Capitolo II

NOZIONI DI CALCOLO DELLE PROBABILITA'

1. Definizione

Nel linguaggio corrente spesso usiamo le parole "probabile" e "improbabile". Per esempio, diciamo:" E' probabile che domani piova", oppure "è improbabile che impariamo l'intero corso di Statistica in un giorno". Alcuni dizionari ci dicono che la parola "probabile" significa "verosimile". Altri cercano di spiegare il significato della parola "probabile" usando la parola "possibile". Cosicché spesso tutte queste definizioni ci conducono ad una sorta di circolo vizioso, mediante il quale si cerca di spiegare il significato di una parola utilizzando altre parole il cui significato è altrettanto difficile da definire quanto il primo.

Diventa dunque necessaria una definizione di probabilità come termine matematico, cioè una definizione abbastanza accurata per scopi matematici. Abbiamo necessità di trovare una misura che ci renda capaci di fare affermazioni numeriche certe, invece di vaghi commenti generali di contenuto eminentemente qualitativo e quindi non quantificabile.

La probabilità è talvolta definita come la misura della solidità o forza della nostra aspettativa riguardo ad un evento futuro. Ma persone diverse possono assegnare gradi di probabilità diversi allo stesso evento futuro oppure la stessa persona esprimerà un diverso modo di pensare riguardo allo stesso evento futuro in momenti di tempo diversi, a seconda dell'umore o del suo stato d'animo. E' quindi desiderabile trovare una definizione di probabilità che fornisca una

buona misura e che non conduca a contraddizioni, cioè non abbia caratteristiche di soggettività.

Se un evento può manifestarsi in un certo numero di modi che sono tutti ugualmente possibili ed una certa proporzione di questi sono classificati come favorevoli al manifestarsi dell'evento, mentre il resto sono classificati come sfavorevoli, il rapporto del numero dei modi favorevoli al numero totale dei modi di manifestarsi dell'evento è chiamato "probabilità che l'evento si manifesti in modo favorevole".

Si possono fare almeno due obiezioni a questa definizione, ma, malgrado ciò, essa può essere applicata in numerosi casi.

La prima obiezione riguarda cosa esattamente intendiamo con le parole "*ugualmente possibili*"? La seconda obiezione è che questa definizione non è applicabile in tutti quei casi in cui sia infinito il numero dei modi di manifestarsi dell'evento. Per questa ragione, dobbiamo enunciare questa definizione in una forma leggermente diversa.

Se un evento che può manifestarsi in due modi diversi si presenta con la stessa modalità un gran numero di volte sotto lo stesso insieme di condizioni, il rapporto del numero di volte o prove in cui l'evento si manifesta con quella modalità, al numero totale delle volte che si è manifestato o prove si avvicinerà ad un limite definito, man mano che il numero totale delle prove aumenta indefinitamente. Questo limite sarà chiamato "*probabilità che l'evento si manifesti con quella modalità*". Talvolta il rapporto tra il numero di casi favorevoli e il numero totale delle prove viene indicato come "*frequenza relativa di eventi favorevoli*", mentre il limite a cui tende la frequenza relativa con il crescere indefinitamente del numero totale delle prove è chiamato "*probabilità che l'evento si manifesti favorevolmente*".

Ancora una volta, nella definizione il significato della espressione "*sotto lo stesso insieme di condizioni*" si presta a critiche, perché è pressocché impossibile riuscire a ripetere un esperimento mantenendo immutate le circostanze nelle quali esso si realizza e l'ambiente nel quale esso si svolge. Ma questa obiezione potrebbe piuttosto essere rivolta contro tutta la scienza sperimentale, perché è certamente difficile ripetere un esperimento esattamente sotto le stesse condizioni.

Per esempio, persino nelle applicazioni della Fisica vi possono essere piccole variazioni nella temperatura, nella pressione barometrica, eccetera, le quali fanno si che gli esperimenti successivi avvengano sotto condizioni generalmente diverse. E se questo rappresenta solo una preoccupazione nel caso delle scienze esatte, diventa invece un problema veramente insormontabile nel caso delle scienze sociali.

Per superare l'ostacolo, quando un ricercatore vuole determinare una legge fisica tramite risultati sperimentali, egli sarà obbligato ad ipotizzare che le variazioni nelle condizioni di realizzazione degli esperimenti siano modeste ed ininfluenti sul risultato. Altrimenti tutta la scienza sperimentale non potrebbe esistere.

La seconda obiezione è contro l'ipotesi che esista un limite verso il quale tende la frequenza relativa. La validità di questa ipotesi non può essere provata teoricamente, ma l'esperienza in molti campi offre un sostegno alla sua ragionevolezza.

Per esempio, se lanciamo una moneta non truccata per 100 volte, il numero di volte che otterremo testa è sempre abbastanza prossimo al valore atteso che è 50. Probabilmente non avremo mai esattamente 50, ma non avremo solo una volta o due volte testa su 100 lanci.

Il significato della parola "*probabilità*" in statistica è talvolta espresso dalle parole: "*frequenza relativa teorica*", "*Valore atteso della frequenza relativa*". Infatti se p indica la probabilità di un evento, l'espressione:

$$p = \underset{N \to +\infty}{\text{Limite}} \ \frac{f}{N}$$

viene abbreviata come segue:

$$E\left(\frac{f}{N}\right) = p$$

in cui E( ) è letto "valore atteso di f/N" ed il termine f/N è proprio la frequenza relativa.

Nella maggior parte dei casi pratici si chiama "frequenza relativa" la probabilità di verificarsi di un evento. Quindi, se un evento può presentarsi n volte e non presentarsi m volte e se ciascuna di queste eventualità è ugualmente possibile, possiamo dire che la probabilità che l'evento si presenti è *n/(n+m)*, mentre la probabilità che l'evento non si presenti è *m/(n+m)*. Qui *n* è la frequenza con cui si presenta l'evento favorevole, mentre *(n+m)* è il numero totale delle prove. Talvolta, invece di dire che la probabilità che un evento si verifichi è *n/(n+m)*, diciamo che il pronostico è di *n* a *m* in favore dell'evento, oppure di *m* a *n* contro l'evento.

Da questa definizione si può facilmente vedere che la somma delle probabilità che l'evento si verifichi o non si verifichi è uguale ad 1. Cioè:

$$\frac{n}{n+m}+\frac{m}{n+m}=\frac{n+m}{n+m}=1.$$

Quindi, se indichiamo con *p* la probabilità che l'evento si presenti e con *q* la probabilità che esso non si presenti, avremo la relazione *p+q=1* oppure *q=1-p*. Quindi si vede che la probabilità *p* è un numero compreso tra 0 ed 1. Inoltre, è assolutamente certo che l'evento in questione *si presenterà* o *non si presenterà*. Non vi sono altre alternative. Quindi la somma delle probabilità che lo stesso evento si presenti o che non si presenti, cioè la probabilità totale, è sempre uguale ad 1.

Certe volte questo caso è anche chiamato "caso della certezza". In altre parole, *la probabilità che si verifichi un evento certo è espressa dall'unità*. Di converso, *è uguale a zero la probabilità che non si verifichi un evento che è certo*. Non esiste alcuna certezza maggiore di 1 e nulla che sia meno probabile di 0.

Vediamo alcuni esempi.

1. La probabilità che io un giorno muoia è $p = 1$, perché è assolutamente certo che un giorno morirò.

2. La probabilità che io muoia domani è un numero p compreso nell'intervallo $0 \le p \le 1$.

3. La probabilità che io viva per sempre è $p = 0$, perché nessun uomo o donna ci si aspetta che viva per sempre.

4. La probabilità che se lancio un dado non truccato ottenga uno qualunque dei numeri tra 1 e 6 è $p = 1$, perché un dado ha solo sei faccie ciascuna delle quali mostra un numero compreso tra 1 e 6 e certamente il risultato dovrà essere uno di questi numeri.

5. La probabilità di ottenere un numero maggiore di 2 lanciando un dado non truccato è $p = 4/6 = 2/3$, perché vi sono sei possibili modi in cui il dado può cadere e di questi, quattro sono favorevoli all'evento indicato.

6. In una lotteria che abbia distribuito 90 biglietti e che preveda di premiare 10 numeri e di non premiare gli altri 80 numeri, la probabilità che un individuo possessore di un biglietto ha di vincere un premio è $p = 10/90$, perché 90 sono tutti i numeri possibili e solo 10 di questi sono associati ad un premio.

7. Da un'urna contenente 6 biglie bianche e 4 biglie rosse, la probabilità di estrarre una biglia rossa è $p = 4/10$, ma la probabilità di estrarre 2 biglie rosse è:

$$p=\frac{C(4,2)}{C(10,2)}=\frac{\frac{4!}{2!2!}}{\frac{10!}{2!8!}}=\frac{4!2!8!}{2!2!10!}=\frac{2}{15}=0{,}13$$

perché il numero complessivo di volte in cui 2 biglie possono essere estratte è C(10,2) ed il numero di volte che le due biglie sono estratte da quelle rosse è C(4,2).

2. Legge additiva e probabilità totale

Supponiamo che vi siano due *eventi che si escludono a vicenda*, cioè se si presenta uno dei due l'altro non si presenta. La probabilità che si presenti il primo evento sia indicata da p_1 e quella relativa al secondo evento sia indicata da p_2. Si esegua una serie di N prove, con il risultato che il primo evento si presenta f_1 volte su N, mentre il secondo si presenta f_2 volte su N. Allora, per la definizione di probabilità data in precedenza, avremo:

$$p_1 = \underset{N\to+\infty}{\text{Limite}}\ \frac{f_1}{N}; \qquad p_2 = \underset{N\to+\infty}{\text{Limite}}\ \frac{f_2}{N}.$$

Ora f_1 ed f_2 sono le frequenze o numero di prove favorevoli collegate al primo ed al secondo evento rispettivamente. Se vogliamo conoscere la probabilità che si verifichino o il primo o il secondo evento, il numero di prove favorevoli è naturalmente pari a $(f_1 + f_2)$ e la probabilità corrispondente, secondo le definizioni precedenti sarà:

$$p_1 + p_2 = \underset{N\to+\infty}{\text{Limite}}\ \frac{f_1 + f_2}{N}.$$

Con un ragionamento simile possiamo dire che *la probabilità che si presenti uno qualsiasi di n eventi che si escludono vicendevolmente è uguale alla somma delle probabilità che si verifichino n eventi separati* e cioè:

$$P\begin{pmatrix}\text{probabilità che si presenti}\\ \text{uno qualunque tra n eventi}\\ \text{che si escludono a vicenda}\end{pmatrix} = p_1 + p_2 + p_3 + \ldots\ldots + p_n .$$

Per esempio, la probabilità che un dado mostri 1 oppure 6 é data dalla uguaglianza $p = 1/6 + 1/6 = 2/6 = 1/3$.

Consideriamo un altro esempio. Supponiamo di avere 15 carte segnate con i primi 15 numeri. Quando una di queste carte viene estratta a caso, qual'è la probabilità di avere un numero multiplo di 2 o di 9? Tra i primi quindici numeri non ve ne è nessuno che sia allo stesso tempo multiplo di 2 e di 9, per cui i due eventi si escludono a vicenda e la risposta è:

$$p_2 + p_9 = \frac{7}{15} + \frac{1}{15} = \frac{8}{15} .$$

Ma se ci domandassimo quale è la probabilità di ottenere un numero che sia multiplo di 2 o di 3, la risposta sarebbe diversa, perché tra i primi 15 numeri 6 e 12 sono multipli sia di 2 che di 3. Quindi i due eventi non si escludono a vicenda.

Per trovare la risposta giusta dobbiamo includere 6 e 12 in una sola delle due frequenze favorevoli e non in entrambe, altrimenti conteremmo 6 e 12 due volte ciascuno, una prima volta come multipli di 2 ed una seconda volta come multipli di 3, il ché porterebbe ad un risultato sbagliato. Quindi la probabilità di avere un multiplo di 2 è 7/15 e quella di avere un multiplo di 3 è 5/15, ma la risposta non è 7/15 + 5/15 = 12/15 = 4/5, bensì 7/15 + 5/15 - 2/15 = 10/15 = 2/3.

La probabilità che si presenti l'uno o l'altro dei due eventi che si escludono a vicenda è generalmente chiamata: Probabilità totale. La stessa espressione è anche usata nel caso di più di due eventi che si escludano vicendevolmente.

3. Eventi composti

Quando due o più eventi si verificano in connessione l'uno con gli altri, il verificarsi congiunto è chiamato evento composto. Parecchi eventi possono verificarsi simultaneamente o in successione. Ma essi possono essere sia indipendenti l'uno dall'altro, sia dipendenti tra loro. *Gli eventi si dicono*

indipendenti se il verificarsi dell'uno non influisce sul verificarsi degli altri. In altre parole, due eventi si dicono mutualmente indipendenti quando la probabilità che si verifichi uno di essi è la stessa sia che l'altro si presenti sia che non si presenti.

4. Legge moltiplicativa e probabilità composte

Siano p_1 e p_2 le probabilità di successo in due *eventi indipendenti*. Per il primo evento il numero dei modi possibili di verificarsi è N_1, mentre il numero dei modi favorevoli è f_1. Per il secondo evento, invece, le quantità suddette saranno indicate da N_2 ed f_2 rispettivamente. Quindi sappiamo che:

$$p_1 = \underset{N_1 \to +\infty}{\text{Limite}} \frac{f_1}{N_1}; \qquad p_2 = \underset{N_2 \to +\infty}{\text{Limite}} \frac{f_2}{N_2}.$$

Si voglia trovare la probabilità dell'evento composto, ossia la probabilità che si verifichino ambedue gli eventi. Questa probabilità sarà indicata con il simbolo p_{12}. Poiché abbiamo ipotizzato che gli eventi siano indipendenti l'uno dall'altro, il numero dei modi favorevoli al verificarsi di ambedue gli eventi è $f_1\ f_2$ (perché ognuno dei modi favorevoli al presentarsi del primo evento deve essere associato con ciascuno dei modi favorevoli al verificarsi del secondo evento). Con un ragionamento simile possiamo dire che il numero dei modi favorevoli al presentarsi del primo evento ed al non presentarsi del secondo è $f_1\ (N_2 - f_2)$. Il numero di modi favorevoli al non presentarsi del primo evento ed al presentarsi del secondo sarà $(N_1 - f_1).f_2$ ed infine il numero dei modi favorevoli al non presentarsi di ambedue gli eventi sarà pari a $(N_1 - f_1).(N_2 - f_2)$.

Inoltre, vi sono $N_1.N_2$ modi possibili di presentarsi dell'evento composto e di questi ultimi $f_1.f_2$ corrisponderanno al caso in cui tutti e due gli eventi si verificano. Di conseguenza:

$$p_1 . p_2 = \underset{\substack{N_1 \to +\infty \\ N_2 \to +\infty}}{\text{Limite}} \frac{f_1.f_2}{N_1.N_2};$$

ma per il teorema concernente il limite di un prodotto potremo scrivere:

$$\underset{\substack{N_1 \to +\infty \\ N_2 \to +\infty}}{\text{Limite}} \frac{f_1 . f_2}{N_1 . N_2} = \underset{N_1 \to +\infty}{\text{Limite}} \frac{f_1}{N_1} . \underset{N_2 \to +\infty}{\text{Limite}} \frac{f_2}{N_2} = p_1 . p_2$$

e quindi potremo formulare il seguente teorema della probabilità composta:
"La probabilità del verificarsi di un evento composto che consiste di due eventi indipendenti è uguale al prodotto delle probabilità di verificarsi relative a ciascuno degli eventi singoli".

Questo teorema può essere generalizzato al caso in cui gli eventi indipendenti siano più di due, per il quale potremo scrivere:

$$p_{123......n} = p_1 . p_2 . p_3 .... p_n .$$

Esempio II.1

La probabilità di ottenere due 1 nel lancio di una coppia di dadi non truccati è:

1/6 . 1/6 = 1/36

infatti, il punteggio che mostra il primo dado non influisce sul risultato ottenuto dal secondo dado. Quindi i due eventi sono indipendenti. Inoltre, quando lanciamo due dadi, ognuna delle 6 facciate del primo dado può associarsi con ognuna delle 6 facciate del secondo dado, per cui il numero totale delle associazioni possibili è pari a 6.6 = 36. Poiché vogliamo avere come risultato due 1, questo può verificarsi solo in un modo e quindi il numero dei modi favorevoli all'evento "uscita di due 1 con il lancio di due dadi non truccati" è pari ad 1. Quindi la probabilità ricercata è uguale a 1/36, che è uguale al prodotto delle probabilità di ottenere 1 separatamente da ogni dado, cioè 1/6.1/6.

Esempio II.2

Se si estraggono 4 biglie da un'urna che ne contiene 6 bianche e 8 rosse e si esegue poi una seconda estrazione di 4 biglie, dopo aver rimesso nell'urna le 4 estratte la prima volta, qual'è la probabilità che la prima estrazione dia come risultato 4 biglie bianche e la seconda 4 biglie rosse?

Il numero dei modi in cui 4 biglie possono essere estratte dal numero complessivo di 14 biglie presenti nell'urna è pari a C(14,4). Inoltre, poiché le 4 biglie estratte la prima volta sono rimesse nell'urna prima di effettuare la seconda estrazione, il risultato della prima estrazione non avrà alcuna influenza su quello della seconda. Quindi i due eventi sono indipendenti.

Il numero dei modi in cui 4 biglie bianche possono essere estratte dall'urna che ne contiene 6 è pari a C(6,4), mentre il numero dei modi in cui 4 biglie rosse possono essere estratte da un'urna che ne contiene 8 è pari a C(8,4).

Pertanto, la probabilità di avere 4 biglie bianche nella prima estrazione è data dalla relazione:

$$\frac{C(6,4)}{C(14,4)} = \frac{\frac{6!}{4!2!}}{\frac{14!}{4!10!!}} = \frac{6!4!10!}{4!2!14!} = \frac{6.5.4.3}{14.13.12.11} = \frac{30}{2002} = \frac{15}{1001}$$

La probabilità di avere 4 biglie rosse nella seconda estrazione sarà:

$$\frac{C(8,4)}{C(14,4)} = \frac{\frac{8!}{4!4!}}{\frac{14!}{4!10!!}} = \frac{8!4!10!}{4!4!14!} = \frac{8.7.6.5}{14.13.12.11} = \frac{140}{2002} = \frac{70}{1001} = \frac{10}{143}$$

Infine, la probabilità dell'evento composto sarà pari a:

$$\frac{15}{1001} . \frac{70}{1001} = \frac{1050}{1002001} \approx \frac{1}{1000}$$

Se vi sono n eventi indipendenti, le probabilità corrispondenti al verificarsi di ciascuno di essi separatamente, come si è visto, sono p_1, p_2, p_3......p_n, mentre la probabilità dell'evento composto è pari a $p_1.p_2.p_3......p_n$. Ma, per esempio, la probabilità che i primi tre eventi si verifichino e gli altri (n-3) non si verifichino, sarà data dal prodotto $p_1.p_2.p_3.(1-p_4)......(1-p_n)$.

Esempio II.3

Un dado viene tirato per 5 volte. Trovare la probabilità che la faccia con il 3 appaia nelle prime 2 prove.

Tale probabilità si otterrà con il seguente prodotto:

$$\frac{1}{6} . \frac{1}{6} . \frac{5}{6} . \frac{5}{6} . \frac{5}{6} = \frac{125}{8376}$$

Ma se la domanda fosse un poco diversa e, ad esempio, ci fosse stato richiesto di trovare la probabilità che la faccia con il 3 appaia proprio 2 volte sui 5 lanci, allora la risposta avrebbe dovuto essere:

$$C(5,2) . \frac{1}{6} . \frac{1}{6} . \frac{5}{6} . \frac{5}{6} . \frac{5}{6} = \frac{5!}{2!3!} . \frac{125}{8376} = \frac{1250}{8376}$$

perché vi sono C(5,2) modi diversi di combinare 2 prove su 5.

Per trovare la probabilità che almeno uno di n eventi indipendenti si verifichi, si deve ragionare come segue: la probabilità che nessuno degli n eventi si presenti è $(1-p_1).(1-p_2)..........(1-p_n)$, se le probabilità di verificarsi dei singoli eventi separatamente sono $p_1, p_2, p_3,......p_n$.

Ma se nessuno degli n eventi non si verifica, questo fatto sta a significare che almeno uno di essi si verificherà e quindi la probabilità cercata sarà pari a:

$$1 - (1-p_1).(1-p_2)..........(1-p_n),.$$

Esempio II.4

Una carta viene estratta a caso da un mazzo di 52 carte e poi è rimessa nel mazzo; viene quindi estratta una seconda carta dallo stesso mazzo.

Qual'è la probabilità che l'asso di spade si presenti almeno una volta?

Per risolvere questo quesito si applica il metodo che segue, il quale permette di trovare con sufficiente facilità la soluzione di parecchi problemi di natura abbastanza simile a quello qui considerato.

La probabilità che l'asso di spade non si presenti in nessuna delle due estrazioni è:

$$\frac{51}{52} \cdot \frac{51}{52} = \frac{2601}{2704}$$

La probabilità di ottenere almeno un asso di spade è quindi:

$$\left(1 - \frac{2601}{2704}\right) = \frac{103}{2704}$$

5 Probabilità condizionata

Consideriamo due eventi x ed y *dipendenti* l'uno dall'altro, cioè l'evento y si verificherà quando si è già verificato l'evento x. Allora la probabilità che l'evento y si verifichi è detta probabilità condizionata ed è di solito indicata con la notazione $p_x(y)$.

In questo caso la legge moltiplicativa della probabilità si scriverà come segue:

$$p(x,y) = p(x).\,p_x(y),$$

in cui $p(x)$ è la probabilità che si verifichi l'evento x e $p(x,y)$ è la probabilità che si verifichino sia x che y. Quindi l'espressione che fornisce la probabilità condizionata è:

$$p_x(y) = \frac{p(x,y)}{p(x)}.$$

Esempio II.5

Supponiamo di estrarre una carta da un mazzo di 52 carte. Senza rimettere la carta estratta nel mazzo, ripetiamo l'operazione con una seconda carta. Qual'è la probabilità che tutte e due le carte siano dello stesso colore?

Prima della prima estrazione, il mazzo contiene carte di due colori nello stesso numero, ma poiché la carta estratta per prima non viene riposta nel mazzo, la seconda estrazione dipenderà dal risultato della prima. Quindi i due eventi sono tra loro dipendenti. In questo caso possiamo usare la formula della probabilità condizionata.

Indichiamo con $p(x)$ la probabilità di estrarre una carta del colore desiderato nella prima estrazione. Allora $p_x(y)$ rappresenterà la probabilità condizionata, cioè la probabilità di estrarre una carta dello stesso colore nella seconda estrazione:

$$p(x) = \frac{1}{2}; \qquad p_x(y) = \frac{25}{51}$$

e quindi la probabilità cercata è pari a:

$$p(x,y) = p(x).\,p_x(y) = \frac{1}{2} \cdot \frac{25}{51} = \frac{25}{102}.$$

6 Speranza matematica

Se una persona ha la probabilità p_1 di ottenere una somma x_1 e p_2 di ottenere la somma x_2 e così via, la sua speranza matematica è data dalla formula:

$$\sum_{i=1}^{i=n} x_i \cdot p_i = x_1 . p_1 + x_2 . p_2 + \dots\dots\dots . + x_n . p_n$$

Esempio II.6

Una persona ha due biglietti di due lotterie diverse. Nella prima lotteria vi sono 5 premi e 95 biglietti non premiati. I premi sono di 200 milioni ciascuno. Nella seconda lotteria vi sono 10 biglietti premiati e 40 non premiati. I premi sono di 100 milioni ognuno. Qual'è la speranza matematica di questa persona?

Per la prima lotteria si ha:

$p_1 = 5/100 = 0{,}05 \qquad x_1 = 200$

mentre per la seconda si ha:

$p_2 = 10/50 = 0{,}20 \qquad x_2 = 100$

Quindi la sua speranza matematica sarà pari a:

$0{,}05 \,.\, 200 + 0{,}20 \,.\, 100 = 10 + 20 = 30$ milioni.

7. Prove ripetute

Se è nota la probabilità che si verifichi un evento in una prova, la probabilità che lo stesso evento si presenti una volta, due volte, eccetera, in n prove è data dai termini successivi dello sviluppo binomiale.

Infatti, indichiamo con p la probabilità che si verifichi l'evento e con q la probabilità dell'evento contrario (cioè che l'evento atteso non si verifichi). Se selezioniamo un particolare insieme di r prove sul numero complessivo delle n prove, la probabilità che l'evento si verifichi in tutte queste r prove e non si presenti nelle altre (n-r) è:

$$p^r \,.\, q^{n-r}.$$

Ma un insieme di r prove può essere selezionato tra le n in C(n,r) modi diversi, tutti ugualmente possibili. Quindi la probabilità che l'evento si presenti in r prove è uguale a:

$$C(n,r).\,p^r.\,q^{n-r} = \frac{n!}{r!(n-r)!}.\,p^r.\,q^{n-r}.$$

Questa espressione rappresenta il termine generale dello sviluppo del binomio di Newton. Infatti, facendo assumere ad r i valori r = 0, 1, 2, ..., n si ottengono tutti i termini che formano lo sviluppo del binomio $(p + q)^n$,:

$$(p+q)^n = C(n,n).\,p^n + C(n,n-1).\,p^{n-1}.\,q + C(n,n-2).\,p^{n-2}.\,q^2 + \ldots\ldots +$$

$$+ C(n,r).\,p^r.\,q^{n-r} + \ldots\ldots + C(n,1).\,p.\,q^{n-1} + C(n,0).\,q^n = \sum_{r=0}^{r=n} C(n,r).\,p^r.\,q^{n-r}.$$

Quindi *la probabilità che l'evento si verifichi* <u>*esattamente*</u> *r volte su n prove è data dall' (n-r+1)-esimo termine della espansione del binomio (p+q)*^{n}*.*

La probabilità che l'evento si verifichi <u>*almeno r volte*</u> *è uguale alla somma dei primi (n-r+1) termini nella espansione del binomio (p+q)*^{n}*.*

<u>Esempio II.7</u>

Se lanciamo un dado 6 volte, la probabilità di avere 1 per 4 volte è pari a:

$$p = \frac{6!}{4!2!}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^2 = 15.\frac{25}{6^6} = \frac{375}{46656}.$$

Ma la probabilità di avere 1 almeno 4 volte è uguale alla somma dei tre termini:

$$C(6,4).\,p^4.\,q^2 + C(6,5).\,p^5.\,q + C(6,6).\,p^6 = 15.\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^2 + 6.\left(\frac{1}{6}\right)^5\left(\frac{5}{6}\right) + \left(\frac{1}{6}\right)^6 =$$

$$= \frac{15.\,25 + 6.\,5 + 1}{6^6} = \frac{375 + 30 + 1}{46656} = \frac{406}{46656}.$$

Se il termine generico del binomio di Newton lo indichiamo con la espressione

$$P_r = C(n,r).\,p^r.\,q^{n-r} = \frac{n!}{r!(n-r)!}.\,p^r.\,q^{n-r}$$

valgono le seguenti formule ricorrenti:

$$P_{r+1} = P_r.\,\frac{\mathbf{p}}{\mathbf{q}}.\,\frac{n - x}{\mathbf{r}+1} \quad \text{e} \quad P_{r-1} = P_r.\,\frac{\mathbf{q}}{\mathbf{p}}.\,\frac{x}{\mathbf{n}-\mathbf{r}+1}$$

I valori C(n,r) sono anche chiamati *coefficienti binomiali*, appunto dato che essi figurano come coefficienti nella relazione che spiega lo sviluppo della potenza n-esima di un binomio.

Se nel binomio generico $(a+b)^n$ poniamo a = b = 1, si ha come risultato che

$$\sum_{r=0}^{r=n} C(n,r) = 2^n$$

Lo sviluppo binomiale della espressione $(p + q)^n$ è di grandissima importanza nella teoria statistica. Essa è la base della distribuzione binomiale che è una delle 3 più importanti distribuzioni teoriche discontinue considerate nella statistica. Studieremo in dettaglio le proprietà della distribuzione binomiale nel seguito.

I coefficienti dello sviluppo del binomio possono essere ottenuti con facilità dal "*triangolo di Pascal*", che si presenta come segue:

Triangolo di Pascal

Valori di n	Coefficienti in $(a+b)^n$
0	1
1	1 1
2	1 2 1
3	1 3 3 1
4	1 4 6 4 1
5	1 5 10 10 5 1
6	1 6 15 20 15 6 1
7	1 7 21 35 35 21 7 1
8	1 8 28 56 70 56 28 8 1
9	1 9 36 84 126 126 84 36 9 1
10	1 10 45 120 210 252 210 120 45 10 1

Esso consiste in una tabella in cui ciascun termine è ottenuto sommando insieme i due termini che si trovano nella riga precedente immediatamente sopra quello cercato. Quindi, per esempio, il quarto termine della ottava riga (56) che figura nella tabella che precede è ottenuto sommando il terzo (21) ed il quarto (35) termine della riga settima immediatamente sopra di esso.

L'utilità di tale tabella consiste nel fatto che, usando il triangolo di Pascal possiamo scrivere immediatamente, per esempio, lo sviluppo del binomio $(a + b)^n$, semplicemente utilizzando i coefficienti esposti nella riga ennesima e questi li possiamo determinare conoscendo quelli della riga (n-1)-esima e così via. Ovviamente la sua utilità è limitata dal fatto che si abbia conoscenza dei termini della riga (n-i)-esima oppure resta limitata solo ai polinomi che non hanno un grado molto elevato. In tutti gli altri casi, comunque, la determinazione dei coefficienti dello sviluppo è facilitata dall'uso dei coefficienti binomiali che figurano nei vari termini dello sviluppo.

Per esempio, utilizzando i coefficienti della settima riga del Ttriangolo di Pascal, si può determinare immediatamente lo sviluppo di $(a+b)^7$.come segue:

$$(a + b)^7 = C(7,7).a^7 + C(7,6).a^6b + C(7,5)a^5b^2 + C(7,4)a^4b^3 + C(7,3)a^3b^4 + C(7,2)a^2b^5 + C(7,1).ab^6 + C(7,0).b^7 =$$

$$= a^7 + 7a^6b + 21.a^5b^2 + 35.a^4b^3 + 35.a^3b^4 + 21.a^2b^5 + 7ab^6 + b^7$$

Si può mostrare che lo sviluppo binomiale vale anche nel caso che i valori di n siano negativi o rappresentino frazioni, ma in questi casi il coefficiente del termine generale deve essere preso nella forma

$$\frac{n(n-1)(n-2).....(n-r+1)}{r!}$$

perché il simbolo C(n,r) non ha significato quando l'esponente del binomio n è negativo o frazionario.

8. Riepilogo delle formule probabilistiche più utili

Siano A e B due fenomeni. A ciascuno si diano n possibilità di manifestarsi. Supponiamo che, dopo queste n possibilità, risulti che:

n_1 volte si è verificato A, ma non B

n_2 volte si è verificato B, ma non A

n_3 volte si sono verificati A e B insieme

n_4 volte non si sono verificati nè A nè B

(ovviamente $n_1+n_2+n_3+n_4=n$). Allora, per definizione:

$$F(A) = \frac{n_1+n_3}{n}$$ = frequenza relativa di A

$$F(B) = \frac{n_2+n_3}{n}$$ = frequenza relativa di B

$$F(A+B) = \frac{n_1+n_2+n_3}{n}$$ = frequenza che si realizzi o A o B o entrambi

$$F(AB) = \frac{n_3}{n}$$ = frequenza congiunta (che si realizzino A e B insieme)

$$F(A|B) = \frac{n_3}{n_2+n_3}$$ = frequenza condizionale (che si verifichi A, sapendo che B si è già verificato)

$$F(B|A) = \frac{n_3}{n_1+n_3}$$ = frequenza condizionale (che si verifichi B, sapendo che A si è già verificato)

E' facile constatare che dalle definizioni predette si deducono le seguenti formule:

$$F(A+B) = F(A) + F(B) - F(AB)$$
$$F(AB) = F(A|B).F(B) = F(B|A).F(A)$$
$$F(A|B) = \frac{F(B|A).F(A)}{F(B)}$$

Quando n, numero di possibilità, è assai elevato e le frequenze sembrano restare costanti, possiamo parlare, invece che di frequenze, di "probabilità".

9. Applicazioni

L'uso degli assiomi, teoremi e definizioni riportati sia per il calcolo combinatorio sia per il calcolo delle probabilità, sarà illustrato ora tramite alcune applicazioni.

Esempio II.8

Si abbia una scatola di 50 lampadine, 5 delle quali sono difettose. Quale è la probabilità di trovare esattamente 2 lampadine difettose in un campione di 4 lampadine prelevato dalla scatola?

I campioni possibili sono C(50,4) e si suppone che essi siano tutti egualmente probabili. La probabilità ricercata si definisce come rapporto tra il numero dei campioni possibili formati di 4 lampadine 2 delle quali difettose ed il numero di tutti i possibili campioni di 4 lampadine estraibili dalla scatola che ne contiene 50.

Per contare quanti possono essere i campioni che hanno 2 lampadine difettose, notiamo che le 2 difettose devono provenire dalle 5 difettose contenute nella scatola e ciò può essere fatto in C(5,2) modi diversi, mentre le altre 2 lampadine non difettose del campione devono essere estratte dalle 45 lampadine non difettose contenute nella scatola e ciò può essere fatto in C(45,2) modi diversi.

In base al principio delle probabilità composte, il numero dei modi in cui due eventi indipendenti possono verificarsi contemporaneamente è pari al prodotto del numero dei modi in cui ciascuno dei due eventi può verificarsi indipendentemente dall'altro. Notiamo, altresì, che, in base al principio delle probabilità totali, in un gruppo di eventi che si escludono a vicenda il numero di volte che può presentarsi uno qualunque di essi è pari alla somma del numero di volte che può presentarsi ciascuno di essi. In questo caso i due eventi che si devono verificare sono la selezione di 2 lampadine difettose fra le 5 presenti nella scatola e la selezione di 2 lampadine non difettose tra le 45 presenti nella scatola. Quindi, potremo selezionare un campione di 4 lampadine di cui 2 difettose in un numero di modi uguale al prodotto C(5,2).C(45,2) e la probabilità ricercata sarà pari a:

$$\frac{C(5,2).C(45,2)}{C(50,4)} = \frac{\frac{5!}{2!3!}.\frac{45!}{2!43!}}{\frac{50!}{4!46!}} = \frac{5.4}{2}.\frac{45.44}{2}.\frac{4.3.2.1}{50.49.48.47} = 0,043$$

In maniera simile possiamo calcolare le probabilità di avere nel campione di 4 lampadine tutte le 4 lampadine non difettose, oppure 1 sola difettosa, 2 sole difettose, 3 sole difettose o tutte e 4 le lampadine difettose. I risultati sono riportati nella Tabella II.8.1.

La discussione che precede è per sua natura di tipo matematico, dato che abbiamo una completa informazione sulla popolazione assoggettata al campionamento, cioè, 50 oggetti tra i quali 5 sono difettosi. Il problema statistico sorge invece quando non conosciamo quanti dei 50 esemplari sono difettosi, ma vogliamo dedurre dal campione la proporzione dei pezzi difettosi presenti nella popolazione. Per esempio, potremmo decidere di ritenere troppo elevata la proporzione dei pezzi difettosi nella popolazione se qualcuno dei 5 pezzi del campione è difettoso. Se ciò accade, possiamo decidere di verificare tutti i 50 pezzi. Dalla tabella II.8.1 vediamo che se venisse adottata questa regola decisionale, nel 64,7 percento dei casi (campioni di 4 elementi) la scatola contenente 5 pezzi difettosi non darebbe luogo all'esame minuzioso anche di tutti i pezzi non campionati.

Tabella II.8.1 - Probabilità di avere x lampadine difettose in un campione di 4 lampadine estratto da una scatola che ne contiene 50, delle quali 5 difettose.

Numero di lampadine difettose x	Frequenza teorica o probabilità a priori di x pezzi difettosi P(x difettose) = f(x)	Frequenza cumulativa o probabilità di avere fino a x lampadine difettose P(difettose ≤ x) = F(x)
4	$\frac{C(5,4).C(45,0)}{C(50,4)} = 0,00002$	1,00000
3	$\frac{C(5,3).C(45,1)}{C(50,4)} = 0,00195$	0,99998
2	$\frac{C(5,2).C(45,2)}{C(50,4)} = 0,04299$	0,99803
1	$\frac{C(5,1).C(45,3)}{C(50,4)} = 0,30808$	0,95504
0	$\frac{C(5,0).C(45,4)}{C(50,4)} = 0,64696$	0,64696
Totale		1,00000

Esempio II.9

Supponiamo che un animale con una malattia particolare abbia il 40 percento di probabilità di guarire se trattato con certe medicine (e, naturalmente, una probabilità del 60 percento di non guarire).

Una interpretazione di questa affermazione è la seguente.

Consideriamo un grandissimo numero di animali che hanno contratto quella malattia; supponiamo che tutti questi animali siano trattati con la medicina e che il 40 percento di essi guarisca. In questo caso possiamo affermare che esiste una probabilità del 40% di scegliere a caso un animale che guarirà, da questo grande numero di animali trattati con quella medicina. La interpretazione di questo esempio è la stessa che nell'esempio II.8, eccetto che il numero di pezzi da cui il campione è scelto è molto grande o infinito.

Una interpretazione alternativa è la seguente.

Vi è il 40 percento di probabilità di guarigione per qualunque animale quando gli sia somministrata quella medicina; cioè, la possibilità di guarire è inerente all'animale stesso piuttosto che rappresentare la probabilità di selezione di un animale che guarirà quando sia stato trattato con quella medicina. Questa seconda interpretazione è analoga al verificarsi di testa nel lancio di una moneta quando la probabilità di avere testa è 40 percento.

Supponiamo che si voglia rispondere alla seguente domanda: qual'è la probabilità che esattamente 2 animali sui 5 trattati con la medicina guarirà? Noi assumiamo 0,4 come la probabilità di guarigione dell'animale scelto per primo, assumiamo 0,4 come la probabilità di guarigione dell'animale scelto per secondo (vale a dire che, come nel lancio di monete, il secondo evento è indipendente dalla prima scelta) e così via. La probabilità di guarigione di uno qualunque dei 5 animali è 0,4 (indipendente dalla guarigione o meno degli altri animali scelti).

Ora possiamo affrontare il problema con una procedura utile per una vasta varietà di problemi di probabilità. L'evento di avere 2 guarigioni su 5 animali trattati scriviamolo come una serie di eventi esclusivi possibili come segue:

"2 guarigioni su 5 animali trattati con la medicina" è la stessa cosa di:
"Animali 1 e 2 guariti e animali 3, 4, 5 non guariti" oppure
"Animali 1 e 3 guariti e animali 2, 4, 5 non guariti" oppure
"Animali 1 e 4 guariti e animali 2, 3, 5 non guariti" oppure
"Animali 1 e 5 guariti e animali 2, 3, 4 non guariti" oppure
"Animali 2 e 3 guariti e animali 1, 4, 5 non guariti" oppure
"Animali 2 e 4 guariti e animali 1, 3, 5 non guariti" oppure
"Animali 2 e 5 guariti e animali 1, 3, 4 non guariti" oppure
"Animali 3 e 4 guariti e animali 1, 2, 5 non guariti" oppure
"Animali 3 e 5 guariti e animali 1, 2, 4 non guariti" oppure
"Animali 4 e 5 guariti e animali 1, 2, 3 non guariti".

Vi sono C(5,2) eventi esclusivi combinati nell'una o nell'altra delle affermazioni precedenti e la probabilità dell'evento combinato è pari alla somma delle probabilità di verificarsi dei singoli eventi.

Poiché le probabilità di guarire per i singoli animali sono indipendenti, la probabilità che il primo ed il secondo animale guariscano e i restanti 3 non guariscano è:

P(il primo guarisce). P(il secondo guarisce). P(il terzo non guarisce). P(il quarto non guarisce). P(il quinto non guarisce) = $(0,4).(0,4).(0,6).(0,6).(0,6) = (0,4)^2.(0,6)^3$.

Tabella II.9.1 - Probabilità di avere esattamente x guarigioni su un campione di 5 prove indipendenti in cui la probabilità di una singola guarigione è 0,4.

Numero di guarigioni nel campione x	Frequenza teorica di avere esattamente x guarigioni P(x guarigioni) = f(x)		Frequenza cumulativa o probabilità di avere fino a x guarigioni P(guarigioni ≤ x) = F(x)
5	$C(5,5) = (0,4)^5$	= 0,01024	1,00000
4	$C(5,4).= (0,4)^4(0,6)$	= 0,07680	0,98976
3	$C(5,3).= (0,4)^3(0,6)^2$	= 0,23040	0,91296
2	$C(5,2).= (0,4)^2(0,6)^3$	= 0,34560	0,68256
1	$C(5,1).= (0,4)^1(0,6)^4$	= 0,25920	0,33696
0	$C(5,0).= (0,6)^5$	= 0,07776	0,07776
Totale			1,00000

La probabilità di uno qualunque dei C(5,2) = 10 eventi sopra elencati sarà la stessa poiché la sola variazione nello scrivere la probabilità è uno scambio nell'ordine in cui i fattori sono scritti. Quindi abbiamo:

$$P(\text{esattamente } 2 \text{ guariti su } 5) = C(5,2)\,(0,4)^2.(0,6)^3 = 0,03456.$$

Similmente possiamo calcolare la probabilità di avere esattamente 3 guarigioni su 5, eccetera. I risultati sono presentati nella tabella II.9.1.

Ancora una volta la discussione che precede è di tipo matematico. Il problema statistico potrebbe qui presentarsi se volessimo stimare la probabilità di guarigione degli animali malati sulla base delle osservazioni fatte su 5 soggetti scelti a caso.

Esempio II.10

Un tale Y sostiene di poter dire se in un contenitore chiuso vi sia acqua o no, tenendo una forcella di legno sopra il contenitore. Per provare la sua capacità, viene suggerita l'esecuzione dell'esperimento come segue. Ad Y vengono presentati 6 insiemi di 3 contenitori ciascuno. Un contenitore di ciascun insieme dei 3 è riempito con acqua, gli altri 2 sono vuoti ed Y viene informato di questo. Egli tenta di scegliere in ogni insieme il contenitore che è pieno di acqua. Se non ha alcuna capacità particolare, egli ha una probabilità di 1/3 di scegliere correttamente il contenitore con l'acqua in qualunque insieme, indipendentemente dalle sue altre scelte. Se invece ha un certo talento nell'avvertire la presenza di acqua, egli dovrebbe fare meglio che indovinando. In altre parole, egli dovrebbe avere una probabilità indipendente maggiore di 1/3 di scegliere correttamente il contenitore con acqua in qualunque insieme di 3 contenitori. Se si specifica la probabilità P che Y scelga correttamente il contenitore con acqua, questo esempio si presenta identico all'esempio II.9 e possiamo dunque calcolare la probabilità che egli ne scelga correttamente x in 6 prove, come in quell'esempio. I risultati di questo calcolo sono riportati nella Tabella II.10.1 per quattro valori di P.

Tabella II.10.1 - Probabilità di scegliere correttamente esattamente x contenitori di acqua in 6 prove con probabilità di successo p costante in ogni prova. [Per esempio: P(4 scelte corrette su 6 prove, con P = 1/3) = $C(6,4)(1/3)^4(2/3)^2$ = 0,0823].

x	P = 1/3	P = 1/2	P = 4/5	P = 1
6	0,0014	0,0156	0,2621	1
5	0,0165	0,0938	0,3932	0
4	0,0823	0,2344	0,2458	0
3	0,2195	0,3125	0,0819	0
2	0,3292	0,2344	0,0154	0
1	0,2634	0,0938	0,0015	0
0	0,0878	0,0156	0,0001	0

Un possibile accertamento dell'abilità di Y può essere fatto convenendo di dire che, se la sua scelta è corretta in almeno 5 delle 6 prove, vi è evidenza che egli abbia una certa capacità di scegliere bene. Se indovina nello scegliere, avremo una probabilità di concludere che ha qualche capacità di scelta (quando in realtà egli non possiede alcuna abilità particolare) uguale a 0,0179 = 0,0014 + 0,0165. Trarre una deduzione sulla sua capacità in base ad un campione del suo lavoro è un problema statistico. La probabilità di concludere che Y ha una certa abilità quando egli in realtà non ne ha affatto è detta anche "livello di significatività" dell'esperimento.

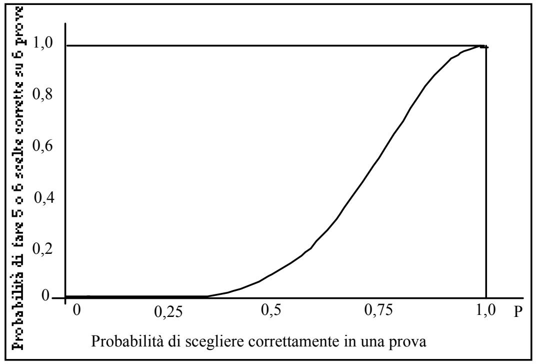

Fig. II.10 - Probabilità che ha Y di operare 5 o 6 scelte corrette su 6 prove con la probabilità P per ogni prova.

Nella Fig. II.10 è rappresentata la probabilità di fare 5 o 6 scelte corrette su 6 prove. La curva tracciata è una curva integrale per verificare l'abilità di Y. L'ordinata della curva fornisce la probabilità di concludere che "Y non cerca di indovinare" se in realtà egli ha la probabilità P di selezionare il contenitore giusto. Se ha il 50% di probabilità di scegliere correttamente, vi è una probabilità pari a 0,1094 di concludere che egli non tenta di indovinare; se ha una probabilità dell'80%, vi è una probabilità pari a 0,6553 di concludere che egli non tenta di indovinare. La curva tracciata nella Fig. II.10 è la curva integrale per verificare l'ipotesi H: P = 1/3, dove la regione critica (Cfr. cap. VI) del test è x = 5, 6.

Va osservato che si potrebbero progettare anche altri esperimenti per verificare la reale capacità di Y. Per esempio, i contenitori potrebbero essere presentati 2 alla volta invece che 3 alla volta oppure potrebbero essere presentati singolarmente, con un lancio di moneta per determinare se ogni contenitore sia vuoto o no.

Si definisce "piano dell'esperimento" il modo specifico in cui si procede alla raccolta dei dati; l'ipotesi da verificare concernente la popolazione (per esempio, P = 1/3 per specificare che Y tenta di indovinare) è chiamata "*ipotesi statistica*" e la regola che indica la procedura dell'esperimentatore (per esempio, rigettare l'ipotesi se risultano 5 o 6 successi) è un "*test statistico*" dell'ipotesi (Cfr. cap. VI e segg.).

Esempio II.11

Un porcellino d'India giallo incrociato con un altro giallo produce un porcellino d'India giallo; uno bianco incrociato con un bianco ne produce uno anch'esso bianco. Ma una cavia gialla incrociata con una bianca produce sempre una prole color crema. I porcellini d'India color crema quando vengono incrociati alla lunga producono il 25 percento di prole gialla, il 50 percento di prole color crema ed il 25 percento di prole bianca. Questo tipo di ereditarietà può essere spiegato attraverso l'uso delle formule genetiche AA, Aa, aa, che rappresentano il giallo, il color crema ed il bianco, rispettivamente. Quando due animali sono incrociati, la prole riceve un fattore indicato con A oppure con a da ognuno dei due genitori. Quindi, se ambedue i genitori sono AA, tutta la prole sarà AA e se tutti e due i genitori sono aa, anche tutta la prole sarà aa.

Se ambedue i genitori sono Aa, allora la loro prole può essere AA, Aa oppure aa. Poiché le cavie color crema sono del tipo Aa, allora si avranno come risultato dell'incrocio tutti e tre i tipi di combinazioni dei fattori e noi possiamo calcolare le relative probabilità come segue, ipotizzando che ambedue i fattori abbiano una uguale probabilità di essere trasmessi da ognuno dei due genitori:

P(nati AA) = P(riceventi A dal padre ed A dalla madre) = (1/2).(1/2) = 0,25

P(nati Aa) = P(riceventi A dal padre ed a dalla madre oppure riceventi a dal padre ed A dalla madre) = = (1/2).(1/2) + (1/2).(1/2) = 0,50.

P(nati aa) = P(riceventi a dal padre ed a dalla madre) = (1/2).(1/2) = 0,25

Utilizzando le ipotesi precedenti e con il ragionamento possiamo calcolare varie probabilità. In un gruppo di 10 nati da genitori del tipo Aa la probabilità di avere esattamente 3 nati color crema è $C(10,3)(0,50)^{10}$, mentre la probabilità di avere esattamente 4 nati di colore giallo è $C(10,4)(0,25)^4(0,75)^6$ e la probabilità di avere esattamente 3 nati color crema e esattamente 4 nati gialli è pari al seguente prodotto: $C(10,3)(0,50)^{10} \cdot C(7,4)(0,50)^4(0,50)^3$.

Se un genitore è AA e l'altro Aa, la probabilità che un nato sia AA è 0,50 e quella che un nato sia Aa è 0,50. Quindi la probabilità che esattamente 4 su 10 nati da questi genitori siano Aa è $C(10,4)(0,50)^{10}$.

Esempio II.12

Progettiamo di piantare le varietà A e B di granturco su coppie di appezzamenti di uguale dimensione in 6 località diverse.

Tabella II.12.1 - Probabilità di x segni più e (6-x) segni meno in 6 prove indipendenti ognuna di probabilità 1/2.
Per esempio, $P(1\ \text{più}) = C(6,1)(1/2)^6 = 0,093750$

x	(6-x)	Probabilità	r
6	0	0,015625	0
5	1	0,093750	1
4	2	0,234375	2
3	3	0,312500	3
2	4	0,234375	2
1	5	0,093750	1
0	6	0,015625	0

Poiché ogni coppia di appezzamenti in ciascuna delle 6 località ha condizioni ambientali uniformi, possiamo supporre che, se le due varietà danno rese uguali, sia 1/2 la probabilità che in nessuna delle località il tipo A dia una resa più elevata del tipo B. Quando la resa del tipo A è maggiore di quella del tipo B in qualunque località registriamo un più, altrimenti registriamo un meno. Supponiamo che la probabilità di un pareggio delle rese sia zero. Possiamo ora calcolare le probabilità: P(nessun più), P(1 più); P(2 più), eccetera. In realtà un calcolo simile è stato fatto nell'Esempio II.10. I risultati sono riportati nella Tabella II.12.1. In questa tavola è indicato con r il numero di volte che si verifica il segno meno frequente. Dalla

Tabella II.12.1 e con un numero di prove $n = 6$ possiamo leggere i valori possibili di r con le rispettive probabilità riportate però come nella Tabella II.12.2.

Notiamo che la Tabella II.12.2 contiene l'elenco completo delle probabilità cumulate P(r o meno di r) per $n = 6$. Quindi non possiamo trovare un valore di r corrispondente a qualunque valore P(r o meno di r) arbitrariamente scelto a meno che non sia elencata la relativa probabilità.

Tabella II.12.2 - Numero di volte che si presenta il segno meno frequente (più o meno), con le rispettive probabilità di verificarsi.

r	P(r) = f(r)	P(r o meno) = F(r)
3	0,31250	1,00000
2	0,46875	0,68750
1	0,18750	0,21875
0	0,03125	0,03125

Ad esempio, possiamo affermare che per il valore 0,01, arbitrariamente scelto e non compreso nella Tabella II.12.2, nessun intero r soddisfa la relazione P(r o meno)$\leq$ 0,01. Il numero $r = 0$ può essere usato come il più grande numero intero relativamente a F(r) = 0,05 e a F(r) = 0,10 poiché P(0 o meno) $\leq$ 0,05 e P(0 o meno) $\leq$ 0,10. Il numero $r = 1$ è il più grande intero che ha la proprietà che P(r o meno) $\leq$ 0,25.

Esempio II.13

Per dati del tipo descritto nell'Esempio II.12, supponiamo di scrivere i segni più e meno nell'ordine che si è osservato durante l'esperimento e per il numero di segni più e meno trovati (per esempio sulle 6 località esaminate siano stati segnati 1 o 2 o 3 etc. segni più per altrettante località e per le restanti il segno meno) calcoliamo la probabilità di ogni singola combinazione. Per esempio, se vi sono tre segni più e tre segni meno, le combinazioni possibili sono date nella Tabella II.13.1.

Tabella II.13.1 - Combinazioni possibili di 3 segni più e 3 segni meno e numero delle sequenze di ugual segno presenti in ciascuna combinazione

+ + + - - - 2	+ - + - + - 6	- + + + - - 3	- + - - + + 4
+ + - + - - 4	+ - + - - + 5	- + + - + - 5	- - + + + - 3
+ + - - + - 4	+ - - + + - 4	- + + - - + 4	- - + + - + 4
+ + - - - + 3	+ - - + - + 5	- + - + + - 5	- - + - + + 4
+ - + + - - 4	+ - - - + + 3	- + - + - + 6	- - - + + + 2

Ogni combinazione è seguita da un numero u, che rappresenta il numero di sequenze dello stesso segno che si hanno nella combinazione di segni considerata (ad esempio, nella combinazione di 8 segni ++++---- si hanno due sequenze, una con quattro più ed una con quattro meno; nella combinazione ++-+---+ si hanno 5 sequenze delle quali una con tre -, una con due +, una con un - e due con un +). Se supponiamo che ogni combinazione sia egualmente probabile, possiamo calcolare la probabilità di qualunque valore di u come nella Tabella II.13.2.

Tabella II.13.2 - Numero delle sequenze che si possono presentare in una combinazione casuale di tre segni più e tre segni meno, con le rispettive probabilità.

u	P(u) = f(u)	P(u o meno)=F(u)
6	2/20	1,0
5	4/20	0,9
4	8/20	0,7
3	4/20	0,3
2	2/20	0,1

Il numero totale delle combinazioni è uguale al numero di volte che tre dei sei posti possono essere scelti per i tre segni più, cioè C(6,3). In generale, se vi sono N segni più ed M segni meno, allora vi sono C(N+M,N) possibili combinazioni. E' più difficile contare le combinazioni in cui u ha un valore determinato. Si dimostra che se u è pari il numero di combinazioni è

$$2C(N-1, u/2 - 1).C(M-1, u/2 - 1)$$

e se u è dispari il numero di combinazioni è

$$C(N-1,u/2 - 1/2).C(M-1, u/2 - 3/2) + C(N-1, u/2 - 3/2).C(M-1, u/2 - 1/2)$$

La concordanza di queste formule con la distribuzione calcolata nella Tabella II.13.2 può essere mostrata ponendo $N = M = 3$. Per esempio, per $u = 2$ il numero di combinazioni è $2C(2,0) = 2$ e per $u = 3$ il numero delle combinazioni è $C(2,1)C(2,0) + C(2,0)C(2,1) = 2 + 2 = 4$, eccetera.

Si deve osservare che stiamo calcolando la probabilità che si abbia un certo numero di sequenze quando siano dati N ed M. Per certi problemi possiamo voler lasciare che anche N ed M varino, per esempio, permettere che il numero così come le combinazioni di più e meno siano una variabile casuale. Questo condurrà a probabilità differenti. Per esempio, se lanciamo una moneta 3 volte, la probabilità che vi sia esattamente due volte ($u = 2$) una stessa faccia (non importa se testa o se croce) è data dalla probabilità che si verifichi una delle seguenti sequenze di testa (T) e croce (C) che si escludono a vicenda: TTC, CTT, CCT, TCC. Poichè la probabilità di una qualunque sequenza è 1/8, la probabilità che si verifichi per due volte la stessa faccia (per esempio due volte T) è uguale a 0,5 e infatti:

P(due volte la stessa faccia) = P(TTC oppure CTT o CCT o TCC) =
= P(TTC) + P(CTT) + P(CCT) + P(TCC) =
= 1/8 + 1/8 + 1/8 + 1/8 = 1/2 = 0,5.

Esempio II.14

Consideriamo due varietà di mais x ed y e supponiamo di indicare con x_1, x_2, x_3 e y_1, y_2, y_3 le rese di sei appezzamenti .

Una volta che si conoscono le rese dei sei appezzamenti, questi ultimi sono ordinati secondo il livello della resa ed a ciascun livello viene assegnato un numero che ne indica l'ordine: 1 al più grande, 2 al maggiore di tutti meno il primo e così via, fino a 6 per il più piccolo.

Sia T' la somma dei posti assegnati a x_1, x_2, x_3, cioè alle rese della prima varietà di mais. La somma di tutti e sei i posti è 21; quindi la somma dei posti assegnati a y_1, y_2, y_3 sarà 21 meno la somma assegnata alla prima varietà.

Se le varietà sono egualmente buone quanto a resa, possiamo considerare che, tra le 20 possibili, qualunque combinazione di posti delle rese degli appezzamenti sia egualmente probabile.

La Tabella II.14.1 elenca gli ordinamenti possibili con i corrispondenti valori di T'.

Tabella II.14.1 - Ordinamenti possibili di tre x e tre y
e somma dei posti assegnati ai valori di x

x x x y y y 6	x y x y x y 9	x y y x y x 11	y y x x x y 12
x x y x y y 7	y x x x y y 9	y x x y y x 11	y x y y x x 13
x x y y x y 8	x y x y y x 10	y x y x x y 11	y y x x y x 13
x y x x y y 8	x y y x x y 10	y x y x y x 12	y y x y x x 14
x x y y y x 9	y x x y x y 10	x y y y x x 12	y y y x x x 15

La Tabella II.14.2 fornisce la distribuzione di T'. La statistica T' è chiamata statistica "somma dei ranghi" (Cfr. paragrafo XI.4.2).

Tabella II.14.2 - Distribuzione di T' per due campioni di tre osservazioni ciascuno.

T'	P(T')	P(T' o meno)
15	0,05	1,00
14	0,05	0,95
13	0,10	0,90
12	0,15	0,80
11	0,15	0,65
10	0,15	0,50
9	0,15	0,35
8	0,10	0,20
7	0,05	0,10
6	0,05	0,05