Laboratorio di R

Simulazioni Monte Carlo: probabilità condizionata

Probabilità condizionata. Se A e B sono due eventi dello spazio campionario Ω (appartenenti alla sigma-algebra $\mathcal{A}$) e $\mathsf{P}(A) > 0$, allora la *probabilità condizionata di B dato A* è definita come

$$\mathsf{P}(B|A) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(A)}.$$

La probabilità condizionata $\mathsf{P}(B|A)$ soddisfa le seguenti tre proprietà (che ricordano i tre assiomi di Kolmogorov):
i) $\mathsf{P}(A|A) = 1$;
ii) $\mathsf{P}(B|A) \geq 0$;
iii) data una successione di eventi $B_1, B_2, \ldots$ appartenenti alla sigma-algebra $\mathcal{A}$ e a due a due incompatibili ($B_i \cap B_j = \emptyset$, per $i \neq j$), si ha che $\mathsf{P}(B_1 \cup B_2 \cup \cdots |A) = \mathsf{P}(B_1|A) + \mathsf{P}(B_2|A) + \cdots$.

Legge del prodotto. Dati k eventi $A_1, A_2, \ldots, A_k$, si ha

$$\mathsf{P}(A_k \cap A_{k-1} \cap \cdots \cap A_1) = \mathsf{P}(A_k|A_{k-1} \cap \cdots \cap A_1) \cdot \mathsf{P}(A_{k-1}|A_{k-2} \cap \cdots \cap A_1) \cdots \mathsf{P}(A_2|A_1) \cdot \mathsf{P}(A_1).$$

Indipendenza tra 2 eventi. Due eventi A e B dello spazio campionario Ω (appartenenti ad $\mathcal{A}$) si dicono *indipendenti* se

$$\mathsf{P}(A \cap B) = \mathsf{P}(A) \cdot \mathsf{P}(B).$$

Se entrambi gli eventi A e B sono non nulli, ovvero se $\mathsf{P}(A) > 0$ e $\mathsf{P}(B) > 0$, valgono le seguenti relazioni di equivalenza

$$\mathsf{P}(A|B) = \mathsf{P}(A) \quad \Leftrightarrow \quad \mathsf{P}(A \cap B) = \mathsf{P}(A) \cdot \mathsf{P}(B) \quad \Leftrightarrow \quad \mathsf{P}(B|A) = \mathsf{P}(B).$$

Inoltre, se A e B sono due eventi indipendenti, allora sono indipendenti anche le coppie di eventi:

$$A \text{ e } \overline{B}, \qquad \overline{A} \text{ e } B, \qquad \overline{A} \text{ e } \overline{B}.$$

Indipendenza fra 3 eventi. Tre eventi A_1, A_2 e A_3 dello spazio campionario Ω (appartenenti ad $\mathcal{A}$) si dicono *indipendenti* se
i) sono a due a due indipendenti;
ii) $\mathsf{P}(A_1 \cap A_2 \cap A_3) = \mathsf{P}(A_1) \cdot \mathsf{P}(A_2) \cdot \mathsf{P}(A_3)$.

Indipendenza fra n eventi. Gli eventi A_1, A_2, …, A_n dello spazio campionario Ω (appartenenti ad $\mathcal{A}$) si dicono *(globalmente) indipendenti* se per ogni sottoinsieme di k eventi vale la relazione

$$\mathsf{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathsf{P}(A_{i_1}) \cdot \mathsf{P}(A_{i_2}) \cdots \mathsf{P}(A_{i_k}).$$

Partizione di eventi. Un insieme di k eventi $C_1, C_2, \ldots, C_k$ (appartenenti ad $\mathcal{A}$) costituiscono una *partizione di eventi* dello spazio campionario Ω se:
i) $C_1 \cup C_2 \cup \cdots \cup C_k = \Omega$;
ii) $C_i \cap C_j = \emptyset$, per ogni $i \neq j$.

Legge delle probabilità totali. Sia $C_1, C_2, \ldots, C_k$ una *partizione di eventi* dello spazio campionario Ω e sia A un evento (appartenente ad $\mathcal{A}$), allora

$$\mathsf{P}(A) = \sum_{j=1}^{k} \mathsf{P}(A|C_j) \cdot \mathsf{P}(C_j).$$

Formula di Bayes. Sia $C_1, C_2, \ldots, C_k$ una *partizione di eventi* dello spazio campionario Ω e sia A un evento (appartenente ad $\mathcal{A}$), allora per ogni evento C_i si ha:

$$\mathsf{P}(C_i|A) = \frac{\mathsf{P}(C_i \cap A)}{\mathsf{P}(A)} = \frac{\mathsf{P}(A|C_i) \cdot \mathsf{P}(C_i)}{\sum_{j=1}^{k} \mathsf{P}(A|C_j) \cdot \mathsf{P}(C_j)}.$$

Indipendenza condizionata tra 2 eventi. Dati tre eventi A, B e C dello spazio campionario Ω (appartenenti ad $\mathcal{A}$) si dice che A e B sono *condizionatamente indipendenti dato* C se

$$\mathsf{P}(A \cap B|C) = \mathsf{P}(A|C) \cdot \mathsf{P}(B|C).$$

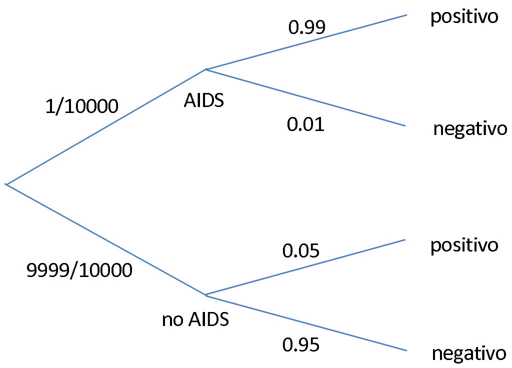
Notiamo che l'indipendenza condizionata tra due eventi A e B, dato C, non implica che questi siano anche (incondizionatamente) indipendenti, cioè che valga la relazione $\mathsf{P}(A \cap B) = \mathsf{P}(A) \cdot \mathsf{P}(B)$.

Simulazioni Monte Carlo. I metodi di simulazione Monte Carlo forniscono un potentissimo strumento per la risoluzione di problemi complessi. Fondamentalmente, si può sempre vedere una simulazione Monte Carlo come un metodo per approssimare un valore atteso attraverso una semplice media aritmetica. Tratteremo questo argomento implementando alcuni semplici esempi con il software R.

Esercizio 1 (occhiali da sole)
Un'azienda deve decidere se produrre ed immettere sul mercato dei nuovi occhiali da sole. Si ritiene che il successo tra i consumatori dei nuovi occhiali dipenda anche dalle condizioni meteorologiche della prossima stagione estiva. Si assuma che per la prossima stagione estiva ci sia una probabilità del 40% che le condizioni meteorologiche siano buone. Nel caso in cui le condizioni meteorologiche sono buone, l'azienda ritiene che ci sia una probabilità dell'80% che le vendite dei nuovi occhiali superino i 2 milioni di euro nell'arco di un anno. Nel caso in cui le condizioni meteorologiche non sono buone, l'azienda ritiene che ci sia una probabilità del 40% che le vendite dei nuovi occhiali superino i 2 milioni di euro nell'arco di un anno.
a) Si disegni il diagramma ad albero che rappresenta le possibili evoluzioni dello scenario per l'azienda e determinare le probabilità associate ad ognuno degli eventi elementari dello spazio campionario (cioè ad ognuno dei nodi terminali del diagramma ad albero).
b) Calcolare la probabilità che le vendite superino i 2 milioni di euro.
c) Sapendo che dopo un anno le vendite hanno superato i 2 milioni di euro, e non avendo nessun'altra informazione, determinare la probabilità (condizionata) che le condizioni meteorologiche siano state buone.

Esercizio 2 (AIDS)
Un soggetto scelto a caso viene sottoposto al test per l'AIDS. Si supponga che nella popolazione un individuo ogni 10000 abbia l'AIDS e che la probabilità che il test risulti positivo quando l'individuo ha l'AIDS sia 0.99, mentre la probabilità che il test risulti positivo quando l'individuo non ha l'AIDS sia 0.05. Sapendo che il soggetto è risultato positivo al test, determinare la probabilità che egli abbia effettivamente l'AIDS.

Soluzione
E' utile organizzare le informazioni del problema nel seguente albero degli eventi



Si considerino quindi i seguenti due eventi: $T = \{\text{il test è positivo}\}$ e $A = \{\text{il soggetto ha l'AIDS}\}$. Si richiede la probabilità condizionata $\mathsf{P}(A|T)$ e per il teorema di Bayes questa è data ha:

$$\begin{aligned}\mathsf{P}(A|T) &= \frac{\mathsf{P}(T|A)\cdot\mathsf{P}(A)}{\mathsf{P}(T|A)\cdot\mathsf{P}(A)+\mathsf{P}(T|\overline{A})\cdot\mathsf{P}(\overline{A})} \\ &= \frac{0.99\cdot(1/10000)}{0.99\cdot(1/10000)+0.05\cdot(9999/10000)} \simeq 0.002.\end{aligned}$$

Si noti che tutte le probabilità di interesse possono essere organizzate nella tabella

	positivo	negativo	
AIDS	$\mathsf{P}(A\cap T)$	$\mathsf{P}(A\cap\overline{T})$	$\mathsf{P}(A)$
no AIDS	$\mathsf{P}(\overline{A}\cap T)$	$\mathsf{P}(\overline{A}\cap\overline{T})$	$\mathsf{P}(\overline{A})$
	$\mathsf{P}(T)$	$\mathsf{P}(\overline{T})$	1

e queste probabilità sono pari a

	positivo	negativo	
AIDS	0.000099	0.000001	0.0001
no AIDS	0.049995	0.949905	0.9999
	0.050094	0.949906	1

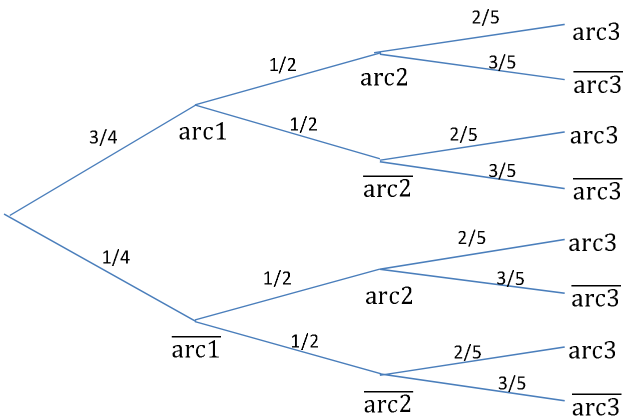
Si noti come la probabilità $\mathsf{P}(T)$ di ottenere un esito positivo al test sia praticamente uguale alla probabilità $\mathsf{P}(\overline{A}\cap T)$ di ottenere un "falso positivo".

Soluzione Monte Carlo

```
t0 <- proc.time()
salute <- c("AMMALATO","SANO"); risultato <- c("POSITIVO","NEGATIVO")
prove <- 100000; num <- 0; den <- 0

for (i in 1:prove) {
  individuo <- sample(salute, size=1, prob=c(1/10000,9999/10000))
  if (individuo=="AMMALATO")
    test <- sample(risultato, size=1, prob=c(0.99,0.01))
  if (individuo=="SANO")
    test <- sample(risultato, size=1, prob=c(0.05,0.95))

  if (test=="POSITIVO") den = den + 1
  # den = den + (test=="POSITIVO")  # alternativa

  if ((individuo=="AMMALATO")&(test=="POSITIVO")) num <- num + 1
  # num = num + ((individuo=="AMMALATO")&(test=="POSITIVO"))  # alternativa
}

cat(num,den,num/den,"\n")
t1 <- proc.time() - t0; t1
```

Esercizio 3 (arcieri)
Tre arcieri scagliano indipendentemente l'uno dall'altro una freccia ciascuno sul medesimo bersaglio. Per il primo tiratore la probabilità di centrare il bersaglio è $3/4$, per il secondo è pari a $1/2$ mentre per il terzo è $2/5$. Sapendo che una sola delle tre frecce ha colpito il bersaglio, calcolare la probabilità che sia stata lanciata dal primo tiratore.

Soluzione
Per prima cosa è utile organizzare le informazioni del problema nel seguente albero degli eventi



dove arc1 sta ad indicare che il primo arciere centra il bersaglio (e $\overline{\text{arc1}}$ che non lo centra); analogamente arc2, $\overline{\text{arc2}}$, arc3 e $\overline{\text{arc3}}$ per il secondo e terzo arciere. È chiaro che lo spazio campionario è formato da otto eventi elementari che possiamo indicare con: arc1-arc2-arc3, arc1-arc2-$\overline{\text{arc3}}$, ..., $\overline{\text{arc1}}$-$\overline{\text{arc2}}$-$\overline{\text{arc3}}$.
Definendo con A l'evento che il primo arciere centra il bersaglio e con O l'evento che viene realizzato un solo centro da parte dei tre arcieri, questi saranno dati, in temini di eventi elementari, da:

$$A = \{\text{arc1-arc2-arc3},\ \text{arc1-arc2-}\overline{\text{arc3}},\ \text{arc1-}\overline{\text{arc2}}\text{-arc3},\ \text{arc1-}\overline{\text{arc2}}\text{-}\overline{\text{arc3}}\},$$
$$O = \{\text{arc1-}\overline{\text{arc2}}\text{-}\overline{\text{arc3}},\ \overline{\text{arc1}}\text{-arc2-}\overline{\text{arc3}},\ \overline{\text{arc1}}\text{-}\overline{\text{arc2}}\text{-arc3}\}.$$

Quindi, la probabilità condizionata che la freccia che ha colpito il bersaglio sia stata lanciata dal primo arciere, dato che una sola delle tre frecce ha colpito il bersaglio, è data da:

$$\begin{aligned}
\mathsf{P}(A|O) &= \frac{\mathsf{P}(A\cap O)}{\mathsf{P}(O)} \\
&= \frac{\mathsf{P}(\text{arc1-}\overline{\text{arc2}}\text{-}\overline{\text{arc3}})}{\mathsf{P}(\text{arc1-}\overline{\text{arc2}}\text{-}\overline{\text{arc3}}) + \mathsf{P}(\overline{\text{arc1}}\text{-arc2-}\overline{\text{arc3}}) + \mathsf{P}(\overline{\text{arc1}}\text{-}\overline{\text{arc2}}\text{-arc3})} \\
&= \frac{\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{3}{5}}{\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{3}{5}+\frac{1}{4}\cdot\frac{1}{2}\cdot\frac{3}{5}+\frac{1}{4}\cdot\frac{1}{2}\cdot\frac{2}{5}} = \frac{9}{14} = 0.6429.
\end{aligned}$$

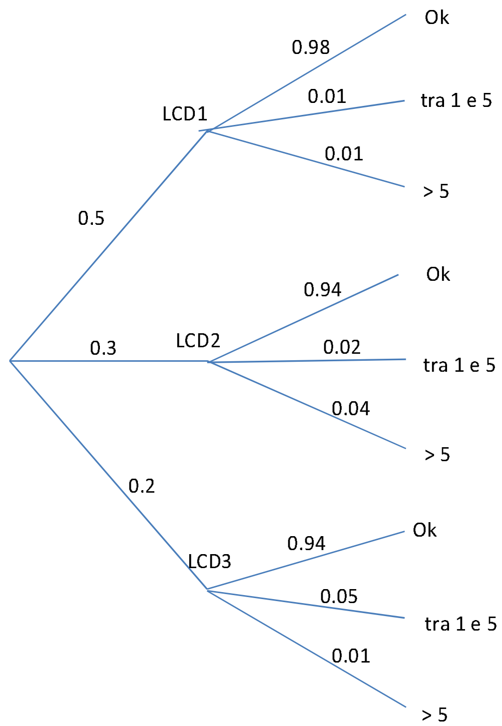
Soluzione Monte Carlo

```
t0 <- proc.time()
tiro <- c("CENTRO","ERRORE")
prove <- 100000; num <-0; den <- 0
```

```
for (i in 1:prove) {
  arciere1 <- sample(tiro, size=1, prob=c(3/4,1/4))
  arciere2 <- sample(tiro, size=1, prob=c(1/2,1/2))
  arciere3 <- sample(tiro, size=1, prob=c(2/5,3/5))
  out <- c(arciere1,arciere2,arciere3)
  if (sum(out=="CENTRO")==1) {
    den <- den + 1
    if (arciere1=="CENTRO") num <- num + 1
  }
}

cat(num,den,num/den,"\n")
t1 <- proc.time() - t0; t1
```

Esercizio 4 (telefoni cellulari)
Un dato costruttore di telefoni cellulari si rifornisce di schermi a cristalli liquidi da tre diversi fornitori LCD1, LCD2 e LCD3. Sul totale dei telefoni assemblati, gli schermi del fornitore LCD1 sono presenti nel 50% dei telefoni, quelli del fornitore LCD2 nel 30% dei telefoni e quelli del fornitore LCD3 nel rimanente 20% dei telefoni. Si sa che non tutti gli schermi utilizzati sono senza difetti. Le specifiche di qualità dei tre fornitori sono le seguenti. Il fornitore LCD1 garantisce che il 98% dei sui schermi sia senza difetti, che l'1% dei sui schermi abbia un numero di pixel difettosi compreso tra 1 e 5, e che l'1% dei sui schermi abbia un numero di pixel difettosi maggiore di 5. Le analoghe proporzioni per il fornitore LCD2 sono 94%, 2% e 4%, mentre per il fornitore LCD3 queste proporzioni sono 94%, 5% e 1%. Supponendo di scegliere a caso un telefono tra quelli costruiti da questo costruttore, determinare:
a) la probabilità che nel telefono prescelto lo schermo abbia più di 5 pixel difettosi;
b) la probabilità che nel telefono prescelto sia stato utilizzato uno schermo del fornitore LCD1, sapendo che lo schermo di questo telefono aveva un numero di pixel difettosi maggiore di 5.

Soluzione
Per prima cosa è utile organizzare le informazioni del problema nel seguente albero degli eventi



Definendo gli eventi LCD1 $=$ {schermo del fornitore LCD1}, LCD2 $=$ {schermo del fornitore LCD2}, LCD3 $=$ {schermo del fornitore LCD3}, e $A=$ {nessun pixel difettoso}, $B=$ {tra 1 e 5 pixel difettosi}, $C=$ {più di 5 pixel difettosi}, tutte le probabilità che ci servono possono essere organizzate nella seguente tabella:

	$A=$ {nessun pixel}	$B=$ {tra 1 e 5 pixel}	$C=$ {più di 5 pixel}	
LCD1	$\mathsf{P}(\mathrm{LCD1} \cap A)$	$\mathsf{P}(\mathrm{LCD1} \cap B)$	$\mathsf{P}(\mathrm{LCD1} \cap C)$	$\mathsf{P}(\mathrm{LCD1})$
LCD2	$\mathsf{P}(\mathrm{LCD2} \cap A)$	$\mathsf{P}(\mathrm{LCD2} \cap B)$	$\mathsf{P}(\mathrm{LCD2} \cap C)$	$\mathsf{P}(\mathrm{LCD2})$
LCD3	$\mathsf{P}(\mathrm{LCD3} \cap A)$	$\mathsf{P}(\mathrm{LCD3} \cap B)$	$\mathsf{P}(\mathrm{LCD3} \cap C)$	$\mathsf{P}(\mathrm{LCD3})$
	$\mathsf{P}(A)$	$\mathsf{P}(B)$	$\mathsf{P}(C)$	1

Utilizzando i dati forniti, queste probabilità risultano pari a:

	$A=$ {nessun pixel}	$B=$ {tra 1 e 5 pixel}	$C=$ {più di 5 pixel}	
LCD1	$0.5 \cdot 0.98 = 0.490$	$0.5 \cdot 0.01 = 0.005$	$0.5 \cdot 0.01 = 0.005$	0.5
LCD2	$0.3 \cdot 0.94 = 0.282$	$0.3 \cdot 0.02 = 0.006$	$0.3 \cdot 0.04 = 0.012$	0.3
LCD3	$0.2 \cdot 0.94 = 0.188$	$0.2 \cdot 0.05 = 0.010$	$0.2 \cdot 0.01 = 0.002$	0.2
	0.960	0.021	0.019	1

Risulta quindi agevole calcolare la probabilità che nel telefono prescelto ci siano più di 5 pixel difettosi. Per la formula delle probabilità totali (si veda la terza colonna della tabella), questa risulta pari a:

$$\begin{aligned}\mathsf{P}(C) &= \mathsf{P}(C|\mathrm{LCD1}) \cdot \mathsf{P}(\mathrm{LCD1}) + \mathsf{P}(C|\mathrm{LCD2}) \cdot \mathsf{P}(\mathrm{LCD2}) + \mathsf{P}(C|\mathrm{LCD3}) \cdot \mathsf{P}(\mathrm{LCD3}) \\ &= 0.01 \cdot 0.5 + 0.04 \cdot 0.3 + 0.01 \cdot 0.2 = 0.019.\end{aligned}$$

Invece, la probabilità condizionata che sul telefono prescelto ci sia lo schermo del fornitore LCD1, dato che questo ha più di 5 pixel difettosi, si può trovare utilizzando la formula di Bayes:

$$\begin{aligned}\mathsf{P}(\mathrm{LCD1}|C) &= \frac{\mathsf{P}(C|\mathrm{LCD1}) \cdot \mathsf{P}(\mathrm{LCD1})}{\mathsf{P}(C)} \\ &= \frac{\mathsf{P}(C|\mathrm{LCD1}) \cdot \mathsf{P}(\mathrm{LCD1})}{\mathsf{P}(C|\mathrm{LCD1}) \cdot \mathsf{P}(\mathrm{LCD1}) + \mathsf{P}(C|\mathrm{LCD2}) \cdot \mathsf{P}(\mathrm{LCD2}) + \mathsf{P}(C|\mathrm{LCD3}) \cdot \mathsf{P}(\mathrm{LCD3})} \\ &= \frac{0.01 \cdot 0.5}{0.01 \cdot 0.5 + 0.04 \cdot 0.3 + 0.01 \cdot 0.2} = \frac{0.005}{0.019} = 0.2632.\end{aligned}$$