Azzolini Riccardo 2020-04-02

Distribuzione di Poisson

1 Distribuzione di Poisson

La distribuzione **di Poisson** di parametro $\lambda > 0$ è definita dalla densità

$$p(x) = \begin{cases} e^{-\lambda}\dfrac{\lambda^x}{x!} & x = 0, 1, \dots \\ 0 & \text{altrimenti} \end{cases}$$

Essa approssima la distribuzione binomiale $B(n, \frac{\lambda}{n})$ per $n \to \infty$, cioè per schemi successo-insuccesso con molte prove, ciascuna delle quali ha una probabilità di successo molto piccola.

Infatti, considerando una variabile $X \sim B(n, \frac{\lambda}{n})$,

$$\begin{aligned} p\{X = x\} = p(x) &= \binom{n}{x}\left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{n!}{x!\,(n-x)!} \cdot \frac{\lambda^x}{n^x}\left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\ &= \frac{\lambda^x}{x!} \cdot \frac{n!}{(n-x)!\,n^x}\left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\ &= \frac{\lambda^x}{x!} \cdot \frac{n(n-1)\cdots(n-x+1)}{n^x}\left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \end{aligned}$$

siccome per $n \to \infty$ valgono i seguenti limiti,

$$\begin{aligned} \frac{n(n-1)\cdots(n-x+1)}{n^x} \sim \frac{n^x}{n^x} &\to 1 \\ \left(1 - \frac{\lambda}{n}\right)^n &\to e^{-\lambda} && \text{(limite notevole } (1 + \tfrac{\alpha}{n})^n \to e^{\alpha}\text{, con } \alpha = -\lambda\text{)} \\ \left(1 - \frac{\lambda}{n}\right)^{-x} &\to 1^{-x} = 1 \end{aligned}$$

si dimostra che

$$p(x) = \frac{\lambda^x}{x!} \cdot \underbrace{\frac{n(n-1)\cdots(n-x+1)}{n^x}}_{\to 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-x}}_{\to 1} \to \frac{\lambda^x}{x!}\,e^{-\lambda}$$

cioè che la densità di X tende effettivamente a quella di Poisson.

Quest'approssimazione è utile in quanto è spesso scomodo gestire coefficienti binomiali per n grandi.

1.1 Problemi modellati da variabili di Poisson

Le variabili aleatorie di Poisson tendono a presentarsi naturalmente quando si vuole contare il numero di occorrenze di un fenomeno in un determinato intervallo di tempo o di spazio, purché tali occorrenze siano tra loro indipendenti. Alcuni esempi sono:

- Il numero di telefonate che giungono a un centralino in un periodo senza affollamento: considerando che le persone che potrebbero telefonare sono molte, ma la probabilità che ciascuna di esse effettivamente chiami è molto bassa (e, si suppone, indipendente dalle altre persone), questo problema è modellato da una variabile binomiale con n grande e p piccolo, che può essere ben approssimata da una variabile di Poisson.
- Il numero di complicazioni post-operatorie per un dato intervento chirurgico in un dato periodo di tempo, purché il numero di interventi effettuati nel periodo sia alto, e la probabilità di complicazione sia piccola: anche in questo caso, si ha una distribuzione binomiale con n grande e p piccolo, ben approssimata da una distribuzione di Poisson.
- Il numero di piante di una determinata specie infestante presenti in un appezzamento di terreno.

Detto in modo più preciso, la distribuzione di Poisson è indicata per modellare fenomeni che soddisfino (ragionevolmente) le seguenti condizioni:

- La probabilità che, in un intervallo di tempo/spazio di ampiezza h (piccola), si abbia $X = 1$ (esattamente un'occorrenza del fenomeno) è approssimativamente proporzionale ad h:
$$P\{X = 1\} = \nu h + o(h)$$
(dove ν è la costante di proporzionalità, e $o(h)$ indica una quantità trascurabile rispetto ad h).
- La probabilità di avere $X \geq 2$ in un intervallo di ampiezza h (piccola) è trascurabile relativamente alla probabilità di avere $X = 1$:
$$\frac{P\{X \geq 2\}}{P\{X = 1\}} = o(h)$$
(al limite per $h \to 0$, ciò significa che non si verificano due o più occorrenze del fenomeno nello stesso preciso istante).

- Il numero di occorrenze del fenomeno in un qualunque intervallo è indipendente dal numero di occorrenze che si hanno in altri intervalli incompatibili con esso (cioè separati, disgiunti da esso).

Teorema: Se un fenomeno verifica le tre condizioni appena enunciate, si può dimostrare che il numero di occorrenze X di tale fenomeno in un intervallo di tempo/spazio di ampiezza t segue la distribuzione di Poisson (di parametro $\lambda = \nu t$, dove ν indica la probabilità di avere un'occorrenza in un intervallo unitario, cioè con $t = 1$).

2 Somma di binomiali indipendenti

Se una variabile aleatoria X è definita come somma di m variabili aleatorie binomiali indipendenti, tutte aventi la stessa probabilità di successo p,

$$X = X_1 + \cdots + X_m \qquad X_i \sim B(n_i, p),\ i = 1, \dots, m$$

si dimostra che la distribuzione di X è $B(n_1 + \cdots + n_m, p)$.

Dimostrazione: Ciascuna variabile binomiale $X_i \sim B(n_i, p)$ può a sua volta essere considerata una somma di n_i variabili di Bernoulli indipendenti di parametro p:

$$X_i = X_{i1} + \cdots + X_{in_i} \qquad X_{ij} \sim B(1, p),\ j = 1, \dots, n_i$$

Allora, la somma di binomiali $X_1 + \cdots + X_m$ equivale alla somma delle Bernoulli di ciascuna delle binomiali:

$$X = X_{11} + \cdots + X_{1n_1} + \cdots + X_{m1} + \cdots + X_{mn_m}$$

Siccome le variabili binomiali di partenza erano indipendenti (per ipotesi), e ognuna di esse è una somma di Bernoulli indipendenti, tutte le Bernoulli considerate sono tra loro indipendenti. Perciò, la variabile X è appunto una somma di Bernoulli indipendenti di parametro p, e di conseguenza ha una distribuzione binomiale, il cui parametro n è dato dal numero di Bernoulli sommate, che è $n_1 + \cdots + n_m$: $X \sim B(n_1 + \cdots + n_m, p)$. $\square$

3 Somma di Poisson indipendenti

Dalla regola della somma di binomiali indipendenti si può ricavare anche una regola per la somma di Poisson indipendenti.

Per iniziare, si considera il caso più semplice: la somma di due Poisson aventi lo stesso parametro λ. Esse sono il limite per $n \to \infty$ della binomiale $B(n, \frac{\lambda}{n})$. Sommando due di queste binomiali, si ottiene $B(2n, \frac{\lambda}{n})$, che, con la sostituzione $m = 2n$, può essere riscritta come $B(m, \frac{2\lambda}{m})$, e allora il suo limite per $m \to \infty$ è la Poisson di parametro 2λ.

Procedendo allo stesso modo nel caso, più generale, di due Poisson (sempre indipendenti) con parametri diversi λ e μ, si ottiene come loro somma la Poisson di parametro $\lambda + \mu$.

3.1 Dimostrazione rigorosa

Una dimostrazione rigorosa di questa regola può essere effettuata operando direttamente sulla densità della distribuzione di Poisson (senza "sfruttare" la binomiale).

Siano U e V due variabili aleatorie di Poisson indipendenti, rispettivamente con parametri λ e μ, cioè aventi le seguenti densità (marginali):

$$p_U(x) = e^{-\lambda}\frac{\lambda^x}{x!} \qquad p_V(x) = e^{-\mu}\frac{\mu^x}{x!}$$

Si vuole calcolare la densità di $U + V$, cioè $P\{U + V = k\}$.

Innanzitutto, si osserva che l'evento $\{U + V = k\}$ può essere scomposto nella seguente unione di eventi elementari disgiunti, corrispondenti a tutte le coppie di possibili valori assunti da U e V tali che la somma sia k:

$$\{U + V = k\} = \bigcup_{i=0}^{k}\{U = i,\ V = k - i\}$$

La probabilità di $\{U = i,\ V = k - i\}$ è data (per definizione) dalla densità congiunta di U e V, che, siccome (per ipotesi) le due variabili sono indipendenti, corrisponde al prodotto delle densità marginali:

$$P\{U = i, V = k - i\} = p(i, k - i) = p_U(i)\, p_V(k - i)$$

Avendo fatto queste osservazioni, si può procedere al calcolo della densità di $U + V$:

$$\begin{aligned}
P\{U + V = k\} &= \sum_{i=0}^{k} P\{U = i,\ V = k - i\} \\
&= \sum_{i=0}^{k} p_U(i)\, p_V(k - i) \\
&= \sum_{i=0}^{k} e^{-\lambda}\frac{\lambda^i}{i!} \cdot e^{-\mu}\frac{\mu^{k-i}}{(k-i)!} \\
&= \sum_{i=0}^{k} e^{-(\lambda+\mu)}\frac{\lambda^i\mu^{k-i}}{(k-i)!\, i!} \\
&= e^{-(\lambda+\mu)}\sum_{i=0}^{k}\frac{\lambda^i\mu^{k-i}}{(k-i)!\, i!}
\end{aligned}$$

Moltiplicando e dividendo per $k!$, si ottiene il coefficiente binomiale $\binom{k}{i}$:

$$
\begin{aligned}
&= e^{-(\lambda+\mu)} \sum_{i=0}^{k} \frac{k!}{k!} \cdot \frac{\lambda^i \mu^{k-i}}{(k-i)!\, i!} \\
&= e^{-(\lambda+\mu)} \sum_{i=0}^{k} \frac{1}{k!} \cdot \frac{k!}{(k-i)!\, i!} \cdot \lambda^i \mu^{k-i} \\
&= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda^i \mu^{k-i}
\end{aligned}
$$

$\displaystyle\sum_{i=0}^{k} \binom{k}{i} \lambda^i \mu^{k-i}$ è lo sviluppo del binomio $(\lambda+\mu)^k$, quindi:

$$
\begin{aligned}
&= e^{-(\lambda+\mu)} \frac{1}{k!} (\lambda+\mu)^k \\
&= e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!}
\end{aligned}
$$

Si è allora dimostrato che la distribuzione di $U+V$ è la Poisson di parametro $\lambda+\mu$. $\square$