Ottavio Serra

Distribuzione di probabilità polinomiale e ipergeometrica[i]

1) **Distribuzione polinomiale.**
a generalizzazione della distribuzione binomiale. Anche questa è di tipo bernouilli, nel senso che la probabilità di successo è la stessa in ogni prova. La formula si ricava estendendo la formula binomiale. Sia A un insieme di n elementi, A_1, A_2, … A_s una sua partizione, n_i il numero di elementi di A_i e p_i la probabilità di successo di un elemento di A_i, con gli ovvi vincoli
$\sum_{i=1}^{s} n_i = n, \sum_{i=1}^{s} p_i = 1.$ (Se s=2, si ha la distribuzione binomiale).
Sia C un sotto-insieme di A avente k elementi (il *campione* estratto); si chiede la probabilità che di questi k elementi k_1 appartengano ad A_1, … k_s appartengano ad A_s. (Ovviamente, deve risultare $k_1+ k_2+ … + k_s=k$ e ogni k_i compreso tra 0 e k).
La distribuzione polinomiale sarà allora la seguente
$$[1] P(k_1, k_2, \dots k_s) = \frac{k!}{k_1! k_2! \dots k_s!} (p_1)^{k_1} (p_2)^{k2} \dots (p_s)^{ks}.$$

Esempio 1. Un'urna contiene 7 palline bianche, 8 nere e 5 rosse. Si estraggono successivamente 10 palline rimettendo ogni volta la pallina estratta nell'urna (si parla di *estrazioni con reinserimento*). Si chiede la probabilità che 4 siano bianche e 4 nere (**non c'è bisogno di specificare che le rimanenti saranno 2 rosse**).

Il reinserimento assicura che le probabilità siano costanti: rispettivamente 7/20, 8/20, 5/20. Risulta

$$P(4,4,(10-4-4)) = \frac{10!}{4!4!(10-4-4)!} \left(\frac{7}{20}\right)^4 \left(\frac{8}{20}\right)^4 \left(\frac{5}{20}\right)^2 \approx 0,0756$$

Nota. Nel caso del totocalcio k=13, s=2, k_1=h= numero dei successi (risultati indovinati), k_2=13-h = nume.ro degli insuccessi (risultati non indovinati), p_1=p=1/3, p2=1-p=2/3. La distribuzione è binomiale: $P(h,13-h) = \binom{13}{h} \left(\frac{1}{3}\right)^h \left(\frac{2}{3}\right)^{13-h} = \frac{13!}{k!(13-h)!} \left(\frac{1}{3}\right)^h \left(\frac{2}{3}\right)^{13-h}$.

(Perché al totocalcio **ho deciso** di assumere 1/3 come probabilità di indovinare un risultato?).

Il problema del totocalcio può essere **modellizzato** come l'estrazione **con reinserimento** da un'urna contenente a paline bianche e b=2a palline nere, per esempio 1 bianca e 2 nere oppure 2 bianche e 4 nere. In ogni caso p=a/(a+b) =a/(3a)=1/3 e 1-p=2/3. Si estraggono 13 palline, (anche se a+b<13 perché c'è il reinserimento). Si chiede qual è la probabilità di ottenere h bianche e, non c'è bisogno di dirlo, 13-h nere (8 bianche e 5 nere, eccetera).

2) **Distribuzione geometrica.**
Anche questa è una distribuzione a probabilità costante; si tratta di calcolare la probabilità che il **primo successo** si presenti al tentativo numero k. Si calcola facilmente $P(k)=q^{k-1}p$.
La probabilità che il primo successo si ottenga nei primi n tentativi è

$P(1 \leq k \leq n)$=p+qp+q^2p+ … +q^{n-1}p = $p\dfrac{1-q^n}{1-q} = 1-q^n$. (Progressione geometrica).

Come al solito, q=1-p. *Come si sarebbe potuto trovare immediatamente il risultato 1-q^n?*

Esercizi. Si supponga che la probabilità (costante) di successo p in una singola prova sia strettamente compresa tra 0 e 1.

(1) Calcolare la probabilità che il primo successo si abbia

(a) nel 3° o nel 4° tentativo,

(b) nel 5° o nel 7° o nel 10° tentativo,

(c) non prima del 5° tentativo.

(2) Calcolare la probabilità che il primo successo si abbia

(a) a un tentativo di ordine dispari: P_1,

(b) a un tentativo di ordine pari: P_2.

(c) Come mai, se i numeri dispari sono tanti quanto i pari, $P_1 > P_2$?

3) Distribuzione ipergeometrica.

Si chiama iper geometrica la distribuzione in cui la probabilità cambia nel corso delle prove. Un esempio è l'estrazione da un'urna **senza reinserimento**, come succede nel gioco della tombola o nelle estrazioni del lotto, del superenalotto e simili.

Per esempio, si abbia un'urna con 20 palline identiche numerate da 1 a 20. La probabilità di estrarre la pallina n° 1 (o n° 13 o qualunque altra) è, a priori, 1/20 alla prima estrazione, 1/19 alla seconda, 1 (ovvio!) alla ventesima. Se però si estraggono 4palline, non importa se in blocco o una alla volta (*senza reinserimento*), la probabilità che tra le 4 ci sia la prescelta (per esempio la n°13) è 4/20. Un calcolo dettagliato fornisce 1/20 se esce alla prima estrazione, (19/20) per (1/19) se non esce alla prima ma esce alla seconda estrazione, e così via. Pertanto si ottiene

$$\frac{1}{20}+\frac{19}{20}\frac{1}{19}+\frac{19}{20}\frac{18}{19}\frac{1}{18}+\frac{19}{20}\frac{18}{19}\frac{17}{18}\frac{1}{17}=\frac{4}{20}.$$

Nel gioco del lotto si estraggono 5 numeri da un'urna che ne contiene 90. Siccome la regola del gioco non tiene conto dell'ordine di estrazione (salvo che per *l'estratto determinato*), lo spazio di probabilità (**il numero delle possibili cinquine**) ha misura $m(S)=\binom{90}{5}=\dfrac{90!}{5!.85!}$ (combinazioni perché non conta l'ordine, semplici perché non c'è reinserimento). Se si gioca un terno secco su una data *Ruota*, si ha successo se i tre numeri giocati fanno parte dei 5 estratti, perciò nell'ipotesi di aver fatto terno secco la misura dell'Evento (*Successo*) è data dal numero delle cinquine che contengono quei tre numeri e dunque nell'urna restano solo 87 numeri **liberi** di combinarsi a 2 a 2: $m(E_3)=\binom{87}{2}$. Segue che la probabilità di fare terno secco su una data ruota è

$$P(3)=\frac{m(E_3)}{m(S)}=\frac{\binom{87}{2}}{\binom{90}{5}}=60\frac{87.86}{90.89.88.87.86}=\frac{6}{9.89.88}=\frac{1}{3.44.89}=\frac{1}{11748}\approx 0{,}0085\%.$$

In un gioco equo il lotto dovrebbe pagare una somma pari a 11748 volte la posta, invece paga solo 4500 volte la posta (controlla su Internet); è vero che ci sono le spese di gestione, ma i casinò di Las Vegas si contentano di molto meno. Morale: **non giocare**!

Generalizziamo. Sia k un numero da1 a 5: La probabilità di indovinare k numeri sula cinquina di una data ruota è

$$[2] P(90,5,k)=\frac{\binom{90-k}{5-k}}{\binom{90}{5}}.$$

Questa, a sua volta, è un caso particolare di una formula più generale. Sia A un insieme (la *popolazione*) di n elementi, A_1, A_2, …. A_s una sua partizione con numero di elementi rispettivamente n_1, n_2, … n_s. Dall'insieme A si estragga, **senza reinserimento**, un campione C di k elementi. Si chiede la probabilità che dei k elementi di C k_1 appartengano ad A_1, k_2 appartengano ad A_2, …, k_s appartengano ad A_s.

Il 1° evento ha misura $\binom{n_1}{k_1}$, il 2° ha misura $\binom{n_2}{k_2}$, …, l'ultimo $\binom{n_s}{k_s}$. Siccome gli eventi sono disgiunti, la probabilità richiesta (detta *distribuzione ipergeometrica multivariata)* è

$$[3] P(k_1,k_2,...,k_s)=\frac{\binom{n_1}{k_1}\cdot\binom{n_2}{k_2}\cdots\binom{n_s}{k_s}}{\binom{n}{k}}.$$

I vincoli sono ovvi: $\sum_{i=1}^{s} n_i = n, \sum_{i=1}^{s} k_i = k, 1 \le k_i \le n_i \forall i = 1..s$

La distribuzione ipergeometrica usuale (**bivariata**) si ottiene per s=2:

$$[4] P(k_1,k-k_1)=\frac{\binom{n_1}{k_1}\cdot\binom{n-n_1}{k-k_1}}{\binom{n}{k}}.$$

Come mai la [2] rientra come caso particolare nella [4]? Siccome si sa che k=5, si tratta di fare una partizione di A in due classi, A_1 di cardinalità $n_1= k_1$ (k_1 da 1 a 5) e di conseguenza $n_2 = 90-n_1$ e $k_2 =5-k_1$. Nella [2] k_1 è k (non c'è confusione perché n=90 e il k della [4] è 5.
Si vede così che l'estrazione al gioco del lotto è un caso particolare di distribuzione ipergeometrica (**bivariata**).
Il lotto consente anche giocate in cui l'utente gioca n_1 numeri (su una ruota), con $n_1>k_1$, sperando che degli n_1 numeri giocati k_1 compaiano nella cinquina estratta. In tal caso la probabilità è data dalla [4].
Esercizio. Si gioca un terno secco su r ruote; calcola la probabilità di fare almeno un terno. Il lotto per ogni terno realizzato paga 4500/r euro per ogni euro della posta; verifica che ancora una volta ha ***rosicchiato*** sul premio.
Il superenalotto segue la distribuzione ipergeometrica bivariata [4].
Da un'urna di 90 numeri se ne estraggono 6 (*senza reinserimento*). Il giocatore può giocare anche più di 6 numeri per aumentare la probabilità di vincere, (naturalmente pagando una posta maggiore), per esempio n_1=10, sperando che tra questi ci siano magari 5 numeri vincenti o addirittura 6. La probabilità di indovinare k dei 6 numeri estratti è

$$P(10,k)=\frac{\binom{10}{k}\cdot\binom{80}{6-k}}{\binom{90}{6}}.$$

La probabilità di fare 6, giocando 6 numeri (n_1=k=6), è

$P(6,6)=\frac{\binom{6}{6}\cdot\binom{84}{0}}{\binom{90}{6}}=\frac{1}{\binom{90}{6}}=\frac{1}{622614630}$, veramente piccola. **(Non giocare!)**

Verifica che, giocando 6 numeri, la probabilità di fare 5 è 504 volte maggiore, ma in ogni caso molto piccola, meno di unna parte su cento milioni. **(Non giocare!)**

Esempio 2. Come nell'esempio 1, l'urna contenga 20 palline identiche salvo che per il color: 7 bianche, 8 nere (e 5 rosse). Se si estraggono 10 palline in blocco (o una dopo l'atra, **senza reinserimento**), la probabilità di averne 4 bianche, 4 nere e 2 rosse segue la distribuzione ipergeometrica multivariata:

$$P(4,4,2)=\frac{\binom{7}{4}\binom{8}{4}\binom{5}{2}}{\binom{20}{10}}=\frac{6125}{46189}\approx 0{,}1326.$$

Questo esempio utilizza gli stessi dati dell'esempio 1, solo che allora si aveva distribuzione polinomiale (**con reinserimento**), mentre ora si ha distribuzione ipergeometrica multivariata, analoga della polinomiale; l'unica differenza e che ora **non c'è reinserimento**.

Esercizi.

(1) Detta p la probabilità di fare terno secco su una ruota, calcolare, giocando i tre numeri su 4 ruote,

(a) la probabilità di non fare alcun terno;

(b) di fare un solo terno;

(c) di farne due;

(d) di farne tre;

(e) di farne quattro;

(f) di farne almeno uno;

(g) di farne almeno due;

(h) di farne al più due.

(2) Giocando 8 numeri al superenalotto, calcolare la probabilità

(a) di fare 5;

(b) di fare 6.

(3) Da un mazzo di carte napoletane (40 carte ripartire in 4 colori di 10 carte ciascuno, delle quali 3 sono figure) si estraggono k carte; calcolare la probabilità che 6 siano figure e 2 siano assi

(a) senza reinserimento, una volta per k=8, una volta per k=10;

(b) con reinserimento, una volta per k=8, una volta per k=10.

(4) Quesito N° 7 della prova di matematica PNI del 2013.

In un gruppo di 10 persone 6 sono con gli occhi azzurri. Se ne scelgono 2 a caso (senza reinserimento). Qual è la probabilità che nessuno delle due abbia gli occhi azzurri?

Prima soluzione. La distribuzione è ipergeometrica, perciò $P = \frac{\binom{6}{0}\binom{4}{2}}{\binom{10}{2}} = \frac{6}{45} = \frac{2}{15}$.

Seconda soluzione. Per comodità chiamo occhi **neri** i **non azzurri**. Usando la probabilità *condizionata*, $\mathrm{P}(\mathrm{A} \cap B) = P(A/B).P(B), ottengo$
P(1°occhi neri ∩ 2°*occhi neri*)= P(2° occhi neri / 1° occhi neri). P(1° occhi neri)= (4/10).(3/9)=2/15.
Calcola le altre probabilità: entrambe le persone con gli occhi azzurri, una con occhi azzurri e l'altra con occhi neri. Quanto dovrebbe fare la somma di queste tre probabilità?
Usa tutti e due i metodi; se qualcosa non torna, ricorda che le persone sono *oggetti* ***distinguibili,*** non sono ***bosoni!***

[i] Per le distribuzioni binomiale, di Poisson e gaussiana vedi nel mio sito *digilander.libero.it/ottavioserra0* in questa stessa cartella *l'articolo Distribuzioni di probabilità e nel*la cartella "*Lezioni allo Scorza*", in particolare, nella sezione "10 in matematica 2010", le lezioni di probabilità. Nella stessa cartella vedi anche "Probabilità 2016".