# Modelli di fenomeni aleatori

Mario Puppi, Valentina Novello

2018-2019

##  1. Il modello dell'urna.

**1.0 Prerequisiti.** Serve sapere come è fatto un modello di un esperimento, spazio dei campioni e distribuzione di probabilità. Inoltre, è utile anche aver viso come si costruisce un modello, con distribuzione assegnata da numeri razionali, a partire dal modello dell'urna. Vediamo alcuni esempi per riassare i concetti base.

**1.1. Estrazione da una popolazione ed osservazione.** E' dato un insieme finito $U$, la *popolazione*, i cui elementi sono gli *individui*. Consideriamo l'esperimento di scegliere un individuo a caso dalla popolazione. Vogliamo descrivere l'esperimento con un modello probabilistico. La prima domanda che ci poniamo è

- Quali sono i risultati possibili dell'esperimento?

La risposta si chiama *spazio dei campioni* dell'esperimento, e nel nostro caso è naturale scegliere la popolazione. Supponiamo che la popolazione sia l'insieme finito

$$U = \{u_1, u_2, \dots, u_n\}$$

Possiamo chiederci ora

- Quale distribuzione di probabilità può assegnare agli eventi dello spazio dei campioni?

Se la scelta dell'individuo avviene in modo *casuale* possiamo supporre che la distribuzione di probabilità sia uniforme. Un esperimento come questo si chiama *estrazione casuale.*

$$u_1 \mapsto \frac{1}{n}, u_2 \mapsto \frac{1}{n}, \dots, u_n \mapsto \frac{1}{n}$$

Conoscere la probabilità sugli *atomi* $\{n\}$ permette di conoscere la probabilità di ogni evento, per la legge della somma.


**1.2 Modelli dell'urna.** L'esperimento che consideriamo ora è l'estrazione casuale di una pallina da un'urna. Le palline formano la popolazione, e un'ipotesi fondamentale è che siano ***individui distinguibili***, poniamo ad esempio che siano numerate da 1 a $n$. Il risultato dell'esperimento è il numero della pallina estratta, uno degli elementi dell'insieme $U = \{1, 2, \ldots, n\}$ e la distribuzione di probabilità sia uniforme:

$$1 \mapsto \frac{1}{n}, 2 \mapsto \frac{1}{n}, \ldots, n \mapsto \frac{1}{n}$$

Il più delle volte, in un esperimento non siamo interessati all'individuo che viene estratto da una popolazione, ma ad una data proprietà dell'individuo. Supponiamo che la popolazione sia un'urna di palline e che la proprietà delle palline sia il colore. Supponiamo che ciascuna delle $n$ palline dell'urna sia colorata con uno tra $k$ colori $c_1, c_2, \ldots, c_k$. Questo significa che la proprietà colore può assumere uno tra $k$ valori distinti. L'esperimento consiste nell'osservare il colore della pallina estratta e non il numero della pallina. Anzi, il numero potrebbe non esserci o non essere osservabile. Questo significa che due palline con lo stesso colore sono ***individui indistinguibili*** dal punto di vista che ci interessa.

Lo spazio dei campioni naturale dell'esperimento è quindi lo spazio dei colori $c_1, c_2, \ldots, c_k$. La distribuzione di probabilità non sarà più uniforme, i colori preponderanti saranno quelli più probabili, avremo così una nuova distribuzione sulla *variabile colore*:

$$c_1 \mapsto \frac{1}{k}, c_2 \mapsto \frac{1}{k}, \ldots, c_k \mapsto \frac{1}{k}$$

dove, i numeri $p_1, p_2, \ldots, p_k$ rappresentano la proporzione dei colori delle palline nell'urna. Chiameremo *osservazione* un esperimento di questo tipo. Un'estrazione casuale è quella particolare osservazione in cui tutte le palline hanno colori distinti.

# 2. Campionamenti

Il campionamento è il tipico esperimento usato in statistica. Sia data una popolazione $U$ e consideriamo una sequenza finita di osservazioni della popolazione, una alla volta.

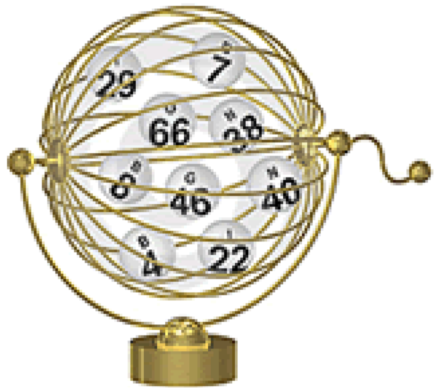
- Quale può essere lo spazio dei campioni dell'esperimento?
- Quale distribuzione di probabilità si può assegnare?

Lo spazio dei campioni dipende dalla modalità con cui avvengono le estrazioni, ne distinguiamo due fondamentali:

- ogni individuo può essere estratto più di una volta (***campionamento con restituzione***)
- ogni individuo non può essere estratto più di una volta (***campionamento senza restituzione***)

Useremo il termine campione per indicare il risultato di un campionamento. Precisamente, parleremo di *campione estratto*, per indicare il risultato di una sequenza di estrazioni, più in generale di *campione osservato* per indicare una sequenza di osservazioni.

**2.1 Campionamento con restituzione.** Un'urna contiene 100 palline e viene fatta una sequenza di 4 estrazioni di una pallina alla volta. Facciamo le seguenti




**Ipotesi** (campionamento con restituzione):

- le palline siano distinguibili, numerate da 1 a 100.
- ad ogni estrazione la pallina estratta viene rimessa nell'urna

Sia $U = \{1, 2, \ldots, 100\}$ l'insieme delle 100 palline dell'urna. Le 4 estrazioni sono 4 esperimenti che indicheremo con $X_1, X_2, X_3, X_4$. Scegliamo lo spazio dei campioni di ciascuna estrazione e la sua distribuzione di probabilità.

- Per ciascuna estrazione lo spazio dei campioni sia $U$ e la distribuzione di probabilità sia quella uniforme:
$$1 \mapsto \frac{1}{100}, 2 \mapsto \frac{1}{100}, \ldots, 100 \mapsto \frac{1}{100}$$
- Qual è lo spazio dei campioni dell'esperimento di campionamento?

Lo spazio dell'esperimento $\Omega$ è il prodotto cartesiano $U \times U \times U \times U = U^4$ Ad esempio, il campione estratto

Ad esempio, il campione estratto (1, 1, 1, 1) è il risultato dell'esperimento che corrisponde all'evento elementare è stata estratta la pallina 1 in ciascuna delle quattro estrazioni. L'evento è stata estratta la pallina 1 nella prima estrazione non è un campione, bensì è l'insieme di tutti i campioni della forma $(1, x_2, x_3, x_4)$, dove $x_2, x_3, x_4$ variano nella popolazione $U$ e lo possiamo indicare con la formula *atomica*
$$\{X_1 = 1\}$$

Abbaimo un linguaggio con cui si può descrivere ogni evento del campionamento. Infatti, ogni campione estratto può essere espresso come intersezione di formule atomiche, ad esempio (2, 51, 73, 29) si può esprimere con la formula

$$\{X_1 = 2\} \cap \{X_2 = 51\} \cap \{X_3 = 73\} \cap \{X_4 = 29\}$$

Per definire la distribuzione di probabilità del campionamento con restituzione si può ragionare in due modi. In ogni caso faremo un'ipotesi che richiama uno dei due principi di indifferenza classici:

- **Ipotesi di equiprobabilità dei campioni.** Si assume che tutti i campioni siano equiprobabili. Siccome i campioni sono in tutto 104 ne risulta la distribuzione uniforme che assegna probabilità $\frac{1}{104}$ ad ogni campione.

- **Ipotesi di indipendenza delle estrazioni.** Si assume che tutte le estrazioni $X_1, X_2, X_3, X_4$ siano indipendenti. Ciò significa che la funzione di probabilità $P$ dell'esperimento ha la proprietà moltiplicativa:

$$P(\{X_1 = 2\} \cap \{X_2 = 51\} \cap \{X_3 = 73\} \cap \{X_4 = 29\}) =$$

$$= P(\{X_1 = 2\})P(\{X_2 = 51\})P(\{X_3 = 37\})P(\{X_4 = 29\})$$

Dato che ogni estrazione è un esperimento con distribuzione uniforme arriviamo alla stessa conclusione dell'ipotesi di equiprobabilità dei campioni:

$$P(\{X_1 = 2\} \cap \{X_2 = 51\} \cap \{X_3 = 73\} \cap \{X_4 = 29\}) = (\frac{1}{100})^4$$

**2.2 Campionamento senza restituzione.** Un'urna contiene 3 palline e viene fatta una sequenza di 2 estrazioni di una pallina alla volta.

- **Ipotesi 1** Assumiamo che le palline siano distinguibili, numerate da 1 a 3.

- **Ipotesi 2 (campionamento senza restituzione)**: ad ogni estrazione la pallina estratta non viene rimessa nell'urna.

Sia $U = \{1, 2, 3\}$ l'insieme delle 3 palline dell'urna. Le 2 estrazioni, sono esperimenti che indicheremo con $X_1, X_2$.

- qual è lo spazio dei campioni di ciascuna estrazione?

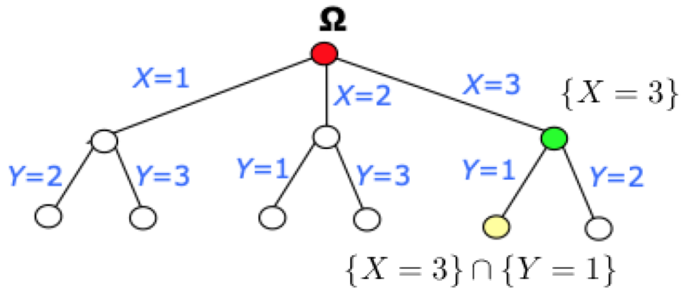
- qual è la sua distribuzione di probabilità?

L'esempio è abbastanza semplice da poter disegnare l'albero degli eventi.

Per comodità scriviamo $X$ e $Y$ per indicare le due estrazioni.

L'evento elementare $\{X = 2\} \cap \{Y = 1\}$ descrive l'estrazione del campione (2, 1).

Notiamo che l'evento $\{X = 2\} \cap \{Y = 2\}$ è impossibile.




Nell'albero degli eventi abbiamo evidenziato il campione osservato $\{X = 3\} \cap \{Y = 1\}$. Gli eventi si possono pensare come informazioni sul risultato dell'esperimento. Lo stato iniziale (conoscenza nulla) è l'evento certo $\Omega$, segue lo stato $\{X = 3\}$, la conoscenza del risultato della prima estrazione, e lo stato finale $\{X = 3\} \cap \{Y = 1\}$ conoscenza completa del risultato dell'esperimento.

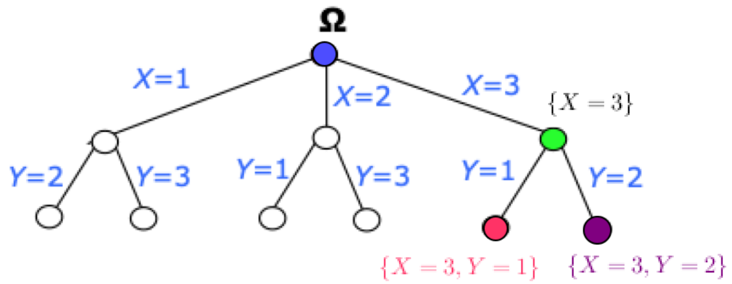
Uno spazio dei campio alternativo per questo esperimento è l'insieme potenza $U^3$. Sono punti campioni anche i risultati impossibili $(x, y)$ con $x = y$, ai quali sarà assegnata probabilità nulla.

**2.2 Distribuzione della probabilità.** Continuiamo lo studio dell'esperimento con due palline $X, Y$ estratte senza restituzione dall'urna $U = \{1, 2, 3\}$.

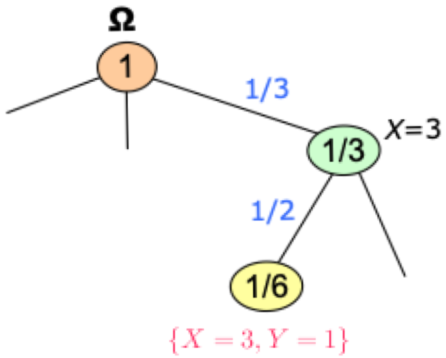
I due esperimenti $X$ e $Y$ non sono indipendenti. Infatti, se $(x, y)$ è un risultato del campionamento, gli eventi $\{X = x\}, \{Y = y\}$ non sono indipendenti. Possiamo però chiederci se siamo in grado di valutare la distribuzione di probabilità sui 6 campioni che sono i risultati possibili dell'esperimento.

Il principio di indifferenza ci suggerisce che la distribuzione sia uniforme con probabilità $\frac{1}{6}$ a ciascuno dei campioni $(x, y)$ con $x \neq y$.

Per rafforzare la comprensione di come si assegnano a priori le probabilità su un esperimento, è utile vedere che si può arrivare allo stesso risultato attraverso un altro ragionamento. Usiamo l'albero delle probabilità per capire come la conoscenza dei risultati di $X$ condiziona la probabilità di $Y$:




L'evento $\{X = 3\}$ condiziona i risultati dell'esperimento $Y$. La distribuzione di $Y$, condizionata dal verificarsi dell'evento $\{X = 3\}$, è una distribuzione uniforme che assegna il valore $\frac{1}{2}$ ciascuno dei due risultati possibili: (3, 1) e (3, 2), mentre assegna il valore 0 a tutti gli altri punti campione. Ne ricaviamo così la probabilità assoluta del campione (3, 1):




# 3.Un caso studio.

Consideriamo il fenomeno del lancio di una moneta. Per matematizzare l'esperimento è necessario fare delle ipotesi sui risultati. Non ci interessa sapere quello che succede durante il movimento in aria della moneta, l'unico dato rilevante è la faccia superiore che appare alla fine, quando la moneta si arresta. Questa semplificazione ha il vantaggio di poter definire l'esperimento come una domanda precisa: tra le due possibili facce quale si realizzerà? Le possibili risposte alla domanda, che sono soltanto due, *testaecroce*, questi sono i risultati possibili dell'esperimento. In generale, un modello di un esperimento probabilistico è una semplificazione ideale di un fenomeno reale.

**Esempio 1.** Lanciamo una moneta per 3 volte e descriviamo quello che è successo ad ogni lancio. Indichiamo con 1, 2, 3 i lanci effettuati, con $C$ oppure $T$ il risultato ottenuto ad ogni lancio.
Ad esempio, la matrice

$$\begin{array}{ccc} 1 & 2 & 3 \\ C & T & C \end{array}$$

ci informa che è uscito $C$ al lancio 1, poi $T$ al lancio 2 e di nuovo $C$ al lancio 3. Questo è un modo per raccontare la storia di un esperimento e la sequenza di $C$ oppure $T$ che abbiamo scritto è il risultato osservato. Possiamo scrivere le cose in un modo più semplice. La lista ordinata $CTC$ ci dà la stessa informazione della matrice.
I risultati possibili dell'esperimento sono le liste di tre simboli $TTC, TTT, \ldots$. Uno solo tra i risultati possibili è il risultato osservato quando l'esperimento viene eseguito. L'insieme di tutti i risultati possibili di un esperimento si chiama lo *spazio dei campioni* dell'esperimento. Nel lancio di una moneta per 3 volte, lo spazio dei campioni $\Omega$ è un insieme con $2^3 = 8$ elementi:

$$\Omega = \{CCC, CCT, CTC, CTT, TCC, TCT, TTC, TTT\}$$

**Esercizio 1.** Quale può essere lo spazio dei campioni dell'esperimento che consiste nel lanciare contemporaneamente 3 monete?

Possiamo fare delle asserzioni sull'esperimento del lancio di una moneta per 3 volte. Ad esempio, dire che abbiamo ottenuto una croce nel terzo lancio. Un'informazione come questa è quello che si chiama un *evento*. Di un evento possiamo farne un oggetto matematico identificandolo con l'insieme di tutti i modi in cui esso può accadere. Ad esempio, l'evento $E$ che afferma che al terzo lancio è uscita croce è l'insieme di tutti i risultati dell'esperimento in cui accade che al terzo lancio si realizza una croce. L'evento $E$ è il sottoinsieme $\{TTC, TCC, CTC, CCC\}$ dello spazio dei campioni.