

# Variabili aleatorie discrete – Casi notevoli

## 1 Variabili aleatorie binomiali

Una variabile aleatoria **binomiale** conta il numero di successi in uno schema successo-insuccesso (con rimpiazzo,<sup>1</sup> che è il tipo di schemi successo-insuccesso visti finora).

### 1.1 Problema: bulloni difettosi

*Problema:* I bulloni prodotti da un'azienda sono difettosi con una probabilità del 20 %, e vengono messi in commercio in confezioni di 3 pezzi ciascuna. Qual è la probabilità che una confezione contenga al più 1 bullone difettoso?

Se si definisce una variabile aleatoria  $X$ , il cui valore è il numero di bulloni difettosi presenti nella confezione, l'evento considerato dal problema è  $\{X \leq 1\}$ , che, siccome  $X$  è discreta (assume i valori 0, 1, 2, 3), può essere riscritto come unione di eventi elementari disgiunti:

$$\begin{aligned}\{X \leq 1\} &= \{X = 0\} \cup \{X = 1\} \\ P\{X \leq 1\} &= P\{X = 0\} + P\{X = 1\}\end{aligned}$$

Supponendo che ciascun bullone sia o meno difettoso indipendentemente dagli altri, questa situazione è uno schema successo-insuccesso: si effettuano  $n = 3$  “prove” (corrispondenti ai bulloni nella confezione), ciascuna delle quali ha una probabilità di “successo”<sup>2</sup> (cioè che il bullone sia difettoso) pari a  $p = 20 \% = 0.2$ , e in questo modo si ottiene complessivamente un certo numero  $k$  di successi (da 0 a 3). Per risolvere il problema, bisogna ricavare la probabilità che  $k$  sia 0 oppure 1.

---

<sup>1</sup>Uno schema successo-insuccesso si dice “con rimpiazzo” se ciascuna prova ha la stessa probabilità di successo, indipendentemente dagli esiti delle prove precedenti. Nella situazione dell'estrazione di palline da un'urna, ciò corrisponde a reinserire (rimpiazzare, appunto) nell'urna la pallina estratta prima di procedere con l'estrazione successiva, in modo che il numero di palline associate al successo e di quelle associate all'insuccesso rimanga invariato.

<sup>2</sup>I termini “successo” e “insuccesso” non vanno interpretati con una qualche connotazione positiva o negativa: sono semplicemente i nomi che si danno, per convenzione, ai due possibili esiti di una prova. Infatti, è indifferente quale dei due esiti sia chiamato “successo” e quale invece “insuccesso”: in questo caso, ad esempio, si potrebbe equivalentemente definire successo un bullone non difettoso, che ha una probabilità dell'80 %, e calcolare la probabilità che il numero di successi sia  $\geq 2$ .

Per calcolare la probabilità di avere  $k$  successi, è necessario considerare tutte le possibili configurazioni di  $k$  bulloni difettosi (successi) e  $n - k = 3 - k$  non difettosi (insuccessi), e sommare le probabilità di ciascuna di queste configurazioni.

1. Siano  $B = \{B_1, B_2, B_3\}$  i tre bulloni in una confezione. Sceglierne una configurazione con  $k$  difettosi significa costruire un sottoinsieme di  $k$  elementi di  $B$ , quindi ci sono  $\binom{n}{k} = \binom{3}{k}$  modi diversi di farlo. Infatti, si osserva che non importa l'ordine: ad esempio, dire che sono difettosi  $B_1, B_3$  è esattamente uguale a dire che lo sono  $B_3, B_1$ .
2. Una volta individuata una configurazione di  $k$  bulloni difettosi e  $n - k$  non difettosi, la probabilità di trovare tale configurazione in una confezione è il prodotto delle probabilità dei singoli bulloni:

$$\overbrace{p^k}^{\text{difettosi}} \underbrace{(1-p)^{n-k}}_{\text{non difettosi}}$$

Complessivamente, la probabilità di avere  $k$  successi su  $n$  prove con probabilità di successo  $p$  è allora:

$$P\{X = k\} = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & k = 0, 1, \dots, n \\ 0 & \text{altrimenti (} k \text{ non è un numero valido di successi)} \end{cases}$$

In generale, una variabile aleatoria con densità data da questa formula è detta **binomiale** di parametri  $n$  e  $p$ , e la sua distribuzione è indicata con  $B(n, p)$ .

Nel caso di questo problema,  $X$  è una variabile aleatoria binomiale di parametri  $n = 3$  e  $p = 0.2$ : in simboli,  $X \sim B(3, 0.2)$ . Si può allora calcolare la probabilità richiesta:

$$\begin{aligned} P\{X \leq 1\} &= P\{X = 0\} + P\{X = 1\} \\ &= \binom{3}{0} 0.2^0 0.8^3 + \binom{3}{1} 0.2^1 0.8^2 \\ &= 1 \cdot 1 \cdot 0.512 + 3 \cdot 0.2 \cdot 0.64 \\ &= 0.512 + 0.384 \\ &= 0.896 \\ &= 89.6 \% \end{aligned}$$

## 1.2 Variabili aleatorie di Bernoulli

Come caso particolare, per  $n = 1$ , la densità della distribuzione binomiale  $B(n, p)$  si riduce a

$$\begin{aligned} P\{X = k\} = p(x) &= \begin{cases} \binom{1}{k} p^k (1-p)^{1-k} & k = 0, 1 \\ 0 & \text{altrimenti} \end{cases} \\ &= \begin{cases} \binom{1}{0} p^0 (1-p)^1 & k = 0 \\ \binom{1}{1} p^1 (1-p)^0 & k = 1 \\ 0 & \text{altrimenti} \end{cases} \\ &= \begin{cases} 1-p & k = 0 \\ p & k = 1 \\ 0 & \text{altrimenti} \end{cases} \end{aligned}$$

cioè

$$p(1) = p \quad p(0) = 1 - p$$

Tale densità corrisponde a una variabile aleatoria che rappresenta una singola prova di uno schema successo-insuccesso: essa assume solo i valori 1 (successo, con probabilità  $p$ ) e 0 (insuccesso, con probabilità  $1-p$ ). Allora,  $B(1, p)$  prende il nome di distribuzione (o legge) **di Bernoulli** (ricordando che gli schemi successo-insuccesso sono anche chiamati schemi di Bernoulli).

Si osserva anche che contare il numero di successi in  $n$  prove corrisponde a sommare i valori di  $n$  variabili di Bernoulli indipendenti, quindi una variabile aleatoria binomiale  $X \sim B(n, p)$  può essere interpretata come una somma

$$X = X_1 + \dots + X_n$$

dove  $X_i \sim B(1, p)$  per  $i = 1, \dots, n$ .

## 1.3 Problema: prenotazioni per un volo

*Problema:* Sapendo che il 30 % dei passeggeri che hanno prenotato non si presentano alla partenza, una compagnia aerea accetta fino a 28 prenotazioni su un volo con una capienza di 24 posti. Qual è la probabilità che almeno un passeggero che ha regolarmente prenotato resti a terra?

Si definisce una variabile aleatoria discreta  $X$  che conta il numero di persone che si presentano, e si considera la probabilità

$$P\{X \geq 25\} = P\{X = 25\} + P\{X = 26\} + P\{X = 27\} + P\{X = 28\}$$

Supponendo che il comportamento (presentarsi o meno alla partenza) dei passeggeri sia indipendente (un'ipotesi non molto realistica: i passeggeri viaggiano spesso in gruppo), il problema corrisponde a uno schema successo-insuccesso, nel quale si sceglie considera successo il presentarsi di un passeggero (che avviene con probabilità  $1 - 30\% = 0.7$ ), quindi  $X \sim B(28, 0.7)$ . La probabilità richiesta viene allora calcolata nel seguente modo:

$$\begin{aligned}
 P\{X \geq 25\} &= P\{X = 25\} + P\{X = 26\} + P\{X = 27\} + P\{X = 28\} \\
 &= \binom{28}{25} 0.7^{25} 0.3^3 + \binom{28}{26} 0.7^{26} 0.3^2 \\
 &\quad + \binom{28}{27} 0.7^{27} 0.3^1 + \binom{28}{28} 0.7^{28} 0.3^0 \\
 &\approx 3276 \cdot 1.34 \cdot 10^{-4} \cdot 0.027 + 378 \cdot 9.39 \cdot 10^{-5} \cdot 0.09 \\
 &\quad + 28 \cdot 6.57 \cdot 10^{-5} \cdot 0.3 + 1 \cdot 4.60 \cdot 10^{-5} \cdot 1 \\
 &\approx 0.0119 + 0.00319 + 0.000552 + 0.0000460 \\
 &\approx 0.0157 = 1.57\%
 \end{aligned}$$

## 2 Variabili aleatorie ipergeometriche

Una variabile aleatoria **ipergeometrica** conta il numero di successi in uno schema successo-insuccesso *senza rimpiazzo*, nel quale le prove hanno un certo numero di possibili risultati considerati “successo”, e altri considerati “insuccesso”, e il risultato di una prova non si può ripresentare nelle successive (come se si estraessero palline da un'urna senza rimpiazzarle), quindi le prove *non* sono indipendenti.

### 2.1 Problema: lettori di nastri

*Problema:* La memoria secondaria di un calcolatore è composta da 30 lettori di nastri, ciascuno dei quali contiene 100 file (quindi ci sono 3000 file in totale). Un programma accede a 28 file diversi. Qual è la probabilità che esso non debba usare il lettore numero 1?

Considerando come successo l'accesso a un file sul lettore 1, e definendo una variabile aleatoria  $X$  che conta tali successi, si è interessati all'evento  $\{X = 0\}$ .

Siccome lo stesso file non può essere letto più volte, il problema è uno schema successo-insuccesso senza rimpiazzo: gli  $n_1 = 100$  file sul lettore 1 corrispondono al successo, e i restanti  $n_2 = 2900$  corrispondono all'insuccesso; si effettuano  $n = 28$  prove, e si vogliono  $k = 0$  successi.

- Tutti i casi favorevoli, nei quali cui si ottengono  $k = 0$  successi, possono essere enumerati scegliendo  $k = 0$  file tra gli  $n_1 = 100$  sul lettore 1, e  $n - k = 28 - 0 = 28$  tra gli  $n_2 = 2900$  sugli altri lettori, quindi il numero di tali casi è:

$$\binom{n_1}{k} \binom{n_2}{n-k} = \binom{100}{0} \binom{2900}{28}$$

- Ogni caso possibile corrisponde a una scelta di  $n = 28$  file tra tutti gli  $n_1 + n_2 = 3000$  disponibili, che può essere effettuata in

$$\binom{n_1 + n_2}{n} = \binom{3000}{28}$$

modi diversi.

Allora, la probabilità di avere  $k$  successi è, in generale,

$$P\{X = k\} = \begin{cases} \frac{\binom{n_1}{k} \binom{n_2}{n-k}}{\binom{n_1 + n_2}{n}} & k = 0, 1, \dots, n \\ 0 & \text{altrimenti} \end{cases}$$

perciò  $X$  è detta variabile aleatoria **ipergeometrica** di parametri  $n_1 = 100$ ,  $n_2 = 2900$  e  $n = 28$ , e, nello specifico, la probabilità di non usare il lettore 1 è:

$$P\{X = 0\} = \frac{\overbrace{\binom{100}{0}}^{=1} \binom{2900}{28}}{\binom{3000}{28}} \approx 0.385 = 38.5 \%$$

*Nota:* Per questa formula, è utile adottare la convenzione che  $\binom{a}{b} = 0$  se  $a < b$ : così, tale formula rimane valida nei casi in cui  $n_1 < k$  e/o  $n_2 < n - k$ , per i quali dà allora probabilità 0, corrispondente al fatto che non siano disponibili abbastanza elementi diversi per ottenere  $k$  successi e/o  $n - k$  insuccessi.

### 2.1.1 Probabilità con rimpiazzo

Se fosse possibile accedere più volte allo stesso file, lo schema successo-insuccesso diventerebbe con rimpiazzo. Per ogni accesso a un file, la probabilità che esso sia sul lettore 1 sarebbe allora sempre  $\frac{1}{30}$ , quindi  $X \sim B(28, \frac{1}{30})$ , e perciò:

$$P\{X = 0\} = \underbrace{\binom{28}{0}}_{=1} \underbrace{\left(\frac{1}{30}\right)^0}_{=1} \left(\frac{29}{30}\right)^{28} \approx 0.387 = 38.7 \%$$

Si osserva che la probabilità ottenuta con rimpiazzo (38.7 %) è molto simile a quella senza rimpiazzo (38.5 %). Infatti, il numero di file (3000) è molto superiore al numero di letture (28), quindi escludere un file dalle letture successive dopo averlo letto una volta cambia solo di poco la probabilità di successo.

### 3 Variabili aleatorie geometriche

In uno schema successo-insuccesso con rimpiazzo, una variabile aleatoria **geometrica** conta il numero di insuccessi che si riscontrano prima di ottenere il primo successo.

Sia  $p$  la probabilità di successo di una singola prova (e quindi  $1 - p$  quella di insuccesso). La probabilità che i risultati delle prime  $k$  prove siano insuccessi è  $(1 - p)^k$ , e la probabilità che poi la  $(k + 1)$ -esima prova abbia successo è  $p$ . Allora, una variabile aleatoria geometrica  $X$  di parametro  $p$  ha la seguente densità:

$$P\{X = k\} = \begin{cases} (1 - p)^k p & k = 0, 1, \dots \\ 0 & \text{altrimenti} \end{cases}$$

*Nota:* La scelta di contare il numero di insuccessi precedenti al successo, e non il numero della prova che dà il primo successo (ad esempio, per il lancio di una moneta, “esce croce 3 volte prima che esca testa”, e non “esce testa per la prima volta al 4° lancio”) ha il vantaggio di ammettere che la variabile assuma il valore  $k = 0$ , permettendo così l’uso di  $k$ , invece che  $k - 1$ , nella formula della densità (ciò semplifica i calcoli). Infatti, ha senso dire che ci sono 0 insuccessi prima del primo successo (significa che il successo è immediato), mentre le prove sono solitamente numerate a partire da 1, quindi non si può avere successo alla prova 0.

#### 3.1 Mancanza di memoria

*Problema:* In uno schema successo-insuccesso, si suppone di non aver avuto alcun successo nelle prime  $k$  prove. Qual è la probabilità di dover attendere ancora  $m$  prove per avere il primo successo, cioè che tale successo avvenga alla  $(k + m)$ -esima prova?

Indicando con la variabile aleatoria  $T$  l’istante di primo successo, e sapendo che  $T > k$  (altrimenti si avrebbe avuto successo in una delle prime  $k$  prove), la richiesta del problema è la probabilità condizionata  $P\{T = k + m \mid T > k\}$ .

In base alla definizione data prima,  $X = T - 1$  (il numero di insuccessi prima del primo successo) è una variabile aleatoria geometrica di parametro  $p$  (dove  $p$  è la probabilità

di successo nella singola prova). Allora, la probabilità richiesta può essere riformulata come:

$$\begin{aligned} P\{T = k + m \mid T > k\} &= P\{T - 1 = k + m - 1 \mid T - 1 > k - 1\} \\ &= P\{X = k + m - 1 \mid X > k - 1\} \\ &= P\{X = k + m - 1 \mid X \geq k\} \end{aligned}$$

Per calcolarla, sarà necessario applicare la definizione di probabilità condizionata:

$$\begin{aligned} P\{X = k + m - 1 \mid X \geq k\} &= \frac{P\{X = k + m - 1, X \geq k\}}{P\{X \geq k\}} \\ &= \frac{P\{X = k + m - 1\}}{P\{X \geq k\}} \end{aligned}$$

(perché  $\{X = k + m - 1\} \cap \{X \geq k\} = \{X = k + m - 1\}$ , per assorbimento).

Considerando che l'evento  $\{X \geq k\}$  corrisponde a un'unione infinita numerabile di eventi elementari disgiunti,

$$\{X \geq k\} = \bigcup_{i=k}^{\infty} \{X = i\}$$

la sua probabilità può essere ricavata nel modo seguente:

$$\begin{aligned} P\{X \geq k\} &= \sum_{i=k}^{\infty} P\{X = i\} \\ &= \sum_{i=k}^{\infty} p(1-p)^i \\ &= \sum_{i=0}^{\infty} p(1-p)^{k+i} \\ &= \sum_{i=0}^{\infty} p(1-p)^k(1-p)^i \\ &= (1-p)^k \sum_{i=0}^{\infty} p(1-p)^i \end{aligned}$$

$p(1-p)^i$  è la densità di probabilità di  $X$ , quindi, per le proprietà delle densità discrete, la somma  $\sum_{i=0}^{\infty} p(1-p)^i$ , che considera tutti gli eventi elementari ( $\{X = 0\}, \{X = 1\}, \dots$ ), deve valere 1, e di conseguenza:

$$P\{X \geq k\} = (1-p)^k$$

Infine, procedendo con il calcolo della soluzione,

$$\begin{aligned}
 P\{X = k + m - 1 \mid X \geq k\} &= \frac{P\{X = k + m - 1\}}{P\{X \geq k\}} \\
 &= \frac{p(1-p)^{k+m-1}}{(1-p)^k} \\
 &= \frac{p(1-p)^k(1-p)^{m-1}}{(1-p)^k} \\
 &= p(1-p)^{m-1}
 \end{aligned}$$

si conclude che

$$P\{X = k + m - 1 \mid X \geq k\} = P\{X = m - 1\}$$

o, equivalentemente, che

$$P\{T = k + m \mid T > k\} = P\{T = m\}$$

cioè che la probabilità di dover aspettare ancora  $m$  prove per ottenere il primo successo è la stessa che si avrebbe se i  $k$  insuccessi iniziali non si fossero verificati. Infatti, siccome le prove sono indipendenti, gli insuccessi nelle prime  $k$  prove non alterano la probabilità di successo nelle prove successive.

Quest'uguaglianza, che in forma generale è

$$P\{X = k + m \mid X \geq k\} = P\{X = m\} \quad \forall k, m = 0, 1, \dots$$

vale per qualunque variabile aleatoria geometrica, ed è chiamata proprietà di **mancanza di memoria**.

## 4 Distribuzione binomiale negativa

La definizione del coefficiente binomiale,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k!} \quad \forall n, k \in \mathbb{N}$$

può essere generalizzata a  $n$  reali:

$$\binom{\beta}{k} = \frac{\beta(\beta-1)\cdots(\beta-k+1)}{k!} \quad \forall \beta \in \mathbb{R}, k \in \mathbb{N}$$

È così possibile definire la distribuzione **binomiale negativa** di parametri  $\alpha \in \mathbb{R}$ ,  $\alpha > 0$  e  $p \in [0, 1]$ , caratterizzata dalla seguente densità di probabilità discreta:

$$p(x) = \begin{cases} \binom{\alpha + x - 1}{x} p^\alpha (1-p)^x & x = 0, 1, \dots \\ 0 & \text{altrimenti} \end{cases}$$



Come significato intuitivo, nel caso in cui  $\alpha \in \mathbb{N}$ , una variabile aleatoria binomiale negativa conta il numero di insuccessi che si hanno prima di ottenere  $\alpha$  successi, in uno schema successo-insuccesso con rimpiazzo nel quale la probabilità di successo è  $p$ . Infatti, per ottenere  $\alpha$  successi e  $x$  insuccessi sono necessarie  $\alpha + x$  prove: i risultati delle prime  $\alpha + x - 1$  saranno una qualsiasi sequenza di  $\alpha - 1$  successi e  $x$  insuccessi, mentre quello dell' $(\alpha + x)$ -esima prova sarà sicuramente l' $\alpha$ -esimo successo. Allora,

- ci sono  $\binom{\alpha+x-1}{x}$  modi diversi di scegliere quali  $x$  prove della sequenza saranno insuccessi,
- ciascuna di queste sequenze ha probabilità  $p^\alpha(1-p)^x$ ,

quindi si ottiene appunto la formula della densità mostrata prima.

Come suggerito da questo significato intuitivo, la distribuzione binomiale negativa è una generalizzazione della distribuzione geometrica, che infatti si ritrova ponendo a 1 il numero di successi ( $\alpha = 1$ ):

$$\binom{1+x-1}{x} p^1 (1-p)^x = \underbrace{\binom{x}{x}}_{=1} p (1-p)^x = p (1-p)^x$$

#### 4.1 Dimostrazione che $p(x)$ è una densità

Per verificare che la funzione

$$p(x) = \begin{cases} \binom{\alpha+x-1}{x} p^\alpha (1-p)^x & x = 0, 1, \dots \\ 0 & \text{altrimenti} \end{cases}$$

sia effettivamente una densità di probabilità discreta, è necessario dimostrare che la sua somma su tutti gli eventi elementari è 1:

$$\sum_{x=0}^{\infty} p(x) = 1$$

La dimostrazione si basa sulla seguente serie:

$$\sum_{x=0}^{\infty} \binom{\alpha+x-1}{x} t^x = \frac{1}{(1-t)^\alpha}$$

Ponendo  $t = 1 - p$ , si ha:

$$\sum_{x=0}^{\infty} \binom{\alpha+x-1}{x} (1-p)^x = \frac{1}{(1-(1-p))^\alpha} = \frac{1}{p^\alpha}$$

Allora:

$$\begin{aligned}\sum_{x=0}^{\infty} p(x) &= \sum_{x=0}^{\infty} \binom{\alpha + x - 1}{x} \overbrace{p^\alpha}^{\text{costante}} (1-p)^x \\ &= p^\alpha \sum_{x=0}^{\infty} \binom{\alpha + x - 1}{x} (1-p)^x \\ &= p^\alpha \frac{1}{p^\alpha} = 1 \quad \square\end{aligned}$$