Data Science per psicologi

Corrado Caudek

2021-10-10

Indice

Capitolo 1

Simbolo di somma (sommatorie)

Le somme si incontrano costantemente in svariati contesti matematici e statistici quindi abbiamo bisogno di una notazione adeguata che ci consenta di gestirle. La somma dei primi n numeri interi può essere scritta come $1+2+\cdots+(n-1)+n$, dove '...' ci dice di completare la sequenza definita dai termini che vengono prima e dopo. Ovviamente, una notazione come $1+7+\cdots+73.6$ non avrebbe alcun senso senza qualche altro tipo di precisazione. In generale, nel seguito incontreremo delle somme nella forma

$$x_1 + x_2 + \cdots + x_n,$$

dove x_i è un numero che è stato definito altrove. La notazione precedente, che fa uso dei tre puntini di sospensione, è utile in alcuni contesti ma in altri risulta ambigua. Pertanto la notazione di uso corrente è del tipo

$$\sum_{i=1}^{n} x_i$$

e si legge "sommatoria per i che va da 1 a n di x_i." Il simbolo $\sum$ (lettera sigma maiuscola dell'alfabeto greco) indica l'operazione di somma, il simbolo x_i indica il generico addendo della sommatoria, le lettere 1 ed n indicano i cosiddetti *estremi della sommatoria*, ovvero l'intervallo (da 1 fino a n estremi inclusi) in cui deve variare l'indice i allorché si sommano gli addendi x_i. Solitamente l'estremo inferiore è 1 ma potrebbe essere qualsiasi altri numero $m < n$. Quindi

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n.$$

Per esempio, se i valori x sono $\{3, 11, 4, 7\}$, si avrà

$$\sum_{i=1}^{4} x_i = 3 + 11 + 4 + 7 = 25$$

laddove $x_1 = 3$, $x_2 = 11$, eccetera. La quantità x_i nella formula precedente si dice l'*argomento* della sommatoria, mentre la variabile i, che prende i valori naturali successivi indicati nel simbolo, si dice *indice* della sommatoria.

La notazione di sommatoria può anche essere fornita nella forma seguente

$$\sum_{P(i)} x_i$$

dove $P(i)$ è qualsiasi proposizione riguardante i che può essere vera o falsa. Quando è ovvio che si vogliono sommare tutti i valori di n osservazioni, la notazione può essere semplificata nel modo seguente: $\sum_i x_i$ oppure $\sum x_i$. Al posto di i si possono trovare altre lettere: $k, j, l, ...,$.

1.1 Manipolazione di somme

È conveniente utilizzare le seguenti regole per semplificare i calcoli che coinvolgono l'operatore della sommatoria.

1.1.1 Proprietà 1

La sommatoria di n valori tutti pari alla stessa costante a è pari a n volte la costante stessa:

$$\sum_{i=1}^{n} a = \underbrace{a + a + \cdots + a}_{n \text{ volte}} = na.$$

1.1.2 Proprietà 2 (proprietà distributiva)

Nel caso in cui l'argomento contenga una costante, è possibile riscrivere la sommatoria. Ad esempio con

$$\sum_{i=1}^{n} ax_i = ax_1 + ax_2 + \cdots + ax_n$$

è possibile raccogliere la costante a e fare $a(x_1 + x_2 + \cdots + x_n)$. Quindi possiamo scrivere

$$\sum_{i=1}^{n} ax_i = a \sum_{i=1}^{n} x_i.$$

1.1.3 Proprietà 3 (proprietà associativa)

Nel caso in cui

$$\sum_{i=1}^{n} (a + x_i) = (a + x_1) + (a + x_1) + \ldots (a + x_n)$$

si ha che

$$\sum_{i=1}^{n} (a + x_i) = na + \sum_{i=1}^{n} x_i.$$

È dunque chiaro che in generale possiamo scrivere

$$\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i.$$

1.1.4 Proprietà 4

Se deve essere eseguita un'operazione algebrica (innalzamento a potenza, logaritmo, ecc.) sull'argomento della sommatoria, allora tale operazione algebrica deve essere eseguita prima della somma. Per esempio,

$$\sum_{i=1}^{n} x_i^2 = x_1^2 + x_2^2 + \cdots + x_n^2 \neq \left(\sum_{i=1}^{n} x_i\right)^2.$$

1.1.5 Proprietà 5

Nel caso si voglia calcolare $\sum_{i=1}^{n} x_i y_i$, il prodotto tra i punteggi appaiati deve essere eseguito prima e la somma dopo:

$$\sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n,$$

infatti, $a_1 b_1 + a_2 b_2 \neq (a_1 + a_2)(b_1 + b_2)$.

1.2 Doppia sommatoria

È possibile incontrare la seguente espressione in cui figurano una doppia sommatoria e un doppio indice:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij}.$$

La doppia sommatoria comporta che per ogni valore dell'indice esterno, i da 1 ad n, occorre sviluppare la seconda sommatoria per j da 1 ad m. Quindi,

$$\sum_{i=1}^{3} \sum_{j=4}^{6} x_{ij} = (x_{1,4} + x_{1,5} + x_{1,6}) + (x_{2,4} + x_{2,5} + x_{2,6}) + (x_{3,4} + x_{3,5} + x_{3,6}).$$

Un caso particolare interessante di doppia sommatoria è il seguente:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j$$

Si può osservare che nella sommatoria interna (quella che dipende dall'indice j), la quantità x_i è costante, ovvero non dipende dall'indice (che è j). Allora possiamo estrarre x_i dall'operatore di sommatoria interna e scrivere

$$\sum_{i=1}^{n} \left(x_i \sum_{j=1}^{n} y_j \right).$$

Allo stesso modo si può osservare che nell'argomento della sommatoria esterna la quantità costituita dalla sommatoria in j non dipende dall'indice i e quindi questa quantità può essere estratta dalla sommatoria esterna. Si ottiene quindi

$$\sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j = \sum_{i=1}^{n}\left(x_i \sum_{j=1}^{n} y_j\right) = \sum_{i=1}^{n} x_i \sum_{j=1}^{n} y_j.$$

Esempio 1.1. Si verifichi quanto detto sopra nel caso particolare di $x = \{2,3,1\}$ e $y = \{1,4,9\}$, svolgendo prima la doppia sommatoria per poi verificare che quanto così ottenuto sia uguale al prodotto delle due sommatorie.

$$\begin{aligned}\sum_{i=1}^{3}\sum_{j=1}^{3} x_i y_j &= x_1y_1 + x_1y_2 + x_1y_3 + x_2y_1 + x_2y_2 + x_2y_3 + x_3y_1 + x_3y_2 + x_3y_3 \\ &= 2\times(1+4+9) + 3\times(1+4+9) + 2\times(1+4+9) = 84,\end{aligned}$$

ovvero

$$(2+3+1)\times(1+4+9) = 84.$$

1.3 Sommatorie (e produttorie) e operazioni vettoriali in R

Si noti che la notazione

$$\sum_{n=0}^{4} 3n$$

non è altro che un ciclo `for`

```
sum = 0;
for (n = 0; n <= 4; n++) {
  sum += 3 * n;
}
```

scritto in C, oppure

```
sum <- 0
for (n in 0:4) {
  sum = sum + 3 * n
}
sum
#> [1] 30
```

scritto in `R`. In maniera equivalente, e più semplice, possiamo scrivere

```
sum(3 * (0:4))
#> [1] 30
```

Allo stesso modo, la notazione

$$\prod_{n=1}^{4} 2n$$

è equivalente al ciclo `for`

```
prod <- 1
for (n in 1:4) {
  prod <- prod * 2 * n
}
prod
#> [1] 384
```

il che si può scrivere, più semplicemente, come

```
prod(2 * (1:4))
#> [1] 384
```