

DISPENSE DI STATISTICA  
di Federico Emanuele Pozzi

SOMMARIO

PARTE III – Statistica descrittiva	
4.1 - <a href="#">Modello gaussiano</a>	2
PARTE IV – Probabilità	
5.1 - <a href="#">Probabilità</a>	11
5.2 - <a href="#">Elementi di calcolo combinatorio</a>	12
5.3 - <a href="#">Problemi riguardanti il calcolo delle probabilità</a>	19
5.4 - <a href="#">Definire e calcolare la probabilità</a>	27
5.5 - <a href="#">Teorema di Bayes e sue applicazioni</a>	35
6.1 - <a href="#">Variabile casuale</a>	40
6.2 - <a href="#">Modello binomiale</a>	43
6.3 - <a href="#">Teorema del limite centrale</a>	48
6.4 - <a href="#">Variabile casuale <math>\chi^2</math></a>	50
PARTE V – Inferenza	
7.1 - <a href="#">Campionamento</a>	56
7.2 - <a href="#">Variabili categoriche dicotomiche</a>	61
8.1 - <a href="#">Brevi richiami di logica</a>	64
8.2 - <a href="#">Saggiare ipotesi</a>	65

## 4.1 – Modello gaussiano

Il discorso che segue serve ad introdurre la gaussiana; se vi interessa solo sapere come è fatta balzate pure due pagine e andate alla prossima freccia (→).

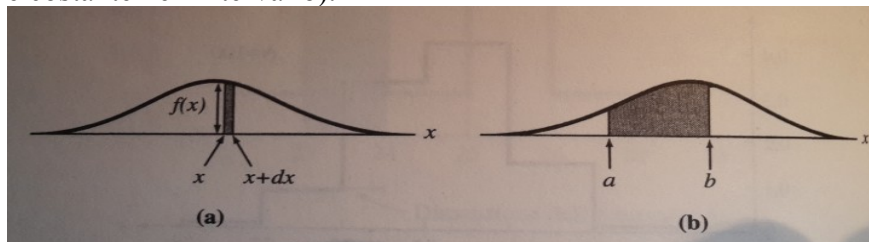
Supponiamo di fare un certo numero di misure di una quantità nota; è ragionevole aspettarsi che queste si distribuiscano simmetricamente intorno alla media delle misure stesse (supponendo che vi siano solo errori casuali). Avremo quindi una distribuzione di frequenze dei valori trovati centrata sulla media e con una certa deviazione standard, che indica il grado di dispersione dei risultati.

Possiamo osservare che la media è definibile come  $\mu_x = \frac{\sum x_k n_k}{N}$ , dove  $n$  è la frequenza assoluta di misure con il relativo valore  $x$ , ed  $N$  il numero totale di misure (la somma di tutti gli  $n$  dà  $N$ ).

Possiamo introdurre la frazione  $F_k = \frac{n_k}{N}$ , che indica la frazione delle  $N$  misure che hanno dato il valore  $k$ -esimo di  $x$ , e dunque rappresenta la densità di frequenza. Notare che la media diventa  $\mu_x = \sum x_k F_k$ , e  $\sum F_k = 1$  per ovvi motivi. L'ultima equazione è detta condizione di normalizzazione (qualunque insieme di numeri la cui somma sia 1 è detto normalizzato).

Se consideriamo  $x$  come una variabile continua, divisa in intervalli  $dx$  sufficientemente piccoli, possiamo scrivere le stesse formule di cui sopra in forma integrale, osservando che in questo caso la densità di frequenza dei vari valori che assume la variabile continua  $x$  sarà definita da una funzione  $f(x)$ , che vogliamo determinare [tale distribuzione dei risultati è detta distribuzione limite, nel senso che è la distribuzione che si ottiene quando il numero di misure effettuate tende ad infinito].

Guardando il disegno qui sotto, risulta abbastanza immediato osservare che la frazione di misure che cade in un intervallo compreso tra  $x$  e  $x+dx$  è  $f(x)dx$  (cioè l'area del rettangolo – considerate che quando  $dx$  è sufficientemente piccolo possiamo assumere che  $f(x)$  vari così poco che possiamo assegnarle valore costante nell'intervallo).



Ma allora è chiaro che la frazione di misure che cade in un intervallo qualsiasi  $[a:b]$  sarà data dall'integrale definito di  $f(x)$  nell'intervallo stesso:  $\int_a^b f(x) dx$ . Notiamo che per come abbiamo

concepito la funzione  $f(x)$ , che per ogni valore di  $x$  dà la frazione di risultati che hanno il valore  $x$ , l'area totale sotto la curva  $f$  avrà valore 1 (è esattamente lo stesso discorso fatto per  $F$  più sopra):

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

Ritornando alla definizione di media con la frazione  $F$ , per una distribuzione di valori molto grande come quella osservata in questo caso possiamo scrivere  $F_k = f(x_k) dx_k$ , e per intervalli molto piccoli possiamo riscrivere la media come  $\mu_x = \int_{-\infty}^{+\infty} x f(x) dx$ ; per considerazioni analoghe, la deviazione standard di questa distribuzione sarà  $\sigma_x^2 = \int_{-\infty}^{+\infty} (x - \mu_x)^2 f(x) dx$  (ricordate che non serve la  $N$  sotto, la nostra distribuzione è normalizzata – la  $N$  è idealmente il denominatore nella funzione  $f(x)$ !).

Ora non ci resta che trovare la funzione  $f(x)$  che abbia le seguenti proprietà:

- essere simmetrica intorno alla media  $\mu$
- essere normalizzata:  $\int_{-\infty}^{+\infty} f(x) dx = 1$

- soddisfare la relazione  $\mu_x = \int_{-\infty}^{+\infty} x f(x) dx$
- soddisfare la relazione  $\sigma_x^2 = \int_{-\infty}^{+\infty} (x - \mu_x)^2 f(x) dx$

Tutto questo discorso serve per farvi capire come si arriva a definire la gaussiana, se non riuscite a capire proprio tutto non preoccupatevi.

Una distribuzione siffatta assume la forma a campana, con un picco in corrispondenza del valore  $\mu$  e pendenze simmetriche ai lati dolci come le colline degli hobbit nel caso in cui la deviazione standard sia ampia o ripide come la torre dell'occhio nel caso contrario (con Sauron non si scherza, o sei preciso o sei fuori).

Questa curva è descritta dalla funzione  $e^{-\frac{x^2}{2\sigma^2}}$ , dove il parametro  $\sigma$  è detto parametro di larghezza e definisce quanto è "larga" la campana (vedremo che corrisponde alla deviazione standard); questa funzione (funzione di Gauss) è il prototipo di tutte le curve a campana, e ci dice che forma assume la nostra distribuzione. Per farla diventare la nostra funzione  $f(x)$  dobbiamo lavorarci su un po':

- La funzione di Gauss è centrata su  $x=0$ . Quindi la prima cosa da fare è centrarla sulla media  $\mu$ , facendo  $e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ .
- La funzione non è ancora normalizzata, e per farlo dovremmo moltiplicarne l'area totale per un fattore  $N$  (che dobbiamo ricavare) tale per cui  $\int_{-\infty}^{+\infty} N e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$ . Vi metto come si arriva alla determinazione di  $N$  in nota<sup>1</sup>; il risultato è  $N = \frac{1}{\sigma\sqrt{2\pi}}$ .
- Abbiamo ottenuto la distribuzione di Gauss:  $G_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ . Ora dobbiamo verificare che soddisfi le due relazioni che riguardano media e deviazione standard. Vi metto in nota solo la dimostrazione per la media<sup>2</sup>, dato che il procedimento per la seconda è analogo, ma l'integrazione per parti è più lunga.

→ La gaussiana  $G_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$  indica la probabilità di ottenere un determinato valore misurando la variabile  $x$ . Rappresenta una curva a campana centrata sulla media  $\mu$  e simmetrica ai lati. In ascissa compare la densità di probabilità, e l'integrale tra in un intervallo della funzione mi dà la probabilità di ottenere un valore che cade nell'intervallo (o in altre parole, la frequenza relativa).

Nella gaussiana la media coincide con moda e mediana.

Il parametro  $\sigma$  è la deviazione standard, ed indica la larghezza della campana. L'indice di

1 Sostituiamo le variabili ponendo  $y=x-\mu$  ( $dx=dy$ ), ottenendo  $\int_{-\infty}^{+\infty} N e^{-\frac{y^2}{2\sigma^2}} dy = 1$ , e ponendo  $z=y/\sigma$ , con quindi

$$dy=\sigma dz, \quad \int_{-\infty}^{+\infty} N e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 \quad ; \text{ l'integrale in } z \text{ è un integrale standard con valore } \sqrt{2\pi}, \text{ ragion per cui}$$

$$N = \frac{1}{\sigma\sqrt{2\pi}} .$$

2  $\mu = \int_{-\infty}^{+\infty} x G_{(\mu,\sigma)}(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$  ; col cambio di variabili  $y=x-\mu$

$$\mu = \frac{1}{\sigma\sqrt{2\pi}} \left( \int_{-\infty}^{+\infty} y e^{-\frac{y^2}{2\sigma^2}} dy + \mu \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2\sigma^2}} dy \right) ; \text{ per la definizione di } y, \text{ il primo integrale è } 0 \text{ (infatti lo } 0 \text{ della}$$

distribuzione in  $y$  corrisponde alla media, ed essendo la funzione simmetrica i contributi in  $-y$  cancellano quelli in  $+y$ ). Il secondo integrale è lo stesso incontrato nella nota 4, e sappiamo già che vale  $\sigma\sqrt{2\pi}$ ; semplificando col denominatore resta l'identità  $\mu=\mu$ . [C.V.D. - applausi].



- l'ultima forma invece considera l'intervallo  $[\mu+t\sigma;+\infty]$ , e in questo caso va da 0.5 (per  $t=0 \rightarrow$  ricordate che la media è anche la mediana, quindi dalla media in poi cadranno esattamente la metà dei valori!) e 0 (per  $t$  che tende ad infinito).

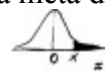


TABLE III. THE NORMAL PROBABILITY INTEGRAL

$\sigma$	$\sigma$	1	2	3	4	5	6	7	8	9	
0.0	0.0	50000	49601	49202	48803	48405	48006	47608	47210	46812	46414
0.1		46017	45620	45224	44828	44433	44038	43644	43251	42858	42465
0.2		42074	41683	41294	40905	40517	40129	39743	39358	38974	38592
0.3		38209	37828	37448	37070	36693	36317	35942	35569	35197	34827
0.4		34458	34090	33724	33360	32997	32636	32276	31918	31561	31207
0.5		30854	30503	30153	29806	29460	29116	28774	28434	28096	27760
0.6		27425	27093	26763	26435	26109	25785	25463	25143	24825	24510
0.7		24196	23885	23576	23270	22965	22663	22363	22065	21770	21476
0.8		21186	20897	20611	20327	20045	19766	19489	19215	18943	18673
0.9		18406	18141	17879	17619	17361	17106	16853	16602	16354	16109
1.0		15866	15625	15386	15151	14917	14686	14457	14231	14007	13786
1.1		13567	13350	13136	12924	12714	12507	12302	12100	11900	11702
1.2		11507	11314	11123	10935	10749	10565	10383	10204	10027	98525
1.3	0.0	96800	95098	93418	91759	90123	88508	86915	85343	83793	82264
1.4		80757	79270	77804	76359	74934	73529	72145	70781	69437	68112
1.5		66807	65322	64255	63008	61780	60571	59380	58208	57053	55917
1.6		54799	53609	52016	51551	50503	49471	48457	47460	46479	45514
1.7		44565	43633	42716	41815	40930	40059	39204	38364	37538	36727
1.8		35930	35148	34380	33625	32884	32157	31443	30742	30054	29379
1.9		28717	28067	27429	26803	26190	25588	24998	24419	23852	23295
2.0		22750	22216	21692	21178	20675	20182	19699	19226	18763	18309
2.1		17864	17429	17003	16586	16177	15778	15386	15003	14629	14262
2.2		13903	13553	13209	12874	12545	12224	11911	11604	11304	11011
2.3		10724	10444	10170	99031	96419	93867	91375	88940	86563	84242

Quest'ultima è la forma che usa il Duca.

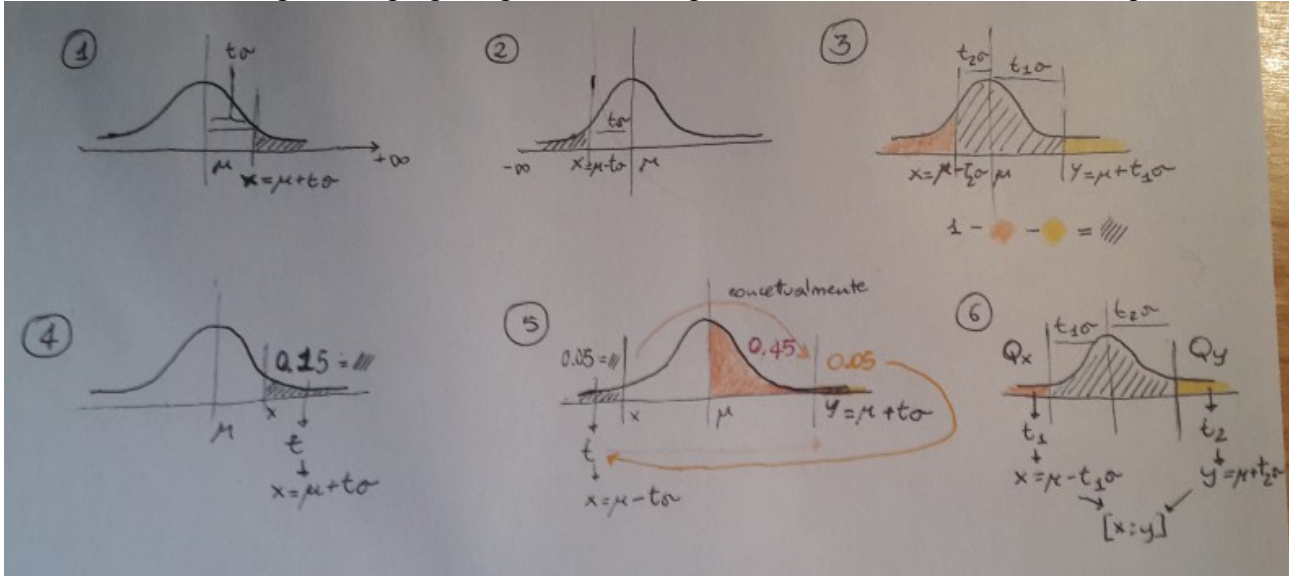
Possiamo usare questa tabella per calcolare la probabilità che un certo valore cada all'interno di un intervallo nelle nostre gaussiane<sup>3</sup>:

- per calcolare la probabilità che una misura cada tra un certo valore di  $x$  maggiore di  $\mu$  e infinito basta che ricordiamo che la  $x$  può essere espressa come  $x=\mu+t\sigma$ ; ricaviamo  $t$  come  $(x-\mu)/\sigma$  e otteniamo il valore di probabilità corrispondente a  $t$  dalla tabella
- per calcolare la probabilità che una misura cada tra un certo valore di  $x$  minore di  $\mu$  e  $-\infty$  basta ricordare che la gaussiana è simmetrica! Infatti posso riscrivere  $x$  come  $x=\mu-t\sigma$ , e ricavare  $t$  come  $(\mu-x)/\sigma$ ; a questo punto ricavo il valore di probabilità corrispondente a  $t$  dalla tabella
- per calcolare la probabilità che una misura cada entro un intervallo  $[x:y]$  devo ricavare le due probabilità tra  $x$  e  $-\infty$  e tra  $y$  e  $+\infty$  come sopra; sommandole, ho il valore della probabilità che la misura cada fuori dall'intervallo. Sottraendo questo valore a 1 ho la probabilità cercata
- supponiamo di voler calcolare quel valore di  $x$  che "lascia fuori" il 15% dei risultati (cioè l'85° centile): in questo caso dobbiamo trovare la  $t$  corrispondente a 0.15, e ricavare il valore di  $x$  come  $x=\mu+t\sigma$
- supponiamo invece di voler calcolare il valore di  $x$  che lascia fuori il 95% dei risultati, cioè il 5° centile; in questo caso non possiamo usare direttamente la tabella, perché arriva al massimo a 50% (considera solo i valori a destra della media), pertanto dobbiamo ragionare sulla simmetria della gaussiana. Infatti, tra il valore  $x$  e la media si distribuiranno il 45% dei risultati (50-5), e di conseguenza ricercheremo quel valore  $x$  tale che tra la media e  $x$  si distribuiscono il 45% dei risultati, ovvero il valore che lascia fuori il 5% dei risultati (corrispondente quindi al 95° centile). Per ovvie ragioni, tale valore sarà equidistante dalla media rispetto a  $x$ . Ricaviamo la  $t$  dalla tabella, e otteniamo  $x$  come  $x=\mu-t\sigma$
- supponiamo ora di voler trovare l'intervallo tra due centili (per esempio il range

3 Ciò significa passare dalla nostra specifica gaussiana alla gaussiana standardizzata, che consiste nel porre  $z=(x-\mu)/\sigma$ , e ricavare poi le  $x$  corrispondenti ai centili come  $x_c=\mu+z_c\sigma$ , con  $z_c$  che è lo  $z$  corrispondente al centile di interesse, cioè il valore corrispondente a quel centile nella gaussiana standardizzata. Siccome mi sembra inutile che un concetto così banale richieda tutti questi passaggi, ho ricavato i risultati e ve li ho presentati nel testo; sugli appunti del Duca troverete una procedura leggermente diversa, ma equivalente; questa mi sembra più semplice, poi vedete voi cosa vi è più utile.

interquartile, oppure i valori di riferimento per una certa variabile di riferimento come la glicemia, che sappiamo giacere tra il 2.5° e il 97.5° centile): ricaviamo i valori  $x$  e  $y$  corrispondenti al primo e al secondo centile di interesse rispettivamente. Per i centili compresi tra 0 e 50° useremo il metodo al punto 5, per quelli tra 50° e 100° quello al punto 6. Vi faccio notare che una volta capito il procedimento, il punto 5 può essere semplificato andando a trovare  $t$  come il valore che dà il centile di interesse.

Sotto trovate dei disegni che spiegano graficamente i procedimenti usati in ciascuno di questi casi:



\*\*\*\*\*

Abbiamo osservato che l'intervallo entro  $[\mu - 1,96\sigma; \mu + 1,96\sigma]$  contiene il 95% dei risultati; considerando una variabile che si distribuisca in maniera gaussiana (supponiamo la glicemia), utilizzare come intervallo di riferimento in clinica tale intervallo ottenuto dalla distribuzione dei valori di glicemia dei sani significa utilizzare un test che distingue sani da malati con una specificità del 95%.

→ la specificità è il rapporto tra veri negativi (cioè non malati secondo il test) e sani; in questo caso noi abbiamo ottenuto la distribuzione dei valori NORMALI di glicemia, e abbiamo detto “bene, consideriamo il 95% centrale di questi valori come un riferimento”, cioè “consideriamo il 95% centrale di questi come veramente sani”.

Nella pratica, quando noi misuriamo il valore della glicemia otteniamo un numero che dobbiamo confrontare con un riferimento per decidere se il soggetto è veramente malato o sano (o se è anormale o normale).

→ considerate che una gaussiana comprende ogni valore possibile, da  $-\infty$  a  $+\infty$ ; infatti la gaussiana tende asintoticamente a 0 (infatti, ogni valore è possibile almeno in teoria, anche se non lo è nella pratica – se troviamo un paziente con glicemia 60000000 mg/dl è il caso di fare una visita dall'oculista, perché abbiamo fatto l'esame ad una torta mentre il paziente sta agonizzando lì vicino – questo ovviamente è un valore molto improbabile, per quanto “possibile”, per cui nella gaussiana avrà una densità di frequenza quasi nulla, ma non nulla). Una specificità di esattamente 1 significa non credere nell'esistenza del diabete!

È chiaro però che un risultato molto distante, come la torta dell'esempio precedente, non sarà considerato veramente sano; quindi nel nostro “test” (che consiste in “misura la glicemia”) il paziente sarà un falso negativo → gli abbiamo misurato la glicemia, e abbiamo visto che ha una glicemia; però cade fuori dall'intervallo normale, quindi non è veramente sano.

Più restringiamo l'intervallo di riferimento, più diminuisce la specificità del test.

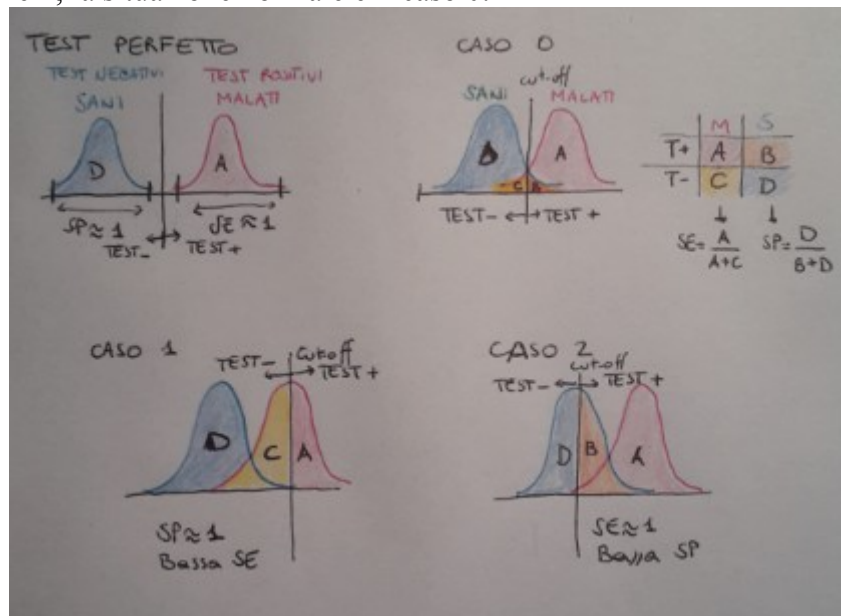
Possiamo ora misurare la glicemia di tutti i diabetici (malati); se prendiamo il 95% dei valori centrali come riferimento per la malattia, ciò equivale ad un test con SENSIBILITA' del 95% → analogamente a prima (ricordare che la sensibilità rappresenta il rapporto tra veri positivi e malati).

Il test: “dimmi la tua glicemia e ti dirò se sei diabetico” sarebbe perfetto nel discriminare tra malati e sani se le curve gaussiane di malati e sani fossero completamente separate (ma ricordate che la gaussiana non è mai nulla; quindi per completamente separate intendiamo “praticamente completamente separate”) come nell'esempio qui sotto → misuro la glicemia, confronto, taaac.

Tuttavia, nella realtà non è così, e ci saranno per esempio diabetici con una glicemia minore di alcuni sani, perché le curve si sovrappongono. Dovrò scegliere quindi un valore di cut-off, cioè un valore della glicemia per cui considero i test con risultato inferiore come negativi e quelli con risultato superiore come positivi. Fate riferimento al disegno per una visualizzazione più facile.

Utilizzare un valore di cut-off significa commettere un errore INEVITABILMENTE (infatti le due curve si sovrappongono e non c'è modo di cambiare questo fatto). Osserviamo che l'area C conterrà soggetti diabetici che il test considera negativi (falsi -), mentre l'area B conterrà soggetti sani che il test considera diabetici (falsi +).

Sensibilità e specificità del test sono in questo caso legate: aumentando la specificità (il che significa trovare un cut-off che minimizzi l'area B) diminuirà la sensibilità (perché aumenterà l'area C e diminuirà l'area A), e aumentando la sensibilità (il che significa minimizzare l'area C con il cut-off) diminuirà la specificità (perché aumenterà B e diminuirà D) → gli estremi sono i casi rispettivamente 1 e 2; la situazione normale è il caso 0.



C'è un modo quantitativo per legare l'andamento di sensibilità e specificità al variare del cut-off? La risposta è ovviamente sì.

Facciamo qualche premessa: sappiamo che la specificità è il rapporto tra veri negativi e sani, e che l'insieme dei sani è la somma di veri negativi e falsi positivi.

Segue che  $SP = \frac{VN}{FP+VN} = \frac{VN+FP-FP}{FP+VN} = 1 - \frac{FP}{FP+VN}$ , con l'ultimo termine che rappresenta ovviamente la probabilità di falso positivo P(FP), esattamente come la sensibilità rappresenta la probabilità di vero positivo).

Da qui:  $P(FP)=1-SP$ .

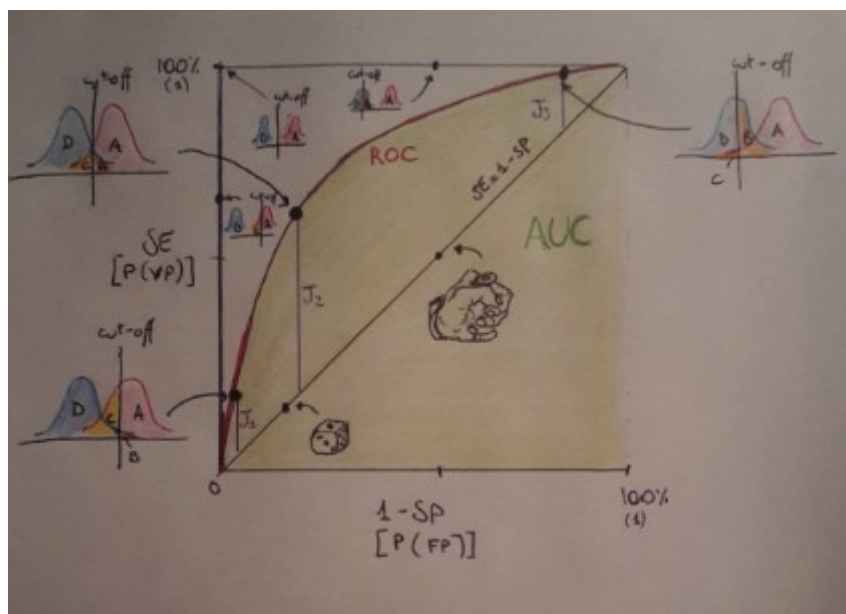
Calcolando le coppie SE e 1-SP per ogni valore di cut-off (come  $\frac{A}{A+C}$  e  $\frac{B}{B+D}$  rispettivamente) posso metterle in un grafico, detto curva **ROC** (receiver operating characteristic), ottenendo ciò che vedete sotto.

Alcune caratteristiche di questo grafico, che non credo dobbiate conoscere ma possono aiutarvi a capirne il senso, sono:

- la bisettrice del grafico è la linea tale per cui  $SE=1-SP$ , cioè la linea del test inutile (classificatore – cioè “test” - random), che può essere paragonato ad un dado ( $SE=16\%$ ,

SP=84%) o ad una moneta (SP=SE=50%)

- la linea blu, che comprende una linea che va verticalmente dall'origine al punto (0,1) e da questo punto al punto (1,1), rappresenta un classificatore perfetto, capace in teoria di separare perfettamente i sani dai malati
- l'area sotto la curva è un indice di quanto buono è il classificatore (cioè, in ultima analisi, di quanto sono separabili i sani dai malati):
  - se l'area è 0.5 non sono in alcun modo separabili → malati e sani sono sovrapposti in base al test: per fare un esempio, se il mio test è “misuro l'altezza di un paziente per stabilire se ha il cancro o no” otterrebbe un'area di 0.5; posso variare il mio cut-off, dicendo per esempio “quelli che sono sopra 2 metri sono malati” oppure “quelli che dicono sono sopra 1.95 metri sono malati” e così via, ma alla fine sempre di fare a caso si tratta, e le curve di malati e sani saranno sovrapposte in base alla mia caratteristica – sia i malati che i sani si distribuiranno allo stesso modo e con la stessa media rispetto al parametro altezza
  - se l'area è 1 malati e sani sono già separati → per esempio, se io fossi nella Terra di Mezzo e il mio test fosse “guardo quanto uno è alto per dire se è un hobbit o un gigante” allora le due curve saranno già separate rispetto all'altezza; infatti, non c'è neanche un hobbit alto come un gigante, né un gigante alto come un hobbit. Variando il cut-off posso solo dimostrare di essere stupido in questo caso (vedi esempi dentro il grafico)
  - in sostanza è intuitivo capire che l'AUC (area under the curve) varia tra 0.5 e 1, e tanto più è vicina a 1, tanto migliore è il test → l'AUC ci aiuta a discriminare tra i test, e ci dice se un test è meglio di un altro, ma non ci dà informazioni su quale cut-off sia meglio!
- L'indice J di Youden ( $J=SE+SP-1$ ) è invece la distanza verticale di un punto sulla curva ROC dalla bisettrice del piano; infatti, dato un punto con  $SE=y$  sulla curva ROC, questo avrà un valore di  $1-SP=x$ , cui a sua volta corrisponderà sulla bisettrice un punto  $y=1-SP$ . Facendo la differenza tra l'y della curva ROC e il corrispondente y della bisettrice (o linea del classificatore random) ottengo proprio  $SE-[1-SP]$ , che è J. Questo indice viene utilizzato per decidere in quello specifico test cui corrisponde quella specifica curva ROC quale cut-off usare: il cut-off prescelto è in genere quello che massimizza J, che in questo senso rappresenta la probabilità di fare una decisione corretta ( $J=1$ ) rispetto allo sparare a caso ( $J=0$ ).  
Notate infatti che il test perfetto, con la ROC in blu, ha cut-off migliore nel punto (0,1), dove  $J=1$ , mentre il test random (test inutile), coincidente con la bisettrice, non ha alcun cut-off migliore, essendo  $J=0$  sempre.  
→ ATTENZIONE: AUC discrimina tra diversi test, J discrimina il miglior cut-off nel singolo test in esame.



I cut-off sono influenzati anche dal fine del nostro test, quindi il discorso fatto sull'indice di Youden va un po' rivisto alla luce dell'applicazione in clinica. Per esempio, un test per determinare se un soggetto ha avuto o meno un infarto sarà preferibile che sia molto sensibile, in modo che al massimo io trattenga falsi positivi, ma non mandi a casa falsi negativi che potrebbero morire nel giro di poche ore.

- La retta che va dal punto (0,1) al punto (1,0), cioè la perpendicolare alla bisettrice, interseca la curva ROC in un punto, che divide la curva stessa in due metà simmetriche nel caso in cui la deviazione standard delle due popolazioni di malati e sani sia la stessa; nel punto di intersezione tra questa retta e la curva ROC si ha che  $SE=SP$  (la retta infatti è  $y=1-x$ , e considerando che  $x$  a sua volta è  $1-SP$  e  $y$  è  $SE$ ...).

Nel caso in cui le due popolazioni presentino diversa deviazione standard (eteroschedastiche – anche detto eteroscedastiche – cioè “che hanno diversa varianza”, contrariamente a omoschedastiche, che sono le curve viste fin'ora) la curva apparirà deformata. Considerate che in genere la devianza è maggiore nei malati (in teoria sono di meno e hanno range più ampi di valori possibili).

- Il coefficiente angolare della tangente alla curva ROC in un punto rappresenta il LR+; dunque la derivata della curva ROC esprimerà la variazione del LR+ al variare del cut-off. Per ulteriori dimostrazioni vedete [questo articolo](#); non è complicatissimo ma capisco che possiate trovarlo tale, soprattutto se non avete una grande infarinatura di analisi matematica. Lo troverete in appendice sintetizzato, alla fine di queste dispense [ma non è fondamentale che lo sappiate, anzi – è solo per curiosità].

Ancora qualche concetto sulla curva ROC:

1. Per la determinazione del cut-off conta anche la prevalenza della malattia. Un esempio pratico chiarirà il concetto.

Possiamo vedere sensibilità e specificità in termini di errore: un test sensibile all'80% commetterà un errore del 20% sui malati (identificandoli come non malati), mentre un test specifico all'80% commetterà un errore del 20% sui sani (identificandoli come malati).

Fatta questa premessa, prendiamo in esame una malattia come il diabete, che ha prevalenza del 10% nella popolazione. Privilegeremo un test più sensibile, uno più specifico o uno ugualmente sensibile e specifico? Si tratta di impostare il nostro cut-off in modo tale da commettere il numero totale di errori minore possibile.

Osservate la seguente tabella: presi 10 soggetti a caso dalla popolazione, 1 avrà il diabete (prevalenza del 10%).

	Specificità	Sensibilità	Errori sani	Errori malati	Errori totali
Test specifico	0,9	0,7	$0,1 \times 9 = 0,9$	$0,3 \times 1 = 0,3$	$0,3 + 0,9 = 1,2$
Test sensibile	0,7	0,9	$0,3 \times 9 = 2,7$	$0,1 \times 1 = 0,1$	$2,7 + 0,1 = 2,8$
Test "uguale"	0,8	0,8	$0,2 \times 9 = 1,8$	$0,2 \times 1 = 0,2$	$1,8 + 0,2 = 2$

Ovviamente il test con cui commetto meno errori totali è quello specifico, perché la malattia ha prevalenza minore del 50%; se avesse una prevalenza del 50% privilegierei il test ugualmente specifico e sensibile, mentre se avesse una prevalenza maggiore del 50% privilegierei il test sensibile (fatevi pure gli esempi, vi aiuteranno a capire).

Questo ovviamente tralasciando casi specifici come l'infarto (vedi sopra): per malattie potenzialmente mortali in breve tempo è chiaro che privilegerò il test sensibile anche se la malattia è a bassa prevalenza (nei limiti del ragionevole).

Tuttavia, abbiamo considerato la prevalenza nella popolazione generale; il lavoro del clinico è anche quello di far variare i cut-off in base ai fattori di rischio. Se viene da me un paziente obeso dovrò considerare che la sottopopolazione degli obesi ha una prevalenza molto maggiore di diabete rispetto alla popolazione generale, dunque per fare un test utile abbasserò il cut-off, aumentando la sensibilità e diminuendo la specificità, per diminuire il totale degli errori commessi.

2. Un'altra considerazione su AUC: in sostanza è il valore, variabile tra 0,5 e 1, che ci dice quanto è informativo un test (cioè quanto è utile).

Tuttavia, vedendo il range di valori che può assumere ci chiediamo se per caso non rappresenti una probabilità. La risposta è affermativa, ma è una probabilità molto particolare: è la probabilità di identificare un malato con la seguente procedura

- Si prendono un malato e un non malato a caso
- Considero malato quello con il valore della variabile maggiore (ma è ovvio che dipende dal test! Ora abbiamo in mente l'esempio della glicemia, ma se stessimo valutando ipotesi contro normotesi gli ipotesi avrebbero valori della variabile minori, e considererei malato quello con il valore minore!), e non malato quello con il valore della variabile testata minore

L'AUC è la probabilità che quello che ho considerato come malato operando in questo modo sia davvero il malato; esprime quindi la capacità del test di discriminare tra sani e malati, ma NON è la probabilità di fare giusto in generale, dato che conta anche la prevalenza di malati nella popolazione (se prendo due soggetti a caso dalla popolazione in una malattia con prevalenza del 10% ho l'81% di possibilità che siano due sani –  $0,9 \times 0,9$  – il 18% che siano uno malato e uno sano –  $0,9 \times 0,1 \times 2$ , dove ho moltiplicato per due volte perché devo contare che posso pescare prima un malato e poi un sano o viceversa – e l'1% che siano due sani –  $0,1 \times 0,1$ ; dunque solo nel 18% dei casi l'AUC mi dirà la probabilità di far giusto).

Per chiarire meglio, pensiamo al test perfetto: malati e sani sono perfettamente separati – se prendo un malato e un sano il malato avrà sempre un valore maggiore della variabile rispetto al sano, per cui la mia probabilità di fare giusto col metodo di cui sopra sarà esattamente 1.

Pensiamo invece al test inutile: se malati e sani sono sovrapposti perfettamente, prendendo un malato e un sano avrò il 50% di probabilità che il malato abbia un valore maggiore della variabile rispetto al sano, e il 50% che non lo abbia. Di conseguenza, la mia probabilità di fare giusto col metodo di cui sopra sarà esattamente 0.5.

\*\*\*\*

## 5.1 – Probabilità

La medicina, la statistica e la probabilità sono intimamente legate.

“La diagnosi è come un bersaglio mobile: ogni volta che si raccoglie un nuovo elemento diagnostico, aumentano le probabilità di una malattia e diminuiscono quelle di un'altra” (Kassirer).

Capita spesso che alla diagnosi non ci si arrivi proprio, nonostante numerosi esami; sta al medico decidere quando è stato fatto abbastanza. A tal proposito sono emersi dei nuovi concetti nella pratica medica:

- Less is more → a volte far meno, o non far nulla, è meglio che fare troppo
- Choose wisely → scegliere saggiamente tra le varie possibilità terapeutiche e tra i vari test
- Slow medicine → <http://www.slowmedicine.it/>

L'origine dell'utilizzo della probabilità in medicina si trova nell'estrema variabilità di sintomi, segni, test. Occorre considerare per esempio che è più probabile imbattersi in segni atipici di una malattia comune piuttosto che in segni tipici di una malattia rara!

Anche le cosiddette “certezze” variano in medicina, perché variano le soglie per discriminare patologici e normali, variano le linee guida, cambiano le conoscenze.

Inoltre, varia anche il giudizio clinico, tanto che uno stesso evento clinico può essere interpretato diversamente non solo da osservatori diversi, ma anche dallo stesso osservatore a distanza di tempo. In genere si può affermare che un'osservazione non è sempre riproducibile, e i fattori a cui imputare questo fatto sono relativi all'esaminatore, all'esaminato e all'esame.

→ per fare un esempio, nel 1945 H. Bawkin studiò la variabilità del giudizio medico, facendo esaminare:

- 389 scolari di 11 anni da un primo gruppo di medici: a 174 di loro venne raccomandata la tonsillectomia (45%)
- 215 di questi furono rivisitati da un secondo gruppo di medici: a 99 di loro si raccomandò la tonsillectomia (46%)
- 116 furono visitati da un terzo gruppo, che raccomandò la tonsillectomia a 51 (44%)

A prescindere dallo stato reale delle sue tonsille, ogni bambino intervistato correva circa il 45% di possibilità di perderle.

Una stessa patologia viene curata in modo diverso da medici diversi, e per limitare l'eterogeneità di comportamento vengono proposte le linee guida.

Tuttavia, anche queste vanno applicate criticamente: In *JAMA (2005)* è stata pubblicata la simulazione di cura di una pz di 79 anni, affetta da artrosi, osteoporosi, diabete di tipo 2, ipertensione e bronchite cronica. Applicando le linee guida specifiche per ogni patologia avrebbe dovuto:

- assumere 12 farmaci 5 volte al giorno,
- seguire 14 diversi suggerimenti su dieta e movimento
- sottoporsi a più visite ed esami di controllo ogni anno

L'incrocio avrebbe prodotto interazioni tra farmaci, tra farmaci e alimentazione, tra farmaci per una malattia e decorso di un'altra. Purtroppo, i pazienti reali sono anche più complessi!

Bisogna cercare un trattamento che permetta al paziente di essere compliant: per esempio, per evitare che il paziente prenda troppi farmaci eviterò di somministrargliene di relativi a patologie “dormienti” nel caso in cui abbia più patologie.

Per quanto riguarda invece la variabilità dell'esaminato:

- In uno studio su un campione casuale 155 pazienti di medici di base, si è constatato che in un quinto delle anamnesi non erano stati riferiti importanti elementi clinici quali melena o pollachiuria (Hoornaert 1977). Le omissioni erano più frequenti non quando i pazienti erano ostili nei confronti del medico, ma quando lo apprezzavano e ne lodavano le doti. Il paziente affezionato al suo medico tende inoltre, per compiacerlo, ad occultare la presenza e la rilevanza anche degli effetti collaterali della terapia (Snyder 1976).

Il problema della variabilità in medicina si estende anche agli studi. Per fare un esempio, gli studi

clinici controllati spesso non fanno il confronto con terapie la cui utilità è già provata ma con placebo, e riguardo la sicurezza d'uso spesso non la rilevano o sono troppo piccoli per evidenziarla.

→ I FANS possono danneggiare lo stomaco. Negli USA il numero di morti per sanguinamento da FANS è pari a quello di morti da AIDS e crimini violenti.

Gli antinfiammatori coxib, con gastrolesività dimezzata, sembravano sicuri per un uso prolungato in pazienti con artrite reumatoide e ulcera. Studi post marketing (dopo la messa in commercio del farmaco) hanno tuttavia messo in luce l'associazione tra uso prolungato di rofecoxib e aumento di infarto del miocardio e ictus.

Per molte malattie vi sono molti fattori di rischio, che ne cambiano la probabilità di verificarsi pur non causandola necessariamente.

Invece che parlare di causa necessaria, in questi casi si parla di complessi necessari, cioè un insieme di condizioni e di fattori che se si presentano causano la malattia, e di complesso sufficiente è l'insieme dei fattori che sono sufficienti a causare la malattia, ma non la causano necessariamente. Ciò spiega perché per esempio il fumo a volte non causa il cancro – il fumo è un componente del complesso sufficiente, non del complesso necessario. Il complesso sufficiente è l'insieme dei fattori che sono sufficienti a causare la malattia, ma non la causano necessariamente.

Occorre osservare che molte patologie sono determinate da più fattori, e che lo stesso fattore potrebbe essere implicato in più malattie con diverso peso.

Anche da ciò nasce il bisogno di utilizzare il calcolo delle probabilità.

La probabilità di un evento è definita come il rapporto fra il numero di esiti di un esperimento stocastico (casuale) in cui l'evento stesso si verifica e il numero totale di esiti, equiprobabili. Il numero totale degli esiti viene definito spazio degli eventi o universo stocastico.

## 5.2 – Elementi di calcolo combinatorio

Non è necessario che seguiate questo paragrafo; è un'aggiunta mia con dimostrazioni non spiegate, ma serve a mio avviso per capire quale approccio usare a seconda del problema. Se non avete voglia di leggere, balzate pure al prossimo paragrafo.

Immaginate di avere un insieme di qualunque tipo: per semplicità supporremo l'insieme  $\{A,B,C,D\}$ , costituito quindi da 4 elementi.

Possiamo chiederci un po' di cose ora:

- In quanti modi posso disporre questi elementi?
- In quanti modi posso ordinare questi elementi?
- In quanti modi posso combinare questi elementi?

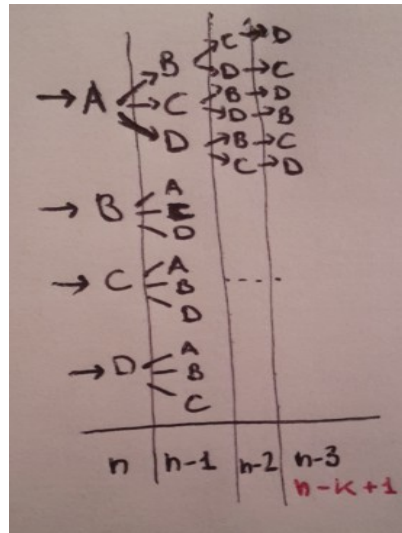
Analizziamo un aspetto per volta, dato che forse non è immediato capire cosa intendiamo con ciascun termine usato.

1. Disposizione: disporre degli elementi vuol dire metterli in una successione lunga un certo numero di elementi. Per esempio, possiamo disporre i nostri 4 elementi in successioni da 3, e per esempio avremo ABC, ACB, AAA, ecc.

Possiamo però distinguere due casi: le disposizioni semplici (in cui gli elementi non si possono ripetere, quindi non potrò disporre i miei elementi come AAA) e le disposizioni con ripetizioni (in cui posso anche ripeterli; ovviamente l'insieme di queste disposizioni conterrà anche l'insieme delle disposizioni semplici).

Analizziamo ora questi due casi, tenendo presente che per una disposizione conta l'ordine, cioè ABC è diverso da BCA anche se gli elementi sono gli stessi, e quindi dovrò contare ENTRAMBE le successioni.

1. Disposizioni semplici: possiamo immaginare un metodo grafico per generare il numero delle nostre disposizioni, cioè il seguente:



In sostanza, vedete che il procedimento usato consiste nell'iniziare la successione con uno dei 4 elementi. A questo punto, per ognuno di questi quattro elementi ho solo altri 3 elementi come alternative possibili per continuare la successione (non posso riusare quello che ho usato al livello precedente); al passaggio successivo ne avrò 2, e a quello successivo solo 1. Ho completato l'albero solo per le disposizioni che iniziano con A, ma è evidente che il procedimento è analogo per tutte.

Un modo elegante per esprimere il numero di alternative possibili ad ogni livello (le nostre colonne) è utilizzare  $n$ , cioè il numero di elementi. Nel primo livello li posso usare tutti, quindi avrò  $n$  alternative. Nel secondo per ogni elemento ne potrò usare  $n-1$ , poi  $n-2$  e così via.

Di conseguenza, il numero di disposizioni al primo livello sarà  $n$ , al secondo sarà  $n(n-1)$ , al terzo  $n(n-1)(n-2)$  e così via.

Occorre notare che io posso decidere di troncare la mia successione quando voglio (per esempio posso voler sapere quante disposizioni lunghe 2 elementi (quindi AB, AC, AD, BA, BC e così via) ci sono; un modo elegante per dirlo è definire la disposizione come di classe  $k$ , con  $k$  numero di elementi della successione. Se io decido che voglio sapere quante disposizioni di questi 4 elementi lunghe 2 simboli ci sono, dirò che cerco le disposizioni di 4 elementi di classe 2.

Sembra ovvio dal grafico che tali disposizioni sono 12, cioè  $4(4-1)$ . Ma come esprimere questo facendo rientrare anche  $k$ ?

Beh, ragioniamo per casi: se  $k$  fosse 4, avremmo  $4(4-1)(4-2)(4-3)=24$  disposizioni. Abbiamo moltiplicato fino a  $n-3$ ; ma 3 rispetto a  $k=4$  è esattamente  $k-1$ . Se  $k$  fosse 3, avremmo  $4(4-1)(4-2)=12$  disposizioni; abbiamo moltiplicato fino a  $n-2$ , e 2 è ancora  $k-1$ .

Generalizzando, moltiplichiamo quindi  $n(n-1)(n-2)\dots(n-(k-1))$ , che è quindi come moltiplicare  $n(n-1)(n-2)\dots(n-k+1)$ .

Introduciamo ora una notazione particolare, il fattoriale:  $n!$ , definito come segue:

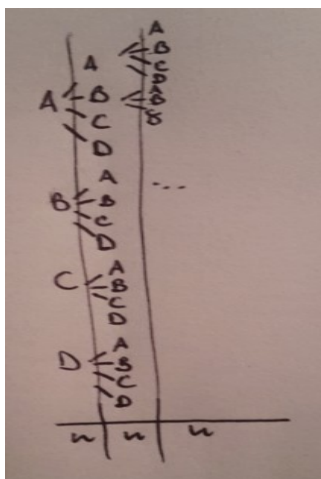
$n! = n(n-1)(n-2)\dots(n-n+1)$ ; si tratta di moltiplicare tutti i numeri compresi tra 1 e  $n$ .

In questo senso possiamo manipolare la nostra formula per le disposizioni semplici moltiplicando per  $(n-k)!/(n-k)!$ . Notiamo quindi che il numeratore di:

$$D_{n,k} = \frac{n(n-1)(n-2)\dots(n-k+1)(n-k)!}{(n-k)!}$$

è esattamente  $n!$  (infatti mancavano da

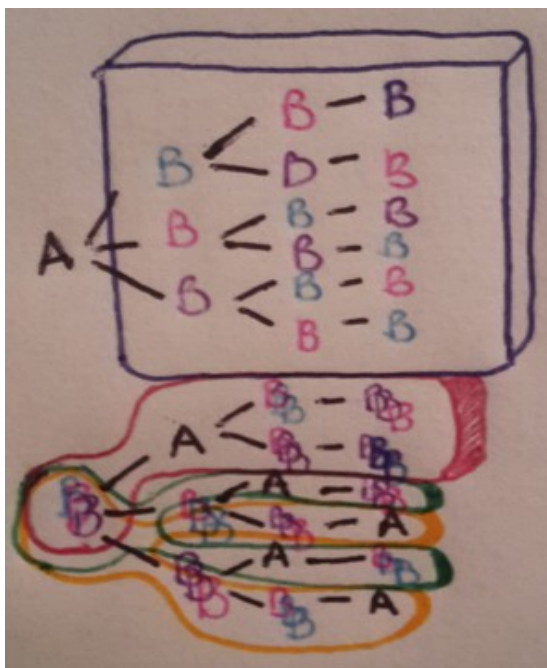
moltiplicare esattamente i numeri tra 1 e  $n-k$ ), e otteniamo  $D_{n,k} = \frac{n!}{(n-k)!}$ , che è il numero di disposizioni semplici di classe  $k$  di  $n$  elementi.



2. Disposizioni con ripetizioni: questo caso è più semplice. Ricostruendo l'albero che abbiamo usato prima, ora per ogni livello abbiamo sempre 4 alternative possibili; quindi, continueremmo a moltiplicare  $n$  per  $n$  esattamente  $k$  volte, il che equivale a scrivere  $D'_{n,k} = n^k$ , che è il numero di disposizioni con ripetizioni di classe  $k$  di  $n$  elementi.
2. Permutazione: permutare significa “cambiare l'ordine”. In questo senso voler conoscere le permutazioni significa voler sapere in quanti modi possibili posso scambiare gli elementi di un insieme. Anche in questo caso posso avere permutazioni semplici (in cui tutti gli elementi sono diversi tra loro) e permutazioni con ripetizione (in cui nell'insieme sono presenti più elementi uguali).

Ora, quando permuto non mi fermo ad una classe  $k$  ovviamente; partendo da una disposizione, che è il mio insieme, per esempio  $\{A,B,C,D\}$ , voglio sapere in quanti modi li posso scambiare. In questo senso una permutazione è come una disposizione di classe  $n$  (cioè con  $k=n$ ). Pertanto, è abbastanza evidente che una permutazione semplice sarà  $P_n = n!$  [occorre notare che  $(n-n)! = 0! = 1$ , dato che il [fattoriale](#) di 0 è [per convenzione](#) 1).

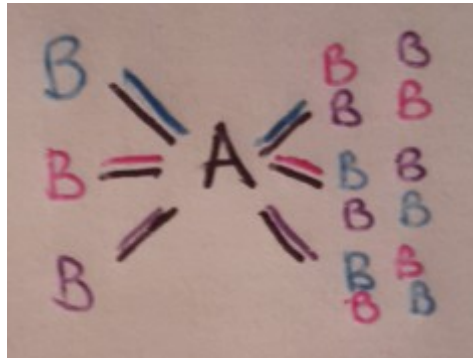
Più complesso e non così intuitivo è il discorso sulle permutazioni con ripetizione: in questo caso mi interessa sapere il numero di permutazioni distinte sapendo che nel mio insieme ci sono elementi che si ripetono; per esempio, il mio insieme è  $\{A,B,B,B\}$ . Osservate la schematizzazione qui sotto:



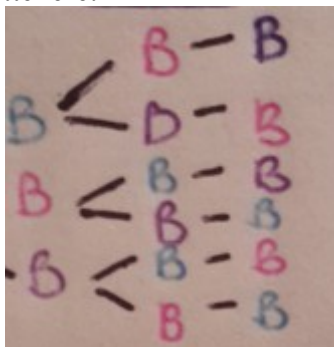
Sostanzialmente, alla seconda riga della prima colonna ho raggruppato le 3 B possibili

(segnate con colori diversi: azzurro, rosa, viola – anche se nella realtà un B è indistinguibile dalle altre ovviamente) in un concetto grafico solo. Ho poi preso le permutazioni uguali, cioè quelle con la stessa successione (sono solo 4 possibili, AB~~BB~~, BA~~BB~~, B~~BA~~B, B~~BB~~A), e le ho racchiuse in figure blu, rosse, verdi e arancioni. Notate che se contate ognuno di questi gruppi contiene esattamente 6 permutazioni uguali.

Il fatto è proprio questo; per ogni permutazione unica (cioè per ogni successione distinta di n elementi con k elementi uguali tra loro) ci sono k! ripetizioni. È facile capire perchè proprio k!. Infatti, prendete il primo caso: ho preso A, e poi ho riquadrato in blu esattamente le permutazioni dell'insieme {B,B,B}, in cui tutti gli elementi sono uguali; in questo caso ho 3! permutazioni, cioè 6, e tutte queste 6 permutazioni sono uguali. Ma in fondo, nel caso per esempio in cui io metta la A al secondo posto, posso valutare le varie alternative come segue:



Cioè posso eliminare la A e ottenere:



Che è esattamente lo stesso insieme di permutazioni del riquadro blu di due disegni fa; le permutazioni sono ancora 3!. Analogo ragionamento per la A al secondo e al terzo posto.

Posso quindi disporre le mie 24 permutazioni in una tabella in modo tale da avere le permutazioni DIVERSE in riga, e le 6 permutazioni UGUALI per ognuna di queste in colonna:

k!	AB BB	BA BB	B BAB	B B B A
	AB BB	BA BB	B BAB	B B B A
	AB BB	BA BB	B BAB	B B B A
	AB BB	BA BB	B BAB	B B B A
	AB BB	BA BB	B BAB	B B B A
	AB BB	BA BB	B BAB	B B B A

In generale, otterrò un rettangolo di area n! (con n elementi dell'insieme) e altezza k! (con k

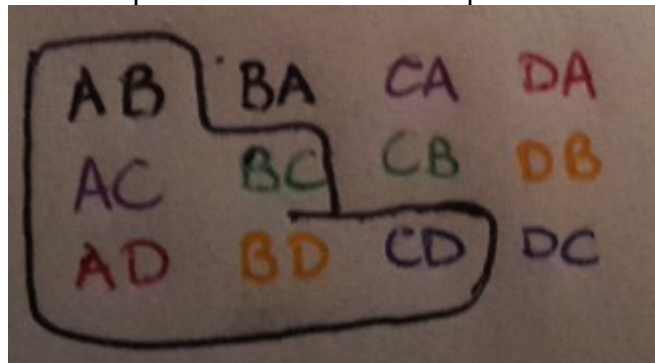
elementi uguali nell'insieme), ragion per cui la mia base (permutazioni diverse) sarà data da area/altezza, ovvero:  $P_n^k = \frac{n!}{k!}$ .

Non vi dimostro la seguente formula, ma è abbastanza semplice: nel caso in cui ci siano più elementi diversi ripetuti  $k_1, k_2$ , ecc volte (per esempio  $\{A, B, B, C, C, C\}$  con  $k_1=2$  e  $k_2=3$ ) le permutazioni con ripetizione diventano  $P_n^{k_1, k_2, \dots} = \frac{n!}{k_1! k_2! \dots}$ .

3. Combinazioni: combinare significa capire in quanti modi posso prendere  $k$  elementi di un insieme di  $n$  elementi in modo che diano una combinazione unica (è un concetto più intuitivo di quanto si pensi, dato che in fondo usiamo le combinazioni per proteggere i nostri dati o i nostri oggetti). Per esempio, voglio sapere in quanti modi posso prendere 2 elementi dall'insieme  $\{A, B, C, D\}$  in modo che diano una combinazione unica. In questo caso NON conta l'ordine, perché  $AB$  è esattamente la stessa cosa di  $BA$  (se l'ordine contasse tornerei infatti al caso delle disposizioni no?).

Anche in questo caso avrò combinazioni semplici (in cui gli elementi della singola combinazione sono tutti distinti) e combinazioni con ripetizione (in cui posso prendere più volte lo stesso elemento dell'insieme).

1. Partiamo dalle semplici: osservate la tabella qui sotto.



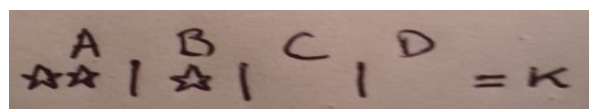
Rappresenta l'insieme delle disposizioni semplici di classe 2 dell'insieme  $\{A, B, C, D\}$ . L'area riquadrata è l'insieme delle combinazioni semplici di classe 2 dello stesso insieme; infatti in quest'area ogni combinazione è unica (esattamente come nell'area complementare). Possiamo però fare un ragionamento intelligente: in effetti, se decido di escludere le combinazioni con gli stessi elementi disposti in ordine diverso significa che decido di escludere le permutazioni semplici (perché gli elementi dell'insieme sono tutti distinti) dei  $k$  elementi. Quindi realizzo che per ogni combinazioni ci saranno  $P_k$  combinazioni uguali, e il totale mi dà esattamente le disposizioni di classe  $k$  di  $n$ . Dunque:  $C_{n,k} \cdot P_k = D_{n,k}$ , ovvero

$$C_{n,k} = \frac{D_{n,k}}{P_k} = \frac{n!}{(n-k)! k!}$$

Tale espressione è indicata anche come coefficiente binomiale, con la seguente notazione:  $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ , e ha [proprietà peculiari](#).

2. Per quanto riguarda le combinazioni con ripetizione il discorso si fa più complesso. Immaginate di avere un insieme  $\{A, B, C, D\}$ , e di voler vedere in quanti modi potete prendere delle combinazioni di 3 elementi senza vincoli sul fatto che siano o meno distinti.

Possiamo immaginare di rappresentare graficamente le combinazioni in questo modo:



Sostanzialmente, la nostra combinazione AAB viene definita da un'equazione detta "a stelle e barre", che ha queste proprietà:

- l'equazione contiene k stelle e n-1 barre
- le barre separano le posizioni relative agli elementi dell'insieme (c'è una barra tra le A e le B, tra le B e le C, tra le C e le D – sono n-1 perché non serve una barra tra le D e ciò che viene dopo le D); le barre non separano elementi uguali
- le stelle corrispondono al numero di volte che viene ripetuto un elemento nella combinazione in esame (quindi nel nostro caso avrò due stelle subito, che significa due A, poi una barra, poi una stella che significa B e poi due barre, dato che tra la barra B-C e la barra C-D e appena dopo la barra C-D non c'è nulla: non ci sono C, né D)
- l'equazione è soddisfatta quando ci sono k+n-1 simboli e quando ci sono k stelle (sembra una stupidata detta così ok, ma è che abbiamo definito un'equazione non proprio convenzionale, in cui dei simboli sono uguali ad un numero, quindi meglio precisare)

Occorre notare che l'equazione a stelle e barre ci dice solo QUANTI singoli elementi ci sono (2 A, 1 B, 0 C, 0 D), non l'ordine di questi (che può essere AAB, ABA, BAA), dato che stiamo valutando combinazioni e non disposizioni! In questo modo ogni soluzione dell'equazione identifica biunivocamente una combinazione.

In ogni caso, il problema si riduce quindi a determinare in quanti modi possono essere disposte le k stelle nelle k+n-1 posizioni, dato che ogni disposizione di k stelle rappresenta una combinazione. Ma questo sappiamo già farlo: infatti è semplicemente una permutazione con ripetizione con esattamente k simboli uguali tra loro e altri n-1 simboli uguali tra loro su k+n-1, dato che ora il nostro insieme è {k stelle, n-1 barre}. Applicando la formula, otteniamo quindi:

$$C'_{n,k} = P_{k+n-1}^{k,n-1} = \frac{(k+n-1)!}{k!(n-1)!} ; \quad \text{tuttavia, osserviamo che riarrangiando}$$

$$C'_{n,k} = \frac{(k+n-1)!}{k!(k+n-1-k)!} = \binom{k+n-1}{k} , \quad \text{oppure, del tutto analogamente:}$$

$$C'_{n,k} = \frac{(k+n-1)!}{(n-1)!(k+n-1-(n-1))!} = \binom{k+n-1}{n-1} \quad (\text{ci si poteva arrivare anche}$$

attraverso le proprietà del coefficiente binomiale, dato che  $\binom{k+n-1}{k} = \binom{k+n-1}{n-1}$ )

Giusto per capire cosa fare, vi metto di seguito un "albero decisionale", che vi spiega come agire di fronte ad un problema che potrebbe richiedere l'utilizzo del calcolo combinatorio.

Come esercizio, provate a partire dalle domande poste come esempio per ogni possibilità e cercate di risalire al procedimento da utilizzare attraverso l'albero decisionale.

Facciamo un esempio partendo dalla domanda: "quante squadre di basket posso formare scegliendo fra 10 giocatori?".

Al passo 1 ci chiediamo: "devo sapere in quanti modi devo ordinare i 10 giocatori?" e la risposta è ovviamente no; devo infatti ricavare squadre da 5.

Al passo 2 ci chiediamo: "devo costruire una successione di k elementi dei 10 giocatori?" e la risposta è sì: devo costruire successioni di 5 giocatori dai 10 iniziali.

Al passo 3 ci chiediamo: "l'ordine conta?" e la risposta è no. Infatti, se prendo la squadra con i giocatori 34567 questa sarà esattamente uguale alla squadra con i giocatori 45637, quindi dovrò contarle come una sola squadra e non come due squadre distinte.

Al passo 4 ci chiediamo: "posso ripetere più volte lo stesso elemento nella squadra?" e ovviamente la risposta è no. Infatti una squadra fatta di 11111 non avrebbe senso, a meno che non ci giochi Kobe Bryant. Dunque, ciò che voglio è una combinazione semplice, e applicando la formula ottengo ben 252 squadre possibili. Buona partita!

# Insieme I di n elementi

$I = \{s_1, s_2, \dots, s_n\}$   
 Dovete sapere in quanti modi potete scartare di posto (ordine) gli elementi di I?

Dovete sapere quanti sottoinsiemi di k elementi di I? [cioè sottoinsiemi "lunghe" k elementi]

II problema non vi risolve col calcolo combinatorio

Sottrazioni con gli stessi elementi al posto in modo diverso sono sottrazioni diverse? Lavoro: la sottrazione  $xy$  è diversa da  $yx$ ? C'è un ordine?

**PERMUTAZIONI SENZA RIPIETIZIONE**  
 $P_n = n!$   
 Quante parole si possono formare con le lettere della parola A?

**PERMUTAZIONI CON RIPIETIZIONE**  
 $\frac{n!}{k_1! k_2! \dots k_r!}$   
 In quanti modi posso scrivere 4 palline, gli  $k_i$  3 volte e una  $k_r$ ?

Fatto ripetere più volte lo stesso elemento nella sottrazione?

Fatto ripetere più volte lo stesso elemento nella sottrazione?

Fatto ripetere più volte lo stesso elemento nella sottrazione?

**COMBINAZIONI CON RIPIETIZIONE**  
 $C_{n+k-1}^k$   
 In quanti modi posso assegnare 2 caramelle tra 3 bambini?

**COMBINAZIONI SENZA RIPIETIZIONE**  
 $C_{n,k} = \binom{n}{k}$   
 Quante squadre di basket posso formare scegliendo fra 10 giocatori?

**DISPOSIZIONI CON RIPIETIZIONE**  
 $D_{n,k} = n^k$   
 Quante password 4 cifre si possono formare?

**DISPOSIZIONI SENZA RIPIETIZIONE**  
 $D_{n,k} = \frac{n!}{(n-k)!}$   
 Quante parole di 4 lettere si possono formare con 5 consonanti?

Detto questo, nel prossimo paragrafo vi sarà più semplice capire perché si fa in un modo e non in un altro. Spero che la vostra comprensione giustifichi lo sforzo fatto per cercare di spiegare nel modo più semplice possibile questi concetti.

### 5.3 – Problemi riguardanti il calcolo delle probabilità

#### PROBLEMA 1: Problema dei dadi dello Chevalier de Méré

“È più conveniente scommettere sull'uscita di almeno un 6 al lancio di un dado 4 volte, o di almeno una coppia di 6 al lancio di una coppia di dadi 24 volte?”

→ Ragionamento stupido: la probabilità che esca un 6 è di  $1/6$ , mentre la probabilità che esca una coppia di 6 è  $1/36$ . Quindi la probabilità della coppia di 6 è  $1/6$  della probabilità che esca un 6; se “do” all'evento “coppia di 6” 6 volte le occasioni di verificarsi rispetto all'evento “esce un 6” i due eventi si verificheranno con la stessa probabilità.

→ Ragionamento corretto: lo dividerò per punti in modo che possiate seguire il procedimento logico.

- Se lancio un singolo dado 4 volte posso ottenere vari esiti: per esempio, possono uscire nell'ordine 1-2-3-4, oppure 6-4-6-3, oppure 3-3-3-3, e così via. Ma come faccio a valutare qual è il numero TOTALE di queste successioni dato questo problema? Se guardate nel paragrafo precedente, vi renderete conto subito che questa domanda vi chiede le disposizioni con ripetizione. Infatti esse rappresentano il totale degli eventi possibili.

Sappiamo quindi che il totale degli eventi (o universo degli eventi) è  $6^4=1296$ .

Vogliamo determinare quante volte si verifica l'evento “almeno un 6”. Iniziamo quindi per semplicità a determinare quando non si verifica “nessun 6”: questo è più semplice, perché equivale a ricavare il sottoinsieme dell'universo degli eventi senza 6 (il che è concettualmente equivalente a ricavare un nuovo universo degli eventi con un dado senza il 6, cioè con un dado a 5 facce). È evidente che questo universo ha dimensione  $5^4=625$ .

Dunque, il numero di eventi in cui ho almeno un 6 è  $1296-625=671$ , e la probabilità di ottenere almeno un 6 sarà data da  $671/1296=0.52$ , che se ci pensate è una probabilità parecchio alta rispetto all' $1/6$  iniziale.

Valutiamo ora la probabilità della coppia di 6 in 24 lanci. Come prima, chiediamoci il numero di eventi possibili, cioè le disposizioni con ripetizione. Consideriamo che lanciando una coppia di dadi ci sono 36 esiti possibili, tutti equiprobabili (in sostanza equivale a lanciare un dado a 36 facce); dunque l'universo degli eventi è  $36^{24}=2.245 \times 10^{37}$  esiti possibili.

L'evento “non esce coppia di 6” equivale a lanciare un dado a 35 facce, senza la faccia “coppia di 6”; dunque occorre in  $35^{24}=1.142 \times 10^{37}$  esiti. L'evento “coppia di 6” occorre quindi nella differenza degli esiti, cioè  $(2.245-1.142) \times 10^{37}=1.103 \times 10^{37}$  volte. La probabilità totale è quindi  $1.103 \times 10^{37} / 2.245 \times 10^{37} = 0.49$ .

Conviene quindi scommettere sul primo evento, “almeno un 6 in 4 lanci”.

#### PROBLEMA 2: Paradosso del compleanno

“Dati 25 studenti nati nello stesso anno, è più probabile che siano nati tutti in date diverse o no?”

→ Soluzione: semplifichiamo il problema. Supponiamo che il nostro anno sia lungo solo 3 giorni, e vogliamo sapere qual è la probabilità che 3 persone siano nate in giorni diversi. Cambiano solo i numeri, non la forma: dunque il problema prevederà una soluzione che usa lo stesso ragionamento del problema enunciato all'inizio.

Iniziamo col determinare l'universo degli eventi, che mettiamo in tabella: indicherò con il colore blu la prima persona, rosso la seconda, verde la terza.

111	112	113	121	122	123	131	132	133
211	212	213	221	222	223	231	232	233
311	312	313	321	322	323	331	332	333

Abbiamo 27 possibilità in totale: ma dobbiamo chiederci se l'ordine conta oppure no. Beh, in questo caso osserviamo che l'evento alla terza casella della prima riga presenta due 1 e un

3, esattamente come l'evento alla terza casella della prima colonna. Ma queste successioni sono in ordine diverso; tuttavia, possiamo notare che questo evento ha quindi più modi di realizzarsi, dato che c'è sia la possibilità che il blu sia nato il 3, sia la possibilità che ad essere nato il 3 sia il rosso, sia la possibilità che sia il verde ad essere nato il 3, con di volta in volta i due rimanenti nati entrambi l'1. In questo caso questa particolare disposizione avrà quindi 3 possibilità di presentarsi, e dovrò considerarle come distinte. Dunque, l'ordine conta<sup>4</sup>, e il mio universo degli eventi sarà dato dalle disposizioni con ripetizione di classe 3 di 3 elementi.

\*\*\*\*\*

Torniamo al nostro problema iniziale.

Dobbiamo considerare quindi l'universo degli eventi come l'insieme delle disposizioni con ripetizione di classe 25 di 365 elementi, cioè  $365^{25}$ .

L'evento "tutti hanno un compleanno diverso" è analogamente a prima l'insieme delle disposizioni semplici di classe 25 di 365 elementi, cioè  $\frac{365!}{(365-25)!}$ .

La probabilità che tutti abbiano un compleanno diverso è dunque il rapporto tra le due (cioè il numero di volte in cui occorre l'evento desiderato sul totale degli eventi possibili):

$$\frac{365!}{365^{25}(365-25)!} = 0.43$$

Dunque, almeno 2 persone condivideranno un compleanno con il 57% di probabilità (infatti l'evento complementare di "sono nati tutti in giorni diversi" è "non sono nati tutti in giorni diversi", ovvero "almeno 2 sono nati nello stesso giorno", e la sua probabilità si può ricavare come  $1-0.43=0.57$ ).

Quindi la risposta è la seguente: è più probabile che non siano nati tutti in date diverse.

→ riguardo a questo problema il Duca segue un ragionamento diverso, che guardando un po' in Internet va per la maggiore; in ogni caso, convergono tutti sulla stessa identica formula che ho usato io, a cui si arriva in due secondi con il calcolo combinatorio (contate dall'ultima riga di asterischi, prima ovviamente non conta).

Se trovate più semplice [questo ragionamento](#) (per cui forse vi conviene vedere la versione [in inglese](#) che è un po' più completa, o [questo video](#), che trovo molto Duchesco nell'irritarmi a livelli inimmaginabili) allora seguitelo – tuttavia ritengo che il mio ragionamento sia più riproducibile rispetto a questo.

→ una cosa interessante in questo problema è trovare il numero di persone minimo per cui la probabilità che almeno 2 siano nate lo stesso giorno è maggiore della probabilità che siano nate tutte in giorni diversi; ovviamente per far ciò dovremmo

calcolare  $\frac{365!}{365^x(365-x)!} < 0.5$  e risolvere in x, e vedere poi l'intero più vicino per

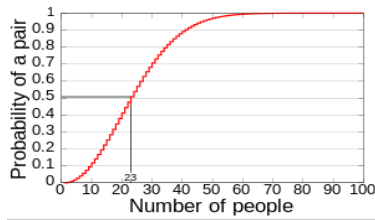
eccesso a x (l'intero per eccesso aumenta il denominatore e quindi rende il risultato sicuramente più piccolo).

Se non avessimo proprio sbatti di risolvere questa disequazione, potremmo calcolare

---

4 Se non riuscite a convincervene, prendete un esempio ancora più semplice: due persone lanciano una moneta a testa. Volete sapere qual è la probabilità che ottengano risultati diversi. Il vostro universo degli eventi è TT, TC, CT, CC. Ma TC è la stessa cosa che CT, e devo quindi contarle come uno solo? In questo caso la probabilità sarebbe 0.33, mentre nell'altro caso sarebbe ben 0.50. Qual è la probabilità giusta? Questo è un altro modo per chiedersi: "devo considerare combinazioni o disposizioni?". È intuitivo considerare che l'evento "esiti diversi" abbia due opportunità distinte di manifestarsi, dunque sarà giusta la seconda. Considerate anche che quando avete definito l'evento di interesse ("esiti diversi") avete considerato di contare tutti i modi in cui questo si verifica! Una buona prova per capire se si tratta di combinazione o disposizione è chiedervi "se ottengo una successione con gli stessi elementi di una che ho già ottenuto, ma con gli elementi scambiati di posto, devo scartarla perché l'ho già contata?". Immaginatevi ora al tavolo a provare quante volte escono "esiti diversi" dal lancio di due monete. Lanciate: esce TT, contate 1 per gli uguali e 0 per i diversi. Esce CC, contate 2 per gli uguali e 0 per i diversi. Esce TC, 2 e 1 rispettivamente. Esce CT: è un'altra volta in cui è capitato il vostro evento in studio ("esiti diversi") quindi dovete contarla, anche se ha gli stessi elementi di TC, scambiati di posto → Dunque, a voi interessano le disposizioni, e l'ordine conta.

la probabilità che almeno 2 persone condividano il compleanno per ogni insieme di  $k$  persone e vedere quando passiamo il 50%. Il risultato è 23.



### PROBLEMA 3: [Monty Hall](#)

“Ci sono 3 porte, con un premio dietro solo una di esse. Il concorrente sceglie una porta, ma non la apre. Il presentatore apre un'altra porta, per far vedere che non c'è nulla dietro, e dopo ciò offre al concorrente di cambiare porta. Dovrebbe accettare?”

→ Soluzione empirica: il premio può trovarsi in una di queste configurazioni:

Porta 1	Porta 2	Porta 3
PREMIO		
	PREMIO	
		PREMIO

Qualunque porta scelga il concorrente, il presentatore aprirà una porta vuota, in modo tale che rimangano una porta vuota e una porta con il premio.

A questo punto, immaginiamo che il concorrente abbia scelto la 1. Abbiamo le seguenti possibilità (fatto salvo che nella 1 è indifferente che il presentatore apra la porta 2 o 3; arbitrariamente diciamo che apre la 3):

Porta 1	Porta 2	Porta 3
PREMIO <b>SCELTA</b>		X
<b>SCELTA</b>	PREMIO	X
<b>SCELTA</b>	X	PREMIO

Nel primo caso se il concorrente cambia porta perde, ma negli altri due casi se la cambia vince. Dunque vince in 2 casi su 3, e gli conviene di più cambiare.

Vale la pena notare, anche se ci può essere resistenza nel realizzarlo, che l'opportunità di cambiare è uguale all'opportunità di aprire entrambe le porte 2 e 3, mentre non cambiando ho solo l'opportunità di aprire la porta che ho scelto (dato che è stata tolta una porta perdente, aprendo la porta che non ho scelto apro quella che il presentatore mi ha lasciato: se lui ha aperto la 3 io apro la 2, se lui ha aperto la 2 io apro la 3, dunque è come se avessi la possibilità di aprirle entrambe, il che mi dà 2 possibilità su 3 di azzeccare quella col premio). Il fatto è che è il presentatore a decidere quale porta aprirmi, e lui ha una conoscenza perfetta, quindi ha probabilità 1 di aprirne una vuota.

→ Se siamo due concorrenti invece? Supponiamo che io ho scelto la 1. Il mio avversario sceglie la 2; lui ha  $2/3$  di probabilità di trovare una porta vuota e  $1/3$  di trovarne una col premio. Supponiamo che la sua scelta si sia rivelata sbagliata.

Ora a me rimangono la porta 1 e la porta 3. Cambio? In questo caso è indifferente, la probabilità di azzeccarla è diventata  $1/2$ .

Il “paradosso” sta qui ed è solo apparente ovviamente.

→ Il modo più semplice e rigoroso di dimostrare perché è più conveniente cambiare è secondo me quello che passa per la probabilità condizionata (e mi sorprende che Q.C. non l'abbia utilizzato, preferendo perdere un sacco di tempo borbottando; la filosofia di un insegnante dovrebbe essere “se la gente non capisce una cosa spiegata in un modo, spiegagliela in un altro”, non “se la gente non capisce una cosa spiegata in un modo, continua a ripetergliela” - ma io sono solo un umile studente con il dono della genialità).

La probabilità condizionata è definita come la probabilità che accada un evento A, a condizione che prima ne sia avvenuto un altro B. Vedremo più avanti una spiegazione rigorosa, ma per ora notate che questa probabilità è esprimibile come il rapporto tra la probabilità che accadano entrambi gli eventi A e B insieme e la probabilità che accada il solo evento B, cioè, in matematiche:  $P(A|B) = \frac{P(A \cap B)}{P(B)}$ .

Nel primo caso si può interpretare tale probabilità in tale modo: la probabilità di vincere cambiando l'iniziale porta 1 sapendo che il presentatore ha aperto la porta 3 è data dal rapporto tra la probabilità "dietro la porta 2 c'è il premio e il presentatore apre la 3" e la probabilità che "il presentatore apre la porta 3".

Queste probabilità possono essere determinate con una semplice tabella di possibilità, ricordando che nel primo caso una volta che il concorrente ha scelto una porta, il presentatore sceglie tra le vuote; la probabilità totale che avvengano entrambi gli eventi è il prodotto delle probabilità dei singoli eventi (vedi più avanti per chiarimenti se già non sai come funzionano le probabilità).

Car location:	Host opens:	Total probability:	Stay:	Switch:
Door 1	Door 2	1/6	Car	Goat
	Door 3	1/6	Car	Goat
Door 2	Door 3	1/3	Goat	Car
Door 3	Door 2	1/3	Goat	Car

In questa tabella che ho trovato il premio è rappresentato da una macchina, e la probabilità che sia dietro una porta qualsiasi è 1/3. Ricordiamo che il presentatore è onniscente e che il concorrente ha scelto la porta 1. Dunque se il premio si trova nella porta 1 il presentatore sa che può aprire la 2 o la 3: avrà quindi il 50% di aprire una delle due (1/2).

La probabilità che il premio sia dietro la 1 e che il presentatore apra la porta 2 è quindi  $\frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$ . Stessa cosa per la porta 3.

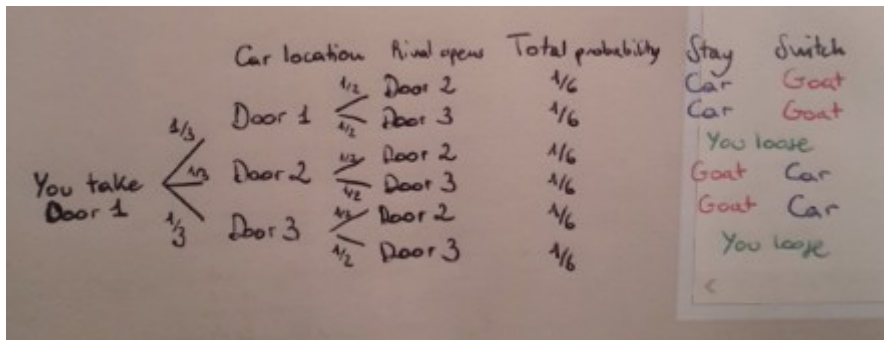
Se il premio è invece dietro la 2, il presentatore sa che potrà aprire solo la 3, e dunque avrà probabilità del 100% (1) di aprirla. La probabilità che il premio sia dietro la 2 e il presentatore apra la 3 sarà quindi  $\frac{1}{3} \cdot 1 = \frac{1}{3}$ .

Dunque qual è la probabilità che il premio sia dietro la 2 e il presentatore apra la 3? Lo abbiamo appena visto, esattamente 1/3. E qual è la probabilità che il presentatore apra la 3? Osserviamo i casi in cui il presentatore apre la 3: nella terza colonna sono la seconda e la terza riga, di cui dovrò sommare le probabilità per ottenere la probabilità totale che il

presentatore apra la 3:  $\frac{\frac{1}{6} + \frac{1}{3}}{\frac{1}{6} + \frac{1}{3}} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$ ; cambiando, abbiamo 2/3 di probabilità di vincere.

Potete anche molto semplicemente sommare le probabilità di vincere non cambiando, cioè  $\frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$ , e quelle di vincere cambiando, cioè  $\frac{1}{3} + \frac{1}{3} = \frac{2}{3}$ .

Consideriamo ora il caso in cui ad aprire un'altra porta sia l'altro concorrente. In questo caso modificheremo il ventaglio delle possibilità ottenendo il seguente:



Io scelgo la 1. La probabilità che il premio sia dietro ognuna delle tre porte è ovviamente  $1/3$ . Questa volta però il mio rivale può scegliere sempre la 2 o la 3, e quindi in alcuni casi vincerà, in altri perderà. Se le volte in cui sappiamo che ha perso ci verrà chiesto di cambiare la scelta sarà assolutamente indifferente. Infatti, come prima  $P(\text{premio nella 2} | \text{rivale apre 3})$  sarà uguale alla probabilità che “il premio sia nella due e il rivale apra la 3”, cioè  $1/6$ , fratto la probabilità che il rivale apra la 3, cioè  $1/3$ . Quindi se cambio vinco in  $1/2$  dei casi, esattamente gli stessi in cui vinco se non cambio: infatti la  $P(\text{premio nella 1} | \text{rivale apre 3})$  è ancora  $1/2$ , così come la  $P(\text{premio nella 3} | \text{rivale apre 2})$ .

In alternativa, posso sommare le probabilità; in  $1/3$  dei casi il rivale perde, in  $1/3$  cambio e vinco, in  $1/3$  non cambio e vinco, a dimostrazione che è del tutto indifferente cambiare o meno. In questo caso l'apparente paradosso nasce dal fatto che essendo il presentatore onniscente (e il rivale no) la sua azione è informativa (so che lui aveva alcune scelte “bloccate”, quindi non tutte le sue scelte erano equiprobabili!), mentre l'azione del rivale non lo è (tutte le sue scelte erano equiprobabili).

Ho provato a darne molte dimostrazioni (ben 4). Nel caso in cui non siate tutt'ora convinti, ce ne sono molte altre, e potete cercarle sulla pagina che vi ho linkato.

#### PROBLEMA 4.

“Incontriamo Antonio a passeggio con un figlio maschio. Sapendo che ha due figli, cambia l'informazione sul genere del secondo sapendo che quello che lo accompagna è il primogenito?”

→ Partiamo dal presupposto che potrebbe non essere a tutti chiaro cosa significa “cambia l'informazione”. Per informazione sul genere intendiamo appunto tutto ciò che sappiamo sul genere del secondo: poiché il genere del secondo può essere sia maschio che femmina, l'informazione si riduce ad una probabilità – mentre prima di sapere che il maschio che lo accompagna è il primogenito il secondo avrà una certa probabilità di essere maschio, dopo aver saputo che il primogenito è un maschio questa probabilità cambia?

È chiaro che in questo caso si tratta di applicare la probabilità condizionata: infatti ci potremmo chiedere “qual è la probabilità che il secondo sia un maschio, a condizione che il primogenito sia un maschio?” e “qual è la probabilità che un figlio sia un maschio, a condizione che l'altro figlio sia un maschio?”. Confrontando le due probabilità vediamo se l'informazione sul genere del secondo cambia nei due casi.

Come al solito, iniziamo a scrivere il nostro universo degli eventi. Una famiglia di due figli con almeno un maschio può avere le seguenti alternative equiprobabili:

	Primogenito	Secondogenito	Probabilità
Caso 1	M	M	0,33
Caso 2	M	F	0,33
Caso 3	F	M	0,33

La probabilità che un figlio sia un maschio a condizione che l'altro figlio sia un maschio sarà per come abbiamo visto primo il rapporto tra la “probabilità che un figlio sia maschio E che l'altro figlio sia maschio” (cioè la probabilità che ci siano due maschi, ovvero  $1/3$ ) e la probabilità che “l'altro figlio sia maschio” (cioè che ci sia almeno un maschio, ovvero

$1/3+1/3+1/3=1$ ). Dunque la probabilità che un figlio sia un maschio a condizione che l'altro figlio sia un maschio è esattamente  $1/3$  (e  $2/3$  è la probabilità che sia femmina ovviamente). La probabilità che un figlio sia un maschio a condizione che il PRIMOGENITO sia un maschio sarà diversa? Applichiamo la probabilità condizionata: è il rapporto tra “probabilità che un figlio sia maschio E che il primogenito sia maschio” (che corrisponde ancora alla probabilità che siano entrambi maschi, cioè  $1/3$ ) e la “probabilità che il primogenito sia un maschio” (cosa che esclude il caso 3: dunque questa probabilità è  $1/3+1/3=2/3$ ). Dunque la probabilità che un figlio sia un maschio a condizione che il primogenito sia un maschio è  $\frac{1/3}{2/3} = \frac{1}{2}$  → l'informazione sul genere del secondo è cambiata.

Notate che un'informazione può cambiare la probabilità di un evento, come avevamo già osservato nel paradosso di Monty Hall.

#### PROBLEMA 5: [Paradosso di Bertrand](#)

“Qual è la probabilità che una corda di un cerchio scelta a caso sia più lunga del lato di un triangolo equilatero inscritto?”

→ Partiamo dal presupposto che didatticamente parlando non ha alcun senso presentarvi questo problema. Voi non sapete nulla di teoria della probabilità, e non si può pretendere certo che capiate intimamente concetti del genere, soprattutto considerando che sono ancora dibattuti e non si è arrivati ad una soluzione del paradosso.

Quello che cercherò di fare è illustrarvi come mai si tratta di un paradosso e quali sono le idee più diffuse in merito, ma pensare di “stimolarvi” in questo modo è da idioti.

[Ho una mia idea sull'insegnamento, e penso che debba procedere per gradi contigui – deduttivamente o induttivamente, ma sempre per gradi. Al Duca evidentemente piace utilizzare il metodo “induttivamente a capocchia di minchia”, continuando a presentarvi cose che non potete sapere e a cui dovrete arrivare senza avere le basi per poter ragionare. Quello che lui fa, e che mi dà davvero fastidio, è analogo al caso in cui un medico chiedesse ad uno studente di medicina a metà del primo anno di ricavare qualcosa da un esame obiettivo: lo studente conosce solo istologia, chimica e fisica, sa poco di anatomia, nulla di fisiologia, patologia, semeiotica – lo studente di medicina ha tutto il diritto di non sentirsi per nulla stimolato e di mandare più o meno manifestamente a fanculo il medico, che magari gli risponderà pure “eh beh beh beh beh, è ovvio che sia così” snocciolandogli nozioni che lo studente NON può conoscere, dato che il medico non gliel'ha spiegate.

Quindi le alternative sono due: o spieghi prima la teoria e poi la fai applicare (metodo deduttivo) oppure non spieghi la teoria e partendo da un problema semplice chiedi agli studenti di ricavare una legge o la teoria stessa (metodo induttivo - “pericoloso” perché ovviamente non si può saltare troppo, altrimenti facendo vedere un fascio di luce e spiegata la gravità tutti potremmo tecnicamente arrivare alla teoria della relatività generale senza sapere una beata mazza di fisica). Shame on you Duke.]

Chiusa questa parentesi/sfogo, passiamo al problema.

Joseph Bertrand ha posto il problema trovando tre differenti soluzioni, che portavano tuttavia a risultati inconsistenti<sup>5</sup>:

1. Metodo degli estremi casuali: la corda è scelta casualmente a partire dai suoi due estremi.
  - Scelgo un punto A sulla circonferenza; faccio ruotare il triangolo finché un

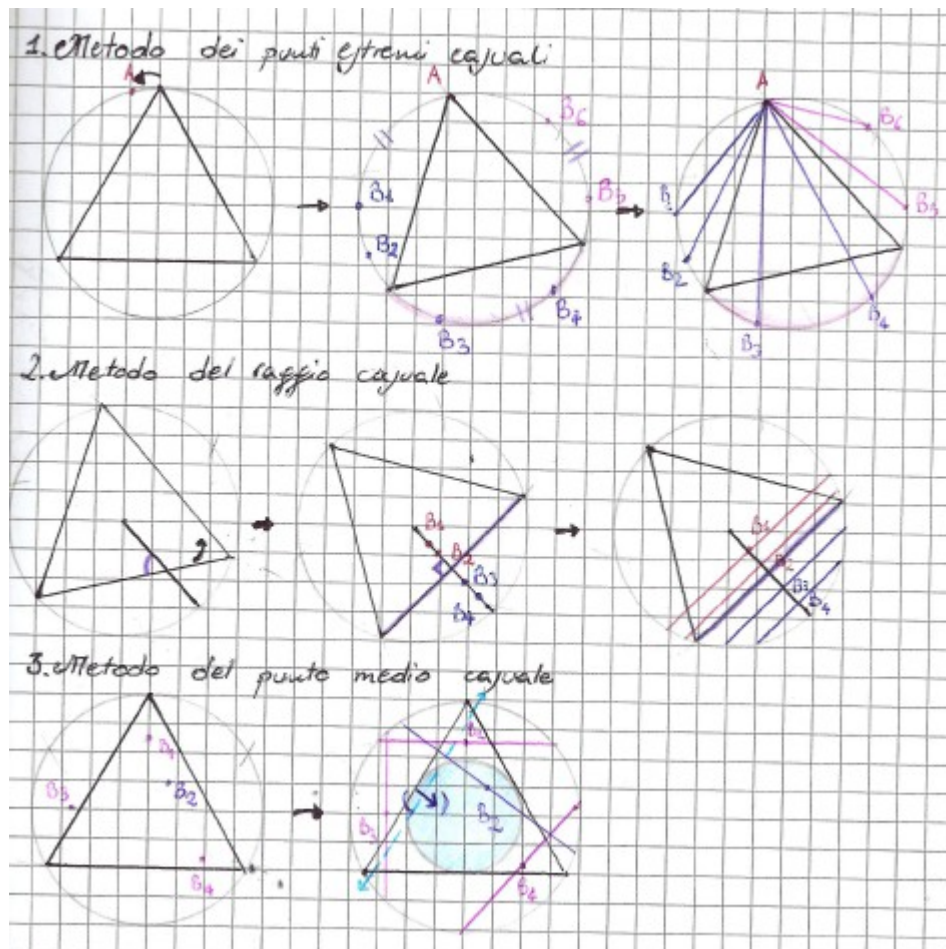
---

5 In medicina troverete spesso il concetto di “consistenza”, soprattutto quando vengono mostrati e confrontati risultati degli studi. Quello della consistenza è un concetto che potete definire molto semplicemente così: se dato uno stesso problema due risultati ottenuti sono uguali o più genericamente in accordo, questi sono consistenti; se discordano non lo sono. Trovare che il fumo è protettivo per il cancro al polmone è inconsistente con gli altri risultati, ma nel caso in cui il risultato trovato fosse: “il fumo è un fattore di rischio” per valutare la consistenza bisognerebbe comparare l'entità del rischio trovato rispetto alle entità dei rischi trovati negli altri studi. Fortunatamente, in questo caso – nel paradosso di Bertrand – due risultati sono consistenti quando hanno lo stesso valore e inconsistenti quando hanno valori diversi.

- vertice non coincide con A.
- Scelgo quindi l'altro estremo della corda sulla circonferenza in un punto qualsiasi.
  - I vertici del mio triangolo equilatero dividono la circonferenza in 3 segmenti circolari uguali: dal disegno vediamo che se il secondo estremo cade sul segmento rosa (B3 e B4) la corda risulta più lunga del lato del triangolo, mentre se il secondo estremo cade sugli altri due segmenti sarà più corta.
  - Dunque la probabilità in questo caso è esattamente  $1/3$ .
2. Metodo del raggio casuale.
- Scelgo un raggio qualsiasi; faccio ruotare il triangolo finché un lato non è ortogonale a tale raggio.
  - Scelgo dei punti casualmente su tale raggio, e traccio le corde che hanno come punti medi questi punti. Tali corde risulteranno perpendicolari al raggio stesso, e parallele al lato viola.
  - Se la corda si trova tra il lato e il centro della circonferenza risulterà più lunga del lato del triangolo, altrimenti minore.
  - Dunque la probabilità in questo caso è  $1/2$ .
3. Metodo del punto medio casuale.
- Scelgo un punto a caso nel cerchio; traccio la corda che ha come punto medio tale punto (c'è una corrispondenza biunivoca tra punto medio e corda, pertanto ogni punto nel cerchio identifica una e una sola corda).
  - Traccio il cerchio inscritto nel triangolo equilatero: tale cerchio avrà raggio pari alla metà del cerchio circoscritto al triangolo, pertanto la sua area (che va come il quadrato del raggio) sarà  $1/4$ .
  - La corda risulta più lunga del lato se interseca il cerchio inscritto. È molto semplice da capire: immaginate di muovere il lato sinistro verso destra, tenendo fermo il cerchio piccolo, lungo la direzione della freccia blu tra parentesi: in questo modo il lato interseca il cerchio, anche per un tratto infinitesimo. Ma potete vedere subito che il segmento diventerà più lungo per poter continuare ad intersecare la circonferenza grande (cioè per continuare ad essere una corda; vedi il tratto in azzurro tratteggiato – nota che il lato del triangolo è esso stesso una corda).
  - Poiché l'area del cerchio piccolo è  $1/4$  di quella del cerchio grande, conterrà un quarto dei punti contenuti nel cerchio grande, dunque un quarto delle corde che hanno tali punti come punti medi.
  - Dunque, la probabilità è in questo caso  $1/4$ .
  - ... un par di ciuffoli: il centro della circonferenza è il punto medio di infiniti diametri (e un diametro è una corda), quindi infiniti diametri passano per il cerchio inscritto. Non c'è corrispondenza biunivoca tra i punti medi dei diametri (che sono sempre uguali al centro della circonferenza) e i diametri stessi. Dunque, qual è la reale probabilità?  
 → si può fare un discorso più generale affermando che l'insieme dei diametri è un infinito non numerabile; un insieme infinito si dice numerabile quando i suoi elementi possono essere messi in corrispondenza biunivoca con i numeri naturali (un esempio è l'insieme dei numeri naturali, o quello dei numeri razionali). L'insieme dei numeri reali non è numerabile (ciò è stato dimostrato da Cantor con l'argomento diagonale – ma meglio non aprire questo discorso che è parecchio incasinato).  
 L'insieme dei diametri non è numerabile, e ciò rende il nostro “universo degli eventi” su cui definire la nostra probabilità uno spazio non

numerabile. Se a questo punto ho perso completamente la vostra attenzione ho ottenuto l'obiettivo: farvi capire che chi pretende di spiegarvi questo concetto senza che voi abbiate le nozioni necessarie (che riguardano per esempio algebra di Borel, misura di Lebesgue, invarianza e in generale teoria della probabilità) didatticamente non capisce nulla. Se l'obiettivo è fare show-off ok, ma se l'obiettivo è farvi capire qualcosa non credo che questo sia il metodo giusto.

Di seguito trovate una schematizzazione dei tre procedimenti:



Il problema è che la domanda non specifica in che modo bisogna scegliere la corda CASUALE. In generale, ci sarà un numero indefinito di possibilità in cui sceglierla, ciascuna che porta a diversi valori di probabilità.

Il problema, in assenza di specificazione, può essere risolto con due approcci:

- Distinguere tra problemi non equivalenti (come ha fatto Marinoff)
- Considerare il problema come ben posto (come ha fatto [Jaynes](#))
  - un problema è ben posto quando:
    - ha soluzione
    - la soluzione è unica
    - il comportamento della soluzione varia in modo continuo al variare delle condizioni iniziali

Il primo caso è ciò che abbiamo fatto noi: trovare diverse soluzioni specificando il modo in cui abbiamo scelto di volta in volta come prendere la corda casualmente. In sostanza con questo approccio il paradosso rimane un paradosso, ovvero il problema è mal posto.

Nel secondo caso invece bisogna agire applicando il principio di “massima ignoranza”, ovvero

evitare di aggiungere informazioni che non sono date nella domanda. In accordo con questo principio, Jaynes nel suo articolo del 1973 dice che la domanda non specifica né la posizione del cerchio, né la dimensione del cerchio. Considerando il problema come ben posto e applicando il principio di massima ignoranza, dovremmo quindi trovare una soluzione unica indipendente da posizione e dimensione del cerchio. Una soluzione che non varia in queste condizioni si dice “invariante per traslazione” e “invariante di scala” (nel suo articolo Jaynes dimostra anche che la soluzione è invariante per rotazione). Per ricavare tale soluzione Jaynes utilizza le equazioni integrali che descrivono l'invarianza in modo da determinare la distribuzione di probabilità (for dummies who want to stay dummies: Jaynes ha cercato una soluzione che andasse bene sempre e l'ha trovata), scoprendo che la soluzione unica coincide con il metodo del raggio casuale.

Ma quale dei due approcci usare? Un articolo uscito giusto l'anno scorso ([Aerts, Sassoli de Bianchi](#)) ha proposto che fatte alcune specifiche particolari il risultato giusto sia quello di Jaynes.

Occorre notare che anche Poincaré (un matematico parecchio importante per usare un eufemismo) era d'accordo con Jaynes ancora prima che Jaynes nascesse, e l'esame di casi di questo tipo ha contribuito allo sviluppo della geometria integrale, che è utilizzata in stereologia, l'interpretazione tridimensionale delle immagini piate. La stereologia ha importanti applicazioni pratiche in medicina (per esempio è servita a determinare che l'area di scambio dei polmoni è circa 75 metri quadrati, ma può anche servire a contare le cellule presenti in un tessuto umano, come i neuroni nel cervelletto). Se a questo punto non vi viene da dire “ma che figata”, e provate solo un senso più o meno acuto di fastidio, allora è proprio certo che non vi potrò mai far piacere la matematica, in qualunque sua declinazione. Sigh.

#### 5.4 – Definire e calcolare la probabilità

La probabilità può apparire un concetto semplice, dato che abbiamo a che fare con essa spesso. Sfortunatamente, non è così. Anzi, è un concetto piuttosto problematico, tale da richiedere tutta una propria teoria. Qui di seguito potremo solo dare qualche nozione base.

Iniziamo col definire cos'è la probabilità:

- La probabilità può essere intesa come il rapporto fra il numero di esiti di un esperimento stocastico in cui l'evento si realizza e il numero di esiti totali, equiprobabili.
  - Ci sono due concetti che ancora non conosciamo bene:
    - Stocastico: la definizione di [processo stocastico](#) è piuttosto complessa da spiegare qui. Prendetela come “esperimento regolato dal caso”.
    - Equiprobabilità: se siete stati attenti avrete notato che nella definizione di probabilità abbiamo usato il concetto stesso della probabilità, attraverso cui possiamo dire che gli esiti totali sono equiprobabili: non bisognerebbe mai usare il definendo in una definizione. Come possiamo definire tuttavia il concetto di equiprobabilità senza ricorrere ancora una volta al concetto di probabilità? Può essere utile utilizzare il [principio di indifferenza](#) di Laplace, che stabilisce che per dire che dei risultati sono equiprobabili bisogna fare in modo che non ci sia una ragione sufficiente perché un risultato sia più frequente di un altro.
      - Il principio stabilisce che se N possibilità non sono distinguibili se non per i loro nomi, allora ad ogni possibilità deve essere assegnata una probabilità di  $1/N$ .
- Alternativamente, possiamo interpretare la probabilità come la “propensione” di un evento a realizzarsi, ovvero come il numero medio di successi per prove.
  - Possiamo introdurre il concetto di “puntata equa” sul realizzarsi di un evento rispetto ad una vincita unitaria, che è un numero reale compreso fra 0 ed infinito, e rappresenta il numero medio di successi per insuccesso. Viene detto ODDS ed è il rapporto tra P e  $1-P$ , con P probabilità del realizzarsi dell'evento.
    - Se avete difficoltà a capire il concetto di puntata equa: immaginate che

vogliate puntare in modo equo sul fatto che lanciando un dado una volta non escano 5 o 6. Ci sono 6 eventi possibili lanciando un dado, e il vostro successo si verificherà in 4 di essi, mentre perderete in due casi.

Il modo più equo di puntare è scommettere 4 euro contro 2. Ciò significa che se vincete, vincerete 2 euro (6 in totale, ma ne avete puntati 4!), e se perderete ne perderete 4 (in questo senso 4 contro 2 – scommettete la perdita contro la possibile vincita). Questo è equo, infatti la probabilità di perdere è  $2/4$  della probabilità di vincere. Ciò significa che lanciando per esempio 3 volte questo dado perdereste una volta i vostri 4 euro, e vincereste in 2 volte i 2 euro: sommando vincite e perdite  $2 \cdot 2 - 1 \cdot 4 = 0$ , cioè non avete né perso né vinto. Dovete considerare che infatti chi avete di fronte e scommette contro di voi scommette in modo complementare: vi darà 2 euro se vincete, ma si prenderà i vostri 4 euro se perdete, e poiché ha un terzo delle probabilità di prendersi i vostri 4 euro ciò sembra giusto.

Quando si scommette lo si fa sul realizzarsi di un evento in UN SOLO tentativo (per esempio una partita di calcio o una corsa di cavalli). Il concetto di “puntata equa” ci permette di trovare quella puntata tale per cui se facessimo infiniti tentativi nessuno dei due scommettitori rivali perderebbe o vincerebbe in totale (cioè non esclude che possa farlo in un tentativo solo!).

Se non siete ancora convinti, fate un esempio diverso: se puntate 3 volte 3 a 2 con l'evento di prima avreste una vincita totale di  $2 \cdot 2 - 1 \cdot 3 = 1$ . Perché chi scommette contro di voi dovrebbe accettare di pagarvi 1 in media ogni 3 volte? Non sarebbe equo, dunque non vorrebbe farlo.

La definizione tuttavia diceva “contro una vincita unitaria” (questo serve semplicemente a confrontare i vari odds; confrontare 4 a 2 e 7 a 5 può non essere immediato). Basta fare la divisione: 4 diviso 2 è uguale a 2 diviso 1, e infatti possiamo anche puntare 2 contro 1 (la puntata, essendo un rapporto, rimane equa allo stesso modo dato che il rapporto non cambia).

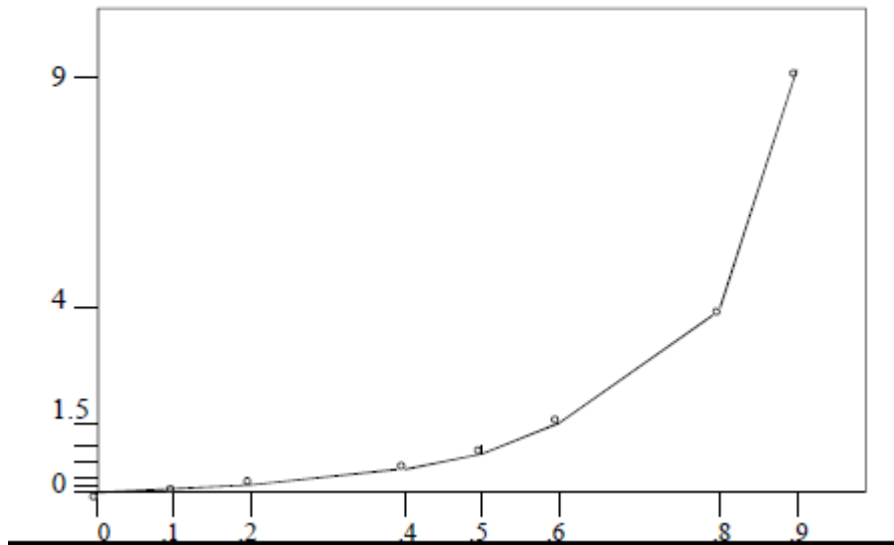
Dunque il nostro odds vale 2; infatti  $P=4/6$ , e  $ODDS = \frac{4/6}{1-4/6} = \frac{4/6}{2/6} = 2$

(che leggiamo “odds di 2” o “odds di 2 a 1” indifferentemente).

- La probabilità è esprimibile, riarrangiando la formula dell'odds, come

$$P = \frac{ODDS}{ODDS + 1} .$$

## ODDS vs P



*Odds sull'asse delle y, probabilità su quello delle x*

- Un altro approccio per interpretare la probabilità è quello soggettivo di [Bayes](#), in cui la probabilità esprime il grado di convincimento (o di fiducia) nella realizzazione di un evento.
  - In questo senso, per un esperimento non irripetibile e con esiti non equiprobabili, la probabilità viene espressa sempre attraverso l'odds. Per fare un esempio pratico, se io dico che scommetto 3 a 1 sulla Juventus che vince il campionato, significa per quanto visto sopra che sono disposto a perdere 3 per vincere solo 1; pertanto assegno all'evento una probabilità di  $3/(3+4)=0.75$  [sono passato dall'odds alla probabilità]. La Juventus non ha REALMENTE una probabilità del 75% di vincere il campionato; tuttavia, non ha neanche la probabilità del 5% di vincerlo (cosa che potrebbe accadrebbe se tutte le squadre avessero la stessa probabilità di vincere –  $1/20$  – ma per esempio il Parma fa più schifo del vomito fecaloide). Poiché le condizioni che determinano la reale probabilità sono troppe e a loro volta soggette a probabilità (come è composta la squadra, decisioni arbitrali, infortuni, organizzazione del calendario, il classico crollo della Roma, la pezzenza delle milanesi, il fatto che al fantacalcio non ho preso manco uno della Juve – cosa che farò l'anno prossimo condannandola alla retrocessione, ecc ecc) non abbiamo un modello che ci permette di calcolarla; dunque in questo caso la probabilità esprime più che altro il grado di fiducia personale nell'evento. Un'altra persona potrebbe infatti assegnare alla vittoria della Juventus una probabilità maggiore o minore.
  - La probabilità diventa così un'entità gnoseologica piuttosto che un'entità ontologica: diventa un qualcosa che esprime un modo di conoscere piuttosto che un modo di essere.

I tre approcci vengono detti rispettivamente classico (o matematico), frequentista e soggettivo, a cui si aggiunge l'approccio assiomatico di Kolmogorov, e conducono a quattro modi diversi di definire e calcolare la probabilità:

- Classico:  $P(E)=n/N$ , con  $N$  risultati equiprobabili dell'evento e  $n$  risultati in un cui si realizza  $E$ .
- Frequentista:  $P(E)=f/N$  per  $N$  che tende all'infinito; la probabilità è la frequenza relativa con cui si realizza l'evento  $E$  al tendere all'infinito del numero di occasioni, assumendo che ad ogni occasione la “propensione” a realizzarsi dell'evento sia costante.
- Soggettivo:  $P(E)=ODDS/(ODDS+1)$ ; la probabilità è espressa dal grado personale di

fiducia nel realizzarsi dell'evento E.

- Assiomatico: la probabilità è definita a partire dai seguenti assiomi<sup>6</sup>
  - La probabilità di un evento è un numero compreso tra 0 e 1:  $0 \leq p \leq 1$  .
  - La probabilità dell'evento certo è 1:  $p(\Omega) = 1$  .
    - $\Omega$  è un simbolo che convenzionalmente esprime l'universo degli eventi: si può leggere l'uguaglianza di prima come “la probabilità che si realizzi un evento presente in  $\Omega$  è 1”.
  - Per due eventi mutuamente esclusivi (cioè che non possono essere contemporaneamente veri) la probabilità che avvenga l'unione dei due eventi è uguale alla somma delle probabilità dei singoli eventi:  $P(A \cup B) = P(A) + P(B)$  .
    - Facciamo un esempio: lanciate un dado, e volete sapere la probabilità che esca un numero minore di 3 o un numero maggiore di 4. La probabilità del primo evento sarà  $2/6$ , così anche la probabilità del secondo evento. Questi eventi non possono accadere contemporaneamente: infatti se esce un numero minore di 3 non può uscire un numero maggiore di 4, e pertanto sono eventi mutualmente esclusivi [state attenti: evento non significa singolo risultato – l'evento è più in generale un qualcosa che può accadere nell'accezione più generale del termine, cioè un INSIEME di risultati! Quando parlate, voi potete dire “domani piove”, e il piovere è un evento, oppure “domani può piovere o grandinare” e anche “piovere o grandinare” è un evento, precisamente l'unione degli eventi “piovere” e “grandinare”].  
Pertanto la probabilità dell'evento “esce un numero minore di 3 o maggiore di 4” è la somma delle probabilità dei due eventi che compongono l'unione, ovvero  $\frac{2}{6} + \frac{2}{6} = \frac{2}{3}$  .

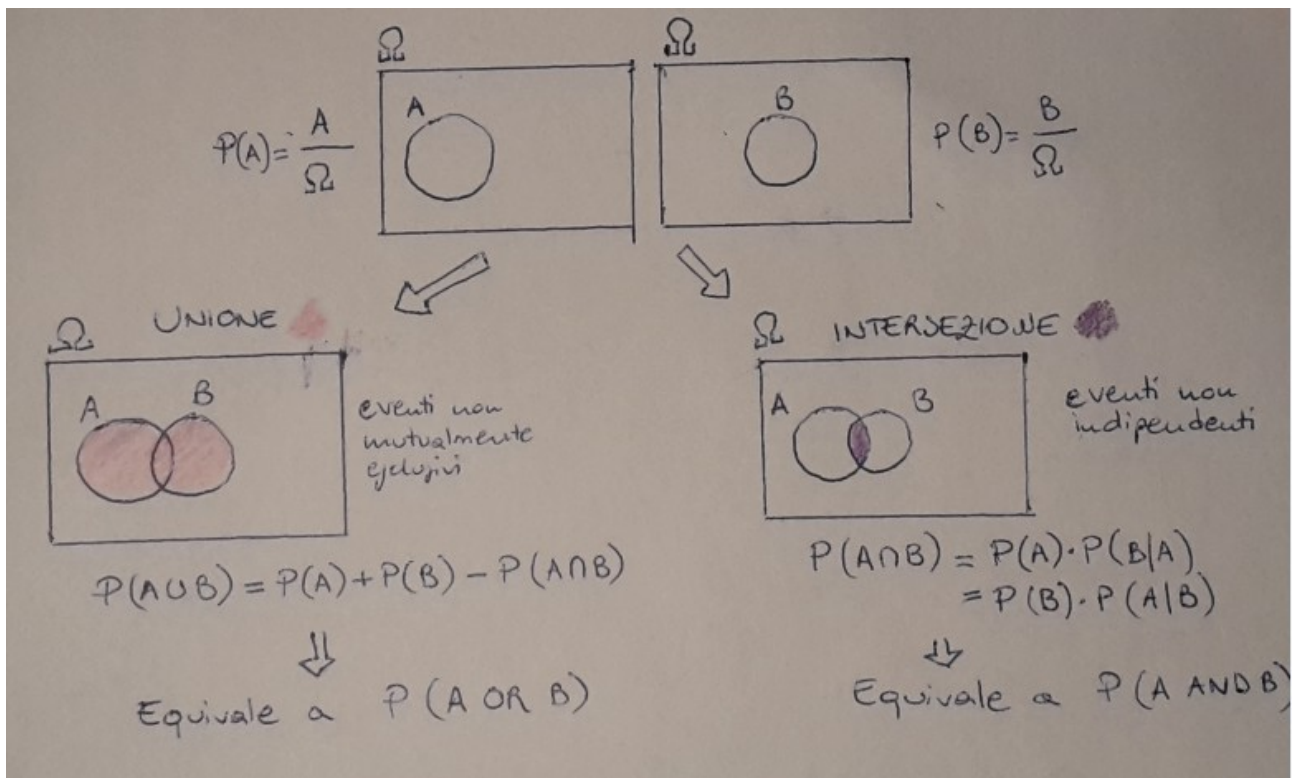
Passiamo ora ad alcune applicazioni del calcolo delle probabilità, per cui ci servirà una minima infarinatura teorica.

Se abbiamo due insiemi di eventi A e B in un universo degli eventi  $\Omega$  le loro singole probabilità possono essere definite rispettivamente come  $A/\Omega$  e  $B/\Omega$  (d'ora in avanti le chiameremo  $P(A)$  e  $P(B)$ ).

Su questi insiemi possiamo definire unione e intersezione come segue:

---

<sup>6</sup> In realtà no, gli assiomi sono di più e sono leggermente differenti, ma siccome non ve ne frega nulla va bene così. Se vi interessa guardate [qui](#).

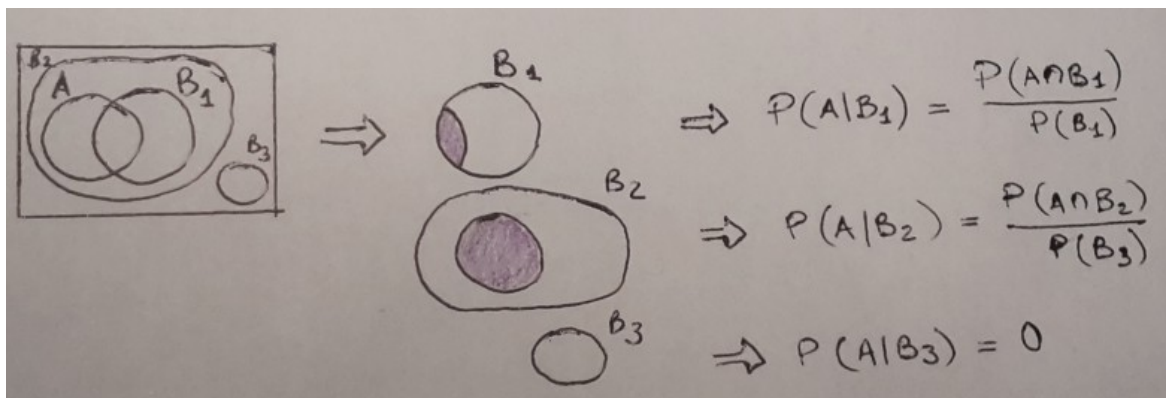


Mentre è abbastanza evidente capire perché la probabilità dell'evento “unione” di eventi debba essere la somma delle probabilità dei singoli eventi meno la probabilità dell'intersezione dei due eventi (infatti, non vogliamo contare 2 volte gli elementi che sono sia in A che in B!), può apparire più complicato capire la formula della probabilità dell'evento “intersezione”.

Tale formula presuppone il concetto di probabilità condizionata  $P(A|B)$ , cioè la probabilità che A si verifichi dato che si è verificato B, definita da Kolmogorov come  $P(A|B) = \frac{P(A \cap B)}{P(B)}$ .

→ Notate che questa è una definizione, non un risultato. Alcuni autori preferiscono considerarlo come un assioma della probabilità (ricordate che gli assiomi sono affermazioni indimostrate/indimostrabili da cui partiamo per elaborare un sistema logico o matematico → ciò significa: parto da certe premesse che non dimostro, perché non ho elementi per farlo – gli assiomi – per giungere a certe conclusioni – il sistema – che saranno giustificate esclusivamente dalla scelta degli assiomi iniziali). La probabilità condizionata è utilizzata anche nel [teorema di Bayes](#), che non vi sto qui a spiegare perché esula dagli scopi di questa introduzione.

Ciò che ora ci preme è capire cosa voglia dire la definizione di probabilità condizionata, posto che non possiamo dimostrarla (è una definizione... in realtà trovate una derivazione formale in [questa pagina](#), ma anche se non la imparate non succede nulla). Significa fondamentalmente calcolare la probabilità di A dopo aver ristretto l'universo degli eventi a B (e a questo punto è chiaro che A sarà ciò che rimane di A dopo aver ristretto l'universo a B, ovvero l'intersezione di A con B):



→ Facciamo un esempio pratico con dei dadi: avete il vostro universo degli eventi che è dato dai possibili risultati del lancio di due dadi. Volete vedere qual è la probabilità che lanciando il dado A ottenga 2 sapendo che la somma dei dadi A e B deve essere minore o uguale a 5: posso scrivere il tutto come  $P(A=2|A+B \leq 5)$ . Posso calcolare questa probabilità con la formula: la probabilità dell'intersezione dei due eventi è 3/36 (calcolata nello spazio degli eventi "Table 1"!!), mentre la probabilità dell'evento  $A+B \leq 5$  nell'universo degli eventi è 10/36. Facendo il

$$\text{calcolo } P(A=2|A+B \leq 5) = \frac{P(A=2 \cap A+B \leq 5)}{P(A+B \leq 5)} = \frac{\frac{3}{36}}{\frac{10}{36}} = \frac{3}{10}$$

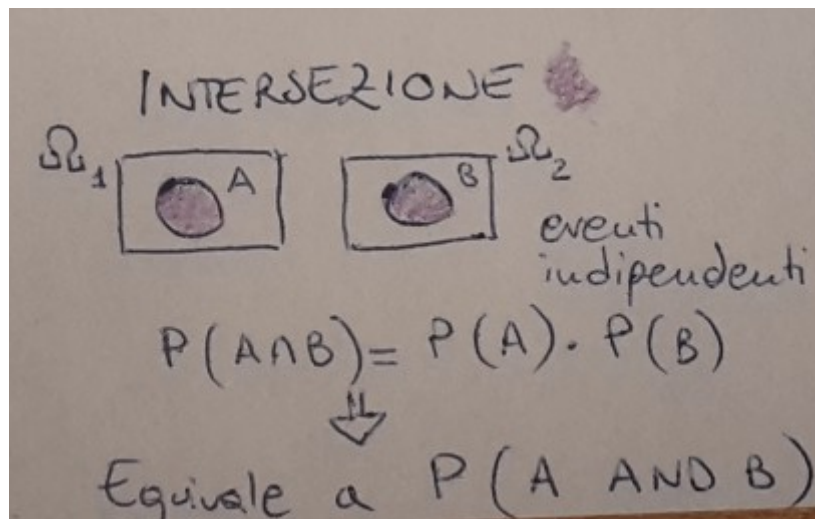
Table 1

+	B=1	2	3	4	5	6
A=1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Ciò è equivalente a restringere l'universo degli eventi "Table 1" all'insieme  $A+B \leq 5$ , che consta di 10 elementi, e calcolare la probabilità di  $A=2$  su questo insieme: poiché avviene in 3 casi, la probabilità è 3/10.

A questo punto, calcolare l'intersezione di un generico evento A e di un generico evento B si riduce a fare  $P(A \cap B) = P(A|B)P(B)$  (o equivalentemente, per il teorema di Bayes,  $P(A \cap B) = P(B|A)P(A)$ ).

→ Ci potremmo chiedere cosa succede per eventi indipendenti: per esempio, qual è la probabilità di ottenere 2 in un dado se con l'altro dado ho ottenuto 5? Ovviamente il primo dado non influenza il secondo; se calcolassi l'intersezione dei due risultati con la regola sopra esposta, come potrei esprimere la probabilità condizionata? Sarebbe esattamente  $P(A=2|B=5) = P(A=2)$ .



Infatti, due eventi sono definiti indipendenti se la loro probabilità congiunta (cioè  $P(A \cap B)$ ) è uguale al prodotto delle loro probabilità. Infatti in tale caso  $P(A \cap B) = P(A)P(B)$ , da cui  $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$ , ma  $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B)$ ; se la probabilità di un evento A è uguale alla probabilità di quell'evento dato un altro evento B (e viceversa), è ovvio che l'evento B non influenza la probabilità che avvenga l'evento A (e viceversa). Quando due eventi sono indipendenti pertanto l'evento intersezione diventa:  $P(A \cap B) = P(A)P(B)$ .

→ potreste essere ragionevolmente confusi: nel penultimo disegno la probabilità condizionata di A dato B3 era nulla, perché era nulla l'intersezione. Ma allora perché nell'ultimo disegno la probabilità condizionata non è nulla, e nemmeno l'intersezione?

Nel penultimo disegno A e B3 erano MUTUAMENTE ESCLUSIVI: cioè erano due eventi disgiunti definiti sullo STESSO universo degli eventi, e il realizzarsi dell'uno ESCLUDE il realizzarsi dell'altro, ragion per cui la loro intersezione era vuota.

Nell'ultimo disegno invece A e B sono INDIPENDENTI: possono essere definiti su universi degli eventi DIVERSI (che possono anche coincidere, ma sono separabili concettualmente; con la somma dei dadi non potevamo farlo – ma questa è una cosa che vi dico io per aiutarvi a capire, non ditela all'esame, perché non è una definizione rigorosa) e l'accadere di uno non influenza affatto il realizzarsi dell'altro.

→ Generalizzando, due eventi si dicono statisticamente indipendenti quando  $P(A \cap B) = P(A)P(B)$ , e mutuamente esclusivi quando  $P(A \cap B) = 0$ .

Notate che:

- Due eventi mutualmente esclusivi sono sempre dipendenti
- Due eventi indipendenti non sono mai mutualmente esclusivi
- Due eventi dipendenti possono essere o non essere mutualmente esclusivi

Riepiloghiamo i risultati ottenuti fin qui per semplicità:

- Probabilità che si verifichi un evento o un'altro:  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ 
  - Se i due eventi sono mutualmente esclusivi:  $P(A \cup B) = P(A) + P(B)$
- Probabilità che si verifichino due eventi:  $P(A \cap B) = P(A|B)P(B) = P(A)P(B|A)$ 
  - Se i due eventi sono indipendenti:  $P(A \cap B) = P(A)P(B)$

Possiamo ora ragionare su due tipi di possibili eventi; l'evento OR e l'evento AND.

- L'evento OR significa: “può accadere A o B”, e per valutarne la probabilità è evidente che dovremmo valutare la probabilità dell'unione di A e B:  $P(A \text{ OR } B) = P(A \cup B)$
- L'evento AND significa “accadono A e B insieme”, e per valutarne la probabilità si valuta la probabilità dell'intersezione di A e B:  $P(A \text{ AND } B) = P(A \cap B)$

Facciamo ora un esempio di problema in ambito medico, che potete facilmente risolvere:

“Sapendo che il 95% dei sani ha la glicemia nella norma, e che il 95% dei sani ha [ALT](#) nella norma, qual è la percentuale dei sani con entrambi i valori nella norma?”

→ Risoluzione: definiamo i due eventi:

- A: Sani con glicemia nella norma →  $P(A) = 0.95$
- B: Sani con ALT nella norma →  $P(B) = 0.95$

Chiediamoci a questo punto cosa dobbiamo trovare: è l'evento AND, che realizza l'evento A e B contemporaneamente.

Dunque, il prossimo passo è capire se i due eventi sono indipendenti o no: poiché la glicemia non influenza il valore dell'ALT, i due eventi sono indipendenti.

Quindi  $P(A \text{ AND } B) = P(A \cap B) = P(A)P(B) = 0,95 \cdot 0,95 = 0,9025$

→ Se ora facessimo un altro test C con SP=95% (cioè il 95% dei sani ha valori nella norma!) indipendente rispetto a glicemia ed ALT, quale sarebbe la probabilità che un soggetto sano abbia valori nella norma per tutti e tre i test? Si tratta ovviamente anche in questo caso di un evento AND tra  $(A \cap B)$  e C; dunque

$$P((A \text{ AND } B) \text{ AND } C) = P((A \cap B) \cap C) = P(A \cap B)P(C) = P(A)P(B)P(C) = 0,95^3 = 0,85$$

→ Pù in generale, potete capire che per una serie di eventi tutti indipendenti tra loro la probabilità che si verifichino tutti decresce, e la legge generale è

$$P(A_1 \text{ AND } A_2 \text{ AND } \dots \text{ AND } A_n) = \prod_{i=1}^n P(A_i) \quad , \text{ dove il simbolo introdotto significa prodotto}$$

di tutti gli elementi.

Se facessi 25 test con SP=95% i sani sarebbero quindi  $0,95^{25} = 0,27$ ; questo significa che circa  $\frac{3}{4}$  dei pazienti sani risulteranno positivi ad almeno un test! Ma pensateci un attimo: quando fate un esame del sangue i valori di riferimento sono indicativi dell'intervallo in cui si distribuisce il 95% dei sani nella popolazione. Pensate a quanti valori indipendenti valutate in un esame del genere: avere un valore sballato non significa pertanto essere automaticamente malato, anche se può far sorgere il sospetto di patologia. Uno può avere un valore sballato ed essere perfettamente sano, in quanto potrebbe far parte di quel 5% dei sani che ha un valore fuori dall'intervallo di riferimento. State attenti quando utilizzate gli esami quindi!

Facciamo un altro esempio, con gli stessi dati del problema precedente:

“Qual è la probabilità che un sano abbia un valore normale in almeno uno dei due test?”

→ Soluzione: in questo caso è evidente che ci viene richiesto l'evento OR, infatti il sano per soddisfare la richiesta deve avere un valore normale in un test O nell'altro.

Dobbiamo dunque chiederci se gli eventi sono mutualmente esclusivi: non lo sono, infatti avere un valore di glicemia nella norma non esclude avere un valore di ALT sballato.

Dunque la probabilità dell'evento OR che mi viene richiesta dal problema corrisponde a:

$$P(A \text{ OR } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,95 + 0,95 - 0,9025 = 0,9975$$

→ Cosa succede per un numero di test non mutualmente esclusivi più alto di 2? Per esempio, se facciamo 25 test indipendenti (dunque non mutualmente esclusivi) qual è la probabilità che un sano abbia almeno uno di essi nella norma?

In questo caso non è così semplice rintracciare la legge (leggete questa parte solo se vi interessa ovviamente, non è importante): ho cercato un po' in internet e la spiegazione più chiara mi sembra essere [questa](#). La formula è:

$$P\left(\bigcup_{i=1}^n E_i\right) = \sum_{k=1}^n (-1)^{k+1} \sum_{\substack{i_1, i_2, \dots, i_k \\ 1 \leq i_1 < i_2 < \dots < i_k \leq n}} P(E_{i_1} \cap E_{i_2} \cap \dots \cap E_{i_k})$$

inclusione ed esclusione applicato alla probabilità.

Una formula del genere è molto probabile che non sappiate neanche come affrontarla: se andate su [questo sito](#) potete ricavare la probabilità dell'unione di tre eventi indipendenti ciascuno con probabilità 0.95, e vedete che è 0.999875.

Se dovessimo trovare la probabilità con 25 test indipendenti applicando la formula che vi ho dato otterremmo un risultato che è praticamente 1 (0.9999999999...). L'applicazione pratica è che l'evento complementare (essere sani e non avere neanche un valore nella norma) ha probabilità praticamente nulla: se ho tutti gli esami sballati è certo praticamente certo che sono malato.

→ Uniamo le due informazioni: qual è la probabilità che io sia sano se su 25 test indipendenti con SP=0,95 ho 3 valori sballati? È una situazione realistica: può capitare che un paziente faccia gli esami del sangue e abbia tre valori sballati – sarà più probabilmente malato o più probabilmente sano? Come vi dovete comportare?

Buon ragionamento!

[P.S.: la soluzione è 0,093. Non ho trovato in internet nulla che sfiori il problema che mi sono e vi ho posto; pertanto, bisogna far da soli.]

[N.B.: ciò che ho scritto dopo le ultime due freccine sono cose che non dovete sapere per nessuna materia e probabilmente per nessuna occasione nella vostra vita, ma è un argomento particolarmente complesso, e siccome ho quasi un'ossessione per le cose incredibilmente complicate volevo rendervene partecipi, sperando che troviate le riflessioni che scaturiscono da ciò che vi ho scritto stimolanti almeno un infinitesimo – e forse è meglio non di più – di quanto lo sono per me.]

## 5.5 – Teorema di Bayes e sue applicazioni

Ricordate quando ho citato il teorema di Bayes? Per ricavarlo si fanno due passaggi molto semplici.

Posto che  $P(A|B) = \frac{P(A \cap B)}{P(B)}$  e che  $P(B|A) = \frac{P(A \cap B)}{P(A)}$ , ricaviamo dalla seconda espressione  $P(A \cap B) = P(B|A)P(A)$ . Sostituendo nella prima espressione otteniamo (e dimostriamo) il teorema di Bayes:  $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ .

Vediamo ora un'applicazione del teorema, in modo tale che capiate a cosa serve.

→ Immaginate di avere un test della glicemia sensibile all'80% e specifico al 95% (ciò significa che avete scelto il vostro cut-off in modo tale che il 95% dei sani sia negativo e l'80% dei diabetici positivo). Sapete che la probabilità di avere il diabete, nella popolazione generale, è del 10%. A questo punto, vi chiedete: “qual è la probabilità che un soggetto diabetico abbia la glicemia sopra la norma?”.

Iniziamo con il definire gli eventi: a partire dalla nostra richiesta possiamo definire l'evento A come “essere diabetico” e l'evento B come “avere la glicemia sopra la norma”.

Ricaviamo così:

- $P(A)=0.10$  [probabilità di essere diabetico]
- $P(B|A)=0.80$  [probabilità di avere la glicemia sopra la norma se si è diabetici – corrisponde alla sensibilità del test]

Dobbiamo calcolare  $P(A|B)$ . Applicando il teorema di Bayes vediamo che questa è pari al rapporto tra  $0.08$  [cioè  $P(B|A)P(A)$ ] e  $P(B)$ . Ma quanto vale  $P(B)$ ?  $P(B)$  è la probabilità di avere la glicemia sopra la norma, cioè di essere un vero positivo o un falso positivo.

La probabilità di essere un vero positivo già la sappiamo: è il prodotto tra la probabilità di avere la glicemia sopra la norma essendo diabetico e la probabilità di essere diabetico [ovvero  $P(B|A)P(A)$ ]; analogamente, la probabilità di essere un falso positivo è il prodotto

tra la probabilità di avere la glicemia sopra la norma essendo sano (cioè  $1-SP=0.05$ ) e la probabilità di essere sano (cioè  $0.90$ ): dunque, ricapitolando:

- $P(\text{Vero+})=P(B|A)P(A)=0.80 \times 0.10=0.08$
- $P(\text{Falso+})=P(B|\text{non } A)P(\text{non } A)=0.05 \times 0.90=0.045$  [non A è la negazione di A: se A è “essere diabetico”, non A è “non essere diabetico”, cioè “essere sano”]
- A questo punto otteniamo:  

$$P(B)=P(\text{Vero+ OR Falso+})=P(\text{Vero+})+P(\text{Falso+})=0.08+0.045=0.125$$
[sono eventi mutuamente esclusivi, pertanto hanno intersezione nulla].
- Quindi  $P(A|B)=\frac{P(A \text{ AND } B)}{P(B)}=\frac{0.08}{0.125}=0.64$  .

Possiamo chiamare  $P(A|B)$  – che ricordate, è la probabilità di essere diabetico avendo la glicemia sopra la norma – come la probabilità A POSTERIORI [nel senso: probabilità di essere diabetico DOPO aver fatto il test]. La probabilità  $P(B|A)$  – probabilità di avere un test sopra la norma essendo diabetico – e  $P(A)$  – probabilità di essere diabetico nella popolazione generale – sono invece probabilità A PRIORI [PRIMA di aver fatto il test: nello specifico, la prima delle due dipende dal cut-off scelto, la seconda dalla prevalenza].

Andando oltre, possiamo evidentemente intendere  $P(A|B)$  come la probabilità di diagnosi: facendo il test, se troviamo un valore della glicemia sopra la norma abbiamo il 64% di probabilità che la diagnosi di diabete che facciamo sia giusta!

Potete notare in questo senso che fare il test aumenta di molto la probabilità di fare una diagnosi ESATTA di diabete: se diagnosticassi il diabete senza fare il test avrei solo il 10% di fare una diagnosi giusta.

Ancora,  $P(A|B)$  esprime quanto è verosimile che il mio paziente abbia il diabete.

Un'applicazione teorica del teorema di Bayes è dunque che ci permette di calcolare la probabilità della causa (“quanto è probabile che la causa di un test sopra la norma sia il diabete?”) a partire dalla probabilità dell'effetto (“se hai il diabete è probabile all'80% che il tuo test sia positivo”). Senza questo strumento statistico, potremmo solo calcolare la probabilità dell'effetto (“test positivo”) a partire dalle cause (“diabete”).

Nel paragrafo precedente abbiamo visto che è possibile esprimere la probabilità anche come odds. Vediamo qualche applicazione di questa possibilità, riutilizzando l'esempio precedente.

Possiamo per esempio riscrivere  $P(A)$  in termini di odds:  $\frac{0.10}{1-0.10}=\frac{1}{9}$  .

A questo punto, possiamo considerare il rapporto:

$$\frac{P(A)P(B|A)}{P(\text{non } A)P(B|\text{non } A)}=\frac{P(A \text{ AND } B)}{P(\text{non } A \text{ AND } B)}=\frac{0.08}{0.045}$$
 : cioè, per quanto visto in precedenza,

corrisponde al rapporto tra veri positivi e falsi positivi.

Ma fermiamoci un momento a considerare ciò che abbiamo fatto.

- $\frac{P(A)}{P(\text{non } A)}$  è un odds: infatti  $P(A)=1-P(\text{non } A)$ : possiamo considerarlo l'odds PRE-TEST, cioè l'odds che un paziente sia diabetico prima di fare il test.
- $\frac{P(B|A)}{P(B|\text{non } A)}$  è il rapporto di SE e 1-SP: infatti  $P(B|A)$  è la probabilità che un diabetico sia positivo al test (sensibilità), mentre  $P(B|\text{non } A)$  è la probabilità che un sano sia positivo al test (1-specificità); può anche essere visto come V+/F+.

Ricordando quanto visto nei capitoli precedenti, questo è il LIKELIHOOD RATIO DEL TEST POSITIVO, o LR+.

- L'ultimo risultato non riuscite a vederlo come un odds: infatti  $0.08 \neq 1-0.045$  . Tuttavia, dividendo numeratore e denominatore per 0.125, che è la  $P(B)$  (ovvero la probabilità che il test sia positivo), osserviamo che questi rapporti definiscono un odds! Infatti la nostra probabilità totale ora diventa 0.125, e non più 1 (in generale

$$odds = \frac{P(evento)}{P(totale) - P(evento)}, \text{ e in quasi tutti i casi consideriamo giustamente } 1 \text{ la nostra}$$

probabilità totale: dividendo numeratore e denominatore per 0.125 abbiamo ristretto il nostro totale ai soli casi in cui il test è positivo, che hanno una probabilità di 0.125; è come se stessimo definendo un “odds condizionato”). Dunque,  $0.08 = 0.125 \cdot 0.045$ .

Chiameremo questo odds l'odds POST-TEST, cioè l'odds che il paziente sia diabetico DOPO aver fatto il test.

Come ben sapete dall'odds è possibile risalire alla probabilità con la formula  $P = \frac{odds}{odds + 1}$ , e

questo ci permette di fare dei confronti tra la probabilità che un soggetto ha di avere la malattia prima del test e la stessa probabilità dopo essere risultato positivo al test. In ultima analisi, ciò ci fa capire se ha senso o no fare il test. Un esempio vi chiarirà il concetto:

→ Supponiamo che una malattia abbia prevalenza del 20% nella popolazione generale, e che un test in grado di determinarla abbia  $LR+ = 10$ . Come varia la probabilità che il paziente sia malato, se risulta positivo al test?

Applicando la formula  $(odds \text{ pre-test}) \cdot (LR+) = odds \text{ post-test}$  possiamo vedere che

$$\frac{0.20}{0.80} \cdot 10 = \frac{2}{0.80} = 2.5; \text{ la probabilità post-test ricavata dall'odds post-test è } \frac{2.5}{2.5 + 1} \approx 0.71.$$

Dunque la probabilità che un paziente sia malato aumenta dal 20% SENZA il test al 71% SE POSITIVO al test: in questo caso è ovvio che farò il test al paziente.

Se invece il  $LR+$  fosse 2? La probabilità post-test diventerebbe il 33% (fate i calcoli come prima). In questo caso il test non mi aiuterebbe molto a capire se il mio paziente è malato, e fargli fare il test potrebbe essere uno spreco di soldi o un rischio ingiustificato nel caso in cui il test sia rischioso.

Da quanto detto fino ad ora è chiaro che un  $LR+$  alto è molto informativo sullo stato di malattia nel caso in cui il test sia positivo, e possiamo usarlo per capire se ha senso o meno fare il test.

Un discorso analogo può essere fatto per  $LR-$ . Partendo dalle stesse considerazioni otterremo una formula del tipo  $(odds \text{ pre-test}) \cdot (LR-) = odds \text{ post-test}$ , e in questo caso otteniamo l'informazione su quanto si abbassa la probabilità che il paziente sia malato nel caso in cui il test sia NEGATIVO [ci aiuta quindi a capire se dobbiamo escludere una determinata malattia per quel paziente].

→ Per fare un esempio: se una malattia ha prevalenza del 30% e un test ha  $LR-$  pari a  $1/18$  [ $SE = 95\%$  e  $SP = 90\% \rightarrow 0.05/0.90 = 1/18$ ], l'odds post-test sarà dato dalla formula:

$$\frac{0.30}{0.70} \cdot \frac{1}{18} = \frac{0.30}{12.6} \approx 0.023.$$

La probabilità post-test sarà quindi  $\frac{0.023}{1.023} \approx 0.022$ .

Poiché se il paziente ha un esito negativo al test la sua probabilità di essere malato si abbassa dal 30% al 2%, avrà senso fare il test per ESCLUDERE la malattia, esattamente come nel caso del test con  $LR+$  alto aveva senso fare il test per CONFERMARE la malattia.

Ricordate:

- $LR+$  alto: test che si può usare per confermare la malattia se positivo
- $LR-$  basso: test che si può usare per escludere la malattia se negativo
  - N.B.: confrontare sempre le probabilità pre- e post-test → se il test ha  $LR+$  alto, ma la probabilità pre-test è già alta forse non ha senso fare il test! Per esempio, se ho probabilità pre-test del 95% e un  $LR+$  di 10 la probabilità post-test diventa 99%. Ma anche 95% non era male, e fare il test non cambia molto!

Vediamo ora qualche applicazione del teorema di Bayes attraverso due semplici esercizi.

ESERCIZIO 1.

“Due urne A e B contengono 20 palline ciascuna, di cui rispettivamente 8 e 4 palline rosse

Lanciando una moneta si sceglie l'urna A se esce testa, l'urna B se esce croce. Quindi si estrae una pallina dall'urna scelta. Qual è la probabilità di aver pescato dall'urna A sapendo che abbiamo estratto una pallina rossa?”

→ Come sempre, quando vi viene posto un problema cercate di essere il più razionali possibile: vi servono sempre tre elementi – richiesta, dati, procedimenti.

- Richiesta: probabilità di aver pescato dall'urna A sapendo che abbiamo estratto una pallina rossa. Può essere ovviamente intesa come una probabilità condizionata, che chiameremo  $P(A|rossa)$ .
- Dati:
  - A → 8 rosse su 20 → probabilità di pescare una rossa  $8/20=0.40$
  - B → 4 rosse su 20 → probabilità di pescare una rossa  $4/20=0.20$ 
    - Queste due probabilità le possiamo vedere come probabilità condizionate:  $P(rossa|A)$  e  $P(rossa|B)$  rispettivamente
  - Probabilità di scegliere l'urna A:  $\frac{1}{2}$  [infatti lanciamo una moneta!]
  - Probabilità di scegliere l'urna B:  $\frac{1}{2}$ 
    - Queste invece sono probabilità semplici (dette marginali):  $P(A)$  e  $P(B)$  rispettivamente
- Procedimento: in questo caso abbiamo dobbiamo trovare una probabilità condizionata, e abbiamo come dati una probabilità condizionata (esattamente “l'inversa” di quella che dobbiamo trovare) ed una probabilità marginale. È abbastanza chiaro che useremo il teorema di Bayes, che possiamo applicare nella forma: 
$$P(A|rossa) = \frac{P(rossa|A)P(A)}{P(rossa)}$$

Tuttavia, per applicare questo teorema ci manca la  $P(rossa)$ , che è una probabilità a priori. Troviamo facilmente questa probabilità: essendoci 12 palline rosse su 40 nel totale delle due urne, la probabilità  $P(rossa)=12/40=0.30$ .

Applicando Bayes 
$$P(A|rossa) = \frac{0.40 \times 0.50}{0.30} = \frac{2}{3} = 0.66$$

Analogamente, 
$$P(B|rossa) = \frac{1}{3} = 0.33$$

→ Osserviamo che in questo caso Bayes era francamente inutile: si poteva ricavare subito la probabilità condizionata. Sapendo che l'insieme delle rosse è dato da 12 palline e che 8 sono nella A e 4 nella B [e che A e B sono equiprobabili] la probabilità di aver estratto l'urna A è  $8/12$ , ovvero 0.33.

## ESERCIZIO 2.

“Abbiamo tre scatole: A, B e C, ciascuna contenente delle palline. Lanciamo un dado: se esce 1 o 2 scegliamo la scatola A, se esce 3 scegliamo la scatola B, se esce 4 o 5 scegliamo la C; se esce 6 rilanciamo il dado. Nelle tre scatole sono contenute palline 10 palline di colore rosso, bianco e verde, con questa distribuzione:

	A	B	C
rosse	3	2	4
bianche	2	6	2
verdi	5	2	4

Assegnare la probabilità a posteriori di aver estratto una scatola sapendo di aver estratto un colore.”

→ Anche in questo caso l'esercizio prevede la stessa tipologia di risoluzione già vista.

Iniziamo col determinare le probabilità di estrazione delle scatole: attenzione a non farvi trarre in inganno dal fatto che utilizziamo un dado – poiché rilanciamo nel caso esca 6, escludiamo tale numero ed è come se il dado avesse solo 5 facce! Dunque le probabilità saranno:

- $P(A)=2/5=0.4$  [infatti A può essere estratta se escono due numeri sui 5 possibili]

- $P(B)=1/5=0.2$
- $P(C)=2/5=0.4$

A questo punto possiamo scrivere le probabilità condizionate di ottenere una pallina di ciascun colore all'interno delle scatole specifiche; semplicemente, per quanto visto sopra, dovremmo calcolare la probabilità di un colore all'interno di ciascuna scatola. Possiamo esprimere il risultato con una tabella:

	A	B	C
P(rossa scatola)	0,3	0,2	0,4
P(bianca scatola)	0,2	0,6	0,2
P(verde scatola)	0,5	0,2	0,4

Possiamo fare anche di più: sapendo le probabilità di estrazione delle scatole  $P(\text{scatola})$  possiamo calcolare la probabilità dell'evento  $P(\text{colore AND scatola})$  – cioè il numeratore del teorema di Bayes – come il prodotto delle due probabilità (sono eventi indipendenti!).

	A	B	C
P(rossa AND scatola)	$0,3 \times 0,4 = 0,12$	$0,2 \times 0,2 = 0,04$	$0,4 \times 0,4 = 0,16$
P(bianca AND scatola)	$0,2 \times 0,4 = 0,08$	$0,6 \times 0,2 = 0,12$	$0,2 \times 0,4 = 0,08$
P(verde AND scatola)	$0,5 \times 0,4 = 0,20$	$0,2 \times 0,2 = 0,04$	$0,4 \times 0,4 = 0,16$

Facendo la somma di tutte queste probabilità possiamo verificare velocemente che corrisponde a 1 (cioè al totale delle possibilità, che sono tutte mutuamente esclusive).

Per applicare il teorema di Bayes ci manca il dato relativo alla probabilità delle palline: sapendo che abbiamo 30 palline in totale, e di queste 9 sono rosse, 10 bianche e 11 verdi (basta sommare in orizzontale) otteniamo:

- $P(\text{rossa})=9/30=0.30$
- $P(\text{bianca})=10/30=0.33$
- $P(\text{verde})=11/30=0.37$

A questo punto possiamo applicare il teorema scrivendo per ogni scatola e pallina

$$P(\text{scatola}|\text{colore}) = \frac{P(\text{colore}|\text{scatola}) P(\text{scatola})}{P(\text{colore})} = \frac{P(\text{colore AND scatola})}{P(\text{colore})} :$$

	P(A colore)	P(B colore)	P(C colore)
rossa	$0,12/0,30=0,40$	$0,04/0,30=0,13$	$0,16/0,30=0,53$
bianca	$0,08/0,33=0,24$	$0,12/0,33=0,36$	$0,08/0,33=0,24$
verde	$0,20/0,37=0,54$	$0,04/0,37=0,11$	$0,16/0,37=0,43$

Un dato interessante è che nonostante le palline bianche nella scatola B fossero 3 volte quelle presenti nella A (e nella C), siccome la probabilità di estrarre la scatola B a priori era la metà la probabilità di estrarre la scatola B data una pallina bianca è solo 1,5 volte la probabilità di estrarre la scatola A (o C) data una pallina bianca [guarda caso, proprio 3/2].

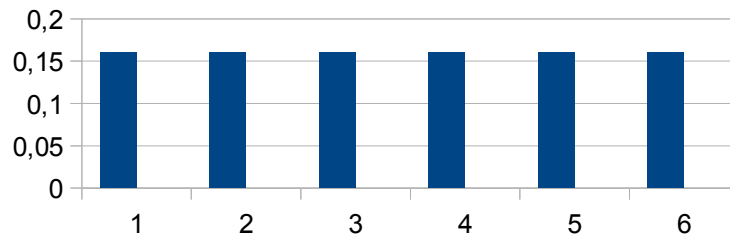
## 6.1 – Variabile casuale

Il Duca a mio avviso ha complicato incredibilmente un concetto che di incredibile ha solo la semplicità. Nella sua definizione una variabile casuale è “una funzione che associa un numero reale ad un universo di eventi, così che ad ogni numero si abbinano un valore di probabilità nel discreto (variabile casuale discreta con associato un “quantum” di probabilità) o nel continuo (variabile casuale continua con associata una densità di probabilità)”. Per me è una definizione oscena, quindi imparatela pure, ma ve ne darò un'altra.

Per me una variabile casuale è una “variabile (continua o discreta) che assume valori distinti a relazione degli esiti possibili di un fenomeno casuale; per ogni valore assunto dalla variabile si associa la probabilità dell'esito corrispondente.”

Comunque prendiate la definizione, le caratteristiche sono sempre le stesse:

- La variabile è una FUNZIONE.
  - Una funzione è una relazione che associa ad un elemento di un insieme (per fare un esempio, le  $x$ ) ad uno ed un solo elemento di un altro insieme (le  $y$ ).
  - Nel nostro caso la variabile casuale è una funzione nel senso che associa ad un numero dell'insieme dei numeri reali (non per forza gli interi quindi!) uno ed un solo elemento dell'insieme degli esiti possibili di un fenomeno casuale.
    - Per esempio, lanciando un dado otteniamo  $\{1,2,3,4,5,6\}$  come universo degli eventi. Una variabile casuale associa un numero reale ad uno ed un solo esito: chiamando la variabile casuale “ $x$ ” assoceremo:  $x=1 \rightarrow$  esito: 1,  $x=2 \rightarrow$  esito: 2, ecc. Questo potrà sembrarvi così intuitivo da chiedervi: ma che senso ha definire una funzione se il valore della variabile è uguale al valore dell'esito? In questo caso praticamente nessuno. Un esempio diverso vi chiarirà il senso del tutto:
      - Definiamo il fenomeno casuale come “estrarre una persona da un elenco di 6 persone: Mario, Gigi, Peach, Yoshi, Toad, Bowser”. Potremmo essere in difficoltà a rappresentare in un grafico la probabilità che uno di questi venga estratto: infatti in un grafico dobbiamo mettere sugli assi dei VALORI, non dei nomi! Dunque, un buon modo di aggirare l'ostacolo è associare a ciascuno dei 6 nomi un numero (che definirà la nostra variabile casuale): per esempio  $x=1 \rightarrow$  Bowser,  $x=2 \rightarrow$  Gigi, ecc. Nulla ci vieta di associare la variabile in modo diverso; potremmo dire  $x=1 \rightarrow$  Mario,  $x=2 \rightarrow$  Peach e così via. L'importante è associare ad ogni numero uno e un solo elemento dell'universo degli eventi.
- La variabile può essere discreta o continua.
  - Nel caso del lancio dei dadi abbiamo una variabile discreta, mentre nel caso delle altezze possibili in un campione di popolazione abbiamo una variabile continua. La differenza tra discreto e continuo spero sia nota; nel caso non lo fosse, il modo più semplice di intenderla è questo  $\rightarrow$  ci sono “spazi” tra i valori? Allora è discreta. Per esempio, tra 1 e 2 c'è 1.2, 1.5, 1.7, ecc.
- Alla variabile casuale  $x$  così definita è associato il valore di probabilità  $y$  dell'esito corrispondente.
  - Abbiamo fatto un'associazione esito  $\rightarrow$  variabile; ora definiamo una funzione che associa ad ogni valore della variabile la probabilità dell'esito a cui quel valore era associato, in modo da avere una funzione finale: esito  $\rightarrow$  variabile  $\rightarrow$  probabilità. In sostanza la variabile casuale è una variabile “di passaggio”.
  - Per fare un esempio pratico, riprendiamo il caso del lancio dei dadi: ogni valore ha probabilità di  $1/6$ . Dunque esito:1  $\rightarrow$   $x=1 \rightarrow 1/6$ ; esito:2  $\rightarrow$   $x=2 \rightarrow 1/6$ , e così via. Otterremo un grafico di questo tipo:



- Osserviamo che se la variabile è discreta la probabilità associata sarà anch'essa discreta. Non potrà avere tutti i valori di probabilità possibili da 0 a 1, in quanto gli elementi a mia disposizione (i valori della variabile) sono discreti, non continui.

Il concetto richiede un po' di conoscenza dell'analisi matematica. Vi riassumo la cosa nel minimo indispensabile per farvi capire (anche se il concetto è intuitivo e potreste aver già capito).

L'insieme dei numeri reali  $R$  comprende tre sottoinsiemi: l'insieme  $Z$  dei numeri interi (1, 2, 3, -5, -1239 ecc), l'insieme  $Q$  dei numeri razionali (quelli esprimibili con frazioni: 2.5, 5/7, 142/73, -56.8 ecc) e l'insieme  $I$  dei numeri irrazionali (quelli che hanno infiniti numeri dopo la virgola; in genere sono le radici di numeri interi, come  $\sqrt{2}$  ma anche il pi greco è un numero irrazionale per esempio). L'insieme degli  $Z$  positivi è detto insieme dei numeri naturali  $N$ . Risulta chiaro che  $Q$  contiene  $Z$ , che contiene  $N$ .

Tutti e tre questi sottoinsiemi sono infiniti. Quando possiamo associare ad ogni elemento di un insieme infinito uno ed un solo elemento dell'insieme dei numeri naturali ( $N$ , l'insieme degli  $Z$  positivi) diciamo che questo insieme infinito è NUMERABILE.

C'è un teorema di analisi matematica molto semplice che afferma che l'insieme dei numeri reali tra 0 e 1 non può essere messo in corrispondenza con i numeri naturali; dunque ci sono più numeri reali tra 0 e 1 che in tutto  $N$ , e l'insieme  $R$  è quindi più infinito di  $N$ . Fin qui tutto intuitivo: se  $R$  contiene  $N$  è decisamente probabile che  $R$  sia più grande.

Un altro teorema afferma tuttavia che  $Q$  e  $Z$  sono numerabili: quindi sono infiniti dello stesso ordine di  $N$ , e già questo è meno intuitivo.

All'atto pratico e per i nostri scopi: una variabile è continua quando è definita in  $R$  (quando cioè può assumere OGNI valore di  $R$  o di un intervallo di  $R$ ), mentre è discreta quando è definita in un insieme numerabile. Nel caso di una variabile discreta possiamo avere quindi anche infiniti valori, ma questo infinito è grande al massimo come l'insieme dei numeri naturali  $N$ .

Cosa c'entra questo col fatto che la probabilità debba essere discreta? Immaginiamo che ad ogni valore della variabile  $x$  sia associato un valore di probabilità  $p$  diverso (abbiamo creato una funzione biunivoca in sostanza). La probabilità  $p$  è definita in teoria nell'intervallo  $[0,1]$  di  $R$ , e quindi dovrebbe essere continua. Ma dato che mettiamo i valori di  $p$  possibili in corrispondenza biunivoca con un insieme (i valori della variabile  $x$ ) che al massimo è numerabile – cioè non più grande di  $N$  – allora anche i valori possibili di  $p$  in questo caso saranno numerabili; dunque, i valori possibili di  $p$  saranno distribuiti in un insieme DISCRETO dell'intervallo CONTINUO  $[0,1]$ . Ciò equivale a dire che “alla variabile discreta è abbinato un valore di probabilità nel discreto”.

Vi faccio notare che il fatto di aver usato “quantum” per definire un elemento di probabilità come ha fatto il Duca non ha alcun senso: un quanto è una quantità fisica o matematica indivisibile (cioè la più piccola possibile in un insieme) – qui non c'entra assolutamente nulla; un elemento di probabilità non è detto che sia il minimo elemento possibile, perché questo presupporrebbe o che tutte le quantità di probabilità siano uguali (questo è ciò che risulta dall'interpretazione rigorosa della sua definizione, ed è ovviamente una cagata) o che ogni probabilità possibile debba essere multipla di una probabilità minima (il che nella migliore delle ipotesi è un concetto inutile che vi

complica solo le idee).

- Capito perché ad una variabile discreta è associata una probabilità nel discreto, è ora abbastanza facile capire che ad una variabile nel continuo è associata una probabilità nel continuo.

Lui dice “densità di probabilità”: se vi ricordate ciò che abbiamo detto parlando della gaussiana, vi sarà facile capire che la densità di probabilità è una funzione  $p(x)$  di una variabile casuale tale per cui la probabilità  $P$  in un intervallo è uguale all'area sotto la

curva in quell'intervallo, cioè  $P(x_1, x_2) = \int_{x_1}^{x_2} p(x) dx$ .

Se queste definizioni implicano una degranolazione immediata dei vostri mastociti e un altrettanto immediato “mastocaz... che mi leggo sta roba”, allora ricordatevi solo questo:

- probabilità nel discreto → istogramma;
- probabilità nel continuo = densità di probabilità → come gaussiana: la probabilità è l'area sotto la curva

#TuttoMoltoBello, ma cosa possiamo fare con una variabile casuale?

Due cose molto semplici: media e varianza.

Per introdurre questo concetto, che ho notato esservi un po' ostico, pensate che un esempio di variabile casuale è dato dai voti che prendete in università. Come fate a calcolarvi la media? Moltiplicate ogni voto per i crediti corrispondenti e poi dividete per il totale dei crediti.

Immaginate di aver preso un 20 da 6 crediti e due 30 in esami da 6 crediti ciascuno. Per fare la media cosa fate?  $\frac{20 \times 6 + 30 \times 12}{18}$ . Ma potete anche scrivere  $\frac{20 \times 6}{18} + \frac{30 \times 12}{18}$ , e più

intelligentemente  $20 \times \frac{6}{18} + 30 \times \frac{12}{18} = 20 \times 0,33 + 30 \times 0,66$ .

Cosa avete fatto in sostanza? Avete moltiplicato il voto (la variabile) per la frequenza relativa con cui avete preso quel voto (cioè per rapporto tra il numero di crediti corrispondenti a quel voto e il totale dei crediti). Ma la frequenza relativa con cui avete preso quel voto non è in un certo senso una probabilità? È sempre definita tra 0 e 1, e a lungo andare vi dà anche un'indicazione su quale voto è più probabile che prendiate al prossimo esame. Infatti per ora avete preso 20 in un terzo dei casi e 30 in due terzi: al prossimo esame supponiamo che andrete più vicino al 30 che al 20!

Poiché la probabilità è una frequenza relativa, per definire la media di una variabile casuale basta che facciamo la somma dei valori della variabile moltiplicati per le rispettive probabilità:

$$\mu_x = \sum x_i p(x_i)$$

A questo punto è facile capire cosa significa la varianza: i nostri voti avranno una certa DISTRIBUZIONE di probabilità, e la varianza sarà un indice della dispersione di questi voti intorno alla media.

Sperando che l'esempio di sopra vi abbia messo nella cornice interpretativa giusta, passiamo a capire cosa fa il Duca.

A partire dalla variabile casuale  $x$  lui definisce il VALORE ATTESO  $E(x)$  e la VARIANZA  $V(x)$ .

Il valore atteso è la media: in effetti, vi aspettate che se fate molti tentativi il valore più frequente (cioè quello più atteso) sia proprio la media. In realtà questo concetto è un po' fuorviante: parlare di valore atteso vi potrebbe far pensare che tale valore esista e sia atteso. Ma per esempio la vostra media in università può essere 25.4 ed è chiaro che non prenderete mai 25.4 in un esame, dato che non è un valore che la vostra variabile “voto” può assumere. Tuttavia, se potesse assumerlo, potreste prendere sempre 25.4 e la media sarebbe uguale alla media dei voti che avete preso. Tutto qui.

La varianza è esattamente ciò che sta sotto radice nella formula della deviazione standard; si può

anche chiamare “scarto quadratico medio”. Ricordate come era la formula?  $\sigma_x^2 = \frac{\sum (x - \mu_x)^2}{n}$ .

Ora, facendo qualche passaggio che vi metto in nota<sup>7</sup>, otteniamo  $\sigma_x^2 = \sum x^2 p(x) - \mu_x^2$ .

Poiché, come potete vedere in nota, il primo termine è la media dei valori al quadrato di x, chiameremo “alla Duca” questo primo termine “valore atteso dei quadrati”.

Otteniamo dunque, nella notazione del Duca:

- Valore atteso di x (media):  $E(x) = \sum x_i p(x_i)$
- Varianza di x:  $V(x) = E(x^2) - E^2(x) = \sum x_i^2 p(x_i) - (\sum x_i p(x_i))^2$ .

## 6.2 – Modello binomiale

Definiamo un tipo particolare di variabile casuale, la variabile casuale binomiale.

Viene definita su un esperimento che abbia le seguenti caratteristiche:

- per ogni prova si hanno solo due possibilità: successo o insuccesso
- la probabilità dell'evento che dà origine al successo è costante
- le prove fatte sono indipendenti

La variabile assumerà quindi valore 1 se otteniamo un successo, 0 se otteniamo un insuccesso (in questo senso è detta binomiale: ha solo 2 valori!).

Il punto è che questo esempio vi spiega chiaramente la necessità di introdurre una variabile casuale: se lanciamo una moneta abbiamo esattamente due eventi possibili, e risulta abbastanza intuitivo assegnare 0 all'uscita croce e 1 all'uscita testa (o viceversa, dipende da come consideriamo il successo). Ma se lanciamo un dado? Abbiamo 6 eventi possibili: possiamo decidere per esempio di assegnare il “successo” all'uscita del 6, e l'insuccesso all'uscita di tutti gli altri numeri. Alla nostra variabile casuale x corrisponderanno quindi i seguenti valori di probabilità:  $x=1 \rightarrow p=1/6$ ;  $x=0 \rightarrow p=5/6$ .

La variabile casuale ci è servita per ridurre un universo degli eventi più grande di due elementi ad una semplice alternativa: successo o insuccesso. Avremmo potuto anche scegliere come “successo” l'uscita di un numero pari, e in quel caso avremmo avuto  $x=1 \rightarrow p=1/2$ ;  $x=0 \rightarrow p=1/2$ .

In questo caso è intuitivo pensare che se facciamo n prove indipendenti (cioè se lanciamo il dado n volte) otterremo n/6 successi e 5n/6 insuccessi.

Tuttavia una variabile del genere non è molto utile, e non ci dice nulla che non sapessimo già prima. In genere noi siamo interessati al numero di successi in un certo numero di prove. Questa richiesta

7 Iniziamo per semplicità dal numeratore:  $\frac{\sum (x - \mu_x)^2}{n}$  può essere riscritto e sviluppato come

$$\sum \left( x - \frac{\sum x}{n} \right)^2 = \sum \left( x^2 + \left( \frac{\sum x}{n} \right)^2 - 2x \frac{\sum x}{n} \right) . \text{ Applicando la sommatoria otteniamo come risultato:}$$

$$\sum x^2 + \sum \left( \frac{\sum x}{n} \right)^2 - 2 \frac{\sum x \sum x}{n} ; \text{ notiamo sommare una quantità sempre uguale per n volte equivale a}$$

moltiplicarla per n, per cui il secondo termine (che è la media al quadrato, e la media è sempre la stessa!) è  $\sum \left( \frac{\sum x}{n} \right)^2 = n \left( \frac{\sum x}{n} \right)^2 = \frac{(\sum x)^2}{n}$ . Il terzo termine è invece  $-2 \frac{\sum x \sum x}{n} = -2 \frac{(\sum x)^2}{n}$ .

Sostituendo nella formula iniziale otteniamo  $\sum (x^2) - \frac{(\sum x)^2}{n}$ . Questo era il numeratore (la devianza).

Dividendo ora per n (per ottenere la varianza) abbiamo  $\frac{\sum (x^2)}{n} - \frac{(\sum x)^2}{n^2} = \frac{\sum (x^2)}{n} - \left( \frac{\sum x}{n} \right)^2$ . Il secondo

termine è la media al quadrato (ovvero il valore atteso al quadrato), mentre il primo termine è la media dei QUADRATI di x, cioè per quanto abbiamo visto prima  $\frac{\sum x^2}{n} = \sum x^2 p(x)$ , e la formula diventa quella scritta su.

definisce la cosiddetta distribuzione binomiale: si tratta di calcolare appunto la probabilità di  $k$  successi in  $n$  prove per ogni valore di  $k$ .

Ritornando all'esempio del dado, potremmo chiederci quale sia la probabilità di ottenere 6 in 4 lanci. Consideriamo un 6 come successo (1) e non 6 come insuccesso (0). In 4 lanci abbiamo le seguenti possibilità:

Numero di successi	Possibilità che realizzano $n$ successi
0	0000
1	1000 – 0100 – 0010 – 0001
2	1100 – 0110 – 0011 – 1001 – 1010 – 0101
3	1110 – 0111 – 1011 – 1101
4	1111

Dovete leggere la colonna di destra come le combinazioni di successi e insuccessi (1 e 0) che realizzano gli  $n$  successi: per esempio per 0 successi c'è solo una combinazione → devono uscire 4 insuccessi di fila (non6-non6-non6-non6). Per 1 successo ci sono invece 4 combinazioni possibili: l'unico 6 può uscire al primo lancio, al secondo, al terzo o al quarto.

Per avere esattamente un successo vuol dire che in un lancio deve uscire un 6, e quindi si deve realizzare un evento con  $p=1/6$ , mentre negli altri 3 lanci non deve uscire un 6, e quindi si devono realizzare 3 eventi con  $p=5/6$  ciascuno. Possiamo dunque scrivere la probabilità di un successo in 4 prove come  $P(6 \text{ AND non } 6 \text{ AND non } 6 \text{ AND non } 6)$ , e poiché tutti i lanci sono indipendenti la probabilità si calcola come il prodotto delle singole probabilità:

$$P(1 \text{ successo in } 4 \text{ prove}) = \frac{1}{6} \frac{5}{6} \frac{5}{6} \frac{5}{6} = \frac{1}{6} \left(\frac{5}{6}\right)^3 .$$

Possiamo fare un analogo ragionamento per la probabilità di 2 successi in 4 prove, ma dobbiamo notare che in questo caso il valore ottenuto andrà moltiplicato per il numero di possibilità di realizzazione dell'evento, che sono 4. Dunque  $P(2 \text{ successi in } 4 \text{ prove}) = 4 \frac{1}{6} \frac{5}{6}^2$  .

Più in generale, osserviamo che il numero di modi in cui si realizzano  $k$  successi in  $n$  prove è dato dal numero di combinazioni possibili, esprimibili attraverso il coefficiente binomiale  $\binom{n}{k}$  .

→ Una precisazione sul perché si usa il coefficiente binomiale in questo caso. Abbiamo detto che rappresenta il numero di combinazioni possibili di  $k$  elementi in  $n$  tentativi. Detto così avreste tutti la tentazione di dire “ok” e accettare sulla fiducia. Siccome a me questa cosa dà un fastidio allucinante, mi sono chiesto: ma è proprio così?

No. Non proprio. Se fossero davvero combinazioni l'ordine non conterebbe: due successioni con gli stessi elementi disposti in modo diverso conterebbero una sola volta! Quindi 1000, 0100, 0010 e 0001 non sarebbero 4 combinazioni, ma solo una. Il coefficiente binomiale tuttavia, che si usa per le combinazioni semplici, ci dà come risultato 4, e non 1. Dunque in questo caso il coefficiente binomiale non ci dà il numero di combinazioni.

Ragionando più attentamente, osserviamo che abbiamo preso un insieme con  $k$  successi e  $n-k$  insuccessi, rappresentabile così:  $\{1,0,0,0\}$ . Ciò che abbiamo messo in tabella sono in realtà le permutazioni di questo insieme, che ha  $k$  elementi ripetuti (gli 1) e  $n-k$  elementi ripetuti (gli 0). Se ricordate la formula delle permutazioni con ripetizioni e la applicate a questo caso ottenete:  $\frac{n!}{k!(n-k)!}$  , che coincide con il coefficiente binomiale  $\binom{n}{k}$  .

Infatti, è facilmente dimostrabile che nel caso di permutazioni di un insieme di  $n$  elementi di cui  $k$  ed  $n-k$  rispettivamente ripetuti tra loro la formula delle permutazioni con ripetizioni è sempre esprimibile con il coefficiente binomiale.

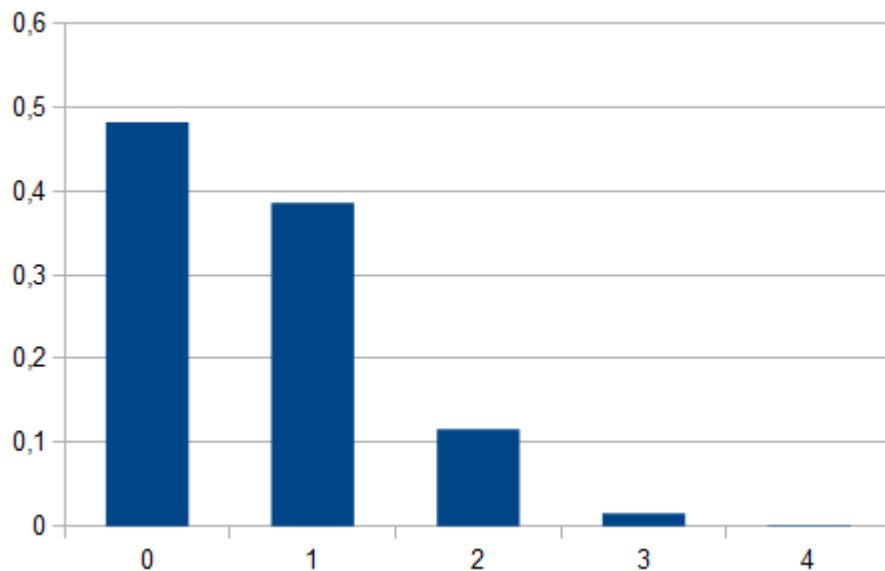
Quindi stiamo attenti: il coefficiente binomiale in questo caso rappresenta le

PERMUTAZIONI di k successi in n tentativi, e non le combinazioni, anche se intuitivamente noi usiamo la parola “combinazione” un po' a casaccio e quindi non siamo portati a notare un'incongruenza.

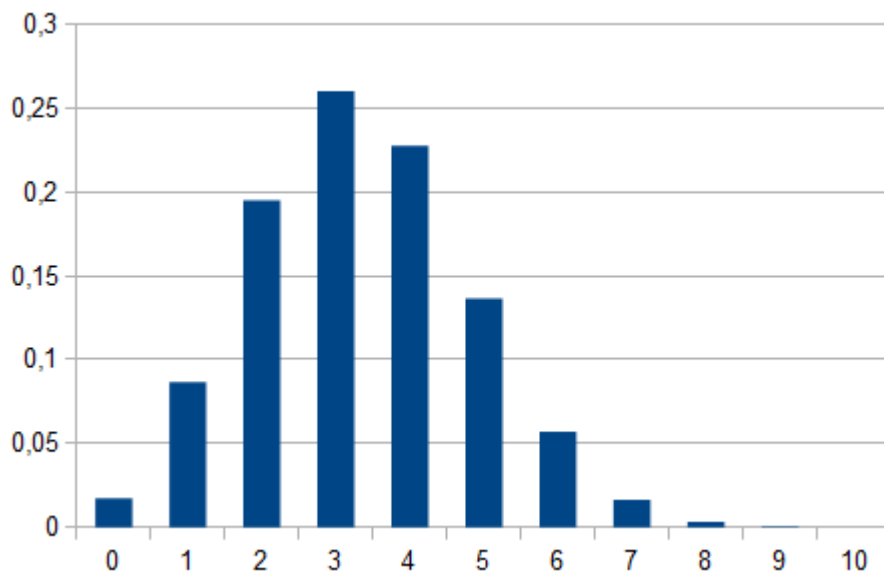
Generalizzando, la formula finale della probabilità binomiale è la seguente:

$$P_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

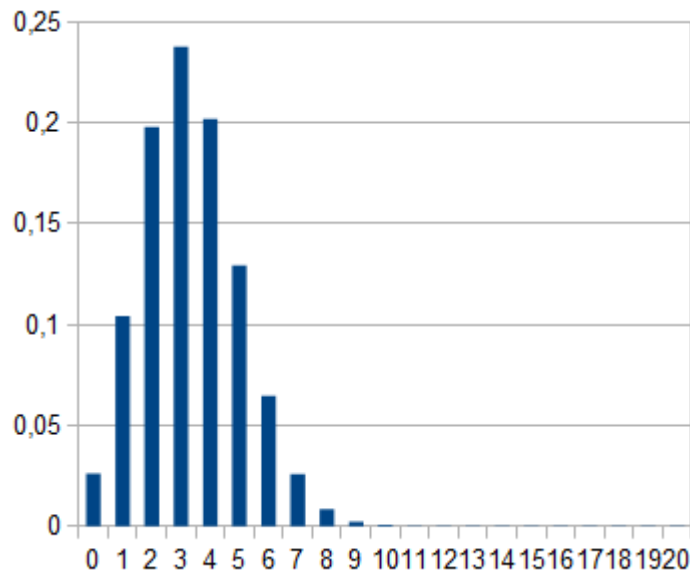
, ed esprime come già detto la probabilità di k successi in n prove, con p probabilità di successo in una singola prova. La sua notazione è spesso indicata come  $B_{n,p}(k)$ . Vi metto di seguito la forma di alcune distribuzioni binomiali in modo che possiate capire che aspetto grafico ha:



*In ordinata probabilità di k successi in 4 lanci di un dado [successo=uscita di un 6]; k in ascissa*



*Illustrazione 1: In ordinata probabilità di k successi in 10 lanci di un dado [successo=uscita di 5 o 6]; k in ascissa*



*Illustrazione 2: Probabilità di ottenere k successi in 20 lanci di due dadi [successo = esce un 7 come somma dei due dadi --> probabilità di successo su singola prova 1/6: 7 esce 6 volte su 36]*

Possiamo calcolare media e varianza di una distribuzione binomiale, secondo i seguenti passaggi:

- Media (valore atteso):  $E(k) = \sum_{k=0}^n k p(k) = \sum_{k=0}^n k B_{n,p}(k) = np$  [[prova](#)]
- Varianza:  $V(k) = \sum_{k=0}^n (k - E(k))^2 p(k) = \sum_{k=0}^n (k - np)^2 B_{n,p}(k) = np(1 - p)$  [la prova è alla pagina 2 dello stesso link].

→ Possiamo fare due osservazioni: innanzitutto possiamo notare che la distribuzione è normalizzata: infatti la somma di tutte le probabilità è uguale a 1. Ciò si dimostra attraverso l'uso della formula di Newton, che esprime l'espansione di un binomio.

Questa formula stabilisce che  $\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p+q)^n$ . Nel nostro caso tuttavia notiamo che  $q=1-p$ , e pertanto la somma di tutte le probabilità binomiali (che insieme danno la distribuzione binomiale) è uguale a  $(p+1-p)^n = 1^n = 1$ .

→ La seconda osservazione riguarda l'utilizzo che facciamo della distribuzione binomiale: perché più in generale ci serve in campo medico? Perché ci permette di saggiare ipotesi di efficacia. Un esempio vi chiarirà il tutto.

Supponiamo che una casa farmaceutica venda un farmaco, il Binozumab, che dice di essere efficace nel curare una malattia che ha una mortalità del 40%. La cosa più logica da fare è testare questo farmaco: prendiamo allo scopo un campione di 100 persone. Se il farmaco non funzionasse, ci aspetteremmo che circa 40 persone muoiano e 60 no. È chiaro che se muoiono 70 persone considereremo il farmaco addirittura dannoso, ma se ne muoiono solo 30 sarà il Binozumab davvero efficace? Come facciamo a decidere?

La risposta è molto semplice: supponiamo che il farmaco non abbia alcun effetto (ipotesi nulla). La probabilità di un successo è in questo caso il 60%; ma su 100 persone quanti successi ci aspettiamo? La risposta è esattamente la media della distribuzione binomiale.

Ricaviamo la distribuzione binomiale per questo problema:

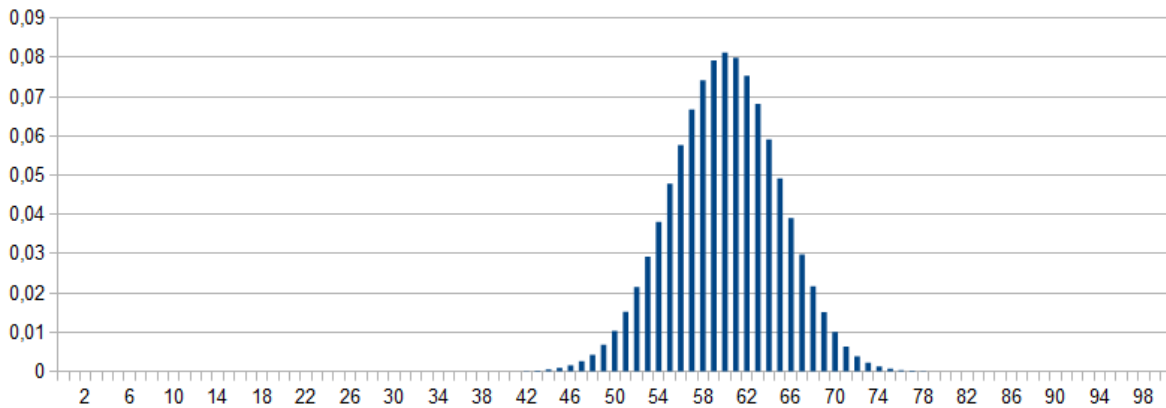


Illustrazione 3: Probabilità di  $k$  sopravvivenze in 100 persone che assumono il farmaco Binozumab supponendo che il farmaco sia del tutto inefficace

Ricaviamo ora media e varianza:

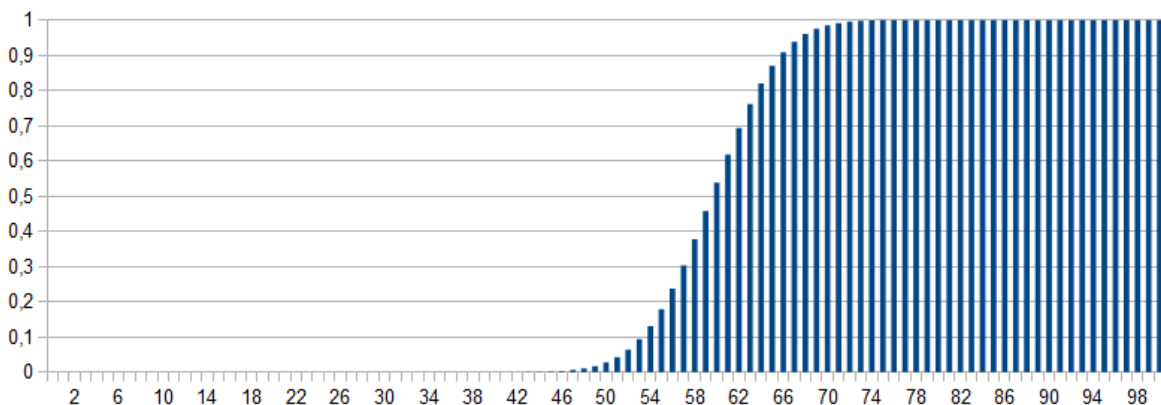
- $E(k)=100 \times 0,6=60$
- $V(k)=60 \times 0,4=24$

Ci attendiamo in media 60 successi se il farmaco non influisce.

Abbiamo però osservato 70 successi: sono abbastanza per dire che il farmaco è efficace? Come facciamo a dirlo? In statistica si dice che un risultato è significativo quando la probabilità che l'evento accada nell'ipotesi nulla è minore del 5%.

In questo caso noi abbiamo una distribuzione di valori, ciascuno con la propria probabilità: possiamo chiederci quale sia la probabilità ottenere  $n$  o meno successi. Poiché la binomiale è normalizzata la somma di tutte le probabilità darà 1, e dunque la domanda che ci siamo posti equivale a trovare il centile corrispondente a  $n$  successi.

La probabilità cumulativa ha il seguente andamento:



Osserviamo che a 70 lanci corrisponde visivamente un centile molto vicino a 100. Poiché il grafico è stato costruito a partire da una tabella posso andare a vedere il valore di probabilità cumulativa corrispondente a 70 successi: è circa 0,98.

Dunque, la probabilità di osservare 70 o meno successi è il 98%, mentre quella di osservare 70 o più successi è ovviamente il 2%, inferiore al 5% che rappresenta il livello di significatività statistica. Risulta quindi chiaro che se il farmaco fosse stato inefficace sarebbe stato molto improbabile osservare 70 successi: dunque, possiamo dire che il farmaco ha evidenza significativa di essere efficace.

Notate che ad intuito 70 successi non sono così lontani da 60, e non ci saremmo aspettati che fosse così improbabile osservarli!

→ Terza considerazione: guardate le illustrazioni 1, 2 e 3. Non vi sembrano assomigliare sempre più a qualcosa che già conoscete? E quel qualcosa non è una curva a campana?

Si dimostra che un numero  $n$  molto grande la distribuzione binomiale è approssimabile ad una distribuzione gaussiana con media  $\mu = E(k) = np$  e deviazione standard

$\sigma = \sqrt{\sigma^2} = \sqrt{V(k)} = \sqrt{np(1-p)}$ . In questo caso diventa quindi molto facile calcolare i centili e le probabilità che un valore osservato risulti statisticamente significativo, dato che possiamo applicare i procedimenti già studiati quando abbiamo affrontato il modello gaussiano. L'errore che si commette è tanto più piccolo quanto più il numero n di tentativi è alto.

→ Tenete presente che in matematica e in generale nelle scienze matematiche la frase “si dimostra” è un modo molto elegante di dire “non c'ho sbatti, fidatevi; se non vi fidate andatevelo a cercare, io manco lo voglio imparare”. Se siete interessati a capire perché uno non dovrebbe avere sbatti di dimostrarvi questa cosa, [eccovi accontentati](#).

### 6.3 – Teorema del limite centrale

Il vero teorema del limite centrale (TLC d'ora in avanti) nella formulazione di [Lindeberg-Lévy](#) è per voi un po' troppo complicato, oltre che non richiesto. Qui di seguito vi metterò una definizione ed una spiegazione per un caso semplificato, e successivamente farò un esempio facile. Non ne darò la dimostrazione perché anche in questo caso semplice questa prevede l'uso di concetti come funzione caratteristica (che a sua volta richiede la trasformata di Fourier) che non vi serve conoscere.

TEOREMA:

Date n variabili indipendenti  $x_i$ , distribuite secondo funzioni di distribuzione di probabilità  $f_i$ , con media  $\mu_i$  e varianza  $\sigma_i^2$ , la funzione di distribuzione di probabilità della variabile S definita come la somma delle variabili  $x_i$ ,  $S = \sum x_i$ , ha come media  $\mu_s = \sum \mu_i$  e come varianza  $\sigma_s^2 = \sum \sigma_i^2$ ; tale distribuzione si approssima alla distribuzione gaussiana  $G_{\sum \mu_i, \sqrt{\sum \sigma_i^2}}(S)$  per  $n \rightarrow \infty$ .

TEOREMA FOR DUMMIES:

Avete un certo numero di variabili casuali  $x_i$ , indipendenti l'una dall'altra: ognuna di queste è distribuita in un certo modo, ha una certa media ed ha una certa varianza. Se prendete una nuova variabile casuale S, definita come la somma di tutte le variabili  $x_i$ , allora questa variabile S avrà una media che è data dalla somma delle medie di tutte le variabili  $x_i$ , ed una varianza data dalla somma di tutte le varianze. Se le variabili  $x_i$  sono molte, la variabile S si avvicina molto ad una gaussiana con media e varianza appena definite.<sup>8</sup>

Facciamo un esempio che vi chiarirà molto il concetto.

Prendete un dado, e definite la vostra variabile casuale  $x_1$  come il valore dei possibili esiti. Ad ogni valore di  $x_1$  sarà quindi associata una probabilità di 1/6. Avremo quindi:

Valore di $x_1$	Esito	$p(x_1)$
1	1	1/6
2	2	1/6
3	3	1/6
4	4	1/6
5	5	1/6
6	6	1/6

Per calcolare media e varianza usiamo le formule viste nel paragrafo 6.1:

- $E(x_1) = \sum x_1 p(x_1) = 3,5$
- $V(x_1) = \sum (x_1)^2 p(x_1) - E^2(x_1) \approx 2,9$

Introduciamo ora una seconda variabile  $x_2$ , associata ancora al lancio di un dado nello stesso modo. Assumerà valori identici, che avranno probabilità identiche, e avrà la stessa media e varianza

<sup>8</sup> Se la variabile finale fosse la variabile differenza x-y la media sarebbe la differenza delle medie e la varianza la SOMMA delle varianze delle singole variabili.

già calcolata.

A questo punto introduciamo la variabile S, somma di  $x_1$  e  $x_2$ .

Cosa vuol dire? È molto semplice: prendete ogni valore di  $x_1$  e sommate ogni valore di  $x_2$ . È esattamente come lanciare 2 dadi e fare la somma, ottenendo il risultato seguente:

$x_1 \mid x_2$	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

I valori in azzurro corrispondono ai valori di  $x_1$ , quelli in giallo a quelli di  $x_2$ , quelli in bianco ai valori di S.

Ma qual è la probabilità associata ai valori di S? Essendo  $x_1$  e  $x_2$  variabili indipendenti, sappiamo che la probabilità che esca un valore di  $x_1$  e un valore di  $x_2$  è data dal prodotto delle probabilità dei due valori, ragion per cui ogni casella bianca della nostra tabella ha probabilità di  $1/36$ . Tuttavia, vediamo che ci sono più valori di S uguali, dunque la probabilità di  $S=7$  sarà per esempio  $6/36$ , cioè  $1/6$ , mentre la probabilità di  $S=2$  sarà solo  $1/36$ .

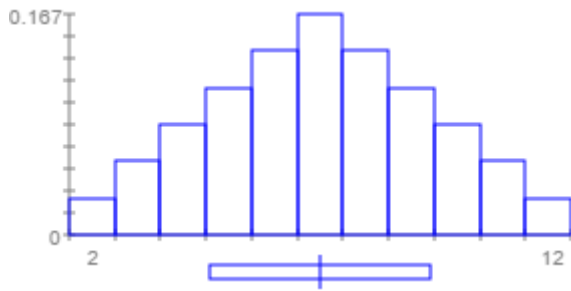
A questo punto possiamo calcolare ancora media e varianza di S, che è la nostra nuova variabile casuale:

- $E(S) = \sum S p(S) = 7$
- $V(S) = \sum (S)^2 p(S) - E^2(S) \approx 5,8$

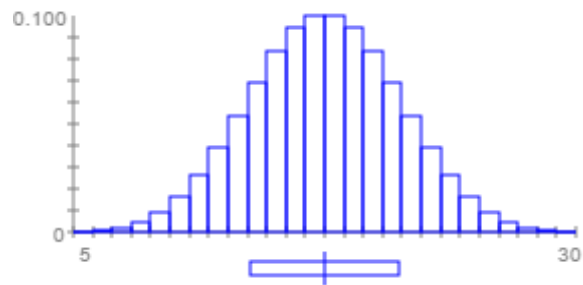
Guarda caso, i valori ottenuti sono esattamente la somma dei valori corrispondenti delle variabili  $x_1$  e  $x_2$ .

Se ora lanciassimo 4 dadi e facessimo ancora la somma, otterremmo una media di 14 e una varianza di circa 11,6. Vi e mi risparmio i calcoli, sperando che abbiate capito il senso del teorema.

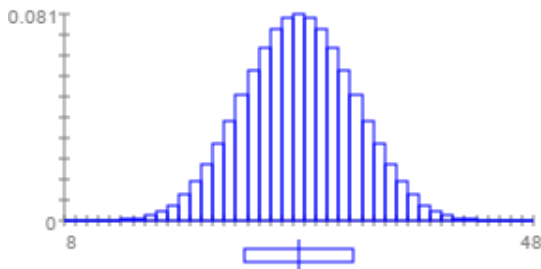
Tuttavia, può essere interessante vedere graficamente come le distribuzioni di probabilità evolvono all'aumentare del numero di variabili casuali sommate; andate su [questo sito](#) e settate come volete n nell'applet – n corrisponde al numero di dadi che decidete di sommare. Osserverete che all'aumentare di n la distribuzione diventa sempre più campaniforme, ovvero tende ad una gaussiana. Vi metto sotto qualche estratto per la consultazione off-line:



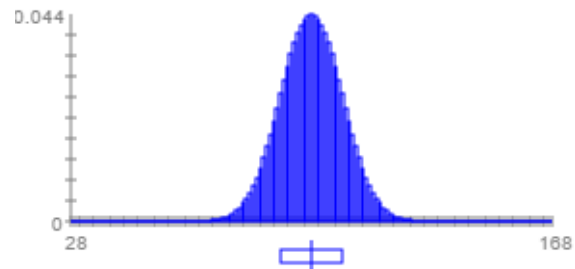
*Somma di 2 dadi*



*Somma di 5 dadi*



*Somma di 8 dadi*



*Somma di 26 dadi*

Il teorema ha importanti applicazioni nel campo delle misurazioni. La maggior parte di ciò che noi misuriamo è la somma di molte variabili casuali. Per esempio, la variabile casuale “altezza” è frutto di un'eredità poligenica, in cui ogni gene (variabile casuale che può assumere diversi valori, gli alleli! Notate che ogni allele è distribuito con una certa frequenza nella popolazione, e dunque avrà una diversa probabilità) aggiunge un certo numero di centimetri all'altezza totale dell'individuo. Ogni gene dà un piccolo contributo che dipende dall'allele, e la somma degli effetti di tutti questi geni dà l'altezza finale. Per il teorema del limite centrale, la somma di questi effetti sarà distribuita in modo gaussiano, ed è per questo motivo che osserviamo una distribuzione a campana dell'altezza nella popolazione.

Bisogna tuttavia notare che la gaussiana è un'approssimazione; più valori distanti dalla media otteniamo, più l'approssimazione con la gaussiana sarà peggiore in queste regioni. Tuttavia, la gaussiana approssima molto bene la variabile somma di variabili nei dintorni della media, e questo spesso ci basta.

Giusto per curiosità, vi faccio notare che ci sono variabili casuali che non sono la somma di altre variabili casuali, ma il loro prodotto. La variabile prodotto non si distribuirà quindi in modo gaussiano, ma in questo caso il logaritmo della variabile prodotto sarà la somma dei logaritmi delle singole variabili casuali; possiamo quindi applicare il TLC al logaritmo della variabile prodotto, che sarà asintoticamente distribuita in modo “log-normale” (log-gaussiano: cioè gaussiano, mettiamo log davanti per ricordarci che è il logaritmo della variabile ad essere distribuito in modo gaussiano, e non la variabile).

## 6.4 – Variabile casuale $\chi^2$

I prossimi due paragrafi saranno abbastanza impegnativi, e non darò dimostrazioni (dato che prevedono l'utilizzo di concetti che sono un po' troppo complicati per i vostri scopi). Ciò significa che potete prendere ciò che scriverò sulla fiducia, cercando di capire più che altro “per quale scopo” introduciamo questi argomenti e tralasciandone il perché causale. Se vi interessa sapere anche da dove saltano fuori queste cose chiedete pure e vi dirò.

La cosa più importante quando parlate di  $\chi^2$  è stare MOLTO attenti alle parole “variabile” e “distribuzione”. Siccome esistono sia una variabile casuale  $\chi^2$  che una distribuzione di

probabilità  $\chi^2$  rischierete di fare tantissima confusione. Cercherò più avanti di adottare alcune notazioni che vi impediranno di fare errori, ma metteteci un po' del vostro leggendo con molta concentrazione.

Iniziamo a parlare della VARIABILE.

La variabile casuale chi quadro ( $\chi^2$ ) è definita come la somma dei quadrati di n variabili casuali indipendenti, ciascuna distribuita secondo una distribuzione gaussiana:  $\chi^2 = \sum z_i^2$ .

Non ho utilizzato la notazione "z" per tali variabili in modo casuale. Se ricordate, z era la "variabile di trasformazione" che avevamo usato per trasformare una gaussiana qualsiasi in una gaussiana standard (con media 0 e deviazione standard 1), e valeva  $z = \frac{(x-\mu)}{\sigma}$ .

Dunque, la nostra variabile  $\chi^2$  potrà essere espressa come  $\chi^2 = \sum_{i=1}^n \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2$ , ma per semplicità e velocità usiamo la notazione con z.

La variabile casuale ha un parametro, n. Ricordate che la variabile casuale è una FUNZIONE, ed una funzione è una funzione di qualcosa. In questo caso è funzione del numero di quadrati di variabili casuali gaussiane standard sommate: considerate che molto semplicemente nel caso n=1 la variabile chi quadro è esattamente uguale alla variabile gaussiana standard al quadrato.

Questo parametro n è detto numero di gradi di libertà (il numero di gradi di libertà è notato come "d", da degrees of freedom; in questo caso n=d).

→ I gradi di libertà sono definibili molto semplicemente come il numero di variabili indipendenti. Più in generale vale la relazione d=n-c, in cui d è il numero di gradi di libertà, n è il numero di variabili e c è il numero di vincoli. Cos'è un vincolo? Un vincolo è una relazione che lega alcune variabili. In pratica quando abbiamo un'equazione che lega più variabili una di queste variabili è determinata dall'equazione, dunque non è "libera" di variare indipendentemente dalle altre.

Facciamo un esempio: se sappiamo che la somma di 3 variabili indipendenti dà un certo valore, poniamo 8, allora le variabili indipendenti sono in realtà 2. Infatti possiamo scrivere  $x_1 + x_2 + x_3 = 8$ , e di conseguenza  $x_1 = 8 - x_2 - x_3$ ; quindi, mentre le variabili  $x_2$  e  $x_3$  possono assumere qualunque valore, la variabile  $x_1$  dovrà assumere un valore che è determinato dai valori delle altre due, e di conseguenza non sarà libera.

In questo caso abbiamo un'equazione che lega le variabili, e dunque un vincolo. Dalla formula di prima ricaviamo che il numero di gradi di libertà è d=3-1=2.

Ora parliamo della DISTRIBUZIONE.

Noi sappiamo che una variabile casuale gaussiana ha una distribuzione gaussiana, data dalla formula  $G_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ ; utilizzando la notazione z otteniamo  $G_{0,1}(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$ .

Ma come si distribuirà il quadrato di z? E la somma di n quadrati di z?

Non c'è un modo per spiegarvi facilmente come si arriva (la spiegazione prevede Jacobiani e altre robe che non è opportuno sapere), ma ciò che dovete sapere è che esiste una distribuzione caratteristica, detta distribuzione  $\chi^2$ ; può essere fonte di confusione utilizzare lo stesso simbolo grafico ( $\chi^2$  appunto) sia per la variabile casuale che per la sua distribuzione di probabilità, pertanto d'ora in avanti utilizzerò la notazione  $f(\chi^2; n)$  per la distribuzione e  $\chi^2(n)$  per la variabile. Potete leggere queste notazioni rispettivamente come "funzione di distribuzione di probabilità di  $\chi^2$  e n" e " $\chi^2$  di n", indicando che i valori di  $\chi^2$  sono ottenuti in funzione di n, e i valori della distribuzione sono ottenuti in funzione dei valori di  $\chi^2$  e di n.

In generale  $f(\chi^2; n) = \frac{(\chi^2)^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}$ , con  $\Gamma(x)$  che rappresenta la funzione gamma, una

roba piuttosto complicata che non vi serve neanche sapere<sup>9</sup>. In realtà non vi serve neanche sapere com'è la distribuzione; ve l'ho messa solo perché abbiate un'idea generale.

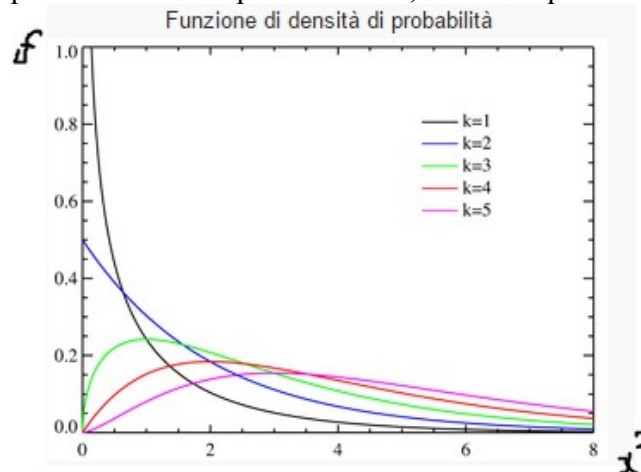
Possiamo ora calcolare media e varianza di questa distribuzione, come facciamo con tutte, e otteniamo:

- $E[\chi^2(n)] = \mu_{\chi^2} = n$
- $V[\chi^2(n)] = \sigma_{\chi^2}^2 = 2n$

Tutto molto semplice vero? Ora possiamo chiederci: e all'atto pratico? Che me ne faccio di sommare i quadrati di una serie di variabili casuali?

Iniziamo col notare che dato un n qualsiasi possiamo ricavare tutti i valori di  $f(\chi^2; n)$ ; infatti, questa distribuzione dipende da n e da  $\chi^2(n)$ , che a sua volta dipende da n (quindi in ultima analisi è una funzione i cui valori dipendono solo da n). Mi direte voi: ma come, e le variabili gaussiane al quadrato non contano? Beh, qui abbiamo introdotto un "trucchetto": considerando le variabili come z, tutte le z saranno uguali dato che hanno tutte distribuzioni gaussiane standard. E noi sappiamo i valori di una distribuzione standard, perché sono messi in tabella (ricordate ciò che abbiamo fatto parlando del modello gaussiano?). È chiaro che in un caso reale non avremo davvero distribuzioni standard, ma avremo un certo numero di distribuzioni gaussiane di cui vogliamo sommare i quadrati: nulla ci vieta tuttavia di passare alla distribuzione standard con le varie notazioni  $z_i$ .

Appurato che i valori di questa funzione dipendono da n, vi metto qui la forma della distribuzione:



Ok, siete molto confusi ora. Cerchiamo di interpretare insieme il grafico.

Inizio col farvi notare che in questo grafico "k" sta per "n", dunque in ascissa avremo i valori di  $\chi^2(k)$  e in ordinata quelli di  $f(\chi^2; k)$ .

Prendiamo la curva con  $k=1$ , ovvero quella per cui c'è una sola variabile gaussiana z elevata al quadrato. Questa corrisponderà alla variabile  $\chi^2$ . Come si distribuiscono i valori di  $\chi^2$ ? Sappiamo che la distribuzione ha come media n (in questo caso "k"); quindi la media dei valori deve essere 1. Tuttavia, vediamo che la probabilità aumenta molto in corrispondenza di valori più bassi di 1: in questo caso la media ( $\chi^2=1$ ) non è il valore più rappresentato → ciò significa che la moda non coincide con la media, e che la curva non è simmetrica.

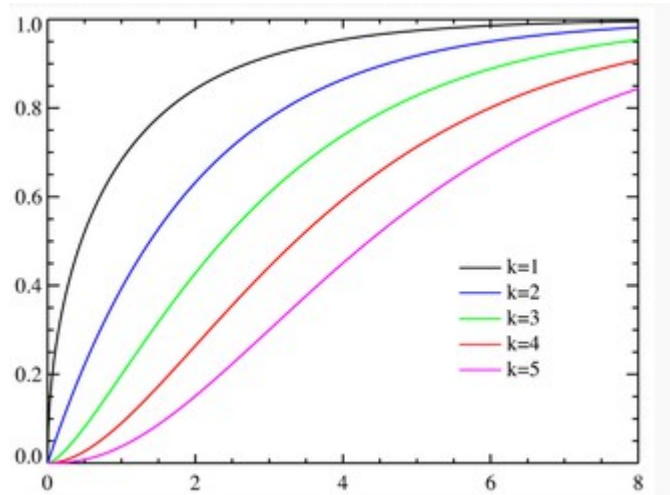
Vediamo che per k sempre maggiori la curva assume una forma sempre più simmetrica; in effetti, per k molto grande la curva si approssima ad una gaussiana.

Se non avete capito perfettamente non importa, era solo per farvi vedere graficamente com'è una distribuzione  $\chi^2$ .

Ora, facciamo un passo in più: possiamo calcolare la funzione di probabilità cumulativa al variare dei valori di  $\chi^2$  per ogni k. Ovviamente all'aumentare del valore di  $\chi^2$  tale funzione

<sup>9</sup>  $\Gamma\left(\frac{n}{2}\right) = \sqrt{\pi} \frac{(n-2)!!}{2^{\frac{n-1}{2}}}$ , con (n-2)!! che indica il [doppio fattoriale](#).

convergerà verso 1.



Vediamo di interpretare questo grafico, perché è abbastanza importante per le applicazioni che si fanno in statistica.

Come sempre, possiamo considerare i valori di probabilità cumulativa associati ai valori di una variabile in termini di centili. Quindi, preso  $k=1$ , qual è la probabilità di osservare un valore di  $\chi^2$  minore o uguale a 2? Vediamo che è circa l'80%. Questo significa che osserveremo un  $\chi^2$  maggiore o uguale di 2 in circa il 20% dei casi.

Ma se  $k=2$  la probabilità di osservare un valore di  $\chi^2$  minore o uguale a 2 diminuisce, e vale circa 60%.

Come regola generale capite facilmente che la probabilità di osservare un  $\chi^2$  maggiore di una certa quantità è esattamente 1 meno la probabilità di osservare un  $\chi^2$  minore di quella quantità.

Ora, perché ho usato il termine “osservato”?

Il  $\chi^2$  è utilizzato anche come test per valutare la significatività di un'ipotesi o la bontà di un modello. Nella sua forma generale il  $\chi^2$  può essere espresso come  $\chi^2 = \sum_{i=1}^n \left( \frac{x_i - \mu_i}{\sigma_i} \right)^2$ , cioè

come  $\sum \left( \frac{\text{valore osservato} - \text{valore atteso}}{\text{deviazione standard}} \right)^2$ ; sappiamo che la sua media sarà esattamente  $n$ , cioè “valore atteso”.

Un esempio spero vi chiarisca tutta la faccenda.

Supponiamo di avere un insieme di dati di cui non conosciamo la distribuzione, ma supponiamo che sia gaussiana. Per esempio, abbiamo misurato le altezze di una popolazione di 200 persone in un'isola e vogliamo verificare l'ipotesi che si distribuiscano secondo una curva gaussiana.

Ora, questa curva avrà una media ed una deviazione standard, che sfortunatamente non conosciamo a priori: dovremo dunque usare i dati per ricavare la media e la deviazione standard. Ma in questo caso abbiamo impostato due equazioni, o “vincoli”.

Infatti, una volta ottenuto il valore della media possiamo riscrivere l'altezza per esempio del primo soggetto come  $h_1 = 200\mu_h - \sum_{i=2}^{200} h_i$ . Allo stesso modo, possiamo utilizzare l'equazione

della deviazione standard per mettere in relazione l'altezza del secondo soggetto con tutte le altre. Quindi, una volta che abbiamo calcolato un parametro (media, deviazione standard...) a partire dai dati abbiamo idealmente posto un vincolo, e dunque abbiamo reso un dato dipendente dagli altri.

Tuttavia, l'altezza non è una variabile discreta, ma continua. Se ricordate il discorso fatto parlando degli istogrammi, vedete che un modo per rappresentare i dati di una variabile continua è dividerli in classi, cioè in intervalli.

Questo è esattamente ciò che faremo ora: divideremo le altezze in intervalli, supponiamo 8. Ma come scegliamo questi intervalli? Visto che abbiamo supposto che i dati si distribuiscano

secondo una gaussiana, possiamo scegliere i dati in termini di distanza espressa in deviazioni standard dalla media. Per ogni intervallo possiamo calcolare il numero di valori osservati (cioè il numero di persone la cui altezza si trova nell'intervallo considerato) e il numero di valori attesi (cioè il numero di persone che ci attendiamo abbiano un'altezza nell'intervallo considerato). Per calcolare questo valore atteso dobbiamo moltiplicare la probabilità che una misura cada nell'intervallo considerato (e questa si trova come già visto parlando della gaussiana) per il numero totale del campione, cioè 200.

Otteniamo quindi una tabella del genere (le divisioni in intervalli sono arbitrarie: abbiamo diviso in 8 intervalli fatti così, nulla ci vietava di dividere in 6 intervalli fatti diversamente – ma in ogni caso gli intervalli non potevano essere meno di 4, e vedremo poi perché):

Intervalli	Intervalli di altezze	Numero osservato	Numero atteso
1	$h < \mu - 1,5\sigma$	14	13,4 [0.0668x200]
2	$\mu - 1,5\sigma \leq h < \mu - \sigma$	29	18,3
3	$\mu - \sigma \leq h < \mu - 0,5\sigma$	30	30
4	$\mu - 0,5\sigma \leq h < \mu$	27	38,8
5	$\mu \leq h < \mu + 0,5\sigma$	28	38,3
6	$\mu + 0,5\sigma \leq h < \mu + \sigma$	31	30
7	$\mu + \sigma \leq h < \mu + 1,5\sigma$	28	18,3
8	$h \geq \mu + 1,5\sigma$	13	13,4

Osserviamo subito che una volta calcolati gli 8 intervalli abbiamo posto un vincolo. Infatti sappiamo che la somma dei valori osservati contenuti negli 8 intervalli deve dare 200; dunque possiamo scrivere l'ultimo intervallo come 200 meno la somma degli altri 7 intervalli.

Quindi in totale abbiamo 3 vincoli (media, deviazione standard, numero totale del campione) e 8 variabili (le nostre classi di dati). Il numero di gradi di libertà, cioè il numero di variabili davvero indipendenti in questo sistema, è  $d=8-3=5$ .

Ora, come facciamo a capire quanto bene si adattano questi dati ad una curva gaussiana? Vediamo che ci sono alcune discrepanze tra i dati attesi ed i dati osservati: nello specifico, la differenza è alta negli intervalli 2, 4, 5 e 7, ma è molto bassa, praticamente nulla negli altri intervalli. Come facciamo a dire che i dati trovati si distribuiscono davvero secondo la distribuzione gaussiana?

Ci può venire in aiuto il chi quadro. Infatti, abbiamo visto che nella sua forma generale questo è esprimibile come  $\sum \left( \frac{\text{valore osservato} - \text{valore atteso}}{\text{deviazione standard}} \right)^2$ , e la somma è fatta su tutte le variabili (in questo caso abbiamo convenuto di considerare i nostri intervalli come variabili).

Possiamo calcolare il valore del chi quadro per il nostro problema. Voi non lo sapete, ma considerando il nostro esperimento come un esperimento di conteggio la deviazione standard della variabile “intervallo” diventa esattamente uguale alla radice quadrata del valore atteso. Non è importante che voi sappiate queste cose, questo esempio serve solo a chiarirvi alcuni concetti nella pratica. Comunque, calcolando il chi quadro per questo esempio otteniamo:

$$\chi^2 = 17,5$$

Ora, ottenere un valore di chi quadro di 17,5 che informazioni ci può dare? Abbiamo visto prima che possiamo definire, dato un valore di chi quadro e il numero di gradi di libertà, la probabilità di ottenere un valore di chi quadro uguale o superiore al valore osservato. È abbastanza intuitivo capire che se questo valore sarà inferiore del 5% il chi quadro sarà significativamente improbabile da osservare; dunque, poiché il chi quadro in questione derivava dal confronto tra i dati osservati e la distribuzione ipotizzata di questi dati, è evidente che una probabilità per il chi quadro minore del 5% significa che la distribuzione ipotizzata non era corretta.

Notiamo che avendo il chi quadro una distribuzione che dipende solo dal numero di gradi di libertà (che prima avevamo detto essere  $n$  per semplicità, ma ora chiameremo più convenientemente “ $d$ ”) i suoi valori sono tabulabili. Abbiamo infatti una tabella che lega il valore di probabilità per il chi quadro al valore della variabile chi quadro e al numero di gradi di libertà. Quella che vi ha dato il Duca è una tabella piuttosto scomoda per gli scopi pratici, ma ora vi spiegherò come applicarla.

TABLE IV. DISTRIBUTION OF  $\chi^2$

Probability.

$\chi^2_{n,p} = \frac{n \cdot s^2}{\sigma^2}$

$n$	.99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01	.001
1	.0157	.0268	.0393	.0548	.0742	.1048	.455	1.074	1.642	2.706	3.841	5.412	6.635	10.827
2	.0201	.0404	.0503	.0638	.0846	.1148	1.386	2.408	3.219	4.605	5.991	7.879	9.210	13.815
3	.115	.185	.235	.284	.344	.414	1.213	2.366	3.078	4.642	6.251	7.879	9.348	16.266
4	.297	.429	.511	.584	.675	.776	1.064	1.928	2.706	3.745	5.191	6.958	8.541	15.491
5	.554	.752	.872	1.064	1.236	1.486	1.676	2.366	3.000	3.838	4.779	5.991	7.378	12.838
6	.872	1.134	1.312	1.501	1.700	1.913	2.204	2.878	3.457	4.351	5.299	6.388	7.879	13.368
7	1.239	1.564	1.772	2.000	2.236	2.490	2.833	3.438	4.015	4.779	5.618	6.625	8.021	14.168
8	1.646	2.032	2.278	2.537	2.813	3.107	3.428	4.015	4.578	5.318	6.168	7.172	8.328	14.838
9	2.088	2.532	2.797	3.078	3.374	3.686	4.015	4.578	5.138	5.878	6.758	7.779	8.938	15.491
10	2.558	3.059	3.337	3.633	3.944	4.270	4.613	5.176	5.735	6.476	7.378	8.397	9.591	16.151
11	3.053	3.609	3.900	4.213	4.539	4.878	5.236	5.795	6.354	7.100	8.015	9.038	10.241	16.811
12	3.571	4.178	4.482	4.813	5.150	5.499	5.866	6.425	6.984	7.730	8.645	9.668	10.811	17.471
13	4.107	4.765	5.078	5.423	5.772	6.126	6.493	7.052	7.611	8.357	9.272	10.295	11.454	18.131
14	4.660	5.368	5.692	6.050	6.413	6.780	7.150	7.709	8.268	9.014	9.929	10.952	12.097	18.791
15	5.229	5.985	6.320	6.690	7.063	7.440	7.819	8.378	8.937	9.683	10.598	11.621	12.744	19.451
16	5.812	6.614	6.962	7.343	7.726	8.113	8.493	9.052	9.611	10.357	11.272	12.305	13.407	20.111

Nel corpo della tabella sono riportati i valori della variabile chi quadro, ragion per cui potete leggere la tabella in questo modo:

→ Avete il valore della variabile chi quadro e il numero di gradi di libertà: supponiamo che essi siano rispettivamente 2 e 7. Allora andrete a trovarvi il valore più vicino a 2 sulla riga  $n=7$ . Vediamo che tale valore è 2.167; a questo valore corrisponde una probabilità dello 0.95. Cosa significa? Significa che dati 7 gradi di libertà la probabilità di osservare un chi quadro maggiore o uguale di 2.167 è del 95%.

Nel nostro esempio avevamo un chi quadro di 17,5 con 5 gradi di libertà; posizionandoci sulla riga  $n=5$  vediamo che il valore di 17,5 sta tra 15.086 e 20.515, cui corrispondono probabilità rispettivamente di 0,01 e 0,001. Poiché la probabilità di osservare un chi quadro di 17,5 con 5 gradi di libertà è situata tra l'1% e lo 0,1%, la nostra variabile “altezza” non si distribuirà come una variabile gaussiana come ci aspettavamo.

Notate in questo senso che mentre la probabilità binomiale ci aiutava a verificare quanto improbabile/probabile fosse l'ipotesi nulla dato un certo numero di eventi osservato, il test del chi quadro ci aiuta a capire quanto improbabile è l'ipotesi che i dati siano distribuiti secondo una certa distribuzione a partire dai dati stessi.

Entrambi sono test di verifica di ipotesi, ma dobbiamo imparare ad usarli nel contesto appropriato.

Capisco che tutta questa discussione sul chi quadro vi possa essere sembrata molto difficile. Non è importante che abbiate capito tutto dell'esempio del test del chi quadro, ma solo che ne abbiate capito il senso generale: è un test che serve a valutare la probabilità che un'ipotesi di distribuzione fatta a partire da dei dati sia corretta.

Vi consiglio comunque di rilegervi ora tutto il paragrafo da capo; alcune cose che non vi erano chiare all'inizio potrebbero essere molto più facili da capire ora. Se ancora non avete capito o avete dei dubbi, potete tranquillamente chiedermi delucidazioni.

## 7.1 – Campionamento (suppongo sia facoltativo per quelli del primo – potete balzare credo)

Tutto ciò che abbiamo visto fin qui ha senso se lo usiamo per ricavare dei risultati, cioè per fare quella che si chiama inferenza. Nella sua accezione più generale, inferire significa trarre conclusioni a partire da premesse: partendo da una certa affermazione A che si sa essere vera, si arriva a concludere che è vera anche una certa affermazione B, e da questa ad una C e così via. In statistica, inferire significa INDURRE verità su una popolazione partendo dall'osservazione di un suo sottoinsieme, detto campione. Definiremo più avanti questi termini.

Spero che sia chiaro il concetto di induzione, ma nel caso non lo fosse ve lo spiego in poche parole: indurre significa passare dal particolare al generale. È il contrario di dedurre, che significa passare dal generale al particolare.

→ Per gli amanti del genere: avete mai notato che in tutte le opere riguardanti Sherlock Holmes si parla di deduzione? Ecco, è un errore formale dovuto al fatto che Conan Doyle non sapesse probabilmente la differenza tra i vari tipi di ragionamento. Sherlock Holmes non deduce praticamente mai (e non servirebbe neanche, dato che nella deduzione le conseguenze sono solo esplicitate, essendo già contenute nelle premesse; dunque la deduzione non fornisce nuove conoscenze), ma piuttosto abduce. Al tempo di Conan Doyle le teorie di quel razzista di Peirce sulle nostre modalità di ragionamento iniziavano ad essere appena conosciute, pertanto è probabile che l'autore di Sherlock Holmes non le conoscesse e continuasse a fare tale errore. Se vi interessa una spiegazione più esaustiva guardate [qui](#).

Ora diamo un po' di definizioni:

- Popolazione: è l'insieme esaustivo delle unità di osservazione a cui si intende estendere l'inferenza.
  - Essendo un insieme molto numeroso (può anche essere l'intera popolazione mondiale) è spesso inaccessibile nella sua totalità – ciò significa che non possiamo analizzare ogni singolo individuo nella popolazione per ragioni pratiche. Di solito non conosciamo i parametri della popolazione, come media, varianza ecc, e pertanto dobbiamo estrarne un campione che ci servirà per stimare tali parametri.
- Campione: è un sottoinsieme della popolazione estratto da essa con una procedura probabilistica.
  - È un insieme di numerosità ridotta che è utilizzato per ricavare statistiche che servano da stimatori dei parametri di interesse.
  - Nel campionamento semplice casuale si tratta di fare un esperimento stocastico, in cui ogni elemento della popolazione ha la stessa probabilità di essere estratto (“campionato”).

Nelle definizioni avrete notato che ci sono tre termini che forse vi potranno essere oscuri (e che il Duca dà per scontati): parametri, statistiche, stimatori. Li definisco brevemente:

- Parametri: sono caratteristiche in un certo senso costanti di una popolazione – media, varianza ecc non sono variabili casuali: una volta definite, rimangono costanti. Tuttavia non sono considerati del tutto costanti, in quanto possono variare al variare della popolazione. Hanno uno status intermedio tra variabile e costante.
- Statistiche: è un concetto un po' vago – potete intenderle come dati ricavati da altri dati applicando i metodi della statistica. La differenza con i parametri è in un certo senso operativa: la media è sia un parametro che una statistica – se pensate alla media come ad una proprietà della popolazione, la state intendendo come parametro, mentre se pensate alla media come ad un qualcosa che vi calcolate dai vostri dati la state intendendo come statistica.
- Stimatori: uno stimatore è una “funzione” che associa ad ogni possibile campione estraibile da una popolazione il valore del corrispondente parametro da stimare, e tale valore è detto appunto “stima”. Si tratta in sostanza di una variabile casuale che varia al variare del campione.
  - Se non avete capito bene la definizione: immaginate di avere una popolazione di tre

persone, alte 160, 170 e 180 cm. Se prendete un campione costituito dalle prime due persone (campione 1) la media sarà 165; un campione costituito dalle seconde due avrà media 175; un campione costituito da prima e terza avrà media 170. Questi valori sono però stime della media reale (che è la media su tutta la popolazione, cioè 170), e ciò che ricaviamo facendo di volta in volta la media del campione è uno stimatore della media della popolazione, che associa ad ogni campione il valore della sua media.

- Uno stimatore utile si suppone abbia le seguenti proprietà: accuratezza, precisione e consistenza [in realtà ci sarebbero altre proprietà, ma il Duca vi mette queste e non voglio complicarvi le cose; almeno questi termini già li conoscete – per la consistenza vedete la nota 9].

Il campionamento casuale può essere effettuato attraverso l'utilizzo di numeri random (casuali). Non serve perdere molto tempo a spiegare come funzionano: un programma vi genera una tabella con le cifre da 0 a 9 disposte in una tabella, e voi potete leggere come vi pare queste cifre per trovare il modo di estrarre un campione dalla popolazione.

→ Esempio veloce: avete una popolazione di 500 persone, da cui volete ricavare un campione di 10 persone. Ovviamente le 500 persone sono numerate da 000 a 499. Utilizzate la tabella dei numeri casuali prendendo le prime 30 cifre di una colonna a caso e raggruppandole a 3 a 3: otterrete numeri come 495, 786, 244, ecc. Siccome i numeri che potete estrarre così vanno da 000 a 999 ma la vostra popolazione va al massimo fino a 499, togliete 500 da ogni numero maggiore 499: 786 diventerà così 286, mentre 244 rimane 244. A questo punto prendete dalla vostra popolazione le persone corrispondenti ai numeri estratti: et voilà, avete il vostro campione, divertitevi.

Dal campione possiamo calcolare le stime di media e varianza: le chiameremo “m” e “s<sup>2</sup>” invece che “μ” e “σ<sup>2</sup>” per rimarcare il fatto che in questo caso sono stimatori che si riferiscono al campione e non alla popolazione, e le chiameremo “media campionaria” e “varianza campionaria”.

La media campionaria “m” si calcola come la media del campione.

La varianza campionaria “s<sup>2</sup>” si calcola invece come 
$$\frac{\sum (x_i - m)^2}{n-1}$$
.

→ Vi chiederete: come mai la varianza al denominatore ha n-1 invece che n?

Ragioniamo: in questo caso abbiamo utilizzato gli scarti quadratici dalla media per calcolare la varianza, ma abbiamo utilizzato come media la media campionaria, cioè quella calcolata a partire dai dati stessi. Questo significa che l'ultimo scarto è già noto [andate a rivedervi il discorso sui gradi di libertà] e di conseguenza i gradi di libertà sono n-1 (l'ultimo scarto è dipendente da tutti quelli precedenti, che sono tra di loro indipendenti).

Per rendervene conto, considerate il caso di n=1: in questo caso la varianza calcolata come conosciamo, con n al denominatore, sarebbe 0; calcolata con n-1 sarebbe invece uguale a 0/0, che è indefinito. È ovvio che se abbiamo un solo valore la varianza non può essere 0, dato che la dispersione non è nota, ma sicuramente non 0; è pertanto chiaro che con la seconda definizione, con n-1 al denominatore, ci dà un'informazione più corretta: avendo un solo elemento, la varianza non è definita.

→ Vi chiederete: ok, ma allora perché usare n invece che n-1? In fondo è sempre così, anche per la popolazione vale lo stesso ragionamento. Vero; il fatto è che quando si valutano le popolazioni n è molto grande, e la differenza tra n e n-1 è spesso irrilevante, quindi tanto vale approssimare.

→ Questa varianza campionaria è uno stimatore “unbiased”: dividere per n introdurrebbe infatti un errore sistematico per le considerazioni già viste sopra e rese evidenti dal caso limite. Dividendo per i gradi di libertà (o “informazioni indipendenti”) si evita tale errore.

→ La varianza campionaria non è il descrittore della varianza del campione, ma lo stimatore della varianza nella popolazione.

Abbiamo detto che gli stimatori (media e varianza) sono delle funzioni del campione: sono cioè delle variabili casuali. Possiamo pertanto costruire un grafico di distribuzione degli stimatori per un dato  $n$  del campione.

→ Il procedimento è semplice: poniamo di avere  $n=10$  su una popolazione di 20 persone. Quanti campioni possiamo avere? Ovviamente si tratta di trovare le combinazioni semplici di classe 10 di un insieme di 20 elementi, pertanto avremo  $20!/(10! \times 10!) = 184756$  campioni. Otterremo che molti di questi campioni avranno una media molto vicina a quella della popolazione, e pochi di questi campioni avranno una media lontana da quella della popolazione: pertanto possiamo porre sull'asse delle  $x$  la media campionaria e sull'asse delle  $y$  il numero di campioni. Otterremo una curva centrata più o meno sulla media della popolazione e più o meno campaniforme.

Possiamo calcolare il valore atteso delle nostre variabili casuali  $m$  e  $s^2$ .

Otteniamo:

- $E(m) = \mu$
- $E(s^2) = \sigma^2$

→ la dimostrazione è un po' [sbatti](#), pertanto più che una dimostrazione vi faccio vedere con un esempio semplice che è proprio così.

Valori della popolazione: 2, 3, 5, 6.

Media della popolazione: 4.

Varianza della popolazione: 2.5.

Campioni possibili da  $n=3$ : [2,3,5] [3,5,6] [2,5,6] [2,3,6].

Medie e varianze dei quattro campioni: [3.33; 2.32] [4.66; 2.32] [4.33; 4.32] [3.66; 4.32].

Media delle medie campionarie: 4.

Media delle varianze campionarie: 3.33.

→ Sorpresa! La media delle varianze campionarie non è uguale alla varianza della popolazione! Beh beh beh beh. Ovvio. Infatti la varianza della popolazione l'ho calcolata usando  $n$  come numeratore (in quel caso 4) mentre la varianza campionaria l'ho calcolata usando  $n-1$ . Se calcolassi la varianza della popolazione con  $n-1$  al denominatore i due valori coinciderebbero.

Il fatto è che qui avevo una popolazione molto piccola, per cui i due modi di calcolare la varianza portano a risultati molto diversi; nella normalità le popolazioni sono grandi, e la media delle varianze campionarie (con  $n-1$  al denominatore) è più o meno coincidente con la varianza della popolazione (calcolata, anche se inaccuratamente, con  $n$  al denominatore).

Poiché i valori attesi degli stimatori in questione sono pari ai parametri della popolazione tali stimatori sono accurati (o "unbiased").

Ma sono anche [consistenti](#)? Perché uno stimatore sia consistente al crescere della dimensione del campione il suo valore atteso deve convergere sul parametro di cui è stima.

Non vi dimostro per ora i risultati seguenti (alcune cose le trovate nel file sbatti di cui sopra), quindi prendeteli per buoni:

- Per  $n$  che tende ad infinito la distribuzione di  $m$  tende ad avere forma gaussiana, con:
  - Media di  $m$  pari a  $\mu$
  - Deviazione standard di  $m$  pari a  $\frac{\sigma}{\sqrt{n}}$ 
    - Se ricordate, questo era stato definito come l'errore standard della media. Considerate che al crescere di  $n$  questo valore diventa sempre più piccolo, pertanto la "campana" è sempre più "appuntita". In termini più rigorosi, lo stimatore " $m$ " è consistente, in quanto converge su  $\mu$  all'aumentare di  $n$ .
    - Inoltre, all'aumentare di  $n$  lo stimatore diventa anche sempre più preciso, perché diminuisce la deviazione standard dello stimatore.
- Se la popolazione campionata è gaussiana la distribuzione di  $s^2$  tende ad essere proporzionale ad una variabile casuale chi quadro a  $n-1$  gradi di libertà. In particolare, si dimostra che il rapporto fra devianza campionaria e varianza della popolazione assume una

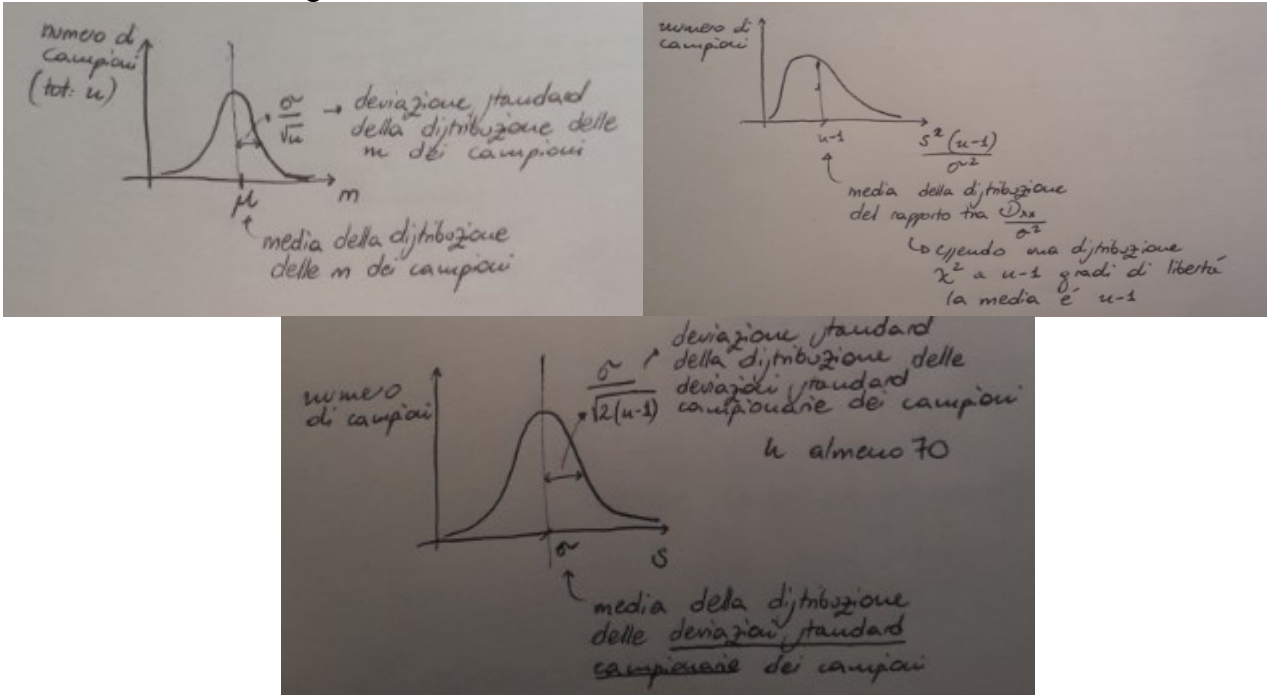
distribuzione chi quadro a  $n-1$  gradi di libertà:  $\frac{D_{xx}}{\sigma^2} = \chi_{n-1}^2$ .

Considerate che la devianza campionaria è ovviamente calcolata come  $D_{xx} = s(n-1)$ , dato che  $s = \frac{\sum (x_i - m)}{n-1}$  e che il numeratore è proprio la devianza del campione.

Inoltre, per  $n$  almeno pari a 70, si dimostra che:

- La media di  $s$  è pari a  $\sigma$
- La deviazione standard di  $s$  è pari a  $\frac{\sigma}{\sqrt{2(n-1)}}$ 
  - Quest'ultimo valore è definito "errore standard della deviazione standard campionaria", e all'aumentare di  $n$  tende molto più velocemente a 0 dell'errore standard della media. In questo senso, lo stimatore della varianza è più consistente (o "robusto") dello stimatore della media.

Può esservi utile vedere graficamente cosa si intende con distribuzione di  $m$  e di  $s^2$ .



Veniamo ora a considerazioni più pratiche. Supponiamo di aver estratto il nostro campione dalla popolazione. Il nostro campione avrà una media campionaria  $m$  ed una varianza campionaria  $s^2$ , mentre la popolazione da cui è stato estratto avrà media  $\mu$  e varianza  $\sigma^2$ . Vogliamo sapere qual è la probabilità di estrarre un campione con quella media e con quella varianza.

→ Detto così sembra un quesito stupido: in fondo se abbiamo la nostra popolazione ed estraiamo un certo campione, per quanto improbabile, perché dovrebbe importarci di sapere quanto è probabile averlo estratto? Il fatto è che l'applicazione reale è meglio inquadrabile se pensiamo a questo: supponiamo di avere una macchina che fa delle bottiglie di birra da 33 cc. Questa macchina opera quindi teoricamente sfornando una popolazione di bottiglie con media 33 cc e deviazione standard tot (facciamo 1 cc). Vogliamo vedere se questa macchina opera bene o se è il caso di cambiarla: estraiamo un campione di un certo numero di bottiglie e le valutiamo. Supponiamo di trovare che questo campione ha una media di 34 cc e una deviazione standard campionaria di 1,5 cc. Sarà il caso di cambiare la macchina? Il ragionamento alla base è questo: se è significativamente improbabile aver estratto un campione del genere (cioè se la probabilità è minore del 5% o dell'1% a seconda del livello di significatività statistica che stiamo usando) allora forse è il caso di cambiare la macchina.

Ora, sapendo che le medie campionarie dei campioni si distribuisce secondo una curva gaussiana

centrata sulla media  $\mu$  della popolazione e con deviazione standard  $\frac{\sigma}{\sqrt{n}}$ , possiamo usare tutto l'armamentario che già conosciamo dalla gaussiana per calcolare la probabilità di ottenere un valore di media del campione maggiore o uguale a "m" data una popolazione con media  $\mu$  e deviazione standard  $\sigma$ .

Dalla gaussiana, se ricordate, potevamo calcolare il valore z (per utilizzare le informazioni date dalla gaussiana standardizzata sulle probabilità), la cui forma generale è:  $\frac{(\text{valore ottenuto} - \text{media})}{\text{deviazione standard}}$ , ovvero in questo caso (gaussiana delle medie dei campioni) è  $z = \frac{|m - \mu|}{\frac{\sigma}{\sqrt{n}}}$ .

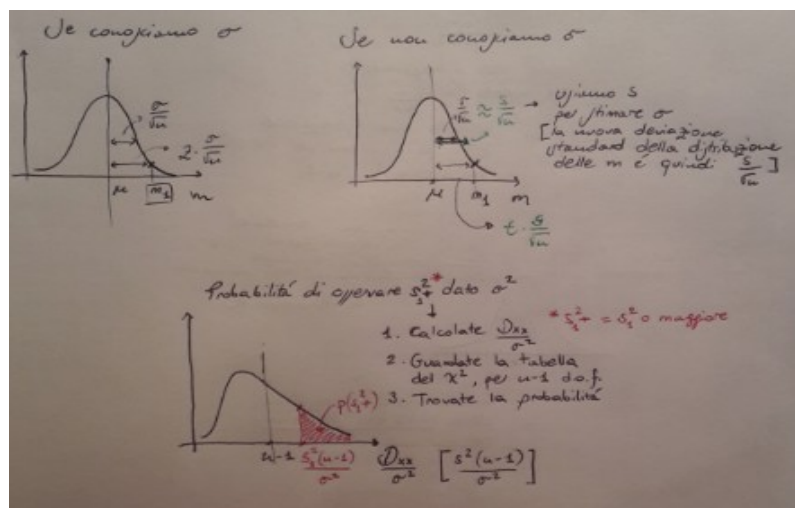
→ Ho messo il valore assoluto in questo caso specifico perché più che sapere quanto è probabile ottenere un valore maggiore o uguale di m ci interessa sapere quanto è probabile ottenere un valore che dista da  $\mu$  almeno come dista m (se m fosse minore di  $\mu$  ci interesserebbe trovare la probabilità di ottenere un valore minore o uguale a m). Essendo la gaussiana simmetrica, la probabilità di ottenere un valore minore di m con  $m < \mu$  è uguale alla probabilità di ottenere un valore maggiore di m con  $m > \mu$ , e dunque tanto vale prendere z come positivo utilizzando il valore assoluto nella differenza al numeratore (ovviamente le due z ottenute senza questo accorgimento sarebbero opposte).

Una volta trovato questo z (che indica quante deviazioni standard -  $\frac{\sigma}{\sqrt{n}}$  - m dista da  $\mu$ ) andiamo a vedere sulle tabelle per la gaussiana standardizzata [era ciò che indicavo come "t" nel paragrafo sulla gaussiana] qual è la probabilità corrispondente.

Se non conosciamo  $\sigma$  della popolazione dovremo usare uno stimatore della deviazione standard della popolazione, ovvero s. In questo caso la z prima ricavata verrà riscritta come  $t = \frac{|m - \mu|}{\frac{s}{\sqrt{n}}}$ ,

dove ho scritto t invece che z perché si tratta proprio di una distribuzione t di Student, con n-1 gradi di libertà (andatevi a rivedere la sezione apposita). Anche in questo caso calcolerò la t e andrò a verificare la probabilità corrispondente di ottenere un valore di t maggiore o uguale a quello trovato. Posso valutare infine la probabilità di ottenere un valore di varianza campionaria pari a s sapendo che la popolazione da cui ho estratto il campione ha varianza  $\sigma$ . In questo caso mi basta ricordare che il rapporto  $\frac{D_{xx}}{\sigma^2}$ , ovvero  $\frac{s^2(n-1)}{\sigma^2}$ , si distribuisce come un chi quadro a n-1 gradi di libertà. La probabilità di trovare una varianza campionaria pari o superiore a  $s^2$  sarà allora ricavabile come la probabilità di ottenere un  $\chi^2 \geq \frac{s^2(n-1)}{\sigma^2}$  con n-1 gradi di libertà.

Per ricapitolare:



Il Duca vi fa degli esempi con dei dati concreti, ma li cambia ogni anno quindi ad un certo punto chissene dei suoi dati. Vi metto io un po' di dati che rendono le cose più semplici.

Sarò molto schematico; supponete di avere:

- Popolazione con media 10 e varianza 4 [quindi deviazione standard 2!]
- Campione "A" estratto di 9 elementi con media 12 e varianza 1
- Campione "B" estratto di 9 elementi con media 9 e varianza 9

Ora facciamo i vari calcoli:

- Campione "A":
  - $z = \frac{(12-10)}{\frac{2}{\sqrt{9}}} = 3 \rightarrow p(z \geq 3) = 0,0013$
  - $t = \frac{(12-10)}{\frac{1}{\sqrt{9}}} = 6 \rightarrow p(t \geq 6) = 0,0002$  [con 8 gradi di libertà]
  - $\chi^2 = \frac{D_{xx}}{\sigma^2} = 1 \frac{(9-1)}{4} = 2 \rightarrow p(\chi^2 \geq 2) = 0,981$  [con 8 gradi di libertà]
- Campione "B":
  - $z = \frac{(10-9)}{\frac{2}{\sqrt{9}}} = 1,5 \rightarrow p(z \geq 1,5) = 0,06681$
  - $t = \frac{(10-9)}{\frac{3}{\sqrt{9}}} = 1 \rightarrow p(t \geq 1) = 0,1733$  [con 8 gradi di libertà]
  - $\chi^2 = \frac{D_{xx}}{\sigma^2} = 9 \frac{(9-1)}{4} = 18 \rightarrow p(\chi^2 \geq 18) = 0,0212$  [con 8 gradi di libertà]

Ho calcolato tutti i valori in [questo sito](#), imbarazzantemente facile da usare. Anche [questo](#) è buono. Come possiamo evincere da questi dati?

- Campione "A" → estrarre un campione con questa media è molto improbabile (sia con gaussiana che con Student la probabilità è sotto il 5% di significatività statistica), mentre è praticamente certo estrarre un campione con questa varianza (è probabile al 98%!!).
- Campione "B" → estrarre un campione con questa media è improbabile, ma non significativamente improbabile (infatti entrambe le statistiche sono superiori al 5%), mentre è significativamente improbabile estrarre un campione con tale varianza (è sotto il 5%).
- In generale → lo stimatore della media e lo stimatore della varianza sono indipendenti (come avevamo già affermato in precedenza): posso sottostimare o sovrastimare la varianza indipendentemente da come stimiamo la media e viceversa. Per quanto riguarda le differenze tra t di Student e gaussiana, queste sono a volte anche ampie (7 volte nel primo caso, 3 nel secondo) e dipendono da come sottostimo o sovrastimo la varianza attraverso la statistica s.
  - Tra parentesi, (questo è il motivo per cui uso i gradi di libertà per calcolare lo stimatore della varianza: in questo modo mi assicuro che la stima di varianza e la stima della media siano indipendenti)

Il Duca mette due esercizi ma sono semplici, non aggiungono nulla e non ho tanta voglia di farli. Probabilmente a fine dispense metterò un po' di esercizi svolti divisi per tipologia.

## 7.2 – Variabili categoriche dicotomiche

Finora abbiamo visto come applicare i modelli statistici al campionamento con una variabile casuale continua. Ma questo è solo un caso particolare. Le variabili categoriche dicotomiche sono

quelle che hanno solo due possibilità: per esempio guarito o non guarito, morto o vivo, maschio o femmina. In una popolazione possiamo avere una distribuzione di una variabile dicotomica, ma essendo appunto una variabile che ha due sole possibilità più che di distribuzione parleremo di proporzione. Per esempio, in una popolazione potrà esserci il 30% di fumatori: 30% è la proporzione della variabile dicotomica “fumatore/non fumatore”. Da questa popolazione possiamo estrarre un campione, e porci la stessa domanda che ci siamo posti prima: quanto è probabile ottenere un campione con quella proporzione di fumatori, data la proporzione di fumatori nella popolazione campionata?

Con una definizione più formale, possiamo definire la proporzione  $\pi$  come un parametro che esprime la percentuale di unità della popolazione campionata caratterizzate da un valore di una variabile nominale dicotomica.

→ “percentuale di unità della popolazione campionata” = percentuale della popolazione

→ “caratterizzate da un valore” = che fumano (o che non fumano, a seconda di ciò che ci interessa)

→ “di una variabile nominale dicotomica” = la variabile “fumatore/non fumatore” non è numerica, ma nominale

Estrarre un campione da una popolazione caratterizzata da una certa proporzione di una variabile dicotomica, osservare la proporzione di quella variabile nel campione e chiederci quanto è probabile fare tale osservazione date le caratteristiche della popolazione di origine equivale a chiederci qual è la probabilità di ottenere una certa percentuale di successi (la proporzione nel campione) in un certo numero di tentativi (la dimensione del campione) data una certa percentuale di successo (la proporzione nella popolazione). Ciò ci indica che il nostro modello di distribuzione di probabilità sarà quello binomiale, e possiamo usare tutto ciò che abbiamo visto quando ne abbiamo parlato per risolvere i nostri problemi (cfr. 6.2).

→ Andate davvero a rivedervi il paragrafo 6.2

→ No, sul serio, andate

→ Muovetevi!! Cosa state ancora qui a leggere?

→ Va be fate un po' come vi pare.

→ Eddai su, ve lo chiedo per favore, poi non capite!

→ Ok, come volete... Ci tornerete su dopo.

→ Però se non capite ci rimango male...

→ Ma la scelta è vostra in fondo...

→ Se non avete sbatti...

→ Proprio zero eh?

→ Va be

Sappiamo già che la media della nostra distribuzione binomiale (cioè la distribuzione dei  $k$  “successi” in  $n$  prove) è  $n\pi$ , mentre la deviazione standard è  $\sqrt{n\pi(1-\pi)}$ .

Se volessimo calcolare tutto in termini di proporzioni dovremmo dividere i nostri  $k$  per  $n$  (un esempio numerico: 3 successi in 10 prove è uguale ad una percentuale di successi di 0.3). Di conseguenza, utilizzando le proporzioni la media e la deviazione standard dovrebbero essere divise

per  $n$  e otterrei  $\mu=\pi$  e  $\sigma=\sqrt{\frac{\pi(1-\pi)}{n}}$  [fatevi due calcoli per la deviazione, quando portare  $n$  sotto radice quadrata al denominatore viene elevato al quadrato!].

Ricordiamo anche che quando i numeri in gioco sono grandi la distribuzione binomiale si approssima ad una gaussiana con media  $\pi$  e deviazione standard  $\sqrt{n\pi(1-\pi)}$ . In particolare, l'approssimazione gaussiana è buona quando  $0,2 < \pi < 0,8$  (cosa che garantisce che la distribuzione binomiale non rimanga troppo asimmetrica) e  $n\pi > 5$  e  $n(1-\pi) > 5$ .

Quando applichiamo l'approssimazione gaussiana passiamo tuttavia dal discreto (numero di successi o proporzione – finita – di successi) al continuo, e in questo caso per calcolare la probabilità di un valore (per esempio 3 successi in 10 prove) dobbiamo applicare la correzione per la continuità, che consiste nel dividere la nostra variabile “successi” in classi.

Per fare un esempio concreto: nel modello binomiale la probabilità di ottenere 0 successi è  $p(k=0)$ ;

nel modello gaussiano la probabilità di ottenere 0 successi è  $p(k < 0,5)$ . Analogamente, la probabilità di ottenere 1 successo nella binomiale è  $p(k=1)$ , mentre nel modello gaussiano è la probabilità della classe che corrisponde ad un successo, ovvero  $p(0,5 < k < 1,5)$ .

Applicando l'approssimazione gaussiana e calcolando  $z$  [questa volta come  $\frac{(k-n\pi)}{\sqrt{n\pi(1-\pi)}}$  ] possiamo calcolare la probabilità di ottenere un numero di successi minore o uguale a  $k$ , o maggiore o uguale a  $k$ , esattamente come abbiamo sempre fatto.

Facciamo un veloce esempio numerico:

“Qual è la probabilità di estrarre da una popolazione con il 30% di fumatori un campione di 60 soggetti di cui meno di 15 siano fumatori?”

→ 60 soggetti è un numero abbastanza grande per applicare l'approssimazione gaussiana. Quale sarebbe la media di fumatori nel campione sulla base del modello? Sarebbe  $n\pi$ , ovvero  $60 \times 0,3 = 18$ .

Poiché stiamo usando l'approssimazione gaussiana non ci chiediamo la  $p(k \leq 14)$ , ma  $p(k \leq 14,5)$  [abbiamo infatti diviso in classi i nostri valori, perché la gaussiana si basa su una variabile continua e non discreta].

Possiamo ora calcolare il nostro  $z$ , ovvero il numero di deviazioni standard per cui 14,5 dista da 18, come  $\frac{(k-n\pi)}{\sqrt{n\pi(1-\pi)}}$ ; quindi avremo  $z = \frac{(14,5-18)}{\sqrt{60 \cdot 0,3 \cdot 0,7}} = -0,99$ . La probabilità di ottenere uno  $z \leq -0,99$  è 0,161, ricavabile dalle tavole.

Dunque la probabilità di estrarre un campione del genere è del 16,1%.

→ Se avessimo usato le proporzioni non sarebbe cambiato nulla: avremmo dovuto dividere tutto per  $n$  del campione (cioè 60). Avremmo ottenuto

$$z = \frac{(0,2417-0,3)}{\sqrt{\frac{0,3 \cdot 0,7}{60}}} = -0,99 \quad [\text{ricordate che dovevate dividere il denominatore per}$$

60 FUORI dalla radice quadrata, quando lo mettete dentro la radice dovete fare il quadrato, e a quel punto  $60/3600$  è uguale a  $1/60$ ].

## 8.1 – Brevi richiami di logica

Senza alcuna pretesa di esaustività, vi scrivo qui di seguito l'abbozzatissima epistemologia tracciata dal Duca.

L'inferenza abbiamo già visto cosa sia. Nel nostro caso specifico possiamo intenderla come l'atto del saggiare ipotesi, attraverso opportuni test.

Per saggiare un'ipotesi bisogna appunto formularle, fare delle osservazioni e applicare un metodo per decidere se le osservazioni confermano o meno l'ipotesi fatta.

Nel tempo il modo in cui gli scienziati e i filosofi concepiscono il metodo scientifico è cambiato. Per [Bacone](#) (vissuto tra il 1500 e il 1600) il metodo scientifico si avvale dell'induzione, e consiste nell'osservazione attenta, senza pregiudizi, dei fenomeni, che genera teorie che spiegano le relazioni tra i fenomeni stessi, producendo conoscenza scientifica. In sostanza Bacone afferma che la nostra mente, quando compie osservazioni scientifiche, deve essere come una “tabula rasa”; questo tuttavia non è praticamente mai possibile.

Secondo Popper, vissuto nel '900, l'osservazione è sempre carica di teoria, e il compito dell'osservazione scientifica è quello di falsificare le teorie ipotizzate per spiegare i fenomeni. In tal senso un'ipotesi che non è falsificabile non è scientifica, e un'osservazione i cui esiti non possono falsificare un'ipotesi è inutile.

→ Ciò comporta che le teorie psicologiche non sono scientifiche, infatti non ci sono osservazioni che possano falsificarle: come posso per esempio falsificare l'ipotesi dell'inconscio di Freud?

→ Per farvi un esempio di osservazione inutile: l'osservazione di che tempo fa oggi non può falsificare l'ipotesi che 44 gatti si dispongano sempre in fila per tre col resto di due.

Secondo Popper noi possiamo fare ipotesi preconcepite, frutto di una fantasia ben educata, non dell'induzione, e l'osservazione può permettere solo di eliminare le ipotesi false, ma non di confermare quelle vere, che potranno essere sempre smentite da una singola osservazione contraria.

Per fare un esempio che applichi questo concetto:

“Avete una serie di carte; su ognuna di queste c'è una lettera su un lato e un simbolo sull'altro. L'ipotesi che volete testare è che se un lato c'è M sull'altro ci sia sempre +. Quali carte girereste per saggiare quest'ipotesi tra le seguenti? [M] [+] [m] [-]”

→ Ovviamente faremo solo le osservazioni che sono utili, cioè quelle in grado di falsificare la nostra ipotesi.

Se giriamo [M] e osserviamo qualcosa che non sia [+] l'ipotesi è falsificata, dunque tale osservazione è utile. In questo caso abbiamo applicato il cosiddetto [modus ponens](#): se l'ipotesi è vera, dietro a [M] deve esserci [+].

Se giriamo [+] e troviamo [m] o qualcosa di diverso da [M] l'ipotesi non è falsificata: infatti l'ipotesi non dice che dietro a [+] deve esserci per forza [M], ma solo che dietro a [M] ci deve essere per forza [+]. L'errore che si commette pensando che se [M] → [+] allora [+] → [M] (con “→” che significa “implica”) è una fallacia logica definita come “[affermazione del conseguente](#)”. Quindi non gireremo questa carta.

Se giriamo [m] e troviamo qualunque cosa non possiamo falsificare l'ipotesi, infatti l'ipotesi non ci dice nulla a riguardo. Supponendo che le alternative possibili siano solo M, m, +, -, con M che è il contrario di m e + che è il contrario di -, affermare che la negazione di [M] (cioè [m]) implica la negazione di [+] (cioè [-]) è un'altra fallacia logica, definita come “[negazione dell'antecedente](#)”. Non gireremo la carta neanche in questo caso.

Se infine giriamo [-] e ci troviamo dietro [M] l'ipotesi è falsificata, in quanto dietro a [M] doveva esserci sempre [+]. Ragionando in questo modo applichiamo il [modus tollens](#): equivale a dire “se è giorno c'è luce; ora non c'è luce, dunque non è giorno”, ovvero la negazione della conseguenza implica la negazione della premessa nel caso di un'ipotesi “forte” (in cui la premessa implica SEMPRE la conseguenza). Quindi gireremo questa carta.

## 8.2 – Saggiare ipotesi

Fin dal primo giorno un tacchino osservò che, nell'allevamento dove era stato portato, gli veniva dato il cibo alle 9 del mattino. E da buon induttivista non fu precipitoso nel trarre conclusioni dalle sue osservazioni e ne eseguì altre in una vasta gamma di circostanze: di mercoledì e di giovedì, nei giorni caldi e nei giorni freddi, sia che piovesse sia che splendesse il sole. Così arricchiva ogni giorno il suo elenco di una proposizione osservativa in condizioni più disparate. Finché la sua coscienza induttivista non fu soddisfatta ed elaborò un'inferenza induttiva come questa: "Mi danno il cibo alle 9 del mattino". Questa concezione si rivelò incontestabilmente falsa alla vigilia di Natale, quando, invece di venir nutrito, fu sgozzato.

Questa metafora, ideata da Bertrand Russell, spiega il cosiddetto induttivismo ingenuo: non si può dare per verificata un'ipotesi continuando ad accumulare osservazioni coerenti con essa – basta infatti una sola osservazione che la smentisca e l'ipotesi è, a volte tragicomicamente, falsa.

Cerchiamo di formalizzare il concetto. Supponiamo di aver fatto un certo numero di osservazioni di cigni bianchi, e di aver quindi indotto ingenuamente che tutti i cigni sono bianchi.

Qual è la probabilità di osservare un cigno nero data quest'ipotesi? Esattamente 0. In questo senso, l'osservazione di un evento dichiarato impossibile da una certa ipotesi è sufficiente per dichiarare con certezza falsa l'ipotesi (ovviamente, l'osservazione deve essere affidabile).

In termini formali: se  $P(\text{osservazione}|\text{ipotesi})=0$  allora "osservazione" → "ipotesi è falsa".

Tuttavia le ipotesi statistiche non sono praticamente mai ipotesi forti: ovvero, la probabilità di fare un'osservazione diversa dall'ipotesi non è nulla. Come ci comportiamo in questo caso? Come possiamo falsificare l'ipotesi? E come possiamo confermarla?

Partiamo dal primo problema: come possiamo falsificare l'ipotesi?

Teniamo presente che NESSUNA osservazione in questo caso potrà smentire con CERTEZZA l'ipotesi, dato che non è un'ipotesi universale. Se l'ipotesi è vera ci sarà sempre una certa probabilità, per quanto piccola, di fare ogni osservazione.

*Diremo che l'ipotesi è falsa (e quindi la rifiuteremo) quando sotto tale ipotesi la probabilità di fare l'osservazione fatta è minore del 5%.*

→ Un esempio vi farà capire il senso di quanto detto.

Supponiamo che la nostra ipotesi sia "F guarisce il 70% dei pazienti".

Scegliamo un campione di 10 soggetti malati: quali osservazioni ci permettono di falsificare l'ipotesi?

Il modello che descrive la probabilità di osservare un certo numero di guariti sul campione di malati è ovviamente quello binomiale, dato che un malato può solo guarire (successo) o non guarire (insuccesso), per il quale dobbiamo usare come probabilità di successo 0,7 (il 70% ipotizzato); la media di successi attesa su 10 tentativi sarà quindi 7 e la deviazione standard della distribuzione 1,449 ( $\sqrt{10 \cdot 0,7 \cdot 0,3}$ ).

Calcoliamo ora la probabilità di osservare k guariti su 10 malati se F ha efficacia di 0,7:

k guariti	Probabilità	Probabilità cumulativa
0	0,0000059049	0,0000059049
1	0,0001377810	0,0001436859
2	0,0014467005	0,0015903864
3	0,0090016920	0,0105920784
4	0,0367569090	0,0473489874
5	0,1029193452	0,1502683326
6	0,2001209490	0,3503892816
7	0,2668279320	0,6172172136
8	0,2334744405	0,8506916541
9	0,1210608210	0,9717524751
10	0,0282475249	1

Ora, se osservassimo 7 guariti su 10 (cioè il 70% di guarigioni) ovviamente non rifiuteremo l'ipotesi iniziale – la probabilità di fare quest'osservazione è del 26%, abbastanza alta. Ma se osservassimo 0 guarigioni su 10? La probabilità di fare un'osservazione del genere se è vera

l'ipotesi che F guarisca il 70% dei malati è dello 0,00059%, cioè è talmente poco probabile che se osservassimo questo rifiuteremmo l'ipotesi iniziale. Tuttavia, nello 0,00059% dei casi avremmo fatto un errore a rifiutarla! Non era un'osservazione impossibile da fare, era solo molto improbabile.

Ma fino a che punto diciamo che un'osservazione ci porta a rifiutare la nostra ipotesi? Vediamo che osservare 4 o meno guarigioni ha una probabilità totale del 4,73% (dovete guardare la cumulativa), sotto quindi il famoso 5% di significatività statistica. Tuttavia, all'estremo opposto anche osservare 10 guarigioni è sotto il 5% - ha infatti il 2,82% di probabilità.

Come dobbiamo considerare il 5% di significatività statistica? Abbiamo 3 opzioni:

- Possiamo considerare l'ipotesi iniziale smentita se facciamo un'osservazione INFERIORE all'ipotesi (cioè tale per cui la proporzione di guarigioni nel campione sia MINORE di 0,7) che abbia meno del 5% di probabilità di essere fatta;
- Possiamo considerare l'ipotesi iniziale smentita se osserviamo una proporzione MAGGIORE all'ipotesi che abbia meno del 5% di probabilità di essere osservata;
- Possiamo considerare l'ipotesi iniziale smentita se osserviamo una proporzione DIVERSA dall'ipotesi (maggiore o minore) che abbia in totale meno del 5% di probabilità di essere osservata → in questo caso quindi prenderemo il 95% delle probabilità centrali e considereremo l'ipotesi rifiutata se la nostra osservazione maggiore o minore dell'ipotesi ha meno del 2,5% di probabilità di essere fatta.

Vediamo cosa comporta ciò:

- Nel primo caso rifiuteremo l'ipotesi se osserviamo 4 o meno guarigioni;
- Nel secondo caso rifiuteremo l'ipotesi se osserviamo 10 guarigioni
- Nell'ultimo caso rifiuteremo l'ipotesi se osserviamo 3 o meno guarigioni (probabilità 1%) oppure 10 guarigioni (2,8% - se vogliamo fare i pignoli tecnicamente 2,8 non è minore di 2,5; ma è così vicino che praticamente la consideriamo come osservazione improbabile al 2,5%).

A rigore l'atteggiamento giusto dovrebbe essere il terzo – in questo caso si cerca la probabilità a due code, cioè si cerca il 5% delle probabilità “esterne” nella curva che descrive la distribuzione di probabilità delle osservazioni. Tuttavia, ad un medico potrebbe interessare di più tutelare il paziente, e quindi evitare di dire che un farmaco ha un'efficacia maggiore di quella che ha in realtà: in questo caso gli interesserà poter smentire l'ipotesi che F abbia un'efficacia così ALTA, e quindi assumerà il primo atteggiamento – cercherà quindi la probabilità ad una coda (nello specifico la sinistra).

Perché uno debba assumere il secondo approccio – probabilità ad una coda destra – in questo caso non è chiaro; ma in altri casi potrebbero valere le stesse considerazioni fatte per il medico di prima.

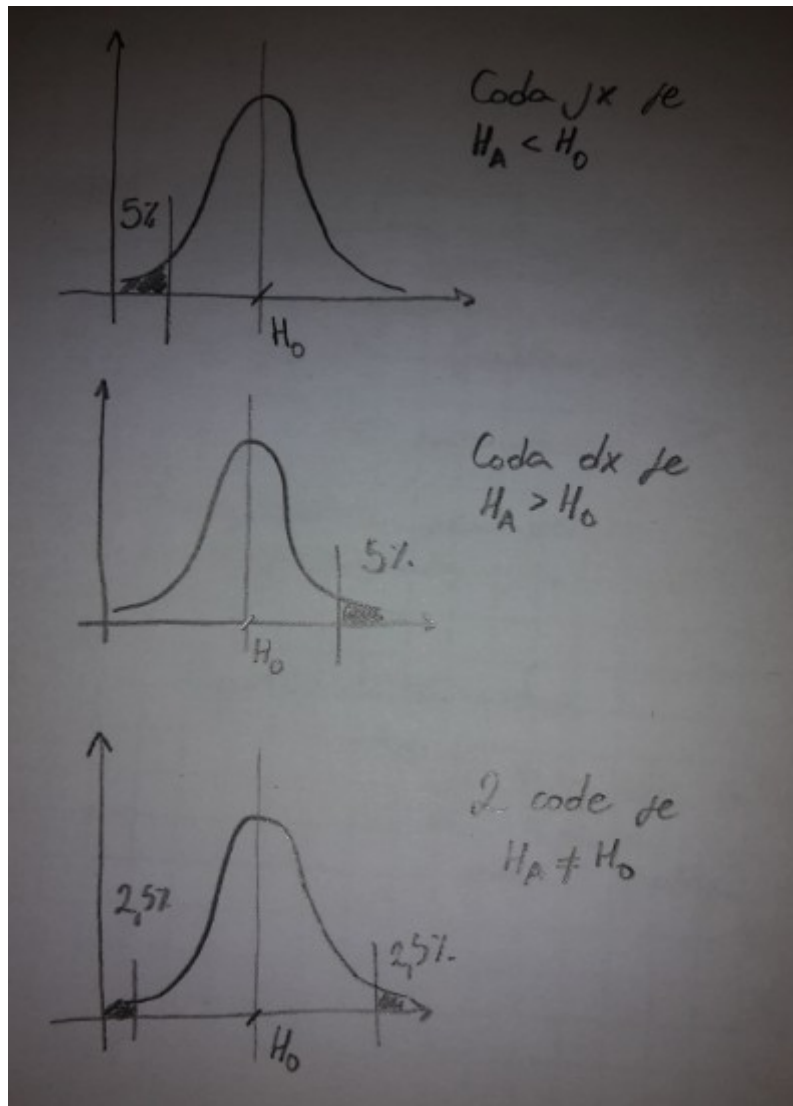
In sostanza, applicando il primo approccio considereremo l'ipotesi rifiutata se  $k < 5$  guarigioni.

È molto importante ricordare che non esiste alcuna osservazione campionaria che possa permettere di considerare falsa con certezza l'ipotesi. Avremo sempre una certa probabilità di fare un errore rifiutando l'ipotesi: ovviamente questo errore è pari alla probabilità di fare un'osservazione che ci porti a rifiutare l'ipotesi quando l'ipotesi è in realtà vera.

Chiameremo tale errore  $\alpha$  (o “errore del primo tipo”). Nel nostro caso (e praticamente sempre per come concepiamo il livello di significatività statistica)  $\alpha = 0,05$ .

Ancora una parola sulle “code”: noi scegliamo la regione di rifiuto (cioè la regione la cui area è il 5% nella distribuzione di probabilità delle osservazioni) in base al tipo di ipotesi alternativa che abbiamo. Se l'ipotesi alternativa è MINORE dell'ipotesi fatta, sceglieremo la “coda sinistra”, cioè i valori minori del valore dell'ipotesi osservabili in meno del 5% dei casi. Se l'ipotesi alternativa è MAGGIORE, sceglieremo la “coda destra”. Se l'ipotesi alternativa è solo DIVERSA, sceglieremo di considerare non falsificanti l'ipotesi il 95% delle osservazioni “centrali” - ovvero di escludere il 2,5% di osservazioni a sinistra e il 2,5% a destra.

Un disegno vi chiarirà il tutto (ma è roba che avete già visto con la gaussiana...).



Facciamo un esempio per confermare di aver capito:

“Si suppone che il fumo sia la causa della malattia M. Vengono seguiti per 5 anni 80 fumatori sani. Nei non fumatori il rischio di M in 5 anni è del 5%. Quale ipotesi saggiamo? Quando possiamo rifiutarla?”

→ Saggiamo l'ipotesi nulla, cioè l'ipotesi che il fumo non abbia effetto.

Intuitivamente vi verrebbe da dire: ma perché? Non dovrei saggiare l'ipotesi che il fumo sia la causa? Beh, provate a ragionarci su: come fareste a saggiarla? In sostanza dovrete trovare osservazioni che falsificano l'ipotesi. Ma l'ipotesi è in questo senso molto vaga! Infatti dovrebbe essere “nei fumatori il rischio di M in 5 anni è maggiore del 5%”; peccato che non potete costruire un modello con una probabilità così indeterminata (pensateci: nel binomiale, che è quello che useremo ora – infatti un soggetto può avere o non avere la malattia! - la probabilità di successo deve essere determinata, non può essere “maggiore del 5%”).

Quindi cercheremo di falsificare l'ipotesi “il fumo non ha effetto” ovvero “nei fumatori la probabilità di M in 5 anni è del 5%” - in questo caso possiamo costruire un modello perché abbiamo una probabilità di successo da saggiare ben definita.

Costruiamo quindi un modello binomiale con 80 tentativi e probabilità di successo 0,05. Quale “coda” ci interesserà? Il fumo non sarà mai protettivo (quindi l'ipotesi alternativa all'ipotesi “il fumo non ha effetto” non può essere “il fumo è protettivo”: in termini formali non ci interessa l'ipotesi alternativa “nei fumatori la probabilità di M in 5 anni è MINORE di 0,05”), e questo esclude la coda sinistra. Per ovvi motivi è esclusa anche la

probabilità a due code. Ci interessa solo la coda destra: infatti il fumo può essere causa di M solo se la probabilità di M nei fumatori è MAGGIORE di 0,05.  
Possiamo ora calcolare la probabilità di osservare M malati su 80:

M	P	Cum
0	.0165	.0165
1	.0695	.0860
2	.1445	.2305
3	.1977	.4282
4	.2003	.6285
5	.1603	.7888
6	.1054	.8942
7	.0587	.9529
8	.0282	.9811
9-80	.0189	1.000

Il numero di M atteso è  $0,05 \times 80 = 4$ . Notiamo che osservare 8 o più M ha una probabilità del 4,71% ( $2,82 + 1,89$ ), dunque sotto il livello di significatività statistica del 5%.

→ Se avessimo voluto considerare l'ipotesi “il fumo è protettivo” avremmo detto che l'ipotesi nulla era smentita osservando 0 M.

→ Se avessimo voluto considerare come ipotesi alternativa “il fumo ha un qualche effetto” avremmo smentito l'ipotesi nulla per  $M=0$  o  $M>8$  (quindi con 8 escluso).

→ Visto che 80 è abbastanza grande avremmo potuto utilizzare anche il modello di Poisson per calcolare i dati (ricordate che Poisson è un'approssimazione del binomiale): avremmo utilizzato ovviamente  $\lambda=4$ , e i dati ottenuti sarebbero stati sostanzialmente uguali.

Chiediamoci ora: ma se NON faccio un'osservazione che mi porta a rifiutare l'ipotesi nulla, posso concludere che l'ipotesi nulla è vera? Ovviamente no (ricordate il tacchino induttivista).

Infatti potrebbe essere vera un'altra ipotesi, e io potrei aver fatto semplicemente un'osservazione improbabile sotto quella ipotesi.

Quindi quando faccio un'osservazione che mi porta a NON rifiutare l'ipotesi nulla (e quindi a darla per vera) rischio di commettere un errore: l'ipotesi nulla potrebbe infatti essere falsa.

Il rischio di commettere tale errore è indicato con  $\beta$  (o “errore di secondo tipo” - errore che si fa quando non si rifiuta un'ipotesi falsa sulla base di un'osservazione non significativa). Ovviamente tale errore sarà pari alla probabilità di fare quella osservazione sotto l'ipotesi alternativa.

Il valore  $1-\beta$  è invece la probabilità di NON fare quell'errore, ovvero la probabilità di non fare quell'osservazione sotto l'ipotesi alternativa; chiameremo questo valore la “potenza” del test che saggia l'ipotesi – è chiaro che la potenza di un test dipende dal tipo di ipotesi alternativa presa in considerazione.

Capisco che questo possa sembrarvi un discorso parecchio complicato, per cui proviamo a vedere se un esempio lo semplifica.

“Supponiamo che F curi il 70% dei malati. Qual è l'errore che farei osservando più di 4 guariti se F in realtà curasse solo il 50%?”

→ Costruiamo i due modelli binomiali con  $\pi$  rispettivamente pari a 0,7 e 0,5.

		$\pi = 0.7$		$\pi = 0.5$	
y	p	Probabilità	Cumulativa	Probabilità	Cumulativa
0	0	.0000059	.0000059	.0009766	.0009766
1	.1	.0001378	.0001437	.0097656	.0107422
2	.2	.0014467	.0015904	.0439453	.0546875
3	.3	.0090017	.0105921	.1171875	.1718750
4	.4	.0367569	.0473490	.2050781	.3769531
5	.5	.1029193	.1502683	.2460937	.6230468
6	.6	.2001209	.3503892	.2050781	.8281249
7	.7	.2668279	.6172171	.1171875	.9453124
8	.8	.2334744	.8506915	.0439453	.9892577
9	.9	.1210608	.9717523	.0097656	.9990233
10	1	.0282475	.9999998	.0009766	.9999999

Se l'ipotesi fosse  $\pi=0,7$  l'osservazione  $y>4$  (con  $y$ =numero di guariti) avrebbe una probabilità del 96,7% (100-4,73) mentre se l'ipotesi fosse  $\pi=0,5$  l'osservazione  $y>4$  avrebbe una probabilità del 62,3% (100-37,7).

Ma se avessi osservato  $y>4$  avrei considerato vera  $\pi=0,7$ . Se nella realtà  $\pi=0,5$  fosse vera, avrei dunque fatto un errore! Tale errore sarebbe del 62,3%, dato che avevo esattamente questa probabilità di fare l'osservazione  $y>4$  sotto l'ipotesi " $\pi=0,5$ ".

Detto in altri termini, la probabilità di fare un'osservazione che mi porti a NON rifiutare  $\pi=0,7$  se in realtà  $\pi=0,5$  è del 62,3% [tale osservazione è quella che, data vera  $\pi=0,7$ , aveva una probabilità del 95%].

Quindi, la potenza del test è  $1-0,623=0,377$ . Se ci fate caso, questa è esattamente la probabilità di fare un'osservazione che ci porti a rifiutare  $\pi=0,7$  se è vera  $\pi=0,5$ .

È chiaro che in questo caso il rischio di commettere un errore del secondo tipo è troppo alto (62%!).

Diremo che un test negativo (cioè che non falsifica l'ipotesi nulla) di BASSA potenza (minore dell'80%) non corrobora l'ipotesi nulla.

Nulla ci vieta comunque di aumentare il campione, in modo da aumentare la potenza, o di cambiare l'ipotesi alternativa per raggiungere lo stesso scopo.

→ Se invece che 10 soggetti ne avessimo avuti 100 le probabilità sarebbero cambiate notevolmente. Sotto l'ipotesi  $\pi=0,7$  osservare  $y>61$  avrebbe avuto una probabilità del 95%, mentre sotto l'ipotesi  $\pi=0,5$  tale osservazione avrebbe avuto una probabilità del 2%! Ciò significa che con un campione di 100 soggetti e un'ipotesi alternativa di  $\pi=0,5$  la potenza del test sarebbe stata del 98%, e non rifiutare l'ipotesi che  $\pi=0,7$  osservando più di 61 guariti avrebbe comportato un rischio di errore molto basso. In questo senso, osservare 61 o più guariti avrebbe corroborato l'ipotesi (escludendo di fatto l'ipotesi alternativa).

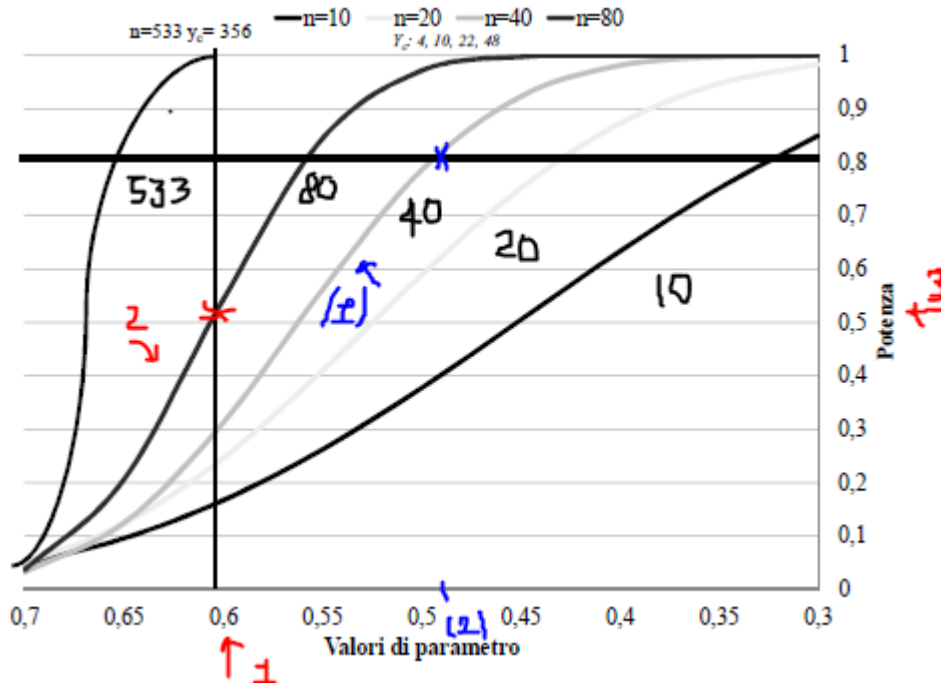
→ Se invece che  $\pi=0,5$  avessimo preso come ipotesi alternativa  $\pi=0,3$  un campione di 10 elementi sarebbe stato sufficiente, in quanto la probabilità di osservare  $y>4$  guariti sotto l'ipotesi  $\pi=0,3$  sarebbe stata solo del 15% - la potenza del test sarebbe stata dunque dell'85%. Il rischio di fare un errore non rifiutando l'ipotesi  $\pi=0,7$  sulla base dell'osservazione di più di 4 guarigioni su 10 sarebbe stato quindi basso, e avremmo concluso che l'ipotesi era corroborata.

Ma come si sceglie l'ipotesi alternativa? Secondo l'FDA una sperimentazione clinica è ben dimensionata se la potenza è almeno dell'80% per un'ipotesi alternativa pari al 50% dell'ipotesi nulla. Ciò significa all'atto pratico:

- Prendi la tua ipotesi ( $\pi=0,7$ ) e considera come ipotesi alternativa il suo 50% ( $\pi=0,35$ );
- Scegli il campione in modo che la potenza sia almeno l'80% (nel nostro caso  $n=14$ );
- Osserva come se non ci fosse un domani;
- Fatti pagare per pubblicare dei risultati falsi e scappa alle Cayman.

Tuttavia, nulla ci vieta di vagliare altre ipotesi alternative. Teniamo presente che data un'ipotesi nulla, un'ipotesi alternativa e una dimensione del campione si può facilmente determinare la potenza del test.

Possiamo tracciare anche un grafico che ci indichi come varia la potenza del test, per una data ipotesi nulla, al variare dell'ipotesi alternativa e del campione. Otterremo le cosiddette “curve di potenza”:



Curve di potenza per test con  $H_0: \pi=0,7$

Come si legge questo grafico? Sull'asse delle x avete i valori di “ipotesi alternativa”; su quello delle y i valori di potenza. Sono tracciate varie linee corrispondenti alle dimensioni dei campioni.

Per ogni linea sono sopra riportati i valori di osservazioni critiche, tali per cui osservare un numero di guarigioni superiore comporta il non rifiuto dell'ipotesi nulla (0,7 nel nostro caso).

Possiamo agire in due modi:

- Linee rosse:
  - 1: scegliamo l'ipotesi alternativa, nel nostro caso 0,6
  - 2: valutiamo un campione (ho scelto la linea n=80, nulla mi avrebbe vietato di scegliere n=533)
  - 3: ricaviamo il corrispettivo valore di potenza (0,5 circa: in questo caso il campione non va bene, se voglio valutare l'ipotesi alternativa scelta dovrò aumentare il numero del campione)
- Linee blu (metodo stupido se posso fare altrimenti):
  - (1): ho un certo campione – per vari motivi: potrebbero non esserci abbastanza soggetti da testare per esempio; nel nostro caso ho preso ad esempio n=40
  - (2): cerchiamo l'intersezione della linea del campione con potenza=80% e ricaviamo la corrispondente ipotesi alternativa – nel nostro caso sarà poco meno di 0,5. Ciò significa che con n=40 potrò escludere, se osservo più di 22 guarigioni, solo le ipotesi alternative da 0,5 in giù. Se osservo più di 22 guarigioni non potrò escludere l'ipotesi 0,6, perché la potenza del test non sarebbe sufficiente.

È importante ribadire che la potenza del test dipende dall'ipotesi alternativa che si assume vera (per valutare l'errore che si commette non considerando falsa l'ipotesi nulla). Si sceglie di solito l'ipotesi alternativa minima rilevante (il discorso fatto prima e conclusosi con la fuga alle Cayman riguardava le indicazioni dell'FDA, quindi quello che si fa nella realtà è ciò che stiamo dicendo ora). L'ipotesi alternativa minima rilevante è quell'alternativa che dal punto di vista pratico è di qualche interesse discriminare dall'ipotesi nulla; in tal modo, se non si può rifiutare l'ipotesi nulla, si

può almeno escludere l'ipotesi alternativa (con rischio di commettere un errore pari a  $\beta$ ).

Facciamo giusto un piccolo esempio cambiando alcune cose, per verificare di aver capito. Riprendiamo il caso della malattia M che colpisce il 5% dei non fumatori in cinque anni; valutiamo ora le ipotesi alternative con un campione di 80 fumatori.

Che ipotesi alternativa possiamo valutare? Supponiamo di voler valutare l'ipotesi (cioè, se ci pensate, l'ipotesi che il fumo comporti un rischio relativo pari a 3 di sviluppare M in 5 anni).

In questo caso otteniamo:

n casi	$\pi_0 = .05$		$\pi_1 = .15$	
	P	$\bar{P}$	P	$\bar{P}$
0	.0165	.0165	.000002	.000002
1	.0695	.0860	.000032	.000034
2	.1445	.2305	.000222	.000256
3	.1977	.4282	.001019	.001275
4	.2003	.6285	.003462	.004737
5	.1603	.7888	.008340	.013077
6	.1054	.8942	.018397	.031474
7	.0587	.9529	.034321	.065795
8	.0282	.9811	.055266	.121061
9-80	.0189	1.000	.878939	1.00000

Ricordatevi che stiamo ragionando considerando la coda DESTRA: abbiamo il 4,7% di possibilità di osservare più di 7 M (1-0.9529), quindi rifiuteremo l'ipotesi nulla (il fumo non ha effetto) se facciamo tale osservazione. Supponiamo ora di osservare MENO di 8 M: ciò ci porterà a NON rifiutare l'ipotesi nulla. Quale errore faremmo se in realtà il fumo avesse un rischio relativo pari a 3? Esattamente il 6,57% (cioè la probabilità di osservare meno di 8 M se  $\pi=0,15$ ), il che significa che il test in questo caso avrebbe una potenza del 93,42%. Si è quindi ragionevolmente certi di non commettere un errore del secondo tipo nei confronti di un rischio relativo di 3.

E se invece volessimo valutare un rischio relativo di 1,5 (cioè  $\pi=0,075$ )? Procediamo nello stesso modo: osserveremmo meno di 8 M nel 74% dei casi. Dunque, non potremmo considerare corroborata con altrettanta forza l'ipotesi nulla.

In sostanza, noi non corroboriamo proprio nulla; possiamo solo escludere ipotesi alternative.

*When you have eliminated all which is impossible, then whatever remains, however improbable, must be the truth.*

The adventure of the Blanched Soldier (Arthur Conan Doyle)

Bene, con questo abbiamo finito la teoria. Manca la parte con gli esercizi di genetica, per la quale vorrei aiutarvi, ma non ho il tempo di scrivere dispense anche su questo. Quindi se avete dei problemi riguardo questa parte e non sapete a chi chiedere scrivetemi pure su facebook (il nome è in cima alle dispense) e cercherò di rispondervi!

Spero che queste dispense vi siano state utili; ciò mi ripagherebbe della fatica che ho fatto a scriverle. Non credo di aver fatto un lavoro perfetto: se qualcosa non vi è chiaro fatemi sapere!

Ora non vi resta che l'esame!

***Che la forza sia con voi, miei giovani padawan!***

