Basi della probabilità

La probabilità si occupa di *fenomeni* aleatori, cioè di un *esperimento* i cui possibili risultati appartengono ad un insieme ben definito e dove l'esito non è prevedibile.
Sia S uno spazio campionario. Una probabilità valida soddisfa i seguenti **assiomi di probabilità**:

1. Le probabilità sono numeri reali non negativi, cioè per tutti gli eventi $E, P(E) \geq 0$.
2. La probabilità dello spazio campione è $1, P(S) = 1$.
3. Le probabilità sono numerabilmente additive: se $UN\_1, UN\_2, \ldots$ sono disgiunti a due a due, allora

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

   Teorema 2.1
   Siano A e B eventi dello spazio campionato S.
4. $P(\emptyset) = 0$
5. Se A e B sono disgiunti ($\cup$), allora $P(A \cup B) = P(A) + P(B)$
6. Se $A \subset B$, allora $P(A) \leq P(B)$
7. $0 \leq P(A) \leq 1$
8. $P(A) = 1 - P(\overline{A})$
9. $P(A - B) = P(A) - P(A \cap B)$
10. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
   Esempio:
   Laniamo un dado, che probabilità ho che esca una determinata faccia?
   Usiamo il punto 3.

$$P(\{1\}) + P(\{2\}) + P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\}) = P(\{1, 2, 3, 4, 5, 6\})$$

   Tutte le probabilità sono uguali e la loro somma è $= 1$, la probabilità che esca una determinata faccia è $\frac{1}{6}$.
   Esempio 2.8:
   Supponiamo che i dadi invece siano due, lo spazio campionato S è dato da:

$$S = \begin{pmatrix} (1,1),(1,2),(1,3),(1,4),(1,5),(1,6) \\ (2,1),(2,2),(2,3),(2,4),(2,5),(2,6) \\ (3,1),(3,2),(3,3),(3,4),(3,5),(3,6) \\ (4,1),(4,2),(4,3),(4,4),(4,5),(4,6) \\ (5,1),(5,2),(5,3),(5,4),(5,5),(5,6) \\ (6,1),(6,2),(6,3),(6,4),(6,5),(6,6) \end{pmatrix}$$

   Gli eventi dove "la somma dei due dadi è 6" è rappresentata da:

$$E = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$$

   La probabilità che la somma dei due dadi sia 6 è data da:

$$P(E) = \frac{|E|}{|S|} = \frac{5}{36}$$

Sia F l'evento "almeno uno dei due dadi è 2", l'evento è rappresentato da:

$$F = \{(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),(1,2),(3,2),(4,2),(5,2),(6,2)\}$$

La probabilità di F è: $P(F) = \frac{11}{36}$
$E \cap F = (2,4),(4,2)$ e $P(E \cap F) = \frac{2}{36}$
$P(E \cap F) = P(E) + P(F) - P(E \cap F) = \frac{5}{36} + \frac{11}{36} - \frac{2}{36} = \frac{14}{36}$
$P(\overline{E}) = 1 - P(E) = \frac{31}{36}$
Simulazioni con `sample`

```
sample(x, size, replace = FALSE, prob = NULL)
```

`x` il vettore di elementi dal quale si sta campionando.
`size` il numero di campioni che si voglio ottenere.
`replace` se si stanno usando rimpiazzi o meno.
`prob` un vettore di probabilità o pesi, associato a `x`.
Per ottenere due numeri casuali tra 1 e 10:

```
sample(x = 1:10, size=2)
```

```
## [1] 2 5
```

`sample` non ritorna mai lo stesso valore 2 o più volte, bisogna usare `replace = TRUE`
Esempio 2.9
4 tipi di sangue con diversa probabilità di distribuzione:

```
bloodtypes <- ("O", "A", "B", "AB")
bloodprobs <- (0.45, 0.40, 0.11, 0.04)
sample(x = bloodtypes, suze = 30, prob = bloodprobs, replace = TRUE)
```

```
##  [1] "A"  "O"  "AB" "A"  "O"  "A"  "A"  "O"  "O"  "O"  "O"  "A"  "O"
## [14] "A"  "A"  "A"  "B"  "A"  "A"  "B"  "A"  "O"  "O"  "A"  "A"  "O"
## [27] "O"  "A"  "A"  "O"
```

```
sim_data <- sample(
    x = bloodtypes, size = 10000,
    prob = bloodprobs, replace = TRUE
)
table(sim_data)
```

```
## sim_data
##    A   AB    B    O
```

```
## 3998  425 1076 4501
```

```
table(sim_data) / 1000
```

```
## sim_data
##      A     AB      B      O
## 0.3998 0.0425 0.1076 0.4501
```

Esempio 2.10

Supponiamo che due dadi a 6 facce vengano lanciati, sommiamo il risultato.
Effettuiamo questo esperimento su 10000 lanci:

```
die1 <- (x = 1:6, size = 10000, replace = TRUE)
die2 <- (x = 1:6, size = 10000, replace = TRUE)
sumDice <- die1 + die2
```

Vediamo i dati

```
read(die1)
```

```
## [1] 1 4 1 2 5 3
```

```
read(die2)
```

```
## [1] 1 6 1 4 1 3
```

Sia E l'evento "la somma dei dadi è 6", e F "almeno uno dei dadi è 2". Definiamo questi eventi dai nostri dati simulati:

```
eventE <- sumDice == 6
read(eventE)
```

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

```
eventF <- die1 == 2 | die2 == 2
read(eventF)
```

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

Usando `mean` troviamo la percentuale con cui si sono verificati gli eventi:

```
mean(eventE) # P(E)
```

```
## [1] 0.1409
```

```
mean(eventF)
```

```
## [1] 0.2998
```

Per stimare $P(E \cap F) = \frac{2}{36} \approx 0,056$:

```
mean(eventE & eventF)
```

```
## [1] 0.0587
```

Non è necessario memorizzare tutti i vettori TRUE/FALSE nelle variabili evento.
Ecco una stima di $P(E \cup F) = \frac{14}{36} \approx 0,389$:

```
mean((sumDice == 6) | (die1 == 6) | (die2 == 6))
```

```
## [1] 0.382
```

Utilizzo di `replicate` per ripetere gli esperimenti

Per simulazioni complesse, seguiamo un flusso:

1. Scrivo il codice per eseguire l'esperimento
2. Ripeto l'esperimento un piccolo numero di volte e controllo i risultati:

- `replicate(100, { ESPERIMENTO })`

3. Ripeto l'esperimento un grande numero di volte e memorizzo il risultato:

- `event <- replicate(10000, { ESPERIMENTO })`

4. Calcolo la probabilità usando `mean`

Probabilità condizionata

Dato uno spazio di probabilità $(S, \mathcal{F}, \mathbb{P})$ e due eventi $E, H \in \mathcal{F}$ con $\mathbb{P}(H) > 0$, si dice *probabilità condizionata* di E dato H la quantità

$$\mathbb{P}(E|H) = \frac{P(E \cap H)}{H}$$

che esprime il grado di fiducia dell'osservatore nel verificarsi di E *supponendo che si verifichi H*.
Supponiamo di lanciare due dadi e uno di essi cade dal tavolo dove non puoi vederlo, mentre l'altro mostra un 4. Vorremmo aggiornare le probabilità associate alla somma dei due dadi in base a queste informazioni. La nuova probabilità che la somma dei dadi sia 2 sarebbe 0, la nuova probabilità che la somma dei dadi sia 5 sarebbe 1/6 perché questa è solo la probabilità che il dado che non possiamo vedere sia un "1" e la nuova probabilità che la somma dei dadi sia 7 sarebbe anche 1/6.
Formalmente abbiamo la seguente definizione:
Sia A e B eventi in uno spazio campionario S, con $P(B) \neq 0$, la *probabilità condizionale* che A dato B sia:

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

Leggiamo $P(A|B)$ come "la probabilità di A dato B".

Valutazioni classiche

Esempio 2.19:
Lanciamo 2 dadi, con quale probabilità entrambe i dadi sono danno 4, sapendo che la loro somma è 8?
$A = \{(4,4)\}$
$B = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}$

$$P(A|B) = P(A \cap B)/P(B) = \frac{1/36}{5/36} = 1/5$$

Invece quale è la probabilità che la somma dei dadi sia 8 sapendo che entrambe i dadi danno 4?
$A = \{(4,4)\}$
$B = \{(4,4)\}$

$$P(A|B) = P(B \cap A)/P(A) = \frac{1/36}{1/36} = 1$$

Quindi:

1. $P((A \cap B)|B) = P(A|B)$
2. $P(A \cup B|B) = 1$

Valutazioni uniformi

Esempio: Rotazione di uno spinner
Facciamo ruotare velocemente uno spinner simmetrico imperniato su un goniometro e ne osserviamo l'angolo di arresto in $]-\frac{\frac{\pi}{2},\pi}{2}]$. Con quale probabilità l'angolo di arresto dello spinner sarà compreso tra $-\frac{\pi}{4} e \frac{\pi}{3}$ (estremi inclusi) supponendo che sia positivo?
Notiamo preliminarmente che le valutazioni uniformi sono diffuse e quindi è *inessenziale* l'inclusione o meno degli estremi nell'intervallo di cui si calcola la probabilità e in quello al quale si condiziona.
Prendiamo $]a,b] =]-\frac{\pi}{2},\frac{\pi}{2}]$ e con $\mathbb{P}$ *uniforme* troviamo:

$$\mathbb{P}\left(\left[-\frac{\pi}{4},\frac{\pi}{3}\right]\middle|\left[0,\frac{\pi}{2}\right]\right)=\frac{\frac{\pi}{3}-0}{\frac{\pi}{2}-0}=\frac{2}{3}$$

visto che $\left[-\frac{\pi}{4},\frac{\pi}{3}\right]\cap\left[0,\frac{\pi}{2}\right]=[0,\frac{\pi}{3}]$.
In questo caso $\mathbb{P}\left(\left[-\frac{\pi}{4},\frac{\pi}{3}\right]\middle|\left[0,\frac{\pi}{2}\right]\right)>\mathbb{P}\left(\left[-\frac{\pi}{4},\frac{\pi}{3}\right]\right)$ perché:

$$\mathbb{P}\left(\left[-\frac{\pi}{4},\frac{\pi}{3}\right]\right)=\frac{\frac{\pi}{3}-\left(-\frac{\pi}{4}\right)}{\frac{\pi}{2}-\left(-\frac{\pi}{2}\right)}=\frac{\frac{7\pi}{12}}{\pi}=\frac{7}{12}<\frac{8}{12}=\frac{2}{3}$$

Valutazioni frequentiste

se le probabilità degli eventi sono frequenze relative di realizzazione in precedenti ripetizioni del fenomeno:

$$\mathbb{P}(E|H)=\frac{k_{E\cap H}}{k_H},\quad E\in\mathcal{F}$$

Essempio:
Lanciamo un dado a sei facce caricato per ottenere 6 con cui, in una sequenza di lanci precedenti, abbiamo ottenuto settantanove volte 6, cinque volte 5, tre volte 4, sette volte 3, cinque volte 2 e una volta 1. Quanto valuteremo la probabilità che un punteggio pari sia primo?
Prendiamo $S=\{1,2,3,4,5,6\},\mathcal{F}=\wp(S)$ e $\mathbb{P}$ specificata da:

$$\begin{array}{lll}\mathbb{P}\{1\}=0.01 & \mathbb{P}\{2\}=0.05 & \mathbb{P}\{3\}=0.07\\ \mathbb{P}\{4\}=0.03 & \mathbb{P}\{5\}=0.05 & \mathbb{P}\{6\}=0.79\end{array}$$

senza considerazioni di simmetria. Posto $A=\text{"primo"}=\{2,3,5\}$ e $B=\text{"pari"}=\{2,4,6\}$, troviamo per la probabilità richiesta:

$$\mathbb{P}(A|B)=\frac{k_{A\cap B}}{k_B}=\frac{k_2}{k_{2,4,6}}=\frac{5}{5+3+79}=\frac{5}{87}\backsimeq 0.057=5.7\%$$

Eventi indipendenti

In uno spazio di probabilità $(S,\mathcal{F},\mathbb{P})$, due eventi $E,F\in\mathcal{F}$ si dicono *stocasticamente indipendenti* sotto $\mathbb{P}$ o semplicemente *indipendenti* quando vale la fattorizzazione:

$$\mathbb{P}(E\cap F)=\mathbb{P}(E)\mathbb{P}(F)$$

Può accadere in due modi:

- **banalmente** per $\mathbb{P}(E)=0$ o $\mathbb{P}(F)=0$
  - perché allora $\mathbb{P}(E\cap F)\leq min\{\mathbb{P}(E),\mathbb{P}(F)\}$ per monotonia e quindi necessariamente $\mathbb{P}(E\cap F)=0$
- **significativamente** con:

  $$\begin{array}{rcl}\mathbb{P}(E|F)&=&\mathbb{P}(E)\\ \mathbb{P}(\mathbb{F}|\mathbb{E})&=&\mathbb{P}(F)\end{array}$$

  se $min\{\mathbb{P}(E),\mathbb{P}(F)\}>0$.
  Scriveremo

  $$E\perp\!\!\!\perp F$$

per indicare che E ed F sono *indipendenti*, più precisamente

$$E \perp\!\!\!\perp_{\mathbb{F}} F$$

per ricordare il ruolo di $\mathbb{P}$.
Altrimenti scriveremo $E \not\perp\!\!\!\perp F$ se E ed F sono *dipendenti*.
In particolare:

- E ed F sono *favorevolmente dipendenti* (sotto $\mathbb{P}$) quando

$$\mathbb{P}(E \cap F) > \mathbb{P}(E)\mathbb{P}(F)$$

cioè $\mathbb{P}(E|F) > \mathbb{P}(E)$ e $\mathbb{P}(F|E) > \mathbb{P}(F)$ nel caso significativo;

- E ed F sono *sfavorevolmente dipendenti* (sotto $\mathbb{P}$) quando

$$\mathbb{P}(E \cap F) < \mathbb{P}(E)\mathbb{P}(F)$$

cioè $\mathbb{P}(E|F) < \mathbb{P}(E)$ e $\mathbb{P}(F|E) < \mathbb{P}(F)$ nel caso significativo;

Esempio:
Lanciamo un ordinario dado a sei facce e ne osserviamo il punteggio. Gli eventi $A = \text{"primi"}$ e $B = \text{"pari"}$ sono indipendenti?
Presa $\mathbb{P}$ classica su tutte le parti di $S = \{1, 2, 3, 4, 5, 6\}$:

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}\{2, 3, 5\} = \frac{3}{6} = \frac{1}{2} \\
\mathbb{P}(B) &= \mathbb{P}\{2, 4, 6\} = \frac{3}{6} = \frac{1}{2} \\
\mathbb{P}(A \cap B) &= \mathbb{P}\{2\} = \frac{1}{6} < \frac{1}{4} = \frac{1}{2} \times \frac{1}{2}
\end{aligned}$$

quindi $A \not\perp\!\!\!\perp B$, in particolare A e B sono *sfavorevolmente dipendenti*.
Tuttavia A e B sono *logicamente indipendenti*, dal momento che che i loro quattro costituenti sono tutti non vuoti:

$$\overline{A} \cap \overline{B} = \{1\}, \overline{A} \cap B = \{4, 6\}, A \cap \overline{B} = \{3, 5\}, A \cap B = \{2\}$$

Invarianza per negazione dell'indipendenza tra due eventi

$$E \perp\!\!\!\perp F \Rightarrow E \perp\!\!\!\perp \overline{F}$$

Dimostrazione:

$$\begin{aligned}
\mathbb{P}(E \cap \overline{F}) &= \mathbb{P}(E) - \mathbb{P}(E \cap F) \\
&= \mathbb{P}(E) - \mathbb{P}(E) - \mathbb{P}(F) \\
&= \mathbb{P}(E)\{1 - \mathbb{P}(F)\} \\
&= \mathbb{P}(E)\mathbb{P}(\overline{F})
\end{aligned}$$

Vale anche il viceversa, perché $\overline{\overline{F}} = F$, quindi:

$$E \perp\!\!\!\perp F \iff E \perp\!\!\!\perp \overline{F} \iff \overline{E} \perp\!\!\!\perp \overline{F} \iff \overline{E} \perp\!\!\!\perp F$$

$\perp\!\!\!\perp$ è una *relazione simmetrica*, quindi affermare che E ed F sono indipendenti sotto $\mathbb{P}$ corrisponde ad affermare che $\mathbb{P}$ si fattorizza su tutti i costituenti di E ed F.

Si noti la differenza: nel caso del dado equilibrato abbiamo scoperto che “pari” e “centrale” sono eventi indipendenti; nel caso della coppia abbiamo *imposto* che F_1 e F_2 siano indipendenti (ed equiprobabili).
Una differenza analoga a quella che passa tra *calcolare* $\mathbb{P}(E|H)$ a partire da $\mathbb{P}(E \cap H)$e *assegnare* $\mathbb{P}(E|H)$ per specificare $\mathbb{P}(E \cap H)$, supponendo che $\mathbb{P}(H) > 0$.
Esempio:
Una coppia ha due figli. Sappiamo che almeno una è femmina. Con quale probabilità sono due femmine? Troviamo subito

$$\mathbb{P}(F_1 \cup F_2) = \mathbb{P}(F_1) + (\overline{F_1} \cap F_2) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$$
$$\mathbb{P}(F_1 \cap F_1|F_1 \cup F_2) = \frac{\mathbb{P}(F_1 \cap F_2)}{\mathbb{P}(F_1 \cup F_2)} = \frac{1/4}{3/4} = \frac{1}{3} \dots$$

dunque $\mathbb{P}(F_1 \cap F_2|F_1 \cup F_2) = \frac{1}{3} \neq \frac{1}{2} = \mathbb{P}(F_1|F_2) = \mathbb{P}(F_2|F_1)!$
Per comprendere questo paradossoa ragioniamo su come sappiamo che almeno una figlia è femmina e ipotizziamo di averla incontrata, introducendo la partizione

$$H = \text{"incontro figlia femmina"}$$
$$\overline{H} = \text{"incontro figlio maschio"}$$

nel diagramma di Venn (in modo tale che $F_1 \cap F_2 \subset H \subset F_1 \cup F_2$).
Supponiamo:

$$\mathbb{P}(H|\overline{F_1} \cap F_2) = p = \mathbb{P}(H \cap F_1 \cap F_2) \quad \text{con} \quad 0 < p < 1$$

Troviamo

$$\begin{aligned}\mathbb{P}(F_1 \cap F_2|H) &= \frac{\mathbb{P}(F_1 \cap F_2 \cap H)}{\mathbb{P}(H)} \\ &= \frac{\mathbb{P}(F_1 \cap F_2)}{\mathbb{P}(H|\overline{F_1} \cap F_2)\mathbb{P}(\overline{F_1} \cap F_2) + \mathbb{P}(F_1 \cap F_2) + \mathbb{P}(H|F_1 \cap \overline{F_2})\mathbb{P}(F_1 \cap \overline{F_2})} \\ &= \frac{1/4}{p\frac{1}{4} + \frac{1}{4} + p\frac{1}{4}} = \frac{1}{1+2p} = \begin{cases} 1 & \text{se } p \to 0 \\ \frac{1}{2} & \text{se } p \to \frac{1}{2} \\ \frac{1}{3} & \text{se } p \to 1 \end{cases}\end{aligned}$$

Specificare $\mathbb{P}$ sui costituenti E ed F supponendoli indipendenti è sempre possibile, ma non sempre (pienamente) appropriato.
Fare diversamente richiede di valutare tipo e forza della dipendenza:

$$\mathbb{P}(E \cap F) = d\mathbb{P}(E)\mathbb{P}(F)$$

con dipendenza sfavorevole per $d < 1$ e favorevole per $d > 1$; il *fattore di dipendenza* d va scelto garantendo la coerenza di $\mathbb{P}$, cioé:

$$max\left\{0, \frac{1}{\mathbb{P}(E)} + \frac{1}{\mathbb{P}(F)} - \frac{1}{\mathbb{P}(E)\mathbb{P}(F)}\right\} < d < min\left\{\frac{1}{\mathbb{P}(E)}, \frac{1}{\mathbb{P}(F)}\right\}.$$

Diremo che E ed F sono *stocasticamente indipendenti* in modo sostanziale quando:

$$E \perp\!\!\!\perp F \text{ con } 0 < \mathbb{P}(E), \mathbb{P}(F) < 1$$

Esempio: interruttori elettrici in serie
Un circuito elettrico è formato da due interruttori *uno di seguito all'altro*: fa passare corrente quando *entrambi* gli interruttori sono chiusi. Supponiamo che (in un dato istante) ciascun interruttore sia chiuso con probabilità $p = 0.8$ indipendentemente dall'altro interruttore. Con quale probabilità (in tale istante) passerà corrente nel circuito?
Posto $C_i =$ "i-esimo interruttore chiuso", $i = 1, 2$, in un diagramma di Venn, assegnamo $\mathbb{P}$ su $\mathcal{F} = \omega(C_1, C_2)$ con $C_1 \perp\!\!\!\perp C_2$ e $\mathbb{P}(C_1) = \mathbb{P}(C_2) = p$. Per l'evento di interesse $C_1 \cap C_2$ troviamo:

$$\mathbb{P}(C_1 \cap C_2) = \mathbb{P}(C_1)\mathbb{P}(C_2) = p^2$$

e quindi $\mathbb{P}(C_1 \cap C_2) = (0.8)^2 = 0.64 < 0.8$.
Esempio: interruttori elettrici in parallelo
Un circuito elettrico è formato da due interruttori *uno di fianco all'altro*: fa passare corrente quando *almeno uno* degli interruttori è chiuso. Supponiamo che (in un dato istante) ciascun interruttore sia chiuso con probabilità $p = 0.8$ indipendentemente dall'altro interruttore. Con quale probabilità (in tale istante) passerà corrente nel circuito?
Probabilità che l'interruttore sia **aperto**: $1 - p = 1 - 0.8 = 0.2$
$\mathbb{P}(\text{nessun interruttore chiuso}) = (0.2)^2 = 0.04$
La probabilità che **almeno uno degli interruttori sia chiuso** (e quindi che passi corrente) è il complementare di questa probabilità:

$$\mathbb{P}(\text{passa corrente}) = 1 - \mathbb{P}(\text{nessun interruttore chiuso}) = 1 - 0.04 = 0.96$$

Simulare la probabilità condizionata

Esempio 2.27

Lanciamo 2 dadi. Si stima la probabilità condizionata che la somma dei dadi sia almeno 10, dato che almeno uno dei dadi è un 6.
Per prima cosa, stimiamo la probabilità che la somma dei dadi sia almeno 10 *e* che almeno uno dei dadi sia 6.

```
eventAB <- replicate(10000, {
  dieRoll <- sample(1:6, 2, replace = TRUE)
  (sum(dieRoll) >= 10) && (6 %in% dieRoll)
})
probAB <- mean(eventAB)
```

Successivamente, stimiamo la probabilità che almeno uno dei dadi sia un 6.

```
eventB <- replicate(10000, {
  die_roll <- sample(1:6, 2, replace = TRUE)
  6 %in% die_roll
})
probB <- mean(eventB)
```

Infine, prendiamo il quoziente.

```
probAB / probB
```

```
## [1] 0.4560601
```

La risposta è:

$$\mathbb{P}(A \cap B)/\mathbb{P}(B) = \frac{5/36}{11/36} = \frac{5}{11} \approx 0.4545$$

Formula di Bayes

Siano H e E due eventi con probabilità non nulle. La probabilità condizionata di H rispetto a E è uguale al prodotto tra la probabilità condizionata di E rispetto a H e la probabilità di H, tutto fratto la probabilità di E:

Formula di Bayes
$H = $"ipotesi"
$E = $"evidenza"

$$\mathbb{P}(H|E) = \frac{\mathbb{P}(E|H) \cdot \mathbb{P}(H)}{\mathbb{P}(E)} \text{ con } \mathbb{P}(H), \mathbb{P}(E) \neq 0$$

Esempio: 20178 US midterm elections
Sappiamo chea nelle elezioni intermedie statunitensi del 2018 il *44%* dei votanti era Republican, il *46%* dei votanti era Gun Owner e il *61%* dei Gun Owner era Republican.
Quale percentuale dei Republican era Gun Owner?
Prendiamo un diagramma di Venn con $R = $"Republican" e $G = $"Gun Ownder", calcoliamo:

$$\mathbb{P}(G|R) = \frac{\mathbb{P}(R|G)\mathbb{P}(G)}{\mathbb{P}(R)} = \frac{0.61 \times 0.46}{0.44} \backsimeq 0.64 = 64\%$$

Esempio: campagna di sceening
Nell'ambito di una campagna di screening che mira a identificare precocemente una malattia rara, della quale sappiamo che colpisce il *2%* della popolazione, un nostro amico è risultato positivo al test.
Il test funziona correttamente sul *99.9%* dei soggetti malati e sul *97.5%* dei soggetti sani.
Rattristati, ma determinati a valutare razionalmente, ci chiediamo quale sia la probabilità che il nostro amico sia malato.
Prendiamo un diagramma di Venn con $M = $"soggetto malato" e $T = $"test positivo", riassumiamo i dati:

$$\begin{aligned} \mathbb{P}(M) &= 0.02 && \text{probabilità di matattia} \\ \mathbb{P}(T|M) &= 0.999 && \text{verosimiglianza dell'ipotesi di malattia} \\ \mathbb{P}(\overline{T}|\overline{M}) &= 0.975 && \text{verosimiglianza dell'ipotesi di salute} \end{aligned}$$

troviamo la *probabilità iniziale* di salute con l'inverso di M:

$$\mathbb{P}(\overline{M}) = 1 - \mathbb{P}(M) = 1 - 0.02 = 0.98 = 98\%$$

Siamo interessati alla *probabilità finale* di malattia $\mathbb{P}(M|T)$ o alla *probabilità finale* di salute $\mathbb{P}(\overline{M}|T) = 1 - \mathbb{P}(M|T)$:

$$\begin{aligned}\mathbb{P}(M|T) &= \frac{\mathbb{P}(T|M)\mathbb{P}(\mathbb{M})}{\mathbb{P}(T|M)\mathbb{P}(M)\mathbb{P}(T|\overline{M})\mathbb{P}(\overline{M})} \\ &= \frac{0.999 \times 0.02}{0.999 \times 0.02 \times 0.025 \times 0.98} \\ &= \frac{1998}{4448} \backsimeq 0.45 = 45\%\end{aligned}$$

di modo che è ancora più probabile che il nostro amico sia sano! Come si concilia questo risultato con le ottime prestazioni del test?
La rarità della malattia attenua la verosimiglianza $\mathbb{P}(T|M)$ e amplifica la verosimiglianza $\mathbb{P}(T|\overline{M})$.
Il test fornisce comunque un segnale forte: la probabilità di malattia *passa dal 2% al 45%* e ciò suggerisce un approfondimento clinico.

Con due spiegazioni

La formula di Bayes per due spiegazioni si può scrivere come:

$$\frac{\mathbb{P}(H|E)}{\mathbb{P}(\overline{H}|E)} = \frac{\mathbb{P}(E|H)}{\mathbb{P}(E|\overline{H})} \times \frac{\mathbb{P}(H)}{\mathbb{P}(\overline{H})}$$

che chiamiamo *formula di Bayes sulla scala dei pronoscici*: il membro di sinistra è il *pronostico finale* (posterior odds) di H, il primo fattore nel membri di destra è il *rapporto di verosimiglianza* (likelihood ratio) tra H e $\overline{H}$, il secondo fattore nel membro di destra è il *pronostico iniziale* (prior odds) di H.
Ogni probabilità $p \in]0, 1[$ corrisponde a un pronostico $r = p/(1-p)$ e si può scrivere come $p = r/(1+r)$ dove $r \in \mathbb{R}_+^* = \{x \in \mathbb{R} | x > 0\}$.
A partire da $\mathbb{P}(M)/\mathbb{P}(\overline{M}) = 2/98 = 1/49 \backsimeq 0.0204$ e $\mathbb{P}(T|M)/\mathbb{P}(T|\overline{M}) = 999/25 \backsimeq 40.0$ troviamo:

$$\frac{\mathbb{P}(H|E)}{\mathbb{P}(\overline{H}|E)} = \frac{999}{25} \times \frac{1}{49} \backsimeq 40 \times 0.0204 = 0.816$$

con la formula di Bayes sulla scala dei prnostici e confermiamo che $\mathbb{P}(H|E) = 0.816/(1+0.816) = 1/1.816 \backsimeq 0.45 = 45\%$.
Notiamo la separazione tra il segnale nei dati e le informazioni iniziali nella fattorizzazione sulla scala dei pronostici:

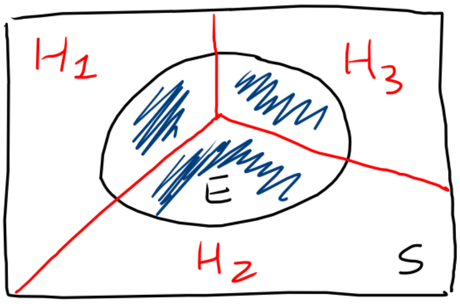
- Il *segnale nei dati* è rappresentato dal fattore $999/25 \simeq 40$;
- Le *informazioni iniziali* sono rappresentate dal fattore $1/49 \simeq 0.0204$.

Con più spiegazioni

Esempio: correzione automatica
L'utente di un computer ha digitato la parola "cartello". Supponiamo che questa sia la manifestazione della sua volontà di digitare una tra le parole "castello", "cartello" o "martello".
$E = \text{"digita cartello"}$

$H_1 =$ "vuole castello"
$H_1 =$ "vuole cartello"
$H_3 =$ "vuole martello"

$$\mathbb{P}(H_2|E) = \frac{\mathbb{P}(E|H_2)\mathbb{P}(H_2)}{\mathbb{P}(E)}$$



$$\mathbb{P}(E) = \mathbb{P}(E \cap H_1) + \mathbb{P}(E \cap H_2) + \mathbb{P}(E \cap H_3)$$

Per la regola delle probabilità totali:

$$\mathbb{P}(E) = \mathbb{P}(E|H_1)\mathbb{P}(H_1) + \mathbb{P}(E|M_2)\mathbb{P}(H_2) + \mathbb{P}(E|H_3)\mathbb{P}(H_3)$$

Dove: (valutazione frequentistica)

$$\begin{aligned}\mathbb{P}(H_1) &= \frac{153}{175} \quad \text{castello}\\ \mathbb{P}(H_2) &= \frac{9}{175} \quad \text{cartello}\\ \mathbb{P}(H_3) &= \frac{13}{175} \quad \text{martello}\end{aligned}$$

$\mathbb{P}(H_1) + \mathbb{P}(H_2) + \mathbb{P}(H_3) = \frac{175}{175} = 1$
(superficiale) Valutazione soggettiva

$$\begin{aligned}\mathbb{P}(E|H_1) &= 0.01 && \text{"r" vicina a "s"}\\ \mathbb{P}(E|H_2) &= 0.95 && \text{senza errori}\\ \mathbb{P}(E|H_3) &= 0.001 && \text{"m" lontana da "c"}\end{aligned}$$

$$\mathbb{P}(H_2|E) = \frac{\frac{9}{175} \cdot \frac{95}{100}}{\frac{153}{175} \cdot \frac{1}{100} + \frac{9}{175} \cdot \frac{95}{100} + \frac{13}{175} \cdot \frac{1}{1000}} \approx 0.847$$

$$\mathbb{P}(H_1|E) = \cdots = 0.152$$

$$\mathbb{P}(H_3|E) = \cdots = 0.001$$

Quindi la formula di Bayes per *tre spiegazioni*:

$$\mathbb{P}(H_i|E) \propto \mathbb{P}(E|H_i)\mathbb{P}(H_i)$$
$$\text{posterioir} \propto \text{likelihood} \times \text{prior}$$

Formula di Bayes con k spiegazioni

$$\mathbb{P}(H_i|E) = \frac{\mathbb{P}(E|H_i)\mathbb{P}(H_i)}{\sum_{i=1}^{k} \mathbb{P}(E|H_i)\mathbb{P}(H_i)} \qquad i = 1, \ldots, k$$

Conteggi

Prendiamo uno spazio campionario S costituito da eventi E, ricordiamo che $\mathbb{P}(E) = \frac{|E|}{|S|}$.

Regola del prodotto:

Se esistono m modi di fare qualcosa, e per ognuno di questi m modi ci sono n modi per fare un'altra cosa, allora ci sono un massimo $m \times n$ modi per fare entrambe le cose.

Combinazioni:

Il numero di modi di scegliere k oggetti distinti da un insieme di n oggetti:

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

$\binom{n}{k}$ in `R`:

```
choose(n,k)
```

Esempio:

Lanciamo una moneta 10 volte, la regola del prodotto ci dice che si sono $2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 2^{10} = 1024$ possibili risultati.

Sia E l'evento "abbiamo ottenuto esattamente 3 volte testa" (HHHTTTTTTT or TTTHTHTTHT, ...). Quale è la probabilità di E? $P(E) =$

$$|E| = \binom{10}{3} = \frac{10 \times 9 \times 8}{3 \times 2 \times 1} = 120$$

quindi $P(E) = \frac{120}{1024} \approx 0.117$.

```
event <- replicate(10000, {
  flips <- sample(c("H", "T"), 10, replace = TRUE)
  heads <- sum(flips == "H")
  heads == 3
})
mean(event)
```

```
## [1] 0.1211
```

```
## Title: Counting
## Author: Luca La Rocca
## Date: 15 October 2024

## Sys.setLanguage("en", unset="it") # uncomment to disable message
translation
rm(list=ls(all=TRUE)) # clean the workspace

prodCart <- function(spaces) # list of vectors
{ # begin function
  S <- expand.grid(rev(spaces))
```

```
  if(ncol(S)>1) S <- S[, ncol(S):1]
  colnames(S) <- paste("Pos", 1:ncol(S), sep="")
  return(S)
} # end function

dispRep <- function(S0, # vector of items available for selection
                     k) # number of items to be selected
  return(prodCart(rep(list(S0), k)))

dispSemp <- function(S0, # vector of items available for selection
                      k) # number of items to be selected
{ # begin function
  if(k==1){ # base case
    S <- as.data.frame(S0)
    colnames(S) <- "Pos1"
  }else{ # recursive case
    S <- Recall(S0[-1], k-1)
    recsize <- nrow(S)
    S <- cbind(rep(S0[1], recsize), S)
    colnames(S) <- paste("Pos", 1:ncol(S), sep="")
    for(i in 2:length(S0)){
      Schunk <- cbind(rep(S0[i], recsize), Recall(S0[-i], k-1))
      colnames(Schunk) <- colnames(S)
      S <- rbind(S, Schunk)
    } # end for
  } # end if-else
  return(S)
} # end function

perm <- function(S0) # vector of items to be permuted
  return(dispSemp(S0,length(S0)))

combSemp <- function(S0, # vector of items available for selection
                      k) # number of items to be selected
{ # begin function
  S <- as.data.frame(t(combn(S0, k)))
  colnames(S) <- paste("Item", 1:ncol(S), sep="")
  return(S)
} # end function

## SPAZI CAMPIONARI ELEMENTARI
cat("Sample space for a coin (head?):", fill=TRUE)
Scoin <- c(FALSE, TRUE)
print(Scoin)
cat("total outcomes =", length(Scoin), fill=TRUE)

cat("Sample space for a French card seed:", fill=TRUE)
Scard <- c("heart", "diamond", "club", "spade")
print(Scard)
cat("total outcomes =", length(Scard), fill=TRUE)

cat("Sample space for a die:", fill=TRUE)
Sdie <- 1:6
```

```r
print(Sdie)
cat("total outcomes =", length(Sdie), fill=TRUE)

## PRODOTTO CARTESIANO DI SPAZI FINITI
cat("Sample space for a coin (head?) and a French card seed:", fill=TRUE)
Scoincard <- prodCart(list(Scoin, Scard))
print(Scoincard)
cat("total outcomes =", length(Scoin)*length(Scard), fill=TRUE)

cat("Sample space for a coin (head?) a French card seed and a die:",
fill=TRUE)
Scoincardie <- prodCart(list(Scoin, Scard, Sdie))
print(Scoincardie)
cat("total outcomes =", length(Scoin)*length(Scard)*length(Sdie), fill=TRUE)

## SPAZI DI DISPOSIZIONI CON RIPETIZIONE
cat("Sample space for two dice:", fill=TRUE)
Stwodice <- dispRep(Sdie, 2)
print(Stwodice)
cat("total outcomes =", length(Sdie)^2, fill=TRUE)

cat("Sample space for three coins (head?):", fill=TRUE)
Sthreecoins <- dispRep(Scoin, 3)
print(Sthreecoins)
cat("total outcomes =", length(Scoin)^3, fill=TRUE)

## SPAZI DI DISPOSIZIONI SEMPLICI
cat("Sample space for an ordered hand of two French queens:", fill=TRUE)
Stwocards <- dispSemp(Scard, 2)
print(Stwocards)
cat("total outcomes =", prod(length(Scard):(length(Scard)-2+1)), fill=TRUE)

cat("Sample space for an ordered hand of three French queens:", fill=TRUE)
Sthreecards <- dispSemp(Scard, 3)
print(Sthreecards)
cat("total outcomes =", prod(length(Scard):(length(Scard)-3+1)), fill=TRUE)

## SPAZI DI PERMUTAZIONI
nlet <- 4 # number of letters
AnagramSimSize <- 2500 # number of simulations
cat("Sample space for a random anagram of", nlet, "distinct letters:",
    fill=TRUE)
AnagramS <- perm(LETTERS[1:nlet])
print(AnagramS)
cat("total outcomes =", factorial(length(AnagramS)), fill=TRUE)
AnagramD <- apply(AnagramS, 1, function(row) !any(row==LETTERS[1:nlet]))
cat("derangement event", fill=TRUE)
print(AnagramS[AnagramD,])
cat("favorable outcomes =", sum(AnagramD), fill=TRUE)
AnagramP <- sum(AnagramD)/nrow(AnagramS)
cat("exact probability of derangement =", AnagramP, fill=TRUE)
cat("how far is it from the limit value?",fill=TRUE)
print(all.equal(exp(-1), AnagramP))
```

```
AnagramDtrials <- replicate(AnagramSimSize,{
  permutation <- AnagramS[sample(1:nrow(AnagramS), 1),]
  !any(permutation==LETTERS[1:nlet])
})
cat("empirical proportion of derangement =", mean(AnagramDtrials),
    "with margin of error =", 1/sqrt(AnagramSimSize), fill=TRUE)

## SPAZI DI COMBINAZIONI SEMPLICI
cat("Sample space for an unordered hand of two French queens:", fill=TRUE)
SunordHand <- combSemp(Scard, 2)
print(SunordHand)
cat("total outcomes =", choose(length(Scard), 2), fill=TRUE)

cat("Sample space for an unordered hand of three French queens:", fill=TRUE)
SunordHand <- combSemp(Scard, 3)
print(SunordHand)
cat("total outcomes =", choose(length(Scard), 3), fill=TRUE)
```

Massa di probabilità e valore atteso

Variabili casuali discrete

In statistica e probabilità, una **variabile casuale** (o **variabile aleatoria**) è una funzione che associa a ciascun elemento di uno spazio campionario S un numero reale. Le variabili casuali sono denotate con lettere maiuscole.
La **funzione di massa di probabilità** (pmf) è una funzione associata a una variabile casuale discreta. Essa fornisce la probabilità che la variabile casuale assuma un determinato valore xx. La pmf di una variabile casuale XX è definita come:

$$p(x) = P(X = x)$$

Teorema 3.1:
Sia p la funzione di massa di probabilità di X.

1. $p(x) \geq 0 \forall x$, la probabilità di ciascun valore x deve essere maggiore o uguale a zero.
2. $\sum_x p(x) = 1$, la somma delle probabilità di tutti i possibili valori che può assumere la variabile casuale deve essere $= 1$.
   Esempio 3.2: Lancio di 3 monete
   Spazio campionario:

   $$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

   Ogni esito è ugualmente probabile, quindi la probabilità di ciascun risultato è:

   $$P(esito) = \frac{1}{8}$$

   Definiamo una variabile casuale X che rappresenta il numero di teste osservate nei tre lanci delle monete. La funzione X può essere descritta come segue:

$$
\begin{aligned}
X(HHH) &= 3 \quad nbsp; (3 \; teste) \\
X(HHT) &= 2 \quad nbsp; (2 \; teste) \\
X(HTH) &= 2 \quad nbsp; (2 \; teste) \\
X(THH) &= 2 \quad nbsp; (2 \; teste) \\
X(TTH) &= 1 \quad nbsp; (1 \; testa) \\
X(THT) &= 1 \quad nbsp; (1 \; testa) \\
X(HTT) &= 1 \quad nbsp; (1 \; testa) \\
X(TTT) &= 0 \quad nbsp; (0 \; teste)
\end{aligned}
$$

L'evento $X = 2$ rappresenta il numero di esiti in cui abbiamo esattamente 2 teste. Gli esiti che soddisfano questa condizione sono:

$$\{HHT, HTH, THH\}$$

possiamo calcolare la probabilità dell'evento $X = 2$:

$$P(X = 2) = P(\{HHT, HTH, THH\}) = P(HHT) + P(HTH) + P(THH)$$

ogni esito ha la stessa probabilità di $\frac{1}{8}$:

$$P(X = 2) = P(HHT) + P(HTH) + P(THH) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$

Valore previsto

La definizione formale del valore atteso per una variabile casuale discreta X con funzione di massa di probabilità (pmf) p è:

$$E[X] = \sum_x x \cdot p(x)$$

dove la somma è presa su tutti i valori possibili della variabile casuale X, a condizione che questa somma esista.

Teorema 3.2 La **Legge dei Grandi Numeri**

La Legge dei Grandi Numeri afferma che la media di n osservazioni di una variabile casuale X converge al valore atteso $E[X]$ quanto $n \to \infty$, assumendo che $E[X]$ sia definito:

$$\text{Se } n \to \infty, \quad \frac{1}{n} \sum_{i=1}^{n} X_i \to E[X]$$

dove X_i è il valore osservato della variabile casuale nella i-esima osservazione.

Esempio 3.9

Usando la simulazione, determiniamo il valore atteso di un rotolo di dado. Ecco 30 osservazioni e la loro media:

```
rolls <- sample(1:6, 30, replace = TRUE)
rolls
```

```
##  [1] 3 3 3 3 1 5 1 2 2 3 3 6 5 1 1 6 6 5 2 6 5 2 5 2 6 1 3 6 6 5
```

```
mean(rolls)
```

```
## [1] 3.6
```

Variabili Casuali Binomiali e Geometriche

Sono modelli comuni e utili per molte situazioni reali. Entrambe coinvolgono esperimenti chiamati **prove di Bernoulli**.

!TODO per esame

Prove di Bernoulli

Una **prova di Bernoulli** è un esperimento che può risultare in due esiti distinti, "successo" e "fallimento". La probabilità di un successo è rappresentata con p, mentre la probabilità di fallimento è quindi:

$$1 - p$$

Variabili Casuali Binomiali

È utilizzata per contare il numero di successi in un numero fisso di prove di Bernoulli. La variabile casuale X che rappresenta il numero di successi in n prove di Bernoulli (finite).

Schemi di Bernoulli finiti

Un **schema di Bernoulli finito** consiste in un numero fisso di prove n, ciascuna con una probabilità di successo p e una probabilità di fallimento $1 - p$. Questo schema è descritto dalla variabile casuale binomiale X, che rappresenta il numero di successi in n prove.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

- $\binom{n}{k}$ *variabile binomiale*: rappresenta il numero di modi in cui possono verificarsi k successi in n prove.
- k è il numero di successi (con $k = 0, 1, 2, \ldots, n$).
- p è la probabilità di successo in ogni prova.
  A volte si scrive $X \sim Binom(n, p)$, X segue una **distribuzione binomiale** con parametri n e p:
- n: numero prove di Bernoulli;
- p: **probabilità di successo** in ogni prova.
  Teorema 3.4:
  Sia X una variabile binomiale casuale, con n prove e p probabilità di successo, allora:

$$E[X] = np$$

Variabili Casuali Geometriche

Conta il numero di prove di Bernoulli fino al primo successo. Se Y rappresenta il numero di prove fino al primo successo (Schemi di Bermoulli infiniti).

Schemi di Bernoulli infiniti

Un **schema di Bernoulli infinito** prevede un numero potenzialmente *infinito* di prove. In questo schema, si continua a eseguire prove di Bernoulli fino a quando non si ottiene il primo successo. La variabile casuale che descrive il numero di prove necessarie fino al primo successo segue una distribuzione geometrica.

$$P(Y=k)=(1-p)^{k-1}p, \quad k=1,2,3,\ldots$$

- p è la probabilità di successo in ciascuna prova.
- $(1-p)^x$ *variabile geometrica*: appresenta la probabilità di ottenere x fallimenti prima di un successo.
  Teorema 3.6:
  Sia X una variabile geometrica casuale, con p probabilità di successo, allora:

$$E[X]=\frac{(1-p)}{p}$$

```
## Title: Binomial and geometric variables
## Author: Luca La Rocca
## Date: 22 October 2024

## Sys.setLanguage("en", unset="it") # uncomment to disable message
translation
rm(list=ls(all=TRUE)) # clean the workspace

plotPMF <- function(x, # vector of values taken with positive probability
                    p, # vector of corresponding probability masses
                    ...) # extra arguments for plot() and points()
{ # begin function
  plot(x, p, type="h", ylim=c(0,max(p)), frame=FALSE, ylab="p(x)", ...)
  points(x, p, pch=20, ...)
  abline(h=0, lty="dotted")
} # end function

plotCDF <- function(cdf, # step function giving the cumulative probabilities
                    main = "", # title of the plot
                    ...) # extra arguments for plot()
{ # begin function
  plot(cdf, verticals=FALSE, ylim=c(0,1), frame=FALSE, pch=20, ylab="F(x)",
       main=main, ...)
  abline(h=c(0,1), lty="dotted")
} # end function

## rolling a die ten times and counting the sixes
Bn <- 10
```

```
Bp <- 1/6

Blo <- 2 # lower threshold
Bhi <- 4 # upper threshold

cat("In n =", Bn, "repeated trials with success probability p =", Bp,
    fill=TRUE)
cat("the probability of getting at least", Blo, "and at most", Bhi,
    "successes is", sum(dbinom(Blo:Bhi, Bn, Bp)), fill=TRUE)
cat("the probability of getting less than", Blo,
    "successes is", pbinom(Blo-1, Bn, Bp), fill=TRUE)
cat("the probability of getting more than", Bhi,
    "successes is", pbinom(Bhi, Bn, Bp, lower=FALSE), fill=TRUE)
cat("and these three probabilities sum to", sum(dbinom(0:Bn, Bn, Bp)),
    fill=TRUE)

Bsimsize <- 10^6
cat("Simulating", Bsimsize, "binomial counts with n =", Bn, "and p =", Bp,
    fill=TRUE)
cat("the proportion of counts that are at least", Blo, "and at most", Bhi,
    "is", mean(rbinom(Bsimsize, Bn, Bp) %in% Blo:Bhi), fill=TRUE)
cat("with margin of error", 1/sqrt(Bsimsize), fill=TRUE)

Bvalues <- 0:Bn
Bmasses <- dbinom(Bvalues, Bn, Bp)

pdf("Rfig10dbinom.pdf", width = 8, height = 5)
plotPMF(Bvalues, Bmasses,
        main=paste("Binomial p.m.f. with n =", Bn, "and p =", round(Bp, 4)))
dev.off()

cbinom <- stepfun(Bvalues, c(0, pbinom(Bvalues, Bn, Bp)))

pdf("Rfig10pbinom.pdf", width = 8, height = 5)
plotCDF(cbinom,
        main=paste("Binomial c.d.f. ( n =", Bn, ", p =", round(Bp, 4), ")"))
dev.off()

## tossing a coin until you get head and counting the tails
Gp <- 1/2
Gmax <- qgeom(0.99, Gp)

Glo <- 3 # lower threshold
Ghi <- 5 # upper threshold

cat("In repeated trials with success probability p =", Gp, fill=TRUE)
cat("the probability of getting at least", Glo, "and at most", Ghi,
"failures",
    fill=TRUE)
cat("before the first success is", sum(dgeom(Glo:Ghi, Gp)), fill=TRUE)
cat("the probability of getting less than", Glo, "failures", fill=TRUE)
cat("before the first success is", pgeom(Glo-1, Gp), fill=TRUE)
cat("the probability of getting more than", Ghi, "failures", fill=TRUE)
```

```
cat("before the first success is", pgeom(Ghi, Gp, lower=FALSE), fill=TRUE)
cat("and the three probabilities sum to",
    pgeom(Ghi, Gp) + pgeom(Ghi, Gp, lower=FALSE), fill=TRUE)

Gsimsize <- 10^6
cat("Simulating", Gsimsize, "geometric counts with p =", Gp, fill=TRUE)
cat("the proportion of counts that are at least", Glo, "and at most", Ghi,
    "is", mean(rgeom(Gsimsize, Gp) %in% Glo:Ghi), fill=TRUE)
cat("with margin of error", 1/sqrt(Gsimsize), fill=TRUE)

Gvalues <- 0:Gmax
Gmasses <- dgeom(Gvalues, Gp)

pdf("Rfig10dgeom.pdf", width = 8, height = 5)
plotPMF(Gvalues, Gmasses, main=paste("Geometric p.m.f. with p =", round(Gp,
4)))
dev.off()

cgeom <- stepfun(Gvalues, c(0, pgeom(Gvalues, Gp)))

pdf("Rfig10pgeom.pdf", width = 8, height = 5)
plotCDF(cgeom, main = paste("Geometric c.d.f. ( p =", round(Gp, 4), ")"))
dev.off()
```

Trasformazioni e variabilità

Scarti quadratici e media

Consideriamo $\mu_X = E[X]$ e $\mu_Y = E[Y]$, le medie attese dei ricavi per i due prodotti. Gli scarti quadratici dai valori attesi per ciascuna variabile sono dati da:

$$(X - \mu_X)^2 \quad \text{e} \quad (Y - \mu_Y)^2$$

Queste espressioni rappresentano la distanza al quadrato tra il ricavo osservato e il suo valore medio. Questi scarti quadratici servono a misurare la dispersione o variabilità attorno alla media senza cancellare gli effetti di valori estremi (perché il quadrato elimina i segni negativi).

Trasformazioni di Variabili Aleatorie

Gli scarti quadratici sono anche esempi di **trasformazioni** delle variabili aleatorie X e Y, in cui la funzione $g(X) = (X - \mu_X)^2$ è applicata a X e produce una nuova variabile aleatoria, $g(X)$, che rappresenta lo scarto quadratico.
Queste trasformazioni possono essere utili in contesti statistici per valutare la distanza dal valore medio. Analogamente, altre trasformazioni come il valore assoluto $|X - \mu_X|$ possono essere utili in altre situazioni per ridurre la sensibilità ai valori estremi.

Legge dello Statistico Inconsapevole

La **legge dello statistico inconsapevole** afferma che per calcolare l'attesa di una funzione di una variabile aleatoria, basta la distribuzione di quella variabile e non serve conoscere esplicitamente la funzione. Per esempio, se vogliamo trovare l'attesa di $(X - \mu_X)^2$, non serve calcolare ogni valore di $(X - \mu_X)^2$ ma è sufficiente conoscere la distribuzione di X e applicare:

$$E[(X - \mu_X)^2] = Var(X)$$

Varianza e Deviazione Standard come Misura della Variabilità

La **varianza** di X, indicata come $Var(X)$, è data dall'attesa dello scarto quadratico:

$$Var(X) = E[(X - \mu_X)^2]$$

La varianza misura quanto i ricavi di X si discostano in media dalla loro media. La **deviazione standard** di $X, \sigma_X = \sqrt{Var(X)}$, fornisce una misura della dispersione dei ricavi di X attorno alla media nello stesso ordine di grandezza dei valori osservati.

Teorema della Linearità della Media

Se consideriamo una combinazione lineare di due variabili aleatorie X e Y, ad esempio $Z = aX + bY$, dove a e b sono costanti, possiamo calcolare la media di Z usando la **linearità dell'attesa**:

$$E[Z] = E[aX + bY] = aE[X] + bE[Y] = a\mu_X + b\mu_Y$$

Invarianza per Traslazione

Se aggiungiamo una costante c alla variabile aleatoria X per ottenere $X' = X + c$, a media di X' è semplicemente traslata di c:

$$E[X'] = E[X + c] = E[X] + c = \mu_X + c$$

La **varianza**, invece, non cambia, perché la dispersione attorno alla media rimane la stessa:

$$Var(X') = Var(X)$$

Esempio:
Vediamo un esempio concreto con dei dati numerici per ogni concetto che abbiamo trattato. Immaginiamo di avere due ricavi giornalieri, X e Y, di due prodotti in euro, osservati su un periodo di 5 giorni. I valori osservati per X e Y sono:

$$X = [120, 135, 150, 160, 140]$$
$$Y = [100, 115, 130, 125, 110]$$

Gli scarti quadratici della media:

$$\mu_X = \frac{120 + 135 + 150 + 160 + 140}{5} = 141$$
$$\mu_Y = \frac{100 + 115 + 130 + 125 + 110}{5} = 116$$

Ora calcoliamo gli scarti quadratici di ciascun valore dalla rispettiva media.
Per X:

$$(120 - 141)^2 = (-21)^2 = 441$$
$$(135 - 141)^2 = (-6)^2 = 36$$
$$(150 - 141)^2 = (9)^2 = 81$$
$$(160 - 141)^2 = (19)^2 = 361$$
$$(140 - 141)^2 = (-1)^2 = 1$$

Per Y:

$$(100 - 116)^2 = (-16)^2 = 256$$
$$(115 - 116)^2 = (-1)^2 = 1$$
$$(130 - 116)^2 = (14)^2 = 196$$
$$(125 - 116)^2 = (9)^2 = 81$$
$$(110 - 116)^2 = (-6)^2 = 36$$

La **varianza** di X si calcola come la media degli scarti quadratici:

$$Var(X) = \frac{441 + 36 + 81 + 361 + 1}{5} = \frac{920}{5} = 184$$

La **deviazione standard** di X è la radice quadrata della varianza:

$$\sigma_X = \sqrt{184} \approx 13.56$$

Analogamente, la **varianza** di Y è:

$$Var(Y) = \frac{256 + 1 + 196 + 81 + 36}{5} = \frac{570}{5} = 114$$

La **deviazione standard** di Y è:

$$\sigma_Y = \sqrt{114} \approx 10.68$$

Consideriamo ora una combinazione lineare dei due ricavi, per esempio $Z = 2X + 3Y$. Per trovare la media di Z, usiamo la **linearità della media**:

$$E[Z] = E[2X + 3Y] = 2E[X] + 3E[Y]$$

Sostituendo i valori delle medie trovate:

$$E[Z] = 2 \times 141 + 3 \times 116 = 282 + 348 = 630$$

Immaginiamo di aggiungere una costante di $c = 10$ euro al ricavo giornaliero di X per ogni giorno. Questo ci dà una nuova variabile aleatoria $X' = X + 10$.
La media di X' sarà:

$$E[X'] = E[X + 10] = E[X] + 10 = 141 + 10 = 151$$

La **varianza** di $X\prime$, invece, resta invariata, poiché aggiungere una costante non cambia la dispersione:

$$Var(X') = Var(X) = 184$$

La **legge dello statistico inconsapevole** ci permette di calcolare l'attesa di una trasformazione di X usando solo la distribuzione di X. Per esempio, se vogliamo calcolare l'attesa dello scarto quadratico $E[(X - \mu_X)^2]$, possiamo notare che questa non è altro che la **varianza** di X. Pertanto, anche senza calcolare esplicitamente ogni scarto, sappiamo che:

$$E[(X - \mu_X)^2] = Var(X) = 184$$

Indipendenza e correlazione

```
## Title: Law of Large Numbers
## Author: Luca La Rocca
## Date: 4 November 2024

## Sys.setLanguage("en", unset="it") # uncomment to disable message
translation
rm(list=ls(all=TRUE)) # clean the workspace

m <- 2500 # size of simulations
n <- 12 # number of Bernoulli trials
p <- 1/6 # success probability
g <- 100 # number of terms to keep when approximating Geom(p)

cat("Sampling m =", m, "observations from a binomial distribution",
fill=TRUE)
cat("with n =", n, "and p =", p, "gives...", fill=TRUE)
x <- rbinom(m, n, p)
cat("...an empirical mean equal to", mean(x),
    "with margin of error", 2*sd(x)/sqrt(length(x)), fill=TRUE)

pdf("Rfig12indepdraws.pdf", width=8, height=5)
plot(1:m, x, type="l", ylim=c(0, 12))
title(main=paste("Independent draws from Binom(n=", n,",p=", round(p, 4),
")",
                 sep=""))
dev.off()

pdf("Rfig12runningmean.pdf", width=8, height=5)
plot(1:m, cumsum(x)/1:m, type="l", ylim=c(0, 12), ylab=expression(bar(x)))
title(main=paste("Running mean from Binom(n=", n, ",p=", round(p, 4), ")",
                 sep=""))
abline(h=n*p, col=gray(0.75))
dev.off()

cat("Sampling m =", m, "observations from a geometric distribution",
fill=TRUE)
cat("with p =", p, "gives...", fill=TRUE)
y <- rgeom(m, p)
cat("...an empirical mean equal to", mean(y),
    "with margin of error", 2*sd(y)/sqrt(length(y)), fill=TRUE)
cat("whereas the population mean is", (1-p)/p,
    "or approximately", sum(0:(g-1)*dgeom(0:(g-1), p)),
    "(adding", g, "terms)", fill=TRUE)
```

```r
plot(1:m, cumsum(2^(y+1))/1:m, type="l", ylab=expression(bar(x)))
title(main=paste("Running mean from the St. Petersburg paradox"))
```