Basi della Probabilità

- La probabilità si occupa di **fenomeni aleatori**: esperimenti con risultati appartenenti a un insieme ben definito, ma non prevedibili.
- Dato uno **spazio campionario** (S), una probabilità valida soddisfa i **tre assiomi di probabilità**:
  1. Le probabilità sono reali e non negative: ($P(E) \geq 0$) per ogni evento (E).
  2. La probabilità dell'intero spazio campione è 1: ($P(S) = 1$).
  3. **Additività numerabile**: se ($A_1, A_2, \ldots$) sono eventi disgiunti, allora:

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

Teorema 2.1: Proprietà delle Probabilità

- Siano (A) e (B) eventi dello spazio campionario (S).
  1. ($P(\emptyset) = 0$)
  2. Se (A) e (B) sono disgiunti, ($P(A \cup B) = P(A) + P(B)$)
  3. Se ($A \subset B$), allora ($P(A) \leq P(B)$)
  4. ($0 \leq P(A) \leq 1$)
  5. ($P(A) = 1 - P(\overline{A})$)
  6. ($P(A - B) = P(A) - P(A \cap B)$)
  7. ($P(A \cup B) = P(A) + P(B) - P(A \cap B)$)

  Esempio 1: Lancio di un Dado
- **Problema**: Qual è la probabilità che esca una determinata faccia?
  - Sappiamo che ($P(\{1\}) + P(\{2\}) + P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\}) = 1$).
  - Poiché tutte le probabilità sono uguali, la probabilità di ottenere una determinata faccia è ($\frac{1}{6}$).

  Esempio 2.8: Lancio di Due Dadi
- **Spazio campione** (S) contiene tutte le combinazioni possibili:

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

  - **Evento** (E): Somma dei due dadi è (6), quindi ($E = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$).
    - Probabilità di (E): ($P(E) = \frac{|E|}{|S|} = \frac{5}{36}$).
  - **Evento** (F): Almeno uno dei due dadi è (2):

$$F = \{(2,1),(2,2),\ldots,(6,2)\}$$

```
- Probabilità di $( F )$: $( P(F) = \frac{11}{36} )$.
```

  - **Intersezione** ($E \cap F$): Eventi comuni a (E) e (F), ($E \cap F = \{(2,4),(4,2)\}$).
    - Probabilità di ($E \cap F$): ($P(E \cap F) = \frac{2}{36}$).
  - **Probabilità di** ($E \cup F$):

$$P(E \cup F) = P(E) + P(F) - P(E \cap F) = \frac{5}{36} + \frac{11}{36} - \frac{2}{36} = \frac{14}{36}$$

  - **Complemento di** (E):

$$P(\overline{E}) = 1 - P(E) = \frac{31}{36}$$

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

- Data uno spazio di probabilità $((S, \mathcal{F}, \mathbb{P}))$ e due eventi $(E, H \in \mathcal{F})$ con $(\mathbb{P}(H) > 0)$, la **probabilità condizionata** di (E) dato (H) è definita come:

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

  Questa quantità rappresenta la probabilità che (E) si verifichi supponendo che si sia verificato (H).

Esempi di Probabilità Condizionata

1. **Lancio di Due Dadi**:
   - Supponiamo di sapere che la somma dei due dadi è 8 e vogliamo calcolare la probabilità che entrambi mostrino 4.
     - **Eventi**:
     - $(A = \{(4,4)\})$: entrambi i dadi mostrano 4.
     - $(B = \{(2,6),(3,5),(4,4),(5,3),(6,2)\})$: eventi in cui la somma è 8.

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

Probabilità Condizionata Uniforme

- **Esempio**: Rotazione di uno Spinner simmetrico.
  - Calcoliamo la probabilità che lo spinner si fermi in un angolo compreso tra $(-\frac{\pi}{4})$ e $(\frac{\pi}{3})$, sapendo che l'angolo è positivo.

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

Valutazioni Frequentiste

- In probabilità frequentista, la probabilità condizionata è la frequenza relativa degli eventi in osservazioni precedenti.
  - Esempio: Lancio di un dado truccato.
    - Sia (A) l'evento "punteggio primo" ({2, 3, 5}) e (B) l'evento "punteggio pari" ({2, 4, 6}).
    - Calcoliamo ($\mathbb{P}(A|B)$) sulla base delle frequenze osservate:

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

Eventi indipendenti

Due eventi ($E, F \in \mathcal{F}$) in uno spazio di probabilità (($S, \mathcal{F}, \mathbb{P}$)) sono *indipendenti* sotto ($\mathbb{P}$) se:

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

Casi di indipendenza:

1. **Indipendenza banale:** quando ($\mathbb{P}(E) = 0$) o ($\mathbb{P}(F) = 0$).
   - Per monotonia, ($\mathbb{P}(E \cap F) = 0$).
2. **Indipendenza significativa:** quando ($\mathbb{P}(E) > 0$) e ($\mathbb{P}(F) > 0$), e si ha:

$$\mathbb{P}(E|F) = \mathbb{P}(E) \quad \text{e} \quad \mathbb{P}(F|E) = \mathbb{P}(F)$$

   Notazioni:

- ($E \perp\!\!\!\perp F$): indica indipendenza tra (E) e (F).
- ($E \perp\!\!\!\perp_{\mathbb{P}} F$): indica indipendenza rispetto a una probabilità ($\mathbb{P}$).
- ($E \not\!\perp\!\!\!\perp F$): indica dipendenza tra (E) e (F).
  Tipi di dipendenza:
- **Dipendenza favorevole:** ($\mathbb{P}(E \cap F) > \mathbb{P}(E)\mathbb{P}(F)$).
- **Dipendenza sfavorevole:** ($\mathbb{P}(E \cap F) < \mathbb{P}(E)\mathbb{P}(F)$).
  Esempio: Lancio di un dado:
  Eventi (A) ("primi") e (B) ("pari"):
- ($\mathbb{P}(A) = \frac{1}{2}$), ($\mathbb{P}(B) = \frac{1}{2}$).
- ($\mathbb{P}(A \cap B) = \frac{1}{6} < \frac{1}{4}$) $\Longrightarrow$ ($A \not\!\perp\!\!\!\perp B$) (dipendenza sfavorevole).

Invarianza per negazione

- Se ($E \perp\!\!\!\perp F$), allora ($E \perp\!\!\!\perp \overline{F}$) e viceversa.
- Indipendenza tra due eventi è una relazione simmetrica.
  Esempi di circuiti elettrici:
- **Serie:** corrente passa se entrambi gli interruttori sono chiusi.
  - ($\mathbb{P}(C_1 \cap C_2) = p^2$).
- **Parallelo:** corrente passa se almeno un interruttore è chiuso.
  - ($\mathbb{P}(\text{corrente}) = 1 - (1-p)^2$).

Simulazione in R per la probabilità condizionata

Calcolo della probabilità condizionata per due dadi:

1. Stima di ($\mathbb{P}(A \cap B)$) con simulazione.
2. Stima di ($\mathbb{P}(B)$).
3. Probabilità condizionata: ($\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$).

```r
# Stima della probabilità che la somma dei dadi sia almeno 10 e almeno uno dei dadi sia un 6
eventAB <- replicate(10000, {
  dieRoll <- sample(1:6, 2, replace = TRUE)
  (sum(dieRoll) >= 10) && (6 %in% dieRoll)
})
probAB <- mean(eventAB)

# Stima della probabilità che almeno uno dei dadi sia un 6
eventB <- replicate(10000, {
  dieRoll <- sample(1:6, 2, replace = TRUE)
  6 %in% dieRoll
})
probB <- mean(eventB)

# Calcolo della probabilità condizionata
```

```
conditional_prob <- probAB / probB
conditional_prob
```

Formula di Bayes

La formula di Bayes permette di calcolare la probabilità di un'ipotesi (H) (ad esempio, una condizione medica) alla luce di un'evidenza (E) osservata (come un risultato positivo al test). Viene formulata così:

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

- $(\mathbb{P}(H|E))$: probabilità dell'ipotesi dato l'evidenza,
- $(\mathbb{P}(E|H))$: probabilità dell'evidenza se l'ipotesi è vera,
- $(\mathbb{P}(H))$: probabilità a priori dell'ipotesi,
- $(\mathbb{P}(E))$: probabilità complessiva dell'evidenza.
  Esempi:

1. **Elezioni USA 2018**:
   - Si vuole calcolare la probabilità che un elettore Republican sia un Gun Owner.
   - Dati: $(\mathbb{P}(R) = 0.44)$, $(\mathbb{P}(G) = 0.46)$, $(\mathbb{P}(R|G) = 0.61)$.
   - Calcolo: $(\mathbb{P}(G|R) = \frac{\mathbb{P}(R|G) \cdot \mathbb{P}(G)}{\mathbb{P}(R)} \approx 0.64 = 64\%)$.
2. **Screening per una malattia rara**:
   - Calcolo della probabilità che un amico sia malato dopo un test positivo.
   - Dati: $(\mathbb{P}(M) = 0.02)$, $(\mathbb{P}(T|M) = 0.999)$, $(\mathbb{P}(\overline{T}|\overline{M}) = 0.975)$.
   - Si trova che $(\mathbb{P}(M|T) \approx 0.45 = 45\%)$, quindi la probabilità di malattia aumenta ma non è definitiva.
3. **Correzione automatica**:
   - L'utente digita “cartello”; si valutano tre ipotesi: "castello", "cartello" e "martello".
   - Le probabilità a posteriori vengono calcolate con la formula di Bayes estesa a tre ipotesi, ottenendo che l'ipotesi "cartello" è la più probabile.

Formula di Bayes con (k) Spiegazioni

Per (k) ipotesi, si generalizza la formula così:

$$\mathbb{P}(H_i|E) = \frac{\mathbb{P}(E|H_i) \cdot \mathbb{P}(H_i)}{\sum_{j=1}^{k} \mathbb{P}(E|H_j) \cdot \mathbb{P}(H_j)} \quad \text{per ogni } i = 1, \ldots, k$$

Questa versione è utile in applicazioni come il riconoscimento automatico e la diagnosi differenziale, dove si confrontano più ipotesi.

Probabilità basata sui Conteggi

La probabilità di un evento E in uno spazio campionario finito S si calcola come:

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

dove $\mid E \mid$ è il numero di risultati favorevoli e $\mid S \mid$ è il numero totale di risultati possibili.

Regola del prodotto:

Se esistono m modi di fare qualcosa e n modi per fare un'altra cosa, il numero di modi per fare entrambe le cose è $m \times n$.
Combinazioni:
Il numero di modi di scegliere k oggetti distinti da un insieme di n oggetti è dato dal coefficiente binomiale:

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

In `R`:

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

Esempio:
Lanciamo una moneta 10 volte, la regola del prodotto ci dice che si sono $2^{10} = 1024$ 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:

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

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

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

```r
## 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
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)
```

```r
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

Una variabile casuale X è una funzione che associa a ciascun risultato di uno spazio campionario S un numero reale, rappresentando quantitativamente l'esisto di un fenomeno aleatorio.

Funzione di Massa di probabilità

La **funzione di massa di probabilità** (pmf) $p(x)$ di una variabile casuale discreta X assegna a ogni valore x la probabilità che X assuma qual valore:

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

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 atteso per variabili discrete

Il valore atteso di una variabile casuale discreta X, con funzione di massa di probabilità p, è:

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

Questa somma pondera ogni valore x di X con al sua probabilità $p(x)$, fornendo una misura media degli esiti attesi.
Teorema 3.2 La Legge dei Grandi Numeri
La Legge dei Grandi Numeri afferma che la media di n osservazioni di X converge al valore atteso $E[X]$ quando n tende a infinito:

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

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 con due esiti distinti, "successo" con probabilità p e "fallimento" con probabilità $1 - p$.

Variabili Casuali Binomiali

Le variabili binomiali contano il numero di successi in un numero finito n di prove di Bernoulli indipendenti, ciascuna con la stessa probabilità di successo p.

Schemi di Bernoulli finiti

Un schema di Bernoulli finito descrive un processo con un numero fisso di prove. La variabile casuale X rappresenta il numero di successi in n prove, e segue una distribuzione binomiale:

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

- $\binom{n}{k}$ *variabile binomiale*: rappresenta il numero di modi in cui possono ottenere k successi su 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:
  Se $X \sim Binom(n, p)$, allora il valore atteso di X è:

$$E[X] = np$$

Variabili Casuali Geometriche

Le variabili geometriche contano il numero di prove di Bernoulli necessarie fino al primo successo.

Schemi di Bernoulli infiniti

Uno schema di Bernoulli infinito si basa su un numero indefinito di prove, eseguite finché non si ottiene il primo successo. La variabile casuale Y che rappresenta il numero di prove 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.
- $(1-p)^{k-1}$ *variabile geometrica*: appresenta la probabilità di ottenere $k - 1$ fallimenti prima del primo successo.
  Teorema 3.6:
  Se $Y \sim Geom(p)$, allora il valore atteso di Y è:

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

```r
## 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
```

```r
## 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

Gli **scarti quadratici** rappresentano la distanza tra un valore osservato e il valore medio (o atteso) di una variabile aleatoria. Vengono utilizzati per misurare la dispersione dei valori rispetto alla loro media, ed essendo elevati al quadrato, eliminano i segni negativi. Gli scarti quadratici di X e Y rispetto alle loro medie μ_X e μ_Y sono:

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

Questi valori servono per calcolare la **varianza** e altre misure di dispersione.'

Trasformazioni di Variabili Aleatorie

Lei **trasformazioni** di una variabile aleatoria come $g(X)=(X-\mu_X)^2$ o $\mid X-\mu_X \mid$ è sono utili per analizzare la **dispersione** o per ridurre l'influenza dei valori estremi. Applicando una funzione a una variabile aleatoria, otteniamo una nuova variabile aleatoria, come nel caso della trasformazione degli scarti quadratici, che misura quanto i valori si discostano dalla media.

Legge dello Statistico Inconsapevole

La **legge dello statistico inconsapevole** ci dice che per calcolare l'attesa di una funzione di una variabile aleatoria, non dobbiamo calcolare esplicitamente i valori della funzione, possiamo usare direttamente la distribuzione della variabile aleatoria:

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

Quindi, possiamo determinare il valore atteso degli scarti quadratici (che definiscono la varianza) senza dover calcolare ogni singolo scarto.

Varianza e Deviazione Standard come Misura della Variabilità

La **varianza** di una variabile aleatoria X è definita come l'attesa degli scarti quadratici:

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

La **deviazione standard** $\sigma_X=\sqrt{Var(X)}$, 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 $Z=aX+bY$, dove a e b sono costanti, la media di Z può essere calcolata come la somma delle media di X e Y, pesate dai rispettivi coefficienti:

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

Invarianza per Traslazione

Se aggiungiamo una costante c alla variabile aleatoria X, ottenendo una nuova variabile $X'=X+c$, la media di X' sarà traslata traslata di c:

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

La **varianza**, invece, non cambia, perché aggiungere una costante non influisce sulla dispersione attorno alla media:

$$\mathrm{Var}(X')=\mathrm{Var}(X)$$

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:

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

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

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

Analogamente, la varianza di Y è:

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

La deviazione standard di Y è:

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

Combinazione lineare

Per una combinazione lineare $Z = 2X + 3Y$. 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:

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

Se aggiungiamo una costante $c = 10$ a X la una nuova variabile aleatoria $X' = X + 10$ avrà:

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

La varianza di $X\prime$, invece, resta invariata:

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

Indipendenza e correlazione

Indipendenza e Correlazione tra Variabili Aleatorie

- Indipendenza: Due variabili aleatorie X e Y sono indipendenti se la probabilità congiunta si separa come prodotto delle probabilità marginali:

  $$P(X = x \text{ e } Y = y) = P(X = x) \cdot P(Y = y)$$

  Questa definizione implica che la realizzazione di una variabile aleatoria non ha alcuna influenza sulla realizzazione dell'altra.
- Incorrelazione: Se due variabili aleatorie sono indipendenti, allora sono anche incorrelate. Questo significa che la loro covarianza è zero:

  $$Cov(X, Y) = 0$$

  L'inverso non è vero: due variabili aleatorie possono essere incorrelate senza essere indipendenti. In altre parole, la covarianza può essere zero anche quando le variabili sono dipendenti, ma non correlati linearmente.

Varianza della Somma di Due Variabili Aleatorie

La varianza della somma di due variabili aleatorie X e Y è data da:

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$$

Se X e Y sono indipendenti, allora la covarianza è zero e la varianza della somma si semplifica a:

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Esempio: Supponiamo che X e Y rappresentino i rendimenti di due azioni indipendenti. Se $Var(X) = 4$ e $Var(Y) = 9$, allora:

$$\mathrm{Var}(X+Y) = 4 + 9 = 13$$

Covarianza

La covarianza tra X e Y misura il grado di dipendenza lineare tra le due variabili:

$$\mathrm{Cov}(X,Y) = E[(X-E[X])(Y-E[Y])] = E(XY) - E(X)E(Y)$$

- $E[X]$ e $E[Y]$ sono i valori attesi (o medie) di X e Y.
- $E[(X-E[X])(Y-E[Y])]$ è l'atteso del prodotto delle deviazioni di X e Y dalle loro medie.
- Se $Cov(X,Y) > 0$, significa che X e Y tendono a crescere insieme.
- Se $Cov(X,Y) < 0$, significa che X e Y tendono a variare in direzioni opposte.

Coefficiente di Correlazione Lineare

Il coefficiente di correlazione lineare è una versione normalizzata della covarianza, che assume valori tra -1 e 1:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot \mathrm{Var}(Y)}}$$

- Un valore di $\rho_{X,Y}$ vicino a 1 indica una forte correlazione positiva.
- Un valore di $\rho_{X,Y}$ vicini a -1 indica una forte correlazione negativa.
- Un valore di $\rho_{X,Y}$ vicini a 0 indica che non c'è relazione lineare.
  Esempio:
  Se $\rho_{X,Y} = 0.8$, significa che X e Y sono fortemente correlati positivamente.

Campioni Casuali e Legge dei Grandi Numeri

La Legge dei Grandi Numeri afferma che, per un campione casuale di grande dimensione $(X_1, X_2, \ldots, X_n)$, la media campionaria tende a convergere alla media della popolazione:

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

- **Campione casuale**: Un campione di variabili casuali $(X_1, X_2, \ldots, X_n)$ è preso da una distribuzione di probabilità, dove ogni XiXi è una realizzazione casuale.
- **Media campionaria**: è la media dei valori osservati nel campione, e rappresenta una stima della media della popolazione.
- La **convergenza** implica che man mano che aumentiamo il numero di osservazioni, la media campionaria si avvicina sempre più alla media teorica $E[X]$.
  Esempio: Se lanciamo una moneta n volte e contiamo il numero di teste, la proporzione di teste tende a 0.5 al crescere di n, che è il valore atteso.

Esempio (Un portafoglio prudente)

Supponiamo di avere un capitale di 9.000 euro investito in due titoli finanziari i cui tassi di rendimento sono X e Y. Il capitale è suddiviso in due terzi e un terzo, con il portafoglio di investimento dato da:

$$T = 6(1+X) + 3(1+Y) = 9 + 6X + 3Y = 9(1+W)$$

dove $W = \frac{2}{3}X + \frac{1}{3}Y$ è il tasso di rendimento del portafoglio. Se $E(X) = 0.03$ e $E(Y) = 0.06$, possiamo calcolare la media del tasso di rendimento del portafoglio:

$$E(W) = \frac{2}{3}E(X) + \frac{1}{3}E(Y) = \frac{0.06+0.06}{3} = 0.04$$

quindi

$$E(T) = 9\{1 + E(W)\} = 9 \times 1.04 = 9.360 \text{ migliaia di euro}$$

Per calcolare la varianza e la deviazione standard del portafoglio, considerando che $Var(X) = 0.02^2$, $Var(Y) = 0.04^2$ e $Cov(X,Y) = -0.0005$, troviamo:

$$Var(W) = \frac{4}{9}Var(X) + \frac{4}{9}Cov(X,Y) + \frac{1}{9}Var(Y)$$
$$Var(W) = \frac{4\times 0.02^2 + 4\times(-0.0005) + 0.04^2}{9} = \frac{0.0012}{9} = \frac{1}{7500}$$

Quindi:

$$Sd(W) = \sqrt{\frac{1}{7500}} \approx 0.011547$$

La deviazione standard del portafoglio è:

$$Sd(T) = 9 \times Sd(W) \approx 0.104 \text{ migliaia di euro}$$

In questo esempio, abbiamo visto che $Sd(W) < Sd(X) < Sd(Y)$.

```r
## 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)

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"))
```

Altre variabili discrete di uso comune

Variabile di Poisson

La distribuzione di Poisson è utilizzata per modellare il numero di eventi che si verificano in un intervallo di tempo o in un'area spaziale, quando gli eventi sono indipendenti e avvengono con una probabilità costante.

Definizione

Una variabile casuale (X) segue una distribuzione di Poisson con parametro $(\lambda > 0)$ se la sua funzione di probabilità è:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$

- (λ) è il parametro di Poisson, che rappresenta sia la media che la varianza della distribuzione.
- (k) è il numero di eventi osservati.
- (e) è la base del logaritmo naturale.

Caratteristiche

- Media: $(E[X] = \lambda)$
- Varianza: $(\mathrm{Var}(X) = \lambda)$
- Distribuzione discreta: (X) può assumere valori interi non negativi (0, 1, 2, …).

Condizioni d'uso

La distribuzione di Poisson è applicabile per modellare:

- Eventi che accadono in un intervallo di tempo o spazio fisso.
- Eventi indipendenti tra loro.
- Eventi con una probabilità costante di occorrenza.

Esempio

Se una biblioteca riceve in media 3 visitatori all'ora $(\lambda = 3)$, la probabilità che ci siano esattamente 5 visitatori in un'ora è:

$$P(X = 5) = \frac{3^5 e^{-3}}{5!}$$

Relazione con la Binomiale

La distribuzione di Poisson può approssimare una distribuzione binomiale quando il numero di prove è grande e la probabilità di successo in ogni prova è piccola, mantenendo costante il valore $(n \cdot p = \lambda)$.

Variabili Ipergeometriche

La **distribuzione ipergeometrica** è utilizzata quando si selezionano campioni da una popolazione finita senza **sostituzione**. A differenza della distribuzione binomiale, che assume la selezione con sostituzione, la distribuzione ipergeometrica tiene conto che, con ogni estrazione, cambia la composizione della popolazione.

Definizione

Una variabile casuale (X) segue una **distribuzione ipergeometrica** con i seguenti parametri:

- (N): la **dimensione della popolazione** (numero totale di oggetti).
- (K): il numero di **successi** (oggetti "di interesse", ad esempio, oggetti di un certo tipo).
- (n): la **dimensione del campione** (numero di oggetti estratti dalla popolazione).
  La funzione di probabilità della variabile casuale (X), che rappresenta il numero di successi nel campione, è:

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

- $\binom{K}{k}$ è il numero di modi in cui si possono scegliere (k) successi dal totale (K) successi.
- $\binom{N-K}{n-k}$ è il numero di modi in cui si possono scegliere $(n-k)$ fallimenti dal totale di $(N-K)$ fallimenti.
- $\binom{N}{n}$ è il numero totale di modi in cui si possono scegliere (n) oggetti dalla popolazione di dimensione (N).

Caratteristiche

- **Supporto**: (X) può assumere valori interi da $(\max(0, n-(N-K)))$ a $(\min(n, K))$.
- **Media**: $(E[X] = \frac{nK}{N})$
- **Varianza**: $(\mathrm{Var}(X) = \frac{nK(N-K)(N-n)}{N^2(N-1)})$

Applicazioni

La distribuzione ipergeometrica è utile in contesti in cui:

- Si estraggono campioni **senza sostituzione** da una popolazione finita.
- Si vuole calcolare la probabilità di ottenere un certo numero di successi (oggetti di interesse) in un campione estratto dalla popolazione.

Esempio

Supponiamo di avere una popolazione di 20 oggetti, di cui 8 sono di tipo "successo" e i restanti 12 sono di tipo "fallimento". Se estrai 5 oggetti senza sostituzione, la probabilità di ottenere esattamente 3 successi è:

$$P(X = 3) = \frac{\binom{8}{3}\binom{12}{2}}{\binom{20}{5}}$$

Confronto con la distribuzione binomiale

La distribuzione ipergeometrica è simile alla distribuzione binomiale, ma differisce per il fatto che gli eventi non sono indipendenti, dato che il campione viene estratto senza sostituzione. La distribuzione binomiale può approssimare la distribuzione ipergeometrica quando il campione è molto piccolo rispetto alla popolazione, in modo che la differenza tra le estrazioni con e senza sostituzione sia trascurabile.

Variabili Binomiali Negative

Le **variabili binomiali negative** sono utilizzate per modellare il numero di prove necessarie prima di ottenere un numero fisso di successi in un processo di Bernoulli. In altre parole, una variabile casuale binomiale negativa misura quante prove devono essere effettuate per ottenere (r) successi, con una probabilità (p) di successo in ogni prova.

Definizione

Una variabile casuale (X) segue una **distribuzione binomiale negativa** con parametri (r) (numero di successi) e (p) (probabilità di successo in ogni prova) se la probabilità che $(X = k)$ (cioè, ottenere il (r)-esimo successo alla (k)-esima prova) è data da:

$$P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, r+2, \ldots$$

- $\binom{k-1}{r-1}$ è il numero di modi in cui i $(r-1)$ successi possono essere distribuiti tra le prime $(k-1)$ prove.
- (p^r) è la probabilità di ottenere (r) successi.
- $(1-p)^{k-r}$ è la probabilità di ottenere $(k-r)$ fallimenti.

Caratteristiche

- **Supporto**: (X) assume valori interi $(k = r, r+1, r+2, \ldots)$, cioè il numero di prove deve essere almeno uguale a (r), poiché non è possibile ottenere (r) successi in meno di (r) prove.
- **Media**: $(E[X] = \frac{r}{p})$, cioè il numero medio di prove necessarie per ottenere (r) successi.
- **Varianza**: $(\text{Var}(X) = \frac{r(1-p)}{p^2})$, che dipende sia dal numero di successi desiderato che dalla probabilità di successo in ogni prova.

Applicazioni

La distribuzione binomiale negativa è utile quando si desidera modellare:

- Il **numero di prove** necessarie per ottenere un numero specifico di successi.
- Eventi che seguono un **processo di Bernoulli** (due possibili esiti, successo o fallimento).
  Alcuni esempi includono:
- Il numero di lanci di una moneta necessari per ottenere 5 teste.
- Il numero di tentativi necessari per ottenere 3 successi in un gioco o esperimento.

Esempio

Supponiamo di lanciare una moneta con probabilità di successo $(p = 0.5)$ (ad esempio, ottenere "testa") e vogliamo sapere quante prove sono necessarie per ottenere 3 teste. Se (X) rappresenta il numero di lanci necessari per ottenere 3 teste, allora (X) segue una distribuzione binomiale negativa con parametri $(r = 3)$ e $(p = 0.5)$.
La probabilità di ottenere il 3° successo al 6° lancio (cioè 2 teste nei primi 5 lanci e la 3ª testa nel 6° lancio) è:

$$P(X = 6) = \binom{5}{2} (0.5)^3 (0.5)^3 = \binom{5}{2} (0.5)^6 = 10 \times \frac{1}{64} = \frac{10}{64} = \frac{5}{32}$$

Relazione con la distribuzione geometrica

La distribuzione binomiale negativa è una generalizzazione della **distribuzione geometrica**. La distribuzione geometrica è il caso particolare della binomiale negativa con $(r = 1)$, cioè quando si cerca il numero di prove necessarie per ottenere il **primo** successo.

Densità di probabilità

La **funzione di densità di probabilità** (PDF) è associata a una **variabile casuale continua** e descrive la probabilità che la variabile casuale assuma un valore all'interno di un intervallo. A differenza delle variabili casuali discrete, per cui si utilizza la funzione di massa di probabilità (PMF), la PDF si applica a **variabili casuali continue**.

Definizione

Per una variabile casuale continua (X), la **funzione di densità di probabilità** $(f_X(x))$ è definita in modo tale che la probabilità che (X) assuma un valore in un intervallo $([a, b])$ è data dall'integrale della PDF su quell'intervallo:

$$P(a \leq X \leq b) = \int_a^b f_X(x)\, dx$$

Le proprietà principali della PDF sono:

1. **Non negatività**: $(f_X(x) \geq 0)$ per ogni (x), cioè la densità di probabilità non può mai essere negativa.
2. **Normalizzazione**: L'integrale della funzione di densità su tutto il dominio deve essere uguale a 1:

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1\$\$ Questo assicura che la probabilità totale sia 1.$$

Interpretazione

La PDF non fornisce direttamente la probabilità che la variabile casuale (X) assuma un valore specifico, ma la probabilità che (X) si trovi in un intervallo $([a, b])$ è data dall'area sotto la curva della PDF tra (a) e (b).

Esempi

1. **Distribuzione Normale** (Gaussiana):
   La distribuzione normale con media (μ) e varianza (σ^2) ha la funzione di densità di probabilità:

   $$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

   È una delle distribuzioni più comuni in statistica e probabilità.
2. **Distribuzione Esponenziale**:
   La distribuzione esponenziale con parametro (λ) ha la PDF:

   $$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

   Viene spesso utilizzata per modellare il tempo tra eventi in un processo di Poisson.
3. **Distribuzione Uniforme**:
   Una variabile casuale (X) che segue una distribuzione uniforme nell'intervallo $([a, b])$ ha la PDF:

   $$f_X(x) = \frac{1}{b-a}, \quad a \leq x \leq b$$

   In questo caso, la probabilità è distribuita uniformemente tra (a) e (b).

Variabili Casuali Continue

Il **valore atteso** (o **media**) di una **variabile casuale continua** (X) è una misura della posizione centrale della distribuzione della variabile. Rappresenta la media ponderata dei valori che (X) può assumere, con i pesi dati dalla sua **funzione di densità di probabilità** (PDF).

Definizione

Per una variabile casuale continua (X) con funzione di densità di probabilità $(f_X(x))$, il **valore atteso** $(E[X])$ è dato dall'integrale della variabile moltiplicata per la sua densità di probabilità:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx$$

- (x) è il valore che (X) può assumere.
- $(f_X(x))$ è la densità di probabilità di (X), che descrive la probabilità di (X) che assume valori in un intervallo infinitesimo.

Caratteristiche

1. **Sommatoria continua**: A differenza delle variabili discrete, dove si sommano le probabilità per ogni valore, per variabili continue si integra la densità di probabilità su tutto il dominio.
2. **Media teorica**: Il valore atteso rappresenta il "centro di massa" della distribuzione e corrisponde alla media dei valori se l'esperimento venisse ripetuto infinite volte.

Proprietà

- Il valore atteso è lineare: Se (a) e (b) sono costanti, e (X) è una variabile casuale continua, allora:

$$E[aX + b] = aE[X] + b$$

Esempio

Supponiamo che (X) segua una distribuzione uniforme nell'intervallo $([0, 1])$, con densità $(f_X(x) = 1)$ per $(x \in [0, 1])$. Il valore atteso di (X) è:

$$E[X] = \int_0^1 x \cdot 1\,dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2}$$

Varianza e Deviazione Standard

La varianza misura la dispersione di una variabile casuale rispetto al suo valore atteso (media). Indica quanto i valori di una variabile casuale si discostano dalla media. Per una variabile casuale (X), la varianza è definita come:

- Per variabili continue:

$$\mathrm{Var}(X) = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 \cdot f_X(x)\,dx$$

  - $(f_X(x))$ è la funzione di densità di probabilità di (X).
    La varianza è sempre positiva e il suo valore è espresso nelle stesse unità al quadrato della variabile casuale.

Deviazione Standard

La deviazione standard è la radice quadrata della varianza e fornisce una misura della dispersione che ha le stesse unità della variabile casuale. La deviazione standard è spesso usata per interpretare quanto una variabile casuale si discosta in media dal suo valore atteso.

$$\text{Deviazione Standard}(X) = \sqrt{\mathrm{Var}(X)}$$

Proprietà della Varianza

1. Varianza di una somma lineare:
   Se (X) e (Y) sono variabili casuali, e (a) e (b) sono costanti, allora:

$$\mathrm{Var}(aX + bY) = a^2 \cdot \mathrm{Var}(X) + b^2 \cdot \mathrm{Var}(Y)$$

   se (X) e (Y) sono indipendenti.
2. Varianza e Deviazione Standard di una Costante:
   La varianza di una costante (c) è zero:

$$\mathrm{Var}(c) = 0$$

   La deviazione standard di una costante è anch'essa zero:

$$\text{Deviazione Standard}(c) = 0$$

Distribuzione normale standard

La distribuzione normale è fondamentale in statistica per i seguenti motivi:

1. Central Limit Theorem (Teorema del Limite Centrale): Molti fenomeni naturali e misurazioni possono essere modellati da distribuzioni normali perché rappresentano la somma di molti fattori indipendenti e piccoli.
   - Esempio: L'altezza degli adulti può essere modellata come una normale, influenzata da fattori genetici ed ambientali.
2. Applicazioni pratiche: Misure biometriche, punteggi di test, indicatori economici, errori di misura scientifici e variazioni nei processi di produzione sono spesso vicini a una distribuzione normale.

Forma Matematica della Distribuzione Normale

$$h(x) = e^{-x^2}$$

Questa funzione produce una curva a forma di campana, ma non è una distribuzione perché non ha area unitaria sotto la curva. L'integrale di questa funzione è:

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$$

Rescalando la funzione, si ottiene la densità di probabilità (pdf) della distribuzione normale standard:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Proprietà della Normale Standard Z:

- Media $(\mu) = 0$
- Deviazione standard $(\sigma) = 1$

Regola empirica

Descrive la proporzione di dati che si trovano entro 1, 2 e 3 deviazioni standard dalla media:

1. Circa il 68% dei dati è entro $\mu \pm \sigma$
2. Circa il 95% dei dati è entro $\mu \pm 2\sigma$
3. Circa il 99.7% dei dati è entro $\mu \pm 3\sigma$

Distribuzione normale generica

La distribuzione normale generica $X \sim Norm(\mu, \sigma)$ si ottiene trasformando una normale standard Z con:

$$X = \sigma Z + \mu$$

- μ: media(centro della distribuzione)
- σ: deviazione standard(larghezza della curva)
  La pdf della distribuzione normale generica è:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

Calcoli con la Normale

- `pnorm(x)` : calcola la probabilità cumulativa $P(Z \leq x)$.
- `dnorm(x)` : calcola il valore della densità $f(x)$.
- `rnorm(n)` : genera n valori casuali da una normale.
- `qnorm(p)` : calcola il quantile per una probabilità p.

Approssimazione normale al binomio

La *distribuzione binomiale* $X \sim Binom(n, p)$ rappresenta la somma di n prove di Bernoulli indipendenti con probabilità di successo p. Quando il numero di prove n è grande, la distribuzione binomiale può essere approssimata da una distribuzione normale con la *stessa media* e *deviazione standard*.

Varianza e Deviazione Standard di una Variabile Binomiale

Per $X \sim Binom(n, p)$:

- La *media* è:

$$\mu = E[X] = np$$

- La *varianza* è:

$$Var(X) = np(1.p)$$

- La *derivazione standard* è:

$$\sigma = \sqrt{np(1-p)}$$

  Questi parametri definiscono completamente la distribuzione binomiale e permettono l'approssimazione normale.

Teorema 4.6: Approssimazione Normale

Per $X \sim Binom(n, p)$, se n è sufficientemente grande, possiamo approssimare X con una variabile casuale normale $Y \sim \mathcal{N}(\mu, \sigma)$, dove:

- $\mu = np$
- $\sigma = \sqrt{np(1-p)}$
  Quando l'approssimazione è valida?
- n deve essere grande.
- p non deve essere troppo vicino a 0 o 1. In genere si richiede che:

$$np \geq 5 \quad \text{e} \quad n(1-p) \geq 5$$

Correzione per la continuità

La distribuzione binomiale è discreta, mentre la normale è continua. Per migliorare l'approssimazione, si utilizza la **correzione per la continuità**:

$$P(a \leq X \leq b) \approx P(a - 0.5 \leq Y \leq b + 0.5)$$

Dove $Y \sim \mathcal{N}(np, \sqrt{np(1-p)})$.
Esempio:
Supponiamo $X \sim Binom(n = 50, p = 0.4)$:

1. Calcoliamo i parametri della normale:

$$\mu = np = 50 \times 0.4 = 20, \quad \sigma = \sqrt{np(1-p)} = \sqrt{50 \times 0.4 \times 0.6} \approx 3.46$$

2. Stimiamo $P(18 \leq X \leq 22)$ usando l'approssimazione normale:
   - Correzione per la continuità: $P(17.5 \leq Y \leq 22.5)$
   - Standardizziamo:

$$Z_1 = \frac{17.5 - 20}{3.46} \approx -0.72, \quad Z_2 = \frac{22.5 - 20}{3.46} \approx 0.72$$

   - Usando una tabelle della normale standard o una funzione come `pnorm` in R:

$$P(17.5 \leq Y \leq 22.5) = P(z_1 \leq Z \leq Z_2) = \Phi(0.72) - \Phi(-0.72) \approx 0.764$$

Variabili casuali uniformi ed esponenziali

Variabili casuali uniformi

Le **variabili casuali uniformi** possono essere **discrete** o **continue**:

1. *Uniforme discreta*: può assumere un numero finito di valori, tutti con probabilità uguale. Ad esempio:
   1. **Dada**: Ogni numero tra 1 e 6 ha la stessa probabilità.
   2. **Moneta**: Se assegniamo 1 alla testa e 0 alla croce, entrambi i valori sono equiprobabili.
2. *Uniforme continua*: Può assumere qualsiasi valore in un intervallo continuo $[a, b]$. La probabilità è uniformemente distribuita su questo intervallo, con una **densità di probabilità (pdf)** data da:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } a \leq x \leq b, \\ 0 & \text{altrimenti.} \end{cases}$$

   Proprietà principali:

- **Media $(E[X])$**: La media è il punto medio dell'intervallo, cioè:

$$E[X] = \frac{a+b}{2}$$

- **Varianza $(Var(X))$**: La varianza è proporzionale alla lunghezza dell'intervallo:

$$Var(X) = \frac{(b-a)^2}{12}$$

  Esempio:
  Se scegliamo un numero casuale uniforme nell'intervallo $[0, 10]$:
- Qual è la probabilità che sia maggiore di 7, sapendo che è maggiore di 6? La variabile condizionata è uniforme su $[6, 10]$. La probabilità è calcolata come:

$$P(X > 7 | X > 6) = \frac{\text{lunghezza intervallo}(7, 10)}{\text{lunghezza intervallo}(6, 10)} = \frac{3}{4}$$

Variabili casuali esponenziali

Una variabile casuale esponenziale misura il **tempo di attesa fino al primo evento** in un processo di Poisson. È descritta da una densità di probabilità:

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

λ è il *tasso* degli eventi(eventi per unità di tempo).
Proprietà principali:

- **Media** $(E[X])$: il tempo medio di attesa è l'inverso del tasso:

$$E[X] = \frac{1}{\lambda}$$

- **Varianza** $(Var(X))$:

$$Var(X) = \frac{1}{\lambda^2}$$

  Proprietà della memoria:
  Le variabili esponenziali hanno la **proprietà di assenza di memoria**, ovvero:

$$P(X > s + t | X > s) = P(X > t)$$

  Significa che il tempo di attesa per un evento non dipende dal tempo già trascorso.
  Esempio:
  Nel caso di una pioggia di meteore:
- Se il tasso λ è 5 meteore all'ora, quanto tempo dobbiamo aspettare mediamente per vedere la prima meteora?

$$E[X] = \frac{1}{\lambda} = \frac{1}{5} = 0.2 \text{ ore}(12 \text{ minuti})$$

Properties of common probability distributions.

RV	PMF/PDF	Range	Mean	Variance	R Root
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$	$0 \le x \le n$	np	$np(1-p)$	`binom`
Geometric	$p(1-p)^x$	$x \ge 0$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$	`geom`
Poisson	$\frac{1}{x!} \lambda^x e^{-\lambda}$	$x \ge 0$	λ	λ	`pois`
Hypergeometric	$\frac{\binom{m}{x}\binom{n}{k-x}}{\binom{m+n}{k}}$	$x = 0, \ldots, k$	kp	$kp(1-p)\frac{m+n-k}{m+n-1}$	`hyper`
Negative Binomial	$\binom{x+n-1}{x} p^n (1-p)^x$	$x \ge 0$	$n\frac{1-p}{p}$	$n\frac{1-p}{p^2}$	`nbinom`
Uniform	$\frac{1}{b-a}$	$a \le x \le b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	`unif`
Exponential	$\lambda e^{-\lambda x}$	$x \ge 0$	$1/\lambda$	$1/\lambda^2$	`exp`
Normal	$\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$	$-\infty < x < \infty$	μ	σ^2	`norm`

Stima delle probabilità

La simulazione permette di stimare la probabilità di eventi legati a variabili casuali attraverso tre passi fondamentali:

1. **Campionamento**: si generano valori casuali dalla variabile distribuita secondo una data legge probabilistica (usando funzioni come `rnorm` per la normale, `runif` per l'uniforme, ecc.).
2. **Valutazione dell'evento**: si calcola un vettore di valori TRUE/FALSE che indica se l'evento è verificato per ogni valore campionato.
3. **Calcolo della proporzione**: si calcola la media dei valori TRUE per stimare la probabilità.
   Esempio:

- Per stimare $P(Z > 1)$, dove Z è normale standard $(N, (0, 1)$, media 0 e deviazione standard $1)$:

```
Z <- rnorm(10000)       # Genera 10.000 campioni
mean(Z > 1)            # Stima la probabilità
```

```
Risultato: circa 0.1588.
```

- Per $P(Z^2 > 1)$, si usa lo stesso principio:

```
Z <- rnorm(10000)
mean(Z^2 > 1)
```

```
Risultato: circa 0.322.
```

Stima di distribuzioni discrete

Per stimare la funzione di massa di probabilità (**pmf**) di una variabile discreta, si utilizza un approccio simile:

1. Si genera un grande campione.
2. Si calcola la frequenza di ogni valore.
3. Si normalizzano le frequenze per ottenere le probabilità.

Esempio: Somma di due dadi

- Se lanciamo due dadi e sommiamo i risultati, la variabile XX rappresenta la somma. La pmf si stima con:

```
X <- replicate(10000, { sum(sample(1:6, 2, replace = TRUE)) })
proportions(table(X))  # Calcola la pmf
```

- La probabilità stimata che la somma sia 11 è 0.0562.
  È possibile visualizzare la pmf con un grafico:

```
plot(proportions(table(X)), main = "Somma di due dadi", ylab = "Probabilità")
```

Teorema del ballottaggio

Il *teorema del ballottaggio* (o *teorema di Bertrand*) è un risultato importante nella teoria della probabilità che riguarda il conteggio delle preferenze in un'elezione. Il teorema afferma che, dato un numero di voti a favore di due candidati (dove uno dei due ha più voti dell'altro), la probabilità che il candidato con il numero maggiore di voti sia sempre avanti durante il conteggio dei voti è determinata dalla differenza tra i voti. In altre parole, se uno dei due candidati ha pp voti e l'altro ha q voti, con $p > q$, la probabilità che il candidato con p voti sia sempre in vantaggio durante il conteggio è data da:

$$\frac{p-q}{p+q}$$

Questo teorema è stato formulato da Joseph Bertrand nel 1887, ed è utile per analizzare il comportamento delle elezioni, soprattutto in contesti in cui il conteggio dei voti avviene gradualmente. Bertrand e altri matematici, come Désiré André, hanno fornito diverse dimostrazioni di questo teorema, tra cui la riflessione e l'induzione matematica. Il teorema assume che ogni voto sia indipendente e che i voti siano distribuiti in modo casuale, senza influenze esterne.