore a 10?

*Sugg.:* Si usi l'approssimazione *normale*:  $\Phi[(9.5-15)/\sqrt{7.5}]$  .

*Risp.:* 2.23% .

- Quanti pezzi occorre estrarre perché sia più del 99% la probabilità di avere almeno 10 pezzi di tipo *A*?

*Risp.:*  $n = 33$  .

7. Una v.c. *lognormale* *Y* ha v.a. 5 e varianza 3. Si determini il punto in cui l'affidabilità è pari al 90%.

*Sugg.:* I parametri della v.c. si ricavano dalle relazioni:

$$\sigma^2 = \log\left(\frac{\text{var}(Y)}{\mathbb{E}^2(Y)} + 1\right) \text{ e } \mu = \log[\mathbb{E}(Y)] - \sigma^2/2.$$

## CAPITOLO 7

### VARIABILI CASUALI CONTINUE: III

In questo capitolo concludiamo lo studio delle v.c. collegate con la v.c. *normale* ed inoltre studieremo una v.c. multipla: la *normale multivariata*.

#### 1. La variabile casuale chi-quadrato ( $\chi^2$ )

La  $\chi^2$  è la prima delle v.c. collegate con la *normale*. Essa è definita come una *somma di v.c. normali standardizzate indipendenti elevate al quadrato*:

$$\chi_v^2 = Z_1^2 + Z_2^2 + \dots + Z_v^2,$$

ove le v.c.  $Z_i \sim iid\mathcal{N}(0, 1)$  e  $v$ , l'unico parametro che la caratterizza, prende il nome di *gradi di libertà* (g.d.l.): motivo di ciò sarà più comprensibile quando useremo questa v.c. in Statistica. Dimostriamo che  $\chi_v^2$  si distribuisce come una particolare *Gamma* con parametro di scala  $1/2$  e parametro di forma  $v/2$ .

1.1 Anche per la dimostrazione di questa proprietà ricorriamo alle f.c. delle singole v.c. che compongono questa somma, o alle f.g.m. poiché esse esistono.

$$\begin{aligned} M_{Z_i^2}(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{z_i^2 t\} \exp\{-\frac{1}{2} z_i^2\} dz_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\{-\frac{1}{2} (\frac{1}{2} - t) z_i^2\} dz_i \\ &= \frac{1}{\sqrt{2\pi (\frac{1}{2} - t)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} y^2} dy = \frac{1}{\sqrt{1 - 2t}} = \left(\frac{1/2}{1/2 - t}\right)^{1/2}, \end{aligned}$$

che risulta la f.g.m. di una *Gamma*( $1/2, 1/2$ ), avendo posto  $y = z_i \sqrt{1/2 - t}$ , trasformazione valida in un intorno di  $t = 0$ , che dà  $dz = dy [(1/2 - t)]^{-1/2}$ .

Da qui abbiamo dimostrato che la v.c.  $Z_i^2 \propto \text{Gamma}(1/2, 1/2)$ . Pertanto:

$$\mathcal{M}_{\chi^2}(t) = \left( \frac{1/2}{1/2 - t} \right)^{v/2}$$

che risulta essere la f.g.m. della  $\text{Gamma}(1/2, v/2)$ .

1.2 Si ha quindi:

$$\mathbb{E}(\chi_v^2) = v \quad \text{e} \quad \text{var}(\chi_v^2) = 2v.$$

Essendo una particolare *Gamma*, la  $\chi^2$  gode della proprietà dell'additività, in particolare sommando  $\chi^2$  indipendenti si ottiene una  $\chi^2$  avente g.d.l. dati dalla somma dei g.d.l. degli addendi. Anche di questa distribuzione si hanno i prontuari, come per la *normale*, occorre però costruire prontuari differenti al variare di  $v$ . In realtà, come vedremo, siamo interessati solo a quelli che possiamo chiamare *prontuari inversi* (dato il valore dell'area cercata, ottenere il valore di  $\chi^2$ ), che sono costruiti in questo modo: si riportano i valori dei 100 $\alpha$ % percentili al variare per colonna di  $v$  e per riga di  $\alpha$ .

1.3 La v.c.  $\chi^2$  può essere ricavata come *forma quadratica* basata su  $n$  trasformazioni lineari di v.c. normali i.i.d.  $\mathbf{A} \mathbf{Z} = (\mathbf{X} - \boldsymbol{\mu}) = \mathbf{Y}$ . Supponiamo che la matrice  $\mathbf{A}$  sia di rango pieno pari a  $n$ . In questo caso la matrice entro la forma quadratica risulta l'inversa della matrice di varianza e covarianza  $\boldsymbol{\Sigma}$  di  $\mathbf{X}$

$$\chi_n^2 \propto \mathbf{Z}^T \mathbf{Z} = \mathbf{Y}^T (\mathbf{A}^{-1})^T \mathbf{A}^{-1} \mathbf{Y} = \mathbf{Y}^T (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{Y} = (\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})$$

poiché, osservando che la matrice di varianza e covarianza di  $\mathbf{Z}$  è  $\mathbf{I}$ , si ha:

$$\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] = \mathbb{E}[\mathbf{Y} \mathbf{Y}^T] = \mathbf{A} \mathbb{E}[\mathbf{Z} \mathbf{Z}^T] \mathbf{A}^T = \mathbf{A} \mathbf{I} \mathbf{A}^T = \mathbf{A} \mathbf{A}^T.$$

Entro la forma quadratica potrebbe anche trovarsi una matrice *idempotente*  $\mathbf{B}$ , che rispetta cioè la proprietà  $\mathbf{B} \mathbf{B}^T = \mathbf{B}$ . Infatti in questo caso la trasformazione  $\mathbf{B} \mathbf{Z} = (\mathbf{X} - \boldsymbol{\mu})$  dà luogo ad una matrice

$$\boldsymbol{\Sigma} = \mathbb{E}[\mathbf{B} \mathbf{Z} \mathbf{Z}^T \mathbf{B}^T] = \mathbf{B} \mathbf{B}^T = \mathbf{B}.$$

## 2. La variabile casuale *t-Student*

La v.c. *t-Student* è definita come un rapporto tra due v.c. *indipendenti*: al numeratore una *normale standardizzata* ed al denominatore la *radice quadrata di una chi-quadrato divisa per i propri g.d.l.*:

$$t = Z / \sqrt{\chi_v^2 / v}.$$

Per determinare la f.d. di questa v.c. si può ricorrere alla seguente procedura: si determini la f.d. congiunta della  $Z$  e della  $\chi_v^2$ , che a causa dell'indipendenza risulta il prodotto delle rispettive f.d., quindi si effettuino le seguenti trasformazioni:  $z = t \sqrt{\chi_v^2 / v}$  e  $\chi_v^2 = y$ . Si ha:



$$f(z, \chi_v^2) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2} z^2\} \frac{1}{2^{n/2} \Gamma(v/2)} \exp\{-\frac{1}{2} \chi_v^2\} (\chi_v^2)^{v/2-1}$$

$$f(t, y) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2} t^2 y/v\} \frac{1}{2^{n/2} \Gamma(v/2)} \exp\{-\frac{1}{2} y\} (y)^{v/2-1} \sqrt{y/v}$$

e usando ancora la trasformazione  $w = \frac{1}{2} y (1 + t^2 / 2)$ , si ottiene:

$$f(t, w) = \frac{1}{\Gamma(v/2) \sqrt{2\pi v}} e^{-w} w^{(v-1)/2} (1 + t^2/v)^{-(v+1)/2}$$

Integrando ora  $w$  tra 0 e  $\infty$  si ottiene la densità marginale rispetto alla  $t$ , che era ciò che si cercava:

$$f(t) = \int_0^\infty f(t, w) dw = \frac{\Gamma[(v+1)/2]}{\Gamma(v/2) \sqrt{\pi v}} (1 + t^2/v)^{-(v+1)/2}.$$

2.2 Come si può osservare questa distribuzione è simmetrica intorno all'asse  $t=0$ , è definita tra  $-\infty$  ed  $\infty$  e presenta due asintoti orizzontali. Anche di questa distribuzione si hanno i prontuari dei percentili al variare di  $v$ .

È possibile dimostrare che per  $v \rightarrow \infty$  la  $t \xrightarrow{L} \mathcal{N}(0, 1)$ .

### 3. La variabile casuale F-Fisher

La v.c. *F-Fisher* è definita come il rapporto tra due chi-quadrato indipendenti, ciascuna divisa per i propri g.d.l.:

$$F = \frac{\chi_{v_1}^2 / v_1}{\chi_{v_2}^2 / v_2}.$$

Anche qui per determinare la f.d. si può scrivere la f.d. congiunta, che per l'indipendenza risulta:

$$f(\chi_{v_1}^2, \chi_{v_2}^2) = f(\chi_{v_1}^2) f(\chi_{v_2}^2),$$

effettuare la trasformazione  $\chi_{v_2}^2 = y$  e  $\chi_{v_1}^2 = Fy$  e quindi integrare via la  $y$ .

3.1 Naturalmente questa è una v.c. definita nel semiasse positivo e dipendente da due parametri: i due g.d.l. del numeratore e del denominatore.

Questa volta i prontuari dei percentili pertanto sono organizzati nel seguente modo: vi è una tabella per ogni livello di  $\alpha$ , tipicamente al 95% ed al 99%, ove si riportano i percentili in corrispondenza di  $v_1$  per colonna e  $v_2$  per riga. L'aspetto della f.d. è simile a quello della *Gamma* avente parametro di forma maggiore di uno. Come è immediato osservare, la v.c. *t-Student* è la radice quadrata di una particolare *F-Fisher* avente g.d.l. del numeratore pari ad uno.

È possibile dimostrare che per  $v_2 \rightarrow \infty$  la  $F \xrightarrow{L} \chi_v^2 / v$ , con  $v = v_1$ .

3.2 La v.c.  $F$  si può ricondurre alla v.c.  $Beta$ , ponendo

$$z = F / (1 + F \nu_1 / \nu_2) \quad \text{e} \quad \alpha = \nu_1 / 2 \text{ e } \beta = \nu_2 / 2.$$

La  $Beta$  è una v.c. continua definita nell'intervallo  $(0, 1)$  avente f.d.

$$f(z) = B(\alpha, \beta) z^{\alpha-1} (1-z)^{\beta-1} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha) \Gamma(\beta)} z^{\alpha-1} (1-z)^{\beta-1}$$

con coefficiente di normalizzazione il numero Beta,  $B(\alpha, \beta)$ , una estensione del coefficiente binomiale ai numeri reali, infatti per  $\alpha$  e  $\beta$  interi risulta:

$$B(k, m) = \frac{(k+m-1)!}{(k-1)! (m-1)!} = (k+m-1) \binom{k+m-2}{k-1}.$$

È facile vedere che:  $\mathcal{E}\{z\} = \alpha / (\alpha + \beta)$ . È tabulata ma per parametri interi vale:

$$\int_0^1 f(z; m; n-m+1) dz = \sum_{i=m}^n \binom{n}{i} \varphi^i (1-\varphi)^{n-i}$$

#### 4. La variabile casuale esponenziale potenziata (normale di ordine $p$ )

Ancora dalla v.c.  $Gamma$  è possibile ottenere una v.c. che costituisce una generalizzazione della v.c. *normale*. Nella ormai vasta letteratura italiana tale v.c. è stata denominata *normale* di ordine  $p$ , mentre nella letteratura anglosassone viene denominata come *power exponential distribution* con una differente espressione dei parametri. Dalla f.d. della  $Gamma$ , fissiamo  $\lambda = 1/2$  e operiamo la trasformazione  $y = x^\alpha \in \mathfrak{R}^+$

$$f(x)dx = \frac{1}{2^\alpha \Gamma(\alpha)} e^{-x/2} x^{\alpha-1} dx = \frac{1}{2^\alpha \alpha \Gamma(\alpha)} e^{-x/2} dx^\alpha = \frac{1}{2^\alpha \Gamma(\alpha+1)} \exp\{-1/2 y^{1/\alpha}\} dy = f(y)dy$$

Se ora si ammette che  $y$  possa assumere anche valori negativi, prendendo il valore assoluto di  $y$ , si deve rinormalizzare la f.d. dividendola per 2. Da cui:

$$f(y) = \frac{1}{2^{\alpha+1} \Gamma(\alpha+1)} \exp\{-1/2 |y|^{1/\alpha}\}.$$

Riparametrizzando con  $\alpha = 1/p$  o con  $\alpha = (1 + \beta)/2$ , si ottiene:

$$f(y) = \frac{1}{2^{(3+\beta)/2} \Gamma\left(\frac{3+\beta}{2}\right)} \exp\{-1/2 |y|^{2/(1+\beta)}\}$$

$$f(y) = \frac{1}{2^{1+1/p} \Gamma(1+1/p)} \exp\{-1/2 |y|^p\} = \frac{1}{2 p^{1/p} \Gamma(1+1/p)} \exp\{-\frac{1}{p} |y|^p\}.$$

4.1 Queste due rappresentazioni della stessa funzione sono state studiate in particolare per  $-1 < \beta \leq 1$  che equivale a  $p \geq 1$ , che danno f.d. simmetriche intorno a  $y = 0$  ed in particolare per  $\beta = 0$  ( $p = 2$ ) danno la v.c. *gaussiana* già studiata; per  $\beta = 1$  ( $p = 1$ ) la *doppia esponenziale* (o distribuzione di *Laplace*) ottenuta ribaltando l'esponenziale intorno all'asse  $f(x) = 0$ ; e per  $\beta \rightarrow -1$  ( $p \rightarrow \infty$ ) si ottiene la v.c. *uniforme* definita nell'intervallo  $[-1, 1]$  già studiata.

Vogliamo sottolineare le relazioni che la legano questa v.c. alla *normale*.  
Dall'esame della varianza e del momento quarto standardizzato che risultano:

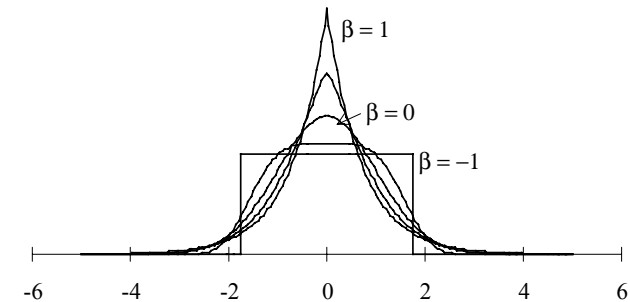
$$var(X) = \frac{2^{2\alpha} \Gamma\left(3\frac{1+\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right)} \quad \text{e} \quad \beta_2 = \frac{\mu_4}{\sigma^4} = \frac{\Gamma(5\alpha) \Gamma(\alpha)}{[\Gamma(3\alpha)]^2},$$

e ricordando che tale momento per la *gaussiana* è pari a 3, si vede che le v.c. con  $p < 2$  ( $\beta > 0$ ) hanno  $\beta_2 > 3$  e si dicono *leptocurtiche* (hanno curtosi maggiore rispetto alla *gaussiana*), le v.c. con  $p > 2$  ( $\beta < 0$ ) hanno  $\beta_2 < 3$  e si dicono *platicurtiche* (hanno curtosi minore rispetto alla *gaussiana*).

Da questo *punto* di vista ci sembra che la parametrizzazione in  $\beta$  sia più semplice mnemonicamente, essendo il segno di  $\beta$  pari al segno di  $\beta_2 - 3$ .

$\beta$	1	0.75	0.5	0.25	0	-0.25	-0.5	-0.75	-1
$\beta_2$	6	5.0293	4.2222	3.5527	3	2.5484	2.1884	1.9234	1

Qui sotto riportiamo il grafico della d.p. di questa variabile per  $\beta = 1, 0.5, 0, -0.5, -1$ , in riferimento a variabili standardizzate, quindi con varianza unitaria.



## 5. La variabile casuale normale multivariata

L'unica v.c. multivariata che studieremo e che sarà richiamata nella statistica è la *normale*. Partiamo dalla più semplice:

$$\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)^T$$

quella in cui le v.c. sono  $iid\mathcal{N}(0, 1)$  (si noti che il vettore di v.c. che costituisce la v.c. multipla è preso come un vettore colonna che per economia di spazio viene scritto come il trasposto, come indicato dall'apice  $^T$ , di un vettore riga).

Pertanto la f.d. congiunta risulta il prodotto delle f.d. univariate:

$$f(z_1, \dots, z_n) = f(z_1) f(z_2) \dots f(z_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2} z_i^2\} = \frac{1}{(2\pi)^{n/2}} \exp\{-\frac{1}{2} \sum_i z_i^2\}$$

che in termini matriciali può esprimersi come:

$$f(\mathbf{z}) = \frac{1}{(2\pi)^{n/2}} \exp\{-\frac{1}{2} \mathbf{z}^T \mathbf{z}\}.$$

Naturalmente il vettore di medie è il vettore nullo  $\mathbf{0}$ , costituito tutto da elementi pari a zero, e la matrice di covarianza è la matrice identità  $\mathbf{I}$ , costituita da uno sulla diagonale principale (le varianze) e zero sui restanti termini (le covarianze).

Si supponga ora di operare un insieme di  $n$  trasformazioni lineari determinate dalla matrice  $\mathbf{A}$  che danno una v.c. del tipo  $\mathbf{Y} = \mathbf{A} \mathbf{Z}$ , ove la matrice di trasformazione  $\mathbf{A}$  per semplicità si intende di rango pieno e quindi invertibile, per cui esiste la trasformazione inversa  $\mathbf{Z} = \mathbf{A}^{-1} \mathbf{Y}$ .

La v.c.  $\mathbf{Y}$  avrà  $E(\mathbf{Y}) = \mathbf{0}$  e  $var(\mathbf{Y}) = \mathbf{A} \mathbf{A}^T = \mathbf{\Sigma}$ . La f.d. congiunta risulta

$$f(\mathbf{y}) = f(\mathbf{z}) |\partial \mathbf{z} / \partial \mathbf{y}| = \frac{|\mathbf{A}^{-1}|}{(2\pi)^{n/2}} \exp\{-\frac{1}{2} \mathbf{y}^T (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{y}\} = \frac{|\mathbf{\Sigma}|^{-1/2}}{(2\pi)^{n/2}} \exp\{-\frac{1}{2} \mathbf{y}^T \mathbf{\Sigma}^{-1} \mathbf{y}\}$$

ove:

$$|\partial \mathbf{z} / \partial \mathbf{y}| = |\mathbf{A}^{-1}| = |\mathbf{A}|^{-1} = |var(\mathbf{Y})|^{-1/2}$$

è lo *Jacobiano* di trasformazione. Ora si fa l'ulteriore sostituzione  $\mathbf{Y} = \mathbf{X} - \boldsymbol{\mu}$ , che comporta uno *Jacobiano* identità.

Pertanto la forma generale della v.c. *multinormale* è

$$f(\mathbf{x}) = \frac{1}{|\mathbf{\Sigma}|^{1/2} (2\pi)^{n/2}} \exp\{-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\}$$

avente  $E(\mathbf{X}) = \boldsymbol{\mu}$  e  $var(\mathbf{X}) = \mathbf{\Sigma}$ , che sono i due parametri che caratterizzano la distribuzione. Si vede che l'argomento dell'esponenziale è una  $\chi^2$  con  $n$  g.d.l.

5.1 Si può notare la seguente particolarità che vale solo per la *normale*:  
l'indipendenza in senso debole implica quella in senso forte.

Infatti se tutte le covarianze sono nulle, la matrice  $\mathbf{\Sigma}$  si riduce ad una matrice diagonale e diagonale è anche la sua inversa e quindi la quantità dentro l'esponenziale si riduce ad una somma di quadrati:

$$(\mathbf{X} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) = \sum_i \sigma_i^{-2} (X_i - \mu_i)^2,$$

cosicché la f.d. congiunta si ottiene come la produttoria delle f.d. univariate:

$$f(\mathbf{x}) = \frac{|\mathbf{\Sigma}|^{-1/2}}{(2\pi)^{n/2}} \exp\{-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{x_i - \mu_i}{2\sigma_i}\right\}^2 = f(x_1) f(x_2) \dots f(x_n)$$

Come abbiamo visto in generale è vero solo il contrario.

In conclusione, per la *normale* multivariata indipendenza in senso forte ed in senso debole coincidono.

## 6. La variabile casuale normale bivariata

Come caso particolare è interessante vedere cosa succede nel caso bivariato. Il determinante moltiplicato e diviso per il prodotto delle due varianze è:

$$|\Sigma| = \text{var}(X_1) \text{var}(X_2) - \text{cov}^2(X_1, X_2) = (1 - \rho^2) \text{var}(X_1) \text{var}(X_2) = (1 - \rho^2) \sigma_1^2 \sigma_2^2$$

ricordando che:

$$\rho = \text{cov}(X_1, X_2) / \sqrt{\text{var}(X_1) \text{var}(X_2)} = \sigma_{12} / \sigma_1 \sigma_2.$$

Mentre l'inversa è:

$$\Sigma^{-1} = \frac{1}{|\Sigma|} \begin{bmatrix} \text{var}(X_2) & -\text{cov}(X_1, X_2) \\ -\text{cov}(X_1, X_2) & \text{var}(X_1) \end{bmatrix} = \frac{1}{1 - \rho^2} \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1 \sigma_2} \\ -\frac{\rho}{\sigma_1 \sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}.$$

Da qui possiamo ottenere l'espressione esplicita della f.d. congiunta:

$$f(x_1, x_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} \times \\ \times \exp \left\{ -\frac{1}{2(1 - \rho^2)} \left[ \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1 \sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right] \right\}$$

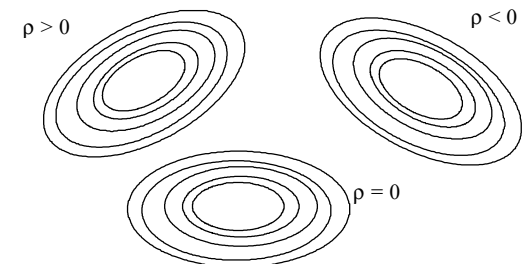
È possibile da qui ricavare le curve di *isoprobabilità*, definite come l'insieme dei punti che hanno la stessa densità di probabilità. Trovandoci nello spazio a tre dimensioni (le due v.c. e l'asse della densità di probabilità), occorre determinare le coppie di punti  $(x_1, x_2)$  che hanno lo stesso valore di  $f(x_1, x_2)$ , fissato anche se generico:

$$\frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1 \sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} = \text{cost.}$$

ed effettuando le standardizzazioni su entrambe le v.c. si ha:

$$z_1^2 - 2\rho z_1 z_2 + z_2^2 = \text{cost.}$$

che rappresenta, per  $|\rho| < 1$ , l'equazione di un'ellisse, avente asse maggiore orientato positivamente per  $\rho > 0$  e negativamente per  $\rho < 0$ ; infine per  $\rho = 0$  l'ellisse ha l'asse principale orientato secondo uno degli assi, come rappresentato nei seguenti diagrammi.



In particolare se le due varianze sono uguali, come nel caso standardizzato, si riduce ad una circonferenza. Attenzione particolare merita il caso in cui  $|\rho| = 1$ . In questo caso l'equazione dei punti di isoprobabilità diventa:

$$(z_1 + z_2)^2 = cost \quad \text{oppure:} \quad (z_1 - z_2)^2 = cost.$$

che rappresenta in entrambi i casi l'equazione di una retta, orientata positivamente o negativamente. Cioè, come sappiamo, nel caso di dipendenza lineare, fissata una v.c., l'altra diventa una v.c. degenere, ossia senza più variabilità: ossia assume uno ed un solo valore e tutta la massa di probabilità condizionata si concentra su quel punto.

6.1 È possibile anche ricavare l'espressione delle f.d. marginali e condizionate. La marginale rispetto a  $x_2$  si ottiene, come sappiamo, integrando rispetto alla  $x_1$ , la f.d. congiunta:

$$\begin{aligned} f(x_2) &= \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[-\rho^2 \frac{(x_2-\mu_2)^2}{\sigma_2^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\} \times \\ &\times \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1 \sigma_2} + \rho^2 \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\} dx_1 = \\ &= \frac{1}{\sqrt{2\pi} \sigma_1 \sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{x_1-\mu_1}{\sigma_1} - \rho \frac{x_2-\mu_2}{\sigma_2}\right]^2\right\} dx_1 \times \\ &\times \frac{1}{\sqrt{2\pi} \sigma_2} \exp\left\{-\frac{1}{2(1-\rho^2)}(1-\rho^2) \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right\} = \frac{1}{\sigma_2 \sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left[\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\} \end{aligned}$$

Pertanto la f.d. marginale è la f.d. *normale* nella sola v.c.  $x_2$ . Analogamente si ricava che la f.d. marginale della  $x_1$  è la sua f.d. univariata. La condizionata ad un valore  $x_1^*$  risulta:

$$\begin{aligned} f(x_2 | x_1^*) &= f(x_2, x_1^*) / \int_{-\infty}^{\infty} f(x_2, x_1^*) dx_2 = \\ &= \frac{\frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1^*-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1^*-\mu_1)(x_2-\mu_2)}{\sigma_1 \sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}}{\frac{1}{\sigma_1 \sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{(x_1^*-\mu_1)^2}{\sigma_1^2}\right]\right\}} = \\ &= \frac{1}{\sigma_2 \sqrt{2\pi} (1-\rho^2)} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{x_2-\mu_2}{\sigma_2} - \rho \frac{x_1^*-\mu_1}{\sigma_1}\right]^2\right\}. \end{aligned}$$

Pertanto la v.c. condizionata risulta una v.c. *normale* avente:

$$\mathcal{E}(X_2 | x_1^*) = \mu_2 + (\rho \sigma_2 / \sigma_1) (x_1^* - \mu_1) \quad \text{e} \quad \text{var}(X_2 | x_1^*) = \sigma_2^2 (1-\rho^2).$$

La funzione che lega il v.a. condizionato della v.c.  $X_2$  al valore condizionante  $m_1$ , funzione che in questo caso risulta *lineare*, si chiama *regressione*.

## Capitolo 8

### LEGGI DI CONVERGENZA STOCASTICA

In questo capitolo esponiamo dapprima la *diseguaglianza di Čebišëv*, quindi alcune definizioni di convergenza delle successioni di v.c., definizioni che costituiscono un ponte verso la Statistica.

#### 1. La diseguaglianza di Čebišëv

Ipotizziamo che la v.c.  $X$  abbia v.a. e varianza finite.

Partendo dalla definizione di varianza si ha:

$$\begin{aligned} \text{var}(X) &= \mathbb{E}[X - \mathbb{E}(X)]^2 = \int_{-\infty}^{\infty} [x - \mathbb{E}(X)]^2 d\mathcal{F}(x) \geq \\ &\geq \int_{-\infty}^{\mathbb{E}(X) - c} [x - \mathbb{E}(X)]^2 d\mathcal{F}(x) + \int_{\mathbb{E}(X) + c}^{\infty} [x - \mathbb{E}(X)]^2 d\mathcal{F}(x) \end{aligned}$$

avendo escluso dall'integrale la parte centrale tra  $[\mathbb{E}(X) - c, \mathbb{E}(X) + c]$ , con  $c$  costante positiva. Minorando ancora il secondo membro, al posto del valore corrente di  $[X - \mathbb{E}(X)]^2$  si può prendere il valore più piccolo che esso può assumere nel rispettivo intervallo di integrazione, valore che sarà:  $x = \mathbb{E}(X) - c$  per il primo e  $x = \mathbb{E}(X) + c$  per il secondo:

$$\begin{aligned}
var(X) &\geq \int_{-\infty}^{\mathbb{E}(X)-c} [\mathbb{E}(X) - c - \mathbb{E}(X)]^2 d\mathcal{H}(x) + \int_{\mathbb{E}(X)+c}^{\infty} [\mathbb{E}(X) + c - \mathbb{E}(X)]^2 d\mathcal{H}(x) = \\
&\int_{-\infty}^{\mathbb{E}(X)-c} (-c)^2 d\mathcal{H}(x) + \int_{\mathbb{E}(X)+c}^{\infty} c^2 d\mathcal{H}(x) = c^2 \text{Prob}[X \leq \mathbb{E}(X) - c] + c^2 \text{Prob}[X \geq \mathbb{E}(X) + c] \\
var(X) &\geq c^2 \text{Prob}[|X - \mathbb{E}(X)| \geq c].
\end{aligned}$$

Da cui l'espressione che massimizza l'area alle code:

$$\text{Prob}[|X - \mathbb{E}(X)| \geq c] \leq var(X) / c^2.$$

Se al posto di  $c$  si assume una costante equivalente, ma relativizzata alla deviazione standard:  $c^2 / var(X) = k^2$ , la stessa disuguaglianza si scrive:

$$\text{Prob}[|X - \mathbb{E}(X)| / \sqrt{var(X)} \geq k] \leq 1 / k^2$$

(naturalmente questa disuguaglianza ha senso solo per  $k > 1$ ).

Questa disuguaglianza è importante dal punto di vista pratico, perché in assenza di una precisa conoscenza sulla forma distribuzionale di una v.c., dà la possibilità di determinare il valore massimo della probabilità alle due code. Ma come vedremo subito è molto importante anche dal punto di vista teorico.

## 2. Convergenza in probabilità

Abbiamo già definito una successione di v.c.  $\{X_n\}$  ed anche la definizione di convergenza in legge: si dice che *una successione di v.c.  $\{X_n\}$  converge in legge ad un'altra v.c.  $Y$  quando il limite della successione delle f.r.  $\{\mathcal{H}(x_n)\}$ , associata alla successione di v.c., è la f.r.  $\mathcal{H}(y)$  della v.c. limite  $Y$* , in ogni punto di continuità della  $\mathcal{H}(x)$ .

Diamo ora un'altra nozione di convergenza basata sui valori che possono essere assunti dalla successione delle v.c.

Si dice che la successione  $\{X_n\}$  *converge in probabilità* verso una v.c.  $Y$  se, fissati comunque due numeri positivi  $\varepsilon$  e  $\delta$  è possibile determinare un intero  $n_{\varepsilon, \delta}$  tale che per ogni  $n > n_{\varepsilon, \delta}$  risulti:

$$\text{Prob}\{|X_n - Y| > \varepsilon\} < \delta$$

ossia

$$\lim_{n \rightarrow \infty} \text{Prob}\{|X_n - Y| > \varepsilon\} = 0.$$

Potremo scrivere tutto ciò come:

$$\lim_{n \rightarrow \infty} X_n \stackrel{P}{=} Y$$



oppure

$$X_n \xrightarrow{P} Y.$$

Supponiamo, nelle stesse condizioni, che la v.c. a cui tende la successione non sia una v.c. ma una costante o, se si vuole, una particolare v.c. che ha perduto proprio la variabilità: tutta la massa di probabilità è concentrata in un punto: è una v.c. degenera. In questo caso si dice che la successione *converge in probabilità* alla costante  $\xi$  e si indica come:

$$\lim_{n \rightarrow \infty} X_n \stackrel{P}{=} \xi \quad \text{oppure} \quad X_n \xrightarrow{P} \xi.$$

Per esempio, dire che  $X_n$  converge in probabilità a  $Y$  significa dire che  $\{X_n - Y\}$  converge in probabilità ad una costante particolare ossia lo zero.

Un teorema assicura che la convergenza in probabilità implica la convergenza in legge, il viceversa è assicurato solo nei confronti della convergenza in probabilità verso una costante.

Si tratta di stabilire che per  $n$  abbastanza grande la generica realizzazione  $x_n$  della v.c.  $X_n$  potrà discostarsi sempre meno dal limite  $\xi$ ; anzi la probabilità, che si discosti di una quantità positiva  $\varepsilon$  fissata, per quanto piccola, è sempre più piccola di un certo  $\delta$  assegnato ed al limite per  $n$  sempre più grande, tende ad annullarsi. Se la generica v.c.  $X_n$  è dotata di v.a. e varianza, ci si rende subito conto che dovrà necessariamente accadere che la varianza dovrà tendere ad annullarsi, poiché la v.c. non dovrà avere più variabilità e dovrà concentrarsi tutta intorno ad un valore, che coinciderà quindi con il v.a. che sarà proprio  $\xi$ . Dalla disuguaglianza di Čebišëv è possibile affermare che se:

$$\lim_{n \rightarrow \infty} E(X_n) = \xi \quad \text{e} \quad \lim_{n \rightarrow \infty} \text{var}(X_n) = 0.$$

allora:

$$\lim_{n \rightarrow \infty} \text{Prob}(|X_n - E(X_n)| > \varepsilon) < \lim_{n \rightarrow \infty} \text{var}(X_n) / \varepsilon^2.$$

ossia:

$$\lim_{n \rightarrow \infty} \text{Prob}(|X_n - \xi| > \varepsilon) = 0.$$

Pertanto quelle sono condizioni *sufficienti* affinché si possa affermare che  $X_n$  converge in probabilità a  $\xi$ .

2.1 Vi è un'altra definizione di convergenza in probabilità (*convergenza forte* o *quasi certa*) che implica la precedente:

$$\text{Prob} \left\{ \bigcup_{r=n}^{n+k} (|X_r - Y| \geq \varepsilon) \right\} < \delta$$

ossia

$$\text{Prob}\{\lim_{n \rightarrow \infty} |X_n - \xi| > \delta\} = 0.$$

Tale convergenza fa riferimento non solo ad *una* generica realizzazione  $X_n$ , ma a *tutte* le possibili che si potranno verificare da un certo posto in poi.

### 3. Convergenza in media quadratica

Si dice che una successione di v.c.  $\{X_n\}$  converge in media quadratica ad una costante  $\xi$ , se:

$$\lim_{n \rightarrow \infty} E(X_n - \xi)^2 = 0.$$

Infatti la quantità:

$$\begin{aligned} E(X_n - \xi)^2 &= E\{[X_n - E(X_n)] + [E(X_n) - \xi]\}^2 \\ &= E\{X_n - E(X_n)\}^2 + \{E(X_n) - \xi\}^2 + 2 \{E(X_n) - \xi\} E\{X_n - E(X_n)\} \\ &= var(X_n) + \{E(X_n) - \xi\}^2 \end{aligned}$$

(si ricordi che il doppio prodotto è nullo per la prima proprietà del v.a.:  $E\{X_n - E(X_n)\} = 0$ ) tende a zero se tendono a zero entrambi gli addendi, non potendo questi essere negativi. Ma, come abbiamo visto, queste sono le condizioni sufficienti per la convergenza in probabilità e quindi la convergenza in media quadratica implica la convergenza in probabilità. Quest'ultima convergenza è molto più agevole da verificare, poiché comporta un esame di solo due quantità caratterizzanti la v.c. e non di tutta la f.r.

### 4. La legge dei grandi numeri

Si prenda in considerazione la v.c. *frequenza relativa*:

$$F = V/n,$$

costituita dal rapporto della v.c. numero di successi  $V$  su  $n$  prove *bernoulliane* effettuate, avente com'è facile accertare

$$E(F) = \phi \quad \text{e} \quad var(F) = \phi \psi / n.$$

Per la disuguaglianza di Čebišëv si ha:

$$Prob[|F - \phi| \geq c] \leq \phi \psi / nc^2$$

Pertanto il limite, prendendo un valore  $c > 0$  piccolo a piacere:

$$\lim_{n \rightarrow \infty} Prob[|F_n - \phi| \geq c] = 0$$

assicura la convergenza in probabilità della v.c.  $F$  al proprio v.a.

Stessa proprietà si può stabilire per la v.c. *media aritmetica* di v.c. i.i.d. aventi v.a. e varianza finite:

$$M = \sum_{i=1}^n X_i / n,$$

che ha com'è noto:  $E(M) = E(X)$  e  $var(M) = var(X) / n$ . Ciò garantisce le condizioni sufficienti per la convergenza in media quadratica. Si ha quindi:

$$\lim_{n \rightarrow \infty} \text{Prob}[|M_n - E(X)| \geq c] = 0$$

che afferma la convergenza in probabilità della v.c. media aritmetica al proprio v.a. Tale legge è nota come *legge debole dei grandi numeri*.

4.1 Se vale la convergenza quasi certa si ha *legge forte dei grandi numeri*.

Applicata alla media aritmetica essa assicura che superato un certo  $n$ , non solo una generica realizzazione  $m_n$  non potrà discostarsi da  $E(X)$  più un dato  $\epsilon$ , se non con probabilità piccolissima, ma anche *tutte* le successive realizzazioni.

4.2 Qual è il senso della legge che abbiamo appena espresso per la media aritmetica e per la sua particolare rappresentazione costituita dalla frequenza relativa? Si faccia riferimento a quest'ultima attraverso un semplice esempio.

Si supponga di giocare ancora una volta a testa e croce e l'uscita di testa rappresenti il successo. Si supponga che il valore di  $E(X) = \phi$  non sia noto. Il nostro desiderio però è di pervenire alla individuazione di tale parametro.

Come possiamo fare, se abbiamo la possibilità di *ripetere* l'esperimento un numero arbitrariamente alto di volte?

Dalla definizione di convergenza in probabilità sappiamo che se avessimo a disposizione un *campione*, ossia una realizzazione di  $n$  prove indipendenti della stessa v.c., o una realizzazione di  $n$  copie indipendenti della stessa v.c., aumentando  $n$ , la frequenza relativa  $f = v / n$  basata su quel campione potrà sempre discostarsi dal valore di  $\phi$  che noi intendiamo determinare, ma la probabilità che tali scostamenti siano superiori ad un certo  $\delta$  fissato, per quanto piccolo, diviene sempre più bassa all'aumentare di  $n$ ; ossia è sempre più probabile che la realizzazione  $f$  si avvicini sempre più a  $\phi$ . Pertanto se abbiamo un campione sufficientemente grande possiamo dire che è trascurabile la divergenza tra il vero valore di  $\phi$  e la *stima* che di esso abbiamo effettuato attraverso il nostro campione.

È la prima volta che, affrontando un problema probabilistico, consideriamo come ignoto qualcosa che riguarda la distribuzione, in particolare un parametro che la caratterizza. Questo è il primo problema di *statistica inferenziale*, cioè in cui, contrariamente alla prospettiva probabilistica, qualcosa sulla distribuzione non è nota ed invece sono noti dei dati sperimentali, che costituiscono il campione.

4.3 Si badi bene ora a quale *non* è il senso della legge prima esposta.

Si supponga di proseguire la partita a testa e croce per un tempo molto lungo e che dopo un certo periodo si osservi una frazione della successione in cui si presentano di seguito un numero molto elevato di *permanenze*, diciamo sette teste di seguito. A questo punto si vuole sapere se esista una strategia da seguire migliore di quella di puntare a caso su una delle due facce. La risposta è inequivocabilmente: *no*. Infatti non ha alcun senso in questa situazione fare

riferimento a *ritardi*, poiché il processo è senza memoria, in quanto ogni lancio è del tutto indipendente dal precedente.

Se si chiedesse la probabilità di quell'evento (una successione di sette teste, seguita da un'ottava testa) *prima* che esso si verificasse, allora la risposta sarebbe ovviamente: la probabilità è molto scarsa; ma dal momento in cui le prime sette teste si sono già verificate (e quindi non ha più senso attribuire a quest'evento alcuna probabilità che non sia unitaria) l'ottava ha la stessa probabilità di verificarsi dell'altra faccia.

Ciò è del tutto ovvio, ma forse un'errata interpretazione della legge dei grandi numeri può condurre all'illusione che, poiché dopo un certo tempo *abbastanza lungo* il numero di teste e di croci *devono* compensarsi, essendo  $\phi = 1/2$ : se sono uscite *troppe* teste dovranno uscire *altrettante* croci.

Vediamo dove stanno gli errori.

Primo: qualunque affermazione basata sulla legge dei grandi numeri è asintotica e quindi non può valere per campioni limitati, anche se molto grandi, come sono tutti i campioni reali.

Secondo: il numero di teste e di croci non devono compensarsi; infatti la legge non afferma che  $v \rightarrow n\phi$ , bensì che  $v/n \rightarrow \phi$ , il che fa una bella differenza: è il rapporto tra due infiniti  $v$  e  $n$  che si stabilizza intorno ad una costante, “oscurando” anche differenze relevantissime tra numero di teste e di croci uscite (cento teste in più o in meno su milioni di lanci sono del tutto plausibili), e non il rapporto tra due grandezze infinite  $v$  ed  $n\phi$  che tendono ad eguagliarsi! In realtà la v.c.  $V$  ha varianza  $n\phi\psi$  che, essendo una *somma* e non una *media* di v.c. i.i.d., tende a *crescere* indefinitamente con  $n$  e non a scomparire, ossia è una v.c. il cui esito è *sempre meno prevedibile*.

Terzo: ammesso che il numero di teste aumenti fino a diventare “sospetto”, l'unica cosa che si dovrà ammettere è che  $\phi = 1/2$  non è il limite di  $v/n$ , in favore di un  $\phi > 1/2$  e quindi, anziché scommettere sulla croce come farebbe un sostenitore della teoria dei ritardi, si dovrebbe scommettere proprio sulla testa con uno pseudo-calcolo di questo tipo: “non so perché escono più frequentemente le teste che le croci, ma dato che è così, e permanendo il guadagno pari alla puntata, è conveniente puntare sulla testa”.

Ovviamente in una partita che veda la presenza di una moneta “equa”, anche questa strategia è del tutto sbagliata, o meglio non vi è ragione di preferirla ad altre, anche le più bizzarre.

Estendere le considerazioni precedenti a giochi quali il lotto, gioco nel quale più frequentemente si sbizzarriscono le malfamate “teorie” dei ritardi proposte da “esperti” di pochi scrupoli, è del tutto immediato. La risposta a coloro che non sono in grado di seguire il ragionamento prima esposto è lo studio già effettuato su *tutta* la storia del lotto italiano, che conferma come i “ritardi” riscontratisi sono *tutti* all'interno della più pura casualità.

## CAPITOLO 9

### NUMERI PSEUDO-CASUALI

La generazione dei numeri casuali e il metodo Montecarlo spesso sono considerati strumenti poveri della statistica. Negli ultimi tempi invece, grazie allo straordinario diffondersi di strumenti di calcolo sempre più economici e sempre più veloci, tali metodi di simulazione sono entrati di diritto nella metodologia statistica, sia nella ricerca teorica che nelle metodologie pratiche.

Oggi ogni package statistico possiede la possibilità di generare numeri casuali e ritengo che questo sia un ottimo strumento per “vedere” una variabile casuale, anche la più complicata, nascere e muoversi secondo i propri desideri.

#### 1. I numeri pseudo-casuali

Può essere nostro desiderio “simulare” una v.c., ossia ottenere una successione di repliche indipendenti (o al solito di realizzazioni di copie indipendenti) che possano essere considerate come *campioni* estratti da quella v.c.

Per una popolazione *bernoulliana* è molto semplice: basta realizzare una partita a testa e croce; ma se il numero di estrazioni desiderate è molto elevato, il gioco può diventare a dir poco molto tedioso.

Se invece la v.c. dalla quale vogliamo fare delle estrazioni è per esempio una v.c. *normale*, si può fare riferimento ad una popolazione reale, ma occorre precedentemente accertarsi che essa segua effettivamente la legge distribuzionale desiderata.

Per ovviare ad entrambi questi inconvenienti è possibile realizzare delle procedure che usano l'elaboratore elettronico come una *pseudo-urna* in grado di generare in tempi rapidissimi milioni di estrazioni che seguono le distribuzioni da noi desiderate.

Perché parliamo di “pseudo-urna”? Perché un elaboratore è in grado di eseguire una successione di operazioni che nulla hanno di casuale ma che, opportunamente combinate tra di loro, danno luogo a risultati che possono considerarsi come se fossero il risultato di operazioni casuali, almeno rispetto a certe caratteristiche precise. Per capire di cosa si tratta, facciamo un esempio. Si prenda un numero a caso compreso tra zero ed uno ed espresso con la precisione di quattro cifre decimali:  $0.abcd$ . Di esso si estraiga la radice quadrata e, delle otto cifre decimali di cui è composto il numero:  $0.efghijk$ , si prendano le quattro centrali:  $ghij$ . Con esse si costituisca un nuovo numero che ha le stesse caratteristiche del primo:  $0.ghij$ : un numero compreso tra zero ed uno di quattro cifre decimali. Tale procedura pertanto può essere iterata, ottenendo dal secondo un terzo numero, da questo un quarto e così via:

$$X_n = f(X_{n-1}),$$

ove con  $f()$  indichiamo proprio l'operazione descritta precedentemente.

Ovviamente quando il primo numero  $X_1$  è stabilito, anche la successione degli altri è determinata, ma se il primo numero e solo il primo è scelto a caso, tutta la successione si può considerare come casuale.

In realtà, quando si parla di numeri *pseudo-casuali*, più che ad una successione finita, si fa riferimento alle proprietà del metodo che li ha generati. Infatti non ha nessuna importanza che si presentino occasionalmente delle successioni “strane” ed anzi anche tali successioni devono poter apparire con la dovuta probabilità. Ciò che occorre stabilire sono le proprietà *a lungo andare* di tali metodi.

Per esempio, si desidera verificare se il metodo su esposto è in grado di fornire numeri pseudo-casuali che possano pensarsi provenienti da una distribuzione uniforme nell'intervallo  $(0, 1)$ . Per verificare ciò, occorre accertarsi che nessuna cifra sia sovra o sottorappresentata rispetto alle altre. Ma evidentemente anche la successione ordinata dei numeri da 0.0001 a 0.9999 gode della stessa proprietà e questa certo non può essere considerata come una successione pseudo-casuale. Bisogna anche accertarsi che non vi siano delle particolari ripetizioni di cifre, di coppie di cifre, di terne di cifre, ecc. Solo ciò ci mette in condizione di considerare il processo come casuale.

Nell'esempio specifico il metodo proposto non è buono, infatti la cifra zero si vede ben presto che risulta sovra rappresentata, ed anzi una volta che si dovesse incorrere in numero del tipo  $0.xy0000zw$  si ottiene come numero successivo 0.0000 dal quale non ci si muove più. Tuttavia per brevissime sequenze tale metodo può ancora andar bene.

1.1 Un'altra caratteristica che si richiede ad un generatore di numeri casuali è quella della completezza. Nel caso sopra esposto si desidera che tutti i 9999 numeri che possono essere generati siano ricavati, prima di ritrovare uno dei numeri già trovati e quindi ripresentare la sequenza già ottenuta.

Il metodo che garantisce le proprietà suddette e che viene più spesso usato è il *generatore congruenziale*. Da un numero *intero* compreso tra 1 e

$$P = 2^n - 1,$$

si effettui il prodotto per un moltiplicatore dato  $\lambda$  e l'operazione di modulo per  $P$ :

$$X_{n+1} = (\lambda X_n) \bmod P.$$

Una delle possibili scelte, che viene usata comunemente nei package statistici è la coppia di numeri:

$$\lambda = 7^5 \quad \text{e} \quad n = 31.$$

Si badi bene che questo metodo consente di generare oltre *due miliardi di numeri tutti diversi* prima di ricominciare la stessa sequenza. Naturalmente si potrà scegliere un numero di partenza a caso e tutta la sequenza si potrà considerare casuale.

Per ottenere numeri pseudo casuali che si possano considerare estratti da una v.c. uniforme nell'intervallo  $(0, 1)$ ,  $U(0, 1)$ , basta dividere ogni numero generato per  $2^n$ .

Ovviamente il numero così ottenuto non è un numero *reale*, così come non lo è la rappresentazione digitale che si ha su computer. Esso è solo un'*approssimazione* di un numero reale, che però per gli scopi eminentemente pratici a cui ciò è finalizzato è più che sufficiente.

## 2. Numeri pseudo-casuali distribuiti secondo una distribuzione data

Una volta ottenuti numeri  $U(0, 1)$ , è possibile ottenere numeri pseudo-casuali che seguano qualunque distribuzione data. Si ricordi infatti che la v.c.  $y$  ottenuta dalla f.r. di qualunque v.c.  $x$ :  $y = F(x)$ , è una v.c.  $U(0, 1)$ .

Pertanto da una successione di v.c.  $y_i$  si possono ottenere i valori  $x_i$  invertendo la f.r.:

$$x_i = F^{-1}(y_i).$$

Per esempio si desideri ottenere numeri pseudo-casuali che seguano una distribuzione esponenziale con fissato  $\lambda$ .

Poiché in questo caso è  $U = F(x) = 1 - e^{-\lambda x}$ , si ha:

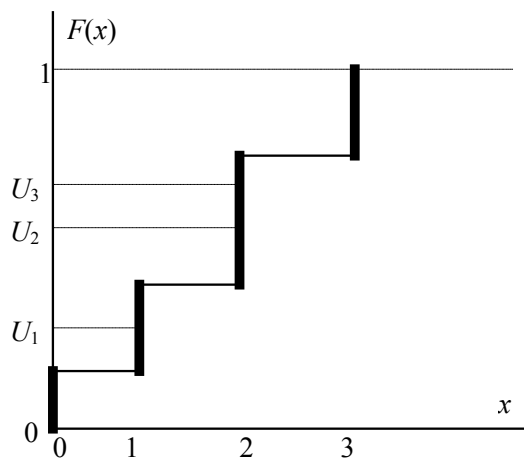
$$X = -\log(U) / \lambda$$

(si noti che nel caso particolare  $1 - U$  ed  $U$  hanno la stessa distribuzione). Per ogni numero  $U_i$  si ottiene un numero  $X_i$  distribuito come richiesto.

2.1 Tuttavia, se dal punto di vista teorico la cosa è del tutto risolta, dal punto di vista pratico è più complessa di quanto sembri.

Per v.c. discrete il problema può essere complesso solo dal punto di vista computazionale, se il numero di punti ove la v.c. è definita è molto grande o illimitato. Ma tuttavia il principio dell'inversione della f.r. resta inalterato.

Un esempio grafico mostra meglio di ogni parola quanto si deve fare: la generazione  $U_1$  dà luogo al numero pseudo casuale  $X=1$ ; mentre le due generazioni  $U_2$  e  $U_3$  danno luogo a  $X=2$ .



2.2 Per v.c. continue invece spesso la f.r. non è esplicitabile, come per esempio la normale, e tanto meno la sua inversa. A questo scopo esistono o degli altri metodi che fanno uso di altre procedure e che qui non trattiamo, o approssimazioni numeriche della  $F^{-1}()$  che risolvono il problema.

Tali approssimazioni sono disponibili in tutte le librerie di subroutine matematiche e statistiche.

Tuttavia sottolineiamo il fatto che l'utente oggi non ha nessuna necessità di tali conoscenze, in quanto i package statistici a disposizione contengono tutti la possibilità di generare numeri pseudo-casuali secondo le più diffuse v.c.

Se ci siamo soffermati tanto sull'argomento è stato per far comprendere al lettore cosa succede *dentro* un package di quel tipo e consentirgli di afferrare meglio il significato di numero pseudo-casuale.



### 3. Il metodo Montecarlo

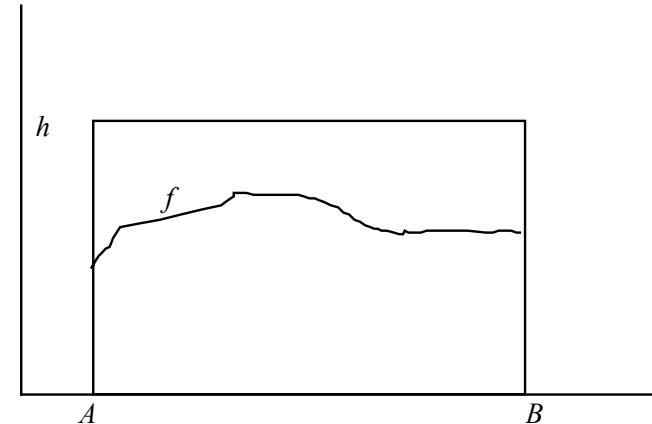
Il metodo Montecarlo consiste nel *simulare* un processo aleatorio replicabile un numero molto grande di volte, verificando così il suo comportamento *a lungo andare*.

Per esempio, si voglia simulare il teorema limite centrale. Si generino campioni di dimensione sufficientemente grande, diciamo  $n = 100$  unità, da una popolazione *bernoulliana* avente  $p$  fissato. Si determini per ogni campione il numero di successi  $v$  ed il valore della realizzazione della v.c. standardizzata

$$z = (v - np) / \sqrt{npq}.$$

Si ripeta l'esperimento un numero  $N$  molto grande di volte, per esempio un milione di volte. Si costruisca la distribuzione di frequenza delle  $z_i$  ( $i=1, \dots, N$ ) e si riporti su un diagramma cartesiano  $z_i$  e  $\text{Prob}(Z = z_i)$  tale distribuzione. A meno che  $p$  non sia molto piccolo esso dovrebbe assumere l'aspetto tipicamente campanulare della distribuzione *normale*.

3.1 Ma non sono solo processi aleatori che possono essere simulati col metodo Montecarlo. Per esempio, si voglia determinare il valore di un integrale definito in un intervallo limitato  $(A-B)$ , ossia l'area sottesa alla curva  $f$ . Fissato un punto  $A < x < B$  ed  $0 < y < h$ , si può stabilire se il punto di coordinate  $(y, x)$  si trovi entro o fuori l'integrale.



Generato un numero molto elevato di coppie di punti  $N$ , seguenti la v.c. bivariata costituita da due v.c. indipendenti uniformi negli intervalli  $x \in (A, B)$  e  $y \in (0, h)$ , si calcoli la proporzione  $v$  di punti che cadono entro l'area sul

numero totale  $N$ . Una *stima* dell'integrale è data dalla frazione dell'area totale considerata proporzionale alla frequenza relativa dei successi:

$$(B - A) h v/N.$$

Naturalmente per un caso così banale di integrale nel piano vi sono strumenti molto più efficaci, ma per integrali con molte più variabili, il metodo di Montecarlo può dare risultati molto precisi e rapidi, soprattutto se integrato con o a supporto di altri metodi.

Ovviamente tale frequenza relativa è solo un'approssimazione della vera proporzione  $p$ , ma tale v.c. come sappiamo ha distribuzione nota: la *binomiale*. Attraverso la statistica saremo in grado di precisare qual è il margine di errore atteso massimo tra questa stima e il valore vero incognito di  $p$  in base all' $N$  prescelto.

3.2 Si supponga di voler determinare sperimentalmente il numero  $\pi$ . Si costruisca quindi un cerchio di raggio unitario. Limitandosi al primo quadrante, si generino coppie di numeri pseudo-casuali uniformi indipendenti  $U_1(0, 1)$  e  $U_2(0, 1)$ . Per ogni coppia si accerti se il punto  $(U_1, U_2)$  cade entro il cerchio o no, verificando se  $U_1^2 + U_2^2 < 1$ . Si contino quindi quanti punti  $v$  sono caduti entro il cerchio diviso il numero di punti osservati  $n$ . Per il teorema limite centrale, la frequenza relativa  $f = v/n$ , si distribuirà normalmente con  $E(f) = \pi/4$  e  $var(f) = \pi(4 - \pi)/(16n)$ ; e quindi:

$$4 v / n \propto \mathcal{N}\left(\pi, \frac{\pi(4 - \pi)}{n}\right).$$

Pertanto si può ammettere che con una probabilità pari al 95%, si stimerà il vero valore di  $\pi$  con un errore non superiore a  $\pm 1.96 \sqrt{\pi(4 - \pi)/n}$ .

Posto quindi tale errore non superiore ad 0.001, ponendo  $\pi$  pari approssimativamente a 3.14, si ottiene:

$$n > 1.96^2 [3.14 (4 - 3.14)] 10^6 > 10\,373\,856.64.$$

Generando quindi col metodo congruenziale 10 373 857 coppie di numeri casuali, partendo dal seme iniziale 97531, si è ottenuta la stima  $\hat{\pi} = 3.141677$ , che differisce come previsto meno di 0.001 dal valore vero.