DISTRIBUZIONE MULTINOMIALE

Consideriamo n prove indipendenti.

Per ogni prova abbiamo r risultati possibili
$$\begin{cases} \text{RISULTATO } 1 & \text{con probabilità } p_1 \geq 0 \\ \vdots & \\ \text{RISULTATO } r & \text{con probabilità } p_r \geq 0 \end{cases}$$

I valori $p_1, \dots, p_r$ sono gli stessi per ogni prova; inoltre $p_1 + \dots + p_r = 1$.

In generale saremo interessati a calcolare probabilità di questo tipo:

$$P\begin{pmatrix} k_1 \text{ volte risultato } 1 \\ \vdots \\ k_r \text{ volte risultato } r \end{pmatrix}$$

dove

$$\begin{cases} k_1, \dots, k_r \geq 0 \text{ interi} \\ k_1 + \dots + k_r = n \end{cases}$$

OSSERVAZIONE

Se si ha $r=2$ si recupera il caso della Binomiale; ad esempio

RISULTATO 1 $\longleftrightarrow$ SUCCESSO $(p_1 = p)$

RISULTATO 2 $\longleftrightarrow$ FALLIMENTO $(p_2 = 1-p)$

(perché $p_1 + p_2 = 1$)

Consideriamo la seguente sequenza di risultati:

$$(\underbrace{R_1,\dots,R_1}_{k_1 \text{ volte}},\ \underbrace{R_2,\dots,R_2}_{k_2 \text{ volte}},\ \dots\dots,\ \underbrace{R_r,\dots,R_r}_{k_r \text{ volte}}).$$

Allora, per indipendenza delle prove, questa sequenza ha probabilità $p_1^{k_1}\cdots p_r^{k_r}$.

Ovviamente si ottiene la stessa probabilità per qualsiasi altra sequenza con "k_1 volte R_1, ..., k_r volte R_r". Quindi la grandezza che vogliamo calcolare è

$$P\begin{pmatrix} k_1 \text{ volte } R_1 \\ \vdots \\ k_r \text{ volte } R_r \end{pmatrix} = \underbrace{\#\left\{\text{sequenze con "}k_1 \text{ volte } R_1, \dots, k_r \text{ volte } R_r\text{"}\right\}}_{\substack{\text{Si può verificare che} \\ = \frac{n!}{k_1!\cdots k_r!}}} \cdot p_1^{k_1}\cdots p_1^{k_r}$$

$\frac{n!}{k_1!\cdots k_r!}$ ↔ è detto COEFFICIENTE MULTINOMIALE (per $r=2$ è un COEFFICIENTE BINOMIALE).

OSSERVAZIONE

Se in ogni prova i risultati $R_1, \dots, R_r$ sono equiprobabili, cioè $p_1 = \dots = p_r = \frac{1}{r}$, si ha

$$P\begin{pmatrix} k_1 \text{ volte } R_1 \\ \vdots \\ k_r \text{ volte } R_r \end{pmatrix} = \frac{n!}{k_1! \cdots k_r!} \left(\frac{1}{r}\right)^{\overbrace{k_1 + \dots + k_r}^{= n}} =$$

$$= \frac{n!}{k_1! \cdots k_r!} \left(\frac{1}{r}\right)^{n}$$

OSSERVAZIONE

Per $r=2$

successo

$$P\begin{pmatrix} k_1 \text{ volte } R_1 \\ k_2 \text{ volte } R_2 \end{pmatrix} = \frac{n!}{k_1!\,(n-k_1)!} p_1^{k_1} (1-p_1)^{n-k_1} = \binom{n}{k_1} p_1^{k_1} (1-p_1)^{n-k_1}$$

(con $0 \le k_1 \le n$).

$\boxed{\begin{array}{l} p_1 = 1 - p_2 \text{ perché } p_1 + p_2 = 1 \\ k_2 = n - k_1 \text{ perché } k_1 + k_2 = n \end{array}}$

Si recupera la BINOMIALE

ESERCIZIO

Un'urna contiene 3 palline bianche, 3 rosse e 2 nere. Si estraggono 4 palline a caso, una alla volta e con reinserimento. Calcolare le probabilità dei seguenti eventi:

1) Viene estratta la sequenza di colori (rosso, nero, nero, bianco).

2) vengono estratte esattamente 2 palline rosse e 1 nera in qualsiasi ordine.

3) vengono estratte esattamente 2 palline rosse in un qualsiasi ordine.

COMMENTO Negli esercizi generalmente sottointenderemo "esattamente" e "in qualsiasi ordine".

SVOLGIMENTO

Si hanno prove indipendenti perché le estrazioni sono <u>con</u> reinserimento.

In ogni prova abbiamo 3 risultati con probabilità $p_B = \frac{3}{8}$, $p_R = \frac{3}{8}$, $p_N = \frac{2}{8}$ (uso notazioni diverse da p_1, p_2, p_3)

1) Con notazioni ovvie la probabilità richiesta è

$$P(R_1 \cap N_2 \cap N_3 \cap B_4) = P(R_1)P(N_2)P(N_3)P(B_4) = p_R\, p_N\, p_N\, p_B = p_B\, p_R\, p_N^2 = \boxed{\frac{3}{8} \cdot \frac{3}{8} \left(\frac{2}{8}\right)^2} = \frac{36}{4096}.$$

<u>OSS</u>

L'espressione $p_B\, p_R\, p_N^2$ può essere visto come $p_1^{k_1} p_2^{k_2} p_3^{k_2}$ con $k_1 = 1$, $k_2 = 1$, $k_3 = 2$ ((1B, 1R, 2N) in un ordine fissato). Abbiamo una parte della formula della multinomiale senza il coefficiente multinomiale (del resto è una sequenza fissata).

2) $P(\text{"2R e 1N" in un qualsiasi ordine}) \overset{\nabla}{=}$ 4 palline estratte

$= P(\text{"1B, 2R, 1N" in qualsiasi ordine}) = \boxed{\frac{4!}{1!\,2!\,1!}} \underbrace{\left(\frac{3}{8}\right)^1 \left(\frac{3}{8}\right)^2 \left(\frac{2}{8}\right)^1}_{= \frac{54}{4096}} = \frac{648}{4096}$

$\frac{4!}{1!\,2!\,1!} = \frac{4 \cdot 3 \cdot \cancel{2}}{\cancel{2}} = 12$

3) $P(\text{"2R" in un qualsiasi ordine}) \overset{\nabla}{=}$ 4 palline estratte, contano solo rosse e non rosse (senza distinguere tra bianche e nere)

$\overset{\nabla}{=} \binom{4}{2} \left(\frac{3}{8}\right)^2 \left(1 - \frac{3}{8}\right)^{4-2} = 6 \left(\frac{3}{8}\right)^2 \left(\frac{5}{8}\right)^2 = \frac{1350}{4096}$,

#ROSSE estratte $\sim Bin\left(n=4, p=p_R=\frac{3}{8}\right)$

OSS. Se volessimo tenere conto dei 3 colori si ha

$P(\text{"2R"}) = \underbrace{P(\text{"2R, 1B, 1N"})}_{} + \underbrace{P(\text{"2R, 2B"})}_{} + \underbrace{P(\text{"2R, 2N"})}_{} = \frac{648 + 486 + 216}{4096} = \frac{1350}{4096}$

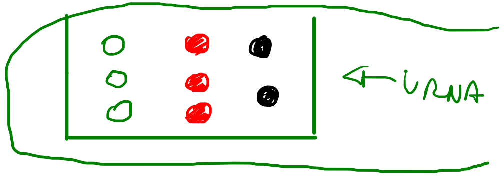
calcolata alla domanda precedente

da calcolare con la multinomiale

Calcoli più complicati, metodo meno conveniente

ESERCIZIO PRECEDENTE CON ESTRAZIONI "SENZA REINSERIMENTO" (ANZICHÉ "CON REINSERIMENTO")

1) $P(R_1 \cap N_2 \cap N_3 \cap B_4) = \underbrace{P(R_1)}_{=\frac{3}{8}} \underbrace{P(N_2 | R_1)}_{=\frac{2}{7}} \underbrace{P(N_3 | R_1 \cap N_2)}_{=1/6} \underbrace{P(B_4 | R_1 \cap N_2 \cap N_3)}_{=\frac{3}{5}} = \frac{3}{8} \cdot \frac{\cancel{2}}{7} \cdot \frac{1}{\cancel{6}} \cdot \frac{\cancel{3}}{5} = \frac{3}{280}$



2) $P(\text{"}2R \text{ e } 1N\text{"}) = P(\text{"}1B, 2R, 1N\text{"}) =$

$$= \frac{\binom{3}{1}\binom{3}{2}\binom{2}{1}}{\binom{8}{4}} = \frac{3 \cdot 3 \cdot 2}{70} = \frac{18}{70} = \frac{9}{35}$$

ROSSE e NON ROSSE SENZA DISTINGUERE BIANCHE E NERE

3) $P(\text{"}2R\text{"}) = \frac{\binom{3}{2}\binom{5}{2}}{\binom{8}{4}} = \frac{3 \cdot 10}{70} = \frac{30}{70} = \frac{3}{7}$

$$\binom{8}{4} = \frac{8!}{4!(8-4)!} = \frac{8 \cdot 7 \cdot 6 \cdot 5 \cdot \cancel{4!}}{4! \, \cancel{4!}} = \frac{\cancel{8} \cdot 7 \cdot \overset{2}{\cancel{6}} \cdot 5}{\cancel{4} \cdot \cancel{3} \cdot \cancel{2}} = 70$$

$$P(\text{"}2R\text{"}) = \underbrace{P(\text{"}2R, 1B, 1N\text{"})}_{=18/70} + \underbrace{P(\text{"}2R, 2B\text{"})}_{= \frac{\binom{3}{2}\binom{3}{2}\binom{2}{0}}{\binom{8}{4}} = \frac{9}{70}} + \underbrace{P(\text{"}2R, 2N\text{"})}_{= \frac{\binom{3}{0}\binom{3}{2}\binom{2}{2}}{\binom{8}{4}} = \frac{3}{70}} = \frac{18+9+3}{70} = \frac{30}{70} = \frac{3}{7}$$

ESERCIZIO

Si lancia 5 volte un dado equo. Calcolare la probabilità che escano (esattamente) "2 volte $\boxed{1}$ e 1 volta $\boxed{3}$" (in un qualsiasi ordine).

RISPOSTA

Abbiamo 5 prove indipendenti (i 5 lanci) e 3 risultati in ogni prova:

$\boxed{1}$ con prob. $p_1 = 1/6$ (2 volte)

$\boxed{3}$ con prob. $p_2 = 1/6$ (1 volta)

$\boxed{2|4|5|6}$ con prob. $p_3 = \frac{4}{6}$ (2 volte)

dedotto

La probabilità richiesta è (applicazione diretta della formula della multinomiale)

$$\frac{5!}{2!\,1!\,2!}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{4}{6}\right)^2 = \frac{5\cdot 4\cdot \cancel{3}\cdot \cancel{2}}{\cancel{2}\cdot\cancel{2}} \quad \frac{1}{6^{\cancel{3}\,2}}\cdot\frac{4}{9} =$$

$$= \frac{5\cdot\cancel{4}}{\cancel{36}\cdot 9} = \frac{5}{81}$$

OSSERVAZIONE

Vediamo un altro modo di ottenere lo stesso risultato (molto più complicato, non conveniente).

Pensiamo a 6 risultati possibili tutti con probabilità $\frac{1}{6}$. Abbiamo le due seguenti situazioni:

→ per $\underbrace{k=2,4,5,6}_{4\text{ casi}}$ $P(\text{"2 volte } \boxed{1}, 1 \text{ volta } \boxed{3}, 2 \text{ volte } \boxed{k}\text{"}) = \frac{5!}{2!\,1!\,2!\,0!\,0!\,0!}\left(\frac{1}{6}\right)^5 = \frac{5\cdot\cancel{4}\cdot 3\cdot 2}{\cancel{2}\cdot\cancel{2}}\left(\frac{1}{6}\right)^5 = \frac{5}{6^4} = \frac{5}{1296}$

→ per $\underbrace{\{k,h\}\subset\{2,4,5,6\}}_{\binom{4}{2}=6\text{ casi}}$ $P(\text{"2 volte } \boxed{1}, 1 \text{ volta } \boxed{3}, 1 \text{ volta } \boxed{k}, 1 \text{ volta } \boxed{h}\text{"})$

$$= \frac{5!}{2!\,1!\,1!\,1!\,0!\,0!}\left(\frac{1}{6}\right)^5 = \frac{5\cdot\cancel{4}^{\,2}\cdot 3\cdot 2}{\cancel{2}}\left(\frac{1}{6}\right)^5 = \frac{10}{6^4} = \frac{10}{1296}$$

La prob. richiesta è $4\cdot\frac{5}{1296} + 6\cdot\frac{10}{1296} = \frac{20+60}{1296} = \frac{80}{1296} = \frac{5}{81}$ (che è lo stesso risultato visto prima)

PROSSIMI ARGOMENTI (ALTRE DISTRIBUZIONI DISCRETE NOTEVOLI)

→ UNIFORME DISCRETA

→ POISSON (nome di un matematico francese)

→ GEOMETRICA E DISTRIBUZIONI COLLEGATE
(queste nelle prossime lezioni)

OSS. Spesso saranno definite a partire dalla espressione della distribuzione (o della densità discreta) senza fare riferimento a $(\Omega, \mathcal{A}, P)$.

DISTRIBUZIONE UNIFORME DISCRETA

lettere minuscole

Si tratta del caso in cui, per un insieme finito $E = \{x_1, \dots, x_n\} \subset \mathbb{R}$, si ha

$$P(X \in B) = \frac{\#(B \cap E)}{\# E} = \frac{\#(B \cap E)}{n} \qquad \forall B \subset \mathbb{R}.$$

ESEMPIO: Si ha questa distribuzione con $E = \{1,2,3,4,5,6\}$ se X è la v.a. che indica il numero che esce lanciando un dado equo.

DISTRIBUZIONE DI POISSON

Una v.a. X ha distribuzione di Poisson con parametro $\lambda > 0$ (in qualche caso scriviamo $X \sim \text{POISSON}(\lambda)$)
se si ha

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad \forall k \in \{0, 1, 2, 3, \dots\}$$

<u>OSS</u>.

La definizione è ben posta se $\sum_{k=0}^{\infty} p_X(k) = 1$. In effetti si ha

$$\sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{= e^{\lambda}} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

ESERCIZIO

Sia $X \sim \text{Poisson}(\lambda = 4)$.

1) Calcolare $P(X > 2)$

2) Calcolare $P(X = k \mid X \leq 2)$ per ogni $k \geq 0$ intero.

SVOLGIMENTO

1) Per eventi di questo tipo si deve passare alla probabilità dell'evento complementare:

$$P(X>2) = 1 - P(X \leq 2) = 1 - \sum_{k=0}^{2} p_X(k) = 1 - \left(\frac{4^0}{0!} e^{-4} + \frac{4^1}{1!} e^{-4} + \frac{4^2}{2!} e^{-4} \right) =$$

$$= 1 - (1 + 4 + 8) e^{-4} = 1 - 13 e^{-4}.$$

2) $$P(X=k \mid X\le 2) = \frac{P(\{X=k\}\cap\{X\le 2\})}{P(X\le 2)} = \begin{cases} \dfrac{P(X=k)}{P(X\le 2)} & \text{per } k\in\{0,1,2\} \ \left(\text{perché } \{X=k\}\subset\{X\le 2\}\right) \\ 0 & \text{per } k\ge 3 \ \left(\text{perché } \{X=k\}\cap\{X\le 2\}=\emptyset\right) \end{cases}$$

ed inoltre (per $k=0,1,2$)

$$\frac{P(X=k)}{P(X\le 2)} = \frac{p_X(k)}{\sum_{j=0}^{2} p_X(j)} = \frac{\frac{4^k}{k!}e^{-4}}{\sum_{j=0}^{2}\frac{4^j}{j!}e^{-4}} = \frac{\frac{4^k}{k!}\cancel{e^{-4}}}{\left(\frac{4^0}{0!}+\frac{4^1}{1!}+\frac{4^2}{2!}\right)\cancel{e^{-4}}} = \begin{cases} \text{per } k=0 & \frac{1}{13} \\ \text{per } k=1 & \frac{4}{13} \\ \text{per } k=2 & \frac{8}{13} \end{cases}$$

<u>OSS</u>. Gli eventi $\{\{X=k\}: k\ge 0\}$ costituiscono una partizione numerabile

Quindi si deve avere $\sum_{k=0}^{\infty} P(X=k \mid X\le 2) = 1$.

In effetti si ha $\sum_{k=0}^{\infty} P(X=k\mid X\le 2) = \sum_{k=0}^{2} P(X=k\mid X\le 2) = \frac{1+4+8}{13} = 1$.

trascuro gli addendi per $k\ge 3$ tutti uguali a zero

APPROSSIMAZIONE DELLA BINOMIALE CON LA POISSON

Sia $\lambda > 0$. Sia $n \geq \lambda$ intero; in questo modo $\frac{\lambda}{n} \in [0,1]$ e non è restrittivo perché poi siamo interessati a considerare un limite per $n \to \infty$.

Prendiamo la densità di $X \sim BIN\left(n, p_n = \frac{\lambda}{n}\right)$.

Allora, per $k \in \{0, 1, \dots, n\}$, si ha

$$\frac{n!}{(n-k)!} = \frac{n(n-1)\cdots(n-k+1)\cancel{(n-k)!}}{\cancel{(n-k)!}}$$

$$p_X(k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{n!}{k!\,(n-k)!}\,\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k} =$$

$$= \frac{\lambda^k}{k!}\underbrace{\frac{\cancel{n}(n-1)\cdots(n-k+1)}{\cancel{n}\,n\cdots\cdots n}}_{=\left(1-\frac{1}{n}\right)\cdots\cdots\left(1-\frac{k-1}{n}\right)}\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to 1} \xrightarrow[n\to\infty]{} \frac{\lambda^k}{k!}e^{-\lambda}$$

$$\left(1-\frac{1}{n}\right) \to 1 \qquad \left(1-\frac{k-1}{n}\right) \to 1$$

densità discreta di una POISSON (λ)

OSS. $\forall k \geq 0$ intero esiste n t.c. $n \geq k$ e quindi questo limite vale $\forall k \geq 0$ intero

Ricapitolando, per $X = X_n \sim BIN\left(n, p_n = \frac{\lambda}{n}\right)$ e $Z \sim POISSON(\lambda)$, si ha

$$\lim_{n \to \infty} p_{X_n}(k) = p_Z(k) \qquad \forall\, k \geq 0 \text{ intero}$$

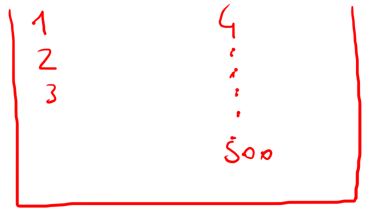
(limite delle densità discrete)

Questo limite ha un interesse teorico che non approfondiremo.

Al contrario c'è un interesse pratico: serve per calcolare valori approssimati di probabilità di eventi legati a v.a. con distribuzione Binomiale "con n grande e p piccolo". Presenteremo di seguito due esempi illustrativi.

ESERCIZIO / ESEMPIO

Sia $X \sim BIN\left(n=1000,\ p=\frac{3}{500}\right)$. Calcolare $P(X \geq 5)$



<u>OSS</u>. Si può pensare di fare riferimento ad un'urna con 500 palline numerate da 1 a 500. Si compiono 1000 estrazioni casuali <u>con</u> reinserimento. Vogliamo calcolare la probabilità di estrarre almeno 5 volte uno dei numeri 1, 2, 3.

RISPOSTA

Il valore esatto della probabilità da calcolare è

$$P(X \geq 5) = \sum_{k=5}^{1000} \binom{1000}{k} \left(\frac{3}{500}\right)^k \left(1 - \frac{3}{500}\right)^{1000-k} = 1 - \sum_{k=0}^{4} \binom{1000}{k} \left(\frac{3}{500}\right)^k \left(1 - \frac{3}{500}\right)^{1000-k}.$$

Non è semplice calcolare un valore approssimato di questa probabilità.

Per trovare un valore numerico approssimato, facciamo riferimento alla approssimazione Poissoniana della Binomiale. Si ha $p_X(k) \approx p_Z(k)$ (per $k \geq 0$ intero)

dove $Z \sim \text{Poisson}\left(\lambda = np = \cancel{1000}^{2} \cdot \frac{3}{\cancel{500}} = 6\right)$

Allora

approssimazione qui

$$P(X \geq 5) = 1 - P(X \leq 4) = 1 - \sum_{k=0}^{4} P(X=k) \cong 1 - \sum_{k=0}^{4} P(Z=k) =$$

$$= 1 - \sum_{k=0}^{4} \frac{6^k}{k!} e^{-6} = 1 - \left(\frac{6^0}{0!} + \frac{6^1}{1!} + \frac{6^2}{2!} + \frac{6^3}{3!} + \frac{6^4}{4!}\right) e^{-6}$$

$$= 1 - (1 + 6 + 18 + 36 + 54) e^{-6} = 1 - 115 \cdot e^{-6}.$$

ALTRO ESEMPIO

Stessa urna dell'esempio precedente.

Si estraggono 200 palline, una alla volta e con reinserimento.

Calcolare la probabilità di estrarre al più 2 volte uno dei numeri 1, 2, 3, 4, 5, 6, 7.

RISPOSTA

La probabilità richiesta è $P(Y \leq 2)$ dove $Y \sim BIN\left(n=200, p=\frac{7}{500}\right)$

Si ha
$$P(Y\leq 2) = \sum_{k=0}^{2} \binom{200}{k} \left(\frac{7}{500}\right)^k \left(1-\frac{7}{500}\right)^{200-k}.$$

Vogliamo trovare un valore approssimato come nell'esempio precedente.

Si ha $p_Y(k) \approx p_Z(k)$ $(\forall k \geq 0$ intero$)$ dove $Z \sim \text{Poisson}\left(\lambda = np = 200 \cdot \frac{7}{500} = \frac{14}{5}\right)$.

Allora

approssimazione qui

$$P(Y \leq 2) = \sum_{k=0}^{2} P(Y=k) \cong \sum_{k=0}^{2} P(Z=k) = \sum_{k=0}^{2} \frac{\left(\frac{14}{5}\right)^k}{k!} e^{-14/5} =$$

$$= \left(\frac{(14/5)^0}{0!} + \frac{(14/5)^1}{1!} + \frac{(14/5)^2}{2!} \right) e^{-14/5} = \left(1 + \frac{14}{5} + \frac{\overset{98}{\cancel{196}}}{\underset{25}{\cancel{50}}} \right) e^{-14/5} = \frac{193}{25} \cdot e^{-14/5}.$$