Un problema di dadi

Michele Impedovo

Riassunto Quante volte, in media, occorre lanciare un dado a *s* facce perché tutte le facce escano almeno una volta? Per risolvere questo problema non è necessario calcolare esplicitamente la probabilità che in *n* lanci di un *s*-dado escano *k* facce. È sufficiente fare ricorso alla *distribuzione geometrica* di probabilità di un'opportuna variabile aleatoria.

Abstract On average, how many times must a *s*-sided die be rolled until each side appears at least once? To solve this problem we do not necessarily know the probability of rolling *k* sides when we throw *n* dice. We will consider the geometric distribution of a random variable.

Michele Impedovo
Liceo Scientifico G. Ferraris Varese
impedovo@tin.it

Il problema del collezionista
Quante volte, in media, occorre lanciare una moneta affinché esca sia TESTA che CROCE?
Quante volte, in media, occorre lanciare un dado perché escano tutte le facce?
Quante figurine, in media, devo comprare per "finire" l'album?

Questo problema, noto come "il problema del collezionista" ammette una soluzione generale bellissima, che vogliamo illustrare.
In generale il problema si può formulare in questo modo: quante volte, in media, occorre lanciare un s-dado (un dado a s facce) perché escano tutte le facce?
Ad un primo approccio si tratta di calcolare, per ogni n, la probabilità $p(n)$ che escano tutte le facce con n lanci (naturalmente risulta $p(0) = p(1) = p(2) = \ldots = p(s-1) = 0$) e quindi calcolare la somma della serie

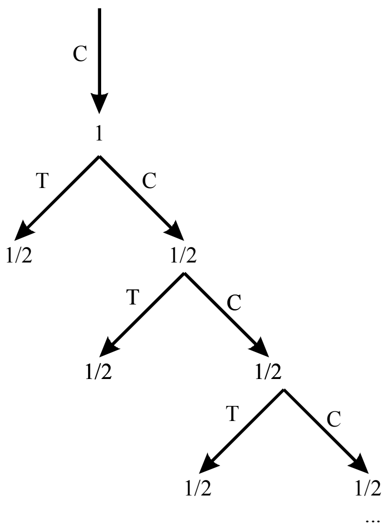
$$\sum_{k=s}^{\infty} k\, p(k)\,.$$

Si tratta in definitiva di introdurre la variabile aleatoria

X = numero di lanci perché escano tutte le facce,

calcolare per ogni n la probabilità $p(X=n)$ e infine calcolare il valor medio E(X) di X.

La moneta: s = 2
Per una moneta equa la soluzione si trova facilmente; si osservi il grafo a fianco, in cui è uscita (per esempio) la faccia C al primo lancio. Allora



$p(X = 2) = 1/2$
$p(X = 3) = 1/4$
$p(X = 4) = 1/8$
…
$p(X = n) = 1/2^{n-1}$

Il valor medio di X è dato dalla somma della serie

$$E(X) = \sum_{k=2}^{\infty} k \frac{1}{2^{k-1}}\,.$$

Come si calcola tale somma? Se poniamo p=1/2, allora

E(X) = $\sum_{k=2}^{\infty} k \frac{1}{2^{k-1}} = \sum_{k=2}^{\infty} k\, p^{k-1} = \sum_{k=2}^{\infty} \frac{d}{dp} p^k = \frac{d}{dp} \sum_{k=2}^{\infty} p^k =$

$= \frac{d}{dp}\left(\frac{1}{1-p} - p^0 - p^1 \right) = \frac{1}{(1-p)^2} - 1\,.$

Abbiamo dimostrato un risultato che ci sarà utile in seguito: $\sum_{k=2}^{\infty} k\, p^{k-1} = \frac{1}{(1-p)^2} - 1\,.$

Se poniamo p=1/2 allora il numero medio di lanci di una moneta equa per ottenere tutte e due le facce è 3.

Se la moneta non è equa ma è di trucco p, cioè esce T con probabilità p e C con probabilità q=1–p, ragioniamo nel seguente modo: X = n equivale al fatto che per n–1 lanci è uscito T e al n-esimo lancio esce C (con probabilità $p^{n-1}q$) oppure per n–1 lanci è uscito C e al n-esimo lancio esce T (con probabilità pq^{n-1}). Quindi

$p(\mathrm{X} = n) = p^{n-1}q + pq^{n-1}$

e risulta

E(X) = $\sum_{k=2}^{\infty} k\left(p^{k-1}q + pq^{k-1} \right) = q\sum_{k=2}^{\infty} kp^{k-1} + p\sum_{k=2}^{\infty} kq^{k-1}$

Per quello che abbiamo dimostrato precedentemente risulta

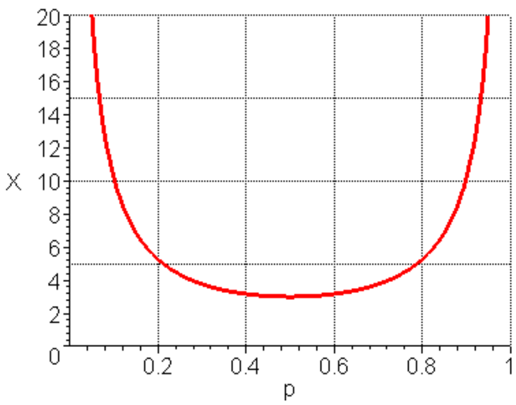
$\sum_{k=2}^{\infty} kp^{k-1} = \frac{1}{(1-p)^2} - 1$

$\sum_{k=2}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2} - 1$

e dunque

E(X) = $q\left(\frac{1}{(1-p)^2} - 1 \right) + p\left(\frac{1}{(1-q)^2} - 1 \right) = \frac{1}{p} + \frac{1}{q} - 1$

dove q=1–p. Il grafico di X al variare di p è il seguente.



Il grafico è ovviamente simmetrico rispetto a p = 0.5, in cui abbiamo il valor medio minimo; se p tende a 0 oppure a 1 allora E(X) tende a infinito, come è ragionevole supporre.

Il problema del tiratore

Apriamo una breve parentesi, per determinare un risultato che ci sarà utile in seguito.

Un tiratore colpisce il bersaglio con probabilità p: quante volte deve tirare per colpire il bersaglio? L'intuito ci dice che se la probabilità di centrare il bersaglio è, per esempio, 1/10, allora occorreranno in media 10 tiri per colpirlo. Ovviamente supponiamo che ogni tiro sia indipendente dai precedenti (e quindi trascuriamo la stanchezza, la delusione, …).

Sia X la variabile aleatoria

X = numero di tiri affinché il tiratore colpisca il bersaglio.

Dimostriamo che E(X) = $1/p$.

Infatti se il bersaglio è colpito per la prima volta dopo n tiri allora vuol dire che il tiratore ha sbagliato per $n-1$ volte (ogni volta con probabilità $q=1-p$) e quindi $p(X = n) = pq^{n-1}$.

Si tratta della *distribuzione geometrica* della variabile aleatoria X: pq^{n-1} è la probabilità che il "tempo di attesa" per il verificarsi dell'evento di probabilità p sia n.

Per il valor medio di X risulta

$$E(X) = \sum_{k=1}^{\infty} kpq^{k-1} = p\sum_{k=1}^{\infty} kq^{k-1}.$$

Poiché sappiamo che

$$\sum_{k=2}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2} - 1$$
allora
$$\sum_{k=1}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2} = \frac{1}{p^2}$$
dunque
$$\mathrm{E(X)} = p\frac{1}{p^2} = \frac{1}{p}.$$
Il tempo medio di attesa per il verificarsi di un evento di probabilità p è $1/p$.
Questo spiega perché la moneta va lanciata in media 3 volte per ottenere sia TESTA che CROCE:
3 = 1 (tempo di attesa per una faccia qualsiasi) + 2 (tempo medio di attesa perché si presenti l'altra faccia, che ha probabilità 1/2).
Lo stesso risultato può essere ottenuto in un modo molto più semplice, senza ricorrere alle serie: il tempo medio di attesa E(X) vale 1 con probabilità p (la probabilità di centrare il bersaglio al primo tiro) e vale 1+E(X) con probabilità $1-p$ (se al primo tiro non ho colpito il bersaglio posso immaginare di ripartire daccapo, e il tempo medio di attesa da quel punto in poi è sempre E(X), dunque il tempo d'attesa totale sarà 1+E(X)). Risulta quindi
$\mathrm{E(X)} = 1\cdot p + (1+\mathrm{E(X)})(1-p)$
da cui
$\mathrm{E(X)} = 1/p$.

Il dado: $s = 6$
Vediamo ora come risolvere il problema nel caso di un dado equo a 6 facce. Per risolvere il problema cambieremo strategia.
Lanciamo il dado una prima volta e otteniamo una certa uscita d_1. Quanti lanci occorre fare per avere un'uscita diversa da d_1? Introduciamo la variabile aleatoria
X_2 = numero di lanci per avere un'uscita diversa da d_1.
Il problema è identico a quello del tiratore: noi "tiriamo" con probabilità p=5/6 di "colpire il bersaglio" e q=1/6 di mancarlo, nel qual caso tireremo di nuovo, e così via. Allora
$\mathrm{E}(X_2) = 6/5 = 1.2$
Quindi occorre 1 lancio per la prima uscita e 1.2 lanci per la seconda uscita: siamo a 2.2 lanci. Il passaggio svolto ci porta subito alla soluzione: se sono già uscite 2

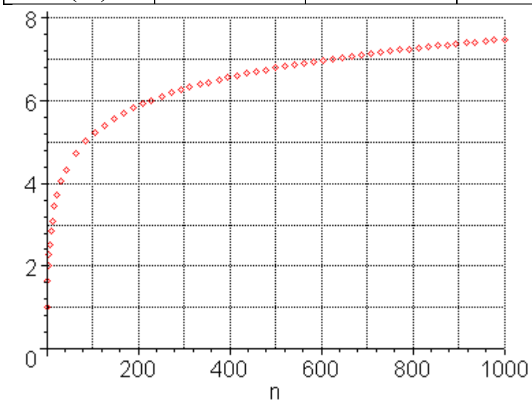
facce uscirà la terza faccia (diversa dalle precedenti) con probabilità p=4/6, e dunque dopo
$E(X_3) = 6/4 = 1.5$ lanci.
La quarta faccia uscirà con probabilità p=3/6, e dunque dopo
$E(X_4) = 6/3 = 2$ lanci.
La quinta faccia uscirà con probabilità p=2/6, e dunque dopo
$E(X_5) = 6/2 = 3$ lanci.
Infine la sesta faccia uscirà con probabilità p=1/6, e dunque dopo
$E(X_6) = 6/1 = 6$ lanci.
Ricordando che la prima faccia ovviamente esce al primo lancio, il che significa
$E(X_1) = 6/6 = 1$
la risposta al problema del collezionista per $s = 6$ è
$E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5) + E(X_6) = \frac{6}{6}+\frac{6}{5}+\frac{6}{4}+\frac{6}{3}+\frac{6}{2}+\frac{6}{1} = 6\sum_{k=1}^{6}\frac{1}{k} = 14.7$.
Occorre lanciare, in media, 14.7 volte un dado per vedere uscire tutte le facce.

Il caso generale
È immediata la generalizzazione.
Abbiamo un s-dado equo. Se indichiamo con Y_s la variabile aleatoria
Y_s = numero di lanci di un s-dado affinché escano tutte le facce
allora
$Y_s = X_1 + X_2 + \ldots + X_s$
Ho un dado a s facce; se sono già uscite k facce diverse, la probabilità che esca la (k+1)-esima faccia è $p = \frac{s-k+1}{s}$ e il valor medio del numero di lanci da effettuare è
$E(X_k) = \frac{s}{s-k+1}$
La soluzione generale è dunque
$E(Y_s) = \sum_{k=1}^{s} E(X_k) = \sum_{k=1}^{s}\frac{s}{s-k+1} = s\sum_{k=1}^{s}\frac{1}{k}$.
Magico, eh?
Volete sapere quante persone occorre prendere in media per avere rappresentati tutti i compleanni dell'anno?

$365\sum_{k=1}^{365}\frac{1}{k} \approx 2365.$

Viene ora voglia di chiedersi quale sia l'andamento di E(Y) al variare di *s*.

s	2	6	10	100	1000
E(Y)	3	14.7	29.3	518.7	7485.5



È sufficiente osservare che

$$\lim_{n\to\infty}\sum_{k=1}^{n}\frac{1}{k} = \ln(n) + \gamma,$$

dove $\gamma = 0.57721566\ldots$ è la cosiddetta costante di Eulero-Mascheroni. Quindi $E(Y_s) \approx s\ln(s)$.

Michele Impedovo