# Capitolo 14

# *MODELLI DI ESTRAZIONE E COLLOCAZIONE*

Il contenuto di questo capitolo è sostanzialmente un'applicazione del procedimento sequenziale del n° 13.2.5 allo studio dei problemi d'estrazione che vengono qui approfonditi – casi particolari si trovano tra gli *Esempi* 4.1.1 e 13.2.3 – anche perché saranno utili come esempi di riferimento nei *Capitoli* 15 (*indipendenza stocastica*) e 16 (*scambiabilità*), Vol. II. Vengono studiate sequenze d'estrazione da un'urna contenente oggetti di *tipo diverso* in tre *modalità d'estrazione – con, senza rimessa* e *con contagio* (rimessa assieme a un oggetto dello stesso tipo) – nell'ipotesi di *scelta a caso* dell'oggetto in ogni estrazione (*modelli base*) e loro *generalizzazioni* (distribuzioni sui *tipi* di oggetto assegnate con funzioni peso). Nel § 14.2 si studiano i modelli con *due alternative* – urna contenente palline (oggetti) bianche e rosse – e per ogni modalità si determina la *distribuzione* del *numero di successi* (estrazioni di pallina bianca) in $n$ estrazioni – la *probabilità* sulla *sua* partizione canonica –. Nelle modalità *con, senza rimessa* e *con contagio* si trovano nell'ordine le distribuzioni *binomiale* (n° 14.2.1), *ipergeometrica* (n° 14.2.2) e *di Pólya* (n° 14.2.3). I modelli con *più* alternative vengono studiati nel § 14.3, ove in corrispondenza alle tre modalità si introducono le distribuzioni *multinomiale,* (n° 14.3.1), *ipergeometrica multipla* (n° 14.3.2) e *di Pólya multipla* (n° 14.3.3). Nel § 14.4 si interpretano i modelli *collocazione* (di classificazione per caratteristica) come modelli d'estrazione di una caratteristica per ciascun oggetto e si studiano quelli corrispondenti alle tre solite modalità, in cui si ritrovano tre modelli della fisica delle particelle – più generale quello relativo al modello con contagio –, introdotti classicamente in convenienti ipotesi di simmetria.

## 14.1 Modelli d'estrazione

Si è fatto cenno nel § 4.1 che il primo approccio al calcolo delle probabilità ha avuto luogo con lo studio di problemi collegati a giochi d'azzardo, i quali, nella maggioranza, si prestano ad essere trattati come problemi d'estrazione di oggetti da un'urna; o perché così si presentano già in origine – com'è nel caso del lotto e di lotterie varie – o dopo convenienti interpretazioni – ad es. nei giochi a dadi, roulette, giochi di carte, problemi di collocazione –.

L'interesse per lo studio dei problemi collegati a **schemi d'estrazione** è solo in piccola parte dovuto alle applicazioni ai giochi d'azzardo. A schemi d'estrazione vengono infatti ricondotti – se non altro in un primo approccio – problemi di rilevante importanza, riguardanti vari aspetti dell'attività umana. È facile intuire le possibilità di applicazione nel campo della statistica per indagini induttive sulla classificazione di "popolazioni". Come detto, le possibilità di applicazione, in senso induttivo e non, sono però molto variegate. Basterà citare alcuni campi, come quello del controllo di qualità, della meccanica statistica, delle assicurazioni, dei problemi di diffusione, compresi quelli epidemici. Indicazioni su taluni di questi modelli proposti saranno date a suo luogo, dopo introdotto l'appropriato modello d'estrazione. Problemi d'induzione statistica, collegati a modelli d'estrazione, sono trattati con qualche dettaglio nel *Complemento* 16.2.4, Vol. II, per spiegare con semplici esempi l'uso corretto (coerente) delle frequenze osservate al fine di ottenere valutazioni di probabilità.

Lo schema d'estrazione che andremo ad esaminare prevede un'urna contenente $m$ oggetti di $t$ tipi, il tipo $h$-esimo di numerosità $m_h$ ($m=m_1+\ldots+m_t$). Vediamo qualche esempio tra quelli considerati nel n° 4.1.1: nel gioco del lotto – primi cinque *Esempi* – ogni pallina è distinguibile dalle altre (è numerata) e riesce perciò $m = t = 90$, ) $m_h = 1$ ($h = 1, \ldots, 90$); situazioni analoghe si hanno quando si lancia un dado, ove gli oggetti sono le 6 facce ($m = t = 6$), e nei problemi di collocazione dell'*Esempio* 13, ove gli oggetti sono 3 cassetti distinguibili ($m = t = 3$); gli oggetti sono ovviamente le carte nei giochi a carte e a seconda dei fini esse vengono considerate indistinguibili se dello stesso rango e allora $m_h = 4$ – ad es. per quasi tutti i punti del poker ed è $t = 8$ e $m = 32$ nell'*Esempio* 6 – oppure dello stesso seme e allora $t = 4$ – ad es. per i punti colore e scala reale nel poker ($m_h = 8$, $m = 32$) e ai fini della distribuzione nel gioco del bridge (*Esempio* 7), in cui $m_h = 13$ e $m = 52$ –; sono palline di colore diverso e per il resto indistinguibili negli *Esempi* 9, 10, 11 – 8 bianche e 2 nere ($t=2$, $m_1=8$, $m_2=2$, $m=10$) – e nell'*Esempio* 12 – 1 bianca, 2 nere, 3 rosse, 4 verdi ($t=4$, $m_1=1$, $m_2=2$, $m_3=3$, $m_4=4$, $m=10$) –.

Circa le modalità d'estrazione, considereremo dapprima modelli che prevedono una sequenza di estrazioni di un oggetto per volta, *a caso in ogni singola estrazione* – come negli *Esempi* del n° 13.2.4 (eccetto che per un quesito) –. Se sono $s$ gli oggetti presenti nell'urna in una di tali estrazioni e sono $s_h$ quelli di tipo $h$, allora ogni oggetto ha probabilità $1/s$ di essere scelto e $s_h/s$ è la probabilità che venga scelto un oggetto di tipo $h$. Inoltre, ogni modalità d'estrazione prevede regole che prescrivono come comportarsi prima di effettuare la prossima estrazione. Abbiamo già incontrato negli esempi le modalità *con rimessa*, in cui è prescritto che l'oggetto estratto venga rimesso nell'urna (reimbussolato), e *senza rimessa*, in cui l'oggetto non viene reimbussolato. Considereremo qui come terza modalità quella *con contagio* o *incremento unitario,* in cui è prescritto che dopo ogni estrazione l'oggetto venga reimbussolato assieme ad uno dello stesso tipo. Studieremo essenzialmente questi tre modelli. Per quelli con rimessa e con contagio daremo anche una versione generalizzata, consistente nel considerare gli oggetti nell'urna tutti distinguibili e sostituendo l'ipotesi di scelta a caso con regole che interpretano convenientemente le due modalità d'estrazione.

Sin qui i modelli menzionati prevedono che in ogni singola estrazione venga scelto un solo oggetto. Nulla vieta naturalmente di considerare anche modelli che prevedono l'estrazione di più oggetti per volta, diciamo $n$ su $m$ presenti nell'urna. Con qualche cautela – al momento di ogni singola estrazione deve essere $n \leq m$ – essi sono però facilmente riconducibili ai modelli precedenti, immaginando un'urna contenente come oggetti la totalità dei sottoinsiemi di $n$ oggetti che si possono formare con gli $m$ a disposizione. Nell'ipotesi di scelta a caso da quest'urna fittizia – è il solo modello d'estrazione a gruppi che prenderemo in considerazione – si può stabilire, come vedremo, una sorta di equivalenza tra una estrazione secondo questo modello e una sequenza di $n$ estrazioni di un oggetto per volta, senza rimessa e a caso.

## 14.2 Estrazioni con due alternative

In questo paragrafo studieremo modelli d'estrazione da un'urna contenente $b$ palline bianche e $r$ rosse e loro generalizzazioni. Posto $E_i$ = *esce pallina bianca all'i-esima estrazione*, indichiamo con $[E_1' \wedge \dots \wedge E_n']_h$ il generico costituente della partizione generata $\mathbb{P}_G(\{E_1, \dots, E_n\})$ con $h$ eventi affermati (e perciò $n-h$ negati). Per ogni $[E_1' \wedge \dots \wedge E_n']_h$ possibile, usando il teorema delle probabilità composte si ha:

$$P([E_1' \wedge \dots \wedge E_n']_h) = P(E_1')\, P(E_2' | E_1') \dots P(E_n' | E_1' \wedge \dots \wedge E_{n-1}') \,. \qquad (46)$$

Come vedremo, in tutte tre le modalità d'estrazione (e loro generalizzazioni) si trova che la probabilità del generico evento elementare $[E_1' \wedge \dots \wedge E_n']_h$ dipende da $h$ e $n-h$, ovvero *da quanti sono gli eventi affermati e quanti quelli negati, ma non da quali*. Posto allora $P([E_1' \wedge \dots \wedge E_n']_h) = \pi(h\,, n-h)$ e indicato con $S_n$ il numero (aleatorio) di eventi affermati (successi) nelle prime $n$ estrazioni, la sua distribuzione di probabilità – quella sulla partizione $\{S_n = 0, \dots, S_n = n\}$ – è data da

$$P(S_n = h) = \sum\nolimits^{(\cdot)} P([E_1' \wedge \dots \wedge E_n']_h) = \binom{n}{h} \pi(h\,, n-h), \quad h = 0\,, \dots, n, \qquad (47)$$

ove la sommatoria si intende estesa ai termini che si ottengono scegliendo in tutti i modi possibili $h$ apici affermati e $n-h$ negati (la notazione è analoga a quella convenuta per la somma logica in 6.3.2*Notazione*).

Come lascia intendere la (46), la valutazione su $\mathbb{P}_G(\{E_1, \dots, E_n\})$ sarà ottenuta con procedura sequenziale. Frammentandola cioè in una sequenza di valutazioni *coerenti* (in accordo con le rispettive modalità d'estrazione) sulle partizioni:

$$\{E_1, \bar{E}_1\},\ \{E_2|E_1', \bar{E}_2|E_1'\},\ \dots,\ \{E_n|E_1'\wedge\dots\wedge E_{n-1}', \bar{E}_n|E_1'\wedge\dots\wedge E_{n-1}'\}.$$

La coerenza di tale valutazione – e perciò anche di quella sulla partizione $\{S_n=0, \dots, S_n=n\}$ – è allora assicurata dalla *b)* di 13.2.5*Teorema*, e questo rende superflua la verifica che $\sum^{(\cdot)} P(E_1'\wedge\dots\wedge E_n') = 1$. Il che è molto utile, perché, come si vedrà, tale verifica è molto semplice nel modello con rimessa e sua generalizzazione, ma non altrettanto negli altri modelli.

### 14.2.1 Estrazione con rimessa. Distribuzioni binomiale.

È il caso più semplice, perché dopo ogni estrazione la composizione dell'urna ritorna nelle condizioni iniziali. In ogni singola estrazione è cioè $b/(b+r)$ la probabilità di estrarre pallina bianca e $r/(b+r)$ quella di estrarre pallina rossa. In termini più precisi, per ogni $i = 2, \dots, n$ riesce:

$$P(E_i|E_1'\wedge\dots\wedge E_{i-1}') = P(E_1) = b/(b+r),\quad P(\bar{E}_i|E_1'\wedge\dots\wedge E_{i-1}') = P(\bar{E}_1) = r/(b+r)$$

A secondo membro della (46) troviamo allora $h$ fattori $b/(b+r)$, $n-h$ fattori $r/(b+r)$ e pertanto si ha: $P([E_1'\wedge\dots\wedge E_n']_h) = [b/(b+r)]^h[r/(b+r)]^{n-h}$, e ciò vale per ogni $n$ e $0 \le h \le n$.

Si trova poi lo stesso risultato se l'urna contiene solo due palline, una bianca e una rossa – o si effettua una sequenza di lanci di una moneta ($E_i$ = *esce testa nel lancio i-esimo*) –, e si interpretano $b$ ed $r$ come pesi (numeri reali positivi e per il resto qualunque) delle distribuzioni di probabilità sulle partizioni $\{E_1, \bar{E}_1\}$ e $\{E_i|E_1'\wedge\dots\wedge E_{i-1}', \bar{E}_i|E_1'\wedge\dots\wedge E_{i-1}'\}$, $i > 1$. Posto $p = b/(b+r)$ e $q = r/(b+r)$, la distribuzione (47) – coerente per 13.2.5*Teorema* come detto in premessa di questo paragrafo – allora diventa

$$P(S_n = h) = \binom{n}{h} p^h q^{n-h},\quad p, q \ge 0,\ p+q = 1,\ h = 0, \dots, n,$$

e ad essa si dà il nome di **distribuzione binomiale**. La verifica diretta della coerenza qui è semplice. Riesce infatti:

$$\textstyle\sum^{(\cdot)} P(E_1'\wedge\dots\wedge E_n') = \sum_{h=0}^n P(S_n = h) = \sum_{h=0}^n \binom{n}{h} p^h q^{n-h} = (p+q)^n = 1$$

Osserviamo ancora che nel gioco a testa e croce con moneta *perfetta* ($p = q = 1/2$) riesce $P(E_1'\wedge\dots\wedge E_n') = 1/2^n$, e perciò i $2^n$ eventi elementari della partizione generata da $\{E_1, \dots, E_n\}$ sono equiprobabili. I problemi che riguardano una partita a testa e croce che prevede $n$ lanci di una moneta perfetta, in cui si suppone che il risultato di ogni lancio non dipenda da risultati di quelli precedenti, possono allora essere studiati ponendosi in ipotesi di simmetria. Le probabilità degli eventi logicamente dipendenti dalla partizione $\mathbb{P}_G(\{E_1, \dots, E_n\})$ possono cioè essere valutate come rapporto di casi favorevoli e casi possibili, come fatto in precedenza.

### 14.2.2 Estrazione senza rimessa. Distribuzione ipergeometrica.

Qui si suppone che la pallina estratta in ogni singola estrazione non venga rimessa nell'urna. Per ogni $n \leq b+r$ riesce allora $[E_1' \wedge \dots \wedge E_n']_h \neq \phi$ se e solo se sono soddisfatte le condizioni $h \leq b$ e $h \leq n$, $n-h \leq r$, $n-h \leq n$, che si traducono nella $max(0, n-r) \leq h \leq min(b, n)$. In questa ipotesi possiamo allora applicare la (46). Poiché l'estrazione avviene senza rimessa, i fattori a secondo membro – in numero di $n$ – sono frazioni di denominatore decrescente da $b+r$, quello di $P(E_1')$, a $b+r-(n-1)$, quello di $P(E_n' | E_1' \wedge \dots \wedge E_{n-1}')$. Il denominatore del risultato è perciò $(b+r)_n$. Circa i numeratori, siamo in grado di dire che $h$ di essi dipendono da $b$ e $n-h$ da $r$, perché il prodotto logico prevede $h$ estrazioni di pallina bianca e $n-h$ di pallina rossa. Percorrendo i fattori da sinistra a destra – come si è fatto del resto per il calcolo del denominatore – i numeratori che dipendono da $b$ sono nell'ordine $b, b-1, \dots, b-(h-1)$, e quelli che dipendono da $r$ sono $r, r-1, \dots, r-(n-h-1)$. Per la proprietà commutativa della moltiplicazione, come numeratore del risultato troviamo allora $(b)_h (r)_{n-h}$. Pertanto: $P([E_1' \wedge \dots \wedge E_n']_h) = (b)_h (r)_{n-h} / (b+r)_n$ e la (47) ora diventa

$$P(S_n = h) = \begin{cases} \binom{n}{h} \dfrac{(b)_h (r)_{n-h}}{(b+r)_n} & \text{se } max(0, n-r) \leq h \leq min(b, n) \\ 0 & \text{altrimenti} \end{cases} \tag{48}$$

e prende il nome di **distribuzione ipergeometrica.**

COMPLEMENTO. *Estrazioni a gruppi di uguale numerosità.*

Gli eventi $S_n = h$ descrivono il risultato della sequenza di $n$ estrazioni di una pallina per volta senza tenere conto dell'ordine. Si può pervenire all'osservazione dello stesso risultato estraendo dall'urna $n$ palline tutte in una volta. È del tutto naturale allora pensare confrontare gli eventi $S_n = h$ con i corrispondenti $E_h^{(n)} =$ *su n palline estratte dall'urna tutte in una volta h sono bianche.* Si tratta di eventi sicuramente diversi, perché sono diverse le modalità d'estrazione e con esse i rispettivi stati d'informazione. Se si suppone che nell'estrazione a gruppi i sottoinsiemi di $n$ palline abbiano la stessa probabilità di essere estratti, viene però naturale ritenere che per ogni $h$ gli eventi $S_n = h$ e $E_h^{(n)}$debbano avere probabilità uguale. Per dare corpo a questa sensazione, numeriamo *idealmente* le palline da 1 a $b+r$ per renderle distinguibili (fisicamente lo sono) – come abbiamo fatto in casi analoghi in precedenza (ad es. in 4.1.1 *Esempi* 9, 10, 11, 12) – e introduciamo la partizione di $\Omega$ negli eventi elementari *esce l'insieme di palline* $\{i_1, \dots, i_n\}$, $1 \le i_1 < \dots < i_n \le b+r$, che giudichiamo equiprobabili. Gli eventi elementari possibili sono allora in numero di $\binom{b+r}{n}$. Quelli favorevoli a $E_h^{(n)}$ si ottengono scegliendo $h$ delle $b$ palline bianche in $\binom{b}{h}$ modi e $n-h$ delle $r$ palline rosse in $\binom{r}{n-h}$ modi. Si ottiene allora:

$$P(E_h^{(n)}) = \binom{b}{h}\binom{r}{n-h}\Big/\binom{b+r}{n}, \quad max(0, n-r) \le h \le min(b, n). \qquad (49)$$

L'espressione che si ricava per la probabilità di $E_h^{(n)}$è formalmente diversa da quella dell'evento $S_n = h$ del modello sequenziale. Ma solo formalmente. È facile verificare infatti che i due risultati sono uguali (esercizio).
In conclusione, se un modello d'estrazione con oggetti di due tipi prevede che le operazioni avvengano in modo sequenziale un oggetto per volta e a caso, ove si abbia interesse per il numero di successi (estrazione di un oggetto di un dato tipo), ma non per l'ordine d'estrazione, le valutazioni di probabilità che si riferiscono a sequenze di $n$ oggetti possono essere fatte sostituendo idealmente al modello originale quello che prevede una sola estrazione di un gruppo di $n$ oggetti tutti in una volta, in condizioni di simmetria. Il che significa immaginare di scegliere a caso un elemento da un'urna contenente come oggetti i sottoinsiemi di $n$ elementi che si possono formare con i $b+r$ oggetti.

### 14.2.3 Estrazione con contagio. Distribuzione di Pólya.

In questo schema si suppone che dopo ogni estrazione la pallina estratta venga rimessa nell'urna assieme ad un'altra dello stesso colore. Allora, contrariamente al caso dell'estrazione *senza rimessa*, la composizione dell'urna viene ora modificata *aumentando* di una pallina il suo contenuto dopo ogni estrazione. Di conseguenza gli eventi $[E_1' \wedge \dots \wedge E_n']_h$ sono possibili per ogni $n$ e $h \le n$, come nel modello senza rimessa. Nella (46) i denominatori dei fattori a secondo membro sono noti, come nei due modelli precedenti, ma ora crescono da $b+r$ a $b+r+n-1$ e il loro prodotto è allora $(b+r+n-1)_n$. Circa il numeratore, il ragionamento è analogo a quello del caso senza rimessa, salvo dover qui incrementare anziché decrementare i fattori, rispettivamente da $b$ a $b+h-1$ e da $r$ a $r+(n-h)-1$. Si ottiene perciò $\pi(h, n-h) = P([E_1' \wedge \dots \wedge E_n']_h) = (b+h-1)_h (r+n-h-1)_{n-h} / (b+r)_n$. Usando la (47) si ha perciò $P(S_n = h) = \binom{n}{h}[(b+h-1)_h (r+n-h-1)_{n-h} / (b+r+n-1)_n]$, da cui si ricava:

$$P(S_n = h) = \binom{b+h-1}{h}\binom{r+n-h-1}{n-h}\Big/\binom{b+r+n-1}{n} \tag{50}$$

che è la **distribuzione di Pólya** relativa al *modello base*.
La formula si generalizza considerando anche in questo caso, come in quello con rimessa, un'urna con due palline, una bianca e una rossa, attribuendo peso $b$ alla bianca e peso $r$ alla rossa per la prima estrazione ed incrementando di $a$ il peso della pallina estratta in ciascuna delle estrazioni successive. Allora riesce $P(E_1)=b/(b+r)$ e $P(\bar{E}_1)=r/(b+r)$. Per il calcolo di $P(E'_{i+1}|E'_1\wedge\ldots\wedge E'_i)$, $1\leq i<n$, occorre e basta conoscere il numero di successi nei primi $i$ colpi. Siano questi $j\leq i$. I pesi di $E'_{i+1}|E'_1\wedge\ldots\wedge E'_i$ e di $\bar{E}_{i+1}|E'_1\wedge\ldots\wedge E'_i$ sono allora $b+ja$ e $r+(i-j)a$, rispettivamente. Le loro probabilità sono perciò nell'ordine $(b+ja)/(b+r+ia)$ e $(r+(i-j)a)/(b+r+ia)$. Per la (46) allora si ottiene

$$P([E'_1\wedge\ldots\wedge E'_n]_h) = \frac{b(b+a)\ldots(b+(h-1)a)\,r(r+a)\ldots(r+(n-h-1)a)}{(b+r)(b+r+a)\ldots(b+r+(n-1)a)} =$$

$$\frac{(-b/a)[(-b/a)-1]\ldots[(-b/a)-h+1](-r/a)[(-r/a)-1]\ldots[(-r/a)-n+h+1]}{[(-b/a)-(r/a)][(-b/a)-(r/a)--1]\ldots[(-b/a)-(r/a)-n+1]}) \,.$$

Posto $p=b/(b+r)$, $q=r/(b+r)$ e $\gamma=a/(b+r)$, quindi $b/a=p/\gamma$ e $r/a=q/\gamma$, da qui si ricava

$$P(S_n = h) = \binom{n}{h}\frac{(-p/\gamma)_h(-q/\gamma)_{n-h}}{(-1/\gamma)_n} = \frac{(-p/\gamma)_h}{h!}\,\frac{(-q/\gamma)_{n-h}}{(n-h)!}\Big/\frac{(-1/\gamma)_n}{n!}$$

e quindi finalmente:

$$P(S_n = h) = \binom{-p/\gamma}{h}\binom{-q/\gamma}{n-h}\Big/\binom{-1/\gamma}{n} \quad \gamma>0,\ p,q\geq 0,\ p+q=1. \tag{50'}$$

La (50') è la formulazione della distribuzione di Pólya unidimensionale. I parametri $p$ e $q$ sono la probabilità iniziale di estrarre pallina bianca e rossa, rispettivamente, e possono essere interpretati come indicatori del bilanciamento iniziale dei due colori. Il parametro $\gamma$ è invece l'incremento percentuale del peso complessivo iniziale in vista della seconda estrazione, e in quanto tale può essere interpretato come un indicatore dell'effetto che ha l'incremento di peso sullo sbilanciamento di cui sopra. È utile osservare per questo che riesce $P(E_2|\bar{E}_1) = p/(1+\gamma) = P(E_1) - [\gamma/(1+\gamma)]\,P(E_1)$ e simmetricamente $P(\bar{E}_2|E_1) = P(\bar{E}_1) - [\gamma/(1+\gamma)]\,P(\bar{E}_1)$. Sicché, $\gamma/(1+\gamma)$ è il decremento percentuale della probabilità che nella seconda estrazione esca l'oggetto di un colore (bianco o rosso), se nella prima estrazione è uscito l'oggetto dell'altro colore (rosso o bianco). Ad esempio, se nella seconda estrazione il peso complessivo dei due oggetti aumenta del 10%, 50%, 100%, le suddette probabilità diminuiscono del 9,09%, 33,33%, 50%, rispettivamente.
Questo modello è stato introdotto da G. Pólya e F. Eggenberger (1923) per

studiare problemi di contagio tra individui di una popolazione colpita da una epidemia. L'osservazione di un individuo sano (ammalato) è una indicazione di regresso (progresso) dell'epidemia, perché rafforza la probabilità di osservare un prossimo individuo sano (ammalato), in misura tanto maggiore quanto più grande è $\gamma$.

## 14.3 Estrazioni con più alternative

Nel caso di una sequenza di estrazioni di un oggetto per volta da un'urna contenente oggetti di $t$ tipi, con ragionamenti del tutto analoghi (differiscono solo per il riferimento a *più* tipi d'oggetto, anziché a *due*), si ottengono risultati simili a quelli trovati nei modelli con due alternative. Ci possiamo perciò limitare a una descrizione essenziale dei modelli e dei relativi risultati. Per descrivere il risultato di una sequenza di $n$ estrazioni di un oggetto per volta consideriamo intanto la sequenza di partizioni $\mathbb{P}_1, \ldots, \mathbb{P}_n$, ove $\mathbb{P}_i = \{\omega_1^{(i)}, \ldots, \omega_t^{(i)}\}$ e $\omega_h^{(i)}$= *nell'estrazione i-esima esce un oggetto di tipo h.* Allora l'evento $\omega_{h_1}^{(1)} \wedge \ldots \wedge \omega_{h_n}^{(n)} \in \mathbb{P}_1 \wedge \ldots \wedge \mathbb{P}_n$ descrive una sequenza di $n$ estrazioni e, in analogia al caso di due alternative, $[\omega_{h_1}^{(1)} \wedge \ldots \wedge \omega_{h_n}^{(n)}]_{n_1, \cdots, n_t}$ è il generico $\omega_{h_1}^{(1)} \wedge \ldots \wedge \omega_{h_n}^{(n)}$ che prevede che l'oggetto di tipo $h$ venga estratto $n_h$ volte (che $n_h$ degli indici $h_1, \ldots, h_n$ siano uguali a $h$), $h = 1, \ldots, t$, $n_1 + \ldots + n_t = n$. Usando il teorema delle probabilità composte ora si trova:

$$P([\omega_{h_1}^{(1)} \wedge \ldots \wedge \omega_{h_n}^{(n)}]_{n_1, \cdots, n_t}) = \\ P(\omega_{h_1}^{(1)}) P(\omega_{h_2}^{(2)} | \omega_{h_1}^{(1)}) \ldots P(\omega_{h_n}^{(n)} | \omega_{h_1}^{(1)} \wedge \ldots \wedge \omega_{h_{n-1}}^{(n-1)}), \quad n_1 + \ldots + n_t = n. \tag{51}$$

Nei tre modelli con e senza rimessa e con contagio e loro generalizzazioni, il risultato analogo a quello del caso $t = 2$ è qui, come vedremo, che la probabilità di $[\omega_{h_1}^{(1)} \wedge \ldots \wedge \omega_{h_n}^{(n)}]_{n_1, \cdots, n_t}$ dipende *da <u>quante</u> sono le estrazioni di oggetti di tipo 1, ... , <u>quante</u> di tipo t, ma non da <u>quali</u>.* In simboli si ha perciò $P([\omega_{h_1}^{(1)} \wedge \ldots \wedge \omega_{h_n}^{(n)}]_{n_1, \cdots, n_t}) = \pi(n_1, \ldots, n_t)$. Indicato allora con $S_n^{(h)}$ il numero aleatorio degli oggetti di tipo $h$ che vengono estratti nelle $n$ estrazioni, $S_n^{(1)} + \ldots + S_n^{(t)} = n$, i costituenti di $\mathbb{P}_1 \wedge \ldots \wedge \mathbb{P}_n$ che compongono l'evento $(S_n^{(1)} = n_1) \wedge \ldots \wedge (S_n^{(t)} = n_t)$ sono $\binom{n}{n_1}\binom{n-n_1}{n_2} \ldots \binom{n-n_1-\ldots-n_{t-1}}{n_t} = n!/n_1! \ldots n_t!$. Pertanto si trova:

$$P((S_n^{(1)} = n_1) \wedge \ldots \wedge (S_n^{(t)} = n_t)) = \frac{n!}{n_1! \ldots n_t!} \pi(n_1, \ldots, n_t), \; n_1 + \ldots + n_t = n. \tag{52}$$

Come nel caso $t = 2$, anche qui la coerenza della valutazione è assicurata da 13.2.5 *Teorema*. In tutti tre i modelli riesce perciò:

$$\sum_{n_1+\ldots+n_t=n} P((S_n^{(1)}=n_1)\wedge\ldots\wedge(S_n^{(t)}=n_t)) = \sum_{n_1+\ldots+n_t=n} \frac{n!}{n_1!\ldots n_t!}\,\pi(n_1,\ldots,n_t) = 1.$$

### 14.3.1 Estrazione con rimessa. Distribuzione multinomiale.

È sufficiente considerare il caso generale, quello in cui l'urna contiene $t$ oggetti distinguibili, $1,\ldots,t$, ai quali vengono attribuite in ogni estrazione probabilità $p_1,\ldots,p_t$ di essere estratti, $p_1+\ldots+p_t = 1$. Usando la (51) si ha allora $\pi(n_1,\ldots,n_t) = p_1^{n_1}\ldots p_t^{n_t}$ e sostituendo nella (52) si ottiene

$$P((S_n^{(1)}=n_1)\wedge\ldots\wedge(S_n^{(t)}=n_t)) = \frac{n!}{n_1!\ldots n_t!}\,p_1^{n_1}\ldots p_t^{n_t},\ \ n_1+\ldots+n_t=n. \qquad (53)$$
$$p_i>0,\ i=1,\ldots,n,\ \ p_1+\ldots+p_n=1.$$

che è la **distribuzione multinomiale** – su $\mathbb{P}_1\wedge\ldots\wedge\mathbb{P}_n$ –.

### 14.3.2 Estrazione senza rimessa. Distribuzione ipergeometrica multipla.

Generalizzando il discorso fatto per il modello senza rimessa nel caso $t=2$, supponiamo qui che l'urna contenga $m_1$ oggetti di tipo $1,\ldots,m_t$ oggetti di tipo $t$ e poniamo $m=m_1+\ldots+m_t$. Poiché l'oggetto estratto non viene rimesso nell'urna, nella (51) i casi possibili sono tutti e soli quelli per cui riesce $0<n_i\leq m_i$ per ogni $i=1,\ldots,t$. Soddisfatte queste condizioni, al denominatore del secondo membro della (51) troviamo $n$ fattori decrescenti di 1 a partire da $m$, e al numeratore ancora $n$ fattori, di cui $n_i$ decrescenti di 1 a partire da $m_i$ (generalmente non consecutivi), $i=1,\ldots,t$. Pertanto si ha $\pi(n_1,\ldots,n_t) = (m_1)_{n_1}\ldots(m_t)_{n_t}/(m)_n$, e perciò si trova

$$P((S_n^{(1)}=n_1)\wedge\ldots\wedge(S_n^{(t)}=n_t)) = \frac{n!}{n_1!\ldots n_t!}\,\frac{(m_1)_{n_1}\ldots(m_t)_{n_t}}{(m)_n},$$

che si può anche scrivere

$$P((S_n^{(1)}=n_1)\wedge\ldots\wedge(S_n^{(t)}=n_t)) = \binom{m_1}{n_1}\ldots\binom{m_t}{n_t}\Big/\binom{m}{n} \qquad (54)$$
$$n_i\leq m_i,\ i=1,\ldots,t,\ n_1+\ldots+n_t=n,\ m_1+\ldots+m_t=m.$$

La (54) è la generalizzazione della (49) e prende il nome di **distribuzione ipergeometrica multipla**. Analogamente a quanto visto nel caso $t = 2$ per la (49), è facile verificare che si perviene allo stesso risultato per la probabilità dell'evento $E^{(n)}_{n_1,\dots n_t}$ = *su n oggetti estratti tutti in una volta da un'urna contenente $m_1$ oggetti di tipo 1, ... , $m_t$ di tipo t, $m=m_1+\dots+m_n$, $n_1$ sono di tipo 1 , ... , $n_t$ di tipo t,* nell'ipotesi che i sottoinsiemi di $n$ oggetti che si possono formare con gli $m$ disponibili abbiano uguale probabilità di essere estratti (scelta a caso dei sottoinsiemi di $n$ oggetti). Pertanto:

> *Se non interessa l'ordine d'estrazione, il modello di scelta sequenziale senza rimessa di n oggetti a caso uno per volta da un'urna che ne contiene m, può essere sostituito nei calcoli dal modello di scelta a caso di un sottoinsieme di n oggetti tutti in una volta.*

Il modello senza rimessa può essere generalizzato in termini di pesi in modo analogo a quelli visti per il modello con rimessa (per $t\geq 2$) e per quello con contagio (per $t=2$). Basta supporre che dopo ogni estrazione i pesi degli oggetti vengano *decrementati* (di -1 nel modello *base*), anziché incrementati o lasciati invariati. Si ottiene così un modello con contagio *negativo* che si presta ad essere studiato assieme a quello con contagio positivo, come in effetti faremo nel prossimo numero.

### 14.3.3 Estrazione con contagio. Distribuzione di Pólya multipla.

Ferme restando le ipotesi fatte nel numero precedente sulla composizione iniziale dell'urna, supponiamo ora che dopo ogni estrazione l'oggetto estratto venga reimbussolato assieme a uno dello stesso tipo. Ragionando in modo analogo al caso $t=2$ allora si ottiene:

$$\pi(n_1,\dots,n_t) = \frac{[m_1\dots(m_1+n_1-1)]\dots[m_t\dots(m_t+n_t-1)]}{m\dots(m+n-1)} = \frac{(m_1+n_1-1)_{n_1}\dots(m_t+n_t-1)_{n_t}}{(m+n-1)_n}$$

Sostituendo nella (52) si trova

$$P((S_n^{(1)}=n_1)\wedge\dots\wedge(S_n^{(t)}=n_t)) = \binom{m_1+n_1-1}{n_1}\dots\binom{m_t+n_t-1}{n_t}\Big/\binom{m+n-1}{n}, \quad (55)$$

$n_1+\dots+n_s=n$, che è la **distribuzione di Pólya multipla** relativa al modello base.

Passando al caso generale, come nel caso con rimessa consideriamo un'urna contenente $t$ oggetti distinti, $1,\dots,t$, una distribuzione di probabilità su $\mathbb{P}_1=\{\omega_1^{(1)},\dots,\omega_t^{(1)}\}$ assegnata mediante la funzione peso $b_1,\dots,b_t$ e un peso $a>0$ destinato a incrementare il peso relativo all'ultimo oggetto estratto prima dell'estrazione successiva. Posto $b=b_1+\dots+b_t$, è $p_1=b_1/b,\dots,p_t=b_t/b$

la distribuzione iniziale. All'atto della $j$-esima estrazione ($j>1$) il peso complessivo della distribuzione di $P_j | \omega_{h_1}^{(1)} \wedge \ldots \wedge \omega_{h_{j-1}}^{(j-1)}$ è $b+(j-1)a$, di cui $b_i + j_i a$ è la parte che compete all'evento $\omega_i^{(j)} | \omega_{h_1}^{(1)} \wedge \ldots \wedge \omega_{h_{j-1}}^{(j-1)}$ se in precedenza è stato estratto $j_i$ volte l'oggetto di tipo $i$, $i = 1, \ldots, t$. Allora riesce:

$$P([\omega_{h_1}^{(1)} \wedge \ldots \wedge \omega_{h_n}^{(n)}]_{n_1,\ldots,n_t}) = \frac{[b_1 \ldots (b_1+(n_1-1)a)] \ldots [b_t \ldots (b_t+(n_t-1)a)]}{b \ldots (b+(n-1)a)} .$$

Moltiplicando a numeratore e denominatore ciascuno degli $n$ fattori per $-1/a$ e ponendo $\gamma = a/b$ – per cui si ha $b_i/a = p_i/\gamma$, $i = 1, \ldots, t$ – da qui si ricava:

$$P([\omega_{h_1}^{(1)} \wedge \ldots \wedge \omega_{h_n}^{(n)}]_{n_1,\ldots,n_t}) = \frac{(-p_1/\gamma)_{n_1} \ldots (-p_t/\gamma)_{n_t}}{(-1/\gamma)_n} = \pi(n_1, \ldots, n_t).$$

Sostituendo nella (52) si trova infine la seguente formula generale della distribuzione di Pólya multipla:

$$P((S_n^{(1)} = n_1) \wedge \ldots \wedge (S_n^{(t)} = n_t)) = \binom{-p_1/\gamma}{n_1} \ldots \binom{-p_t/\gamma}{n_t} \Big/ \binom{-1/\gamma}{n}. \qquad (55')$$

$$\gamma > 0,\ p_i > 0,\ \ i = 1, \ldots, n,\ \ p_1 + \ldots + p_t = 1,\ \ n_1 + \ldots + n_t = n.$$

Come preannunziato al termine del numero precedente, si possono considerare anche modelli con contagio negativo. Manteniamo per questo le ipotesi precedenti sulla composizione dell'urna e sulla funzione peso della distribuzione iniziale, ma diamo adesso ad $a$ il significato di decremento di peso. Con ragionamento analogo a quello del caso di incremento positivo ora si trova:

$$P([\omega_{h_1}^{(1)} \wedge \ldots \wedge \omega_{h_n}^{(n)}]_{n_1,\ldots,n_t}) = \frac{[b_1 \ldots (b_1-(n_1-1)a)] \ldots [b_t \ldots (b_t-(n_t-1)a)]}{b \ldots (b-(n-1)a)} ,$$

purché sia $b_i-(n_i-1)a \geq 0$, $i = 1, \ldots, n$. Moltiplicando qui a numeratore e denominatore ogni fattore per $1/a$ ( invece che per $-1/a$) si ricava

$$P((S_n^{(1)} = n_1) \wedge \ldots \wedge (S_n^{(t)} = n_t)) = \binom{p_1/\gamma}{n_1} \ldots \binom{p_t/\gamma}{n_t} \Big/ \binom{1/\gamma}{n}. \qquad (56)$$

ove il significato di $p_i$, $n_i$, $n$ e $\gamma$ è quello che hanno nella (55').

### 14.3.4 Esempio.

Riprendiamo 4.1.1 *Esempio* 12 – urna contenente 1 pallina bianca (oggetto di tipo 1), 2 nere (oggetti di tipo 2), 3 rosse (oggetti di tipo 3) e 4 verdi (oggetti di tipo 4) – e poniamo gli stessi quesiti, con riferimento però a tutte tre le modalità d'estrazione. Si tratta di calcolare la probabilità che in una sequenza

di 4 estrazioni di una pallina per volta le palline estratte siano tutte bianche (evento $E_1$), dello stesso colore (evento $E_2$), di colore diverso (evento $E_3$), una bianca, due verdi e una di altro colore (evento $E_4$).
Usando la notazione dei paragrafi precedenti abbiamo $t=4$, $m_1=1$, $m_2=2$, $m_3=3$, $m_4=4$, $m=10$ e $p_1=1/10$, $p_2=2/10$, $p_3=3/10$, $p_4=4/10$. I quattro eventi inoltre si scrivono:

$$
\begin{aligned}
E_1 &= (S_4^{(1)}=4) = (S_4^{(1)}=4) \wedge (S_4^{(2)}=0) \wedge (S_4^{(3)}=0) \wedge (S_4^{(4)}=0) \\
E_2 &= \vee_{i=1}^{4} (S_4^{(i)}=4) = \vee_{i=1}^{4} [(S_4^{(i)}=4) \wedge \wedge_{h\neq i}^{1,4} (S_4^{(h)}=0)] \\
E_3 &= \wedge_{i=1}^{4} (S_4^{(i)}=1) \\
E_4 &= (S_4^{(1)}=1) \wedge (S_4^{(4)}=2) \wedge ((S_4^{(2)}=1) \vee (S_4^{(3)}=1)) = \\
&\quad (S_4^{(1)}=1) \wedge (S_4^{(4)}=2) \wedge ([(S_4^{(2)}=1) \wedge (S_4^{(3)}=0)] \vee [(S_4^{(2)}=0) \wedge (S_4^{(3)}=1))
\end{aligned}
$$

Per calcolare le loro probabilità useremo la (53) per il modello con rimessa, la (54) per il modello senza rimessa e la (55) per quello con contagio.

In relazione all'evento $E_1$ riesce $n_1=4$, $n_2=n_3=n_4=0$. Allora si ha:

$$P(E_1) = \frac{4!}{4!0!0!0!} \left(\frac{1}{10}\right)^4 \left(\frac{2}{10}\right)^0 \left(\frac{3}{10}\right)^0 \left(\frac{4}{10}\right)^0 = 0{,}0001$$

nel modello con rimessa, $P(E_1)=0$ nel modello senza rimessa perché $n_1=4>1=m_1$ e

$$P(E_1) = \binom{4}{4}\binom{1}{0}\binom{2}{0}\binom{3}{0}\Big/\binom{13}{4} = \frac{1}{715} = 0{,}0014$$

in quello con contagio (unitario).

Come si vede, ai fini del calcolo si possono trascurare i fattori corrispondenti agli eventi $S_4^{(i)}=0$, cosa che d'ora in poi faremo sistematicamente.

*Probabilità di $E_2$.*
Riesce $P(E_2)=\sum_{i=1}^{4} P((S_4^{(i)}=4) \wedge \wedge_{h\neq i}^{1,4} (S_4^{(h)}=0))$. Nelle modalità con rimessa e con contagio nell'ordine si ha:

$$P(E_2) = \sum_{i=1}^{4} \frac{4!}{4!} \left(\frac{i}{10}\right)^4 = \frac{1+16+81+256}{10000} = 0{,}0354,$$

$$P(E_2) = \sum_{i=1}^{4} \left[\binom{i+3}{4}\Big/\binom{13}{4}\right] = \frac{1+5+15+35}{715} = 0{,}0783.$$

Nel modello senza rimessa riesce $(S_4^{(1)}=4) = (S_4^{(2)}=4) = (S_4^{(3)}=4) = \phi$ e perciò $E_2=(S_4^{(4)}=4)$ e allora si trova $P(E_2)=\binom{4}{4}/\binom{10}{4}=1/210=0{,}0048$.

*Probabilità di $E_3$.*
L'evento $E_3$ è possibile in tutti i tre modelli e si ha:

$$P(E_3) = \frac{4!}{1!\,1!\,1!\,1!}\,\frac{1}{10}\,\frac{2}{10}\,\frac{3}{10}\,\frac{4}{10} = 0{,}0576 \qquad \text{(con rimessa),}$$

$$P(E_3) = \binom{1}{1}\binom{2}{1}\binom{3}{1}\binom{4}{1}\Big/\binom{13}{4} = 0{,}0336 \qquad \text{(con contagio),}$$

$$P(E_3) = \binom{1}{1}\binom{2}{1}\binom{3}{1}\binom{4}{1}\Big/\binom{10}{4} = 0{,}1143 \qquad \text{(senza rimessa).}$$

*Probabilità di $E_4$.*
Anche l'evento $E_4$ è possibile in tutti i tre modelli. Ora si trova:

$$P(E_4) = \frac{4!}{1!\,2!\,1!}\,\frac{1}{10}\left(\frac{4}{10}\right)^2\frac{2}{10} + \frac{4!}{1!\,2!\,1!}\,\frac{1}{10}\left(\frac{4}{10}\right)^2\frac{3}{10} = 0{,}0960 \qquad \text{con rimessa),}$$

$$P(E_4) = \binom{1}{1}\binom{5}{2}\binom{2}{1}\Big/\binom{13}{4} + \binom{1}{1}\binom{5}{2}\binom{3}{1}\Big/\binom{13}{4} = 0{,}0699 \quad \text{(con contagio),}$$

$$P(E_4) = \binom{1}{1}\binom{4}{2}\binom{2}{1}\Big/\binom{10}{4} + \binom{1}{1}\binom{4}{2}\binom{3}{1}\Big/\binom{10}{4} = 0{,}1429 \quad \text{(senza rimessa).}$$

## 14.4 Modelli di collocazione

Sono numerose le circostanze in cui interessa indagare sulla ripartizione degli *individui* di una popolazione in classi in ragione di certe loro *caratteristiche.* Ad esempio, sulla ripartizione degli *italiani* per *sesso,* o *peso,* o *altezza* o *classe di reddito;* oppure sulla ripartizione degli *articoli* di una qualche produzione in base a fissati *livelli di qualità;* oppure ancora sulla suddivisione dei *sinistri* denunciati a una compagnia di assicurazione per *giorni della settimana* o, se si tratta di assicurazione auto, per *cilindrata* e/o *zona di circolazione;* e cosi via.

In questa e in analoghe circostanze, un'indagine capillare permetterebbe ovviamente di pervenire ad una *distribuzione statistica* della popolazione in base alle caratteristiche desiderate, e non vi sarebbe allora alcunché d'incerto. Ma è altresì evidente, che in presenza di popolazioni numerose l'osservazione e classificazione di ogni individuo è spesso praticamente irrealizzabile, o per motivi economici e/o per ragioni di tempestività (si pensi ad esempio alle previsioni dei risultati elettorali, fatte subito dopo la chiusura dei seggi). In altre casi può anche essere che non sia troppo oneroso pervenire ad una conoscenza completa della distribuzione statistica, ma che ciò abbia interesse in una prospettiva induttiva: in vista cioè di utilizzare l'informazione acquisita per ottenere valutazioni nei riguardi delle distribuzioni statistiche di situazioni analoghe future, o comunque incerte. Come ad esempio quella della classificazione degli incidenti per cilindrata e zona di circolazione, sicuramente non proibitiva con l'uso dei moderni elaboratori elettronici e di evidente interesse per la compagnia ai fini di una valutazione del fabbisogno per la copertura degli impegni per il prossimo o i prossimi esercizi.

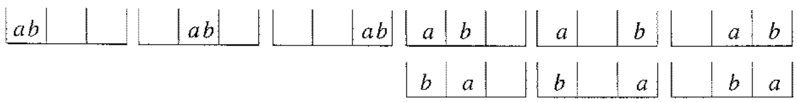
Come detto all'inizio, e come ormai facilmente s'intende, gli esempi potrebbero continuare numerosi; ma non è il caso di insistere. Interessa invece segnalare, qui, che tutte queste situazioni si prestano ad essere descritte utilizzando uno schema comune. Suddivisa la caratteristica che interessa esaminare in «livelli» – maschio e femmina per il sesso, scale numeriche per il peso (altezza, reddito), giorni della settimana (o mesi) per le denuncie di sinistro, eccetera – si può infatti immaginare di pervenire alla classificazione della popolazione assimilando i «livelli» a cassetti e collocando un individuo (oggetto) per volta nel cassetto corrispondente al suo livello. La scelta del cassetto per un individuo è obbligata se si conosce il suo livello (se è stato osservato). Altrimenti, quando il modello è usato per descrivere una situazione incerta, i criteri di scelta saranno probabilistici.

Un *modello di collocazione* quale quello appena delineato può essere ricondotto facilmente a un *modello d'estrazione*. Se sono $n$ gli individui della popolazione e $s$ i livelli della caratteristica, tale modello prevede infatti che il risultato di una collocazione si ottenga operando $n$ *scelte* di un cassetto (una per ogni individuo). Il modello di collocazione diventa allora modello d'estrazione se immaginiamo un'urna contenente gli $s$ cassetti e usiamo gli $n$ individui come indici delle estrazioni numerandoli, dando a loro un *nome* e un *ordinamento*. I cassetti (oggetti nell'urna) vanno considerati a due a due distinti, perché ciascuno di essi è usato per individuare un livello della caratteristica che si vuole classificare. Sicché, l'urna che serve a descrivere il modello di collocazione mediante un modello d'estrazione, contiene $s$ oggetti di $s$ tipi. Le modalità di collocazione, tradotte in modalità d'estrazione, saranno poi quelle che permetteranno di dare una struttura probabilistica al modello. Di attribuire cioè la probabilità di scelta dei singoli cassetti per la collocazione sequenziale degli $n$ oggetti.

### 14.4.1 Aspetti descrittivi dei modelli di collocazione.

Poiché i modelli di collocazione possono venire interpretati come modelli d'estrazione con più alternative, per le notazioni ci riferiremo a quelle dl precedente § 14.3, adeguando però la terminologia. Occorre per questo anzitutto dare individualità ai cassetti e lo facciamo numerandoli da 1 a $s$. Ciò posto, il generico costituente $\omega_h^{(i)}$ della partizione $\mathbb{P}_i$, che nel § 14.3 è definito dalla proposizione *nell'estrazione i-esima esce un oggetto di tipo h,* si esprime qui in modo più appropriato con la proposizione *l'oggetto i-esimo viene collocato nel cassetto h.* I costituenti della partizione $\mathbb{P}_1 \wedge \ldots \wedge \mathbb{P}_n$ descrivono il risultato della collocazione in modo dettagliato. Il generico di essi, $\omega_{h_1}^{(1)} \wedge \ldots \wedge \omega_{h_n}^{(n)}$, dice cioè che il primo individuo viene collocato nel cassetto $h_1$, il secondo nel cassetto $h_2$, ... , l'$n$-esimo nel cassetto $h_n$. Il che si può indicare in modo più suggestivo ponendo $(h_1, \ldots, h_n) = \omega_{h_1}^{(1)} \wedge \ldots \wedge \omega_{h_n}^{(n)}$ e dando ai costituenti il nome di **collocazioni**. Le collocazioni vengono così descritte dalle disposizione con ripetizione $(h_1, \ldots, h_n)$ di classe $n$ ($n$-ple ordinate) dei numeri $1, \ldots, s$. Data $(h_1, \ldots, h_n)$, la corrispondente collocazione si ottiene ponendo l'oggetto $i$-esimo nel cassetto $h_i$, $i = 1, \ldots, n$. Viceversa, data una collocazione, si risale alla disposizione che la descrive ponendo al posto $i$-esimo della corrispondente disposizione il numero del cassetto che contiene l'oggetto $i$. Il numero massimo delle collocazioni possibili – quello dei costituenti di $\mathbb{P}_1 \wedge \ldots \wedge \mathbb{P}_n$ – è quindi $s^n$.

Vediamo di chiarire la questione riprendendo in esame 4.1.1 *Esempio* 13, in cui è prevista la collocazione di $n = 2$ palline $a$, $b$ (1, 2 nella notazione precedente) in $s = 3$ cassetti 1, 2, 3, che abbiamo ivi rappresentato graficamente nell'ordine con |__|__|__|. Nell'ipotesi che sia possibile collocare entrambe le palline anche in un solo cassetto le collocazioni possibili – già viste allora, ma che conviene qui riportare – sono:



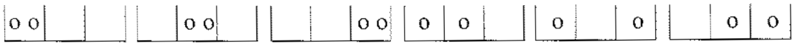
Le corrispondenti descrizioni in termini di disposizioni con ripetizione – di classe 2 dei numeri (cassetti) 1, 2 ,3 – sono le seguenti:

| | | | | | |
|---|---|---|---|---|---|
| (1,1) | (2,2) | (3,3) | (1,2) | (1,3) | (2,3) |
| | | | (2,1) | (3,1) | (3,2) |

Usando i simboli dei modelli d'estrazione gli individui (oggetti) $a$ e $b$ diventano indici della prima e seconda estrazione, alle quali vengono associate le partizioni $\mathbb{P}_a = \{\omega_1^{(a)}, \omega_2^{(a)}, \omega_3^{(a)}\}$ e $\mathbb{P}_b = \{\omega_1^{(b)}, \omega_2^{(b)}, \omega_3^{(b)}\}$. Si ha allora $\omega_1^{(a)} \wedge \omega_1^{(b)} = (1,1)$, $\omega_1^{(a)} \wedge \omega_2^{(b)} = (1,2)$, $\omega_3^{(a)} \wedge \omega_1^{(b)} = (3,1)$, eccetera.

In 4.1.1 *Esempio* 13 si è considerato anche il caso in cui i cassetti possono contenere al più una pallina. Le prime tre collocazioni – quelle corrispondenti alle disposizioni (1,1), (2,2), (3,3) – diventano allora impossibili.

Sin qui il discorso riguarda collocazioni di oggetti *distinguibili.* Se questi sono *indistinguibili,* non è possibile tenere conto dell'ordine in cui gli oggetti vengono collocati. La rappresentazione grafica delle collocazioni possibili deve allora essere modificata. Con riferimento al nostro esempio si avrà allora la seguente rappresentazione:



In generale, la partizione che descrive la collocazione di $n$ oggetti indistinguibili in $s$ cassetti è la $\{(S_n^{(1)}=n_1)\wedge\ldots\wedge(S_n^{(s)}=n_s)\ :\ n_1+\ldots+n_s=n\}$, nella quale il generico fattore $S_n^{(i)}=n_i$ significa ora *il cassetto i-esimo contiene* $n_i$ *oggetti* (in termini d'estrazione: *i successi di tipo i sono* $n_i$). Per avere una notazione più adatta ai problemi di collocazione viene naturale di porre $[n_1, \ldots, n_s]=(S_n^{(1)}=n_1)\wedge\ldots\wedge(S_n^{(s)}=n_s)$. Per indicare la generica **occupazione** dei cassetti si viene così a disporre di un simbolo meno ingombrante, in cui si legge che il primo cassetto è occupato da $n_1$ oggetti (che è vuoto se $n_1=0$), il secondo da $n_2$ oggetti, ... , l's-esimo da $n_s$ oggetti. Per questo motivo daremo a $n_1, \ldots, n_s$ il nome di *numeri di occupazione*. Circa il numero delle possibili occupazioni, esso è dato da $\binom{s+n-1}{n}$, con $n=n_1+\ldots+n_s$. Si tratta delle combinazioni con ripetizione di $s$ elementi di classe $n$ (gli $s$ cassetti che vengono scelti $n$ volte). Proveremo questo risultato indirettamente in 14.4.2*Nota*. Osserviamo ancora qui, invece, che il numero delle occupazioni possibili del nostro esempio è $\binom{3+2-1}{2}=6$ e che con la nuova notazione esse si scrivono:

$$[2, 0, 0],\ [0, 2, 0],\ [0, 2, 0],\ [1, 1, 0],\ [1, 0, 1],\ [0, 1, 1].$$

### 14.4.2 Le statistiche della fisica delle particelle.

In questo numero riportiamo a titolo di esempio alcuni modelli di occupazione che vengono utilizzati in fisica per studiare il comportamento di sistemi di particelle subatomiche: elettroni, fotoni, neutroni, nuclei, protoni, ... . Per pervenire ad una modellizzazione, lo spazio delle fasi del sistema viene suddiviso in un numero convenientemente grande di *celle* – i cassetti – destinate a ricevere le particelle (gli individui di una popolazione). Il comportamento dei sistemi non è comune a tutti i tipi di particelle. Per questo motivo i fisici hanno proposto, per lo studio di questi sistemi, più di un modello. Tutti vengono comunque ricondotti ad un comune schema descrittivo. Le particelle sono considerate ovviamente indistinguibili. Non sarebbe infatti realistico supporre di essere in grado di identificarle materialmente, dando a loro un nome. Nulla impedisce però di immaginare di farlo, se ciò serve per descrivere le modalità di collocazione e per introdurre nel modello la struttura probabilistica. Noi lo faremo per tutti i tre modelli che andremo a studiare qui di seguito. In primo approccio descriveremo cioè l'operazione di collocazione mediante la partizione $\mathbb{P}_1 \wedge \ldots \wedge \mathbb{P}_n$, spiegandola così come se le particelle fossero distinguibili e collocate nelle celle una alla volta in un ordine assegnato. La struttura probabilistica viene allora introdotta stabilendo la modalità d'estrazione sequenziale delle celle. Noi considereremo qui le tre modalità d'estrazione base – con e senza rimessa, con contagio –, trovando per questa via tre modelli classici della fisica delle particelle, che sono stati però introdotti originariamente assumendo opportune ipotesi di simmetria. Poiché le particelle sono indistinguibili, ciò che concretamente interessa calcolare, nelle tre modalità d'estrazione suddette, è la distribuzione di probabilità sull'insieme della occupazioni $\{[n_1, \ldots, n_s]_\mathrm{n} : n_1 + \ldots + n_s = n\}$. Per un adeguato commento dei modelli interesserà tuttavia considerare anche la distribuzione di probabilità sull'insieme delle collocazioni, cioè sulla partizione $\mathbb{P}_1 \wedge \ldots \wedge \mathbb{P}_n$. Ricordiamo in proposito che abbiamo indicato con $\pi(n_1, \ldots, n_s)$ la probabilità del suo generico costituente $[\omega_{h_1}^{(1)} \wedge \ldots \wedge \omega_{h_n}^{(n)}]_{\mathrm{n}_1, \ldots, \mathrm{n}_s}$, che qui dice, tenendo conto dell'ordine, che $n_1$ particelle vengono collocate nella cella 1, … , $n_s$ nella cella $s$.

*Statistica di Maxwell-Boltzmann.*

Supponiamo che la cella assegnata ad ogni particella sia scelta a caso tra *tutte* le $s$ disponibili, occupate e non. Il corrispondente modello d'estrazione è allora quello sequenziale con rimessa; esso prevede perciò che si eseguano $n$ estrazioni con reimbussolamento di una cella per volta, da un'urna che ne contiene $s$ distinguibili. La distribuzione di probabilità sull'insieme delle occupazioni si ottiene allora ponendo nella (53) $p_i = 1/s$, $i = 1, \ldots, s$. Si trova:

$$P([n_1, \ldots, n_s]) = \frac{n!}{n_1! \ldots n_s!} \left(\frac{1}{s}\right)^n, \quad n_1 + \ldots + n_s = n.$$

In questo modello riesce $\pi(n_1, \ldots, n_s) = (1/s)^{n_1} \ldots (1/s)^{n_s} = 1/s^n$. Le $s^n$ collocazioni (gli eventi elementari di $\mathbb{P}_1 \wedge \ldots \wedge \mathbb{P}_n$) sono dunque tutte possibili ed equiprobabili. Questa condizione di equiprobabilità – conseguenza qui della modalità di collocazione – viene assunta nel modello originale direttamente per ragioni di simmetria. Anche se può sembrare a prima vista un'ipotesi ragionevole, si è rivelata però in realtà inadeguata ad interpretare in modo soddisfacente il comportamento di qualsiasi sistema di particelle subatomiche. L'ipotesi di simmetria sulle collocazioni – e perciò anche quella di collocazione sequenziale con rimessa – non corrisponde cioè al comportamento osservato di nessun sistema di particelle.

*Statistica di Fermi-Dirac.*
Supponiamo che la cella assegnata ad ogni particella sia scelta a caso tra quelle *vuote* al momento della sua collocazione. Allora, le collocazioni possibili sono tutte e sole quelle che prevedono celle occupate al più da una particella. Il corrispondente modello d'estrazione è quello sequenziale senza rimessa. Perché l'operazione non sia impossibile, bisogna perciò che sia $n \leq s$ (ma in questi modelli $n$ va pensato in realtà molto minore di $s$). La distribuzione di probabilità sull'insieme delle occupazioni si ottiene allora ponendo nella (54) $t=m=s$, $n_i \leq m_i = 1$ – sicché riesce $\binom{m_i}{n_i}=1$ –, $i=1, \ldots, s$, e quindi si trova:

$$P([n_1, \ldots, n_s]) = 1\Big/\binom{s}{n}, \;\; n_i \leq 1,\, i=1, \ldots, s,\, n_1+\ldots+n_s = n \leq s.$$

In questo modello le occupazioni sono dunque equiprobabili. Sono tali anche le collocazioni possibili, che sono in numero di $(s)_n$ – si può assegnare una cella in $s$ modi alla particella 1, in $s-1$ modi alla particella 2, ..., in $s-n+1$ modi alla particella $n$ – e riesce $\pi(n_1,\ldots,n_t) = 1/(s)_n$ come si vede ponendo $t=m=s$ e $m_1=\ldots=m_s=1$ nell'espressione di $\pi(n_1,\ldots,n_t)$ del n° 14.3.2.
Come già segnalato, la condizione di simmetria per le occupazioni è assunta come ipotesi. A titolo di curiosità segnaliamo ancora che questo modello è ad esempio adatto per spiegare il comportamento di sistemi di elettroni, di neutroni e di protoni. Più in generale, le particelle che obbediscono alla statistica di Fermi-Dirac sono dette fermioni.
Vale anche la pena di osservare che il problema che stiamo trattando può essere interpretato come una generalizzazione dell'estrazione del lotto su una ruota. Al posto di estrarre 5 dei 90 numeri presenti in un'urna, qui si immagina di scegliere $n$ di $s$ celle disponibili. Se queste non sono distinguibili – com'è nel nostro caso – o non interessa l'ordine d'estrazione si ottiene che le combinazioni di classe $n$ di $s$ oggetti distinguibili (qui di celle) sono equiprobabili.

*Statistica di Bose-Einstein.*
Nel modello Fermi-Dirac, una cella non può essere occupata da più di una particella. Possiamo pensare che questo accada perché le particelle sono dotate di una forza repulsiva in grado di impedire alle altre particelle l'ingresso in una cella già occupata (da una particella). Il modello che andiamo ora ad esaminare si lascia invece descrivere immaginando che le particelle si attraggano con una forza traducibile nel corrispondente modello d'estrazione in termini di contagio unitario. Che il problema di occupazione possa cioè essere assimilato a un modello d'estrazione di Pólya da un'urna contenente $s$ celle, in cui è 1 sia il peso iniziale di ciascuna cella sia l'incremento di peso della cella estratta in ogni singola estrazione. La distribuzione di probabilità sull'insieme delle occupazioni si ottiene ora ponendo nella (55) $t=m=s$, $m_i=1$, $i=1, \ldots, s$, e

quindi si ricava:

$$P([n_1, \ldots, n_s]) = 1 \Big/ \binom{s+n-1}{n}, \quad n_1 + \ldots + n_s = n$$

Come nel precedente modello di Fermi-Dirac, anche in questo le occupazioni possibili sono equiprobabili. Ora però il loro numero è maggiore, perché ogni cella può essere occupata da più particelle (anche da tutte). Questo modello si è mostrato adatto per la descrizione del comportamento di alcuni sistemi di particelle, quali ad esempio i fotoni e gli atomi di idrogeno. Le particelle che obbediscono alla statistica di Bose-Einstein vengono denominate bosoni.

NOTA. *Formule combinatorie.*

Sappiamo che le valutazioni di probabilità sulla partizione $\mathbb{P}_1 \wedge \ldots \wedge \mathbb{P}_n$ che si ottengono con procedura sequenziale sono coerenti (13.2.5 *Teorema*) e che sono perciò coerenti anche quelle che si ottengono a partire da queste per la partizione $\{(S_n^{(1)} = n_1) \wedge \ldots \wedge (S_n^{(s)} = n_s) : n_1 + \ldots + n_s = n\}$, meno fine della prima. Vale 1 perciò la somma delle probabilità dei loro eventi elementari. Questo risultato può essere usato efficacemente come strumento per dimostrare in via probabilistica talune formule di tipo combinatorio. Vediamo qualche esempio.

Nel caso delle statistiche di Fermi-Dirac e di Bose-Einstein le occupazioni sono combinazioni di $s$ elementi di classe $n$, semplici le prime e con ripetizione le seconde. Poiché sono equiprobabili, la loro somma – che deve essere 1 – è uguale alla probabilità comune per il loro numero, e questo è allora il reciproco delle rispettive probabilità. Ciò dimostra che $\binom{s}{n}$ è il numero delle combinazioni semplici di $s$ elementi di classe $n$ e $\binom{s+n-1}{n}$ le combinazioni con ripetizione.

Posto nella (48) $n_0 = max(0, n-r)$ e $n_1 = min(b, n)$ e sommando per $h$ compreso tra $n_0$ e $n_1$ si ottiene la formula combinatoria:

$$\sum_{h=n_0}^{n_1} \binom{n}{h} (b)_h (r)_{n-h} = (b+r)_n .$$

Ponendo $p_i = b_i/(b_1 + \ldots + b_t)$ a secondo membro della (53) e sommando per $n_1 + \ldots + n_t = n$ si trova la seguente generalizzazione dello sviluppo del binomio di Newton.

$$\sum_{n_1 + \ldots + n_t = n} \frac{n!}{n_1! \ldots n_t!} b_1^{n_1} \ldots b_t^{n_t} = (b_1 + \ldots + b_t)^n,$$

Analogamente, sommando ancora per $n_1 + \ldots + n_t = n$ a secondo membro della (55') si ricava la formula combinatoria:

$$\sum_{n_1+\ldots+n_t=n} \binom{-p_1/\gamma}{n_1}\cdots\binom{-p_t/\gamma}{n_t} = \binom{-1/\gamma}{n}, \quad \gamma>0, p_i>0, \; p_1+\ldots+p_t=1.$$

COMPLEMENTO.

Il modello di Fermi-Dirac prevede che le particelle si respingano con intensità tale da impedire che una cella possa essere occupata da più di esse. Per quel che riguarda l'aspetto probabilistico si suppone classicamente che le occupazioni possibili – in numero di $\binom{s}{n}$ – siano equiprobabili. Come abbiamo visto, questa valutazione è equivalente a quella che si ottiene assimilando il problema di collocazione a uno di estrazione sequenziale (delle celle) senza rimessa, che può anche essere interpretato come problema di estrazione con contagio negativo *unitario* (n° 14.3.3). La valutazione della distribuzione di probabilità sulle occupazioni con procedura sequenziale consente di introdurre modelli con contagio negativo in cui le particelle si respingano con forza più debole di quella prevista dal modello di Fermi-Dirac, capace cioè di ostacolare l'ingresso delle particelle in celle già occupate, ma non di impedirlo. Analogamente, l'assegnazione della probabilità con procedura sequenziale permette di introdurre modelli con contagio positivo in cui le particelle si attraggano con forza graduata a piacere dall'entità del contagio. Per i modelli di contagio negativo si userà la formula (56) – purché siano soddisfatte per $n$ e per i numeri di occupazione $n_i$ le limitazioni ivi previste – e per quelli di contagio positivo la (55') – che non pone limiti per il numero delle particelle e per i numeri di occupazione (salvo ovviamente che per la somma di questi, che deve essere uguale al numero delle particelle – (n° 14.3.3).
A titolo di esempio consideriamo il problema di collocazione di 4 particelle in 6 celle e andiamo a calcolare la probabilità delle occupazioni in cinque ipotesi sull'entità del contagio, comprendenti i tre modelli (statistiche) studiate sopra. Osserviamo anzitutto che qualunque sia l'entità del contagio (negativo o positivo), la distribuzione iniziale è uniforme, riesce $p_i = 1/6$ e si può pensare assegnata attribuendo peso 1 alle singole celle – come già fatto del resto per il modello di Bose-Einstein –. Circa il contagio, lo esprimeremo in percentuale di 1 (del peso comune delle celle). Esso risulta perciò -1% nella statistica di Fermi-Dirac, 0% in quella di Maxwell-Boltzmann e 1% in quella di Bose-Einstein. Per questi tre casi il calcolo della distribuzione sulle occupazioni si effettua usando le formule messe in evidenza sopra per ciascuno di essi. Per gli altri due modelli abbiamo scelto un contagio negativo e uno positivo di entità pari a 0,15. Abbiamo allora $a=0{,}15$, $b_1=\ldots=b_6=1$, $b=6$, $1/\gamma=b/a=40$ e $p_i/\gamma=20/3$. Riesce $b_i-(4-1)a=0{,}55$, e perciò ogni cella ha la possibilità di essere estratta 4 volte anche nel modello con contagio negativo pari a -0,15 – le quattro particelle possono essere collocate anche in un solo

cassetto –. Usando la (56) troviamo allora che la corrispondente distribuzione sulle occupazioni è

$$P([n_1, \dots, n_4]) = \binom{20/3}{n_1} \dots \binom{20/3}{n_4} \Big/ \binom{40}{4}, \quad n_1 + \dots + n_4 = 4.$$

Quella relativa al modello con contagio +0,15 è analoga (basta sostituire 20/3 con -20/3 e 40 con -40), come segue usando la (55'). In tutti i modelli, poi, la $P([n_1, \dots, n_4])$ è una funzione simmetrica dei numeri di occupazione. Le occupazioni che hanno lo stesso insieme di numeri $\{n_1, \dots, n_4\}$ hanno cioè uguale probabilità. È sufficiente perciò riportarne una per tipo, come abbiamo fatto nella tabella più avanti, elencandole in ordine di «concentrazione» decrescente. Circa il loro calcolo, ci limitiamo qui ad indicare quello della probabilità di una delle $\binom{6}{3}\binom{5}{1} = 15$ occupazioni che concentrano tre particelle in una cella e la rimanente in un'altra cella. Per l'occupazione [3, 1, 0, 0, 0, 0] e contagio -0,15 si trova allora:

$$P([3, 1, 0, 0, 0, 0]) = \binom{20/3}{3}\binom{20/3}{1}\binom{20/3}{0}\binom{20/3}{0} \Big/ \binom{40}{4} =$$

$$\left( \frac{(20/3)(17/3)(14/3)}{3!} \times \frac{(20/3)}{1!} \times 1 \times 1 \right) \Big/ \frac{40 \cdot 39 \cdot 38 \cdot 37}{4!} = 0{,}00214.$$

Se il contagio è 0,15 si ha invece:

$$P([3, 1, 0, 0, 0, 0]) = \binom{-20/3}{3}\binom{-20/3}{1}\binom{-20/3}{0}\binom{-20/3}{0} \Big/ \binom{-40}{4} =$$

$$\left( \frac{(-20/3)(-23/3)(-26/3)}{3!} \; \frac{(-20/3)}{1!} \right) \Big/ \frac{(-40)(-41)(-42)(-43)}{4!} = 0{,}00339.$$

Come si vede nella tabella, e com'è ovvio che sia, $P([4, 0, 0, 0, 0, 0])$ cresce al crescere del contagio e $P([1, 1, 1, 1, 0, 0])$ decresce. Crescono anche le

| $a$ | $P([4, 0, 0, \dots])$ | $P([3, 1, 0, \dots])$ | $P([2, 2, 0, \dots])$ | $P([2, 1, 1, \dots])$ | $P([1, 1, 1, 1, \dots])$ |
|---|---|---|---|---|---|
| -1 | 0 | 0 | 0 | 0 | 0,06667 |
| -0,15 | 0,00029 | 0,00214 | 0,00390 | 0,00919 | 0,02161 |
| 0 | 0,00077 | 0,00309 | 0,00463 | 0,00926 | 0,01852 |
| 0,15 | 0,00214 | 0,00399 | 0,00529 | 0,00920 | 0,01601 |
| 1 | 0,00794 | 0,00794 | 0,00794 | 0,00794 | 0,00794 |

probabilità delle occupazioni [3, 1, 0, 0, 0, 0], [2, 2, 0, 0, 0, 0], meno concentrate della prima, mentre la probabilità della [2, 1, 1, 0, 0, 0] è prima crescente e poi decrescente.