Capitolo 5 – LA FORMULA DELLA SOMMA

Una domanda molto interessante può essere:
"Se si lanciano due dadi, qual è la probabilità che la somma dei numeri delle due facce sia una certa somma p?"
La risposta a questa domanda non è particolarmente difficile. Per soddisfare questo quesito si enumerano i casi che soddisfano l'evento richiesto e si divide il numero di eventi favorevoli per il numero di casi possibili.
Nel caso in cui il numero di dadi che si lanciano è piccolo il problema risulta essere relativamente semplice, ma quando si vuole calcolare la probabilità di ottenere una data somma con un numero di dadi più grande (ad esempio 10), il conteggio dei casi favorevoli richiede certamente un tempo molto lungo.
La domanda che ci si pone adesso è la seguente:
"E' possibile ricavare una formula generale che fornisca la probabilità di avere una certa somma p lanciando n dadi, ciascuno con s facce (s=6, nel caso di dadi comuni), ottenibile in tempo ragionevole?"
Cercando di trovare una risposta a questa domanda, sono state fatte delle ricerche nel campo del calcolo combinatorio che ha riportato allo sviluppo di una formula generale calcolabile utilizzando il linguaggio di programmazione R.
L'idea di base è che presi dei dadi che vengono lanciati ripetutamente si va a prestare attenzione alla somma dei lanci, così da potere confrontare la frequenza relativa ottenuta tramite la simulazione in R con la probabilità teorica e valutare come essa si possa utilizzare nei giochi d'azzardo.
I dadi sono usati in molti giochi d'azzardo come un metodo per scegliere numeri casuali su cui scommettere (Craps, Cavaliere de Mère ecc…) e sono usati nei giochi da tavolo per determinare il numero di spazi da muovere, i risultati di un conflitto, ecc…
Prima di lanciarci all'applicazione reale della formula, vanno fatta un analisi preliminare.

5.1 Generalizzazione del problema della somma di risultati di lanci di dadi

Se volessimo determinare la probabilità di ottenere una somma uguale a 31 con un lancio di 10 dadi cosa succederebbe?
Lo spazio campionario U risulta essere la disposizione composta seguente:

$$D^*_{6,10} = 6^{10} = 60466176$$

Essendo 6^{10} un numero molto grande è logico pensare che calcolare il numero dei casi favorevoli richiederà un tempo molto lungo.

La domanda quindi che ci viene naturale è:
"Esiste una formula generale per determinare il numero dei possibili esiti favorevoli lanciando n dadi a s facce non truccati per ottenere una certa somma p?"
La risposta a questa domanda è **sì**.

L'origine e la particolarità di questa formula non è un calcolo combinatorio qualunque, poiché il suo sviluppo è stato ricostruito da concetti base che uniti

tra loro ci permettono di dimostrare e capire la correttezza della formula.
Si parte dall'idea di utilizzare il modello delle combinazioni con ripetizione.
Questa scelta è stata effettuata poiché questo modello ci permette di risolvere i problemi del tipo: *"trova il numero di possibili modi in cui si possono distribuire 3 biglie in 5 scatole (ricordando che ci sono esiti in cui delle scatole sono vuote)"*.
Ma esso risulta avere una analogia al problema del lancio di 10 dadi con somma 31, in particolare si tratterà di determinare il numero dei modi in cui si possono distribuire 31 biglie (somma) in 10 scatole (dadi):

$$C^*_{n,k} = \binom{n+k-1}{k} = C^*_{10,31} = \binom{10+31-1}{31} = \binom{40}{31} = \frac{40!}{31! \cdot 9!} = 273438880$$

5.2 Primi ostacoli

Prestando più attenzione, si nota che una scatola può essere vuota o contenere qualche biglia o perfino contenerle tutte, lasciando le altre scatole vuote.
Ciò non è permesso nel problema dei dadi poiché, banalmente, un dado a 6 facce ha come valore minimo 1 (problema 1) e ha come valore massimo 6 (problema 2).
Soddisfare la prima condizione è facile. Basterà assegnare ad ogni scatola a priori 1 biglia, quindi il problema si trasforma nell'assegnare le 21 biglie rimanenti a 10 scatole.
Soddisfare la seconda condizione è un po' più complesso poiché impostare un limite massimo non è così banale come può sembrare, poiché è come se si stesse valutando un dado a 22 facce.
Il passo successivo è quello di calcolare i casi favorevoli in un risultato temporaneo introducendo delle correzioni necessarie per soddisfare il vincolo che ogni dado reale può avere è un valore massimo di 6.
Quindi, per determinare il numero corretto dei possibili esiti che possono presentare i comuni dadi a 6 facce, sarà necessario eliminare, dal numero dei possibili esiti, il numero dei casi proibiti derivanti dalle scatole fuorilegge che non rispecchiano il vincolo del valore massimo di 6 biglie.
Considerando che, per rispettare il valore minimo di 1 biglia per ogni scatola, essa ha già preliminarmente ricevuto 1 biglia, il numero di biglie che una scatola potrà ricevere dovrà essere compreso tra 0 e 5. Le scatole fuorilegge saranno quindi quelle che riceveranno (dopo la pre-assegnazione di 1 biglia) un numero di biglie ≥ 6.

5.3 Rimozione dei casi proibiti

Trovare un metodo per la rimozione dei casi proibiti è la parte più difficile, poiché il numero di distribuzioni valide (o legali) è ottenibile attraverso la sottrazione:

num. di distribuzioni valide = num. di distribuzioni con valori illimitati – num. di distribuzioni proibite.

Essa non è banale come può sembrare. Il valore temporaneo del risultato senza la correzione dei casi proibiti risulta essere:

$$C^{*}{}_{10,21} = \binom{10+21-1}{21} = \binom{30}{21}$$

L'attenzione ora si sposta nel cercare i casi proibiti da sottrarre per ottenere le distribuzioni valide. Un primo approccio è quello di *forzare* una scatola ad essere una scatola fuorilegge, quindi assegnandogli a priori un numero di biglie ≥ 6. Nel caso in cui assegnassimo alla prima scatola 6 biglie, considerando le 10 biglie già assegnate all'inizio, allora ci saranno altri 15 biglie da assegnare. Quindi le combinazioni avranno il valore seguente:

$$C^{*}{}_{10,15} = \binom{10+15-1}{15} = \binom{24}{15} = 1307504$$

Considerando però che la scelta di forzare una scatola fuorilegge può avvenire in 10 modi diversi (10 possibili scelte della scatola fuorilegge), il numero totale di distribuzioni proibite generate in questo modo sarà:

$$10 \cdot \left(C^{*}{}_{10,15}\right) = 10 \cdot \binom{10+15-1}{15} = 10 \cdot \binom{24}{15} = 13075040$$

Chiamiamo questo gruppo di distribuzioni proibite $\boldsymbol{A_1}$.

La procedura sopra indicata potrebbe essere riassunta dai seguenti 4 passi eseguiti ciclicamente:

1. si sceglie la 1ª scatola;
2. si assegnano ad essa 6 biglie;
3. si distribuiscono le rimanenti 15 biglie a tutte le 10 scatole (inclusa quella attualmente scelta come fuorilegge);
4. si passa alla scatola successiva e si riparte dal passo (2);

Il ciclo terminerà dopo il raggiungimento della 10ª scatola.

Durante la distribuzione delle 15 biglie restanti, si possono avere delle assegnazioni che possono portare un'altra scatola ad essere fuorilegge, questo ci porta alla conclusione che il gruppo di distribuzione A_1 è il gruppo che include tutte le distribuzioni proibite con almeno 1 scatola fuorilegge.

Sfortunatamente sorge un altro problema: nel gruppo generato sono presenti duplicati di distribuzioni proibite.

In pratica il gruppo A_1 non è un insieme ma un multi-insieme, ovvero se in una distribuzione in cui la scatola 2 fosse scelta come scatola fuorilegge, una sua possibile distribuzione porta ad esempio la scatola 9 fuorilegge, quindi quando sarà il turno della scatola 9 come fuorilegge una sua possibile distribuzione porterà la scatola 2 ad essere fuorilegge, ma questa distribuzione è stata prodotta 2 volte. Questo ragionamento è applicabile per ogni combinazione semplice di 2 scatole ($C_{10,2}$). Quindi per trovare il numero di distribuzioni duplicate nel gruppo $\boldsymbol{A_1}$ è necessario trovare il numero di distribuzioni avente almeno 2 scatole fuorilegge.

Quindi costruiremo il gruppo con almeno 2 scatole fuorilegge che prenderà il nome di gruppo $\boldsymbol{A_2}$ nel quale sono contenute le disposizioni in cui sono preassegnate 6 biglie a 2 scatole.

Poiché nel gruppo $\boldsymbol{A_2}$ sono assegnate a due scatole 6 biglie, considerando le 10 biglie già assegnate all'inizio, allora ci saranno altre 9 biglie da assegnare. Quindi le combinazioni avranno il valore seguente:

$$C^{*}{}_{10,9} = \binom{10+9-1}{9} = \binom{18}{9} = 48620$$

La scelta di queste due scatole pero può essere fatta in $\binom{10}{2} = 45$ modi diversi quindi il numero totale di distribuzioni proibite che vengono generate in questo modo è:

$$\binom{10}{2} \cdot (C^*{}_{10,9}) = 45 \cdot \binom{10+9-1}{9} = 45 \cdot \binom{18}{9} = 2187900$$

Ciò che si nota è che anche il gruppo ***A2*** è un multi-insieme, ovvero ha anch'esso distribuzioni duplicate.
Una cosa sulla quale bisogna fare attenzione è che un suo possibile duplicato viene prodotto quando si fa riferimento a 2 scatole fuorilegge e se ne crea una 3ª, però, facendo riferimento a quest'ultima e ad una delle 2 scatole scelte inizialmente, durante l'assegnamento delle restanti 9 biglie, si crea un'altra duplicazione con l'altra scatola che non abbiamo scelto. Quindi nel gruppo ***A2*** abbiamo $\binom{3}{2} = 3$ sequenze duplicate per ogni doppia preassegnazione.
Quindi, per contare il numero delle distribuzioni duplicate all'interno del gruppo ***A2***, è necessario trovare il numero di distribuzioni aventi almeno 3 scatole fuorilegge.
Quindi costruiremo il gruppo con almeno 3 scatole fuorilegge che prenderà il nome di gruppo ***A3*** nel quale sono contenute le disposizioni in cui sono preassegnate 6 biglie a 3 scatole.
Poiché nel gruppo ***A3*** sono assegnate a 3 scatole 6 biglie, considerando le 10 biglie già assegnate all'inizio, allora ci saranno altre 3 biglie da assegnare. Quindi le combinazioni avranno il valore seguente:

$$C^*{}_{10,3} = \binom{10+3-1}{3} = \binom{12}{3} = 220$$

La scelta di queste tre scatole pero può essere fatta in $\binom{10}{3} = 120$ modi diversi quindi il numero totale di distribuzioni proibite che vengono generate in questo modo è:

$$\binom{10}{3} \cdot (C^*{}_{10,3}) = 220 \cdot \binom{10+3-1}{3} = 220 \cdot \binom{12}{3} = 26400$$

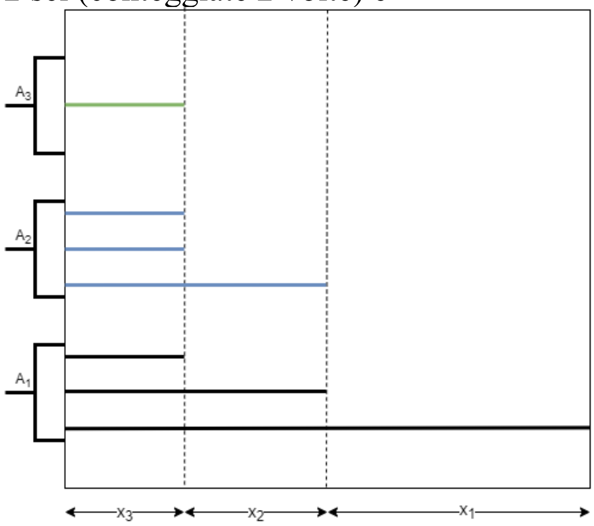
Sembra non ci sia più fine a questa continua necessità di trovare il numero di distribuzioni con sempre più scatole fuorilegge ma, fortunatamente ora le cose cambiano.
A differenza dei due gruppi precedenti, nel gruppo ***A3*** sicuramente non sono presenti duplicati, inquanto sicuramente nell'assegnamento delle 3 biglie rimanenti non si può creare un'ennesima scatola fuorilegge, infatti ci accorgiamo che se si prendessero 4 scatole fuorilegge allora avremmo già 24 biglie assegnate, il che non è possibile inquanto il massimo valore di biglie da assegnare è 21.
E' importante ricordare che ci sono sequenze duplicate del gruppo ***A3*** non solo nel gruppo ***A2*** ma anche nel gruppo ***A1***, ovvero sicuramente ci saranno delle disposizioni in cui scegliendo una scatola fuorilegge, se ne creano altre 2,
ad esempio la sequenza {1,1,**<u>6</u>**,0,0,6,0,6,0,1} scegliendo come scatola fuorilegge quella alla 3° posizione. La stessa sequenza sarebbe conteggiata anche quando la scatola preassegnata è quella in posizione 6 oppure quella in posizione 8.

Ci sono quindi $\binom{3}{1} = 3$ sequenze duplicate per ogni preassegnazione di 6 biglie ad una scatola.
Queste sequenze appartengono sia al gruppo ***A₁*** che al gruppo ***A₂*** che all'insieme ***A₃***.
Detto ciò, il processo è finito e non ci resta che calcolare finalmente le distribuzioni valide.
Per far ciò definiamo 3 variabili:
x_3 numero delle distribuzioni uniche con esattamente 3 sei;
x_2 numero delle distribuzioni uniche con esattamente 2 sei;
x_1 numero delle distribuzioni uniche con esattamente 1 sei;
Dunque, possiamo dire che:

- $\boldsymbol{A_3}$ è già formato da distribuzioni uniche con esattamente 3 sei e quindi:
$$x_3 \;=\; |A_3|$$
- $\boldsymbol{A_2}$ include 3 volte le distribuzioni con esattamente 3 sei. Il numero di distribuzioni uniche con esattamente 2 sei si può ottenere sottraendo $3x_3$ dalla cardinalità di $\boldsymbol{A_2}$:
$$x_2 \;=\; |A_2| - \;3x_3 = \;|A_2| - \;3|A_3|$$
- $\boldsymbol{A_1}$ include 2 volte le distribuzioni uniche con esattamente due sei e 3 volte le distribuzioni uniche con esattamente 3 sei.
Quindi il numero di distribuzioni con esattamente 1 sei è:
$$x_1 = |A_1| - 2x_2 - 3x_3 = \;|A_1| - 2(|A_2| - \;3|A_3|) - 3|A_3|$$
$$= \;|A_1| - 2|A_2| + \;3|A_3|$$

Finalmente è possibile contare il numero delle possibili distribuzioni proibite, <u>senza distribuzioni duplicate:</u>
$$x_1 + x_2 + x_3 = \;|A_1| - 2|A_2| + \;3|A_3| + |A_2| - \;3|A_3| + \;|A_3|$$
$$= \;|\boldsymbol{A_1}| - |\boldsymbol{A_2}| + |\boldsymbol{A_3}|$$
La figura seguente visualizza il numero di sequenze duplicate nei gruppi $\boldsymbol{A_1}$, $\boldsymbol{A_2}$ e $\boldsymbol{A_3}$ il risultato ottenuto $x_1 + x_2 + x_3 = |\boldsymbol{A_1}| - |\boldsymbol{A_2}| + |\boldsymbol{A_3}|$.

Il gruppo $\boldsymbol{A_1}$ include tutte le distribuzioni con esattamente 1 sei (conteggiate una volta) ma anche le distribuzioni con 2 sei (conteggiate 2 volte) e quelle con 3 sei (conteggiate 3 volte).
Allo stesso modo il gruppo $\boldsymbol{A_2}$ include tutte le distribuzioni con esattamente 2 sei (conteggiate 1 volta) ma anche quelle con 3 sei (conteggiate 3 volte).
In generale nel multi-insieme $\boldsymbol{A_k}$ le distribuzioni con esattamente j sei sono conteggiate $\binom{j}{k}$ volte.



Lo stesso risultato può anche essere derivato algebricamente risolvendo un sistema di incognite x_1, x_2, x_3 per come sono state precedentemente definite.

5.4 Conclusioni per la formula

Dopo aver trovato il numero di distribuzioni proibite e dopo aver eliminato da questo numero le distribuzioni duplicate possiamo finalmente applicare la relazione:

num. di distribuzioni valide = num. di distribuzioni con valori illimitati – num. di distribuzioni proibite.

dove il numero di distribuzioni con dadi illimitati è:

$$C^*{}_{10,21} = \binom{10+21-1}{21} = \binom{30}{21} = 14307150$$

ed il numero di distribuzioni proibite (senza duplicati) è:

$$|\boldsymbol{A_1}| - |\boldsymbol{A_2}| + |\boldsymbol{A_3}| = \binom{10}{1}\binom{24}{15} - \binom{10}{2}\binom{18}{9} + \binom{10}{3}\binom{12}{3} = 10913540$$

e il numero delle distribuzioni permesse è quindi:

$$N^{10}{}_{31} = \binom{30}{21} - \left(\binom{10}{1}\binom{24}{15} - \binom{10}{2}\binom{18}{9} + \binom{10}{3}\binom{12}{3}\right) = 3393610$$

Concretamente, considerando il numero di possibili esiti ($6^{10} = 60466176$), possiamo affermare che la probabilità di avere la somma 31 con un lancio di 10 dadi è:

$$P(31{,}10{,}6) = \frac{3393610}{60466176} = 0.05612 \rightarrow\ 5.612\%$$

Al termine di questo lungo procedimento per costruire l'espressione che fornisce la probabilità di ottenere una somma 31 con un lancio di 10 dadi,
le espressioni che, come abbiamo verificato, ci permettono di calcolare la probabilità di avere una somma p con il lancio di n dadi a s facce ci permette di acquisire una migliore comprensione sul ruolo dei diversi elementi che costituiscono la formula che è stata derivata, che risulta essere:

$$P(p,n,s) = \frac{1}{s^n} \cdot \sum_{k=0}^{k_{max}} (-1)^k \binom{n}{k}\binom{p-s\ \cdot k-1}{p-s\ \cdot k-n}$$

Sembra complessa, ma è frutto di tutte le nozioni precedentemente descritte, infatti gli elementi che la compongono sono:

- $\frac{1}{s^n}$ corrisponde a $P(U)$, ovvero alla probabilità del numero di casi possibili;
- $\binom{n}{k}\binom{p-s\ \cdot k-1}{p-s\ \cdot k-n}$ corrisponde al calcolo dei casi proibiti, con la valutazione del pre-inserimento di 1 biglia nelle n scatole, e le loro possibili permutazioni;
- $(-1)^k$ corrisponde al calcolo per ottenere il vero insieme delle distribuzioni proibite(senza duplicati), la cui cardinalità come nell'esempio era $x_1 + x_2 + x_3$, la cui tecnica usata è necessaria per rimuovere i casi duplicati basata sul <u>principio di inclusione ed esclusione</u>, applicato a multi-insiemi.
- $k_{max} = \left[\frac{p-n}{s}\right] = [x]$ dove $[x]$ è la parte intera del numero risultante, che corrisponde al numero massimo dei casi proibiti calcolabili.

Infine. per mostrare che la formula è corretta non ci basta che applicarla:

$$P(31,10,6) = \frac{1}{6^{10}} \cdot \sum_{k=0}^{3} (-1)^k \binom{10}{k} \binom{31-6k-1}{31-6k-10} =$$

$$= \frac{1}{6^{10}} \cdot \left(\binom{10}{0}\binom{30}{21} - \binom{10}{1}\binom{24}{15} + \binom{10}{2}\binom{18}{9} - \binom{10}{3}\binom{12}{3} \right) = 0.05612 \rightarrow 5.612\%$$

Essa risulta coincidere con il calcolo precedentemente effettuato.

5.5 Approfondimento: un'interpretazione duale

Esiste un modo più semplice per calcolare le combinazioni per il nostro problema.
Per ottenere il numero di distribuzioni con dadi illimitati nei capitoli precedenti sono state usate le
combinazioni con ripetizione: $C^*_{10,21} = \binom{10+21-1}{21} = \binom{30}{21}$.
Ad essa è stato associato i possibili modi di distribuire 21 biglie in 10 scatole.
Sapendo che $\binom{n}{k} = \binom{n}{n-k}$ quindi $\binom{30}{21} = \binom{30}{9}$, sfruttando la formula a proprio favore, si nota che è possibile applicare un discorso analogo usando le combinazioni con ripetizione:

$$C^*_{10,21} = \binom{10+21-1}{21} = \binom{30}{21} = \binom{30}{9} = \binom{22+9-1}{9} = C^*_{22,9}$$

Da ciò si ricava un'interpretazione duale del problema originario:

$$C^*_{n,k} = \binom{n+k-1}{k} = \binom{n+k-1}{n-1} = C^*_{(k+1),(n-1)}$$

Tutte le sequenze ottenibili dall'interpretazione duale data, sono "collegabili" ad una sequenza che fa parte del problema originario.
Il nuovo problema può esser rappresentato come:
date 30 biglie numerate, se ne estraggono 9 senza reinserimento, e di queste si elencano tutte le possibili 9 estrazioni disposte in ordine crescente.

Una delle possibili combinazioni potrebbe essere:

$$\{5,11,17,22,23,24,26,28,29\}$$

Questa sequenza non è altro che una <u>somma cumulata</u> dal quale possiamo ricavare una sequenza appartenente al problema originario.
La sequenza originaria sarà costruita prendendo l'i-esima biglia della sequenza e sottraendola a quella precedente.
Applicando questo ragionamento alla sequenza precedente otteniamo:

$$\{5,6,6,5,1,1,2,2,1\}$$

Poiché il numero di lancio di dadi deve corrispondere a 10 e non a 9, ed avendo calcolato una somma parziale che sicuramente sarà ≤ 30, per ottenere 31 si utilizza una biglia virtuale di valore 31, che si aggiunge alla sequenza ottenendo:

$$\{5,11,17,22,23,24,26,28,29,31\}$$

Infine otteniamo:

$$\{5,6,6,5,1,1,2,2,1,2\}$$

Che è proprio quello che ci aspettavamo.
Questa interpretazione duale sembra più semplice di quella presentata precedentemente: infatti in questo caso la condizione che ogni biglia debba contribuire con almeno un punto alla somma non necessita nessun aggiustamento artificiale ed è automaticamente soddisfatta dal fatto che due numeri naturali consecutivi presenti nella lista ordinata differiscono sempre di almeno una unità.
Esiste quindi una corrispondenza biunivoca tra i possibili esiti del lancio di 10 dadi con somma 31 e le sequenze ordinate costruite come specificato sopra.
Però c'è un problema: particolari sequenze all'interno dell'interpretazione duale sono collegate a delle sequenze proibite all'interno del problema originario.
Prendiamo in considerazione per esempio la sequenza seguente:

$$\{1,2,3,4,5,6,7,8,9,31\}$$

Essa risulta:

$$\{1,1,1,1,1,1,1,1,1,\underline{\mathbf{22}}\}$$

Con questa interpretazione duale una distribuzione proibita avviene ogniqualvolta nella sequenza dei 9 numeri naturali (con il numero 31 aggiunto alla fine) il primo numero o la differenza tra due elementi consecutivi è maggiore di 6. Ad ogni modo anche l'interpretazione duale non risolve il problema dei dadi fuorilegge.