

5.1 Elementi di base della probabilità

5.1.1 Spazio dei campioni

Un *esperimento* è un processo il cui risultato è soggetto ad incertezza.

Esistono due tipi di esperimenti: *deterministici* e *casuali*.

Per un esperimento E , l'insieme di tutti i possibili risultati è detto **spazio dei campioni** S .

Esempio 5.1 Si individui lo spazio dei campioni per il lancio di una moneta

Soluzione. Nel lancio di una moneta, se T indica ‘testa’ e C indica ‘croce’, allora $S = \{T, C\}$. Per generare lo spazio dei campioni in R per uno o due lanci di una moneta:

```
> library(prob)
>
> tosscoin(1)
```

```
##      toss1
## 1        H
## 2        T
```

```
> tosscoin(2)
```

```
##      toss1 toss2
## 1        H      H
## 2        T      H
## 3        H      T
## 4        T      T
```

Esempio 5.2 Si individui lo spazio dei campioni per il lancio di un dado non truccato

Soluzione. Il dado ha sei facce e quindi sarà $S = \{1, 2, 3, 4, 5, 6\}$. In R possiamo generare lo spazio dei campioni con il comandi

```
> library(prob)
>
> rolldie(1)
```

Esempio 5.3 Si individui lo spazio dei campioni per l'estrazione di palline da un'urna.

Soluzione.

È il tipo di esperimento più classico e importante con i numeri casuali, perché presenta diverse variazioni possibili e introduce al calcolo combinatorio. Un'urna contiene n palline indistinguibili da cui se ne estraggono k alla volta. Ci sono diverse possibili strategie: per esempio si possono reinserire o no le palline nell'urna fra un'estrazione e l'altra (*con reinserimento*), oppure, se $k > 1$, si può considerare oppure no l'ordine con cui vengono estratti i k numeri ogni volta (*con ordinamento*). Consideriamo il caso di $n = 3$ palline presenti nell'urna con l'estrazione di $k = 2$ palline per volta.

- Con ordinamento e reimmissione:

```
> library(prob)
>
> urnsamples(1:3, size = 2, replace = TRUE,
+           ordered = TRUE)

##      X1 X2
## 1    1  1
## 2    2  1
## 3    3  1
## 4    1  2
## 5    2  2
## 6    3  2
## 7    1  3
## 8    2  3
## 9    3  3
```

- Con ordinamento, senza reimmissione:

```
> library(prob)
>
> urnsamples(1:3, size = 2, replace = FALSE,
+           ordered = TRUE)
```

```
##      X1 X2
## 1    1  2
## 2    2  1
## 3    3  1
## 4    3  1
## 5    2  3
## 6    3  2
```

- Senza ordinamento, con reimmissione:

```
> library(prob)
>
> urnsamples(1:3, size = 2, replace = TRUE,
+           ordered = FALSE)
```

```
##      X1 X2
## 1    1  1
## 2    1  2
## 3    1  3
## 4    2  2
## 5    2  3
## 6    3  3
```

- Senza ordinamento, senza reimmissione:

```
> library(prob)
>
> urnsamples(1:3, size = 2, replace = FALSE,
+           ordered = FALSE)
```

```
##      X1 X2
## 1    1  2
## 2    1  3
## 3    2  3
```

5.1.2 Eventi

Un *evento* è un insieme di risultati, cioè un sottospazio dello spazio dei campioni.

In R, si può utilizzare la funzione `subset` per estrarre un sottoinsieme da un insieme.

Esempio 5.4 Considero gli eventi: ‘Estraggo una carta di cuori o di quadri’.

Soluzione.

```
> library(prob)
>
> S <- cards()
> subset(S, suit == c('Heart', 'Diamond'))
```

##	rank	suit
## 14	2	Diamond
## 16	4	Diamond
## 18	6	Diamond
## 20	8	Diamond
## 22	10	Diamond
## 24	Q	Diamond
## 26	A	Diamond
## 27	2	Heart
## 29	4	Heart
## 31	6	Heart
## 33	8	Heart
## 35	10	Heart
## 37	Q	Heart
## 39	A	Heart

Esempio 5.5 Considero gli eventi: ‘Estrarre una carta che sia un 2 oppure un 3 di qualsiasi seme’.

Soluzione.

```
> library(prob)
>
> S <- cards()
> subset(S, rank %in% 2:3)
```

##	rank	suit
## 1	2	Club
## 2	3	Club
## 14	2	Diamond
## 15	3	Diamond
## 27	2	Heart

```
## 28      3    Heart
## 40      2    Spade
## 41      3    Spade
```

Esempio 5.6 Considero l'evento: 'Escono gli stessi numeri nel lancio di due dadi'.

Soluzione.

```
> library(prob)
>
> S <- rolldie(2)
> subset(S, X1 == X2)

##      X1 X2
## 1      1  1
## 8      2  2
## 15     3  3
## 22     4  4
## 29     5  5
## 36     6  6
```

Si utilizzano quindi le operazioni fra insiemi per definire gli eventi. In particolare, dati due insiemi A e B si definiscono le seguenti operazioni (con relativa implementazione in R) che danno come risultato l'insieme C :

- **Unione.** Si indica con $C = A \cup B$. Un elemento $v \in C$ se $v \in A$ oppure $v \in B$.
- **Intersezione.** Si indica con $C = A \cap B$. Un elemento $v \in C$ se $v \in A$ e $v \in B$.
- **Differenza.** Si indica con $C = A - B$. Un elemento $v \in C$ se $v \in A$ e $v \notin B$.

```
> library(prob)
>
> S <- cards()
> A <- subset(S, suit == 'Heart')
> B <- subset(S, rank == 3)
> union(A, B)

##      rank      suit
## 2         3      Club
## 15        3 Diamond
```

```
## 27      2    Heart
## 28      3    Heart
## 29      4    Heart
## 30      5    Heart
## 31      6    Heart
## 32      7    Heart
## 33      8    Heart
## 34      9    Heart
## 35     10    Heart
## 36      J    Heart
## 37      Q    Heart
## 38      K    Heart
## 39      A    Heart
## 41      3    Spade
```

```
> intersect(A, B)
```

```
##      rank  suit
## 28      3 Heart
```

```
> setdiff(A, B)
```

```
##      rank  suit
## 27      2 Heart
## 29      4 Heart
## 30      5 Heart
## 31      6 Heart
## 32      7 Heart
## 33      8 Heart
## 34      9 Heart
## 35     10 Heart
## 36      J Heart
## 37      Q Heart
## 38      K Heart
## 39      A Heart
```

```
> setdiff(B, A)
```

```
##      rank  suit
## 2       3   Club
## 15      3 Diamond
## 41      3   Spade
```

Esistono altre funzioni R per identificare sottoinsiemi:

- Controlla il vettore elemento per elemento:

```
> library(prob)
>
> x <- 1:5
> y <- 3:6
> y %in% x

## [1] TRUE TRUE TRUE FALSE
```

- Controlla se tutto il primo vettore è contenuto nel secondo:

```
> library(prob)
>
> x <- 1:5
> y <- 3:6
> isin(x, y)

## [1] FALSE
```

5.1.3 Definizione e proprietà della probabilità

Dato un esperimento e il suo spazio dei campioni, S , obiettivo della *probabilità* è assegnare ad ogni evento A un numero $P(A)$, detto **probabilità dell'evento** A che dia una misura della possibilità che si verifichi l'evento A .

Quindi la probabilità dell'evento A è una funzione:

$$P : S(A) \longrightarrow R$$

dove $S(A)$ è l'insieme di tutti i possibili eventi A .

La probabilità soddisfa i seguenti 'assiomi' di Kolmogorov:

1. $P(A) \geq 0 \quad \forall A$
2. $P(S) = 1$
3. Se A_1, A_2, \dots, A_n sono eventi disgiunti, allora:

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i), \forall n$$

5.1.3.1 Alcune proprietà

Detto A^C l'insieme complementare di A nello spazio dei campioni S , e detti A e B due eventi, valgono le seguenti proprietà.

1. $P(A^C) = 1 - P(A)$.

Dimostrazione. Poiché $A \cup A^C = S$ e $A \cap A^C = \emptyset$,

$$1 = P(S) = P(A \cup A^C) = P(A) + P(A^C)$$

2. $P(\emptyset) = 0$.

Dimostrazione $\emptyset = S^C$.

3. Se $A \subset B$ allora $P(A) < P(B)$.

Dimostrazione. $B = A \cup (B \cap A^C)$ perciò $P(B) = P(A \cup (B \cap A^C)) = P(A) + P(B \cap A^C) \geq P(A)$ poiché $B \cap A^C \neq \emptyset \Rightarrow P(B \cap A^C) \geq 0$.

- 4.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Generalizzando a n insiemi $A_i, i = 1, \dots, n$:

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^n P(A_i \cap A_j) + \dots (-1)^{n-1} P\left(\bigcap_{i=1}^n A_i\right).$$

5.1.3.2 Il modello Equally Likely (ELM).

In alcuni tipi di esperimenti tutti gli eventi hanno uguale probabilità di accadere, allora il modello utilizzato, detto Equally Likely Model (ELM), calcola la probabilità di un evento A come:

$$P(A) = \frac{\text{numero di casi favorevoli}}{\text{numero di casi possibili}} = \frac{\#A}{\#S}.$$

Esempi classici di esperimenti di questo tipo sono il lancio di una moneta, il lancio di un dado, l'estrazione di una carta da un mazzo, ecc.

In R, per visualizzare la probabilità relativa ad un evento calcolata con ELM, si può utilizzare la funzione `probspace`, che aggiunge una colonna all'output contenente la probabilità calcolata per ogni elemento.

```
> library(prob)
>
> S <- rolldie(1)
> p <- rep(1/6, times = 6)
>
> probspace(S, probs = p)
```

```
##      X1      probs
## 1     1 0.1666667
## 2     2 0.1666667
## 3     3 0.1666667
## 4     4 0.1666667
## 5     5 0.1666667
## 6     6 0.1666667
```


In realtà molte funzioni, fra cui anche `rolldie` hanno un parametro di input che calcola la probabilità di ogni elemento dell'insieme risultato secondo un modello ELM:

```
> library(prob)
>
> rolldie(1, makespace = TRUE)

##      X1      probs
## 1  1 0.1666667
## 2  2 0.1666667
## 3  3 0.1666667
## 4  4 0.1666667
## 5  5 0.1666667
## 6  6 0.1666667
```

La funzione `urnsamples` NON ha il parametro `makespace`.

5.1.4 Metodi di conteggio

Nell'applicare il modello ELM è quindi necessario *contare* il numero di elementi di un insieme. Consideriamo quindi alcune 'tecniche' di conteggio.

Principio della moltiplicazione. Se un esperimento si compone di due fasi, di cui la prima può essere realizzata in n_1 modi e la seconda in n_2 modi, allora l'esperimento si può realizzare in $n_1 n_2$ modi. Generalizzando, se ci sono m fasi realizzabili rispettivamente in n_1, n_2, \dots, n_m modi, allora l'intero esperimento potrà essere realizzato in $n_1 n_2 \dots n_m$ modi.

Esempio 5.7 Se si deve compilare un piano di studi con 3 esami a scelta, di cui il primo in una lista di 5 esami, il secondo di 3 esami e il terzo di 7 esami, quanti piani di studio diversi si potranno avere?

Soluzione. Si potranno avere $5 \cdot 3 \cdot 7 = 105$ diversi piani di studio.

Numero di permutazioni. Il possibile numero di *permutazioni* di n elementi è $n!$.

5.1.4.1 Insiemi ordinati.

Torniamo all'esperimento di estrazione di k palline da un'urna con n indistinguibili palline, ciascuna numerata con un diverso numero in $[1, n]$. Se considero insiemi

ordinati (in cui cioè due insiemi con gli stessi k elementi ma estratti in ordine diverso sono considerati DIVERSI), il numero di risultati che posso avere dall'esperimento è:

- n^k se faccio l'esperimento *con* reimmissione
- $n(n-1) \dots (n-k+1)$ nel caso *senza* reimmissione

Il conteggio del numero di insiemi dell'esperimento dell'estrazione delle palline dall'urna si può realizzare utilizzando la funzione `nsamp`.

```
> library(prob)
>
> nsamp(n = 3, k = 2, replace = TRUE, ordered = TRUE)

## [1] 9

> nsamp(n = 3, k = 2, replace = TRUE, ordered = FALSE)

## [1] 6
```

Esempio 5.8 Si deve scrivere il programma di una due giorni di gare di nuoto. Ci sono in totale 10 gare da suddividere 3 nel primo giorno e 7 nel secondo. Quanti diversi programmi si possono fare per il primo giorno?

Soluzione. $10 \cdot 9 \cdot 8 = 720$

```
> library(prob)
>
> nsamp(n = 10, k = 3, replace = FALSE, ordered = TRUE)

## [1] 720
```

Esempio 5.9 Lancio un dado 3 volte. Quante possibili sequenze di numeri posso avere?

Soluzione. $6^3 = 108$.

5.1.4.2 Insiemi non ordinati

Sempre nell'esperimento precedente dell'estrazione di palline dall'urna, se ora considero insiemi *non ordinati* (in cui cioè due insiemi con gli stessi k elementi ma estratti in ordine diverso sono considerati UGUALI), il numero di risultati che posso avere dall'esperimento è:

- con reimmissione

$$\frac{(n-1+k)!}{(n-1)!k!},$$

- senza reimmissione (*coefficiente binomiale*)

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

In R:

```
> library(prob)
>
> nsamp(n = 3, k = 2, replace = FALSE, ordered = TRUE)

## [1] 6

> nsamp(n = 3, k = 2, replace = FALSE, ordered = FALSE)

## [1] 3
```

La funzione `nsamp` si può usare anche vettorialmente. Per esempio:

```
> library(prob)
>
> n <- c(10, 3, 5)
> k <- c(3, 3, 2)
>
> nsamp(n, k, replace = FALSE, ordered = TRUE)

## [1] 720    6    20
```

Esempio 5.10 Nel programma di gare di nuoto precedente, nella seconda giornata si devono decidere due gare da mettere nel pomeriggio. Quante possibilità di scelta ci sono?

Soluzione.

$$\binom{7}{2} = 21.$$

Esempio 5.11 Lancio un dado 3 volte. Quante possibili sequenze di numeri posso avere, senza considerare il diverso ordine dei numeri nella sequenza?

Soluzione.

$$\frac{8!}{5!3!} = 56$$

In R ci sono il comando `factorial` per calcolare il fattoriale e `choose` per calcolare il coefficiente binomiale.

Esempio 5.12 Qual è la probabilità che in un insieme di n persone, almeno due abbiano lo stesso segno zodiacale? (supponendo che la probabilità di avere un certo segno zodiacale per ogni persona segua il modello ELM).

Soluzione. Sia quindi A l'evento: 'due persone hanno lo stesso segno zodiacale', A^C (complementare di A): 'non ci sono due persone che hanno lo stesso segno zodiacale'. Vogliamo quindi determinare $P(A)$ oppure $P(A^C)$. Lo spazio dei campioni S contiene 12^n elementi, poiché ogni persona ha 12 diverse possibilità. Possiamo calcolare $P(A)$ come:

$$P(A) = 1 - P(A^C) = 1 - \frac{\#A^C}{\#S}.$$

L'insieme complementare è costituito dagli eventi in cui le persone NON hanno lo stesso segno zodiacale, poiché la prima persona ha 12 diverse possibilità, la seconda ne ha 11 e così via. Allora si ha che $\#A^C = 12 * 11 * 10 \dots (12 - n + 1)$. Se c'è una sola persona nell'insieme la probabilità è:

$$P(A) = 1 - 12/12 = 0.$$

Se ci sono almeno 13 persone la probabilità è:

$$P(A) = 1 - \frac{12 \cdot 11 \dots 1}{12^n} = 1.$$

Se ci sono, per esempio, 6 persone la probabilità è:

$$P(A) = 1 - \frac{12 \cdot 11 \dots 7}{12^6} = 0.777 \dots$$

5.1.5 Probabilità condizionata ed eventi indipendenti

Supponiamo di estrarre due carte da un mazzo. Qual è la probabilità di estrarre due carte di cuori? Per quanto riguarda la prima carta, $A = \{\text{la prima carta è di cuori}\}$ la probabilità è $1/4$. Per quanto riguarda la seconda, $B = \{\text{la seconda carta è di cuori}\}$, la probabilità di B dipende ovviamente dalla prima carta estratta. Se la prima carta NON era di cuori, la probabilità è $13/51$, ma se la prima carta era di cuori la probabilità diventa $12/51$. Quindi:

$$P(B|A) = \frac{12}{51}.$$

La seconda probabilità è quindi condizionata dal primo evento. Si parla allora di *probabilità condizionata*.

Definizione. Dati due eventi A e B , si definisce probabilità condizionata di B dato A e si scrive $P(B|A)$:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \text{ if } P(A) > 0.$$

Non consideriamo la teoria sviluppata nel caso $P(A) = 0$ in questa sede.

Nell'esempio precedente con le carte, $P(A \cap B) = 13/52 * 12/51$, $P(A) = 13/52$, allora $P(B|A) = 12/51$.

Esempio 5.13 Lancio una moneta tre volte. Lo spazio dei campioni è:

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

Sia $A = \{\text{escono due } T\}$ e $B = \{\text{escono una } T \text{ e una } C\}$. Calcolare $P(A)$ e $P(B)$.

Soluzione. Si ha che: $P(A) = 4/8 = 1/2$, $P(B) = 6/8 = 3/4$ e $P(A \cap B) = 3/8$. Quindi:

$$P(B|A) = \frac{3/8}{1/2} = \frac{3}{4}$$
$$P(A|B) = \frac{3/8}{3/4} = \frac{1}{2}$$

In R, si usa la funzione `prob` con il parametro di input `given`.

Esempio 5.14 Si supponga di lanciare un dado due volte e si considerino gli eventi: $A = \{\text{la somma dei numeri è } 9\}$, $B = \{\text{il primo dado è } 3\}$.

Soluzione.


```
> library(prob)
>
> S <- rolldie(2, makespace = TRUE)
> A <- subset(S, X1 + X2 == 9)
> B <- subset(S, X1 = 3)
> Prob(A, given = B)

## [1] 0.1111111

> Prob(B, given = A)

## [1] 1
```

Per quanto riguarda le proprietà, valgono per la probabilità condizionata esattamente le stesse proprietà che per la probabilità viste in precedenza.

Esempio 5.15 Consideriamo un'urna con 8 palline, di cui 5 bianche e 3 nere. Si estraggono due palline dall'urna. Sia $A = \{\text{la prima pallina è bianca}\}$, $B = \{\text{la seconda pallina è bianca}\}$. Calcolare la probabilità che si estraggano due palline bianche.

Soluzione.

$$P(A \cap B) = \frac{5}{8} \cdot \frac{4}{7} \simeq 0.3571 \dots$$

in R:

```
> S <- rep(c("bianco", "nero"), times = c(5, 3))
> U <- urnsamples(S, size = 2, replace = FALSE,
+               ordered = TRUE)
> P <- probspace(U)
> Prob(P, isrep(P, "bianco", 2))

## [1] 0.3571429
```

Consideriamo ora il seguente esempio.

Esempio 5.16 Si supponga di lanciare una moneta due volte. Lo spazio dei campioni è: $S=\{TT,TC,CT,CC\}$. Considerati gli eventi: $A=\{ \text{il primo lancio è } T \}$, $B=\{ \text{il secondo lancio è } T \}$, si calcoli $P(B|A)$.

Soluzione. Si ha $A \cap B = \{TT\}$ e $P(A \cap B) = 1/4$. Il calcolo di $P(B|A)$ dà

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/4}{1/2} = \frac{1}{2} = P(B)$$

Quindi l'evento A NON condiziona l'evento B . I due eventi sono quindi *indipendenti*.

Definizione. Due eventi A e B sono detti *indipendenti* se:

$$P(A \cap B) = P(A)P(B).$$

In caso contrario, sono detti *dipendenti*.

Seguono alcune proprietà.

Se A e B sono eventi indipendenti, si ha:

1. A e B^C sono indipendenti.
2. A^C e B sono indipendenti.
3. A^C e B^C sono indipendenti.

Se si hanno più di due eventi, si ha:

Definizione. A_1, A_2, \dots, A_n sono eventi indipendenti se vale:

$$P(A_1 \cap A_2 \dots \cap A_n) = P(A_1)P(A_2) \dots P(A_n)$$

Esempio 5.17 Lancio 5 monete. Qual è la probabilità di avere almeno una T ?

Soluzione. Se $A_i=\{\text{esce } T \text{ nell}'i\text{-esimo lancio}\}$, allora ogni A_i è indipendente. Sia $A=\{\text{esce almeno una } T\}$, allora $A^C=\{\text{escono tutte } C\}$.

$$\begin{aligned} P(A) &= 1 - P(A^C) \\ &= 1 - P(A_1^C \cap A_2^C \dots \cap A_5^C) \\ &= 1 - P(A_1^C)P(A_2^C) \dots P(A_5^C) \\ &= 1 - \left(\frac{1}{2}\right)^5 \end{aligned}$$


```
> library(prob)
> S <- tosscoin(5, makespace = TRUE)
> A <- subset(S, isrep(S, vals = "T", nrep = 5))
> 1-Prob(A)

## [1] 0.96875
```
