

STATISTICA

1. IL CAMPO DI STUDIO E LE FINALITÀ DELLA STATISTICA.

"Statistica" è un termine con un significato amplissimo sia per la varietà delle sue applicazioni sia per le possibilità di impiego dei suoi strumenti. Gran parte del pensiero scientifico ne adotta il linguaggio e le proposte metodologiche. Le definizioni che vengono date di Statistica sono numerosissime. Ciò non deve meravigliare, poiché la Statistica soddisfa esigenze di scienze diverse e trae origine da vari filoni di ricerca. Giova ricordare al riguardo che lo stesso termine "statistica", derivato da "Stato" nel senso di comunità politicamente organizzata, non rispecchia il contenuto attuale della disciplina, essendo stato proposto nel secolo XVIII per indicare una materia di studio, nata nelle università tedesche, consistente nella "descrizione comparata" degli Stati, mediante l'analisi, da un punto di vista essenzialmente qualitativo, dei fattori storico-geografici, socio-economici e politici.

Da quanto detto emerge che se volessimo fornire una definizione della Statistica ci troveremmo presto in difficoltà. Preferiamo, pertanto, rinunciare a questo compito e limitarci a presentare il **campo di studio e le finalità della Statistica**.

Per delineare il campo di studio e le finalità della statistica può essere utile portare un esempio: ci proponiamo di indicare al lettore i paragrafi del presente lavoro dove potrà trovare gli approfondimenti del caso e le relative indicazioni bibliografiche.

L'esempio che intendiamo discutere esamina alcuni aspetti del comportamento di un imprenditore che ha costituito una impresa per la produzione e la vendita di prodotti dolciari ed intende espandere la sua attività economica aprendo un nuovo stabilimento.

Un problema come quello che si va costruendo può richiedere non solo l'intuito dell'imprenditore per la valutazione delle convenienze dell'operazione ma anche l'intervento di persone particolarmente qualificate per predisporre gli studi e i piani necessari. Possiamo dire che spetta alla statistica fornire metodi che aiutino ad analizzare i dati ed a fornire modelli che mettano in grado chi deve prendere decisioni, nel nostro caso l'imprenditore, di farlo nelle migliori condizioni.

Il problema dell'imprenditore è quello di aprire o non aprire il nuovo stabilimento ed a ciò sono collegati i guadagni o le perdite dovute al successo o all'insuccesso dell'iniziativa. Ci troviamo di fronte ad un problema di **teoria delle decisioni** (si veda il paragrafo 4.4 ed il 4.6.5).

Le basi della moderna teoria delle decisioni sono state poste da Wald (1950). Fino agli inizi degli anni cinquanta la Statistica non era considerata la disciplina "naturale" per lo studio dei problemi di decisione. Infatti la maggior parte degli studiosi era ancora impegnata a dare costrutto teorico ai fondamenti della statistica e non era in grado di fornire una solida rete concettuale per ulteriori sviluppi e più pratiche applicazioni. (a tal proposito si veda il paragrafo 2). Da allora l'interesse per questi problemi è andato via via aumentando tanto da far dire a molti studiosi moderni che lo scopo principale, se non l'unico, della statistica è quello dello studio delle decisioni sotto incertezza.

Nell'esempio possiamo immaginare che la prima preoccupazione dell'imprenditore, che supponiamo intenda restringere il proprio mercato alla zona in cui avrà sede il nuovo stabilimento, sia quella di avere informazioni sulla zona in questione.

Egli è interessato alle caratteristiche della popolazione, delle famiglie, delle abitazioni, dell'apparato industriale e delle aziende agricole, alla spesa per generi alimentari e bevande, ai consumi delle famiglie, al reddito delle famiglie, ecc.... Su una tale massa di dati si può condurre una **analisi esplorativa** che metta in evidenza la struttura dei dati, ne esamini la qualità e li predisponga in una forma utile alle ulteriori analisi (si veda il paragrafo 2.1). Sui dati a disposizione si può altresì

condurre una **analisi descrittiva** tesa a: classificare i dati, fornirne le **distribuzioni** e le loro rappresentazioni (si veda il paragrafo 3.2.1); calcolare indici sintetici delle principali caratteristiche delle distribuzioni (medie, indici di variabilità, indici di concentrazione, indici di forma (si veda il paragrafo 3.2.2); studiare le relazioni tra le diverse variabili anche calcolando indici opportuni (si veda il paragrafo 3.2.3.).

L'imprenditore può essere interessato, per una localizzazione ottimale dello stabilimento, alla individuazione di aree omogenee sotto l'aspetto agricolo per i prodotti di base necessari alla produzione che ha intenzione di avviare. Per fare ciò è utile ricorrere ad una tecnica per l'**analisi di dati multivariati** quale la **cluster analysis** (si veda il paragrafo 4.4).

In molti casi è necessario ricorrere a rilevazioni parziali mediante sondaggio, con il metodo delle interviste, per l'impossibilità di eseguire una rilevazione totale. Vale la pena sottolineare che, anche nei casi ove una rilevazione totale fosse possibile, un sondaggio è spesso più conveniente. Nel nostro esempio è abbastanza agevole immaginare che dati, quali quelli sul consumo di dolci, possano essere ottenuti mediante una **indagine campionaria** su una piccola parte della popolazione della zona interessata. Si porrà quindi il problema di ricavare, sulla base dei dati campionari, informazioni sull'intera popolazione. Le conclusioni a cui si giunge sulla base di un campione saranno "corrette" solo quando il "**campione**" riproduce al meglio la popolazione. Un campione che riproduce al meglio le caratteristiche della popolazione è detto rappresentativo. Come estrarre, allora, il campione in modo che sia rappresentativo? E' chiaro che se gli individui da intervistare venissero scelti in modo da rappresentare soltanto una categoria (es. lavoratori dipendenti) i risultati sarebbero falsati. Si può evitare tale distorsione facendo sì che gli elementi del campione siano indipendenti dal carattere studiato ed il modo migliore è quello di scegliere le unità da inserire nel campione in modo **casuale**. In molte situazioni l'efficienza si migliora sostituendo alla scelta casuale sul complesso della popolazione una scelta casuale entro singole classi (es. suddividendo i comuni in agricoli ed industriali) e si ha allora un **campione stratificato** (si veda il paragrafo 5.5). Sui dati campionari è possibile impiegare le tecniche di analisi esplorativa e descrittiva (citate in precedenza), tuttavia se volessimo utilizzare le informazioni fornite dal campione per trarre conclusioni relative alla popolazione avremmo un problema di **inferenza statistica**.

Si ha una inferenza statistica quando, sulla base della informazione fornita dall'osservazione di alcuni fatti, si formulano supposizioni o previsioni riguardanti fatti rimasti incerti.

Le metodologie induttive oggi in uso appaiono riconducibili a due diverse impostazioni, di differente origine storica (impostazione frequentista-oggettivistica, impostazione bayesiana). Queste due diverse impostazioni non sempre convivono pacificamente tra gli statistici teorici, anche se in questi ultimi anni l'antica polemica sembra attenuarsi e sopravvive soprattutto attorno alla definizione ed alla natura -soggettiva od oggettiva - della probabilità. Per più di un secolo, tuttavia, a partire dalla metà dell'800, la polemica tra le due scuole è stata piuttosto aspra.

Supponiamo, ora, che l'imprenditore desideri fare inferenza sul consumo medio annuo di dolci (indichiamolo con θ). Il valor medio nel campione può costituire una buona indicazione del valor medio θ nell'intera popolazione. Avremmo però potuto osservare solo persone che consumano pochissimi dolci e quindi sottovalutare θ . Come fare allora per vedere se la media campionaria è una "buona stima" di θ ?. Per fare ciò bisogna introdurre una esplicita assunzione che colleghi le osservazioni con θ . Tale connessione può essere stabilita se riusciamo a specificare la probabilità di osservare ogni possibile risultato campionario, probabilità che sarà, in generale, diversa a seconda del valore di θ . Tale legge di probabilità si dice **modello statistico**. Risulta pertanto evidente uno stretto legame con la **teoria della probabilità** (si veda il paragrafo 4.1)

Il modo tradizionale di procedere per risolvere un problema di inferenza statistica è proprio quello di assegnare un **modello statistico** (si veda il paragrafo 4.3) e di servirsi delle osservazioni per ottenere informazioni su di esso.

Nell'esempio siamo interessati alla **stima** del consumo medio annuo di dolci delle famiglie della zona. Nell'ipotesi che ci si possa impegnare sulla legge di distribuzione del consumo delle famiglie siamo di fronte ad un problema di **stima parametrica**. La stima può essere effettuata in due modi differenti: **stima puntuale o per intervallo**. Nel primo caso si assegna un valore plausibile al

parametro sulla base del campione e tali stime vengono usualmente corredate da misure di attendibilità; nel secondo caso si fornisce un intervallo il quale verosimilmente contiene il valore del parametro incognito. Nella impostazione classica (frequentista) il parametro incognito è riguardato come una costante non nota e la costruzione di una stima si basa sulla sola informazione campionaria.(si veda il paragrafo 4.6). Uno dei metodi più usati nella teoria della stima è quello della **massima verosimiglianza** (si veda il paragrafo 4.6.1).

E' facile d'altra parte pensare che l'imprenditore abbia una idea sulla distribuzione del consumo medio annuo, idea che si è fatta sulla base delle proprie conoscenze, delle proprie impressioni, delle proprie aspettative e delle proprie informazioni. In tale circostanza possiamo assimilare il parametro ad una variabile aleatoria dotata di una distribuzione (**distribuzione iniziale**) valutata soggettivamente dall'imprenditore. I dati campionari aumentano la conoscenza sul parametro θ . Tale impostazione si pone nel filone di ricerca che va sotto il nome di **bayesiana** ed il problema della stima si riduce ad un problema di sintesi della **distribuzione finale** (si veda il paragrafo 4.5).

Una domanda interessante è la seguente: se alcune assunzioni fatte sulla forma del modello fossero false i metodi utilizzati fornirebbero ancora buoni risultati ?. La risposta si può trovare nell'analisi della **robustezza** (si veda il paragrafo 5.5).

Può essere altresì utile ai fini dell'analisi controllare mediante **test non parametrici (test di accostamento)** se la forma della distribuzione della popolazione da cui è stato estratto il campione è proprio del tipo specificato (si veda il paragrafo 4.6.4).

Nel caso non ci si possa impegnare sulla legge di distribuzione della popolazione sono necessari metodi di **stima non parametrici**. In tale circostanza è possibile utilizzare delle **tecniche di ricampionamento** per poter valutare meglio le proprietà degli stimatori. proposti (si veda il paragrafo 5.2.).

Sul consumo medio di dolci delle famiglie della zona potremmo anche avanzare alcune congetture o ipotesi (ad esempio : se $\theta \geq A$ siamo costretti a raddoppiare l'unità produttiva). In tali situazioni ricorriamo alla **teoria della prova delle ipotesi** (classica o bayesiana). La teoria della prova delle ipotesi riguarda intuitivamente il seguente problema: le osservazioni fatte sono consistenti con l'ipotesi avanzata o no? (si veda il paragrafo 4.5.3 per l'impostazione bayesiana ed il paragrafo 4.6.3 per quella classica).

Un problema delicato è quello della corretta interpretazione da dare ad "accettare" o "rifiutare" una ipotesi. Certamente non possiamo dire di ritenere vera una ipotesi o che i dati osservati conducono a ritenerla vera, possiamo solamente dire che la scelta di una ipotesi da una indicazione di comportamento. Nel nostro problema le ipotesi da verificare ($\theta \geq A$ e $\theta < A$) sono collegate ad una scelta tra due azioni: raddoppiare o meno l'unità produttiva. Sebbene il collegamento tra azione ed ipotesi possa essere molto stretto non è corretto identificare azione ed ipotesi. Accettare una ipotesi non significa ritenerla vera, pensando necessariamente falsa l'alternativa, ma comportarsi, ai fini di una ben delimitata questione, come se quell'ipotesi fosse vera.

Un dato importante nel nostro problema può essere quello riguardante l'esistenza di altri stabilimenti del settore e l'esame della loro produzione. E' agevole immaginare che tali dati riguardanti la produzione riguardino anni diversi. In tale circostanza ci troviamo di fronte ad una serie storica ed utilizzando i metodi tradizionali di **analisi delle serie storiche** potremmo provare ad individuare il trend, le variazioni stagionali (ad esempio si vendono più dolci in inverno che in estate, si vendono più dolci in prossimità delle festività), alcuni cambiamenti ciclici e le restanti fluttuazioni.

Supponiamo ora che in base ai dati ottenuti sia stata presa la decisione di aprire il nuovo stabilimento.

E' chiaro che i problemi non sono finiti ma ne sorgono altri di natura diversa ed una possibile risposta a tali quesiti si può dare utilizzando metodi statistici appropriati. Al nostro imprenditore non basta la stima della domanda iniziale ma interessa anche il tasso di crescita della domanda attesa per il prossimo futuro. Questa previsione dipende evidentemente dalla crescita della popolazione, dal livello del reddito dell'area considerata, dalla competitività, dalla pubblicità e da

altre variabili. Una proposta per risolvere tale problema può essere data dall'**analisi della regressione** e dalle **tecniche di analisi multivariata** ad essa collegate (si veda il paragrafo 5.1).

Se nella produzione di un determinato tipo di dolce vengono impiegate macchine diverse diventa importante verificare mediante un gruppo di assaggiatori se il gusto rimane invariato. In alcune circostanze non solo vengono utilizzati diversi tipi di macchinario ma anche diversi tipi di farina. Le domande che ci poniamo sono le seguenti: sul gusto vi sono degli effetti dovuti al macchinario ? ; vi sono degli effetti dovuti ai diversi tipi di farina ?; vi sono macchinari che possono essere utilizzati solo per un tipo di farina ?. A queste domande si può rispondere con i metodi dell' **analisi della varianza** (si veda il paragrafo 5.1.2).

Dopo un primo periodo l'imprenditore decide di immettere sul mercato un nuovo prodotto. Vengono provate diverse ricette e mediante un gruppo di assaggiatori si valuta la desiderabilità delle diverse alternative. Questo è un classico problema di **disegno ed analisi degli esperimenti**.(si veda il paragrafo 5.3) E' forse utile rilevare che l'esperimento oltre a consentire accurati confronti tra le diverse ricette deve anche fornire le prove della significatività statistica delle differenze. Ne segue che tutti i trattamenti devono essere ripetuti in condizioni invariate- entro i limiti delle possibilità pratiche-affinchè i confronti tra i vari risultati possano offrire elementi attendibili per giudicare il significato delle differenze osservate.

Per disposizione di legge è necessario controllare che il peso effettivo delle confezioni e quello dichiarato non sia molto diverso (le differenze esistono sia per la variabilità della produzione, sia per altre cause come l'umidità, ecc.) Per un controllo di questa natura si può utilizzare sia un campione di ampiezza fissata sia un **campionamento sequenziale**, determinando pertanto un procedimento di arresto che indichi quando interrompere il campionamento e prendere una decisione. Giova osservare che a volte il procedimento sequenziale è più economico di quello con ampiezza campionaria fissata perché richiede un numero inferiore di prove. Il campionamento sequenziale costituisce quindi un indubbio vantaggio (si veda il paragrafo 5.4).

Supponiamo infine che l'imprenditore voglia sapere se il nuovo prodotto immesso sul mercato è stato gradito o non gradito dai consumatori. Viene estratto un campione e viene rilevata solamente la presenza dell'una o dell'altra modalità. Il problema è quello di valutare la probabilità che, in un certo numero di prove future, la modalità "gradito" si presenti secondo una frequenza assegnata, noti che siano i risultati di prove passate. Tale problema è di natura squisitamente **previsiva** (si veda il paragrafo 4.7)

Evidentemente anche altri problemi, oltre a quelli elencati, possono essere studiati con l'ausilio di opportune tecniche statistiche, si pensi al controllo di qualità, alla selezione del personale, all'analisi delle vendite giornaliere, alle scorte, ecc...

Dopo aver discusso a grandi linee, questo esempio pensiamo di poter affermare che:

il campo di studio della STATISTICA è costituito dal complesso degli strumenti concettuali, delle tecniche e dei procedimenti matematici che servono alla descrizione, all'analisi di fatti osservati e alla miglior utilizzazione delle informazioni che essi forniscono ai fini della previsione di fatti non osservati.

Esempi come quello riportato possono essere numerosissimi ed interessare moltissimi campi di applicazione.

Da ciò discende che vi sono importanti connessioni tra la statistica ed i vari campi di applicazione della stessa. Basti pensare come molte tecniche statistiche siano state suggerite proprio dalle scienze in cui venivano applicate .

A scopo puramente espositivo nel seguito seguiremo l'usuale distinzione della Statistica in **STATISTICA DESCRITTIVA** e **STATISTICA INFERENZIALE** (paragrafi 3 e 4).

2 UNO SGUARDO STORICO.

Le origini della statistica sono state collocate in epoche diverse a seconda del significato attribuito al termine e della prospettiva storiografica adottata. La storia dei fondamenti della statistica è

strettamente connessa con il concetto di probabilità e con la storia di tutte le altre scienze. Si potrebbe dire che la statistica nasce con l'idea di numero. Il contare è infatti una operazione caratteristica della vita quotidiana, con la quale abbiamo spesso a che fare per motivi di ordine pratico. E' indubbio allora che l'idea di contare e di classificare popolazioni umane, animali, oggetti è tanto antica quanto quella di numero. Tuttavia uno dei principali filoni a cui si suole far risalire gli inizi della nostra disciplina nacque in Inghilterra ad opera di John Graunt (1620-1674) e di William Petty (1623-1687), fondatori dell' "aritmetica politica" (la disciplina che, più tardi, assumerà il nome di demografia). Graunt nel 1662 pubblicò un libro dal titolo "Natural and Political Observations on the Bills of Mortality". L'opera è interessante non solo per le informazioni raccolte e le regolarità riscontrate (rapporto tra i sessi, maggiori nascite maschili nel rapporto 14 a 13 a Londra e 15 a 14 in campagna, costruzione di tavole di mortalità) ma anche per i collegamenti con i principi del calcolo delle probabilità introdotti da Gerolamo Cardano (1501-1576), Blaise Pascal (1623-1662), Pierre de Fermat (1601-1665), Christian Huygens (1629-1695). Ricordiamo inoltre il tedesco Süssmilch (1707-1767) che raccolse ed elaborò, con il metodo utilizzato da Graunt, i dati demografici allora disponibili e diede un carattere di generalità ad alcune regolarità demografiche già riscontrate da Graunt stesso.

In astronomia Roger Cotes (1682-1716) verificò che la miglior stima della parte sistematica di un gruppo di osservazioni affette da errore, riguardanti un fenomeno fisico, è la loro media aritmetica. Tale idea diede origine al principio dei minimi quadrati e alla nozione di distribuzione degli errori. Il principio dei minimi quadrati venne formulato da Adrien Marie Legendre (1752-1797), esteso da Pierre Simon Laplace (1749-1827) e da Karl Friedrich Gauss (1777-1855). Nello stesso periodo Abraham de Moivre (1667-1754) derivò una approssimazione dei termini della distribuzione binomiale (una prima versione del teorema del limite centrale), Laplace derivò analoghe formule asintotiche. Jakob Bernoulli (1654-1705) formulò la legge debole dei grandi numeri e Thomas Simpson (1710-1761) calcolò l'esatta distribuzione della somma di variabili aleatorie indipendenti ed identicamente distribuite aventi lo stesso supporto.

Nel 1763 fu pubblicato postumo il famoso saggio di Thomas Bayes (1702-1761) con il quale si gettavano le basi di un filone di idee che costituiscono il nucleo di quella che viene chiamata "impostazione bayesiana dell'inferenza statistica".

La statistica del secolo XIX è dominata dalla figura del belga Adolphe Quételet (1796-1874) autore di un gran numero di pubblicazioni e che può essere, a giusta ragione, considerato il fondatore della statistica moderna. Le ricerche di Quételet sui caratteri antropometrici delle popolazioni umane furono continuate, fino ai primi decenni del nostro secolo, da alcuni scienziati inglesi, biologi o genetisti come Francis Galton (1822-1911), Karl Pearson (1857-1936), W.S. Gosset (1876-1937), Ronald Aylmer Fisher (1890-1962). A questi studiosi, interessati alle applicazioni nel campo delle scienze naturali, va riconosciuto il merito di aver fondato il complesso dei metodi detti di "inferenza statistica oggettivistica".

Un passo decisivo per la costruzione della rigorosa autonomia metodologica della statistica si deve a Fisher i cui contributi principali riguardano : la teoria della stima statistica, del campionamento e del disegno degli esperimenti.

Altri studiosi, invece, fra cui, nel nostro paese, va ricordato Corrado Gini (1884-1965) coltivarono principalmente il ramo della statistica descrittiva che mira ad accertare le proprietà dei fenomeni entro determinati limiti spaziali e temporali, vale a dire nell'ambito delle osservazioni effettivamente eseguite.

Nel 1928 Egon Pearson (1895-1980) e Jerzy Neyman (1894-1981) pubblicarono un saggio nel quale per la prima volta si fa esplicita menzione di due tipi di errore (di prima e di seconda specie) cui necessariamente si va incontro quando, sulla base delle osservazioni campionarie, si deve decidere l'appartenenza del campione a una delle due popolazioni (due ipotesi) di possibile provenienza. Attraverso questa impostazione ci si avvia alla costruzione di una teoria oggettivistica dell'induzione. E' questo l'indirizzo di ricerca cui aderisce ancora oggi la maggior parte degli studiosi. Vedremo successivamente la differenza nel concepire il significato del ragionamento induttivo tra Fisher e la scuola di Neyman-Pearson.

La caratteristica saliente di questi sviluppi è sempre quella di giungere ad una impostazione oggettivistica della statistica.

Dopo la seconda guerra mondiale sotto la spinta dei lavori di A. Wald (1902-1950) e di Jimmy Savage (1917-1971) ha avuto inizio un processo di revisione della inferenza statistica classica ed ha cominciato ad affermarsi un indirizzo basato su una concezione soggettiva di probabilità detta "inferenza bayesiana". I principali sostenitori di questo filone di ricerca sono: F.P. Ramsey (1903-1930), Bruno de Finetti (1906-1985), Jimmy Savage e, con sfumature differenti Harold Jeffreys (1891-1989.); tra i principali divulgatori delle idee "neo-bayesiane" segnaliamo Dennis V. Lindley e Morris De Groot (1931-1989).

In questi ultimi decenni l'impiego degli elaboratori elettronici ha aperto nuovi ed interessanti indirizzi di ricerca. Molte tecniche studiate in passato solo dal punto di vista teorico (es: regressione robusta, analisi delle serie storiche, modelli log-lineari, modelli lineari generalizzati, metodi di classificazione, ecc..) vengono utilizzate grazie alla disponibilità dei programmi adeguati. I metodi grafici, inoltre, hanno reso più agevole l'interpretazione di molte tecniche statistiche multivariate ed i programmi statistici per minicomputers hanno permesso l'utilizzazione diffusa di molte tecniche statistiche.

3.STATISTICA DESCRITTIVA.

Nella statistica descrittiva trovano collocazione le metodologie riguardanti la raccolta e sistemazione delle principali informazioni attinenti al fenomeno studiato rappresentato dall'aggregato dei dati a disposizione.

I metodi statistici che sono stati sviluppati in letteratura e che fanno parte della statistica descrittiva si pongono essenzialmente due scopi:

- a) effettuare una analisi dei dati per scoprirne la struttura e le anomalie (**ANALISI ESPLORATIVA**)
- b) operare una sintesi dei dati in modo da far emergere e chiarire le caratteristiche essenziali (**ANALISI DESCRITTIVA**).

3.1.ANALISI ESPLORATIVA.

L'analisi esplorativa dei dati riguarda l'organizzazione, la sintesi e la presentazione dei dati allo scopo di renderli più comprensibili e per meglio scoprirne la struttura sottostante e determinarne le tendenze.

Tali obiettivi sono sempre stati centrali nella ricerca scientifica. Nel periodo 1960-1970 l'analisi esplorativa dei dati ha avuto una rinascita ed il fautore principale di questa rinascita è John Wilder Tukey il cui libro *Exploratory data analysis* (1970, 1977) ha fornito una dettagliata introduzione al problema. Lo sviluppo dell'analisi esplorativa è stato aiutato dalla crescente capacità degli elaboratori elettronici nel calcolo e probabilmente ancora più rilievo ha avuto la possibilità di una efficiente rappresentazione grafica. I calcolatori hanno altresì reso possibile la raccolta e l'organizzazione di una enorme massa di dati.

Generalmente, in letteratura, l'analisi dei dati viene suddivisa in :

1. analisi preliminare o analisi esplorativa dei dati;
2. analisi definitiva.

Mentre l'analisi definitiva è spesso basata su un modello probabilistico e può comprendere problemi riguardanti la stima dei parametri e la prova delle ipotesi (per approfondimenti si rinvia ai paragrafi 4 e seguenti), l'analisi preliminare si pone lo scopo di chiarire la struttura dei dati e di ottenere una semplice descrizione degli stessi.

Seguendo Chatfield(1988) possiamo dire che nell'analisi preliminare dei dati si possono individuare le seguenti fasi:

1. predisposizione dei dati in una forma utile alle analisi successive;

2. verifica della qualità dei dati. E' importante accertare l'esistenza di errori, osservazioni mancanti o altre particolarità;
3. esame di eventuali modifiche da apportare ai dati (ad esempio eliminazione di errori o trasformazione di una o due variabili);
4. calcolo di alcuni indici sintetici quali la media o lo scarto quadratico medio (si veda il paragrafo 4.3.2.)

Durante il procedimento di "esplorazione" dei dati quattro temi vengono affrontati separatamente e spesso combinati tra loro : analisi della resistenza, analisi dei residui, riespressione dei dati e rappresentazione dei dati.

Analisi della resistenza. Il problema è quello di verificare se un cambiamento arbitrario in qualsiasi piccola parte dei dati provoca cambiamento sensibile nelle tecniche di analisi. Ad esempio, nel sintetizzare i dati si può utilizzare sia la media che la mediana, ma mentre la mediana è molto resistente la media è fortemente non resistente. Per tale motivo numerose tecniche di analisi esplorativa dei dati si basano sulla mediana.

E' necessario distinguere tra resistenza e robustezza.

La robustezza implica, generalmente, insensibilità all'allontanamento dalle assunzioni che sorreggono il modello probabilistico sottostante (si veda il paragrafo 5.6).

Analisi dei residui. L'atteggiamento dell'analisi esplorativa dei dati è quello di sostenere che l'analisi di un insieme di dati non sia completa senza un attento esame dei residui. L'enfasi posta sull'analisi dei residui riflette la tendenza delle analisi resistenti a fornire una netta separazione tra il comportamento prevalente ed il comportamento non usuale dei dati. Quando la maggior parte dei dati segue un determinato disegno, quel disegno determina un adattamento resistente. I residui allora tengono conto di ogni drastico allontanamento dal disegno, come pure delle abituali fluttuazioni. Residui non usuali suggeriscono di esaminare le circostanze che sottostanno a quelle osservazioni.

Riespressione dei dati. Spesso una riespressione dei dati osservati tramite l'uso di una diversa scala può aiutare a semplificare l'analisi dei dati stessi. L'analisi esplorativa sottolinea i benefici che si possono avere, nel caso in cui la scala originaria non sia soddisfacente, con un cambiamento di scala. Ciò può aiutare a creare simmetria, variabilità costante, additività degli effetti, dipendenza dalla struttura dei dati. Le trasformazioni usate nell'analisi esplorativa provengono, molto spesso, dalla famiglia di funzioni note come "trasformazioni di potenza" che trasformano y in y^p , unitamente alle "trasformazioni logaritmiche".

Rappresentazione dei dati. Il contributo maggiore all'analisi esplorativa è stato dato dalle nuove tecniche grafiche. E' indubbio infatti che i metodi grafici assolvono un ruolo importante in tutti gli aspetti dell'indagine statistica. Non solo le tecniche grafiche sono di ausilio al ricercatore ma in molti casi sono essenziali. Tukey nel suo libro afferma che " *il risultato più grande di un grafico si raggiunge allorché ci costringe ad osservare quello che non ci saremmo mai aspettati di vedere*".

I grafici esplorativi sono usati come ausilio per diagnosticare le caratteristiche dei dati e per suggerire, in alcuni casi, analisi statistiche appropriate ed anche modelli.

Per una discussione di queste problematiche si può vedere: Chatfield (1985), (1986), McNeil (1977), Erickson e Nosanchuck (1977).

3. 2. ANALISI DESCRITTIVA.

Un aspetto importante dell'analisi descrittiva è quello di operare una sintesi di tutte le rilevazioni eseguite. I dati raccolti sulla popolazione di riferimento vengono sintetizzati attraverso opportuni indici rappresentativi con l'ovvia accortezza che tale riduzione non faccia perdere troppe informazioni rilevanti contenute nei dati originali.

L'insieme degli argomenti che vanno sotto il nome di analisi descrittiva sono i seguenti:

- 1) classificazione dei caratteri statistici, loro distribuzioni e rappresentazioni;
- 2) determinazione di indici sintetici delle principali caratteristiche delle distribuzioni unidimensionali (medie, indici di variabilità, indici di concentrazione, indici di forma);
- 3) determinazione di indici sintetici delle principali caratteristiche delle distribuzioni bidimensionali e multidimensionali (indici di dipendenza, coefficienti di correlazione e simili.).

3.2.1. Classificazione dei caratteri statistici, loro distribuzioni e rappresentazioni.

Il termine "distribuzione " ha significati diversi. Noi faremo riferimento alle distribuzioni di frequenza (distribuzioni statistiche) e alle distribuzioni di probabilità. Nel seguito esamineremo l'accezione statistica poiché meglio si presta a porre in risalto le principali peculiarità del concetto generale di distribuzione che potranno essere poi riferite anche a quello di distribuzione di probabilità. La nozione di distribuzione nasce dall'esigenza di sintetizzare un'ampia massa di dati riferentesi ad un carattere qualitativo (ad esempio: sesso, stato civile, professione, ecc..) o carattere quantitativo (ad esempio, : età, reddito, statura, numero di figli, ecc) inerente una certa collettività (popolazione) i cui elementi, di natura qualsiasi, sono genericamente detti "unità statistiche". L'operazione che dà luogo alla distribuzione statistica è la classificazione. Essa costituisce lo strumento mediante il quale non solo le osservazioni vengono organizzate e compendiate senza perdita di informazioni sulle caratteristiche essenziali del fenomeno in esame ma anche il mezzo più idoneo per le successive elaborazioni.

Uno schema di classificazione, nella sua forma più semplice, consiste in una regola che assegna ogni unità statistica ad una classe di equivalenza: tutti gli oggetti appartenenti ad una stessa classe sono considerati uguali, mentre gli oggetti appartenenti a classi distinte vengono considerati diversi. Ad ognuna delle classi di equivalenza si può associare un numero che deve intendersi come etichetta o nome della classe. Il processo consistente nell'assegnare un numero al fenomeno osservato costituisce la misurazione; i numeri assegnati sono indicati con il nome di "classi di misura". La regola con cui vengono assegnati i numeri in relazione al legame intercorrente tra essi e le vere grandezze del fenomeno studiato, caratterizza il tipo di misurazione e determina il livello di misura.

La classificazione tradizionale prevede quattro livelli di misura : nominale, ordinale, di intervallo e di rapporto.

1. scala **nominale**: i dati vengono semplicemente etichettati e classificati secondo gruppi significativi;

2. scala **ordinale**: le osservazioni sono ancora una volta raggruppate in classi ma in maniera tale da poter stabilire relazioni d'ordine tra di esse;

3. scala **per intervallo**: in questo caso è possibile valutare non solo l'ordine fra le classi ma anche verificare se le differenze tra due valori qualsiasi della scala sono di grandezza nota. In questo tipo di scala il punto zero e l'unità di misura sono arbitrari;

4. scala **di rapporti**: è il più alto livello di misurazione, nel quale il ricercatore può confrontare non solo le differenze tra le misure ma anche le relative grandezze. Lo zero della misura coincide con lo zero della grandezza.

Per sintetizzare l'insieme delle osservazioni effettuate sovente il ricercatore associa ad ognuna delle possibili classi di misura la **frequenza** (assoluta o relativa) ossia il numero delle osservazioni le cui misure appartengono alla classe. L'insieme delle classi di misura e delle loro frequenze costituisce la **distribuzione statistica**.

Qualunque sia la natura delle classi di misura, ossia a livello nominale, ordinale, intervallo o rapporti è sempre possibile costruire una distribuzione statistica purché ogni osservazione cada in una ed una sola classe di misurazione. E' consuetudine tuttavia assegnare nomi differenti alle distribuzioni a seconda del procedimento di misurazione adottato; così si dà il nome di **mutabile statistica sconnessa** a quella in cui le classi di misurazione sono a livello nominale, **mutabile statistica rettilinea** a quella in cui le classi di misurazione sono a livello ordinale, mentre a quella le cui classi sono a livello di intervallo o di rapporto di **variabile statistica**.

La rappresentazione di una distribuzione statistica può effettuarsi nei modi seguenti:

- 1) elencando tutte le coppie (classi di misura, frequenza associata);
- 2) graficamente;
- 3) esprimendo la regola che associa ad ogni classe una frequenza (espressione analitica della distribuzione).

Limitiamoci, per ora, alla considerazione di un solo carattere quantitativo che sarà opportuno distinguere in discreto o continuo a seconda che esso sia suscettibile di assumere solo i valori di una certa scala (numero di figli, numero di addetti, ecc.) oppure un qualunque valore reale di uno o più intervalli (es. pesi, stature, redditi ecc.). Per un carattere discreto la sintesi avviene con la costruzione di una tabella nella quale si fa corrispondere ad ogni possibile valore del carattere il numero degli elementi della collettività che posseggono quel valore (frequenza assoluta). Quando il carattere è continuo risulta necessario considerare non più le singole classi di misura, bensì insieme di classi di misura, generalmente intervalli. Con questo tipo di operazione si ottiene una variabile statistica che procede per classi di intervalli e che è considerata continua.

Nel seguito indicheremo con la lettera N il numero delle unità statistiche della popolazione classificate rispetto ad un generico carattere X , con x_i la generica classe di misura, con $N(x_i)$ la frequenza assoluta. Se k sono le classi di misura avremo:

$$\sum_{i=1}^k N(x_i) = N.$$

Spesso le numerosità vengono sostituite con le frequenze relative ottenute dividendo per N le frequenze assolute, le indicheremo con $p(x_i)$:

$$p(x_i) = \frac{N(x_i)}{N}.$$

Evidentemente si ha:

$$\sum_{i=1}^k p(x_i) = 1.$$

Facciamo osservare che una distribuzione statistica ammette sempre una interpretazione come **distribuzione di probabilità** : se si sceglie a caso una delle N unità statistiche della popolazione, in modo che ogni unità abbia probabilità uguale a $1/N$ di essere estratta, la probabilità che l'unità prescelta appartenga alla classe di misura x_i è :

$$p(x_i) = \frac{N(x_i)}{N}.$$

Strumenti significativi per apprezzare, anche in modo qualitativo, le caratteristiche di una variabile statistica sono le rappresentazioni grafiche. Per le variabili statistiche la rappresentazione è effettuata ricorrendo ad un sistema di assi cartesiani ortogonali. Per quelle discrete si utilizza il diagramma delle frequenze, ottenuto riportando in ascissa le classi di misura ed in ordinata le frequenze corrispondenti (assolute o relative). Per le variabili statistiche che procedono per classi di intervalli, ossia per quelle continue, la rappresentazione grafica è ottenuta mediante l'istogramma. Per la sua costruzione si riportano sull'asse delle ascisse i vari intervalli e su ogni intervallo si costruisce un rettangolo avente area uguale o proporzionale alla frequenza della classe stessa.

Un concetto che, in un certo senso, unifica la nozione di variabile discreta e continua è quello di funzione di ripartizione (f.r.) o funzione cumulativa delle frequenze. Tale funzione, indicata con $F(x)$, è definita nel modo seguente :

$$F(x) = \text{frequenza relativa dei valori } \leq x = \text{Fr}(X \leq x)$$

Essa è cioè una funzione che ad ogni intervallo del tipo $(-\infty, x]$ x reale qualsiasi, assegna la frequenza relativa dei valori appartenenti all'intervallo stesso e, per costruzione, possiede i seguenti requisiti:

- i) è funzione non decrescente di x ;
- ii) è continua da destra per ogni x ;
- iii) $\lim_{x \rightarrow -\infty} F(x) = 0$; $\lim_{x \rightarrow +\infty} F(x) = 1$

Viceversa, ogni funzione definita su tutto l'asse reale a valori nell'intervallo chiuso $[0, 1]$ che soddisfa i precedenti requisiti è la funzione di ripartizione di una qualche variabile statistica.

Nel caso di variabili continue l'istogramma assumerà andamenti differenti a seconda del numero e ampiezze degli intervalli. In generale, come è facile immaginare, un istogramma basato su ampi raggruppamenti mostrerà pochi larghi gradini con forti dislivelli al passaggio da ciascuno a quello successivo. E' presumibile che passando a raggruppamenti più stretti i salti divengano più piccoli ed il profilo dell'istogramma più regolare, così da suggerire di lasciarlo tracciando una curva che rispetti le aree entro ogni suddivisione, ossia può essere descritto con una funzione sufficientemente regolare $y=f(x)$. L'interpretazione, in sintonia con l'istogramma risiede nel fatto che $f(x)dx$ rappresenta la frequenza relativa dei valori dell'intervallo $(x, x+dx)$. Questa funzione è detta densità di frequenza e lo stesso istogramma non è che la rappresentazione di una particolare densità di frequenza costante a tratti. Formalmente la densità di frequenza rappresenta, nei punti in cui esiste, la derivata della corrispondente funzione di ripartizione e quindi la frequenza media che compete all'intervallo $(x, x+h)$ quando $h \rightarrow 0+$.

Viceversa :

$$F(b)-F(a) = \int_a^b f(x)dx = \text{frequenza relativa all'intervallo } (a, b].$$

Le **distribuzioni statistiche multidimensionali** sorgono, come quelle unidimensionali, da un processo di classificazione: pertanto gli N elementi (unità statistiche) vengono classificati considerando più di un carattere. Nel seguito tratteremo quelle bidimensionali.

Nel caso di una distribuzione bidimensionale discreta il modo più semplice per rappresentare il risultato della classificazione consiste nel riportare i dati sotto forma di tabella a doppia entrata. Sulla riga madre vengono riportati i valori osservati di un carattere X ($x_i, i=1, 2, \dots, k$) e sulla colonna madre i valori dell'altro carattere Y ($y_j, j=1, 2, \dots, h$). All'intersezione della riga i .ma con la colonna j .ma sarà disposta la frequenza relativa della coppia $(x_i, y_j) i=1, 2, \dots, k; j=1, 2, \dots, h$:

$$p(x_i, y_j) = \text{Frequenza}(X=x_i, Y=y_j)$$

oppure la numerosità $N(x_i, y_j)$ (frequenze assolute).

Evidentemente si ha :

$$p(x_i, y_j) = \frac{N(x_i, y_j)}{N}.$$

Per caratteri entrambi continui, il risultato a cui si giungerà sarà quello di una corrispondenza tra intervalli bidimensionali (del piano) e relative numerosità (o frequenze relative). Forme miste si avranno nel caso in cui i caratteri considerati non abbiano stessa natura.

La tabella a doppia entrata rappresenta la distribuzione congiunta di (X, Y) in quanto riporta, per ogni possibile determinazione (x, y) di (X, Y) la corrispondente frequenza. Si osservi che la distribuzione congiunta di (X, Y) induce un'unica distribuzione marginale per X ed un'unica distribuzione marginale per Y . D'altro canto, fissate due distribuzioni marginali, esistono molte distribuzioni congiunte che le ammettono come marginali. La distribuzione marginale è la distribuzione che si otterrebbe classificando le unità statistiche solo rispetto all'uno o all'altro carattere. Oltre alle distribuzioni marginali per X ed Y , che sono chiaramente unidimensionali, possiamo ricavare altre distribuzioni unidimensionali da una tabella a doppia entrata, le cosiddette

distribuzioni subordinate o condizionate ($Y \& X=x_i$), ($X \& Y=y_j$). Esse rappresentano, ad esempio nel primo caso, la distribuzione di un carattere Y nelle sole unità statistiche che hanno un valore assegnato dell'altro carattere $X=x_i$.

La f.r. di una distribuzione bidimensionale è definita mediante $F(x, y)$ = frequenza relativa delle coppie con la prima coordinata $\leq x$ e la seconda $\leq y$.

Essa è cioè la funzione che ad ogni intervallo del piano $\{(-\infty, x], (-\infty, y]\}$, con x e y reali qualunque, fa corrispondere la frequenza relativa. Se $F(x, y)$ è la f.r. di una distribuzione bidimensionale, allora possiede i seguenti requisiti:

i) è una funzione non decrescente in ciascuna variabile;

ii) è continua da destra in ciascuna variabile;

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

iv) $\lim_{x \rightarrow +\infty, y \rightarrow +\infty} F(x, y) = 1$

v) $\Delta F = F(x+h, y+k) - F(x+h, y) - F(x, y+k) + F(x, y) \geq 0$ per ogni $h, k > 0$.

Viceversa, ogni funzione di due variabili definita su tutto il piano reale con i requisiti precedenti è la funzione di ripartizione di una qualche distribuzione bidimensionale.

La funzione di ripartizione della marginale X è data da :

$$F_X(x) = \lim_{y \rightarrow +\infty} F(x, y)$$

in modo analogo si ha:

$$F_Y(y) = \lim_{x \rightarrow +\infty} F(x, y)$$

Qualora esista il limite

$$\lim_{h \rightarrow 0, k \rightarrow 0} \frac{\Delta F}{kh} = f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$$

esso rappresenta la densità di frequenza della distribuzione bidimensionale.

Come abbiamo detto in precedenza una distribuzione bidimensionale non risulta caratterizzata dalla coppia delle distribuzioni marginali, cionondimeno in molti importanti problemi connessi allo studio delle relazioni statistiche è utile considerare la classe di tutte le funzioni di ripartizione bidimensionali che posseggono f.r. marginali assegnate F_X, F_Y . Tale classe prende il nome di classe

di Fréchet. Precisamente : assegnate le f.r. marginali si dirà classe di Fréchet relativa a F_X e F_Y

l'insieme di tutte le f.r. bidimensionali $F(x, y)$ tali che le marginali siano proprio quelle assegnate.

Per indicare tale classe utilizzeremo il simbolo $\Gamma(F_X, F_Y)$. Si considerino ora le due particolari

funzioni :

$$F_{XY}^+ = \min(F_X(x), F_Y(y))$$

$$F_{XY}^- = \max(F_X(x) + F_Y(y) - 1, 0)$$

definite a partire dalle sole funzioni marginali. Si dimostra (Hoeffding 1940, Fréchet 1951) che

esse costituiscono due particolari f.r. che appartengono alla classe. Per le funzioni F_{XY}^- e F_{XY}^+ si

ha una interessante disuguaglianza:

$$F_{XY}^-(x, y) \leq F(x, y) \leq F_{XY}^+(x, y)$$

Le funzioni F_{XY}^- e F_{XY}^+ , dette, rispettivamente, elemento minimo ed elemento massimo della classe di Fréchet, hanno, come vedremo successivamente, un particolare ed importante significato. Il loro calcolo è talvolta operazione complessa che però risulta grandemente semplificato dall'uso

delle cosiddette tabelle di cograduazione e contrograduazione di Tommaso Salvemini (1939, 1943) quando le f.r. marginali sono f.r. che procedono per soli salti (variabili statistiche discrete).

Un problema importante per scopi descrittivi è quello della rappresentazione analitica di una distribuzione statistica. Il problema della rappresentazione analitica consiste nel trovare una funzione matematica che possa rappresentare nel modo migliore un fenomeno collettivo. Una classe di funzioni di densità particolarmente interessante per la loro attitudine a descrivere distribuzioni unimodali continue è quella delle funzioni le cui rappresentazioni grafiche prendono il nome di curve di Karl Pearson (1894, 1895). L'idea nasce dalla constatazione che le distribuzioni di frequenza sono caratterizzate sovente da una sola moda (si veda il paragrafo 3.2.2) e dal fatto di tendere a zero gradatamente alle due estremità dell'asse x. Ciò posto se $f(x)$ indica la funzione di densità la classe di funzioni può ottenersi risolvendo la seguente equazione differenziale:

$$\frac{df(x)}{dx} = \frac{(x-a)f(x)}{b_0 + b_1x + b_2x^2}$$

A questa famiglia di funzioni appartengono molte delle distribuzioni interessanti la statistica ed il calcolo delle probabilità ; in particolare quando $b_1 = b_2 = 0$ la densità che si ottiene è detta "**normale**" o "gaussiana". Tale distribuzione ha la seguente funzione di densità:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \text{ con } x \in \mathbb{R}, m \in \mathbb{R}; \sigma > 0$$

Per una completa descrizione di tali curve si veda Elderton e Johnson (1969).

3.2.2. Indici sintetici delle principali caratteristiche delle distribuzioni unidimensionali.

Svariate sono le circostanze in cui è utile o necessario predisporre una o più *misure descrittive* che in qualche modo sintetizzino le caratteristiche essenziali della distribuzione statistica. Ad esempio per determinare il reddito totale di una popolazione non occorre conoscere la forma della distribuzione dei redditi, ma basta il reddito medio ed il numero N degli individui; per conoscere l'accrescimento di una popolazione in un certo numero di anni non occorre la distribuzione dei singoli accrescimenti annui ma basta un accrescimento medio.

La posizione o dislocazione della distribuzione di frequenza rappresenta, per esempio, una caratteristica che può far differire tra loro distribuzioni simili. Risulta quindi importante analizzare la posizione di ogni distribuzione statistica servendosi di alcuni valori sintetici definiti **misure della posizione**.

I parametri moda, mediana e media sono indicatori del centro della distribuzione e pertanto possono essere interpretati come misure che, in qualche modo, informano intorno alla "dislocazione" della distribuzione.

MODA. Per distribuzioni discrete è rappresentato dalla classe di misura a cui è associata la massima frequenza. Se X è una distribuzione con funzione di densità $f(x)$. M_0 è un valore modale della distribuzione se $f(x)$ ha un massimo locale in M_0 . In molte circostanze è importante conoscere se una distribuzione è unimodale, bimodale o multimodale. La proprietà di unimodalità da origine ad altre proprietà, ad esempio, a molte disuguaglianze.

MEDIANA. Data una distribuzione statistica la mediana M_e è un qualunque valore in corrispondenza del quale risulta:

$$\text{Frequenza}(X \leq M_e) \geq 1/2 ; \text{Frequenza}(X \geq M_e) \geq 1/2$$

Sostanzialmente : M_e lascia alla sua sinistra e alla sua destra almeno il 50% della massa unitaria.

Se $F(x)$ non è strettamente crescente, M_e può non essere unica e parleremo di classe mediana. Per una funzione $F(x)$ continua e strettamente crescente, M_e è unica ed è la soluzione di $F(M_e) = 1/2$.

MEDIA. Uno dei significati che si dà alla parola "media" di due o più grandezze è quello, piuttosto vago, che indica un valore compreso tra il valore minimo ed il valore massimo delle grandezze considerate. Questa definizione di media, dovuta a Cauchy, risulta eccessivamente generica: in primo luogo perchè non offre nessun metodo di scelta fra gli infiniti valori possibili, ma soprattutto perchè non allude agli scopi statistici a cui una media deve soddisfare, ed inoltre perchè esistono funzioni da qualificare come medie i cui valori non sono necessariamente intermedi fra quelli noti. Dobbiamo riconoscere all'insigne cultore di geometria Oscar Chisini il merito di aver individuato e formulato il motivo fondamentale della ricerca di una media.

Con Chisini (1929) possiamo dire che la ricerca di una media ha come scopo quello di semplificare una data questione sostituendo in essa a due o più quantità una sola quantità che le sintetizzi, senza alterare la visione d'insieme del fenomeno considerato.

Date n grandezze omogenee x_1, x_2, \dots, x_n , di cui interessa considerare una funzione $y=f(x_1, x_2, \dots, x_n)$, con $x_i \in I \subset \mathbb{R}$ per $i=1, 2, \dots, n$, se la funzione $\varphi(x)=f(x, x, \dots, x)$ è invertibile su I , diciamo che l'unica radice $m_0 \in I$ dell'equazione (nell'incognita m) $\varphi(m)=f(x_1, x_2, \dots, x_n)$ è la MEDIA di x_1, x_2, \dots, x_n , agli effetti della valutazione di f .

Cioè:

$$m_0 = \varphi^{-1}(f(x_1, x_2, x_3, \dots, x_n)).$$

Esempi tipici di medie si ottengono in corrispondenza di particolari forme della funzione invariante $f(x_1, x_2, \dots, x_n)$.

Precisamente:

i) $f(x_1, x_2, \dots, x_n) = \sum_i c_i x_i$ con $c_i \geq 0$ non tutti nulli, dà luogo alla media aritmetica ponderata.

ii) $f(x_1, x_2, \dots, x_n) = \sum_i c_i \frac{1}{x_i}$ con $x_i > 0$ per ogni i , $c_i \geq 0$ non tutti nulli, dà luogo alla media

armonica (ponderata).

iii) $f(x_1, x_2, \dots, x_n) = \prod_i x_i^{c_i}$ con $x_i > 0$ per ogni i , $c_i \geq 0$ non tutti nulli, dà luogo alla media

geometrica (ponderata).

iv) $f(x_1, x_2, \dots, x_n) = \sum_i c_i x_i^k$, $x_i > 0$ per ogni i , $c_i \geq 0$ non tutti nulli, dà luogo alla media di potenza

di ordine k .

Vediamo ora di dare una definizione di media secondo Chisini che possa essere valida per una variabile statistica qualunque, caratterizzata dalla funzione di ripartizione $F(x)$ avente come supporto S (il supporto di F è l'insieme dei punti x tali che $F(x+\epsilon) - F(x-\epsilon) > 0$ per ogni $\epsilon > 0$). Tale proposta è dovuta a B.de Finetti (1931). Indichiamo con $U_a(x)$ la f.r. degenerare su a , così definita:

$U_a(x) = 0$ per $x < a$, $U_a(x) = 1$ per $x \geq a$. Una media verrà individuata in relazione ad una certa

circostanza espressa mediante un numero dipendente da F : $\Phi(F)$.

Data la funzione di ripartizione $y=F(x)$, avente come supporto $S \subseteq R$, se la funzione $z = \psi(t) = \Phi[U_t]$ definita su S , è ivi invertibile, diciamo che l'unica radice m_0 dell'equazione $\psi(m) = \Phi(F)$ è la media di F agli effetti della valutazione di Φ .

Evidentemente $m_0 = \psi^{-1}(\Phi(F))$.

E' interessante analizzare che una media posseda alcune proprietà come la consistenza, l'internalità, la monotonia, l'associatività.

Una media $M(\cdot)$ è detta *consistente (riflessiva)* se $M(U_a) = a$ per ogni $a \in R$. Indicato con S il supporto di F , $M(\cdot)$ è detta *interna* se $M(F)$ è compresa tra gli estremi inferiore e superiore di S , per ogni F . $M(\cdot)$ è detta *monotona* se $M(F) \leq M(G)$ per ogni coppia di f.r. (F, G) tali che $F(x) \geq G(x)$ per ogni reale x . Una media $M(\cdot)$ è *omogenea* se: $M[F(\alpha x)] = (1/\alpha)M[F(x)]$ per ogni F e per ogni α diverso da 0.

Una media è *traslativa* se: $M[F(x-\alpha)] = \alpha + M[F(x)]$ per ogni F e α reale.

E' facile convincersi che ogni media nel senso di Chisini è riflessiva e che, in generale, non gode delle altre proprietà.

Una proprietà che merita un discorso a parte è l'associatività.

Una media $M(\cdot)$ è detta *associativa* se, rispetto a tutte le terne di f.r. (F_1, F_2, G) con $M(F_1) = M(F_2)$, accade che $M[tF_1 + (1-t)G] = M[tF_2 + (1-t)G]$ per ogni $t \in [0, 1]$.

In altri termini, una media associativa resta invariata se si altera una parte della distribuzione senza alterare la media parziale. Per fissare le idee, si pensi ad una collettività composta da una quota t di lavoratori dipendenti e da una quota $(1-t)$ di lavoratori autonomi e sia, ad esempio, F_1 la distribuzione del reddito fra i lavoratori dipendenti e G la distribuzione del reddito fra i lavoratori autonomi. La distribuzione del reddito nella collettività sarà espressa dalla f.r. $H = tF_1 + (1-t)G$. La proprietà di associatività assicura di poter calcolare il reddito medio della collettività sostituendo la distribuzione effettiva F_1 con una distribuzione F_2 avente la stessa media.

Vediamo ora l'espressione analitica di una generica media associativa.

Sia $M(\cdot)$ una media consistente e strettamente monotona sulla classe delle f.r. con supporto dato da $[a, b]$ ($-\infty < a < b < +\infty$). Allora $M(\cdot)$ è associativa se, e soltanto se, esiste una funzione g continua e strettamente crescente, tale che:

$$M(F) = g^{-1} \left[\int_a^b g(x) dF(x) \right].$$

Tale caratterizzazione è stata ottenuta indipendentemente nel 1930 da A.N. Kolmogorov e da M. Nagumo per n grandezze omogenee e poi estesa al caso di distribuzioni da B. de Finetti nel 1931. Tale tipo di media era già stato considerato da Bonferroni (1924) come vera e propria definizione. Tra le medie esprimibili mediante l'espressione data le sole che riescano omogenee sono quelle date dalle espressioni:

$$M(F) = \left\{ \int_a^b x^k dF(x) \right\}^{1/k} \quad k \neq 0$$

$$M(F) = \exp \left(\int_a^b \log x dF(x) \right)$$

ossia le medie di potenze e la media geometrica .

Mentre le sole che riescano traslative sono quelle date dalle espressioni:

$$M(F) = \frac{1}{c} \left(\log \int_a^b \exp(cx) dF(x) \right) \quad c \neq 0$$

$$M(F) = \int_a^b x dF(x)$$

ossia le medie "esponenziali" e la media aritmetica.

Tra le medie di potenze di grande rilievo sono : la m. aritmetica (con $k=1$), quella quadratica (con $k=2$), quella armonica (con $k=-1$).

Si deduce inoltre che l'unica media contemporaneamente omogenea e traslativa è la media aritmetica (Bemporad 1926).

Ricordiamo infine che, per quanto concerne la definizione di media nei casi in cui non è indicata, preventivamente, una circostanza che debba rimanere inalterata dopo il calcolo della media vi è una proposta di Herzog (1961). Tale autore suppone che esista una funzione che misuri l'alterazione apportata al fenomeno considerato dai vari valori che la media potrebbe assumere, tale funzione determina la media, che è appunto quel valore che la rende minima. In tal modo la media aritmetica è il valore di θ per cui risulta minima la funzione

$$g(\theta) = \int (x-\theta)^2 dF(x)$$

e la mediana è la media per cui risulta minima la funzione

$$g(\theta) = \int |x-\theta| dF(x) .$$

Le misure della posizione, precedentemente esaminate, assolvono solo in modo parziale al compito di descrivere sinteticamente le caratteristiche di una distribuzione, in quanto ci informano solo su un aspetto particolare della distribuzione stessa e cioè sulla posizione. Se si vuole descrivere sinteticamente ma più compiutamente una distribuzione statistica mediante pochi valori si devono perciò analizzare anche altre caratteristiche.

L'analisi della variabilità dei dati rappresenta, ad esempio, una fase essenziale nello studio delle caratteristiche di ogni distribuzione.

La variabilità è la proprietà che ha un fenomeno collettivo di manifestarsi con valori o modalità differenti. Seguendo l'indicazione di Corrado Gini (1955) possiamo distinguere lo studio della variabilità in:

a) studio della dispersione ;

b) studio della mutua variabilità.

Si parla di dispersione tipicamente nel caso di misure ripetute di una stessa grandezza che dovrebbero teoricamente coincidere ed invece differiscono a causa di errori di osservazione. In questo caso il problema che interessa è quello di avere un indice che misuri le deviazioni delle quantità rilevate dalla grandezza effettiva del carattere ossia si vuole sapere di quanto le quantità rilevate differiscono da una grandezza di riferimento (polo) solitamente una media.

Si parla di mutua variabilità allorchè le intensità differiscono tra loro perchè sono effettivamente diverse (es: statura, peso, reddito, ecc...). In questo secondo caso si vuole sapere di quanto le grandezze effettive differiscono tra di loro .

Sia F la funzione di ripartizione di una data distribuzione ed A il polo di riferimento rispetto al quale si voglia calcolare la dispersione. Allora una vasta classe di indici di dispersione può essere individuata mediante la seguente espressione:

$$\left\{ \int |x - A|^k dF(x) \right\}^{1/k} ; \quad k > 0$$

Quando A coincide con la media aritmetica m e $k=2$ si ha, dalla precedente formula generale, il noto *scarto quadratico medio* che si indica con σ ed è definito nel modo seguente:

$$\sigma = \left\{ \int |x - m|^2 dF(x) \right\}^{1/2}.$$

σ^2 è la varianza (nel seguito verrà indicata con $\text{Var}(X)$).

Un indicatore di mutua variabilità, la cui portata statistica è stata particolarmente valorizzata dagli statistici italiani, è rappresentato dalla differenza media (con ripetizione):

$$\Delta(F) = \int \int |x - y| dF(x) dF(y) = 2 \int F(x)(1 - F(x)) dx.$$

Strettamente legato alla variabilità è un altro aspetto tipico delle distribuzioni relative a caratteri trasferibili: **la concentrazione**. La teoria classica della concentrazione riguarda caratteri trasferibili, almeno idealmente, da una unità del collettivo statistico ad un'altra; ad esempio: il reddito prodotto da ciascuna unità, la manodopera facente capo ad una impresa e così via. L'analisi della concentrazione ha lo scopo di evidenziare se, in una data situazione concreta, la distribuzione del carattere è più o meno lontana da quella situazione ideale, di equiripartizione, in cui ogni soggetto detiene la stessa quota del carattere. Vediamo ora gli elementi su cui si fonda, tradizionalmente, questa analisi.

Nella formulazione più classica, si parte da un collettivo di n unità e si considerano le singole

grandezze ordinate in modo non decrescente $x_1 \leq x_2 \leq \dots \leq x_n$ nel rispetto del vincolo $\sum_{i=1}^n x_i = nm$ con

n ed m (media) fissati. In questo caso si ha:

$F_i = \frac{i}{n}$ e $Q_i = \frac{1}{nm} S_i$ dove $S_i = \sum_{j=1}^i x_j$. L'insieme delle coppie ordinate $(\frac{i}{n}, \frac{S_i}{nm})$ $i=1, 2, \dots, n$ è la

funzione di concentrazione di Lorenz. Una misura sintetica della concentrazione può costruirsi rapportando la somma

$$C_n = n-1 \left\{ \frac{i}{n} - \frac{S_i}{nm} \right\}$$

al suo valore relativo alla situazione di massima concentrazione: $S_1 = \dots = S_{n-1} = 0$, $S_n = nm$. L'indicatore così ottenuto è il rapporto di concentrazione di Gini:

$$R = \frac{2C_n}{(n-1)} = \frac{\Delta(F)}{2m}$$

Una formulazione più generale valevole anche per variabili continue si basa sui seguenti elementi.

$F(x)$ = frequenza con cui si presenta, nel collettivo, una modalità $\leq x$, con $x \in \mathbb{R}$; per la natura del problema, è naturale assumere che il supporto di $F(x)$ sia un sottoinsieme di $[0, \infty)$; si assume inoltre che esista finita la media aritmetica di $F(x)$;

$Q(x) = \frac{1}{m} \int_0^x t dF(t)$ rappresenta la frazione del carattere complessivo posseduta dalle unità nelle

quali il carattere è presente con modalità $\leq x$. $Q(x)$ ha tutte le proprietà di una funzione di ripartizione. Se il carattere è il reddito, allora $Q(x)$ è la frazione del reddito complessivo nelle mani di detentori di redditi $\leq x$.

La curva di concentrazione di Lorenz è definita parametricamente dalle coppie $(F(x), Q(x))$.

Una definizione alternativa è la seguente

$$L(p) = \frac{1}{m} \int_0^p F^{-1}(z) dz \quad (0 \leq p \leq 1)$$

dove $F^{-1}(z) = \inf\{x: F(x) \geq z\}$.

E' opportuno notare che dall'espressione precedente e dal fatto che $F^{-1}(z)$ è non decrescente discende che $L(p)$ è convessa e la sua derivata è uguale ad uno in $p=F(m)$. Il rapporto di concentrazione di Gini è dato da:

$$R=1-2 \int_0^1 Q(x)dF(x) = 1-2 \int_0^1 L(p)dp .$$

Alcuni aspetti di una distribuzione colpiscono immediatamente colui che si appresta a descriverla. Alludiamo qui alla **forma della distribuzione**. Per non allargare troppo il discorso ci limitiamo alle distribuzioni unimodali. E' proprio la moda e la sua posizione rispetto agli estremi del campo di variabilità una delle caratteristiche di maggior risalto. Se il valore modale è situato vicino all'estremo inferiore parleremo di **asimmetria** positiva mentre nel caso esso sia situato vicino all'estremo superiore parleremo di asimmetria negativa. Dato che per distribuzioni simmetriche è sicuramente soddisfatta la relazione: $M_e = M_o = m$ il confronto tra le misure della posizione può essere utilizzato come misura della simmetria di una distribuzione. La prima di queste misure è stata proposta da K. Pearson (1895) ed è data da:

$$s_k = \frac{(m - M_o)}{\sigma}$$

dove m , M_o e σ indicano rispettivamente la media, la moda e lo scarto quadratico medio. E' chiaro che tale indice è nullo in caso di perfetta simmetria. Dal momento che tale indice può essere impiegato solamente in presenza di curve unimodali, Pearson ritenne opportuno cercare anche un indice alternativo. Avendo constatato che, in presenza di curve simmetriche, i momenti centrali di ordine dispari ($\bar{m}_k = \int (x-m)^k dF(x)$ indica il momento centrale di ordine k) si annullano, propose:

$$\beta_1 = \left(\frac{\bar{m}_3}{\sigma^3} \right)^2$$

Dal momento che β_1 è sempre positivo e, quindi, incapace di cogliere il verso della asimmetria, allo stesso venne preferita una sua variante ottenuta eliminando l'elevamento al quadrato.

Un altro aspetto della forma della distribuzione è dato dalla **curtosi** (grado di appuntimento della distribuzione). La curtosi è una misura di un tipo di allontanamento dalla distribuzione normale. La curtosi è misurata dal rapporto tra il momento centrale di ordine quarto ed il quadrato della varianza, ossia dal seguente indicatore:

$$\beta_2 = \frac{\bar{m}_4}{\sigma^4} .$$

Tale indice è uguale a 3 nella distribuzione normale. Rapporti più grandi di 3 indicano un eccesso di valori in un intorno della media. Rapporti minori di 3 corrispondono a curve più piatte della normale (es. la distribuzione uniforme).

3.2.2. Indici sintetici delle principali caratteristiche delle distribuzioni bidimensionali.

Fra i criteri ed i procedimenti messi a punto dalla metodologia statistica per la ricerca delle relazioni tra variabili unidimensionali e delle relazioni interne tra le componenti delle variabili multidimensionali, particolare importanza assumono quelli che discendono, in vario modo, dalla teoria della **regressione** e della **correlazione** e dalla teoria della **connessione**. Insieme formano la teoria della **dipendenza statistica**.

La teoria della regressione e della correlazione studia la dipendenza nel caso in cui le due variabili siano entrambe espresse in forma quantitativa; la teoria della connessione invece considera lo stesso problema nel caso in cui le grandezze abbiano forma anche qualitativa.

Il concetto cruciale quando si studiano due caratteri simultaneamente è quello di **INDIPENDENZA STATISTICA**.

Sia (X, Y) una distribuzione statistica bidimensionale con f.r. $F_{X, Y}$ dove X ha f.r. F_X e Y ha f.r. F_Y

Diciamo che X e Y sono statisticamente indipendenti quando

$$F_{X, Y}(x, y) = F_X(x) F_Y(y) \quad \text{per ogni } (x, y) \in \mathbb{R}^2$$

Se tale uguaglianza non vale per almeno una coppia, allora X e Y si dicono statisticamente dipendenti. Se (X, Y) è una distribuzione statistica discreta la condizione precedente di indipendenza è equivalente a:

$$p_{X, Y}(x, y) = p_X(x)p_Y(y) \quad \text{per ogni } x, y;$$

mentre se (X, Y) è continua tale condizione diventa

$$f_{X, Y}(x, y) = f_X(x)f_Y(y) \quad \text{per ogni } x, y.$$

La nozione di indipendenza statistica è simmetrica nei due caratteri, nel senso che se Y è indipendente da X allora X è indipendente da Y .

Il termine **connessione** sta ad indicare assenza di indipendenza statistica. Soprattutto se i fenomeni sono qualitativi l'informazione sulla connessione esistente tra i due fenomeni viene sintetizzata per mezzo di indici di connessione basati sulle sole frequenze. Tali indici, non negativi per convenzione, dovranno essere tanto più piccoli quanto più siamo vicini ad una situazione di indipendenza, e tanto più grandi quanto più siamo vicini ad una situazione di dipendenza perfetta. Per consentire agevoli confronti tali indici dovranno essere possibilmente normalizzati fra 0 e 1. Gli indicatori di connessione proposti in letteratura sono molti noi ci soffermeremo brevemente sull'indice proposto da Karl Pearson. Tale indice è espresso in funzione delle contingenze. Si definisce **contingenza** valutata in (x_i, y_j) $i=1, 2, \dots, k; j=1, 2, \dots, h$ la quantità:

$$c(x_i, y_j) = p(x_i, y_j) - p_X(x_i)p_Y(y_j)$$

La contingenza misura la differenza esistente fra le frequenze osservate e quelle teoriche in caso di indipendenza ed è quindi evidente che una media di queste contingenze può fornire una misura di quanto i dati osservati si allontanino dall'ipotesi di indipendenza.

L'indice di connessione di Pearson, denominato χ^2 (si legga "chi-quadrato"), è costruito utilizzando una trasformazione quadratica delle contingenze ed è così definito:

$$\chi^2 = \sum_{i,j=1}^{k,h} \frac{[c(x_i, y_j)]^2}{p_X(x_i)p_Y(y_j)}$$

Per costruzione $\chi^2 \geq 0$. Inoltre $\chi^2 = 0$ se e solo se $c(x_i, y_j) = 0$ per ogni i e j , ossia se e solo se sono statisticamente indipendenti. Inoltre si verifica la seguente disuguaglianza:

$$\chi^2 \leq \min\{(k-1), (h-1)\}$$

e si ha:

a) se $k \leq h$ e $\chi^2 = k-1$ allora esiste perfetta dipendenza funzionale di X da Y e quindi X è funzione di Y .

b) se $h \leq k$ e $\chi^2 = h-1$ allora esiste perfetta dipendenza funzionale di Y da X e quindi Y è funzione di X .

Evidentemente se $k=h$ e $\chi^2 = k-1$ allora la funzione che lega X a Y è biunivoca e quindi anche Y è funzione di X .

Possiamo ora ricavare, normalizzando l'indice χ^2 , un indice relativo di connessione, a valori nell'intervallo $[0, 1]$, nel modo seguente:

$$\chi^{2*} = \frac{\chi^2}{\min\{(k-1), (h-1)\}}$$

Evidentemente $\chi^{2*} = 0$ se e solo se esiste indipendenza statistica tra X e Y e $\chi^{2*} = 1$ se e solo se Y è funzione di X ($h \leq k$), oppure X è funzione di Y ($k \leq h$).

Quando il carattere è quantitativo e dunque la scala di misurazione è numerica l'analisi finora condotta può essere ulteriormente approfondita utilizzando valori sintetici delle distribuzioni condizionate.

In tale situazione lo studio della dipendenza di Y da X (o analogamente di X da Y) può essere condotto da diversi punti di vista :

- 1- guardando, come ha fatto C.Gini (1943), alla dissomiglianza tra le distribuzioni condizionali (o subordinate) e la distribuzione marginale;
- 2- guardando, come ha suggerito V. Castellano (1957), alla dissomiglianza esistente tra le distribuzioni condizionali prese due a due;
- 3- guardando, secondo la proposta di K.Pearson, alla dispersione dei valori della Y intorno alla media delle distribuzioni condizionali oppure ricercando come variano le medie delle distribuzioni parziali al variare di x.

La via più seguita è senza dubbio quella della sostituzione delle distribuzioni condizionate con le medie delle stesse, $m(x) = M\{Y|X=x\}$, che nel loro insieme costituiscono la "**FUNZIONE DI REGRESSIONE**" di Y su X. Alternativamente si pensi di descrivere Y per mezzo di una opportuna funzione di X, $h(X)$. Un criterio per saggiare l'efficacia con cui $h(X)$ descrive Y è fornito dall'errore quadratico medio: $M\{Y-h(X)\}^2 = \zeta$. Tale criterio dice che quanto più piccolo è ζ tanto più accurata è la descrizione di Y realizzata da $h(X)$. Pertanto per una proprietà della media aritmetica discende che la funzione di regressione rappresenta la trasformata di X che meglio descrive Y nel senso da minimizzare l'errore quadratico medio.

Una seconda nozione di indipendenza si basa sulla funzione di regressione ed è quella di **INDIPENDENZA REGRESSIVA O INDIPENDENZA IN MEDIA**. Si dice che Y è regressivamente indipendente da X quando risulta $m(x) = m(Y)$ essendo $m(Y)$ la media aritmetica di Y. Si noti che la relazione di indipendenza regressiva non è simmetrica in quanto, ad esempio, Y può essere regressivamente indipendente da X senza che X sia regressivamente indipendente da Y. Facciamo altresì notare che dall'essere Y regressivamente indipendente da X non segue necessariamente che X e Y siano statisticamente indipendenti. Poichè l'indipendenza statistica implica l'indipendenza regressiva segue logicamente che la dipendenza regressiva di Y da X implica connessione di Y e X. Il grado di dipendenza di Y da X può essere misurato seguendo le direzioni a cui si è fatto cenno in precedenza. In particolare:

- a) ricercando l'intensità della connessione (dipendenza) dei valori medi mediante l'indice quadratico di connessione proposto da C.Gini;
- b) ricercando la dispersione che si verifica nelle distribuzioni subordinate mediante il **rapporto di correlazione** proposto da K.Pearson.

Sebbene il ragionamento che sta alla base della definizione dei due indici sia diverso, essi hanno una stessa espressione. In questa sede presentiamo l'idea di Pearson.

Il rapporto di correlazione di Pearson prende in considerazione le varianze delle distribuzioni subordinate $\text{Var}(Y|X=x)$. E' chiaro che se volessimo fare una previsione del valore della variabile dipendente Y sulla base della conoscenza del valore assunto dalla variabile indipendente X, potremmo adottare la funzione di regressione. E' però altrettanto chiaro che la previsione è tanto migliore quanto più bassa è la variabilità delle distribuzioni subordinate intorno alla funzione di regressione. Se le distribuzioni subordinate non avessero alcuna variabilità la dipendenza di Y da X sarebbe perfetta.

Il rapporto di correlazione di Pearson viene definito:

$$\eta^2_{YX} = 1 - \frac{M\{\text{Var}(Y|X)\}}{\text{Var}(Y)}.$$

Per cogliere meglio il significato di tale indicatore risulta necessario l'uso della formula di scomposizione della varianza. Sia (X, Y) una distribuzione statistica bidimensionale con $\text{Var}(Y)$ esistente, allora vale la seguente:

$$\text{Var}(Y) = \text{Var}(m(X)) + M\{\text{Var}(Y|X)\}.$$

E' immediato verificare che $0 \leq \eta^2_{YX} \leq 1$. Inoltre η^2_{YX} vale 0 se e solo se $\text{Var}(m(X))=0$, equivalentemente $m(x)=M(Y)$. In altre parole, l'indice vale 0 se e solo se la funzione di regressione di Y su X è costante al variare di X. D'altro canto vale 1 se e solo se $M(\text{Var}(Y|X))=0$, equivalentemente se e solo se, per ogni x, Y è degenerare sul valore $m(x)$. In tal caso Y è funzione di X valendo $Y=m(X)$. Di conseguenza il rapporto di correlazione di Pearson passando da 0 a 1 segnala un progressivo allontanamento da una situazione di indipendenza regressiva verso una di dipendenza funzionale.

In generale la funzione di regressione di Y su X non è rappresentabile per mezzo di una semplice espressione analitica. Sorge pertanto l'esigenza di semplificare la struttura della funzione di regressione, sostituendo a $m(x)$ un'altra più semplice funzione $m^*(x)$, la quale sia, in qualche senso, abbastanza prossima a $m(x)$. Tale procedimento comporta la sostituzione dei valori $m(x)$ con i valori $m^*(x)$ e quindi rientra nella categoria dei cosiddetti procedimenti di perequazione. Tuttavia è prassi in statistica contraddistinguere tale procedimento con il termine interpolazione. Una delle espressioni analitiche più semplici è senz'altro quella lineare.

Sia dunque V la classe delle funzioni lineari. L'obiettivo è selezionare, all'interno di V, una funzione $m^*= a + bX$ la quale costituisca la miglior interpolante lineare di una assegnata funzione di regressione. Come criterio di ottimalità impiegheremo l'errore quadratico medio, sceglieremo quella $m^*(x)$ i cui parametri a e b sono tali da rendere minima la quantità: $M\{m(X)-a-bX\}^2$. Tale procedimento di minimo conduce alla "retta di regressione" dei minimi quadrati ed i parametri sono dati da:

$$a = M(Y) - \left[\frac{\text{Cov}(X, Y)}{\text{Var}(X)} \right] M(X)$$

$$b = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

Ottenuta l'interpolante è necessario verificare la "bontà" di questa interpolazione. Si utilizza al riguardo un indice del tutto analogo al rapporto di correlazione di Pearson definito da

$$d^2 = \frac{\text{Var}(m^*(X))}{\text{Var}(Y)} = 1 - \frac{M(Y-m^*(X))^2}{\text{Var}(Y)}$$

detto "**coefficiente di determinazione**". L'interpolazione sarà tanto migliore quanto più alta è la quota della varianza di Y spiegata da $m^*(x)$. E' facile convincersi che $0 \leq d^2 \leq 1$. Inoltre poiché la funzione di regressione minimizza l'errore quadratico medio, vale la disuguaglianza $\eta^2_{YX} \geq d^2$ ed i due indici coincidono se e solo se la funzione di regressione è lineare. I concetti esposti di funzione di regressione e di interpolante lineare possono essere generalizzati al caso riguardante più di due caratteri in cui interessa lo studio della dipendenza di uno di essi (v.dipendente) da più altre (v.indipendenti)

Un altro aspetto dello studio delle relazioni statistiche è quello della correlazione che presenta notevoli punti di contatto con il concetto di regressione.

Francis Galton (1822-1911) è generalmente riconosciuto come il fondatore dell'analisi della correlazione. Sebbene anche altri autori, tra i quali Carl Friedrich Gauss (1777-1855), August Bravais (1811-1863), Francis Edgeworth (1845-1926) trattarono distribuzioni normali multivariate fu Galton a riconoscere la necessità di misurare la correlazione in variabili bidimensionali. Galton nel 1877 introdusse i concetti di regressione e correlazione.

L'analisi della correlazione ha lo scopo di determinare la forza della relazione lineare esistente tra distribuzioni statistiche. Noi accenneremo alla correlazione tra distribuzioni statistiche bidimensionali, coloro che sono interessati alla correlazione tra più variabili statistiche possono far riferimento alle trattazioni che considerano le correlazioni canoniche, analisi fattoriale, regressione multipla, analisi multivariata, serie storiche

Il coefficiente di correlazione lineare tra X e Y è definito da:

$$\rho = \frac{\text{Cov}(X, Y)}{[\text{var}(X)\text{var}(Y)]^{1/2}}$$

dove $\text{Cov}(X, Y) = M[(X-M(X))(Y-M(Y))]$ si dice covarianza.

Se entrambe le distribuzioni sono costanti la correlazione non è definita. Il coefficiente di correlazione lineare, ρ , può anche definirsi a partire dai coefficienti angolari delle rette di regressione dei minimi quadrati di Y su X e di X su Y . Precisamente, indicati con β_{YX} e β_{XY} tali coefficienti si ha:

$$\rho = \pm (\beta_{YX} \beta_{XY})^{1/2}$$

con il segno che coincide con quello della covarianza. Il coefficiente di correlazione soddisfa la disuguaglianza $-1 \leq \rho \leq +1$, risulta uguale a $+1$ o -1 se e solo se X e Y sono legate linearmente; cioè se esistono due costanti a e $b \neq 0$ tali che $Y = a + bX$ con $\rho = -1$ se $b < 0$ e $\rho = +1$ se $b > 0$. Se $\rho = 0$, X ed Y si dicono non correlate. Se X e Y sono indipendenti statisticamente allora $\rho = 0$, ma non vale necessariamente il viceversa. L'aggiunta di costanti a X e Y non altera il valore di ρ . Si può dimostrare che $\rho^2 = d^2$ dove d^2 è il coefficiente di determinazione precedentemente definito.

L'analisi della correlazione lineare è un aspetto particolare dell'analisi della **concordanza** tra due variabili. Esaminare la concordanza tra due caratteri, significa appurare se è più o meno forte la loro tendenza ad associarsi, nel collettivo statistico considerato, in modo che all'aumentare dell'uno, aumenti anche l'altro. Evidentemente, un esame del tipo indicato è sovente stimolato dalla necessità di conoscere, a fini previsivi, come varia un fenomeno in relazione ad un altro. Ciò che spesso viene sottovalutato è il seguente vincolo implicato nell'analisi della concordanza: lo studio della distribuzione congiunta dei due caratteri è fatto sotto l'ipotesi che restino ferme le distribuzioni marginali. Questo fatto è rilevante perchè la possibilità dei due caratteri di associarsi in modo perfettamente concordante dipende dalla forma di tali distribuzioni e, pertanto la nozione di perfetta concordanza è, necessariamente, relativa alle distribuzioni marginali. Quale allora, in simili circostanze, la situazione di massima concordanza compatibile con le marginali assegnate?. Lo strumento più idoneo per rispondere a questo quesito è rappresentato dalla nozione di classe di Fréchet (si veda il paragrafo 3.2.1). La motivazione che la classe di Fréchet è il punto di partenza per lo studio della concordanza può trovarsi in Salvemini (1939, 1943) e Dall'Aglio (1956).

Sulla classe di Fréchet $\Gamma(F_X, F_Y)$ relativa alle due f.r. marginali F_X, F_Y si può definire un ordinamento parziale.

Siano F_{XY}, G_{XY} due f.r. bidimensionali appartenenti a $\Gamma(F_X, F_Y)$. Se

$$F_{XY}(x, y) \leq G_{XY}(x, y) \text{ per ogni } x, y, R$$

e la disuguaglianza vale in senso forte per almeno una coppia (x, y) allora si dirà che F_{XY} è meno concordante di G_{XY} .

Per cogliere, in modo sintetico, il comportamento concorde delle due distribuzioni X ed Y possiamo definire sulla classe Γ una funzione a valori reali, $C(F_{XY})$, in modo tale che essa rispetti

l'ordinamento parziale definito sulla classe.

La funzione $C(F_{XY})$ è detta indice assoluto di concordanza se dall'essere F_{XY} meno concordante di G_{XY} discende

$$C(F_{XY}) \leq C(G_{XY}).$$

$C^*(F_{XY})$ è detto indice relativo di concordanza quando, oltre alla precedente proprietà di monotonia, soddisfa la seguente doppia disuguaglianza: :

$$-1 \leq C^*(F_{XY}) \leq +1$$

essendo i valori -1 e +1 raggiunti rispettivamente quando $F_{XY} = F_{XY}^-$ e $F_{XY} = F_{XY}^+$. Un indice relativo di concordanza è dato dall'indice quadratico di OMOFILIA proposto da C.Gini. Sia (X, Y) una distribuzione statistica bidimensionale con f.r. F_{XY} . Si chiama indice quadratico di omofilia la quantità:

$$O(F_{XY}) = \frac{\text{Cov}(X, Y)}{\text{Cov}_{F+}(X, Y)} \quad \text{se } \text{Cov} \geq 0$$

$$O(F_{XY}) = - \frac{\text{Cov}(X, Y)}{\text{Cov}_{F-}(X, Y)} \quad \text{se } \text{Cov} \leq 0$$

dove Cov_{F+} e Cov_{F-} rappresentano la covarianza tra X ed Y rispettivamente calcolata in corrispondenza dell'elemento minimo e massimo della classe di Fréchet.

3.2.3 Parametri descrittivi della distribuzione ed ordinamenti parziali.

Il procedimento che abbiamo seguito per presentare, nel paragrafo precedente, l'indice di omofilia (vale a dire definendo un ordinamento parziale delle funzioni di ripartizione e ricercando un indice coerente con tale ordinamento) rientra in una proposta più generale per definire indici statistici. Pertanto la richiesta è che le misure utilizzate per valutare quantitativamente le proprietà delle distribuzioni siano coerenti con qualche ordinamento parziale corrispondente alla stessa proprietà. Ne sono un esempio le classiche misure di posizione, dispersione, simmetria, curtosi e le misure di dipendenza.

L'idea di ordinare le distribuzioni rispetto ad una determinata proprietà è dovuta a Mann e Whitney (1947) che introdussero in letteratura la nozione di ordinamento stocastico e di ordinamento per la posizione. Nel 1948 Birnbaum propose un ordinamento di dispersione. Gli ordinamenti di simmetria e curtosi sono stati proposti da Van Zwet (1964) e l'ordinamento star-shaped da Barlow e Proschan (1966). In quattro fondamentali articoli Bickel e Lehmann (1975, 1976, 1979) discussero in maniera esauriente i problemi inerenti le misure di posizione e di variabilità. Tali impostazioni sono state unificate, nel caso univariato, da Oja (1981).

Presentiamo ora l'idea di Bickel e Lehmann per definire sia le misure della posizione sia le misure della variabilità basandosi sul concetto di ordinamento stocastico.

Si dice che la v.a. X è dominata da Y ($X <_{st} Y$), se :

$$F_X(x) \geq F_Y(x) \quad \text{per ogni } x .$$

e la disuguaglianza vale in senso forte per almeno un x .

Seguendo l'impostazione di Bickel e Lehmann (1975) possiamo dire che un parametro di posizione $\mu(F_X)$ o $\mu(X)$ definito sopra una opportuna classe di distribuzioni dovrebbe soddisfare le seguenti condizioni:

a) $X <_{st} Y \Rightarrow \mu(X) \leq \mu(Y)$ (coerenza con l'ordinamento di dominanza stocastica)

b) $\mu(aX+b) = a\mu(X)+b$

c) $\mu(-X) = -\mu(X)$.

Le tre proprietà precedenti sono indipendenti. Molte delle misure della posizione (si veda il paragrafo 3.2.2) soddisfano le condizioni a) b) c) (ad es: la media, la mediana).

In analogia con quanto fatto per la definizione di una misura di posizione si può definire una misura della dispersione come un funzionale che soddisfa certe condizioni di invarianza e che in aggiunta possieda la proprietà di assumere un valore più grande per F_X che per F_Y se F_X è più dispersa di F_Y . Per distribuzioni simmetriche, la dispersione è usualmente intesa come la vicinanza della distribuzione al centro di simmetria μ .

Seguendo Bickel e Lehmann (1976) possiamo dire che F_Y è più dispersa attorno ξ che F_X attorno a μ se :

$$|X - \mu| \leq_{st} |Y - \xi.$$

Per distribuzioni asimmetriche, F_Y è detta più diffusa che F_X se (si veda Bickel e Lehmann (1979)) :

$$F_X^{-1}(v) - F_X^{-1}(u) \leq F_Y^{-1}(v) - F_Y^{-1}(u)$$

per ogni $0 < u < v < 1$.

Una misura della dispersione $\sigma(X)$ è allora definita in modo tale che soddisfi le seguenti proprietà:

a) $\sigma(X) \leq \sigma(Y)$ quando F_Y è più dispersa che F_X (coerenza con l'ordinamento di dispersione)

b) $\sigma(aX+b) = |a|\sigma(X)$ per ogni a e b .

Le classiche misure della dispersione soddisfano a queste proprietà.

Anche per gli indici di simmetria e curtosi si può procedere in maniera analoga (Si veda van Zwet 1964).

Analogo discorso si può fare per le misure della dipendenza.

L'intuitivo concetto di dipendenza positiva (negativa) di una v.s. Y su un'altra v.s. X è che a valori grandi di Y corrispondono valori grandi di X . (nella letteratura statistica italiana lo studio della dipendenza positiva equivale allo studio della concordanza)

Sono molte le famiglie di distribuzioni bidimensionali con una naturale interpretazione di dipendenza con segno. Le più note sono quelle in regressione, associazione e dipendenza in quadrante. Le definizioni di queste famiglie si trovano in Lehmann (1966), Esary, Proschan e Walkup (1967) e Esary e Proschan (1972). Una discussione dettagliata di tali concetti è stata fatta da Yanagimoto (1973). In questo contesto ci sembra utile ricordare solamente la dipendenza in quadrante.

Due distribuzioni si dicono dipendenti in quadrante positivamente (negativamente) se:

$$F_{XY}(x, y) \geq (\leq) F_X(x)F_Y(y) \quad \text{per ogni } (x, y) \in \mathbb{R}^2$$

La precedente può anche essere riformulata nel modo seguente :

$$(X|Y>y) \geq_{st} X \quad \text{per ogni } y \in \mathbb{R}$$

dove $(X|Y>y)$ indica la distribuzione condizionata di X dato $Y>y$.

Sia ora Q^+ (Q^-) la famiglia di tutte le distribuzioni bidimensionali con dipendenza in quadrante positiva (negativa).

Possiamo allora ordinare la famiglia Q^+ (Q^-) in modo tale che per ogni (X, Y) e (X', Y') appartenenti a Q^+ (Q^-) sia

$$(X, Y) \leq_{Q^+} (X', Y') (\leq_{Q^-})$$

se

$$i) F_X = F_{X'}, F_Y = F_{Y'}$$

$$ii) (Y|X>x) \leq_{st} (Y'|X'>x') (\geq_{st}), \text{ per ogni } x \in \mathbb{R}$$

Un indice di dipendenza coerente con l'ordinamento testè introdotto è dato dall'indice di omofilia del paragrafo precedente.

3.4. TECNICHE PER L' ANALISI DI DATI MULTIVARIATI.

In questo paragrafo intendiamo segnalare un insieme di tecniche di analisi statistica multivariata che formano un corpo unitario nell'impostazione più generale che possiamo denominare "Analyse des données" facendo appunto riferimento al termine usato dalla scuola statistica francese che a partire dagli anni sessanta, sotto l'influenza di Jean-Paul Benzecri, ha sviluppato tale tipo di analisi. Per una trattazione sistematica si veda Benzecri (1973) e Cailliez e Pages (1976).

Lo scopo di queste tecniche è quello di analizzare dati multivariati per far scaturire ipotesi (piuttosto che provarle) ed aiutare a scoprire la struttura sia tra le variabili che tra gli individui. In particolare esse possono essere utili per ridurre la dimensione del problema statistico e nel fornire grafici dei dati. Nell'ambito di tale indirizzo statistico, le tecniche in questione fanno parte delle "Analisi lineari" da una parte e dei "metodi di classificazione" dall'altra.

Sia \mathbf{X} una variabile aleatoria p -dimensionale con vettore delle medie $\boldsymbol{\mu}$, dove $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ e $\boldsymbol{\mu}' = [E(X_1), E(X_2), \dots, E(X_p)]$. La matrice ($p \times p$) di varianze e covarianze di \mathbf{X} è data da:

$$\Sigma = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})']$$

dove l'elemento (i, j) di Σ è la covarianza tra X_i e X_j se $i \neq j$ ed è la varianza di X_i se $i = j$, ed E indica il valor medio (speranza matematica).

La matrice di correlazione di \mathbf{X} , indicata con P , è tale che l'elemento (i, j) misura la correlazione tra X_i e X_j , allora la diagonale di P è formata da tutti uno.

Supponiamo di avere n osservazioni sulla variabile p -dimensionale. Indichiamo la matrice dei dati con X . Il vettore delle medie campionarie è \bar{x} . La matrice di varianze e covarianze campionaria S

può essere calcolata in maniera ovvia: $S = \frac{(X - \mathbf{1}\bar{x}')(X - \mathbf{1}\bar{x}')}{n-1}$ dove $\mathbf{1}$ indica un vettore ($n \times 1$) di uno.

Dalla S discende immediatamente la matrice di correlazione campionaria R .

ANALISI DELLE COMPONENTI PRINCIPALI. Questo metodo, proposto agli inizi del secolo da Karl Pearson (1901), è stato completamente sviluppato da Harold Hotelling negli anni '30.

L'idea è quella di trasformare le p variabili osservate in p nuove variabili, ortogonali fra loro, dette **componenti principali**, che sono una combinazione lineare delle variabili originarie ($\mathbf{a}'\mathbf{X} = \sum a_i X_i$) e che sono scelte in modo da spiegare quanta più variabilità sia possibile. Allora la prima componente, $\mathbf{a}_1'\mathbf{X}$, è scelta con la varianza massima, con il vincolo $\mathbf{a}_1'\mathbf{a}_1 = 1$. Matematicamente \mathbf{a}_1 è l'autovettore della matrice delle varianze e covarianze di \mathbf{X} (o alla matrice di correlazione) che corrisponde all'autovalore più grande.

Molto spesso i coefficienti delle diverse componenti principali sono gli autovettori di S (o di R) e la varianza di ciascuna componente è data dall'autovalore corrispondente. È noto che la somma degli autovalori di una matrice quadrata, reale e semi definita positiva è uguale alla somma degli elementi sulla diagonale (detta traccia). Allora per una matrice di correlazione, i cui elementi sulla diagonale sono tutti uguali ad 1, la traccia è p così che la proporzione tra varianza totale spiegata dalla prima componente è $\frac{\lambda_1}{p}$, dove λ_1 è il più grande autovalore di R . Si trova sovente che le prime due o le

prime tre componenti "spiegano" la maggior parte della variabilità dei dati originali, in tal modo l'effettiva dimensione del problema statistico in esame è molto più piccola di p . In particolare un diagramma delle prime due componenti è molto spesso utile per determinare gruppi di individui o dati anomali (outliers). Notiamo che se le X sono linearmente dipendenti (e.g. diciamo $X_3 = X_1 + X_2$) allora S (o R) sarà singolare e semi definita positiva e pertanto uno o più autovalori saranno nulli.

ANALISI DEI FATTORI. Delineata nel primo ventennio del nostro secolo nell'ambito della ricerca psicometrica da C. Spearman, al quale è generalmente attribuita la paternità, K. Holzinger, C. Burt, G.H. Thomson, e poi sviluppata da L.L. Thurstone, l'analisi dei fattori ha uno scopo simile a quello dell'analisi delle componenti principali vale a dire ridurre la dimensione del problema. In tale metodologia la quantità di riferimento non è la varianza ma la covarianza (o correlazione) che, a differenza dell'altra, si presta, anche sul piano teorico, ad essere interamente "spiegata" in termini ridotti, postulando un particolare modello. L'analisi è basata su un proprio modello statistico che considera m ($< p$) fattori latenti non osservabili, $\{f_i\}$, tali che

$$X_j = \sum_i \lambda_{ij} f_i + e_j.$$

I pesi sono detti coefficienti di saturazione (factor loadings). Le $\{f_j\}$ sono detti fattori comuni e le $\{e_j\}$ fattori specifici. La quota di variabilità di X_j spiegata dai fattori comuni è detta "comunalità" della j -esima variabile. Evidentemente non diamo i dettagli dell'analisi ma facciamo osservare che molti programmi di computer permettono di ruotare i fattori per renderne più facile l'interpretazione. Osserviamo infine che l'analisi dei fattori è spesso confusa con l'analisi delle componenti principali, particolarmente quando quest'ultima è utilizzata per fornire i valori di partenza nel modello di analisi dei fattori in costruzione.

GRADUAZIONE MULTIDIMENSIONALE (MULTIDIMENSIONAL SCALING) Tale metodologia riguarda le relazioni tra unità statistiche. L'idea è quella di costruire una "mappa" generalmente a due dimensioni, di un insieme di individui data una misura di somiglianza o di dissomiglianza tra ogni coppia di unità statistiche. Lo scaling classico è appropriato quando le dissimilarità sono approssimativamente delle distanze Euclidee. Allora vi è una dualità con le componenti principali poiché lo scaling classico è essenzialmente un autovettore di XX' mentre la componente principale è un autovettore di $X'X$. Lo scaling ordinale, detto anche non metrico, usa solamente le proprietà ordinali delle dissomiglianze ed impiega una procedura numerica iterativa.

CLUSTER ANALYSIS Si propone di partizionare un collettivo di unità statistiche in gruppi o cluster che sono in un certo senso vicini tra di loro. Vi sono una vasta gamma di procedimenti che dipendono da differenti criteri, o differenti algoritmi numerici o da differenti obiettivi. Ad esempio, alcuni metodi assegnano le unità statistiche ad un fissato numero di gruppi, mentre altri lasciano che il numero dei gruppi sia determinato dai dati. Altre procedure si prefiggono di determinare una completa struttura gerarchica dei dati.

ANALISI DELLE CORRISPONDENZE E' la tecnica più interessante ed originale dell'"Analyse des données" e nella sua forma più completa è stata proposta da Benzecri (1973). Essa tratta di tabelle di dati a due o più dimensioni, con il vincolo che tali dati siano espressi in forma di numeri interi non negativi per i quali abbia senso la somma per riga, colonna, strato, ecc. E' ovvio che l'applicazione più immediata ed usuale è quella concernente l'analisi di tabelle di contingenza bi- o pluri-dimensionali, nelle quali i dati numerici oggetto di analisi sono le frequenze iscritte nelle varie caselle, relative alle combinazioni di modalità di due o più caratteri statistici.

Data una matrice $(n \times p)$ non negativa di dati, si determinano le somme per riga e per colonna e sia R la matrice diagonale $(n \times n)$ delle somme per riga e C la matrice diagonale $(p \times p)$ delle somme per colonna. Allora $R^{-1}X$ e $C^{-1}X$ sono le matrici dei profili-riga e dei profili-colonna dove dividiamo semplicemente ogni osservazione mediante l'appropriata somma della riga o colonna. L'analisi consiste essenzialmente nello studio di un autovettore di $(R^{-1}X)(C^{-1}X')$ o $(C^{-1}X')(R^{-1}X)$, qualsiasi sia la matrice più piccola.

CORRELAZIONI CANONICHE La teoria delle correlazioni canoniche è stata introdotta in letteratura da Hotelling (1936) ed è una metodologia molto utile nell'analisi esplorativa di dati multivariati. L'analisi delle correlazioni canoniche riguarda la partizione di un insieme di variabili in due sottoinsiemi, un insieme x ed un insieme y . Lo scopo è quello di determinare le combinazioni lineari $v = a'x$ e $\phi = b'y$ in modo da avere tra v e ϕ la massima correlazione possibile. Tali combinazioni lineari possono fornire utili informazioni sulle relazioni esistenti tra due insiemi di variabili. L'analisi delle correlazioni canoniche pur presentando notevoli punti di contatto con l'analisi delle componenti principali, si differenzia da questa in quanto risponde alla seguente domanda: fino a che punto uno dei due gruppi di variabili spiega il comportamento dell'altro?

ANALISI DISCRIMINANTE. Nelle situazioni più disparate ed in una vastissima gamma di campi il ricercatore (e, magari incoscientemente, anche l'uomo comune) si trova a dover classificare una unità statistica, caratterizzata da un insieme di misure ma di provenienza non nota, in una di diverse popolazioni date. Il primo a parlare di analisi discriminante fu R.A. Fisher nel 1936; ma poiché si tratta in definitiva di precisare una regola o criterio di assegnazione, la cui formulazione dipende

dall'ottica prescelta, si svilupparono ben presto metodi alternativi. I metodi di analisi discriminante tendono a risolvere i due seguenti problemi:

- a) individuare criteri statistici adeguati a stabilire a quale o più delle popolazioni date si deve assegnare una unità statistica sulla base delle caratteristiche rilevate ;
- b) valutare il grado di discriminazione, cioè calcolare la probabilità di classificazioni errate in base alle quali si attribuisce ad una popolazione un individuo che invece dovrebbe essere classificato in un'altra o viceversa.

Ad esempio se le popolazioni indicano diverse malattie e x_1, x_2, \dots, x_n le misure dei sintomi di un paziente, il problema consiste nel diagnosticare la malattia del paziente sulla base dei sintomi.

Segnaliamo infine alcuni testi introduttivi utili all'approfondimento delle problematiche trattate in questo paragrafo: Mardia, Kent e Bibby (1979), Chatfield e Collins (1980), Greenacre (1986).

4. IL PROBLEMA DELL'INDUZIONE IN STATISTICA.

4.1. Probabilità ed induzione.

"L'induzione è il procedimento che dai particolari porta all'universale" questa definizione di Aristotele ha trovato concordi tutti i filosofi. Aristotele stesso vede nell'induzione una delle due vie attraverso le quali riusciamo a formare le nostre credenze; l'altra è la deduzione. Doveva arrivare Hume (*A Treatise of Human Nature* 1739) a risvegliare filosofi e scienziati dal "sonno dogmatico", a porre per la prima volta, in pieno settecento, il problema logico della conoscenza. La storia del pensiero da Hume in poi è la storia di una eterna disputa intorno al significato del conoscere induttivo. Questa disputa, tuttavia, non ha certo interrotto l'avanzamento prodigioso che vi è stato in tutte le scienze empiriche. Il problema filosofico dell'induzione è il seguente: siamo razionalmente giustificati nel ragionare da esempi ripetuti di cui abbiamo avuto esperienza ad esempi di cui non abbiamo avuto esperienza? La risposta di Hume è seccamente negativa. Nè, dopo Hume, chi ha raccolto la sua sfida è riuscito a fare molto meglio. La ragione di questa sequenza di sconfitte consiste essenzialmente nel fatto che la sfida di Hume non può essere vinta sul terreno proposto da Hume. Chiediamoci allora perchè mai ci serviamo del ragionamento induttivo qualunque ne sia la caratterizzazione migliore. Lo stesso Hume riconosce che il ruolo principale del ragionamento induttivo è quello di costituire "*la guida nella vita*" e cioè di orientarci nelle nostre decisioni pratiche. Pertanto il problema della scienza è lo stesso del vivere comune: imparare dall'esperienza. Chiediamoci allora: con il ragionamento deduttivo (facendo discendere il particolare dal generale) riusciamo ad ampliare il campo delle nostre conoscenze?

Evidentemente no. Quello che permette di andare oltre o di accrescere la nostra conoscenza è il ragionamento induttivo.

In questa sede non ci addentriamo nell'analizzare le varietà di atteggiamenti che si sono via via succeduti nel pensiero scientifico o il dibattito epistemologico su questo tema.

La preoccupazione di tutti è comunque quella di "imparare dall'esperienza" ma non si è capaci di trovare un accordo sul modo.

Già nell'"*Ars Conjectandi*" (1773) di Bernoulli, nel saggio di Bayes (1763) "*Essay towards solving a problem in the doctrine of chances*" e nell'"*Essay philosophique sur les probabilités*" (1814) di Laplace, probabilità ed induzione appaiono quasi categorie inscindibili.

Il problema allora da esaminare è quello di verificare se la teoria della probabilità possa svolgere un ruolo unificante nel ragionamento induttivo. E' chiaro, infatti, che se per ragionamento intendiamo qualcosa di fondato sulla logica allora il "ragionamento induttivo" lo è solo nell'ambito della logica probabilistica (che costituisce la logica dell'incerto), cioè della teoria della probabilità.

Ancora oggi si assiste ad un acceso dibattito sui fondamenti del ragionamento induttivo ed ad una divisione nell'impostazione e nella risoluzione di questioni più propriamente tecniche. Analoghe divisioni si ripetono nel campo della teoria della probabilità; anzi il dibattito a cui assistiamo in statistica è principalmente originato dalle diverse concezioni della probabilità. Si può dire, estremizzando, che ad ogni idea di probabilità corrisponda una proposta induttiva. La causa

nascosta di molte discussioni sull'inferenza statistica è proprio rappresentata dal mancato riconoscimento di tale diversità di atteggiamenti sulla probabilità.

Per quanto riguarda l'aspetto più propriamente matematico, le idee sulla probabilità non sono fortemente discordanti (l'unico punto in discussione è se le funzioni debbano essere finitamente additive o completamente additive). Il nucleo della discordia concerne invece il significato della nozione di probabilità e può essere sintetizzato nel ricordare che da un lato vi sono concezioni secondo cui, per elevare a rango scientifico la probabilità, si deve depurarla di tutti gli elementi soggettivi che la riguardano e la caratterizzano; dall'altro vi sono quelle concezioni che, non solo, non ritengono tali elementi come causa di disturbo, ma ne fanno il punto di partenza per la definizione e la conseguente teoria matematica. Dal punto di vista filosofico, le concezioni del primo tipo (OGGETTIVE) ritengono la probabilità come elemento del mondo fisico, esistente fuori di ciascuno di noi; le concezioni del secondo tipo (SOGGETTIVE) ritengono che la probabilità, esprimendo in ogni caso una opinione personale, venga ad esistenza solo nel momento in cui, il giudizio del soggetto viene a contatto di date circostanze concrete, incerte. E' così che la probabilità acquista significato solo in funzione di un dato individuo in specificate situazioni.

Al dibattito sui fondamenti della probabilità hanno contribuito moltissimi studiosi tra cui ricordiamo: R.von Mises, J.M.Keynes, F.P.Ramsey, B.de Finetti e R. Carnap. Altro filone di studio riguarda l'enunciazione in forma assiomatica delle proprietà sostanziali da stabilire come punto di partenza per una teoria formale della probabilità, in quest'ambito segnaliamo le ricerche di F.P.Cantelli, A.Rényi e, soprattutto di A.N.Kolmogorov.

L'inferenza statistica classica fa riferimento alla concezione di probabilità con un significato oggettivo e quindi indissolubilmente ancorata alla nozione di frequenza, mentre l'inferenza bayesiana trova un valido e rigoroso riferimento principalmente nella concezione soggettiva della probabilità.

Soffermiamoci allora brevemente sulle definizioni di probabilità frequentista e soggettivista facendo riferimento rispettivamente a von Mises ed a de Finetti.

La definizione di von Mises passa attraverso l'introduzione di due proprietà a cui devono soddisfare le successioni di eventi oggetto di studio designate con il termine "*collettivo*":

1. le frequenze relative con cui ciascuno dei possibili risultati si presenta in un numero finito di prove, tendono a dei valori limite, detti probabilità o frequenza limite, al tendere del numero delle prove all'infinito;
2. tali frequenze limite rimangono inalterate se, anziché la successione originale se ne considera una estratta da essa in un modo qualunque ma ben precisato.

Dall'impiego di questi due assiomi discende che la probabilità è una funzione semplicemente additiva sugli esiti delle prove, a valori reali in $[0, 1]$, essendo unitaria la probabilità dell'evento certo.

Secondo i soggettivisti la probabilità rappresenta una misura del grado di plausibilità che un individuo assegna ad un dato evento (incerto). Nell'impostazione di de Finetti, quando si parla di evento si intende un fatto singolo. La definizione numerica del grado di plausibilità assegnato da un individuo ad un evento ben determinato avviene mediante lo schema della scommessa che svolge nella teoria il ruolo di strumento di misura perchè in ogni occasione è possibile valutare la probabilità di un evento riferendosi ad una scommessa ipotetica che lo coinvolge e la definizione che ne discende è operativa.

Dato un evento A si dice probabilità di A, $P(A)=p$, il prezzo p che un individuo è disposto a pagare (ricevere) per riscuotere (pagare) un importo monetario unitario se A si verifica e nulla se A non si verifica.

L'alternativa della precedente definizione sta a significare che la valutazione della probabilità non deve cambiare se l'individuo scambia il suo ruolo da scommettitore a banco. Ma tale valutazione, oltre alla precedente restrizione, deve anche soddisfare al principio di COERENZA secondo il quale la scommessa non deve dar luogo a guadagni (perdite) certi.

Dopo aver esposto sommariamente due delle impostazioni che partono dal significato di probabilità presentiamo brevemente l'impostazione assiomatica dovuta a Kolmogorov (1933). La convinzione di fondo di Kolmogorov si regge sul fatto che la teoria della probabilità può essere sviluppata

assiomaticamente prescindendo dalle questioni inerenti il significato di probabilità. Kolmogorov mostrò che la teoria della probabilità si occupa di entità astratte che non necessitano, almeno a livello di sviluppo della teoria, di alcuna interpretazione. Il punto di partenza è la considerazione di un insieme di elementi, detto spazio dei risultati, indicato con Z , e di una classe di sottoinsiemi (eventi) di Z . Quando l'insieme dei risultati Z è finito oppure infinito ma numerabile ogni suo sottoinsieme può considerarsi evento. Se però Z è infinito non numerabile sorgono delle difficoltà dovute al fatto che non tutti i suoi sottoinsiemi possono considerarsi eventi. Allora, in vista dell'assegnazione della probabilità, dobbiamo restringere la classe di sottoinsiemi di Z che devono essere considerati eventi. Tale nuova classe, indicata con A , deve tuttavia essere sufficientemente ampia da permettere di operare con le operazioni di unione, intersezione e complemento sui suoi elementi senza uscire dalla classe stessa. Una tale famiglia di sottoinsiemi di Z è detta σ -algebra. Comunque sia fissata la classe A , i suoi elementi sono detti insiemi misurabili ed è solo a questi che sarà possibile attribuire la qualifica di eventi ed assegnare una probabilità. La coppia (Z, A) si dice spazio misurabile.

La definizione di probabilità P viene fissata mediante i seguenti assiomi:

- 1) ad ogni elemento $E \in A$ si associa un numero reale $P(E) \geq 0$
- 2) per l'evento Z si ha: $P(Z) = 1$
- 3) per ogni successione $E_1, E_2, \dots, \in A$ a due a due incompatibili si ha:

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

4.2. Le fasi dello sviluppo dell'inferenza statistica.

Da un punto di vista storico si possono individuare tre fasi nello sviluppo dell'inferenza statistica:

- a) in una prima fase vi fu una rapida affermazione della teoria bayesiana, molta leggerezza nell'applicazione e conseguente discredito ed abbandono di tale concezione;
- b) in una seconda fase vi fu la ricerca di altre vie per affrontare ed impostare i problemi dell'inferenza statistica (sviluppo dell'impostazione classica o oggettivistica);
- c) nella terza fase affiorano le insufficienze dell'impostazione classica e vi è una revisione del giudizio sulle teorie bayesiane.

Nella prima fase, importanza decisiva ebbero le discussioni sull'impostazione data da Bayes. Il punto cruciale di tutte le discussioni fu quello che viene spesso indicato come "postulato di Bayes" o "principio di indifferenza" quello della uniforme distribuzione a priori.

Caduta in discredito l'induzione statistica bayesiana, a causa della insoddisfazione per il principio di indifferenza a cui si era fatto ricorso in maniera indiscriminata, gli statistici si impegnarono nella ricerca di procedimenti convenienti di inferenza statistica di tipo nuovo, che non facessero riferimento, almeno in via esplicita, alla formula di Bayes.

La preoccupazione di giungere ad una spiegazione oggettiva della statistica costituisce la principale idea direttiva durante il grandissimo sviluppo che la nostra disciplina ha avuto in questo secolo. Si può parlare di un vero e proprio rinascimento della statistica che incomincia specialmente in Inghilterra. Le radici di questa ripresa sono rintracciabili negli studi di Galton, K. Pearson, Fisher, Gosset e di altri eminenti studiosi. In seguito si è sviluppata una scuola di cui si sogliono nominare, come principali esponenti J. Neyman e E. Pearson.

La caratteristica di tutto questo movimento era l'insistenza sulla oggettività. Il ragionamento induttivo derivante dal teorema di Bayes veniva rifiutato nell'assunto che le probabilità iniziali non esistono. La ragione di questo atteggiamento sta nel fatto che si intendeva la probabilità come frequenza o meglio come il "limite" di una frequenza.

La questione invece della distinzione se si tratti di pensare o agire in senso induttivo, è una questione che divide in due i sostenitori della scuola classica. Da una parte vi è Fisher e dall'altra Neyman e E. Pearson. J. Neyman distingue nettamente la stima dei parametri dalla prova delle ipotesi ed afferma che nessun test statistico può consentire di rifiutare una ipotesi se non attraverso

il paragone con un'altra ipotesi. Neyman e Pearson partono dall'assunto che nessun test può dare evidenza alla verità o falsità di una singola ipotesi. L'ipotesi viene respinta solo a vantaggio di un'altra più verosimile alla luce dei dati, facendo dell'atto inferenziale una "regola di condotta". Neyman osserva che accettare una ipotesi non significa necessariamente affermare che quell'ipotesi è vera ma solo che si adotterà una azione connessa con quell'ipotesi in un particolare problema. Per questo motivo Neyman parla di "comportamento induttivo" in contrapposizione all'inferenza induttiva elaborata da Fisher e rifiuta l'impostazione bayesiana proprio sostenendo che non si tratta di fare un ragionamento induttivo.

La sempre più stretta connessione tra stima e decisioni non poteva non preparare ad una logica decisionale. Sarà Wald a contribuire, sebbene involontariamente, a superare la prevenzione degli oggettivisti contro l'impostazione bayesiana soprattutto rendendo più effettivo il vedere i problemi secondo un comportamento induttivo. Wald nel 1937 propone l'analisi sequenziale e nel 1950 approda alla teoria delle decisioni. Uno dei risultati più importanti della sua teoria mostra che, sotto condizioni molto generali, le "decisioni ammissibili" sono quelle bayesiane vale a dire "la classe di tutte le soluzioni di Bayes e dei limiti delle soluzioni bayesiane" è una classe completa cioè una classe che contiene almeno una funzione di decisione "uniformemente migliore" di qualunque altra non appartenente ad essa". Ciò significa che la soluzione ottimale da dare ad un problema di decisione va ricercata tra le soluzioni di Bayes. Si può dire, come conclusione che Wald con la sua impostazione aveva incontrato tutto ciò che occorreva per obbligare il ritorno alla concezione bayesiana.

4.3. I problemi tipici dell'inferenza statistica.

Se effettuiamo un esperimento E , in una data situazione, possiamo ottenere evidentemente diversi risultati. I dati che si ottengono effettivamente dall'esperimento costituiscono pertanto solo un punto di uno spazio di possibili dati. È abbastanza naturale allora impostare l'analisi su tale spazio, insieme dei possibili risultati sperimentali, che indichiamo con Z . Per effettuare una analisi statistica occorre dotare Z di una struttura matematica. Lo spazio Z viene reso misurabile tramite l'associazione ad esso di una appropriata σ -algebra A . In questo contesto un **MODELLO STATISTICO PROBABILISTICO** per l'esperimento E è una terna (Z, A, P) , ove P è una famiglia di misure di probabilità sullo spazio misurabile (Z, A) . L'elemento essenziale del modello statistico è proprio P . I dati $z, z \in Z$ che si ricavano da E , una volta che l'esperimento è stato condotto, costituiscono la premessa per inferire su P . Le misure di probabilità P possono essere indicizzate da un parametro $\theta, \theta \in \Theta$:

$$P = (P_\theta : \theta \in \Theta)$$

cioè ad ogni θ è associata una misura di probabilità P_θ che assegna la probabilità ai sottoinsiemi di A . Θ è detto spazio dei parametri. Nel seguito supporremo che le misure di probabilità possano essere descritte mediante una funzione di densità $p(z|\theta)$. In questo caso il modello statistico può essere presentato nella forma semplificata :

$$(Z, p(z|\theta), \theta \in \Theta).$$

Per ogni scelta di θ in Θ selezioniamo una funzione di densità che regola il meccanismo aleatorio che genera i risultati sperimentali. Quando Θ è un sottoinsieme di uno spazio euclideo, il modello viene detto parametrico, altrimenti viene detto non parametrico.

È ovvio che se non si avessero incertezze intorno ai possibili valori di θ , vale a dire Θ contiene un solo elemento, non si avrebbe alcun problema statistico poiché saremmo in grado di calcolare la probabilità di un qualsiasi evento connesso con l'esperimento. In generale invece il modello prevede una duplice incertezza:

a) circa il valore di θ

b) circa il dato sperimentale regolato dalla funzione di densità $p(z|\theta)$, una volta scelto θ .

Supponiamo di effettuare un esperimento E e di ottenere come dato sperimentale: $z^* \in Z$. Evidentemente, ora, non c'è più incertezza sul dato sperimentale mentre rimane l'incertezza su θ . I

tipici problemi che si presentano nell'inferenza statistica sono quelli di **STIMA** e di **PROVA DELLE IPOTESI**.

Il problema della stima consiste nello scegliere un valore $\theta \in \Theta$ che plausibilmente ha generato, attraverso il modello, l'osservazione z^* .

Il problema di prova delle ipotesi consiste nel ripartire l'insieme dei valori ammissibili Θ in due sottoinsiemi Θ_1 e Θ_2 e decidere se il parametro θ che ha generato l'osservazione z^* appartiene a Θ_1 oppure a Θ_2 .

I problemi di stima e di prova delle ipotesi nella formalizzazione appena data considerano gli aspetti **IPOTETICI** del problema statistico. Nella sostanza in tali problemi si rivolge l'attenzione alla verifica della attendibilità di leggi che governano un dato fenomeno.

Un altro problema statistico è quello che nel 1920 Karl Pearson chiamava "The fundamental problem of practical statistics". Tale problema consiste nella previsione di nuove osservazioni sulla base di altre passate, problema che, a nostro avviso è meglio individuabile nei suoi connotati probabilistici, potendo essere ridotto alla determinazione di una **DISTRIBUZIONE PREDITTIVA**.

Di seguito verranno esposte le linee generali di alcuni dei metodi principali proposti per affrontare i problemi di stima (puntuale e per intervallo) e di verifica delle ipotesi statistiche sia nell'impostazione bayesiana sia nell'impostazione oggettivistica. Prima tuttavia presentiamo il criterio dell'utilità attesa elemento centrale della teoria delle decisioni.

4.4. Il criterio dell'utilità attesa.

Le tappe fondamentali in questa direzione di studio sono rappresentate dal contributo di Daniele Bernoulli (1738), dal saggio di F.P. Ramsey (1926), dall'appendice all'opera di J. von Neumann e O. Morgenstern (1944) sulla teoria dei giochi ed il comportamento economico, dalla sistemazione assiomatica dovuta a L.J. Savage (1954). L'argomento è stato trattato altresì in numerosi articoli da B. de Finetti (1952, 1964, 1967). Negli ultimi anni diversi sono i trattati di statistica che si richiamano ai lavori citati, tra essi ricordiamo quelli di T.S. Ferguson (1967), M.H. De Groot (1970), J.O. Berger (1985). Si veda altresì P.C. Fishburn (1970, 1982).

Supponiamo che lo statistico consideri uno spazio D delle possibili decisioni (es: stimatori di parametri, test di prova delle ipotesi, ecc...). La decisione viene assunta tenuto conto dell'espressione funzionale $q(\theta, d)$, definita su $D \times \Theta$, che esprime le conseguenze (guadagni o perdite) in forma monetaria derivanti dalla scelta della decisione d quando l'ipotesi θ è quella vera.

Poiché siamo incerti sulla vera realizzazione del parametro θ la funzione $q(\theta, d)$ è aleatoria, pertanto il problema della scelta ottimale della funzione d si può formulare come problema di ordinamento delle distribuzioni di probabilità di questi guadagni (perdite). Una decisione d è tanto migliore (peggiore) quanto maggiore (minore) risulta $q(\theta, d)$. Generalizzando il problema della decisione può essere concepito come quello di scegliere la migliore distribuzione di probabilità sul guadagno aleatorio tra quelle della classe Ξ delle distribuzioni unidimensionali ammissibili.

Assumiamo che in tale classe vi siano tutte e soltanto le distribuzioni che hanno supporto limitato $[a, b]$, e che tale classe sia chiusa rispetto alle combinazioni lineari convesse di suoi elementi. Il problema è dunque quello di individuare una sintesi a valori reali degli elementi di Ξ . Supponiamo che ad ogni elemento F (funzione di ripartizione) di Ξ sia possibile associare un numero reale $G(F)$, funzionale della funzione di ripartizione F . Supponiamo inoltre che il funzionale $G(F)$ soddisfi le seguenti proprietà: consistenza, monotonia ed associatività.

Tali condizioni, in virtù di un risultato di Nagumo-Kolmogorov-de Finetti (si veda il paragrafo 3.2.2), sono sufficienti per stabilire che esiste una funzione u continua e strettamente crescente, tale che:

$$G(F) = u^{-1} \left(\int_a^b u(x) dF(x) \right).$$

L'espressione precedente rappresenta la forma più generale che deve assumere $G(F)$ come strumento di decisione in condizioni di incertezza se si vuole che siano rispettate le condizioni dette in precedenza. La funzione $u(x)$ è detta **FUNZIONE DI UTILITÀ** e $u(G(F))$ è l'indice di utilità su Ξ . Pertanto la scelta tra guadagni aleatori è affidata alla speranza matematica di una funzione dei guadagni la cui forma è legata all'avversione al rischio di chi decide. La determinazione di u è quindi soggettiva: ogni decisore, a seconda del suo comportamento di fronte al rischio, esprimerà una propria funzione di utilità.

Il problema che ci eravamo posti era quello di fissare un criterio per scegliere nell'ambito di Ξ la distribuzione migliore. Fissata da parte del decisore la funzione di utilità, sia F_d la distribuzione di $q(\theta, d)$, si tratta di scegliere, se esiste, la decisione d^* per la quale risulti

$$G(F_{d^*}) = \sup_{d \in D} (F_d)$$

con $G(F)$ definito in precedenza, ovvero in modo che risulti massimo l'indice di utilità

$$u(G(F)) = \int_a^b u(x) dF(x).$$

ossia massimizzare l'utilità attesa.

4.5. Impostazione bayesiana nella risoluzione dei problemi di stima e di prova delle ipotesi.

Sia $(Z, p(z|\theta), \theta \in \Theta)$ il modello statistico. Dato θ , $p(z|\theta)$ regola l'incertezza su z (dati sperimentali); supponendo ora di poter fare altrettanto per θ assumiamo una legge di probabilità per il parametro θ che, per semplicità, assumeremo data sotto forma di densità: $h(\theta)$, $\theta \in \Theta$ che chiameremo **DENSITÀ INIZIALE**. Una volta ciò sia avvenuto il problema è ancora quello di esprimere una misura di plausibilità dei valori di θ . Vale a dire: abbiamo inizialmente una certa probabilità che θ appartenga ad un dato insieme B calcolata mediante la densità $h(\theta)$, osserviamo un risultato sperimentale z e ci chiediamo quale sia la probabilità che $\theta \in B$ dopo aver appreso tale risultato sperimentale. Come determinarla? Attraverso una **DENSITÀ DI PROBABILITÀ FINALE**. Per la determinazione esiste una risposta standard data dalla formula di Bayes, cioè:

$$h(\theta|z) = \frac{h(\theta)p(z|\theta)}{\int h(\theta)p(z|\theta)d\theta}$$

dove $h(\theta)$ è la densità iniziale, $p(z|\theta)$ è la verosimiglianza e $h(\theta|z)$ è la densità di probabilità finale ed esprime una densità di probabilità sui possibili valori di θ , da cui:

$$\Pr\{\theta \in B|z\} = \int_B h(\theta|z)d\theta.$$

Come si vede oltre alla verosimiglianza, una analisi completa dei problemi statistici, richiede l'individuazione della distribuzione iniziale del parametro. Un problema molto delicato ma di fondamentale importanza è, perciò, quello della scelta della distribuzione iniziale A tal proposito si veda de Finetti e Savage (1962), Edwards, Lindman e Savage (1963), Jeffreys (1961), Raiffa e Schlaifer (1961), Savage (1962). La scelta della distribuzione iniziale è soggettiva. Alcuni Autori R.von Mises (1942), I.J.Good (1956) H.Robbins(1955) hanno introdotto in letteratura

l'impostazione "**empirico bayesiana**.", impostazione in cui la distribuzione iniziale ha una interpretazione frequentista e la scelta è determinata dai dati.

La distribuzione finale esaurisce il problema dell'inferenza statistica in termini bayesiani. Esistono, tuttavia, situazioni in cui è utile ricercare una conveniente sintesi della distribuzione finale. La determinazione di questa sintesi può avvenire mediante il ricorso alla teoria delle decisioni.

In base al criterio dell'utilità attesa discusso nel paragrafo precedente si sceglie la decisione d in modo che risulti massimo l'indice di utilità:

$$u(G(F)) = \int u(q(\theta, d)) dH(\theta|x)$$

con u funzione di utilità espressa dal decisore che si suppone nulla in 0. E' consuetudine in statistica ragionare sulla perdita espressa in forma monetaria anzichè sui guadagni. Pertanto posto $-u(q(\theta, d)) = l(d, \theta)$ (funzione di perdita) il problema della scelta della decisione ottimale è ricondotto a quello della ricerca della decisione d^* che rende minimo, quando esiste, il danno atteso finale (rischio finale):

$$\rho(d) = \int l(\theta, d) dH(\theta|x)$$

ossia tale che:

$$\rho(d^*) = \inf_{d \in D} \int l(\theta, d) dH(\theta|x)$$

Tale regola di decisione prende il nome di **REGOLA DI DECISIONE BAYESIANA** ed il procedimento viene detto "analisi bayesiana in forma estensiva".

Vale la pena di osservare che tale metodo soddisfa al principio di verosimiglianza (cioè l'inferenza dipende esclusivamente dal risultato sperimentale acquisito e non dalle proprietà dello spazio campionario) poiché x è fissato.

Si noti infine che la disponibilità di osservazioni campionarie non apporta modifiche qualitative al problema decisionale. Sia in presenza, sia in assenza di osservazioni un problema decisionale si risolve nella ricerca del minimo della funzione di rischio bayesiano. Segnaliamo, inoltre che la scelta della decisione ottimale può essere rappresentata come media alla Herzog (si veda paragrafo 3.2.2) della distribuzione finale.

4.5.1. STIMA PUNTUALE. Nei problemi di stima puntuale si ha $\Theta = D$ cioè la coincidenza dello spazio parametrico con lo spazio delle decisioni. Pertanto il problema diviene: scegliere tra i possibili valori di un parametro quel valore $\theta^* \in \Theta$ che rende minimo il danno atteso finale, ossia tale che:

$$\rho(\theta^*) = \int_{\Theta} l(\theta, \theta^*) dH(\theta|x) = \inf_{y \in \Theta} \int l(\theta, y) dH(\theta|x)$$

θ^* è detto "stima bayesiana" di θ . E' ovvio che la scelta di opportune funzioni di danno consente di determinare in forma standard le stime bayesiane. Ad esempio: quando la funzione di danno è quadratica, del tipo $C(\theta - \hat{\theta})^2$, la stima bayesiana coincide con il valore atteso della distribuzione finale; quando la funzione di danno è del tipo $l(\theta, \hat{\theta}) = k|\theta - \hat{\theta}|$ con $k > 0$ otteniamo come stima bayesiana un valore mediano della distribuzione finale..

4.5.2. STIMA PER INTERVALLO. Molte volte il problema della sintesi della distribuzione finale non viene risolto ricercando un singolo valore $\theta = \hat{\theta}(x_1, x_2, \dots, x_n)$ ma imponendo che la probabilità che θ appartenga ad un insieme I sia pari ad una costante assegnata:

$$\Pr\{\theta \in I|x\} = \int_I h(\theta|x) d\theta = 1 - \alpha \quad \alpha \in (0, 1)$$

$(1 - \alpha)$ è la probabilità che θ appartenga all'insieme I (dato x). Limitiamoci ora ad una classe di sottoinsiemi di Θ : la classe degli intervalli $I = (a, b)$ sottoinsieme di Θ con $a < b$. Definiamo **intervallo di confidenza bayesiano** quello relativamente al quale si abbia:

$$\int_a^b h(\theta|x) d\theta = 1 - \alpha$$

con α assegnato. Se non si assegnano "a priori" determinati requisiti, di cui devono godere gli intervalli, il problema presenta un notevole grado di indeterminatezza. Ad esempio se la funzione di densità è continua, di intervalli di questa natura ne esistono infiniti. E' possibile risolvere il problema seguendo due vie:

a) ricercando il valore di a e b tali che la lunghezza $l=b-a$ sia minima con il vincolo **Errore.**

b) ricercando intervalli centrali.

4.5.3. VERIFICA DELLE IPOTESI. Questo problema è caratterizzato dal fatto che la classe delle decisioni D contiene due elementi $\{d_0, d_1\}$ e lo spazio dei parametri è partizionato in due sottoinsiemi Θ_0 e Θ_1 in modo che se il parametro θ appartiene a Θ_0 si ritiene appropriata la decisione d_0 e se $\theta \in \Theta_1$ si ritiene appropriata d_1 . Si usa indicare le ipotesi statistiche con l'enunciato:

$$H_0 = \theta \in \Theta_0 \quad \text{ipotesi di nullità}$$

$$H_1 = \theta \in \Theta_1 \quad \text{ipotesi alternativa}$$

Pertanto scegliere una ipotesi corrisponde per quanto si è detto all'assunzione della decisione d_0 o d_1 .

La particolare struttura dello spazio delle decisioni induce una partizione sullo spazio campionario in due sottoinsiemi S_0 e S_1 , tali che, se il campione appartiene ad S_0 si prende la decisione d_1 (si rifiuta l'ipotesi H_0), se il campione appartiene a S_1 si prende la decisione d_0 (si accetta l'ipotesi H_0).

Un problema delicato è quello della corretta interpretazione da dare a "accettare l'ipotesi H_0 " ed a "rifiutare l'ipotesi H_0 ". Tale dizione non ha il significato di ritenere "vera" una ipotesi e neppure che i dati osservati conducono a ritenerla vera. L'unica interpretazione ammissibile è quella che attribuisce alla scelta di una ipotesi il significato di comportamento conforme alla supposizione della verità di quell'ipotesi.

Dato un modello $p(x|\theta)$, $\theta \in \Theta$ supponiamo si voglia controllare l'ipotesi $H_0: \theta \in \Theta_0$ contro l'ipotesi $H_1: \theta \in \Theta_1$ con $\Theta_0 \cap \Theta_1 = \emptyset$ sulla base di un risultato sperimentale. Il problema è, in generale, risolto calcolando le probabilità finali delle due ipotesi:

$$\Pr(\theta \in \Theta_0 | x) \quad \text{e} \quad \Pr(\theta \in \Theta_1 | x)$$

ed eventualmente mettendole a confronto. Il ricercatore è lasciato poi libero di decidere per l'una o l'altra ipotesi. Potrebbe, ad esempio, decidere di prendere una determinata decisione, dicendo che $\theta \in \Theta_0$ se la probabilità finale $\Pr(\theta \in \Theta_0 | x)$ è maggiore della $\Pr(\theta \in \Theta_1 | x)$. In molte circostanze si usa immaginare che, oltre all'insieme di decisioni $D: \{d_0, d_1\}$ vi sia un danno conseguente alla scelta delle decisioni. Una tipica funzione di danno $l(\theta, d)$ potrebbe essere la seguente:

$$l(\theta_0, d_0) = 0 \text{ per } \theta \in \Theta_0 \quad ; l(\theta_1, d_0) = a > 0 \text{ per } \theta \in \Theta_1$$

$$l(\theta_0, d_1) = b > 0 \text{ per } \theta \in \Theta_0 \quad ; l(\theta_1, d_1) = 0 \text{ per } \theta \in \Theta_1$$

Utilizzando lo schema della teoria delle decisioni con una funzione di danno di questo tipo è necessario calcolare il rischio finale bayesiano per le decisioni d_0, d_1 . Si sceglierà la decisione che avrà un rischio bayesiano più piccolo. Conviene sottolineare che i due criteri sono equivalenti se a e b sono uguali.

4.5.3. PROBLEMI NON PARAMETRICI. L'impostazione bayesiana ai problemi d'inferenza è stata assai fruttuosa nell'analisi dei problemi parametrici : più difficile è stata invece la sua applicazione ai problemi formulati in modo non parametrico. In questi ultimi lo " spazio dei parametri " è di solito uno spazio funzionale; in particolare quello delle funzioni di ripartizione con supporto dato o più in generale, quello di tutte le funzioni di ripartizione. E' proprio connessa a questa particolare natura dello spazio dei parametri la complessità accennata. E' di J.E.Rolph (1968) il primo tentativo diretto di depositare sul particolare spazio parametrico una misura di probabilità utilizzando i momenti; crediamo però che migliore impostazione generale presentino i lavori successivi. Ci riferiamo ai lavori di T.S.Ferguson (1973, 1974) e C.Antoniak (1974). Ferguson introdusse una particolare misura di probabilità aleatoria retta da un processo di Dirichlet che bene si adatta come distribuzione sullo spazio delle misure di probabilità su spazi misurabili astratti ed in particolare su R^k . Antoniak ne considerò una generalizzazione che permette di affrontare problemi di inferenza non parametrica non adatti ad essere trattati col processo di Dirichlet. K.A.Doksum (1974) introdusse infine una classe molto generale di processi aleatori: i processi neutrali a destra. Essi possono essere definiti a partire da un qualunque processo ad incrementi indipendenti, quasi certamente non decrescente; quello di Dirichlet su R corrisponde ad un particolare processo ad incrementi indipendenti. Molte sono le applicazioni del processo suddetto (analisi della varianza, modello lineare , analisi discriminante, teoria della credibilità , modelli con punto di cambiamento, ecc..).

4.6. Impostazione classica (oggettivistica) nella risoluzione dei problemi di stima e di prova delle ipotesi.

Il tratto distintivo tra l'impostazione classica e quella bayesiana dei problemi di inferenza risiede, nella sostanza, nell'uso della regola di Bayes e nelle conseguenze che ne derivano. Gli statistici classici rifiutano l'uso del teorema di Bayes a causa dell'interpretazione che assegnano alla probabilità che ha, per essi, un significato oggettivo e quindi è indissolubilmente ancorata alla nozione di frequenza. Da ciò dunque la rinuncia alla predetta regola, non potendosi sempre assegnare alla distribuzione iniziale tale carattere oggettivo, e la pretesa di concludere il ragionamento induttivo solo con il riferimento alla verosimiglianza e/o alle proprietà dello spazio campionario.

Gli oggettivisti affermano con forza che le valutazioni personali non devono entrare nella trattazione dell'incertezza. Così facendo tuttavia, sono costretti ad inventare caso per caso "test" per la conferma delle ipotesi arrivando ad una proliferazione di metodi per estrarre delle conclusioni da un insieme di dati e quindi con la " pretesa " di prendere le decisioni in base soltanto all'esito delle osservazioni che hanno fornito i dati.

La scuola classica si avvale di due principi fondamentali : quello della verosimiglianza e quello della ripetizione dell'esperimento. Il primo stabilisce che tutte le informazioni che un generico campione può fornire sono contenute nella funzione di verosimiglianza e altresì questa, per un campione dato, misura la plausibilità dei possibili modi di essere del fenomeno. Il secondo principio considera il campione effettivamente ottenuto alla stregua di uno dei possibili campioni che si sarebbero potuti ricavare ripetendo un gran numero di volte, nelle stesse condizioni, l'operazione di campionamento. Nell'ambito dell'impostazione classica le ipotesi non sono viste come realizzazioni di un ente aleatorio ma come possibili stati di natura tra i quali ve n'è uno vero che è compito dello

statistico rivelare. Vale altresì la pena di sottolineare che all'interno della impostazione oggettivistica si riscontrano due atteggiamenti riguardo al ruolo della teoria delle decisioni. Da una parte vi è Fisher, secondo cui la teoria delle decisioni deve limitarsi a trattare i problemi pratici e non i problemi scientifici per i quali è inadeguata, dall'altra vi è Wald (fondatore della teoria delle decisioni) che tende ad inquadrare ogni problema statistico nell'ambito di questa teoria.

Presentiamo i metodi utilizzati all'interno di tale impostazione per risolvere i problemi di stima puntuale e per intervallo e di prova delle ipotesi.

Nella impostazione classica l'obiettivo della teoria della stima è quello di assegnare un valore ad uno o più parametri di una distribuzione sulla base di un gruppo parziale di osservazioni. I metodi di stima vengono classificati in "parametrici" e "non parametrici" a seconda delle ipotesi che vengono adottate sulla famiglia delle distribuzioni su cui si desidera indagare: si parla di problemi parametrici se gli elementi della famiglia differiscono tra loro solo per i valori dei parametri; non parametrici allorchè differiscono oltre che per i parametri per la forma funzionale delle distribuzioni considerate. La stima di un parametro può essere effettuata in due modi diversi: stima puntuale e stima per intervallo. Il problema della stima puntuale consiste nell'assegnare un valore plausibile al parametro non noto θ sulla base di un campione e di fornire possibilmente anche una misura della precisione della stima effettuata. Il metodo di stima per intervallo equivale a fornire un insieme, ed in particolare un intervallo, il quale plausibilmente contiene il "vero" valore del parametro.

4.6.1. STIMA PUNTUALE. Sia X una variabile aleatoria e si supponga che la sua funzione di ripartizione (f.r.) F dipenda da un certo numero di parametri $\theta \in \Theta$, mentre sia nota la forma analitica $F(x, \theta)$. Sia inoltre (X_1, X_2, \dots, X_n) un campione estratto dalla f.r. $F(x, \theta)$, cioè una ennupla di variabili aleatorie mutuamente indipendenti ed identicamente distribuite secondo la legge F . Limitiamoci, per semplicità, al caso in cui Θ è un sottoinsieme di R .

Nel problema della stima è necessario determinare uno stimatore. Si dice stimatore di θ una funzione $T=T(X_1, X_2, \dots, X_n)$ se T applica R^n su Θ . Il valore che assume per una particolare realizzazione $X_1=x_1, X_2=x_2, \dots, X_n=x_n$ è detto stima di θ .

La questione centrale della teoria della stima puntuale è quella di ricercare uno stimatore T di θ che fornisca "buoni" risultati. Considerare buono o non buono un dato stimatore è in larga misura dovuto a certe attese che si nutrono circa gli stimatori. La teoria classica della stima ha predisposto un certo numero di proprietà di cui deve godere uno stimatore per essere considerato buono. Le principali proprietà sono: "non distorsione", "consistenza", "efficienza" e "sufficienza". Vediamone le definizioni.

a) **NON DISTORSIONE** Uno stimatore T è detto non distorto per θ se la speranza matematica di T coincide con θ : $E_{\theta}T(X_1, X_2, \dots, X_n) = \theta$.

b) **CONSISTENZA.** La proprietà di consistenza richiede che se (X_1, X_2, \dots, X_n) è un campione estratto da F , allora la successione degli stimatori T_n converga in probabilità verso il parametro θ , cioè,

$$\lim_{n \rightarrow \infty} \Pr \{ |T_n - \theta| > \varepsilon \} = 0$$

per ogni $\varepsilon > 0$ e per ogni $\theta \in \Theta$.

c) **SUFFICIENZA.** Vediamo innanzitutto di cogliere il significato della importante proprietà della sufficienza. Il lavoro dello statistico consiste nell'interpretare, dopo ogni esperimento, i dati che ha raccolto e di trarre alcune valide conclusioni sulla popolazione in esame. E' chiaro che se in luogo di considerare la sequenza dei valori osservati si fissa l'attenzione su una funzione campionaria perdiamo parte dell'informazione acquisita. Questa perdita dovuta alla "riduzione dei dati" può essere inessenziale per il problema che ci interessa, in tal caso diremo che lo stimatore è "sufficiente". Una definizione sufficientemente generale è la seguente.

Uno stimatore T è sufficiente per θ (o per la famiglia di f.r. F) se la distribuzione condizionale di (X_1, X_2, \dots, X_n) dato $T = t$, è indipendente da θ (eccetto per un insieme di probabilità nulla di valori t).

La definizione data di sufficienza è dovuta a Fisher (1922) e non è di facile applicazione poiché presuppone la conoscenza dello stimatore che si vuole provare essere sufficiente. Un risultato che permette di superare questo inconveniente è il cosiddetto teorema di fattorizzazione dovuto a Neyman (1935) e perfezionato successivamente da Halmos e Savage (1949) e Bahadur (1954). Nella versione relativa a variabili aleatorie discrete (o assolutamente continue) tale risultato può enunciarsi dicendo che uno stimatore T è sufficiente se, e solo se, la funzione di probabilità dell'ennupla campionaria (o la funzione di densità) può fattorizzarsi secondo

$$p(x_1, \dots, x_n; \theta) = g(T(x_1, \dots, x_n); \theta)h(x_1, \dots, x_n)$$

in cui h è una funzione non negativa delle sole x_i e non dipende da θ mentre g è una funzione non negativa di θ e di $T(x_1, \dots, x_n)$.

Esiste un'ampia famiglia di distribuzioni che ammette stimatori sufficienti (anzi sufficienti minimali). Essa è la "famiglia esponenziale" a cui appartengono le più note distribuzioni di interesse statistico come, ad esempio, quella binomiale, di Poisson, gaussiana ecc... Questa famiglia è stata ampiamente studiata e caratterizzata da vari autori e una trattazione generale ne è stata fatta da E. Lehmann (1959, 2° ed. '86).

d) **EFFICIENZA**. L'indicatore assunto per il confronto degli stimatori è l'errore quadratico medio di campionamento $E(T-\theta)^2$. Nel caso in cui lo stimatore è non distorto tale errore quadratico coincide con la varianza. Allora uno stimatore T_1 non distorto per θ è migliore di un'altro stimatore, pure non distorto per θ , T_2 , se $\text{Var}_\theta(T_1) < \text{Var}_\theta(T_2)$, qualunque sia θ .

Interessante sarebbe determinare stimatori non distorti che avessero varianza uniformemente minima. Una disuguaglianza che permette di determinare tali stimatori è quella dovuta a Cramer (1946) e Rao (1945).

Giova mettere in evidenza che i concetti di non distorsione e di sufficienza risultano essenziali per la ricerca degli stimatori a varianza uniformemente minima come è stato esplicitamente posto in evidenza da Rao (1949), Blackwell (1947) e da Lehmann e Scheffé (1950) (per approfondimenti si veda Lehmann (1983)).

Quando, con i procedimenti accennati non è possibile determinare stimatori ottimali, sono state proposte altre metodologie con l'obiettivo di costruire stimatori che, almeno asintoticamente, posseggano buoni requisiti. Tra esse emerge quella connessa al "**metodo della massima verosimiglianza**". Tale metodo è stato introdotto in letteratura da Fisher.

La funzione di verosimiglianza, nel caso di un campione di ampiezza n , è data da (ponendo $\mathbf{x} = (x_1, x_2, \dots, x_n)$)

$$L(\mathbf{x}, \theta) = \prod_{i=1}^n f(x_i, \theta) \quad \theta \in \Theta$$

Si dirà che $T = \theta(X_1, \dots, X_n)$ è uno stimatore di massima verosimiglianza se θ è un punto di massimo (stretto) della funzione di verosimiglianza, cioè se $\hat{\theta}$ è tale che

$$L(\mathbf{x}, \hat{\theta}) = \sup_{\theta \in \Theta} L(\mathbf{x}, \theta).$$

Le proprietà degli stimatori di massima verosimiglianza sono ampiamente trattate da Zacks (1971).

4.6.2. **STIMA PER INTERVALLO.** La teoria degli intervalli di confidenza è dovuta principalmente a J.Neyman. Presentiamo la metodologia per la costruzione dell'intervallo di confidenza per un singolo parametro θ . Sia $T(X)$ con $X=(X_1, X_2, \dots, X_n)$ uno stimatore di θ . Assumiamo che $T(X)$ sia una variabile aleatoria continua. Allora, date le probabilità α_1 e α_2 , è possibile determinare $h_1(\theta)$ e $h_2(\theta)$ tali che

$$\Pr\{ T(X) \leq h_1(\theta) \mid \theta \} = \alpha_1$$

e

$$\Pr\{ T(X) \geq h_2(\theta) \mid \theta \} = \alpha_2$$

Per ogni particolare valore di θ la probabilità che $T(X)$ sia compresa tra $h_1(\theta)$ e $h_2(\theta)$ è $1-\alpha_1-\alpha_2$. Se $h_1(\theta)$ e $h_2(\theta)$ sono funzioni strettamente monotone di θ allora esistono le funzioni inverse ($\theta_1(T), \theta_2(T)$). Pertanto le seguenti affermazioni sono equivalenti:

$$1) h_1(\theta) \leq T(X) \leq h_2(\theta)$$

$$2) \theta_1(T) \leq \theta \leq \theta_2(T)$$

Allora $(\theta_1(T), \theta_2(T))$ è un intervallo di confidenza per θ a livello $1-\alpha_1-\alpha_2$.

L'affermazione che la probabilità che $\theta_1(T)$ e $\theta_2(T)$ racchiudono il valore di θ è $(1-\alpha_1-\alpha_2)$ non ha in generale senso una volta che T sia stato calcolato per una particolare realizzazione campionaria per il fatto che, nella teoria frequentista, θ è un numero certo (non aleatorio) sebbene non noto. L'interpretazione che si deve dare è quella di una proposizione che a lungo andare risulta vera nel $100(1-\alpha_1-\alpha_2)$ per cento dei casi.

4.6.3. **VERIFICA DELLE IPOTESI.** L'impostazione classica e la sistemazione di questo capitolo della statistica induttiva è opera principalmente di J.Neyman e E.Pearson (1928, 1933). Il testo classico di sistemazione di tale teoria è *Testing Statistical Hypotheses* di Lehmann (1959, 2ed.1986). Il problema può essere formulato nel modo seguente:

Sia X una variabile aleatoria dotata di funzione di densità $f(x, \theta)$. Sia (x_1, x_2, \dots, x_n) un punto nello spazio campionario e θ un punto nello spazio dei parametri Θ . Definiamo l'ipotesi nulla $H_0 = \theta \in \Theta_0$ e l'ipotesi alternativa con $H_1 = \theta \in \Theta_1$.

Vediamo ora un test non casualizzato per provare le ipotesi suddette.

Qualsiasi test non casualizzato divide lo spazio campionario in due sottoregioni, S_0 , regione di rifiuto ed il suo complemento, S_1 , regione di accettazione. Scegliere un test equivale a scegliere una regione S_0 . Utilizzando un test si può giungere ad una decisione corretta o si può commettere uno dei due errori seguenti:

- rifiutare l'ipotesi quando è vera (errore di prima specie);
- accettare l'ipotesi quando è falsa (errore di seconda specie).

Evidentemente sarebbe desiderabile proporre un test in modo da minimizzare questi due tipi di errore. Sfortunatamente quando il numero delle osservazioni è fissato, le probabilità di errore non possono essere controllate simultaneamente. In generale si assegna un limite superiore alla probabilità di rifiutare H_0 quando è vera e si prova a minimizzare l'errore di seconda specie subordinatamente a questa condizione. Il ricercatore seleziona un numero α tra 0 e 1, detto *livello di significatività* ed impone la condizione che

$$\Pr\{ X \in S_1 \mid \theta \in \Theta_0 \} \leq \alpha$$

Tale condizione assicura che al variare dei valori di θ in Θ_0 la probabilità può variare ma l'errore di 1^a specie non può superare α . Poichè vi possono essere molte regioni S_1 dobbiamo trovare quella che massimizza

$$\Pr\{X \in S_1 \mid \theta \in \Theta_1\}$$

Questa probabilità, fissata una regione S_1 , è una funzione di θ ed è detta *potenza del test* ed indica la probabilità di rifiutare l'ipotesi nulla che è proprio quello che vogliamo fare quando $\theta \in \Theta_1$. Considerata come funzione di θ per ogni $\theta \in \Theta$, la probabilità suddetta è detta *funzione di potenza*. La scelta del livello di significatività è in generale arbitraria.

Consideriamo ora la struttura di un test casualizzato.

Per ogni valore x un test sceglie tra due decisioni, rifiutare o accettare, con certe probabilità che dipendono da x e che indicheremo con $\phi(x)$ e $1 - \phi(x)$ rispettivamente. Se il valore di X è x , è messo in atto un esperimento aleatorio con due possibili risultati R e R^* , le probabilità dei quali saranno $\phi(x)$ e $1 - \phi(x)$. Se il risultato dell'esperimento è R , l'ipotesi è rifiutata, altrimenti è accettata. Un test casualizzato è, naturalmente, completamente caratterizzato da una funzione ϕ , la *funzione critica*, con $0 \leq \phi(x) \leq 1$ per ogni x . Se ϕ assume solamente i valori 0 e 1 ritorniamo al caso di un test non casualizzato. L'insieme dei punti x per cui $\phi(x) = 1$ è allora la regione di rifiuto, così che in un test non casualizzato ϕ è semplicemente la funzione indicatore della regione critica.

Se la distribuzione di X è P_θ , e si utilizza la funzione critica ϕ , la probabilità di rifiuto è

$$E_\theta \phi(X) = \int \phi(x) dP_\theta(x).$$

Il problema è selezionare ϕ in modo tale da massimizzare la potenza $B_\phi(\theta) = E_\theta \phi(X)$ per ogni θ appartenente a Θ_1 sotto il vincolo $E_\theta \phi(X) \leq \alpha$ per ogni $\theta \in \Theta_0$.

Un test che soddisfa ai requisiti suddetti è detto uniformemente più potente. In generale test uniformemente più potenti non sempre esistono. La domanda che ci si pone è la seguente: come costruire il test quando non esiste quello uniformemente più potente?. In tali casi è necessario introdurre nel problema alcune specifiche assunzioni sui criteri di prova o sulle situazioni considerate. Si possono seguire due vie:

- possiamo avanzare alcune particolari assunzioni sul modello in esame e ricercare test uniformemente più potenti per tali modelli. Questo modo di procedere limita il campo di applicazione dei test ottenuti anche se non sono necessarie grandi limitazioni.
- possiamo lasciare il modello di partenza molto generale ma porre alcune restrizioni ai test che si prendono in considerazione. Le restrizioni da imporre sono suggerite essenzialmente da motivi pratici. Test uniformemente più potenti possono esistere all'interno di questa classe ristretta di test e non esistere nella classe più ampia di tutti i possibili test. Le restrizioni atte a limitare l'insieme di tutti i possibili test a particolari sottoinsiemi sono: la non distorsione, la somiglianza e l'invarianza (si veda Lehmann 1986).

4.6.4. PROBLEMI NON PARAMETRICI. I problemi di natura non parametrica sono caratterizzati dalla mancanza di specificazione della forma analitica delle f.r. coinvolte. Nozione essenziale per affrontare i problemi di stima in ambito non parametrico è quella di funzione stimabile con relativo grado e nucleo. Nel contesto esaminato grande importanza riveste una classe di funzioni campionarie dette U-statistics, mediante le quali è generalmente possibile ottenere stimatori non distorti e a varianza uniformemente minima. Tale classe di funzioni è stata proposta da W. Hoeffding (1948); per una trattazione completa si veda Fraser (1957), Lee (1990).

Per quanto riguarda i test, in ambito non parametrico il primo problema che è stato affrontato è quello della misura della correlazione tra due variabili, interpretata come misura di indipendenza nel caso gaussiano. Il coefficiente di correlazione basato sui ranghi è stato proposto da C. Spearman nel 1904 mentre quello di M. Kendall nel 1938. I problemi di confronto di omogeneità tra due campioni sono stati risolti sotto l'ipotesi classica di normalità da F. Wilcoxon (1945). L'estensione a k popolazioni sulla base di k campioni indipendenti è dovuta a W. Kruskal e W. Wallis (1952, 1953). Il proliferare delle tecniche ha portato gli statistici a studiare l'efficacia degli strumenti non parametrici rapportandoli a quelli ottimali in un certo contesto. I lavori di E. Pitman dal 1938 al 1948 hanno dato origine al concetto di "efficienza asintotica relativa".

La ricerca di una legge di probabilità che sia la base di un modello statistico è, per definizione, un problema non parametrico: a fianco del celebre test del χ^2 proposto da K. Pearson (1900) si trovano numerose altre procedure dovute a H. Cramer (1928), R. Von Mises (1931), A. Kolmogorov (1933).

4.6.5. LA TEORIA DELLE DECISIONI COME IMPOSTAZIONE UNIFICANTE NELLA RISOLUZIONE DEI PROBLEMI DI STIMA E DI PROVA DELLE IPOTESI.

Un approccio unificante ai problemi di stima puntuale e per intervallo e di verifica delle ipotesi è fornito dalla teoria delle decisioni di Wald (1950). Seguendo lo stesso Wald possiamo dire che tale impostazione generalizza e semplifica la teoria di Neyman e Pearson unificando, cioè, trattando problemi considerati distinti nella teoria di Neyman e Pearson come casi particolari del problema della teoria delle decisioni. Come abbiamo visto nel paragrafo precedente nell'impostazione bayesiana contrasti di questo tipo non sussistono nel senso che una risposta coerente al problema induttivo, a prescindere dalle decisioni che si intendono prendere, è fornita dalla distribuzione finale e/o predittiva; quando però esigenze di svariata natura lo richiedono, l'impostazione bayesiana consente di inserire gli elementi della decisione e fornisce un criterio coerente per la scelta di quella ottimale.

Un importante concetto all'interno della teoria delle decisioni di Wald è quello di funzione di decisione ammissibile.

La funzione di decisione d_1 si dice "ammissibile" se non esiste un'altra funzione di decisione d_2 tale che :

$$R(\theta, d_2) \leq R(\theta, d_1) \text{ per ogni } \theta$$

e

$$R(\theta, d_2) < R(\theta, d_1) \text{ per almeno un } \theta.$$

La seguente espressione

$$R(\theta, d) = E\{l(\theta, d)\} = \int_{-\infty}^{\infty} l(\theta, d) f(x, \theta) dx$$

è detta "**funzione di rischio**" e misura la perdita attesa quando la funzione di decisione adottata è d e l'ipotesi vera è θ .

Come criteri di scelta della funzione di decisione ottimale si possono citare, nuovamente, quello bayesiano in base al quale si sceglie la funzione di decisione che rende minima $E(R(\theta, d))$ essendo la speranza matematica calcolata rispetto alla distribuzione iniziale sul parametro; e quello "**minimax**" consistente nell'assumere la decisione che rende minimo il funzionale di d

$$M(d) = \text{Max}_{\theta} R(\theta, d)$$

Nella ricerca di procedure statistiche ottimali è interessante individuare classi di funzioni di decisione per le quali se si considera una decisione ad essa esterna, è possibile determinare una decisione della classe con un rischio minore. Una simile classe è detta "completa". La si dirà completa minimale quando non contiene una sottoclasse propria completa. Se esiste una classe completa minimale allora essa coincide con quella delle funzioni di decisione ammissibili ed in più, si dimostra (sotto condizioni molto generali) che le decisioni ammissibili sono quelle bayesiane rispetto ad opportune distribuzioni iniziali o sono limiti di soluzioni bayesiane e che ogni funzione di decisione bayesiana è ammissibile.

4.7. SULL'IMPOSTAZIONE PREVISIVA DELL'INFERENZA.

Si è andata profilando recentemente (si veda: Cifarelli e Regazzini 1980-81, 1982), nell'ambito bayesiano, una impostazione dell'inferenza **completamente predittiva** che costituisce una alternativa allo schema tradizionale basato sul concetto di modello statistico. Le motivazioni che

inducono molti studiosi ad accogliere il punto di vista previsivo completo dell'inferenza possono essere ricondotte ai seguenti punti:

a) scopo dell'indagine scientifica è quello di fornire strumenti adeguati di previsione e non quello di scoprire leggi "vere" che dovrebbero governare i fenomeni;

b) osservabilità, come requisito di ogni condizione su cui si intenda esprimere giudizi di probabilità.

D'altra parte l'impostazione ipotetica pone, come scopo primario dell'induzione statistica, l'individuazione del "vero" meccanismo aleatorio che genera i dati e, generalmente, non pone problemi circa la natura delle ipotesi associate al modello (significato fisico, osservabilità). La riflessione e la discussione sulla liceità dell'impostazione ipotetica viene da lontano (si veda de Finetti 1937) ma ciò nonostante tale impostazione ha dominato gli studi statistici di questo secolo.

Uno statistico opera in una visione predittiva dell'inferenza quando fissa come obiettivo la valutazione della probabilità di un evento collegato a certi fatti futuri subordinatamente alle determinazioni di certe osservazioni. Supponiamo ad esempio di avere una successione di v.a. $\{X_n\}$ la cui distribuzione congiunta è specificata e di essere interessati a fare inferenza su X_{n+1} , X_{n+2} , ..., data l'esperienza $E = (X_1, X_2, X_3, \dots, X_n)$. Il processo di apprendimento dall'esperienza è rappresentato dal calcolo della distribuzione del processo condizionatamente ad E . In tale situazione solamente il passato ed futuro di risultati sperimentali sono coinvolti, senza qualsiasi ulteriore intermediazione di alcun modello. E' detta *predittiva* la distribuzione

$$\Pr\{X_{n+1} \leq x_{n+1}, X_{n+2} \leq x_{n+2}, \dots, X_{n+k} \leq x_{n+k} | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} \quad (4.7.1)$$

qualunque siano n e k interi positivi. Se la legge di probabilità della successione dei numeri aleatori $\{X_n\}$ è assegnata la determinazione della predittiva segue gli usuali procedimenti del calcolo delle probabilità. Non è necessario pertanto ricorrere al modello statistico a patto che si sia potuta assegnare la legge del processo. Ciò non è sempre agevole. E' immediato osservare che per poter valutare la probabilità di eventi futuri sulla base di eventi passati (noti) occorrerà stabilire un legame logico di dipendenza tra tutti gli eventi (passati e futuri). Uno schema che negherebbe l'essenza del ragionamento induttivo è quello di indipendenza stocastica. Al di là dell'indipendenza si annoverano forme di dipendenza più o meno strette che possono variamente caratterizzare la dipendenza della previsione da quanto si è osservato.

Scartata l'ipotesi di indipendenza stocastica perchè la sua adozione negherebbe il senso del ragionamento induttivo, la scelta dovrà ricadere su una condizione che, ragionevolmente, possa essere ritenuta soddisfatta in diverse circostanze concrete nel cui ambito lo statistico viene invitato ad intervenire. La suddetta condizione dovrebbe però soddisfare un'altra esigenza: quella di far comprendere con chiarezza e semplicità formale come giuoca l'esperienza, rappresentata dalle osservazioni sperimentali, sulla previsione degli esiti delle prove future. Tenuto conto che le situazioni in cui le prove dello stesso fenomeno si svolgono in condizioni invariate sono molte le f.r. $F_n(x_1, x_2, \dots, x_n)$ dovrebbero essere tali da garantire l'invarianza della (4.7.1), qualunque siano n ed r , rispetto alle permutazioni degli elementi del vettore (x_1, x_2, \dots, x_n) delle osservazioni. E' questa una delle condizioni caratterizzanti la **SCAMBIABILITA'** del processo $\{X_n\}$ che, perciò potrà

essere accolta come ipotesi di lavoro. Entro lo schema della scambiabilità vi è una stretta connessione tra problema impostato in modo ipotetico e problema previsivo. Infatti attraverso il **teorema di rappresentazione** di de Finetti (de Finetti 1937) può essere chiarita la relazione esistente tra i due punti di vista o, più precisamente, che sotto opportune condizioni l'assegnazione di una distribuzione predittiva implica "l'esistenza" di un modello statistico tradizionale.

Infatti, se la successione delle v.a. $\{X_n\}$ è scambiabile e P indica la misura di probabilità del processo corrispondente, indichiamo con B un evento relativo ad un numero qualsiasi delle X_n , allora esiste una misura μ sullo spazio Y delle funzioni di ripartizione su R tale che

$$P(B) = \int_Y P_F(B) d\mu(F)$$

dove $P_F(B)$ è la probabilità di B calcolata sotto l'assunzione che tutte le X_n sono indipendenti ed identicamente distribuite secondo F , μ è la misura di probabilità iniziale che ora ha il significato di misura limite del processo \hat{F}_n .

Se $I(A)$ è la funzione indicatore dell'insieme A , indichiamo con

$$\hat{F}_n = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x) \quad x \in \mathbb{R}$$

la funzione di ripartizione empirica.

Il teorema di rappresentazione fornisce una preziosa indicazione per la valutazione della distribuzione iniziale. Conviene chiarire che il teorema stabilisce che la valutazione di μ si basa su opinioni relative a fatti osservabili cioè \hat{F}_n ma non autorizza a considerarla come legge di probabilità di un effettivo numero aleatorio.

In conclusione possiamo dire che il raccordo logico tra l'impostazione previsiva e quella ipotetica è dato dal teorema di rappresentazione di de Finetti.

All'interno dell'impostazione previsiva è possibile riformulare alcuni concetti fondamentali dell'inferenza statistica quale il concetto di riassunto esaustivo. Infine poichè l'impostazione previsiva e l'impostazione ipotetica sono da considerarsi facce di uno stesso problema i riassunti esaustivi a fini previsivi svolgono una importante funzione nello studio delle relazioni tra le due impostazioni.

5. ALCUNI ARGOMENTI SPECIFICI.

5.1. I MODELLI LINEARI.

5.1.1. MODELLO LINEARE DELLA REGRESSIONE FUNZIONALE.

Si ha un modello lineare quando un carattere quantitativo Y , detto anche variabile dipendente o endogena, è considerato come parzialmente spiegato o determinato da un certo numero di altri caratteri, X_1, X_2, \dots, X_k detti variabili indipendenti o esplicative o esogene, tramite una relazione stocastica della forma

$$Y = \sum_{i=1}^k \beta_i X_i + e_i \quad i=1, 2, \dots, k$$

nella quale i β_i sono k parametri non noti ed e_i è una variabile aleatoria che dovrebbe tener conto dell'eventuale esistenza di altre variabili, che per vari motivi non possono essere identificate, e di eventuali effetti accidentali.

Il modello può essere scritto in forma matriciale, posto di avere n osservazioni $(Y_j, X_{ij}), i=1, 2, \dots, k; j=1, 2, \dots, n$:

$$\mathbf{Y} = \mathbf{X}\mathbf{\beta} + \mathbf{e} \quad (5.1.1)$$

in cui \mathbf{Y} è un vettore ($n \times 1$) delle $n > k$ osservazioni sulla variabile dipendente, $\mathbf{\beta}$ vettore ($k \times 1$) non noto, \mathbf{e} vettore ($n \times 1$) degli errori ed \mathbf{X} la matrice nota e non stocastica di tutte le osservazioni su tutte le variabili indipendenti X_1, X_2, \dots, X_k . Naturalmente, se nel modello compare un addendo costante, allora la matrice \mathbf{X} avrà la prima colonna costituita da elementi eguali ad uno. Il modello è lineare, ma occorre sottolineare che tale qualifica non discende dalla natura lineare di \mathbf{Y} rispetto alle variabili indipendenti, bensì dalla linearità rispetto ai parametri. Il modello lineare è variamente denominato a seconda della natura delle quantità che servono a definirlo, cioè $(\mathbf{Y}, \mathbf{X}, \mathbf{\beta}, \mathbf{e})$. Si

hanno così: il modello della regressione funzionale (modello definito con le condizioni enunciate in precedenza), il modello della regressione condizionale, il modello della regressione stocastica, il modello della regressione con variabili affette da errore, il modello della regressione con parametri aleatori, il modello del disegno degli esperimenti. Trattiamo brevemente il modello della regressione funzionale ed assumiamo che siano soddisfatte le ipotesi seguenti:

a) $E(\mathbf{e})=0$; b) $E(\mathbf{e}\mathbf{e}') = \sigma^2 \mathbf{I}_n$, $\sigma^2 > 0$; c) $\text{rango}(\mathbf{X}) = k$

La prima condizione è di immediata percezione e non è strettamente necessaria poichè alla sua violazione si può ovviare pensando modificata la costante del modello. La seconda condizione impone che le componenti della variabile n -dimensionale \mathbf{e} siano non correlate ed in più che posseggano la stessa varianza (in generale non nota). La terza condizione è semplicemente la condizione di indipendenza lineare tra le colonne della matrice \mathbf{X} . Essa è cioè la condizione di assenza di multicollinearità.

Un procedimento assai noto per la stima di β e che assicura certe proprietà ottimali agli stimatori è quello dei minimi quadrati. esso consiste nel determinare il vettore β che minimizza la funzione :

$$Q(\beta) = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta).$$

Tale metodo porta a stimare β con

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

Il teorema di Gauss-Markov assicura che nella classe degli stimatori lineari e non distorti lo stimatore ottenuto con il metodo dei minimi quadrati è ottimale.

Lo stimatore non distorto per σ^2 è :

$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})}{(n-k)}.$$

Ricordiamo che sovente risulta inaccettabile l'ipotesi b) di omoschedasticità e non correlazione tra gli errori. Qualora sia $E(\mathbf{e}\mathbf{e}') = \Sigma$ (simmetrica, definita positiva, nota) lo stimatore

$$\hat{\beta}_A = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \mathbf{X}'\Sigma^{-1}\mathbf{Y}$$

è detto stimatore dei minimi quadrati generalizzati o di Aitken e gode delle medesime proprietà di $\hat{\beta}$.

Come si vede il metodo dei minimi quadrati non è in alcun modo legato ad una particolare ipotesi riguardante la distribuzione dei residui. Se però si vogliono approfondire ulteriormente le proprietà delle stime e risolvere altri problemi connessi al modello lineare occorre ipotizzare la forma della distribuzione di \mathbf{e} . L'ipotesi comunemente adottata per \mathbf{e} è quella di normalità, cioè :

$$\mathbf{e} \approx N(0, \sigma^2 \mathbf{I}_n).$$

Con tale ipotesi è ora possibile non solo fornire gli stimatori puntuali per β e σ^2 ma anche: costruire gli intervalli di confidenza per ogni coefficiente β_i $i=1, \dots, k$, costruire le regioni di confidenza di un gruppo o di tutti i coefficienti, controllare ipotesi riguardanti i singoli o gruppi di coefficienti.

Facciamo osservare che con l'ipotesi di normalità dei residui lo stimatore ottenuto con il metodo dei minimi quadrati coincide con lo stimatore di massima verosimiglianza.

C.Stein (1956), trattando il cosiddetto problema delle k medie :

$$\mathbf{Y} = \mathbf{z} + \mathbf{u}$$

in cui \mathbf{u} è un vettore aleatorio normale con valore atteso nullo e matrice di varianze e covarianze $\sigma^2 \mathbf{I}_k$, \mathbf{z} è un vettore di parametri $k \times 1$, mostrò che la stima dei minimi quadrati di \mathbf{z} quando k è maggiore od uguale a 3 e la funzione di danno è quadratica, non è ammissibile, vale a dire esistono altri stimatori con funzione di rischio non maggiore per ogni \mathbf{z} . Tale risultato fu successivamente generalizzato da Brown (1966) al caso di una vasta classe di funzioni di danno. Poiché, trasformando opportunamente le variabili, il modello (5.1.1) può scriversi nella forma delle k medie, il risultato di Stein si applica anche alla stima dei minimi quadrati di β .

Si deve forse a risultati di tale natura la spinta degli statistici a tentare altri approcci ed in particolare quello bayesiano. Ed infatti, lo stesso Stein, in un lavoro successivo (1962), dopo aver proposto uno

stimatore ammissibile (ridge) si preoccupò di porre in luce come tale stimatore potesse essere interpretato come una approssimazione di quello bayesiano nella consueta ipotesi di danno quadratico. D.V.Lindley (1962) fornì analoga interpretazione introducendo, in "nuce" quello che successivamente verrà detto "modello gerarchico". L'approccio bayesiano al modello lineare della regressione, sia per quanto riguarda i problemi di natura ipotetica sia per quanto concerne l'aspetto predittivo, procede, come di consueto, con l'assegnazione di una distribuzione iniziale sui parametri che compaiono nel modello statistico e quindi con la determinazione di quella finale per il tramite della formula di Bayes. La teoria base del modello lineare da un punto di vista bayesiano si può trovare in Broemeling (1985).

5.1.2. ANALISI DELLA VARIANZA E DELLA COVARIANZA.

La tecnica utilizzata per effettuare un confronto tra le medie di diverse popolazioni è detta analisi della varianza. L'elemento principale è il confronto tra la variabilità tra le medie delle popolazioni e la variabilità tra popolazioni. La ragione di tutto ciò discende dal fatto che anche se non vi fossero differenze tra le medie delle popolazioni, tutte le osservazioni non sono identiche e ciò per la natura casuale delle misurazioni. Se la variabilità tra le popolazioni e la variabilità tra le medie sono dello stesso ordine di grandezza allora possiamo concludere che non vi sono differenze significative tra le medie delle popolazioni.

Nell'analisi della varianza i modelli matematici impiegati sono tradizionalmente classificati in tre categorie: modello ad effetti fissi, modello ad effetti aleatori, modello misto. L'impostazione classica dell'analisi del modello ad effetti fissi porta alle procedure del test t e del test F. È stato dimostrato che tali procedure hanno una naturale interpretazione bayesiana.

Facciamo osservare che il modello di analisi della varianza può essere dedotto dal modello lineare generale. In tale situazione gli elementi della matrice X sono scelti in modo da includere o escludere l'appropriato parametro per ogni osservazione e così sono 0 o 1. Ogni colonna di X può essere considerata come una variabile indicatrice e X è generalmente detta "matrice disegno".

L'analisi della covarianza indica una metodologia statistica mediante la quale, utilizzando i principi dell'analisi della varianza e della regressione, può essere effettuata una particolare analisi delle variazioni simultanee che si presentano in due o più variabili correlate.

Tale metodologia si propone, anzitutto, di correggere -eliminando l'effetto della regressione- le differenze fra le variazioni medie dei valori di due o più gruppi e poi di provare la significatività di queste differenze corrette.

5.1.3 MODELLO LINEARE GENERALIZZATO.

Molti modelli statistici possono essere scritti nella forma generale seguente:

osservazione = componente sistematica + componente aleatoria

o in simboli $Y_i = u_i + e_i$ dove Y_i indica l' i -esima osservazione della v.a. Y , $u_i = E(Y_i)$ ed $E(e_i) = 0$. Nel modello lineare generale visto sopra, u_i è assunta funzione lineare delle variabili esplicative, ossia

$u_i = x_i \cdot \beta$, dove x_i è il vettore delle variabili esplicative per l' i -esima osservazione e β il vettore dei

parametri. Gli errori sono assunti indipendenti e distribuiti normalmente. Nel modello lineare generalizzato la distribuzione degli errori è pensata più generale e si assume che diverse funzioni di u_i siano combinazioni lineari dei β . Più precisamente, il modello lineare generalizzato assume che:

1) le v.a. Y_i siano indipendenti ed identicamente distribuite con distribuzione appartenente alla famiglia esponenziale (ricordiamo che la famiglia esponenziale include la normale, gamma, esponenziale, binomiale e la distribuzione di Poisson come casi particolari).

2) esista una funzione legame, g , (che deve essere monotona e differenziabile) tale che: $g(u_i) = x_i' \beta$.

La quantità $\eta_i = x_i' \beta$ è detta "previsore lineare" e pertanto $g(u_i) = \eta_i$.

Se le Y_i sono distribuite normalmente e g è la funzione identità è facile verificare che il modello lineare generalizzato si riduce al modello lineare generale visto prima, tuttavia il modello lineare generalizzato può descrivere molti altri problemi. Vediamone due esempi.

a) Se le Y_i si distribuiscono secondo la distribuzione di Poisson e g è una funzione logaritmica allora otteniamo il modello "**log-lineare**". Tale modello è applicato nello studio delle tabelle di contingenza.

b) Un'altra importante applicazione è quella riguardante i dati binari. Supponiamo che ogni Y_i sia distribuita secondo una binomiale con parametri n_i e p_i e che p_i dipenda dai valori delle variabili esplicative. Allora $u_i = n_i p_i$ e nell'uso comune vi sono due funzioni legame.

La trasformazione "logit" di p_i è definita da $\log\left[\frac{p_i}{1-p_i}\right]$ e pertanto la funzione legame è:

$$g(u_i) = \log\left[\frac{u_i}{n_i - u_i}\right] = \log\left[\frac{n_i p_i}{n_i - n_i p_i}\right] = \log\left[\frac{p_i}{1-p_i}\right].$$

Se $g(u_i) = x_i' \beta$ è una funzione lineare delle variabili di previsione allora l'analisi risultante è detta "**regressione logistica**".

Una funzione legame alternativa è la trasformazione "probit" data da:

$$g(u_i) = \Phi^{-1}\left(\frac{u_i}{n_i}\right) = \Phi^{-1}(p_i)$$

dove Φ è la f.r. della distribuzione normale standardizzata. Il risultato dell'analisi viene detto "**analisi probit**".

I modelli lineari generalizzati sono stati introdotti in letteratura da J.A. Nelder e R.W. Wedderburn nel 1972 ma l'applicazione in maniera estensiva è molto più recente. Le applicazioni sono state facilitate dalla disponibilità di packages (es: GLIM). Una trattazione completa di tali modelli si trova nel libro di McCullagh e Nelder (1983, 2ed 1989).

5.1.4. MODELLO LINEARE DINAMICO

Il tema della previsione ha da sempre affascinato gli statistici ed occupa un posto importante nella teoria dei processi stocastici e nell'analisi delle serie storiche. I primi lavori su tale argomento sono stati di Kolmogorov e Wiener. Kolmogorov (1941) ha affrontato il problema per processi aleatori stazionari a tempo discreto, mentre Wiener (1949) ha analizzato il caso di processi a tempo continuo. Kalman (1960, 1963) e Kalman e Bucy (1961) hanno esteso i risultati di Wiener e Kolmogorov ai processi aleatori non stazionari. Tali risultati hanno giocato un ruolo importante nell'ambito dei programmi spaziali e sono diventati uno strumento fondamentale di analisi nella teoria del controllo ingegneristico.

La generalità dei risultati e la facile programmazione su computer ha reso molto popolare il filtro di Kalman. Il filtro viene utilizzato per una varietà di fini anche se la sua funzione base è quella di stimare lo stato di un sistema. Per rendersi conto delle possibilità applicative di tale strumento basta scorrere le riviste statistiche ed ingegneristiche degli ultimi venti anni.

Il lavoro di Kalman esplicita la connessione tra il problema della stima dei parametri ed i **modelli lineari dinamici o modelli stato spazio**. Tali modelli godono della proprietà che implica l'indipendenza del futuro di un processo dal suo passato, dato il presente (proprietà markoviana).

Il modello lineare dinamico è descritto da due equazioni. La prima equazione mette in evidenza come vengono generate le osservazioni ed è detta **equazione delle osservazioni**

$$y_t = F_t \theta_t + v_t$$

dove t = indice del tempo ($t=1, 2, \dots$); y_t = vettore ($m \times 1$) delle osservazioni del processo fatte al tempo t ; θ_t = vettore ($n \times 1$) dei parametri del processo al tempo t ; F_t = matrice ($m \times n$) di variabili indipendenti note al tempo t ; v_t = vettore ($m \times 1$) di variabili aleatorie.

L'aspetto dinamico dei parametri è incorporato nella **equazione del sistema**:

$$\theta_t = G_t \theta_{t-1} + w_t$$

dove G_t = matrice ($n \times n$) spiega come i parametri si evolvono nel tempo; w_t = vettore ($n \times 1$) che indica lo sviluppo stocastico dei parametri nel tempo.

Come si può notare il modello è formato da una equazione che specifica come il processo y_t è dipendente stocasticamente dai parametri θ_t e da un'altra equazione che specifica come il processo dei parametri si evolve nel tempo. Tali modelli sono anche detti **modelli stato-spazio**. Il problema che ci si pone è sempre quello di determinare stime ottimali di θ_t sulla base delle osservazioni del processo. Tali stime si possono ottenere con diverse metodologie: mediante il teorema delle proiezioni ortogonali; mediante l'utilizzo del teorema della correlazione normale; mediante il metodo dei minimi quadrati generalizzati; mediante la diretta applicazione del teorema di Bayes. A nostro avviso l'impostazione bayesiana permette di cogliere meglio l'essenza del problema (si veda Harrison e Stevens 1976).

Assumiamo che v_t si distribuisca secondo $N(0, V_t)$ e w_t si distribuisca secondo $N(0, W_t)$ e che v_t e w_t siano stocasticamente indipendenti ed inoltre che le matrici di varianze e covarianze V_t W_t siano note. Assumiamo inoltre che θ_0 abbia distribuzione $N(m_0, C_0)$. Applicando il teorema di Bayes si ottiene che la distribuzione finale di θ_t è $N(m_t, C_t)$ con

$$m_t = G_t m_{t-1} + R_t F_t' (V_t + F_t R_t F_t')^{-1} (y_t - F_t G_t m_{t-1}) = G_t m_{t-1} + K_t (y_t - F_t G_t m_{t-1})$$

$$C_t = R_t - R_t F_t' (V_t + F_t R_t F_t')^{-1} F_t R_t$$

$$\text{con } R_t = (G_t C_{t-1} G_t' + W_t)$$

Tali equazioni ricorsive sono note come filtro di Kalman. Il filtro di Kalman è pertanto un metodo esplicito per adattare le stime e le previsioni man mano che aumentano le informazioni.

La quantità

$$K_t = R_t F_t' (V_t + F_t R_t F_t')^{-1}$$

viene detta "matrice guadagno" e mette in evidenza come avviene l'aggiornamento della stima.

Il procedimento fornisce la stima bayesiana ottenuta in forma "sequenziale". Al tempo $t=0$ si ha che la stima è determinata interamente dall'opinione iniziale; dopo aver osservato la prima osservazione, tale valore viene inserito nella stima di θ , ossia in m_1 e così via. Si può dire che il processo determina in maniera iterativa la distribuzione iniziale al tempo t e quella finale.

Una sistemazione della teoria di varie classi di modelli dinamici si può trovare nel libro di West e Harrison (1989).

5.2) METODI DI RICAMPIONAMENTO.

Nei paragrafi precedenti abbiamo visto che uno dei problemi classici dell'inferenza statistica è quello per cui si vuole stimare una caratteristica di una funzione di ripartizione F , $\theta(F)$, quando sono note le determinazioni di n variabili aleatorie X_1, \dots, X_n che si suppone essere indipendenti ed identicamente distribuite con funzione di ripartizione F .

Si vuole, per esempio, conoscere la media o la varianza di F . Normalmente la stima della caratteristica di F in considerazione viene effettuata con l'uso di una funzione $T(X_1, \dots, X_n)$ (stimatore) delle osservazioni. La scelta della funzione T è effettuata sulla base di criteri di ottimalità che in ultima analisi dipendono dalla funzione di ripartizione di $T(X_1, \dots, X_n)$. Anche quando si abbia a disposizione un'unico stimatore T , il problema della conoscenza della funzione di ripartizione di $T(X_1, \dots, X_n)$ rimane comunque di notevole interesse; tramite essa si può, per esempio, valutare la bontà della stima fornita da T , o costruire regioni di confidenza per la caratteristica di F che si vuole stimare.

E' comunque spesso difficile risolvere il problema della determinazione della funzione di ripartizione di T in modo analitico soprattutto se la funzione di ripartizione F non appartiene ad un modello parametrico noto o la caratteristica di F che si vuole stimare non è una semplice funzione dei parametri del modello.

Negli ultimi anni si è assistito nella letteratura statistica ad un fiorire di interessi rivolti all'uso di alcuni **metodi di ricampionamento** il cui scopo è appunto quello di stimare la funzione di ripartizione di una statistica delle osservazioni difficilmente calcolabile da un punto di vista analitico. I metodi di ricampionamento più noti sono quelli del **jackknife** e del **bootstrap**.

IL METODO DEL JACKKNIFE. Tale tecnica è stata introdotta nella metodologia statistica da Quenouille (1949, 1956) con lo scopo di ridurre la distorsione di uno stimatore nel contesto dell'analisi delle serie temporali. Successivamente Tukey (1958) suggerì di utilizzare tale tecnica per sviluppare un metodo generale di costruzione di intervalli di confidenza approssimati e conìò il termine "jackknife".

La metodologia di jackknife richiede di ricalcolare lo stimatore che interessa per tutti gli $\frac{n!}{[k!(n-k)!]}$ sottoinsiemi di ampiezza k delle n osservazioni. k è, in generale, uguale a $n-1$. Partendo dal campione (X_1, X_2, \dots, X_n) si costruiscono dei sottocampioni sopprimendo la i -esima osservazione $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ e si definisce lo stimatore T_{n-1}^i basato su quest'ultimo sottocampione.

Si definisce "pseudo valore" d'ordine i di T_n lo stimatore $J_i(T) = nT_n - (n-1)T_{n-1}^i$ e la media degli pseudo valori viene detto **stimatore jackknife**.

IL METODO DEL BOOTSTRAP. Nel 1979 Efron ha introdotto una tecnica di ricampionamento, da lui chiamata di bootstrap, con lo specifico scopo di stimare alcune caratteristiche della funzione di ripartizione di $T(X_1, \dots, X_n)$, come la distorsione o la varianza, quando esse sono difficilmente calcolabili in modo analitico e la funzione di ripartizione F è totalmente o parzialmente sconosciuta. La semplicità concettuale della tecnica proposta da Efron è bilanciata soltanto dal richiedere un uso spesso intensivo delle capacità di calcolo fornite dai moderni elaboratori elettronici. Fu comunque ben presto evidente che il bootstrap poteva essere proficuamente usato per stimare la stessa funzione di ripartizione di $T(X_1, \dots, X_n)$. Nei dodici anni che ci separano dal primo articolo di Efron, un gran numero di lavori sono apparsi nella letteratura statistica e probabilistica con lo scopo di giustificare da un punto di vista teorico questo metodo che contemporaneamente riscuoteva notevoli successi in campo applicativo.

L'idea di base del procedimento è molto semplice. In sostanza si tratta di una tecnica di ricampionamento che fa uso prevalentemente dell'elaboratore elettronico e fornisce soluzione

numerica a problemi la cui complessità rende proibitivo il ricorso all'analisi statistica tradizionale. L'idea è quella di ricampionare in (X_1, X_2, \dots, X_n) e di studiare il comportamento dello stimatore T .

L'algoritmo del bootstrap si può così riassumere:

Fase 1) Sulla base del campione (X_1, X_2, \dots, X_n) estratto dalla popolazione con funzione di ripartizione F si calcola la funzione di ripartizione empirica

$$\hat{F}_n = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x).$$

Fase 2) Condizionatamente a \hat{F}_n , si estrae un campione bernoulliano di ampiezza N da (X_1, X_2, \dots, X_n) e si ottiene $(X_1^*, X_2^*, \dots, X_N^*)$ (N può essere uguale ad n);

Fase 3). Si approssima il comportamento di T con quello di $T(X_1^*, X_2^*, \dots, X_N^*)$ che è lo stimatore di bootstrap.

Giova sottolineare che quest'ultima fase è generalmente ripetuta M volte (M relativamente grande) per ottenere una approssimazione mediante il metodo di Monte Carlo. In questo caso si osservano M campioni e così si ottengono $T_1^*, T_2^*, \dots, T_M^*$ stimatori.

L'aspetto rilevante di queste tecniche è dovuto al fatto che si prestano a moltissime applicazioni (es: modelli di regressione, disegni campionari complessi, econometria, analisi delle serie storiche, ecc...). Le ricerche in quest'area sono numerosissime, infatti ogni nuovo numero di qualsiasi rivista di statistica non manca di riportare contributi teorici o applicativi delle due suddette tecniche. Questa larga possibilità applicativa risiede nel fatto che sia il jackknife che il bootstrap sono in grado agevolmente di risolvere problemi complessi, per i quali le difficoltà di natura teorica precludono l'uso dei metodi tradizionali.

Segnaliamo infine che vi sono diversi altri metodi di ricampionamento, molti di questi sono descritti nell'eccellente studio di Efron del 1982. La versione bayesiana del bootstrap è stata introdotta in letteratura da Rubin (1981).

5.3) DISEGNO DEGLI ESPERIMENTI.

E' stato Fisher tra il 1920 ed il 1930 a definire il ruolo della statistica nel disegno degli esperimenti e, viceversa, il ruolo del disegno degli esperimenti in statistica. La storia del disegno degli esperimenti è un esempio emblematico dell'utilità della statistica nello sviluppo nelle scienze. Ancora oggi la ricerca segue due vie: da una parte vi è la continuazione del lavoro di Fisher (in questi **disegni "classici"** la struttura è in generale semplice ma le unità dell'esperimento possono avere una variabilità molto alta e pertanto la divisione delle unità in blocchi è importante) dall'altra vi è la presa in esame degli esperimenti tecnologici in cui la variabilità delle unità è virtualmente ignorata (negli esperimenti tecnologici la struttura dei trattamenti può essere complessa, dal momento che dipende dal modello matematico del sistema che si sta studiando). Quest'ultimo aspetto tecnologico ha portato alla teoria del disegno degli esperimenti ottimale. Negli ultimi anni tre sono i filoni di ricerca che hanno avuto maggior attenzione: **disegno ottimo, disegno per mezzo dell'elaboratore elettronico, disegni mistura**. Tralasciando il disegno "classico" (**disegno a blocchi, disegno fattoriale**) presentiamo brevemente questi tre settori.

DISEGNO OTTIMO La motivazione tradizionale sottostante la teoria del disegno ottimo è che gli esperimenti devono essere disegnati in modo da ottenere la miglior stima possibile del modello. L'ottimalità di un disegno fu considerata per la prima volta da Smith (1918) e successivamente ripresa da Wald (1943), Hotelling (1944) ma i maggiori contributi in quest'area sono dovuti a Kiefer (1958, 1959) e Kiefer e Wolfowitz (1959, 1960). Il libro di Silvey (1980) fornisce una esauriente presentazione dei risultati classici nella teoria del disegno ottimo. Per applicare il disegno ottimo è necessario un criterio per poter confrontare gli esperimenti ed un algoritmo per ottimizzare il criterio sull'insieme dei possibili disegni sperimentali. Il criterio classico è stato derivato nell'ambito

del modello lineare in cui si assume che i dati sperimentali possano essere rappresentati dall'equazione:

$$Y_i = \mathbf{f}(\mathbf{x}_i)' \boldsymbol{\beta} + e_i \quad i=1, 2, \dots, n$$

dove Y_i è la risposta dell'iesima prova dell'esperimento, \mathbf{x}_i è un vettore delle variabili di previsione per l'iesima prova, \mathbf{f} è un vettore di p funzioni, $\boldsymbol{\beta}$ è un vettore di p parametri non noti, e_i è l'errore sperimentale dell'iesima prova. Un modo naturale per misurare la qualità dell'inferenza rispetto ad un singolo parametro dipende dalla varianza dello stimatore del parametro. Se gli errori sono non correlati ed hanno varianza costante σ^2 , la matrice di varianza e covarianza degli stimatori ottenuti con il metodo dei minimi quadrati $\hat{\boldsymbol{\beta}}$ è: $\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$, dove \mathbf{X} è la matrice ($n \times p$) la cui iesima riga è $\mathbf{f}(\mathbf{x}_i)'$. Ci limitiamo al caso in cui \mathbf{X} abbia rango pieno. Un altro modo utile per misurare la bontà della stima dipende dalla varianza della previsione di Y dato \mathbf{x} :

$$d(\mathbf{x}) = \sigma^2 \mathbf{f}(\mathbf{x})' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{f}(\mathbf{x}).$$

Entrambe queste varianze dipendono dalla matrice di ordine ($p \times p$) $(\mathbf{X}'\mathbf{X})^{-1}$ e suggeriscono che un buon esperimento è tale se rende piccola questa matrice in qualche senso. Poichè non vi è un unico metodo per ordinare queste matrici sono stati proposti vari funzionali per misurare quanto è piccola la matrice. I più noti sono:

1. D-ottimalità.-Un disegno è detto D-ottimale se minimizza il determinante di $(\mathbf{X}'\mathbf{X})^{-1}$;
2. A-ottimalità.-Un disegno è detto A-ottimale se minimizza la traccia di $(\mathbf{X}'\mathbf{X})^{-1}$;
3. E-ottimalità. -Un disegno è detto E-ottimale se minimizza il massimo autovalore di $(\mathbf{X}'\mathbf{X})^{-1}$;
4. G-ottimalità.- Un disegno è detto G-ottimale se minimizza il max $d(\mathbf{x})$, dove il massimo è preso su tutti i possibili vettori \mathbf{x} ;
5. I λ ottimalità. un disegno è detto I_λ ottimale se minimizza **Errore.** dove λ è una misura di probabilità sullo spazio delle variabili previsive.

DISEGNO DEGLI ESPERIMENTI PER MEZZO DELL'ELABORATORE. Come dicevamo prima, l'idea base nel disegno ottimo è quella di scegliere un disegno che ottimizzi alcuni criteri inferenziali sull'insieme dei disegni considerati. In pratica questa ottimizzazione può essere difficoltosa o impossibile da risolvere analiticamente. La prima ricerca fatta utilizzando un elaboratore elettronico nella risoluzione del problema fu quella di Box e Hunter (1965) sui disegni non lineari.

DISEGNI MISTURA. In molte situazioni sperimentali la risposta dipende dall'ammontare relativo delle variabili previsive e non dall'ammontare assoluto. La natura di questi esperimenti può essere espressa considerando i seguenti vincoli: se X_1, X_2, \dots, X_k indicano le k variabili previsive, misurate come proporzioni, allora per ogni prova sperimentale abbiamo:

$$0 \leq X_j \leq 1 \text{ per ogni } j, \text{ e } \sum X_j = 1.$$

Sebbene la particolarità degli esperimenti mistura sia riassunta dal vincolo precedente, la teoria proposta può essere utilizzata in generale per ogni problema in cui vi sono uno o più vincoli lineari sulle variabili previsive. La risoluzione di questi problemi è stata formulata da Scheffé (1958).

Giova sottolineare che le applicazioni del disegno degli esperimenti vanno dal classico settore dell'agricoltura a quello più moderno e sorprendente dell'analisi numerica, alla teoria dei sistemi e al controllo statistico di qualità. L'accessibilità degli elaboratori ha altresì permesso l'utilizzo dei concetti del disegno ottimo in situazioni non standard.

5.4) ANALISI SEQUENZIALE.

Nei metodi statistici tradizionali (parametrici e non parametrici, univariati e multivariati), che costituiscono il fulcro dell'analisi statistica, l'ampiezza campionaria è sempre fissata a priori. L'analisi sequenziale attinge alle tecniche in cui l'ampiezza del campione e la composizione dei dati non è predeterminata ma può dipendere, in qualche maniera specificata, dai dati stessi man mano che essi divengono disponibili nel corso dell'indagine.

Gli elementi che caratterizzano il metodo sequenziale sono due:

- 1) un "procedimento di arresto" che indichi allo statistico come osservare i valori estratti dalla popolazione e quando interrompere il campionamento;
- 2) ad ogni stadio del campionamento un "procedimento decisionale" che indichi quale azione scegliere.

L'obiettivo principale dell'analisi sequenziale è quello di determinare un procedimento di arresto e un procedimento decisionale che soddisfino a certi criteri suggeriti dalla natura del problema inferenziale.

La moderna teoria dell'analisi sequenziale iniziò durante la seconda guerra mondiale per rispondere alla richiesta di metodi più efficienti nell'indagine campionaria.

Fu Wald a sviluppare il test del rapporto di probabilità sequenziale ed a fornire semplici approssimazioni delle probabilità di errore. Per una analisi dettagliata delle procedure sequenziali nell'ambito della teoria delle decisioni si può vedere Wald (1947), Ferguson (1967), Berger (1985).

Nell'area della stima sequenziale si possono ricordare i lavori di Thompson (1933), Neyman (1934), Hotelling (1941) ma il primo procedimento sequenziale per la stima puntuale fu proposto da Haldane (1945) e da Stein (1945) per la stima intervallare. Presentiamo ora brevemente il problema nella prova delle ipotesi.

Sia X_1, X_2, \dots, X_n , un campione bernoulliano estratto da una popolazione X avente funzione di densità $f(x, \theta)$ e sia θ un parametro non noto. Il problema è quello di provare, sotto opportuni criteri, una ipotesi $H_0: \theta \in \Theta_0$ contro l'ipotesi alternativa $H_1: \theta \in \Theta_1$ dove Θ_0 e Θ_1 sono sottoinsiemi disgiunti di Θ . Per specifici valori di $\theta \in \Theta_0$ e $\theta \in \Theta_1$ indichiamo $f(x)$ con $f_0(x)$ e $f_1(x)$ rispettivamente. Un "test sequenziale" di H_0 contro H_1 è definito dalla coppia aleatoria (N, D) dove N indica la variabile di arresto ed è definita da un insieme di procedimenti che interrompono il campionamento con $N=n$ osservazioni (x_1, x_2, \dots, x_n) per ogni $n \geq 1$. La funzione di decisione D è definita da un insieme di procedimenti, dopo che l'evento $N=n$ si è verificato, per accettare H_0 (i.e. $D=H_0$) o accettare H_1 (i.e. $D=H_1$) sulla base di (x_1, x_2, \dots, x_n) . Se $\Pr [N < \infty] = 1$ per ogni θ diremo che il procedimento sequenziale termina sicuramente. Quest'ultima è evidentemente una proprietà molto desiderabile. La quantità

$$E_\theta[N] = \sum n \Pr[N=n]$$

è detta numero campionario medio (NCM) richiesto per il procedimento. Un buon procedimento dovrebbe avere un numero medio piccolo per tutti i θ . Le funzioni seguenti:

$$P(\theta) = \Pr[D=H_1]$$

$$Q(\theta) = \Pr[D=H_0]$$

sono rispettivamente la funzione di potenza e la funzione operativa caratteristica del test. Se il procedimento termina sicuramente, è chiaro che $P(\theta) + Q(\theta) = 1$ per tutti i θ . Gli errori di prima e di seconda specie del test sono definiti: $\alpha(\theta) = P(\theta)$ per $\theta \in \Theta_0$ e $\beta(\theta) = Q(\theta)$ per $\theta \in \Theta_1$

Un buon test dovrebbe avere valori piccoli di $\alpha(\theta)$ e $\beta(\theta)$. Criteri per determinare test ottimali possono variare da problema a problema ma sono generalmente espressi in termini di N , NCM, probabilità di errore, e costo dell'esperimento.

Dal momento che con il test sequenziale, in determinate circostanze, si arriva a campioni di ampiezza molto grande e dal momento che le approssimazioni di Wald sono spesso poco accurate

vi è stata una proliferazione di procedure alternative basate sulla simulazione e sul calcolo numerico.

5.5 INDAGINI CAMPIONARIE.

Le indagini campionarie costituiscono una delle aree di ricerca più antiche e più importanti della statistica. In tale settore si è avuto di recente una rinascita nell'attività di ricerca. Una distinzione di base nei metodi di campionamento è tra "*campionamento probabilistico*" e "*non probabilistico*". Con il campionamento probabilistico ogni elemento della popolazione ha una probabilità nota (diversa da zero) di essere selezionato. Il campionamento probabilistico richiede un piano di campionamento mediante il quale il campione è estratto. Nella forma più semplice il piano può essere una lista degli elementi della popolazione ma, più in generale, è uno strumento per identificare gli elementi. Il maggior vantaggio del campionamento probabilistico è dovuto al fatto che per derivare le proprietà degli stimatori può essere impiegata la teoria statistica. Nessuno sviluppo teorico è invece possibile nel campionamento non probabilistico. Le forme più comuni di campionamento non probabilistico sono:

- a) **campionamento per convenienza o per "caso"**. Gli elementi sono scelti per convenienza o per caso (es: campione di volontari, avventori di un magazzino, intervistati all'angolo di una strada)
- b) **campionamento fatto in base al giudizio di esperti**.
- c) **campionamento per quota**. (molto usato nelle indagini di mercato). Agli intervistatori sono assegnate quote di persone da intervistare per differenti tipi di intervista. (es: ad un intervistatore viene richiesto di intervistare 5 uomini di età superiore ai 42 anni, 8 donne disoccupate, 10 casalinghe).

Sotto opportune condizioni ognuno di questi metodi può dare risultati utili. Giova sottolineare tuttavia che tali metodi non sono riconducibili alla teoria dei campioni in quanto non è configurabile un esperimento casuale che li generi. La sola via per dare una valutazione di questi metodi di campionamento è quella di ricondursi a situazioni in cui i risultati sono noti o per l'intera popolazione o relativamente ad un campione probabilistico e di effettuare alcuni confronti.

I principali metodi di campionamento probabilistico sono:

- 1) **campionamento casuale semplice**. Con questo metodo ogni possibile insieme di n differenti elementi, $S=(x_1, x_2, \dots, x_n)$ estratto da una popolazione di N elementi è egualmente probabile. Il campionamento può avvenire con o senza "reimmissione". La metodologia più utilizzata nelle indagini campionarie è quella del campionamento senza reimmissione. Il campionamento senza reimmissione è altresì utilizzato nella teoria del bootstrap per stimare gli errori non parametrici standard.
- 2) **campionamento sistematico**. Con questo metodo si seleziona ogni k -esimo elemento della lista della popolazione, iniziando da un numero compreso tra 1 e k scelto casualmente.
- 3) **campionamento stratificato**. Tale campionamento consiste nel suddividere la popolazione in un numero finito di gruppi o strati e all'interno di ogni strato viene selezionato un campione. Utilizzando questo procedimento le ampiezze campionarie di ogni strato sono scelte a priori per raggiungere gli obiettivi dell'indagine. Sovente è utilizzata la stessa proporzione campionaria in tutti gli strati (stratificazione proporzionale).
- 4) **campionamento "cluster (a grappoli)**. Come nel campionamento stratificato la popolazione è formata da un insieme di gruppi. Tuttavia, con il campionamento cluster viene selezionato solamente un campione dei gruppi i quali in seguito vengono completamente esaminati.
- 5) **campionamento a due o più stadi**. Si seleziona dapprima un campione di gruppi e successivamente un campione di elementi tra i gruppi selezionati (campionamento a due stadi). Evidentemente tale procedura può essere estesa a più stadi: al primo stadio si seleziona un campione di gruppi, al secondo stadio un campione tra i gruppi del primo stadio e così via.

6) **probabilità proporzionale all'ampiezza campionaria.** Un piano più efficiente di campionamento a più stadi e con ampiezze diverse dei gruppi consiste nello scegliere i gruppi ad ogni stadio con probabilità di inclusione proporzionale alla loro ampiezza.

7) **campionamento a due o più fasi.** Consideriamo una popolazione di N elementi, inizialmente non stratificata. Nella prima fase si estrae un campione casuale di ampiezza n . Si conduce una analisi su questo campione per ottenere informazioni su di una variabile X che può essere utilizzata per classificare le n unità ad esempio in k strati o essere utile per stimare un'altra variabile Y . Nella seconda fase si estrae un sotto-campione dagli n elementi selezionati nella prima fase. Questa seconda fase porterà a selezionare m_i ($i=1, 2, \dots, k$) unità in ognuno degli strati o m unità dagli n elementi della prima fase.

5.6) ROBUSTEZZA.

Box (1953) introdusse il termine "robusto" nella letteratura statistica per descrivere le procedure che forniscono buoni risultati anche quando sono violate le assunzioni sulle quali queste procedure si basano. Seguendo una linea di ricerca iniziata da K.Pearson (1931), Box esaminò gli effetti sull'analisi della varianza e sul test di Bartlett della rimozione dell'assunzione di normalità sulla quale tali tecniche si basano. Pearson nel 1931 scoprì che l'analisi della varianza era "robusta" alla violazione di tale assunzione ma avanzò l'ipotesi che tale conclusione poteva non essere vera nel confronto delle stime delle varianze allorché i campioni fossero indipendenti. Box nel 1953 trovò, invece, che il test di Bartlett è molto sensibile all'allontanamento dalla normalità. Questo risultato lo portò a sconsigliare l'utilizzo del test di Bartlett come test preliminare nell'analisi della varianza (una pratica raccomandata da molti statistici)

L'esame della sensibilità di molte tecniche standard alle assunzioni di base e la ricerca di nuove tecniche che sono meno sensibili è stato uno dei punti centrali della ricerca statistica di queste ultime decadi (si veda Huber 1981).

Le definizioni di robustezza avanzate nella letteratura statistica sono diverse. Sia P la legge di probabilità di un modello statistico, un procedura è robusta se :

- ammette una grande efficienza assoluta per tutte le alternative a P ;
- ammette una grande efficienza relativa rispetto ad una procedura ottimale sotto P ;
- ammette una grande efficienza assoluta su un insieme ben specificato di leggi;
- la distribuzione dello stimatore sul quale è basata la procedura varia "poco" allorché P è sottoposta a piccole alterazioni;
- essa è poco sensibile a raggruppamenti di dati, a troncamenti o a degli arrotondamenti;
- essa è poco sensibile all'abbandono dell'ipotesi di indipendenza.

Queste definizioni non forniscono una visione esaustiva del problema ma si basano tutte sul medesimo spirito. Seguendo P.Huber (1972) possiamo dire che la robustezza è una sorta di assicurazione: sono pronto a pagare una perdita di efficienza da 5 a 10 % rispetto al modello ideale per proteggermi dai cattivi effetti di piccole deviazioni dal modello. In generale, seguendo F.Hampel e P.Huber possiamo dire che un procedimento statistico è robusto se le sue prestazioni vengono alterate di poco per piccole modificazioni della legge di distribuzione del modello statistico.

Conviene sottolineare che nella statistica classica la robustezza, intesa come insensibilità a piccole variazioni delle assunzioni, concerne principalmente la robustezza rispetto a variazioni nella distribuzione assunta e al trattamento degli outliers (per una ampia trattazione si veda Huber (1981) e Hampel e al.(1986)).

Nell'ambito della statistica bayesiana la robustezza analizza la sensibilità degli output bayesiani (stime, decisioni, ecc.) alle variazioni degli input: distribuzione a priori, verosimiglianza e nell'approccio decisionale la stessa funzione di perdita (si veda Berger 1984).

5.7) SERIE STORICHE .

Un gruppo di osservazioni ordinate cronologicamente è detto comunemente SERIE STORICA. Tali serie vengono osservate in connessione a fenomeni molto diversi che vanno da quella dei prezzi di una merce ai mm. di pioggia giornaliera caduti in una determinata area.

I metodi tradizionali di analisi delle serie storiche riguardano la scomposizione della serie in : trend, componente stagionale, cambiamenti ciclici e componente erratica.

L'idea fondamentale per lo studio delle serie storiche è quella di pensare la serie stessa come scaturente da osservazioni condotte su una famiglia di variabili aleatorie $\{X(t); t \in T\}$; cioè, per ogni $t \in T$, $X(t)$ viene pensato come una osservazione effettuata su una variabile aleatoria. Tale famiglia di variabili è detta **PROCESSO STOCASTICO**. La serie storica potrà allora essere considerata come una particolare realizzazione del processo stocastico. Lo scopo che l'analisi delle serie storiche persegue è di duplice natura : da una parte quello di tentare di spiegare il meccanismo che genera la serie storica, dall'altra quello di prevedere il comportamento della serie nello sviluppo futuro eventualmente collegandolo a quello di altre serie storiche.

Un modo per descrivere un processo stocastico consiste nello specificare la distribuzione congiunta di $X(t_1), \dots, X(t_n)$ per ogni insieme di tempi t_1, \dots, t_n e per ogni valore di n . Purtroppo questo non sempre è agevole; un modo più utile per descrivere un processo stocastico è quello di dare i momenti del processo, in particolare, media, varianza e funzioni di autocovarianza.

Una importante classe di processi stocastici è formata dai processi stazionari.

In modo intuitivo possiamo dire che una serie storica è stazionaria se non vi sono cambiamenti sistematici in media (non vi è un trend), se non vi sono cambiamenti sistematici della varianza e se le variazioni periodiche sono state rimosse.

Formalmente una serie storica è detta *strettamente stazionaria* se la distribuzione congiunta di $X(t_1), \dots, X(t_n)$ è la stessa della distribuzione congiunta di $X(t_1+\tau), \dots, X(t_n+\tau)$ per ogni t_1, \dots, t_n, τ . In altre parole, spostando l'origine del tempo di un ammontare τ non vi sono effetti sulla distribuzione congiunta, la quale dipende solamente dagli intervalli tra t_1, \dots, t_n . La definizione data vale per ogni valore di n . Nella pratica, tuttavia, conviene dare una definizione di stazionarietà meno restrittiva.

Un processo è detto "*stazionario del secondo ordine o debolmente stazionario*" se ha media e varianza costante e le altre proprietà non mutano nel tempo; in particolare l'autocovarianza tra X_t e X_{t+k} , data da

$$\gamma(k) = E[(X_t - m(t))(X_{t+k} - m(t))]$$

dove $m(t) = E(X_t)$, dipende solamente dal tempo intercorrente, k , tra X_t e X_{t+k} .

Il coefficiente di autocorrelazione di ritardo k è dato da $\rho_k = \frac{\gamma_k}{\gamma_0}$ e l'insieme dei coefficienti $\{\rho_k\}$ è detto funzione di autocorrelazione. Lo "spettro" di un processo stazionario discreto è la trasformata di Fourier di $\{\gamma_k\}$. Per le serie storiche multivariate è di notevole interesse lo studio delle relazioni in e tra le serie. Ad esempio la funzione di cross-correlazione tra due serie stazionarie è una funzione che misura la correlazione tra X_t e Y_{t+k} . Vi sono molte classi di modelli sia stazionari che non stazionari. Il più semplice modello è dato dal processo aleatorio puro o "**white noise**" che è definito da una successione di variabili aleatorie indipendenti ed identicamente distribuite con media zero e varianza costante.

Sia Z_t un processo aleatorio puro. Un processo autoregressivo di ordine p (AR(p)) è definito da:

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + Z_t$$

dove $\alpha_1, \dots, \alpha_p$ sono costanti.

Un processo a media mobile di ordine q (MA(q)) è definito da :

$$X_t = Z_t - \theta_1 Z_{t-1} - \dots - \theta_q Z_{t-q}$$

dove $\theta_1, \dots, \theta_q$ sono costanti.

Utilizzando un operatore lineare, B , (backward shift) per cui $BX_t = X_{t-1}$ possiamo combinare il processo $AR(p)$ con il processo $MA(q)$ per ottenere un processo autoregressivo a media mobile $ARMA(p, q)$ nella forma

$$\alpha(B)X_t = \theta(B)Z_t \quad (5.7.1)$$

dove $\alpha(B) = 1 - \alpha_1 B - \dots - \alpha_p B^p$ e $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ sono polinomi in B di ordine p e q rispettivamente. Il processo $ARMA$ è *stazionario* se tutte le p radici dell'equazione caratteristica associata alla componente AR , cioè $\alpha(B)=0$, sono in modulo superiori a 1 e, inoltre, è *invertibile* se tutte le q radici dell'equazione caratteristica associata alla componente MA , cioè $\theta(B)=0$, sono in modulo superiori a 1.

L'invertibilità assicura che vi sia un unico processo MA per una data funzione di autocorrelazione. I processi non stazionari possono essere resi stazionari prendendo le differenze tali che

$$W_t = (1-B)^d X_t$$

Sostituendo W_t a X_t nella (5.7.1) otteniamo il processo $ARMA$ integrato di ordine p, d, q , indicato con $ARIMA(p, d, q)$ (per approfondimenti si veda Box-Jenkins 1970; Priestley 1981)

Un'altra classe di modelli di notevole interesse è la classe dei modelli stato-spazio (si veda il paragrafo 5.1). In tali modelli la previsione viene ottenuta utilizzando un procedimento di natura ricorsiva detto filtro di Kalman. Molti modelli di serie storiche possono essere riscritti nella formulazione di modello stato-spazio (si veda Harvey 1989).

6.UNO SGUARDO SULLA FRONTIERA DELLA STATISTICA.

Presentiamo infine alcune recenti direzioni di ricerca. E' evidente tuttavia che queste poche righe non possono dare l'idea dell'abbondanza delle ricerche statistiche sia metodologiche che applicate.

La statistica, in questi ultimi anni, si è sviluppata in molte direzioni. Gran parte dello sviluppo è dovuto ai miglioramenti tecnologici avvenuti. In particolare i progressi nella strumentazione e nelle comunicazioni hanno portato a nuove metodologie nella raccolta dei dati ed hanno posto nuovi problemi. La tecnologia ha portato ad un rapido sviluppo delle tecniche di analisi di dati multivariati e sulle tecniche di ricampionamento.

METODI DI RICAMPIONAMENTO. Le ricerche sui metodi di ricampionamento sono numerosissime e c'è da aspettarsi che, nei prossimi anni, metodi quali il jackknife, la cross-validation ed il bootstrap vengano estesi a nuovi modelli e che vengano analizzate le condizioni ottimali per permetterne un utilizzo appropriato.

METODI GRAFICI. Metodi grafici di tutti i tipi basati su elaboratori elettronici si sono sviluppati rapidamente e sono stati adottati in campi sempre più diversi ed è evidente che nuove applicazioni chiamano nuove tecniche grafiche. I metodi grafici sono stati studiati e sviluppati non solo in connessione con la descrizione e l'esplorazione dei dati ma anche per la costruzione di modelli e sono stati utilizzati per inferenza e decisioni come nelle carte di controllo.

INFERENZA BAYESIANA. Nell'ambito della statistica bayesiana sono stati sviluppati, in questi ultimi anni, soprattutto i metodi bayesiani robusti. Questi metodi analizzano la sensibilità delle conclusioni bayesiane (stime, decisioni ...) alle variazioni della distribuzione iniziale, verosimiglianza e provano a minimizzare questa sensibilità. Lo sviluppo dell'analisi della robustezza è avvenuta di pari passo anche all'interno dell'impostazione frequentista. L'analisi bayesiana trarrà sicuramente, da questi studi sulla robustezza, un nuovo impulso. Sono state altresì sviluppate nuove procedure numeriche e nuovi metodi grafici di analisi. Molta attenzione è stata data in questi ultimi anni ai metodi per l'analisi di modelli non parametrici e semiparametrici.

Il lavoro di Gelfand e Smith (1990) ha dato luogo ad un rapido sviluppo delle applicazioni e della teoria del campionamento sostitutivo, noto anche come algoritmo del Gibbs sampler. Tale

problematica è simile a quella introdotta da Tanner e Wong (1987) e Geman e Geman (1984) ed è inoltre simile all'algoritmo Metropolis (Metropolis e altri (1953)) Questa metodologia permette di ottenere una stima delle distribuzioni marginali a posteriori del parametro, o dei parametri di interesse. Il campionamento sostitutivo permette di trattare problemi multiparametrici attraverso una serie di problemi ad un solo parametro. In questo modo, gli scogli tipici dell'integrazione numerica multivariata, come la scelta di griglie e, di conseguenza, la necessità di effettuare trasformazioni parametriche, sono superati.

INFERENZA STATISTICA CLASSICA E TEORIA DELLE DECISIONI. Come abbiamo visto il cuore della statistica classica è la teoria inferenziale di Neyman e Pearson e la teoria delle decisioni di Wald. In questi ultimi anni si sono sviluppati gli studi intesi ad approfondire le proprietà delle procedure statistiche in modelli multiparametrici sotto forma di proprietà di ammissibilità, di teoremi per determinare classi complete di stimatori, test. In particolare con la scoperta, agli inizi degli anni 70, della relazione tra ammissibilità e processi diffusione si è potuto meglio valutare l'ammissibilità di determinate procedure.

Una importante area di ricerca è la teoria asintotica dell'inferenza statistica ed in particolare lo studio di più alti ordini di approssimazione anche là dove viene generalmente applicato il primo ordine. Strettamente collegata a questa area vi è l'approccio geometrico-differenziale alla statistica. Le teorie generali di convergenza sono state estese e rifinite.

INDAGINI CAMPIONARIE. Le ricerche più recenti hanno posto l'attenzione sull'effetto del disegno di campionamento sulla validità di procedure standard e sull'analisi di come semplici modificazioni delle procedure tengano conto del disegno in esame. Molti sono i lavori riguardanti disegni campionari complessi e problemi associati a dati incompleti. Molta attenzione è stata dedicata a disegni campionari longitudinali, all'esame del fondamentale problema degli errori di non-campionamento e quello associato con le non-risposte ed agli errori di risposta.

ROBUSTEZZA ED ANALISI ESPLORATIVA. Gli statistici hanno sempre cercato di valutare gli effetti di dati anomali (outliers) sulle loro inferenze. Sono stati sviluppati pertanto metodi di adattamento dei modelli che sono robusti contro gli outliers e che sono insensibili alla presenza degli outliers senza il bisogno di individuare tali osservazioni anomale. Questi metodi sono stati sviluppati e provati con successo nell'analisi della regressione e nel modello lineare generale. Le nuove procedure impiegate per adattare modelli non parametrici e semi parametrici non sono stati trattati con l'attenzione alla robustezza. La costruzione di procedure robuste in tale settore è ancora suscettibile di approfondimenti..

Strettamente collegata con i metodi robusti e con i metodi grafici vi è stata la crescita dell'analisi esplorativa dei dati. Questa metodologia è basata in larga parte sulle idee di J.W.Tukey.

SERIE STORICHE. I metodi di analisi delle serie storiche sia nel dominio temporale sia nel dominio delle frequenze si sono notevolmente sviluppati e perfezionati. Esistono oggi molte tecniche che si concentrano sull'utilizzo dei modelli ARMA. Le proprietà di questi modelli sono molto note e le idee di base sui problemi di stima e sulle altre procedure inferenziali sono ben definite. Gli sforzi recenti si rivolgono in molte direzioni. Una di queste direzioni è l'analisi della bontà delle procedure di inferenza quando il processo è sulla frontiera della stazionarietà. Un'altra direzione di ricerca consiste nello studiare i criteri utili alla scelta del numero dei parametri richiesti in un modello, il problema della determinazione dell'ordine. Una terza direzione è lo studio dei processi lineari non gaussiani. Molta attenzione è stata data ai metodi bayesiani ed ai modelli stato-spazio. Molti modelli di serie storiche (vedi i processi ARMA) possono essere riscritti sotto forma di modelli stato spazio, tale circostanza ha portato a sviluppare tecniche di inferenza basate sulla esatta valutazione della funzione di verosimiglianza ed ha permesso la costruzione di modelli più generali, che sono simili ai processi ARMA ma incorporano parametri che variano nel tempo.

ANALISI SEQUENZIALE. In questi ultimi anni i test sequenziali sono diventati molto soddisfacenti per lo sviluppo della teoria dei rinnovi non lineare che permette di calcolare con una buona approssimazione numerose quantità che prima richiedevano enormi calcoli. La teoria dei rinnovi non lineare ha altresì contribuito allo sviluppo di altre aree, e.g.stima sequenziale e i problemi dei punti di cambiamento.

Una seconda area che in questi ultimi anni ha visto un notevole progresso è quella dei problemi "multi -arm bandit". Questa è una speciale classe di problemi di programmazione dinamica, che permettono di determinare una assegnazione ottimale dello sforzo in una varietà di modelli di code. Tuttavia, gli algoritmi di programmazione generale dinamica richiedono un tempo di calcolo enorme. L'introduzione del "test di allocazione dinamico" ha ridotto di molto i problemi computazionali.

Il filtro di Kalman introdotto da Kalman (1960) e Kalman -Bucy (1961) è stato applicato in molti problemi di controllo stocastico. Questa teoria richiede la conoscenza della dinamica del processo sottostante. Se non si conosce la dinamica del processo è necessario ad ogni stadio stimare i parametri non noti del sistema. I risultati del filtro si basano sulla formulazione stato-spazio di un modello dinamico. All'inizio degli anni settanta il filtro di Kalman e l'approccio stato-spazio hanno ricevuto una certa attenzione anche da parte degli economisti.

INFERENZA NEI PROCESSI STOCASTICI. L'area di inferenza nei processi stocastici è piuttosto ampia con un impatto notevole sulle applicazioni ed è il centro di molte ricerche. I problemi di inferenza riguardano i più importanti e tradizionali processi stocastici (catene di Markov, processi di diffusione, processi di punto). Per i processi di punto i progressi sono stati molto rapidi grazie alla loro struttura, maneggevolezza ed al persistente contatto con le applicazioni. Fino alla metà degli anni 1970 l'inferenza per i processi di punto fu più una collezione di tecniche che un campo di ricerca coerente. L'introduzione della teoria delle martingale ha rivoluzionato l'intera area ed ha permesso l'unificazione, almeno concettuale, dell'intera problematica

SISTEMI ESPERTI PROBABILISTICI. Negli anni recenti, lo sviluppo della tecnologia informatica ha fortemente influenzato anche le discipline statistiche. Si vuole far riferimento qui, in particolare, a quella branca dell'informatica che va sotto il nome di *Intelligenza Artificiale*. Esistono due articolazioni di questa tecnologia in relazione alla Statistica. La prima prevede lo sviluppo di sistemi esperti per assistere l'utente nella fase di impostazione e realizzazione di una analisi statistica (scelta del modello, individuazione delle trasformazioni da apportare alle variabili, riconoscimento di dati anomali etc.). L'altra direzione prevede, invece, la realizzazione di sistemi esperti probabilistici nei vari campi di applicazione (medico, industriale, bancario, etc.) Si tratta, in estrema sintesi, di rappresentare un sistema complesso di variabili e l'associata "base di conoscenza" sotto forma di un grafo, i cui nodi rappresentano le variabili mentre i collegamenti tra i nodi indicano possibili forme di "causalità" non deterministica. Invero la struttura del grafo consente di rappresentare in modo compatto le relazioni di indipendenza condizionale probabilistica tra le variabili e ciò, a sua volta, permette l'acquisizione dell'evidenza empirica e la conseguente propagazione dell'incertezza nella rete in modo efficiente e probabilisticamente coerente. L'aspetto statistico interviene allorché la distribuzione congiunta dei nodi del grafo viene rappresentata in forma parametrica. Oltre alle procedure di acquisizione e propagazione dell'evidenza empirica, viene attivato un meccanismo di aggiornamento sequenziale delle distribuzioni dei parametri che consente una revisione costante delle strutture della rete in relazione ai casi osservati.

7.RIFERIMENTI BIBLIOGRAFICI.

- AA.VV., 1978, *I fondamenti dell'inferenza statistica*, Dipartimento Statistico, Università di Firenze.
- Amari, S., 1990, *Differential Geometrical Methods in Statistics*, Springer Verlag, N.Y.
- Antoniak, C., 1974, Mixtures of Dirichlet processes with applications to bayesian nonparametric problems, *Annals of Statistics*, 2, 1152-1174.
- Bahadur, R.R., 1954, "Sufficiency and statistical decision functions", *Annals of Mathematical Statistics*, 25, 423-462.
- Barlow, R.E. e Prochan, F., 1966, *Mathematical Theory of Reliability*, J.Wiley, N.Y.
- Barnett, V., 1982, *Comparative Statistical Inference*, J. Wiley N.Y.
- Bemporad, G., 1926, Sul principio della media aritmetica, *Rendiconti dell'Accademia Nazionale dei Lincei*, VI-3, 87-91.

- Benzecri, J.P., 1973, "L'analyse des données", Tome 1: *La Taxinomie*, Tome 2; *L'analyse des correspondences*, Dunod, Parigi..
- Berger J., 1984, The robust bayesian viewpoint (with discussion) In *Robustness in Bayesian Analysis* (J.Kadane ed.), North-Holland, Amsterdam.
- Berger J., 1985, *Statistical Decision Theory and Bayesian Analysis*, Springer Verlag, N.Y.
- Bernoulli D., 1738, Specimen theoriae novae de mensura sortis, *Commentarii Academiae Scientiarum. Imperialis. Petropolitanae*, Tomus V, 175-192. (ripr. in *Econometrica*, 1954, 22, 23-25).
- Bickel, P.J. e Lehmann, E.L., 1975, Descriptive statistics for nonparametric models. I Introduction, II Location, *Annals of Statistics*, 3, 1038-1069.
- Bickel, P.J. e Lehmann, E.L., 1976, Descriptive statistics for nonparametric models. III Dispersion *Annals of Statistics*, 4, 1139-1158.
- Bickel, P.J. e Lehmann, E.L., 1979, Descriptive statistics for nonparametric models. IV Spread, in *Contributions to Statistics* (J.Jureckova, ed.), Reidel Publishing Company., 33-40.
- Birnbaum, Z.W., 1948, On random variables with comparable peakedness, *Annals of Mathematical Statistics*, 19, 76-81.
- Blackwell, D., 1947, Conditional expectation and unbiased estimation, *Annals of Mathematical Statistics*, 18, 105-110.
- Bonferroni, C.E., 1924, La media esponenziale in matematica finanziaria, *Annuario del Regio Istituto Superiore di Scienze Economiche e Commerciali di Bari*, AA 23-24, 1-14.
- Box, G., 1953, Non-normality and tests on variances, *Biometrika*, 40, 318-335.
- Box, G. e Hunter J.S., 1965, Multifactor experimental Designs for exploring response surfaces, *Annals of Mathematical Statistics*, 28, 195-242.
- Box, G., e Jenkins, G., 1970, 2ed.1976, *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco.
- Broemeling, L.D., 1985, *Bayesian analysis of linear models*, Marcel Dekker Inc., N.Y.
- Brown, L.D., 1966, On the admissibility of invariant estimators of one or more location parameters, *Annals of Mathematical Statistics*, 37, 1087-1136.
- Caillez, F. e Pages, J.P., 1976, *Introduction à l'analyse des données*, S.M.A.S.H., Paris.
- Castellano, V., 1957, Contributi alle teorie della correlazione e della connessione tra due variabili, *Metron*, 18.
- Chatfield, C., 1985, The initial examination of data (with discussion)", *Journal of the Royal Statistical Society*, ser.A, 148, 214-253.
- Chatfield, C., 1986, "Exploratory data analysis", *European Journal of Operations. Research.*, 23, 5-13.
- Chatfield, C., 1988, *Problem solving. A statistician's guide*, Chapman and Hall, Londra.
- Chatfield, C., 1989, *The Analysis of Time Series. An introduction*", Chapman and Hall, Londra.
- Chatfield, C. e Collins, A.J., 1980, *Introduction to Multivariate Analysis*, Chapman e Hall, London.
- Chisini, O., 1929, "Sul concetto di media", *Periodico di Matematiche*, 1929, IX, 2, 106-116.
- Cifarelli, D.M. e Regazzini, E., 1980-81, Sul ruolo dei riassunti esaustivi ai fini della previsione in contesto bayesiano. *Rivista di Matematica per le Scienze Economiche e Sociali*, 3, 109-125 e 4, 3-11.
- Cifarelli, D.M. e Regazzini, E., 1982, Some considerations about mathematical statistics teaching methodology suggested by the concept of exchangeability. In G. Koch and F. Spizzichino, eds, *Exchangeability in Probability and Statistics*, pagg 185 - 205.
- Cochran, W.G., 1963, *Sampling Techniques*, J. Wiley, N.Y.
- Cochran, W.C. e Cox, G.M., 1957, *Experimental Design*, J. Wiley, N.Y.
- Cox, D.R e Miller, H.D., 1968, *The Theory of stochastic processes*, Chapman e Hall, London.
- Cramer, H., 1928, On the composition of elementary errors" *Skand. Aktuaris*, 11, 13-74 e 141-180.
- Cramer, H., 1946, "A contribution to the theory of statistical estimation", *Skand Aktuarietidskr*, 29, 85-94.
- De Groot, M.H., 1970, *Optimal Statistical Decisions*, Mc Graw - Hill, N.Y.
- De Groot, M.H., 1986 *Probability and Statistics*, Reading Mass, .N.Y.

- de Finetti, B. e Emanuelli F., 1967, *Economia delle Assicurazioni*, UTET, Torino.
- de Finetti, B. e Savage, L.J., 1962, Sul modo di scegliere le probabilità iniziali, *Biblioteca del Metron*, 1, 81-154.
- de Finetti, B., 1931, Sul concetto di media, *Giornale Istituto Italiano degli Attuari*, 2, 369-396.
- de Finetti, B., 1931, Sul significato soggettivo della probabilità, *Fundamenta Mathematicae*, , 289-329.17
- de Finetti, B., 1937, La prévision: ses lois logiques, ses sources subjectives, *Annales de l'Institut Henri Poincaré*, VII, I, 1-68.
- de Finetti, B., 1952, Sulla preferibilità, *Giornale degli Economisti e Annali di Economia*, 11, 685-709.
- de Finetti, B., 1959, *Induzione e Statistica*, CIME, Cremonese.
- de Finetti, B., 1964, *Teoria delle Decisioni*, in *Lezioni di Metodologia Statistica per Ricercatori*, vol.VI, Università degli Studi di Roma.
- de Finetti, B., 1970, *Teoria della probabilità. Sintesi introduttiva con appendice critica*, Einaudi., Torino.
- de Finetti, B., 1972, *Probability, Induction and Statistics*, J Wiley N.Y.
- Doksun, K., 1974, Tailfree and neutral random probabilities and their posterior distribution *Annals of Probability*, 2, 183-201.
- Edwards, W., Lindman, H. e Savage, L.J., 1963, Bayesian statistical inference for psychological research, *Psychological Review*, 70, 3, 193-242.
- Efron, B., 1979, Bootstrap methods: another look at the jackknife *Annals of Statistics*, 1, 1-26.
- Efron, B., 1982, *The jackknife, the bootstrap and other resampling plans*, SIAM..
- Elderton, W.P.e Johnson, N.L. 1969, *System of Frequency Curves*, Cambridge University Press, Cambridge
- Erickson, B.H. e Nosanchuck, R. A., 1977, *Understanding Data*, Mc Graw - Hill, N.Y.
- Esary, I.D. e Proschan, F., 1972, Relation among some concepts of bivariate dependence *Annals of Mathematical Statistics*, 43, 651-655.
- Esary, I.D., Proschan, F. e Walkup, D.W., 1967, Association of random variables with applications *Annals of Mathematical Statistics*, , 38, 1466-1474.
- Ferguson, T.S., 1967, *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, N.Y.
- Ferguson, T.S., 1973, A bayesian analysis of some nonparametric problems, *Annals of Statistics*, 1, 209-230.
- Ferguson, T.S., 1974, Prior distributions on spaces of probability measures *Annals of Statistics*, 2, 615-619.
- Fishburn, P.C., 1970, *Utility Theory for Decision Making*, J.Wiley, N.Y.
- Fishburn, P.C., 1982, *The Foundations of Expected Utility*, Reidel, Dordrecht.
- Fisher, R.A., 1922, The goodness of fit of regression formulae and distribution of regression coefficient" *Journal of the Royal. Statistical Soc.iety*, ser. A, 222, 309-368.
- Fisher, R.A., 1990, *Statistical Methods, Experimental Design and Scientific Inference*. A re-issue of *Statistical methods for research workers.- The design of Experiments statistical- Methods and Scientific Inference"*, Oxford Science Publications.Oxford.
- Fraser, D.A.S., 1957, *Nonparametric methods in statistics*, J. Wiley, N.Y.
- Fréchet, M., 1951, Sur les tableaux de corrélation dont les marges sont données, *Annals Univ. de Lion.*, 14, 53-77.
- Galton, F., 1877, Typical laws of heredity., *Proc.R.Inst.G.Brit.*, 8, 282-301.
- Galton, F., 1886, Regression towards mediocrity in hereditary stature", *Journ of the Anthropological Inst.*, Miscellanea, XV, 246-263.
- Galton, F., 1889, *Natural Inheritance"* Macmillan, Londra.
- Gelfand, A.E. e Smith, A.F.M., 1990, Sampling based approaches to calculating marginal densities, *Journal of the American Statistical Association*, 85, 398-409.

- Geman, S. e Geman, D. , 1984, Stochastic relaxation, Gibbs distributions and the bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.
- Gini, C., 1914-1915, Indici di omofilia e di rassomiglianza e loro relazione col coefficiente di correlazione e con gli indici di attrazione, in *Atti Ist. Veneto*, vol. LXXIV.
- Gini, C., 1943, Intensità e precisione della connessione statistica, *Statistica*, 7.
- Gini, C., 1955, *Variabilità e Concentrazione*, Libreria Veschi, Roma.
- Good, I.J., 1956, On the estimation of small frequencies in contingency tables, *Journal of the Royal Statistical Society, ser. B*, 18, 113-1.
- Greenacre, M. 1984, *Theory and Applications of Correspondence Analysis*, Academic Press, London.
- Haldane, J. B. S., 1945, On a method of estimating frequencies, *Biometrika*, 33, 222-225.
- Halmos, P.R. e Savage, L.J., 1949, Application of the Radon-Nikodim Theorem to the theory of sufficient statistics *Annals of Mathematical Statist.*, 20, 225-241.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. e Stahel, W.A., 1986, *Robust statistics: the approach based on influence funtions*, J. Wiley , N.Y.
- Harrison, P.J. e Stevens, C.F., 1976, Bayesian forecasting (with discussion), *Journal of the Royal Statistical Society, . Ser B*, 205-247.
- Harvey, A.C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.
- Herzel, A., 1961, Sulla definizione dei concetti di media e di interpolazione, *Metron*, XXI, 126-138.
- Hoeffding, W., 1940, Masstabinvariante korrelation-theorie, *Schriften of Mathematical Inst. Univ. Berlin.*, 5, 181-233.
- Hoeffding, W., 1948, A class of statistics with aymptotically normal distribution, *Annals of Mathematical Statistics*, 19, 293-325.
- Hotelling, H., 1933, Analysis of a complex statistical variable into principal components, *Journal of Educational Psychology*, 26.
- Hotelling, H., 1935, The most predictable criterion, *Journal of Educational Psychology*, 26.
- Hotelling, H., 1936, Relation between two sets of variables *Biometrika*, 28.
- Hotelling, H., 1941, Experimental determination of the maximum of a function, *Annals of Mathematical Statistics*, 12, 20-45.
- Hotelling, H., 1944, Some improvements in weighing and other experimental techniques, *Annals of Mathematical Statistics*, 15, 297-306.
- Huber, P., 1972, Robust Statistics: a review, *Annals of Mathematical Statistics*, 43, 1041-1067.
- Huber, P., 1981, *Robust statistics*, J. Wiley, N.Y.
- Jeffreys, H., 1961, *Theory of Probability* (III ed.), Clarendon Press, Oxford.
- Kalman, R.E. e Bucy, R.S., 1961, New results in linear filtering and prediction theory *Trans. ASME J. Basic Engineering*, 83, 95-108.
- Kalman, R.E., 1960, A new approach to linear filtering and prediction problems, *Trans. ASME, J. Basic Engineering*, 82, 35-45.
- Kalman, R.E., 1963, New methods in Wiener filtering theory, in G.L. Bogdanoff J.Kozin (eds) *Proc of first Symp. on Eng. Appl. of Random Function Theory and Probability*, J. Wiley, N.Y.
- Kendall, M. , 1938, A new measure of rank correlation, *Biometrika*, 30, 81.
- Kendall, M.G. e Stuart, A. e Ord, J.K., 4 ed. 1983, *The advanced theory of Statistics*, vol 1-2 -3, C. Griffin Londra.
- Kiefer, J., 1958, On the nonrandomized optimality and randomized non-optimality of simmetrical design, *Annals of Mathematical Statistics*, 29, 657-699.
- Kiefer, J., 1959, Optimum experimental designs, *Journal of the Royal Statistical Society, Ser. B*, 21, 272-319.
- Kiefer, J. e Wolfowitz, J., 1959, Optimum design in regression problems *Annals of Mathematical Statistics*, 30, 271-294.
- Kiefer, J. e Wolfowitz, J., 1960, The equivalence of two extremum problems, *Canadian Journal of Mathematics*, 12, 360-366.

- Kolmogorov, A. N., 1933, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer Verlag, Berlin.
- Kolmogorov, A.N., 1930, Sur la notion de Moyenne, *Rendiconti dell'Accademia Nazionale dei Lincei*, 9, 388-391.
- Kolmogorov, A.N., 1933, Sulla determinazione empirica di una legge di distribuzione, *Giornale Istituto Italiano degli Attuari*, 4, 83-91.
- Kolmogorov, A.N., 1941, "Interpolation and extrapolation of stationary random sequences", *Bull. de l'Academie des Sciences de U.S.S.R.*, Ser Marth, 5, 3-15.
- Kotz, S. e Johnson, N. L., 1982-1989, *Encyclopedia of Statistical Sciences*, vol 1-9, J.Wley, N.Y.
- Kruskal, W. e Wallis, W., 1952, Use of ranks in one criterion variance analysis, *Journal of the American Statistical Association.*, 47, 583-612.
- Kruskal, W. e Wallis, W., 1953, Use of ranks in one criterion variance analysis, *Journal of the American Statistical Association*, 48, 907-911.
- Lecoutre, J.P.e Tassi, P., 1987, *Statistique non parametrique et robustesse*, Economica, Parigi.
- Lee, A.J., 1990, U-Statistics, Marcel Dekker, Inc., N.Y.
- Lehman, E.L., 1966, Some concepts of dependence, *Annals of Mathematical Statistics*, 37, 1137-1153.
- Lehmann, E.L., 1959, 2ed 1986, *Testing Statistical Hypoteses*, J. Wiley, N.Y.
- Lehmann, E.L., 1983, *Theory of Point Estimation*, J. Wiley, N.Y.
- Lehmann, E.L. e Scheffé, H., 1950, Completeness, similar regions and unbiased estimation, *Sankya*, Ser. A, 10, 305-340.
- Lindley, D.V., 1962, Discussion on Prof. Stein paper, *Journal of the Royal Statistical Society*, Ser.B, 24, 265-296.
- Lindley, D.V., 1970, *Introduction to Probability and Statistics*, I.Probability, II.Inference., Cambridge University Press, Cambridge.
- Lindley, D.V., 1972, *Bayesian Statistics. A Review*, SIAM. Philadelphia.
- Mann, H.B.e Whitney, D.R., 1947, On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18, 50-60.
- Mardia, K.V., 1970, *Families of bivariate distribution* Griffin and co., London.
- Mardia, K.V., Kent, T. e Bibby, M., 1979, *Multivariate analysis*, Academic Press, London.
- Maritz, J.R. e Lwin, T. , 2ed. 1989, *Empirical Bayes Methods*, Chapman e Hall, London.
- Mc Cullagh, P. e Nelder, J.A., 1983, 2 ed.1989, *Generalized Linear Models*, Chapman and Hall, Londra.
- Mc Neil, D. R., 1977, *Interactive data analysis*, J. Wiley, N.Y.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., e Teller, E. , 1953, Equations of state calculations by fast computing machines, *J. Chem. Phys.*, 27, 720-733.
- Nagumo, M., 1930, "Uber eine Klasse von Mittelwerten", *Japanese Journal of Mathematics.*, 7, 72-79.
- Nelder, J.A. e Wedderburn, R.W., 1972, Generalized linear models, *Journal of the Royal Statistical Society.*, A, 135, 370-384.
- Neyman, J. e Pearson, E.S., 1928, On the use and interpretation of certain test criteria *Biometrika*, 20 A, 175-240, 263-294.
- Neyman, J. e Pearson, E.S., 1933, On the problem of the most efficient tests of statistical hypothesis *Phil. Trans. Roy Soc Ser. A*, 231, 289-337.
- Neyman, J., 1934, On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection, *Journal of the Royal Statistical Society*, 97, 558-625.
- Neyman, J., 1935, Su un teorema concernente le cosidette statistiche sufficienti, *Giornale dell'Istituto Italiano degli Attuari.*, 64, 320-334.
- Oja, H., 1981, On location, scale, skewnes and kurtosis of univariate distridutions, *Scandinavian Journal of Statistics.*, 8, 154-168
- Owen, D.B., 1976, *On the History of Statistics and probability*, Marcel Dekker. N.Y.

- Pearson, K., 1894, Contribution to the mathematical theory of evolution, *Philos. Trans. R. Soc. London*, A185, 719-810.
- Pearson, K., 1895, Contribution to the mathematical theory of evolution II. Skew variation in Homogeneous material *Philos. Trans. R. Soc. London*, A186, 343-414.
- Pearson, K., 1900, On a criterion that a given system of deviation from a probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen in random sampling, *Phil. Mag. Ser.*, 157-175.
- Pearson, K., 1901, On lines and planes of closet fit to system of points in space, *Philosophical Magazine and Journal of Science*, 2, VI.
- Pearson, K., 1920, Notes on the History of correlation, *Biometrika*, 13.
- Pearson, K., 1920, The fundamental problem of practical statistics, *Biometrika*, 13, 1-16.
- Prietsley, M.B., 1981, *Spectral Analysis and Time series*, vol 1-2, Academic Press, Londra.
- Quenouille, M., 1949, Approximate tests of correlation in time series *Journal of the Royal Statistical Society*, . Ser B., 11.
- Quenouille, M., 1956, Notes on bias in estimation, *Biometrika*, 43, 353-360.
- Raiffa, H. e Schlaifer, R., 1961, *Applied statistical decision theory*, Graduate School of Business Administration, Harvard University.
- Ramsey, F.P., 1925, *Truth and probability* in *The foundations of mathematics and other logical essays*, Kegan, Londra, 1931.
- Rao, C.R., 1945, Information and accuracy attainable in the estimation of statistical parameters, *Bull. Calcutta Math. Soc.*, 17, 81-91.
- Rao, C.R., 1949, A note on unbiased and minimum variance estimates, *Calcutta Statist. Bull.*, 3, 36.
- Rao, C.R., *Linear Statistical Inference and its Applications*, J. Wiley, N.Y.
- Robbins, H., 1955, An empirical Bayes Approach to Statistics, in *Proc. 3rd. Berkeley Symposium of Math. Statist. and Prob.* vol. 1, 157-163. University of California Press. Berkeley.
- Rolph, J.E., 1968, Bayesian estimation of mixing distributions, *Annals of Mathematical Statistics*, , 39, 1289-1302.
- Rubin, D.B., 1981, The bayesian bootstrap, *Annals of Statistics*, 9, 130-134.
- Salvemini, T., 1939, Sugli indici di omofilia, in *Atti della prima riunione della società italiana di Statistica*, Pisa, 1939.
- Salvemini, T., 1943, Sul calcolo degli indici di concordanza tra due caratteri quantitativi, *Atti VI Riun. Sc. Soc. It. Statist.*
- Salvemini, T., 1959, *Regressione e correlazione*, Boringhieri, Torino.
- Savage, L.J., 1954, *The Foundations of Statistics*, J. Wiley, N.Y.
- Savage, L.J., 1962, *Bayesian Statistics. In Decision and Information Process* " Mc Millan, N.Y.
- Scheffé, H., 1956, A mixed model for the analysis of variance, *Biometrika*, 40, 87-104.
- Scheffé, H., 1956, Alternative models for the analysis of variance, *Annals of Mathematical Statistics.*, 27, 251-271.
- Scheffé, H., 1958, Fitting straight lines when one variable is controlled, *Journal of the American Statistical Association*, 53, 106-117.
- Scheffé, H., 1959, *The Analysis of Variance*, J. Wiley, N.Y.
- Serfling, R.J., 1980, *Approximation of Mathematical Statistics*, J. Wiley.
- Siegmund, D., 1985, *Sequential Analysis*, , Springer Verlag, N.Y.
- Silvey, S.D., 1980, *Optimal Design*, Chapman and Hall, N.Y.
- Smith , K., 1918, On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations, *Biometrika*, 12, 1-85.
- Spearman, C., 1904, The proof and measurement of association of two things, *Amer. J. Psychol.*, 15, 72-101.
- Stein , C., 1945, A two sample test for a linear hypothesis whose power is independent of the variance" *Annals of Mathematical Statistics*, 16, 243-258.

- Stein, C., 1956, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *Proc. III Berkley Simp. on Math. Stat. and Prob.*, I, 197-206, Berkley Univ. of Cal. Press.
- Stein, C., 1962, "Confidence sets for the mean of a multivariate normal distribution", *Journal of the Royal Statistical Society, Ser.B*, 24, 265-296.
- Stigler, S.M., 1986, *The History of Statistics*, Harvard University Press.
- Tanner, M.A. e Wong, W.H. , 1987, The calculation of posterior distributions by data augmentation (with discussion), *Journal of the American Statistical Association*, 82, 528-550.
- Thompson, 1933, .On the likelihood that one unknown probability exceeds another in view of the evidence of two samples, *Biometrika*, 25, 285-294.
- Thurstone, L.L., 1931, Multiple factor analysis, *Psychological Review*, 38.
- Tukey, J., 1958, Bias and confidence in not-quick large sample *Annals of Mathematical Statistics*, 29, 614.
- Tukey, J.W., 1970 2 ed. 1977, *Exploratory data analysis*, Addison, Wesley.
- Van Zwet, W.R., 1964, *Convex transformations of random variables*, Math. Centrum, Amsterdam.
- Von Mises , R., 1931, *Wahrscheinlichkeitrechnung*, Leipzig.
- Von Mises R., 1957, *Probability, Statistics and Truth*, Dover Publications, Inc., N.Y.
- Von Neumann, J.e Morgenstern, O., 1944, *Theory of games and economic behavior*, Princeton University Press.
- Wald, A., 1943, On the efficient design of statistical investigations, *Annals of Mathematical Statistics*, 42, 1035-1053.
- Wald, A., 1947, *Sequential Analysis*, J. Wiley, N.Y.
- Wald, A., 1950, *Statistical Decision Functions*, J. Wiley, N.Y.
- West, M. e Harrison, J., 1989, *Bayesian forecasting and Dynamic models*", Springer Verlag, N.Y.
- Wiener, N., 1949, *Extrapolation, Interpolation of stationary time series*, MIT Press, Cambridge Wiley, N.Y.
- Wilcoxon, F., 1945, Individual comparison by ranking methods, *Biometrika*, 1, 80-83.
- Yanagimoto, T., 1973, Families of positively dependent random variables, *Annals Inst. Statist. Math.*, 24, 559-573.
- Zacks, S., 1971, *The Theory of Statistical Inference*, J. Wiley, N.Y.

PIETRO MULIERE
Dipartimento di Economia Politica
e Metodi Quantitativi
Università di Pavia.

Pavia, settembre 1991.