5 novembre 2021

Corso di aggiornamento professionale/perfezionamento

Ente organizzatore CISA con collaborazione dell'Università di Firenze

La razionalità nel giudicare

Modulo 3

La prova statistica nel processo penale:
DNA e dattiloscopia

Giuristi

Paolo Garbolino

(ordinario di logica e filosofia della scienza, Università IUAV di Venezia)

Giuseppe Gennari

(giudice del Tribunale di Milano)

Quantitativi

Luigi Vannucci

(già ordinario di Matematica per le Applicazioni economiche e finanziarie, Università di Firenze)

Fabio Corradi

(ordinario di Statistica, Università di firenze)

Il ruolo che la statistica può svolgere nell'attività giudiziaria

di Luigi Vannucci

Disponibile per gli iscritti al corso a www.fondazioneforensefirenze.it
FONDAZIONE per la FORMAZIONE FORENSE
dell'ORDINE degli AVVOCATI di FIRENZE

Da matematico non gioco in casa se considero il titolo del modulo 3 "Prova statistica nel processo penale: DNA e dattiloscopia" e mi limiterò (sic, visto l'improbo compito) a illustrare l'ambito d'impiego della statistica in generale e a fare intravedere gli spazi per un suo utilizzo nella attività giurdiziaria.

1 Le sigle che suddividono la conoscenza

L'analisi delle problematiche che ogni realtà (fisica, chimica, biologica, sociale, giuridica, economica, finanziaria, organizzativa, gestionale...) può proporre è "raccontata" per sé e agli altri in prima approssimazione usando i vocaboli di una lingua, come l'italiano, l'inglese, il cinese... , ma quando questa realtà vuole essere davvero conosciuta per eventualmente agire su di essa per il conseguimento di auspicati obiettivi occorre corredare la lingua di **lingaggi tecnici**. In ambito accademico i diversi linguaggi tecnicisi richiamano spesso con sigle di tre lettere: le elenco nel seguente ordine, separandone alcune da due punti, anziché dalla sola virgola

MAT; STA, INF; FIS, CHI, GEO; BIO, MED, FAR, BOT, ZOO, VET;
SOC, JIU, ECO, STO; FIL; LET; ART...

MAT sta per **matematica**. MAT è isolata con un punto e virgola perchè la matematica è il solo linguaggio tecnico che non insiste necessariamente, se non quando la si applica, su una realtà empirica. Il matematico puro può porsi appassionanti problemi per realtà convenzionali che i matematici stessi creano, senza mai affrontare questioni legate a realtà empiriche di tipo fisico, chimico, biologico, sociale, giuridico, economico, organizzativo...

Si dà il caso che le invenzioni/scoperte prodotte dai matematici, puri o applicati che siano, si rivelino poi estremamente utili per foggiare modelli matematici atti a descrivere con compiutezza molte problematiche della realtà empirica per facilitarne la loro efficace soluzione.

In particolare i modelli matematici rappresentativi delle varie realtà considerabili si distinguono in **descrittivi** e **descrittivi & normativi**: i primi hanno il solo scopo di rappresentare come opera la natura (in contesti di fisica, di chimica, di biologia, di sociologia, di economia...); i secondi, oltre a rappresentare il funzionamento naturale di dati sistemi, prevedono anche che **uno o più soggetti siano interessati** a trarne vantaggi con appropriate scelte individuali.

STA sta per **statistica** e INF sta per **informatica**. Anche queste discipline non hanno una realtà empirica specifica a cui si riferiscono e approntano strumentazioni di ordine generale per essere più performanti in ogni ambito applicativo.

La statistica in particolare aiuta, con la cosiddetta **statistica descrittiva** l'ideatore dei modelli con l'approntamento di **indicatori di sintesi** atti a rappresentare la realtà spesso archiviata in una caterva di dati disordinati e confusi e in un secondo momento, con la cosiddetta **statistica inferenziale**, si fa carico di validare i modelli, che il matematico applicato costruisce, analizzando la distanza tra le previsioni che i modelli consentono e i dati informativi, vecchi e nuovi via via acquisiti.

L'informatica con la sua strumentazione in continuo ampliamento nelle pìù diverse direzioni si rivela ormai indispensabile non solo in ogni ambito della ricerca scientifica, ma anche capace di incidere significativamento sul funzionamento dei sistemi e in particolare di quelli socio-economici, **creando nuove forme di realtà empirica**, con l'insorgenza di nuovi problemi da risolvere: il pianeta Terra non ha mai avuto in un lasso di tempo di così pochi decenni le trasformazioni a cui stiamo assistendo anche nel nostro quotidiano modo di vivere.

È da sottolineare come il **connubio tra informatica e statistica** abbia prodotto l'efficace tecnica di analisi e risoluzione dei problemi detta **simulazione**: l'informatica dà la possibilità di sperimentare in pochi secondi i modelli ideati quando questi non siano analizzabili dal **punto di vista analitico** e neanche dal **punto di vista numerico esatto** con la produzione di un voluminoso insieme di risultanze, che solo la statistica consente di sintetizzare e di procedere ad adeguate induzioni.

Per le altre sigle lascio al lettore intuire e commentare: dico solo che ho distinto gli **ambiti di indagine in freddi** (FIS, CHI, GEO), **della vita** (BIO, MED, FAR, BOT, ZOO, VET), **sociali** (SOC, JIU, ECO, STO), **filosofici, letterari, artistici** (FIL; LET; ART...). È pleonastico ricordare che ogni scienziato (fisico, ..., biologo, ..., sociologo, ..., filosofo, ...) non è detto che debba ricorrere ad ausili esterni al proprio ambito, ma può già lui essere contemporaneamente matematico, statistico, informatico...

2 Statistica

Sono troppe le **definizioni della statistica** (già nel 1934 Willcox ne aveva elencate ben 115). Per Yule (George Udny Yule 1871-1951) essa è un **metodo per l'esposizione e l'interpretazione dei dati quantitativ**i; per Fisher (Ronald A. Fisher 1890-1962) essa è una **branca delle matematiche applicate**, quella che studia i fenomeni osservazionali; per Boldrini (Marcello Boldrini 1890-1969) è l**a storia empirica delle scienze**; per Kendall (Maurice Kendall 1907-1983) è la **metodologia del trattamento dei dati riferiti a collettivi**; per Wald (Abraham Wald 1902-1950) è **la teoria delle decisioni statistiche**. In tutte queste definizioni ritroviamo aspetti «curati» dalla statistica come oggi la si intende.

2.1 Statistica descrittiva

Si considera acquisita una collezione di dati empirici, dati riferibili ai più svariati campi di applicazione: demografia, biometria, economia, scienze naturali, sperimentazioni in laboratorio, processi produttivi, indagini di mercato, controllo di qualità... Qui si innesta la cosiddetta **statistica descrittiva** che è quell'insieme di regole da seguire per arrivare a presentare **sintesi efficaci per la conoscenza di una data realtà**. Sono innumerevoli le esemplificazioni in cui si ha necessità di tali sintesi: i censimenti, i sondaggi di opinione, gli exit pool, l'auditel, le tavole di sopravvivenza, gli incidenti automobilistici o ogni altro sinistro oggetto di contratti assicurativi, le serie storiche delle quotazioni dei titoli azionari o dei prezzi, i decessi secondo le cause di morte, il peso e la statura dei coscritti alla leva militare, i nati per sesso, i parti plurimi, la composizione del nucleo familiare al decesso del pensionato o dell'attivo, il numero di particelle emesse da una data sostanza in un prefissato intervallo di tempo, le frequenze delle estrazioni dei numeri sorteggiati nel lotto, l'esito di terapie mediche, la statistica stellare e chi più ne ha più ne metta.

Non sto qui a ricordare tutte le «trovate» che nel tempo sono diventate patrimonio comune di tutti coloro che professionalmente curano queste sintesi: diagrammi, istogrammi, raffigurazioni dinamiche come per le previsioni metereologiche in televisione, in scala logaritmica, in doppia scala logaritmica, in coordinate polari, rappresentazioni con poligoni, rappresentazioni con gradazioni di colore (queste ultime due per superare le difficoltà che insorgono quando alla resa dei conti bisogna comunque agire su un foglio di carta o su un video per «raccontare» di punti che stanno in spazi a più di due dimensioni). Colpiscono nell'ambito delle indagini sulle quotazioni dei titoli le cosiddette candelette giapponesi, largamente utilizzate nell'analisi tecnica delle serie delle quotazioni dei titoli.

2.1.1 Indicatori di sintesi per variabili quantitative

A proposito di sintesi conoscitive relative a masse di dati empirici mi pare rilevante l'impostazione del problema della **media secondo Chisini** (Oscar Chisini 1889-1967). Diceva Chisini che se al posto di n dati rilevati, $x_1, x_2, ..., x_n$, se ne vuol indicare uno soltanto, x, questo dovrebbe essere determinato in funzione dell'**invariante** che si intende associare ai dati stessi. Se l'invariante può essere rappresentato da una funzione dei dati, $f(x_1, x_2, ..., x_n)$, allora la media idonea è la soluzione (o una delle soluzioni) dell'equazione in x

$$f(x, x, ..., x) = f(x_1, x_2, ..., x_n)$$

Per esempio si dovrà usare la **media aritmetica** se

$$f(x_1, x_2, ..., x_n) = x_1 + x_2 + ... + x_n$$

si dovrà usare la **media geometrica** se

$f(x_1, x_2, ..., x_n) = x_1 \cdot x_2 \cdot ... \cdot x_n$

si dovrà usare la **media armonica** se

$f(x_1, x_2, ..., x_n) = \frac{1}{x_1} + \frac{1}{x_2} + ... + \frac{1}{x_n}$

e così via.

Talvolta l'**indice statistico** considerato deve soddisfare certe naturali proprietà come quelle ad esempio di essere **interno ai dati**, **monotono rispetto ad ogni dato, invariante per traslazioni**, **invariante per scala...** .

Talvolta la sintesi dei dati può riguardare la ricerca di funzioni analitiche o empiriche che poi possano utilizzarsi a fini di **interpolazione, perequazione, estrapolazione**: si segnalano metodologie basate su **medie mobili** per la perequazione dei dati, il **metodo dei minimi quadrati** per ricercare funzioni analitiche con cui procedere poi a perequare o a interpolare o ad estrapolare.

2.1.2 Un esempio di collezione di dati relativi all'attività giudiziaria

Il Ministero della Giustizia ha un ufficio statistica che produce sistematicamente una messe imponente di dati su cui l'arte di individuare gli appropriati indicatori di sintesi potrebbe essere esercitata con il fine non solo di illustrare quanto storicamente si è verificato, ma anche di trarre suggerimenti per azioni volte a incrementare l'efficacia del sistema giustizia. A titolo esemplificativo si riporta il **Monitoraggio della giustizia penale in Italia - Anni 2003 - 2021 - Fonte: Ufficio I - Statistiche Giudiziarie - Ministero della giustizia**. Si tratta del numero di procedimenti penali pendenti a fine periodo.

Da notare che i dati sono solo **dati di giacenza a certe date** e non consentono di valutare i volumi di procedimenti completati nei singoli periodi: occorrerebbe considerare anche i **dati di flusso in entrata e in uscita** di ogni periodo e disaggregati per tipo di organo giudicante per quantificare l'efficenza del sistema giudiziario nel suo complesso e delle sue sottoaggregazioni. Ovviamente tale quantificazione dovrebbe tenere in considerazione anche il **numero di magistrati** e la **complessità dei percorsi giudiziari dei vari segmenti della giustizia penale**.

TAB. 1 - Monitoraggio della giustizia penale in Italia - Anni 2003 - 2021
Fonte: Ufficio I - Statistiche Giudiziarie - Ministero della giustizia
Dati al 20 luglio 2021

Numero di procedimenti penali pendenti a fine periodo

Anno	Corte Cassazione	Corte di appello	Tribunale ordinario	Tribunale minorenni	Giudice di pace	totale
2003	31.140	130.395	1.139.953	36.061	60.379	1.397.928
2004	30.953	135.322	1.184.940	39.847	76.113	1.467.175
2005	32.862	140.822	1.165.732	38.671	81.208	1.459.295
2006	37.439	154.844	1.194.511	40.407	105.987	1.533.188
2007	33.177	157.996	1.195.646	38.568	116.857	1.542.244
2008	28.340	170.307	1.198.005	34.608	116.396	1.547.656
2009	25.560	199.470	1.184.591	35.939	137.177	1.582.737
2010	29.381	219.392	1.221.444	37.653	147.913	1.655.783
2011	30.561	238.008	1.239.629	39.553	154.306	1.702.057
2012	31.289	249.319	1.293.360	42.768	160.709	1.777.445
2013	31.871	266.475	1.313.246	43.126	164.967	1.819.685
2014	34.143	260.748	1.302.395	43.040	161.039	1.801.365
2015	35.984	257.504	1.313.577	42.274	153.220	1.802.559
2016	30.354	268.445	1.187.734	38.539	133.439	1.658.511
2017	30.236	275.596	1.165.339	38.890	118.759	1.628.820
2018	24.609	271.247	1.157.500	40.145	108.421	1.601.922
2019	23.579	263.317	1.152.240	41.810	101.073	1.582.019
2020	24.473	271.640	1.185.957	40.901	108.167	1.631.138
2021 I tr.	21.940	266.454	1.178.927	40.756	105.167	1.613.244

2.2 Statistica inferenziale

Una volta ottenuti i dati su una realtà di cui interessa approfondire la conoscenza (e talvolta occorre partire addirittura dalla progettazione di ogni dettaglio del tipo di indagine da realizzare) si distinguono poi due casi: quello in cui i dati riguardano un **sottoinsieme proprio di elementi dell'intera collettività** e quello in cui i dati riguardano **tutti gli elementi della collettività**. Se nel secondo caso il problema è ancora quello della efficace organizzazione dei dati in coerenza con gli obiettivi conoscitivi da perseguire, nel primo caso si apre l'ardua questione di inferire (si parla appunto di **statistica inferenziale**) corrette valutazioni sull'intera collettività partendo dalle informazioni offerte dal cosiddeto **campione**, tenuto conto dello **schema di campionamento** seguito: si segnalano ad esempio il campionamento bernoulliano (o con reibussolamento), il campionamento senza reimbussolamento, il campionamento a grappoli, il campionamento a stadi successivi, il campionamento stratificato, ecc.

Tralascio l'illustrazione dei termini appena introdotti, che necessiterebbe di decine di ore, schematizzo i quattro i punti di vista che si possono considerare quando si parla di statistica inferenziale

- **aggiornare le probabilità delle ipotesi** sulla scorta di evidenze empiriche acquisite: stime di massima verosimiglianza, inferenza bayesiana.
- **stimare grandezze di interesse per una intera collettività** disponendo dei dati riferiti a un sottoinsieme proprio della stessa: teoria degli stimatori.
- **considerare due o più ipotesi in alternativa** e fissare delle soglie di probabilità per la loro accettazione/rifiuto: test di ipotesi.
- rappresentare in una tabella a doppia entrata **guadagni/perdite in corrispondenza di decisioni da prendere**, in riga, e stati del mondo in alternativa, in colonna e fissare **criteri di scelta**: decisioni statistiche.

2.2.1 Aggiornare le probabilità delle ipotesi sulla scorta di nuove informazioni

Si applicano tutti i classici risultati del calcolo delle probabilità che brevemente richiamo, senza soffermarmi sulla simbologia peraltro già introdotta dai proff. L. Peccati e M. Galeotti nel Modulo 1 "Le regole della probabilità: il ragionamento del giudice tra prova scientifica e massima di esperienza". Per esemplificare

(i) $P(\overline{E}) = 1 - P(E)$

detta **probabilità dell'evento contrario**.

(ii) $P(E_1 + E_2 + ... + E_n) =$

$= \sum P(E_i) - \sum P(E_iE_j) + \sum P(E_iE_jE_h) + ... + (-1)^{n-1}P(E_1E_2...E_n)$

detta anche **formula (o principio) di inclusione-esclusione**, dove le sommatorie vanno estese rispettivamente a tutti i possibili prodotti di $1, 2, ..., n$ eventi diversi da scegliersi tra gli n eventi considerati e i termini di ogni sommatoria sono pertanto rispettivamente $\binom{n}{1}, \binom{n}{2}, ..., \binom{n}{n}$. Nel caso in cui gli n eventi siano **a due a due incompatibili** (sono allora incompatibili anche tre a tre, quattro a quatro e così via) si ha come caso particolare della (ii) la cosiddetta additività finita della probabilità per n eventi a due a due incompatibili

$$P(E_1 + E_2 + ... + E_n) = \sum_{i=1}^{n} P(E_i).$$

Si ricordano anche

$P(E_1 + E_2 + ... + E_n) = 1 - P(\overline{E_1}\ \overline{E_2}...\overline{E_n})$

$P(E_1 E_2 ... E_n) = 1 - P(\overline{E_1} + \overline{E_2} + ... + \overline{E_n})$

diretta conseguenze delle leggi (ben note nella logica e nell'insiemistica) di de Morgan (Augustus de Morgan 1806-1871), che possono essere valide alternative al principio di inclusione-esclusione. Altri esempi

$P(E_1 + E_2) = P(E_1) + P(\overline{E_1} E_2)$

$P(E_1 \triangle E_2) = P(E_1) + P(E_2) - 2P(E_1 E_2)$

e chissà quanti altri si potrebbero utilmente considerare.

Nell'inferenza il ruolo basilare è giocato dalla **probabilità condizionata**. L'**evento** E **condizionato all'evento** H, in simboli $E \mid H$, è l'evento si verifica se si verifica il prodotto HE, non si verifica se $H\overline{E}$ e si "annulla" se $\overline{H}$. Basta un unico livello di condizionamento essendo

$(E \mid H_1) \mid H_2 = E \mid H_1 H_2$

Se nella classe degli eventi da probabilizzare si aggiungono gli eventi condizionati allora per la valutazione di probabilità di questi eventi condizionati si ha

$$P(E\,|H\,) = \frac{P(HE)}{P(H)} \text{ se } P(H) > 0$$

Si noti che tutti gli eventi possono essere considerati condizionati essendo ovviamente $E = E\,|\Omega$ dove Ω è l'evento certo. Valgono le tre seguenti proprietà nella valutazione delle probabilità in cui intervengono eventi condizionati

(iii) $P(EH) = P(H)P(E \mid H) = P(E)P(H \mid E)$

(iv) $P(E_1 E_2 ... E_n) =$
$= P(E_1) \cdot P(E_2 \mid E_1) \cdot P(E_3 \mid E_1 E_2) \cdot ... \cdot P(E_n \mid E_1 E_2 ... E_{n-1})$

nota anche come teorema delle **probabilità composte;**

(v) $P\,(E) = P(H_1)P(E \mid H_1) + P(H_2)P(E \mid H_2) + ... + P(H_n)P(E \mid H_n)$

se $H_1, H_2, ..., H_n$ costituiscono una partizione dell'evento certo. Questo risultato è noto anche come teorema delle **probabilità totali** e segue dalle ovvie uguaglianze

$E = E\Omega = E\,(H_1 + H_2 + ... + H_n) = EH_1 + EH_2 + ... + EH_n$

Dalla $P(EH) = P(H)P(E \mid H) = P(E)P(H \mid E)$ in precedenza considerata segue il celebre **teorema di Bayes,** Thomas Bayes (1702-1761), riportato postumo in *An Essay towards Solving a Problem in the Doctrine of Chances* 1763.

(vi) $P(H \mid E) = P(H) \cdot \dfrac{P(E \mid H)}{P(E)}$

detto anche **teorema delle probabilità delle cause** o teorema delle probabilità a posteriori. Il teorema mostra come deve modificarsi la probabilità dell'«ipotesi» H se si sa che si è verificata l'« evidenza» E: $P(H \mid E)$ è anche detta **probabilità a posteriori**, $P(H)$ è detta allora **probabilità a priori**, $P(E \mid H)$ **verosimiglianza** e $\dfrac{P(E \mid H)}{P(E)}$ **rapporto di verosimiglianza (*likehood ratio*)**.

Un'altra formulazione del teorema è utile nelle applicazioni quando le ipotesi H_1, H_2, ..., H_n, con $n \geq 2$, costituiscono una partizione dell'evento certo. Per ogni $i = 1, 2, ...n$ si ha

$$\text{(vii) } P(H_i\,|E\,) = P(H_i)\frac{P(E \mid H_i)}{P(E)} =$$

$$= P(H_i) \cdot \frac{P(E \mid H_i)}{P(H_1)P(E \mid H_1) + P(H_2)P(E \mid H_2) + ... + P(H_n)P(E \mid H_n)}$$

o anche (usando la simbologia per grandezze proporzionali)

$$P(H_i \mid E) \propto P(H_i)P(E \mid H_i)$$

giocando $\dfrac{1}{P(E)}$ come **fattore di normalizzazione**.

Se poi le ipotesi in alternativa hanno tutte la stessa probabilità allora, e solo allora, la probabilità a posteriori è proporzionale alla sola verosimiglianza

$$P(H_i \mid E) \propto P(E \mid H_i)$$

Bayes postulò tale assunzione, nel caso in cui il valutatore non abbia elementi per comportarsi diversamente giudicando tutte le ipotesi ugualmente probabili.

2.2.2 Stimatori puntuali e intervallari

Limitandosi al caso di **campionamento bernoulliano**, **ex ante** si tratta di considerare n variabili aleatorie, $X_1, X_2, ..., X_n$, tutte indipendenti e distribuite come Y la variabile statistica (o aleatoria) associata alla collettività. **Ex post**, ovvero dopo l'effettivo campionamento, si hanno i dati riferiti al campione selezionato: $x_1, x_2, ..., x_n$.

Il problema, beninteso per ogni tipo di campionamento, è come usare i dati per stimare momenti della variabile Y, per esempio $E\,[Y]$ e $\sigma^2\,[Y]$, o valori da dare ai parametri di una distribuzione teorica, ipotizzata adeguata per rappresentare l'incertezza su Y.

Uno **stimatore** s è una funzione dei dati osservati: **ex post è deterministico**, **ex ante è una variabile aleatoria**. Per uno stimatore varie proprietà possono essere richieste e tra le più comuni si segnalano

1) la **centratura** o **non distorsione**, ovvero, se si indica con a il valore (reale deterministico) da stimare, qualunque sia la dimensione (si usa impropriamente dimensione invece di cardinalità) campionaria deve essere

$$E\left[s(X_1, X_2, ..., X_n)\right] = a$$

2) la **consistenza**, ovvero, se si indica ancora con a il valore da stimare, al crescere della dimensione campionaria deve risultare

$$\lim_{n \to +\infty} E\left[s(X_1, X_2, ..., X_n)\right] = a$$

3) l'**efficienza**, ovvero, se si indica ancora con a il valore da stimare, fissata una dimensione campionaria, n, uno stimatore s_1 si dice più efficiente di s_2 se

$$E\left[(s_1(X_1, X_2, ..., X_n) - a)^2\right] \leq E\left[(s_2(X_1, X_2, ..., X_n) - a)^2\right]$$

Si osservi che se s_1 e s_2 sono stimatori centrati allora la disuguaglianza può scriversi

$$\sigma^2\left[s_1(X_1, X_2, ..., X_n)\right] \leq \sigma^2\left[s_2(X_1, X_2, ..., X_n)\right]$$

Si dice **momento campionario di ordine** r associato ad un campione di numerosità n (qui quello bernoulliano) lo stimatore

$$\frac{1}{n}\sum_{i=1}^{n} X_i^r$$

Si dice **varianza campionaria** associata ad un campione di numerosità n (qui quello bernoulliano) lo stimatore

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j)^2$$

Se si usano tali stimatori, in altre parole se si usa il **metodo dei momenti campionari**, giova tenere conto che

$$E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = E\left[Y\right]$$

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j)^2\right] = \frac{n-1}{n}\sigma^2\left[Y\right]$$

che potranno suggerire le seguenti **stime puntuali ex post**

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

quale stima di $E\left[Y\right]$ o quale stima del parametro m se per Y si fosse fatta l'ipotesi di normalità, o quale stima del rapporto $\dfrac{1}{\beta}$ se per Y si fosse fatta l'ipotesi di distribuzione esponenziale e così via, e

$$A = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j)^2$$

quale stima di $\sigma^2[Y]$ o quale stima del parametro σ^2 se per Y si fosse fatta l'ipotesi di normalità e così via.

Talvolta, anche per motivi prudenziali, si sostituisce o **si accompagna una stima puntuale con una stima intervallo.** Per esempio avendo determinato $\overline{x}$ e A con i dati di un campione si potrà proporre, pensando che la distribuzione dello stimatore sia «quasi» normale,

$$\left[\overline{x} - \gamma_\varepsilon\sqrt{\frac{A}{n}}, \overline{x} + \gamma_\varepsilon\sqrt{\frac{A}{n}}\right]$$

quale intervallo di stima fiduciaria a livello $1-\varepsilon$ per $E[Y]$, ovvero

$$P\left(E[Y] \in \left[\overline{x} - \gamma_\varepsilon\sqrt{\frac{A}{n}}, \overline{x} + \gamma_\varepsilon\sqrt{\frac{A}{n}}\right]\right) = 1-\varepsilon$$

dove γ_ε è il **frattile della distribuzione normale standard** che verifica l'equazione in x

$$\int_0^x \frac{1}{\sqrt{2\pi}}\, e^{-\frac{t^2}{2}}\, dt = \frac{1-\varepsilon}{2}$$

Alcuni frattili notevoli sono di seguito riportati in Tab. 2

Tab. 2 - Frattili della distribuzione normale standard

$\gamma_{0.5} = 0.6744898$	$\gamma_{0.25} = 1.1503496$
$\gamma_{0.1} = 1.6448540$	$\gamma_{0.05} = 1.9599648$
$\gamma_{0.025} = 2.2414042$	$\gamma_{0.01} = 2.5758326$
$\gamma_{0.005} = 2.8070399$	$\gamma_{0.0025} = 3.0233531$
$\gamma_{0.001} = 3.2905538$	$\gamma_{0.0005} = 3.4808079$
$\gamma_{0.00025} = 3.6623585$	$\gamma_{0.0001} = 3.8908206$

Un altro **esempio di stima intervallo riguarda i frattili di una distribuzione**. Se i dati di un campione bernoulliano, associato ad una variabile continua, posti in ordine crescente (si parla di statistica dell'ordine) sono $x_{(1)} < x_{(2)} < ... < x_{(n)}$ allora una **stima intervallo della mediana** di Y, indicata con $y_{\frac{1}{2}}$ e supposta unica, è data da (come facilmente si giustifica pensando alla probabilità che tutti i dati siano o sotto o sopra la mediana)

$$P\left(y_{\frac{1}{2}} \in \left[x_{(1)}, x_{(n)}\right]\right) = 1 - 2\left(\frac{1}{2^n}\right) = 1 - \frac{1}{2^{n-1}}$$

L'intervallo fiduciario di appartenenza della mediana **può essere ridotto di ampiezza se si aumenta la probabilità di errore**. Un semplice ragionamento combinatorico conduce alla

$$P\left(y_{\frac{1}{2}} \in \left[x_{(1+j)}, x_{(n-j)}\right]\right) = 1 - 2\left(\sum_{i=0}^{j} \binom{n}{i} \frac{1}{2^n}\right)$$

che vale per $j = 0, 1, ..., \lfloor \frac{n-1}{2} \rfloor$: usuale simbolo per parte intera. Per $j = 0$ si ha il caso precedente, ovviamente.

Quanto detto per la mediana può valere **per un qualsiasi frattile**, o VAR della teoria del rischio dove l'acronimo VAR sta per *value at risk*. Se si indica y_α l'ascissa, ipotizzata unica, tale che $F_Y(y_\alpha) = \alpha$ allora vale, ipotizzando anche che αn sia intero,

$$P\left(y_\alpha \in \left[x_{(\alpha n - j_1)}, x_{(\alpha n + j_2)}\right]\right) =$$
$$= 1 - \sum_{i=0}^{\alpha n - j_1 - 1} \binom{n}{i} \alpha^i (1-\alpha)^{n-i} - \sum_{i=\alpha n + j_2}^{n} \binom{n}{i} \alpha^i (1-\alpha)^{n-i}$$

che con l'**approssimazione normale della binomiale** può stimarsi utilizzando i valori tabullati della normale standard

$$P\left(y_\alpha \in \left[x_{(\alpha n - j_1)}, x_{(\alpha n + j_2)}\right]\right) \simeq \Phi\left(\frac{j_2}{\sqrt{n\alpha(1-\alpha)}}\right) - \Phi\left(\frac{-j_1 - 1}{\sqrt{n\alpha(1-\alpha)}}\right)$$

dove $\Phi(x) = \int_{-\infty}^{x} \frac{\exp\left(-\frac{t^2}{2}\right)}{\sqrt{2\pi}} dt$ è appunto la distribuzione normale standard.

Si noti che le stime basate sui momenti campionari e sulla statistica dell'ordine non fanno assunzioni sulla famiglia di appartenenza della distribuzione della variabile su cui si vogliono inferire informazioni: si parla di **statistica inferenziale non parametrica**. Una tecnica di stima puntuale del valore dei parametri di una distribuzione, che presuppone invece **assunzioni in merito alla famiglia parametrica** di appartenenza, è il **metodo di massima verosimiglianza** che consiste nel deteminare il vettore dei parametri, θ, che realizza appunto la massima verosimiglianza

$$\max_{\theta \in \Theta} f(x_1, \theta) f(x_2, \theta) ... f(x_n, \theta)$$

dove f è la densità di probabilità da selezionare in una famiglia parametrica e che viene univocamamente individuata una volta fissato θ.

Se g è poi la densità a priori per la variabile aleatoria Θ si parla di **stimatori bayesiani** quando si determini θ, determinazione di Θ, dalla **massimizzazione della densità di probabilità ex post**

$$\max_{\theta \in \Theta} g(\theta) f(x_1, \theta) f(x_2, \theta) ... f(x_n, \theta)$$

2.2.3 Test di ipotesi

Considerare due o più ipotesi in alternativa e fissare delle soglie di probabilità per la loro accettazione/rifiuto è lo scopo dei test di ipotesi. Limitandosi al caso più semplice si consideri l'ipotesi che i dati empirici si riferiscano ad un campione estratto da una popolazione ipotizzata adeguatamente modellata con una predeterminata distribuzione di probabilità, ipotesi H, e si vuol misurare la plausibilità di tale ipotesi.

A titolo puramente illustrativo presento il test di accettazione dell'ipotesi di Pearson (o test chiquadro - Egon S. Pearson 1895-1980). Data l'ipotesi H (la specificazione di una distribuzione di probabilità) si partisce l'insieme delle determinazioni della variabile aleatoria in k sottointervalli, sia $p_j > 0$ la probabilità che la variabile aleatoria appartenga al $j - mo$ intervallo per $j = 1, 2, ..., k$ e sia invece f_j la frequenza dei dati osservati, complessivamente n, nello stesso intervallo: allora il test di Pearson prevede di considerare il valore

$$\chi^2 = \sum_{j=1}^{k} \frac{(f_j - np_j)^2}{np_j}$$

e di controllare su apposite tabulazioni (*ex ante* il test χ^2 è una variabile aleatoria ben approssimata da una distribuzione della famiglia delle chiquadro) se il χ^2 associato al campione è minore o uguale di un opportuno frattile della distribuzione. Ovviamente tale frattile viene scelto in base alla probabilità, che si ritiene accettabile, di rifiutare l'ipotesi H (**si parla di errore di prima specie**): tanto minore è tale probabilità quanto maggiore è la soglia sotto la quale si accetta l'ipotesi H. Talvolta, se è possibile incrementare la dimensione del campione, si fissano due soglie, α_1 e α_2, con $\alpha_1 < \alpha_2$ e se $\chi^2 \leq \alpha_1$ si accetta H, se $\chi^2 \geq \alpha_2$ si rifiuta H e se $\alpha_1 < \chi^2 < \alpha_2$ si decide di incrementare ulteriormente la dimensione del campione.

Per per apprezzare i valori tabulati, cartacei una volta oggi digitalizzati, del test χ^2 **può essere utile ricordare i primi due momenti della variabile** χ^2. Risulta come **agevolmente si verifica** che (somma di k variabili aleatorie binomiali, le f_j, trasformate in $\frac{(f_j - np_j)^2}{np_j}$)

$$E\left[\chi^2\right] = \sum_{j=1}^{k} \frac{np_j\,(1 - p_j)}{np_j} = k - 1$$

mentre è assai **più laborioso mostrare** che

$$\sigma^2\left[\chi^2\right] = 2\,(k-1) - \frac{k^2 + 2k - 2}{n} + \sum_{j=1}^{k} \frac{1}{np_j}$$

che nel caso equiprobabile $p_j = \frac{1}{k}$ per $j = 1, 2, ..., j$ equivale a

$$\sigma^2\left[\chi^2\right] = \frac{2\,(n-1)}{n} \cdot (k-1)$$

2.2.4 Decisioni statistiche

Nella teoria delle decisioni nel cosiddetto caso finito a fronte di un numero $n > 1$ di ipotesi in alternativa si considera un numero $m > 1$ di possibili decisioni pure da prendere e si costruiscono tabelle di guadagni (o di perdite, basta cambiare segno al guadagno) nelle cui $m \times n$ celle si indicano i guadagni g_{ij} con $i = 1, 2, ..., m$ e $j = 1, 2, ..., n$.

Si parla di decisioni statistiche quando si ipotizzano più ipotesi in alternativa e si deve valutare quale ipotesi è da scegliere in un dato contesto: la questione è quindi riportata a un problema di **decisione** da prendere (si rilegga la «definizione» della statistica di Wald). Se n sono le ipotesi in alternativa si costruisce la tabella delle perdite, il cui generico termine si indica con l_{ij} («l» da *loss*), $i, j = 1, 2, ..., n$, che rappresenta la perdita che si ha se si decide per H_i ed è invece vera H_j. In base ai valori contenuti in tale tabella si sceglie l'ipotesi (anche in modo misto?!) basandosi su adeguati criteri decisionali.

Nel caso più classico di due ipotesi in alternativa si pone $l_{11} = l_{22} = 0$, l_{21} pari alla probabilità di rifiutare H_1 (o accettare H_2) se è vero H_1 (**errore di primo tipo** se H_1 è considerata l'ipotesi «principale») e l_{12} pari alla probabilità di accettare H_1 (o rifiutare H_2) se è vero H_2 (**errore di secondo tipo** se H_1 è considerata l'ipotesi «principale»).

3 La statistica applicata alle indagini giudiziarie con esemplificazioni virtuali

Tutte le metodologie statistiche, descrittive e inferenziali finora richiamate, potrebbero trovare applicazioni in ambito giudiziario, ma per scarsità di tempo **mi limito di seguito a illustrare il solo problema dell'aggiornamento delle probabilità di ipotesi sulla base di informazioni aggiuntive**, sottoparagrafo 2.2.1, in ipotetiche virtuali situazioni giudiziarie, sottolineando il ruolo della correlazione che sussiste tra gli eventi, condizionati o meno, di volta in volta considerati.

3.1 Come si modifica la probabilità di un'ipotesi con l'acquisizione di uno o due indizi

Si presenta un semplice esempio-esercizio per mostrare come si deve aggiornare con coerenza la probabiltà di un'ipotesi se si tiene conto di informazioni pertinenti aggiuntive.

Supponiamo che un fatto, un reato, la presenza di una persona..., si compia in un dato intervallo di tempo in un certo intervallo spaziale 1-dimensionale.

Fissata una discretizzazione per il tempo (in ascissa) e uno per lo spazio (in ordinata) si suppone che l'evento certo (associato al fatto considerato) sia scomponibile in $10 \times 10 = 100$ **eventi elementari tutti ugualmente probabili.** Con questa assunzione se si considera la rappresentazione di *Fig.*1

Fig. 1

9										
8										
7										
6			E	E						
5			E	E						
4			E	E						
3										
2										
1										
0										
SPAZIO *TEMPO*	0	1	2	3	4	5	6	7	8	9

l'insieme-evento E, da interpretare "il fatto associato all'evento E si è verificato nelle sei celle contrassegnate da E", ha probabilità

$$P(E) = \frac{6}{100}$$

Si supponga ora che un reato C sia stato commesso in una delle 100 celle dello spazio-tempo considerato. La presenza nello scenario configurato di uno degli indiziati del reato potrebbe essere documentata da due rilevatori (si pensi a fotocamere, telecamere, celle telefoniche...). Il rilevatore A ha raggio di azione nelle $3 \times 8 = 24$ celle del rettangolo di vertici $A \equiv [(0,0);(2,7)]$, il rilevatore B ha raggio di azione nelle $10 \times 5 = 50$ celle del rettangolo di vertici $B \equiv [(0,5);(9,9)]$ e si suppone che i due rivelatori non possano fornire dettagli sulle singole celle di A rispettivamente di B in cui la rilevazione è avvenuta.

Se consideriamo gli eventi

$A \equiv$ "Il rilevatore A ha documentato la presenza dell'indiziato nel rettangolo di vertici $A \equiv [(0,0);(2,7)]$"

$B \equiv$ "Il rilevatore B ha documentato la presenza dell'indiziato nel rettangolo di vertici $B \equiv [(0,5);(9,9)]$"

$C \equiv$ "Il reato considerato è stato commesso dall'indiziato nel rettangolo $C \equiv [(2,3);(4,7)]$"

la rappresentazione grafica di quello che può verificarsi in termini dei possibili prodotti associati ai tre eventi A, B, C in ognuna delle 100 celle è

Fig. 2

9	B	B	B	B	B	B	B	B	B	B
8	B	B	B	B	B	B	B	B	B	B
7	AB	AB	ABC	BC	BC	B	B	B	B	B
6	AB	AB	ABC	BC	BC	B	B	B	B	B
5	AB	AB	ABC	BC	BC	B	B	B	B	B
4	A	A	AC	C	C					
3	A	A	AC	C	C					
2	A	A	A							
1	A	A	A							
0	A	A	A							
SPAZIO *TEMPO*	0	1	2	3	4	5	6	7	8	9

Per esempio nella cella-evento elementare $(1,7)$, troviamo il prodotto AB: la probabilità $\frac{1}{100}$ associata alla cella contribuisce per $\frac{1}{100}$ alla probabilità del prodotto AB che è $\frac{9}{100}$ essendo 9 le celle in cui si verifica AB ovvero quelle del rettangolo $AB \equiv [(0,5);(2,7)]$; nella cella-evento elementare $(2,5)$, troviamo il prodotto ABC: la probabilità $\frac{1}{100}$ associata alla cella contribuisce per $\frac{1}{100}$ alla probabilità del prodotto ABC che è $\frac{3}{100}$ essendo 3 le celle in cui si può verificare ABC ovvero quelle dell'insieme $ABC \equiv [(2,5);(2,7)]$.

L'evento C è un'ipotesi di lavoro per circoscrivere tempo e luogo in cui il reato è stato commesso: $C \equiv [(2,3);(4,7)]$ è l'unione di 15 celle e all'inizio delle indagini ogni cella ha associato un evento elementare equiprobabile. Con queste assunzioni è immediato stimare le seguenti probabilità degli eventi A, B, C

$$P(A) = \frac{24}{100},\ P(B) = \frac{50}{100} \text{ e } P(C) = \frac{15}{100}$$

degli eventi prodotto AB, AC, BC

$$P(AB) = \frac{9}{100},\ P(AC) = \frac{5}{100} \text{ e } P(BC) = \frac{9}{100}$$

dell'evento prodotto ABC

$$\text{e } P(ABC) = \frac{3}{100}$$

Si noti che i due eventi A e B, considerabili quali eventi-indizi per valutare la posizione giudiziaria dell'indiziato del reato, sono **negativamente correlati** essendo

$$P(AB) = \frac{9}{100} < P(A) \cdot P(B) = \frac{24}{100} \cdot \frac{50}{100} = \frac{12}{100}$$

Ci si chiede ora come debba essere modificata la probabilità a priori, $P(C) = \frac{15}{100}$, se si tiene conto delle informazioni ottenibili dai due rilevatori. Applicando i classici risultati sulla valutazione delle probabilità degli eventi condizionati (probabilità composte, teorema di Bayes) e tenuto conto delle valorizzazioni delle probabilità suggerita da *Fig.2* si ottengono le seguenti valutazioni di probabilità

$$P(A \mid C) = \frac{P(AC)}{P(C)} = \frac{5}{15}$$

$$P(B \mid C) = \frac{P(BC)}{P(C)} = \frac{9}{15}$$

$$P(AB \mid C) = \frac{P(ABC)}{P(C)} = \frac{3}{15}$$

$$P\left(C \mid A\right) = \frac{P(AC)}{P(A)} = \frac{5}{24} = P(C) \cdot \frac{P(A \mid C)}{P(A)} = \frac{15}{100} \cdot \frac{\frac{5}{15}}{\frac{24}{100}} = \frac{15}{100} \cdot \frac{25}{18} = \frac{5}{24}$$

$$P\left(C \mid B\right) = \frac{P(BC)}{P(B)} = \frac{9}{50} = P(C) \cdot \frac{P(B \mid C)}{P(B)} = \frac{15}{100} \cdot \frac{\frac{9}{15}}{\frac{50}{100}} = \frac{15}{100} \cdot \frac{6}{5} = \frac{9}{50}$$

$$P\left(C \mid AB\right) = \frac{P(ABC)}{P(AB)} = \frac{3}{9} = \frac{1}{3} = P(C) \cdot \frac{P(AB \mid C)}{P(AB)} = \frac{15}{100} \cdot \frac{\frac{3}{15}}{\frac{9}{100}} =$$

$$= \frac{15}{100} \cdot \frac{20}{9} = \frac{1}{3}$$

Per inciso si osservi che qui accade che $A \mid C$ e $B \mid C$ sono stocasticamente indipendenti essendo

$$P(A \mid C) \cdot P(B \mid C) = \frac{5}{15} \cdot \frac{9}{15} = \frac{3}{15} = P(AB \mid C)$$

Si può osservare che in questo esempio-esercizio l'acquisizione di ognuna delle tre informazioni A o B o AB fa crescere la valutazione della probabilità che l'indiziato abbia commesso il reato, che passa da $\frac{15}{100}$ rispettivamente a

$$\frac{5}{24}, \frac{9}{50}, \frac{1}{3}$$

con il seguente ordinamento dei valori

$$P\left(C \mid AB\right) = \frac{1}{3} > P\left(C \mid A\right) = \frac{5}{24} > P\left(C \mid B\right) = \frac{9}{50}$$

Ovviamente in tutti e tre i casi è maggiore di 1 il relativo rapporto di verosimiglianza (*likehood ratio*) e risulta, in concordanza con l'ordinamento ottenuto per le tre probabilità condizionate di $C \mid AB$, $C \mid A$ e $C \mid B$,

$$\frac{P(AB \mid C)}{P(AB)} = \frac{20}{9} > \frac{P(A \mid C)}{P(A)} = \frac{25}{18} > \frac{P(B \mid C)}{P(B)} = \frac{6}{5}$$

Muovendo nel quadrato-evento certo i tre eventi A, B e C è facile produrre le varie alternative che possono presentarsi nel confronto di $P(C)$ con $P(C \mid A)$, $P(C \mid B)$, $P(C \mid AB)$ o equivalentemente nel confronto dei tre rapporti di verosimiglianza con il valore 1.

Per contestualizzare e con l'intento di misurare l'effetto della correlazione degli eventi-indizi A o/e B si consideri ora la seguente rappresentazione, in *Fig.* 3, ottenuta dalla precedente con una traslazione di 1 in alto di tutte le celle di A

Fig. 3

9	B	B	B	B	B	B	B	B	B	B
8	AB	AB	AB	B	B	B	B	B	B	B
7	AB	AB	ABC	BC	BC	B	B	B	B	B
6	AB	AB	ABC	BC	BC	B	B	B	B	B
5	AB	AB	ABC	BC	BC	B	B	B	B	B
4	A	A	AC	C	C					
3	A	A	AC	C	C					
2	A	A	A							
1	A	A	A							
0										
$SPAZIO$ $TEMPO$	0	1	2	3	4	5	6	7	8	9

Da *Fig.* 3 si deduce ora che

$P(A) = \frac{24}{100}$, $P(B) = \frac{50}{100}$, $P(C) = \frac{15}{100}$

$P(AB) = \frac{12}{100}$, $P(AC) = \frac{5}{100}$ e $P(BC) = \frac{9}{100}$

$P(ABC) = \frac{3}{100}$

L'unica probabilità cambiata è quella di AB. Essendo

$$P(AB) = \frac{12}{100} = \frac{24}{100} \cdot \frac{50}{100} = P(A) \cdot P(B)$$

i due eventi-indizi A e B sono ora **stocasticamente indipendenti**. Nulla cambia per

$$P(A \mid C) = \frac{P(AC)}{P(C)} = \frac{5}{15}$$

$$P(B \mid C) = \frac{P(BC)}{P(C)} = \frac{9}{15}$$

$$P(AB \mid C) = \frac{P(ABC)}{P(C)} = \frac{3}{15}$$

(gli eventi $A \mid C$ e $B \mid C$ continuano ad essere stocasticamente indipendenti) né cambiano le probabilità di C condizionate al singolo indizio

$$P(C \mid A) = \frac{P(AC)}{P(A)} = \frac{5}{24} = P(C) \cdot \frac{P(A \mid C)}{P(A)} = \frac{15}{100} \cdot \frac{\frac{5}{15}}{\frac{24}{100}} = \frac{15}{100} \cdot \frac{25}{18} = \frac{5}{24}$$

$$P(C \mid B) = \frac{P(BC)}{P(B)} = \frac{9}{50} = P(C) \cdot \frac{P(B \mid C)}{P(B)} = \frac{15}{100} \cdot \frac{\frac{9}{15}}{\frac{50}{100}} = \frac{15}{100} \cdot \frac{6}{5} = \frac{9}{50}$$

ma cambia la probabilità di C condizionata al prodotto dei due indizi, riducendosi da $\frac{1}{3}$ a $\frac{1}{4}$

$$P(C \mid AB) = \frac{P(ABC)}{P(AB)} = \frac{3}{12} = \frac{1}{4} = P(C) \cdot \frac{P(AB \mid C)}{P(AB)} = \frac{15}{100} \cdot \frac{\frac{3}{15}}{\frac{12}{100}} =$$

$$= \frac{15}{100} \cdot \frac{5}{3} = \frac{1}{4}$$

L'ordinamento dei valori della probabilità di C condizionate è ancora lo stesso ma con il primo valore cambiato

$$P(C \mid AB) = \frac{1}{4} > P(C \mid A) = \frac{5}{24} > P(C \mid B) = \frac{9}{50}$$

Tra l'altro, avendosi in questa ipotesi l'indipendenza stocastica sia di A e di B che quella di $A \mid C$ e $B \mid C$, si può fattorizzare la probabilità $P(C \mid AB)$ come segue

$$P(C \mid AB) = P(C) \cdot \frac{P(AB \mid C)}{P(AB)} = P(C) \cdot \frac{P(A \mid C)}{P(A)} \cdot \frac{P(B \mid C)}{P(B)} =$$

$$= \frac{15}{100} \cdot \frac{25}{18} \cdot \frac{6}{5} = \frac{1}{4}$$

mettendo così in evidenza come la probabilità a posteriori di $C \mid AB$ possa ottenersi come prodotto della probabilità a priori, $P(C) = \frac{15}{100}$, e dei due rapporti di verosimiglianza $\frac{P(A \mid C)}{P(A)} = \frac{25}{18}$ e $\frac{P(B \mid C)}{P(B)} = \frac{6}{5}$.

Se la traslazione in alto di tutte le celle di A è non di 1 ma di 2 unità allora la rappresentazione dell'incertezza è lquella di *Fig.* 4

Fig. 4

9	AB	AB	AB	B	B	B	B	B	B	B
8	AB	AB	AB	B	B	B	B	B	B	B
7	AB	AB	ABC	BC	BC	B	B	B	B	B
6	AB	AB	ABC	BC	BC	B	B	B	B	B
5	AB	AB	ABC	BC	BC	B	B	B	B	B
4	A	A	AC	C	C					
3	A	A	AC	C	C					
2	A	A	A							
1										
0										
$SPAZIO$ $TEMPO$	0	1	2	3	4	5	6	7	8	9

Da *Fig.* 4 si deduce ora che

$P(A) = \frac{24}{100}$, $P(B) = \frac{50}{100}$, $P(C) = \frac{15}{100}$

$P(AB) = \frac{15}{100}$, $P(AC) = \frac{5}{100}$ e $P(BC) = \frac{9}{100}$

$P(ABC) = \frac{3}{100}$

L'unica probabilità che cambia è ancora quella di AB. Essendo

$$P(AB) = \frac{15}{100} > \frac{24}{100} \cdot \frac{50}{100} = P(A) \cdot P(B)$$

i due eventi-indizi A e B sono ora **positivamente correlati**. Nulla cambia per

$$P(A \mid C) = \frac{P(AC)}{P(C)} = \frac{5}{15}$$

$$P(B \mid C) = \frac{P(BC)}{P(C)} = \frac{9}{15}$$

$$P(AB \mid C) = \frac{P(ABC)}{P(C)} = \frac{3}{15}$$

gli eventi $A \mid C$ e $B \mid C$ continuano ad essere stocasticamente indipendenti, né cambiano le probabilità di C condizionate al singolo indizio

$$P\left(C \mid A\right) = \frac{P(AC)}{P(A)} = \frac{5}{24} = P(C) \cdot \frac{P(A \mid C)}{P(A)} = \frac{15}{100} \cdot \frac{\frac{5}{15}}{\frac{24}{100}} = \frac{15}{100} \cdot \frac{25}{18} = \frac{5}{24}$$

$$P\left(C \mid B\right) = \frac{P(BC)}{P(B)} = \frac{9}{50} = P(C) \cdot \frac{P(B \mid C)}{P(B)} = \frac{15}{100} \cdot \frac{\frac{9}{15}}{\frac{50}{100}} = \frac{15}{100} \cdot \frac{6}{5} = \frac{9}{50}$$

ma cambia la probabilità di C condizionata al prodotto dei due indizi, riducendosi ora fino a $\dfrac{1}{5}$

$$P\left(C \mid AB\right) = \frac{P(ABC)}{P(AB)} = \frac{3}{15} = \frac{1}{5} = P(C) \cdot \frac{P(AB \mid C)}{P(AB)} = \frac{15}{100} \cdot \frac{\frac{3}{15}}{\frac{15}{100}} =$$

$$= \frac{15}{100} \cdot \frac{4}{3} = \frac{1}{5}$$

L'ordinamento dei valori della probabilità di C condizionate agli indizi cambia ed è ora

$$P\left(C \mid A\right) = \frac{5}{24} > P\left(C \mid AB\right) = \frac{1}{5} > P\left(C \mid B\right) = \frac{9}{50}$$

e quello che qui si verifica è vero ma paradossale: il prodotto logico di due indizi, AB, ognuno dei quali fa crescere la probabilità a priori di C, può indurre a valorizzare la probabilità di C meno di quanto si otterrebbe con un singolo indizio!

Sempre riferendosi all'esempio considerato **si potrebbe dover considerare l'evento condizionato** $C \mid A + B$ immaginando che almeno uno dei due rilevatori abbia documentato la presenza dell'indiziato. In tale ipotesi per la probabilità dell'evento considerato sussiste

$$P\left(C \mid A + B\right) = P(C) \cdot \frac{P(A + B \mid C)}{P(A + B)} =$$

$$= P(C) \cdot \frac{P(A \mid C) + P(A \mid C) - P(AB \mid C)}{P(A) + P(B) - P(AB)}$$

Nelle tre ipotesi prefigurate si determinerebbero i seguenti valori:

- nell'ipotesi di *Fig.* 2 (eventi A e B negativamente correlati)

$$P\left(C \mid A + B\right) = \tfrac{15}{100} \cdot \frac{\frac{5}{15} + \frac{9}{15} - \frac{3}{15}}{\frac{24}{100} + \frac{50}{100} - \frac{9}{100}} = \frac{11}{65}$$

con l'ordinamento dei valori

$$P\left(C \mid AB\right) = \frac{1}{3} > P\left(C \mid A\right) = \frac{5}{24} > P\left(C \mid B\right) = \frac{9}{50} > P\left(C \mid A + B\right) = \frac{11}{65}$$

- nell'ipotesi di *Fig.* 3 (eventi A e B stocasticamente indipendenti)

$$P\left(C \mid A + B\right) = \tfrac{15}{100} \cdot \frac{\frac{5}{15} + \frac{9}{15} - \frac{3}{15}}{\frac{24}{100} + \frac{50}{100} - \frac{12}{100}} = \frac{11}{62}$$

con l'ordinamento dei valori

$$P\left(C \mid AB\right) = \frac{1}{4} > P\left(C \mid A\right) = \frac{5}{24} > P\left(C \mid B\right) = \frac{9}{50} > P\left(C \mid A + B\right) = \frac{11}{62}$$

- nell'ipotesi di *Fig.* 4 (eventi A e B positivamente correlati)

$$P\left(C \mid A+B\right)=\tfrac{15}{100}\cdot\frac{\frac{5}{15}+\frac{9}{15}-\frac{3}{15}}{\frac{24}{100}+\frac{50}{100}-\frac{15}{100}}=\frac{11}{59}$$

con l'ordinamento dei valori

$$P\left(C \mid A\right)=\frac{5}{24}>P\left(C \mid AB\right)=\frac{1}{5}>P\left(C \mid A+B\right)=\frac{11}{59}>P\left(C \mid B\right)=\frac{9}{50}$$

3.2 Come si modifica la probabilità di un'ipotesi con l'acquisizione di più di due indizi

Generalizzando a un numero qualsiasi di eventi-indizi, $A_1, A_2, ..., A_m$, con $m > 2$ vale in generale per l'evento C (un reato che si è consumato in un ben definito modo) condizionato al prodotto degli m indizi

$$P\left(C \mid A_1A_2...A_m\right)=P(C)\cdot\frac{P(A_1A_2...A_m \mid C)}{P(A_1A_2...A_m)}$$

L'ipotesi che gli m eventi-indizi siano stocasticamente indipendenti e che lo siano anche condizionatamente a C consente di fattorizzare la probabilità $P\left(C \mid A_1A_2...A_m\right)$ ovvero rappresentarla come prodotto della probabilità a priori di C per gli m rapporti di verosimiglianza

$$P\left(C \mid A_1A_2...A_m\right)=P(C)\cdot\frac{P(A_1 \mid C)}{P(A_1)}\cdot\frac{P(A_2 \mid C)}{P(A_2)}\cdot ...\cdot\frac{P(A_m \mid C)}{P(A_m)}$$

Quando $\dfrac{P(A_h \mid C)}{P(A_h)}>1$ l'indizio $h-$mo aumenta la probabilità di $P(C)$ e viceversa se $\dfrac{P(A_h \mid C)}{P(A_h)}<1$. Se $\dfrac{P(A_h \mid C)}{P(A_h)}=1$ gli eventi A_h e C sono stocasticamente indipendenti e l'indizio $h-$mo non produce variazioni alla probabilità *ex ante* di C.

Se si ha soltanto l'indipendenza stocastica degli eventi-indizi $A_1, A_2, ..., A_m$ condizionatamente a C allora la fattorizzazione è solo per $P(A_1A_2...A_m \mid C)$ e si ha

$$P\left(C \mid A_1A_2...A_m\right)=P(C)\cdot\frac{P(A_1 \mid C)\cdot P(A_2 \mid C)\cdot ...\cdot P(A_m \mid C)}{P(A_1A_2...A_m)}$$

Se si ha soltanto l'indipendenza stocastica degli eventi-indizi $A_1, A_2, ..., A_m$ allora la fattorizazione è solo per il denominatore e si ha

$$P\left(C \mid A_1A_2...A_m\right)=P(C)\cdot\frac{P(A_1A_2...A_m \mid C)}{P(A_1)\cdot P(A_2)\cdot ...\cdot P(A_m)}$$

Sempre considerando un numero qualsiasi di eventi-indizi, $A_1, A_2, ..., A_m$, con $m > 2$ vale in generale per l'evento C (un reato che si è consumato in un ben definito modo) condizionato alla somma degli m indizi

$$P\left(C \mid A_1+A_2+...+A_m\right)=P(C)\cdot\frac{P(A_1+A_2+...+A_m \mid C)}{P(A_1+A_2+...+A_m)}$$

L'ipotesi che gli m eventi-indizi siano stocasticamente indipendenti e che lo siano anche condizionatamente a C consente di rappresentare la probabilità $P\left(C \mid A_1 + A_2 + ... + A_m\right)$ mediante il complemento a 1 della fattorizzazione delle probabilità degli eventi contrari

$$P\left(C \mid A_1 + A_2 + ... + A_m\right) =$$
$$= P(C) \cdot \frac{1 - [(1 - P(A_1 \mid C) \cdot (1 - P(A_2 \mid C)) \cdot ... \cdot (1 - P(A_m \mid C))]}{1 - [(1 - P(A_1) \cdot (1 - P(A_2)) \cdot ... \cdot (1 - P(A_m))]}$$

Se si ha soltanto l'indipendenza stocastica degli eventi-indizi $A_1, A_2, ..., A_m$ condizionatamente a C allora la "fattorizzazione" è solo per $P(A_1{+}A_2{+}...{+}A_m \mid C)$ e si ha

$$P\left(C \mid A_1 + A_2 + ... + A_m\right) =$$
$$= P(C) \cdot \frac{1 - [(1 - P(A_1 \mid C) \cdot (1 - P(A_2 \mid C)) \cdot ... \cdot (1 - P(A_m \mid C))]}{P(A_1 + A_2 + ... + A_m)}$$

Se si ha soltanto l'indipendenza stocastica degli eventi-indizi $A_1, A_2, ..., A_m$ allora la fattorizazione è solo per il denominatore e si ha

$$P\left(C \mid A_1 + ... + A_m\right) = P(C) \cdot \frac{P(A_1 + A_2 + ... + A_m \mid C)}{1 - [(1 - P(A_1) \cdot (1 - P(A_2)) \cdot ... \cdot (1 - P(A_m))]}$$

3.3 Come si modifica la probabilità di un'ipotesi con l'acquisizione di decine, centinaia... di indizi

Concludo il mio intervento segnalando le valutazioni induttive di probabilità nel caso in cui gli indizi possano diventare molto numerosi, come nei casi in cui si utilizzano i dati genetici di particolari indiziati messi a confronto con il materiale biologico repertato.

Conviene introdurre la nozione di indicatore dell'evento E, in simboli

$$|E| = \begin{cases} 1 \text{ se } E \text{ si verifica} \\ 0 \text{ se E non si verifica} \end{cases}$$

Con questa posizione sono equivalenti gli eventi

E e $|E| = 1$, E_1E_2 e $(|E_1| + |E_2| = 2)$, $E_1 + E_2$ e $(|E_1| + |E_2| \geq 1)$,
$E_1E_2...E_n$ e $(|E_1| + |E_2| + ... + |E_n| = n)$,
$E_1 + E_2 + ... + E_n$ e $(|E_1| + |E_2| + ... + |E_n| \geq 1)$ ...

Riferendosi a un possibile ambito giudiziario si considerino gli eventi

$C \equiv$ "Il materiale biologico repertato è quello dell'indiziato XY"

Per $h = 1, 2, ..., m$ sia

$A_h \equiv$ "Il dato genetico $h-$mo del materiale biologico repertato è uguale al corrispondente dato genetico $h-$mo dell'indiziato XY"

Indicando con $\overline{C}$ l'evento contrario di C, si ricorda che (utilizzando ancora una delle versioni del teorema di Bayes) si può rappresentare la probabilità dell'evento $C \mid |A_1| + |A_2| + ... + |A_m| = j$ con j che può assumere i valori da 0 a m, anche nel modo seguente

$$P\left(C \mid |A_1| + ... + |A_m| = j\right) = P(C) \cdot \frac{P(|A_1| + ... + |A_m| = j \mid C)}{P(|A_1| + ... + |A_m| = j)} =$$

$$= P(C) \cdot \frac{P(|A_1| + ... + |A_m| = j \mid C)}{P(C) \cdot P(|A_1| + ... + |A_m| = j \mid C) + P(\overline{C}) \cdot P(|A_1| + ... + |A_m| = j \mid \overline{C})} =$$

$$= P(C) \cdot \frac{P(|A_1| + |A_2| + ... + |A_m| = j \mid C)}{P(C) \cdot P(|A_1| + ... + |A_m| = j \mid C) + (1 - P(C)) \cdot P(|A_1| + ... + |A_m| = j \mid \overline{C})}$$

Per contestualizzare, se si ipotizza l'indipendenza stocastica degli eventi $A_1, A_2, ..., A_m$ sia condizionatamente a C che a $\overline{C}$ e che per ogni $h = 1, 2, ..., m$ sia

$$P(A_h \mid C) = p > q = P(A_h \mid \overline{C})$$

allora si può considerare che

$$P\left(C \mid |A_1| + |A_2| + ... + |A_m| = j\right) =$$

$$= P(C) \cdot \frac{\binom{m}{j} p^j (1-p)^{m-j}}{P(C) \cdot \binom{m}{j} p^j (1-p)^{m-j} + (1 - P(C)) \cdot \binom{m}{j} q^j (1-q)^{m-j}} =$$

$$= P(C) \cdot \frac{p^j (1-p)^{m-j}}{P(C) \cdot p^j (1-p)^{m-j} + (1 - P(C)) \cdot q^j (1-q)^{m-j}}$$

$$= P(C) \cdot \frac{1}{P(C) + (1 - P(C)) \cdot \left(\frac{q}{p}\right)^j \left(\frac{1-q}{1-p}\right)^{m-j}}$$

Per esemplificare se si assume $m = 20$, che la probabilità di C sia $P(C) = \frac{1}{100000}$ (si suppone che XY sia 1 dei 100000 indiziabili) allora al variare di p e q si possono ottenere informazioni come le seguenti di Fig. 5 in merito alle probabilità condizionate $P\left(C \mid |A_1| + |A_2| + ... + |A_{20}| = j\right)$

Fig. 5 - Valori di $P\left(C \mid |A_1| + |A_2| + ... + |A_{20}| = j\right)$

p	q	$j = 14$	$j = 15$	$j = 16$	$j = 17$	$j = 18$	$j = 19$	$j = 20$
0.75	0.25	0.06157	0.37126	0.84163	0.97952	0.99768	0.99974	0.99997
0.75	0.30	0.00767	0.05133	0.27473	0.72614	0.94887	0.99236	0.99890
0.75	0.35	0.00139	0.00770	0.04145	0.19418	0.57311	0.88207	0.97656
0.75	0.40	0.00034	0.00156	0.00698	0.03068	0.12467	0.39059	0.74254
0.75	0.45	0.00011	0.00041	0.00151	0.00551	0.01993	0.06941	0.21477
0.75	0.5	0.00004	0.00013	0.00041	0.00123	0.00368	0.01096	0.03218

Occorrerebbe ancora molto spazio per **illustrare tutti e quattro i punti di vista dell'inferenza statistica** e mi limito, traendo spunto dalla tabella di *Fig.* 5, a un'ultima sottolineatura relativa alle connessioni che tra di loro possono essere individuate. Qui è semplice osservare come l'**aggiornamento delle probabilità delle ipotesi possa essere usato per decidere fissato che sia un criterio di decisione**: ad esempio se si stabilisce una soglia di probabilità al 99% per accettare C allora, guardando ai valori di $P\left(C \mid |A_1| + |A_2| + ... + |A_{20}| = j\right)$ di *Fig.* 5, questo può aversi solo con $q = 0.25$ per valori di j maggiori di 17 o con $q = 0.30$ per valori di j maggiori di 18.

Riferimenti bibliografici

Sarebbero innumerevoli e mi limito a dire che:

- il paragrafo 1 è una mia personale visione di come è suddivisa oggi la curiosità dei ricercatori;
- i richiami teorici del paragrafo 2 sono tratti da: Luigi Vannucci, *Compendio di calcolo delle probabilità e d'inferenza statistica con esercizi svolti*, Pitagora Editrice Bologna, 2003;
- il paragrafo 3 è invece quanto ho prodotto per questo Corso di aggiornamento professionale/perfezionamento, in particolare per il modulo 3 "La prova statistica nel processo penale: DNA e dattiloscopia".

Indice