

LABORATORIO 4 - Probabilità elementare

STATISTICA E LABORATORIO (CDL in INTERNET OF THINGS, BIG DATA, MACHINE LEARNING)

Anno Accademico 2023-2024

Dado

Si suppone di lanciare un dado equilibrato.

```
Omega<-c(1,2,3,4,5,6)
set.seed(11)
sample(Omega,1) # lancio di un dado regolare
```

```
## [1] 2
```

```
set.seed(11)
sample(Omega,3, replace=TRUE) # tre lanci di un dado regolare;
```

```
## [1] 2 2 1
```

```
# replace=TRUE per avere estrazioni con reinserimento
```

```
# lancio del dado regolare fino ad ottenere 6 per la prima volta  
set.seed(11)  
ris<-sample(Omega,1)  
ris
```

```
## [1] 2
```

```
while (ris!=6) {ris<-sample(Omega,1)  
  print(ris)  
}
```

```
## [1] 2  
## [1] 1  
## [1] 1  
## [1] 5  
## [1] 4  
## [1] 6
```

Lotteria

Una lotteria è costituita da 1000 biglietti, di cui 5 vincenti. Si scelgono a caso 10 biglietti. Si vuole determinare la probabilità di $A = \text{"un biglietto è vincente"}$.

```
# probabilita' di un biglietto vincente  
choose(5,1)*choose(995,9)/choose(1000,10)
```

```
## [1] 0.04821976
```

```
# probabilita' di almeno un biglietto vincente  
1-choose(5,0)*choose(995,10)/choose(1000,10)
```

```
## [1] 0.0491063
```

```
# probabilita' di due biglietti vincenti  
choose(5,2)*choose(995,8)/choose(1000,10)
```

```
## [1] 0.0008793878
```

Poker

Un giocatore di poker ha in mano 5 carte; si determini la probabilità che abbia 2 assi e 3 jack.

```
#il numero di modi in cui si possono distribuire 2 assi da 4 carte  
choose(4,2)
```

```
## [1] 6
```

```
#il numero di modi in cui si possono distribuire 3 jack da 4 carte  
choose(4,3)
```

```
## [1] 4
```

```
#il numero totale di mani da 5 carte è  
choose(52,5)
```

```
## [1] 2598960
```

```
#la probabilità di ottenere 2 assi e 3 jack è  
choose(4,2)*choose(4,3)/choose(52,5)
```

```
## [1] 9.234463e-06
```

Indirizzi

Una rete aziendale è costituita da un server e da dieci PC. Quando un PC accede alla rete, riceve un indirizzo IP scelto in modo casuale tra 200 disponibili.

Nell'ipotesi che tutti e dieci i PC accedano insieme alla rete, quale è la probabilità che il server abbia assegnato almeno due indirizzi IP identici?

```
nind <- 200 # numero indirizzi disponibili
npc <- 10 # numero pc
# probabilita che vengano assegnati almeno due indirizzi identici
1-choose(nind,npc)*prod(1:npc)/(nind^npc)
```

```
## [1] 0.2043925
```

Due palline

Si consideri l'estrazione di due palline da un'urna contenente dieci palline nere e cinque bianche.

```
# urna con 10 palline nere e 5 bianche  
urna<-c(rep("N",10),rep("B",5))  
set.seed(11)  
# estrazione di due palline senza reinserimento  
sample(urna,2)
```

```
## [1] "N" "N"
```

```
# estrazione di due palline con reinserimento  
sample(urna,2,replace = TRUE)
```

```
## [1] "N" "N"
```



```
# estrazione di cinque palline senza reinserimento  
sample(urna,5)
```

```
## [1] "N" "N" "B" "N" "B"
```

```
# estrazione di cinque palline con reinserimento  
sample(urna,5,replace = TRUE)
```

```
## [1] "N" "B" "B" "N" "B"
```

Banca

Una filiale di un istituto bancario ha 1210 clienti titolari di conto corrente. L'ufficio crediti distingue tra buoni e cattivi clienti, tenendo conto delle eventuali insolvenze. Inoltre, sono noti i dati sull'eventuale possesso della carta di credito. Le informazioni disponibili vengono sintetizzate nella seguente tabella

```
banca <- rbind(cbind(rep("cattivo",60),rep("si",60)),
               cbind(rep("buono",520),rep("si",520)),
               cbind(rep("cattivo",21),rep("no",21)),
               cbind(rep("buono",609),rep("no",609)))
banca <- as.data.frame(banca) # matrice trasformata in data frame
colnames(banca) <- c("tipo","cc") # nomi delle colonne
banca$tipo<-ordered(banca$tipo, levels=c("cattivo","buono"))
banca$cc<-factor(banca$cc, levels=c("si","no"))
str(banca)
```

```
## 'data.frame':    1210 obs. of  2 variables:
## $ tipo: Ord.factor w/ 2 levels "cattivo"<"buono": 1 1 1 1 1 1 1
## $ cc : Factor w/ 2 levels "si","no": 1 1 1 1 1 1 1 1 1 1 ...
```

```
tab<-table(banca$cc,banca$tipo) # tabella di contingenza  
#(distribuzione di frequenza assoluta congiunta)  
tab
```

```
##  
##      cattivo buono  
##  si      60    520  
##  no      21    609
```

Si sceglie casualmente un cliente e si vuole valutare l'eventuale indipendenza tra gli eventi $A = \text{"si sceglie un buon cliente"}$ e $B = \text{"si sceglie un possessore di carta di credito"}$.

```
# distribuzione di probabilita' congiunta di "cc"  
# e "tipo"; distribuzione di frequenza relativa congiunta  
# in alternativa, prop.table(tab)
```

```
tab/sum(tab)
```

```
##  
##          cattivo          buono  
##   si 0.04958678 0.42975207  
##   no 0.01735537 0.50330579
```

```
# distribuzione di probabilita' marginale di "cc"  
# (frequenza relativa marginale)  
margin.table(tab,1)/sum(margin.table(tab,1))
```

```
##  
##          si          no  
## 0.4793388 0.5206612
```

```
# distribuzione marginale di "tipo"  
# (frequenza relativa marginale)  
margin.table(tab,2)/sum(margin.table(tab,2))
```

```
##  
## cattivo buono  
## 0.06694215 0.93305785
```

```
# distribuzione condizionata di "tipo/cc=si"  
# (frequenza relativa condizionata)  
tab[1,]/sum(tab[1,])
```

```
## cattivo buono  
## 0.1034483 0.8965517
```

```
# distribuzione condizionata di "tipo/cc=no"  
# (frequenza relativa condizionata)  
# in alternativa, prop.table(tab,1)  
tab[2,]/sum(tab[2,])
```

```
##      cattivo      buono  
## 0.03333333 0.96666667
```

```
# distribuzione condizionata di "cc/tipo=buono"  
# (frequenza relativa condizionata)  
tab[,1]/sum(tab[,1])
```

```
##      si      no  
## 0.7407407 0.2592593
```

```
# distribuzione condizionata di "cc/tipo=cattivo"  
# (frequenza relativa condizionata)  
# in alternativa, prop.table(tab,2)  
tab[,2]/sum(tab[,2])
```

```
##           si           no  
## 0.4605846 0.5394154
```

```
# distribuzione marginale di "cc"  
ccmarg<-margin.table(tab,1)/sum(margin.table(tab,1))  
# distribuzione marginale di "tipo"  
tipomarg<-margin.table(tab,2)/sum(margin.table(tab,2))  
# vettore interpretato come matrice  
ccmarg <- as.matrix(ccmarg)  
# vettore interpretato come matrice  
tipomarg <- as.matrix(tipomarg)
```

```
# probabilita' congiunta di "cc" e "tipo" in caso di indipendenza
prob_ind <- ccmarg%*%t(tipomarg)
prob_ind
```

```
##          cattivo      buono
## si 0.03208797 0.4472509
## no 0.03485418 0.4858070
```

```
# probabilita' congiunta effettiva di "cc" e "tipo"
tab/sum(tab)
```

```
##
##          cattivo      buono
## si 0.04958678 0.42975207
## no 0.01735537 0.50330579
```


Un'azienda produce il 30% dei suoi articoli con una prima linea di produzione, che fornisce 8 pezzi difettosi su 100, mentre il restante 70% con una seconda linea, che fornisce 5 pezzi difettosi su 100. Si sceglie a caso un articolo, senza sapere da quale linea provenga, e viene scartato perché difettoso. Si vuole calcolare la probabilità che provenga dalla prima linea di produzione. Indicato con $B =$ "l'articolo selezionato è difettoso" e con $A_i =$ "l'articolo selezionato proviene dall' i -esima linea", $i = 1, 2$, si ha che $P(B|A_1) = 8/100$, $P(B|A_2) = 5/100$, $P(A_1) = 3/10$, $P(A_2) = 7/10$. La probabilità che l'articolo difettoso provenga dalla prima linea di produzione è

```
pa1=3/10
pa2=7/10
pbdata1=8/100
pbdata2=5/100
pb=pa1*pbdata1+pa2*pbdata2
pa1datob=pa1*pbdata1/pb
pa1datob
```

```
## [1] 0.4067797
```

Una popolazione presenta il 32% di fumatori. È noto che il 25% dei fumatori e il 5% dei non fumatori è affetto da una patologia respiratoria cronica. Si sceglie a caso un individuo dalla popolazione. Quale è la probabilità che sia affetto dalla patologia? Se l'individuo risulta ammalato, quale è la probabilità che sia fumatore? Indicato con $B = \text{"l'individuo scelto è ammalato"}$ e con $A = \text{"l'individuo scelto è fumatore"}$, si ha che $P(B|A) = 0.25$, $P(B|A^c) = 0.05$, $P(A) = 0.32$, $P(A^c) = 0.68$.

```
##Probabilità di B P(B)
```

```
pa=0.32
```

```
pac=1-pa
```

```
pbdatoa=0.25
```

```
pbdatoac=0.05
```

```
pb=pa*pbdatoa+pac*pbdatoac
```

```
pb
```

```
## [1] 0.114
```

Probabilità di A condizionata dall'evento B $P(A|B)$

```
padatob=pa*pbdatoa/pb  
padatob
```

```
## [1] 0.7017544
```