Dipartimento di Ingegneria dell'Impresa
Corso di Informatica
Inferenza statistica e teoria dell'informazione

Tesina di inferenza statistica e teoria dell'informazione

Preparato da: Federica Magliocca e Teresa Alomba

Abstract

In questa tesina analizzeremo un problema probabilistico e sfrutteremo R, un ambiente per la gestione, l'analisi dei dati e la visualizzazione di grafici, per verificare i risultati ottenuti.
Prima di procedere illustreremo argomenti che ci saranno utili per la risoluzione del nostro problema.
Daremo alcuni cenni di teoria della probabillità, definendo, in particolare, la funzione di probabilità e la funzione di ripartizione; parleremo della distribuzione continua uniforme e spiegheremo come si trasforma una variabile aleatoria.
Infine descriveremo il problema, mostreremo le simulazioni fatte e trarremo le nostre conclusioni.

Contents

1 Cenni di probabilità

1.1 Distribuzione di probabilità

Una distribuzione di probabilità è un modello matematico che collega i valori di una variabile alle probabilità che tali valori possano essere osservati.
Le distribuzioni di probabilità vengono utilizzate per modellizzare il comportamento di un fenomeno di interesse in relazione alla popolazione di riferimento, ovvero alla totalità dei casi di cui lo sperimentatore osserva un dato campione.
In questo contesto la variabile di interesse è vista come una **variabile casuale** (o **variabile aleatoria**) la cui legge di probabilità esprime il grado di incertezza con cui i suoi valori possono essere osservati.
In base alla scala di misura della variabile di interesse **X**, possiamo distinguere due tipi di distribuzioni di probabilità:

- **distribuzioni continue**: la variabile viene espressa su un scala continua;
- **distribuzioni discrete**: la variabile viene misurata con valori numerici interi.

Formalmente, le distribuzioni di probabilità vengono espresse da una legge matematica detta **funzione di densità di probabilità** (indicata con $f(x)$) o **funzione di probabilità** (indicata con $p(x)$) rispettivamente per le distribuzioni continue o discrete.

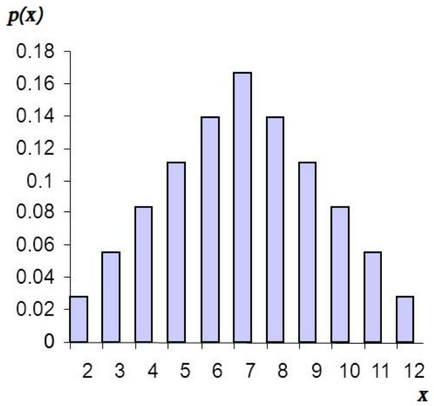

Figure 1.1: Esempio di distribuzione di probabilità discreta

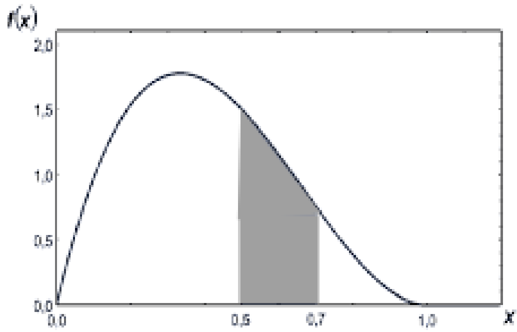

Figure 1.2: Esempio di distribuzione di probabilità continua

1.2 Funzione di ripartizione

Data una variabile casuale X, la funzione di ripartizione di X è la funzione che associa a ciascun valore di x ila probabilità che X assuma valori minori o uguali ad x ed è così definita:

$$F : \mathbb{R} \to [0, 1] \quad F(x) = P(X < x)$$

Una funzione F è una valida funzione di ripartizione se è non decrescente, continua a destra e

$$F(x) \geq 0 \ \forall x$$
$$\lim_{x \to +\infty} F(x) = 1$$
$$\lim_{x \to -\infty} F(x) = 0$$

Se X è una variabile casuale assolutamente continua la funzione di ripartizione di X può essere espressa come funzione integrale:

$$F(x) = \int_{-\infty}^{x} f(u)du$$

dove f è la funzione di densità di X. Se X è una variabile casuale discreta

$$F(x) = \sum_{x_i \leq x} p(x_i)$$

dove $p(x) = P(X = x)$ è la funzione di probabilità di X.

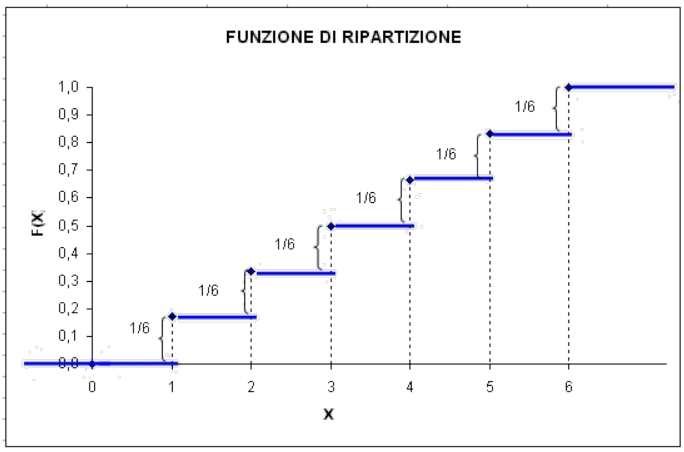

Figure 1.3: Funzione di ripartizione di una v.a. X che rappresenta il lancio di un dado a 6 facce

2 Distribuzione uniforme

2.1 Discreta

La **distribuzione discreta uniforme** su un insieme finito S è la distribuzione di probabilità $\boldsymbol{U(S)}$ che attribuisce a tutti gli elementi di S la stessa probabilità p di verificarsi.

In particolare:

$$1 = P(S) = \sum_{s \in S} P(s) = \sum_{s \in S} p = p|S| \tag{1}$$

Quindi,

$$P(s) = \frac{1}{|S|} \tag{2}$$

2.2 Continua

In teoria della probabilità la **distribuzione continua uniforme** è una distribuzione di probabilità continua che attribuisce la stessa probabilità a tutti i punti appartenenti ad un dato intervallo $[\boldsymbol{a,b}]$ contenuto nell'insieme.
Tale distribuzione viene solitamente indicata con $\boldsymbol{U(a,b)}$.
La sua densità di probabilità è:

$$f(x) = \frac{1}{b-a} \; su \; [a,b] \tag{3}$$

la sua funzione di ripartizione è:

$$F(x) = \frac{x-a}{b-a} \; per \; x \in [a,b]; \tag{4}$$

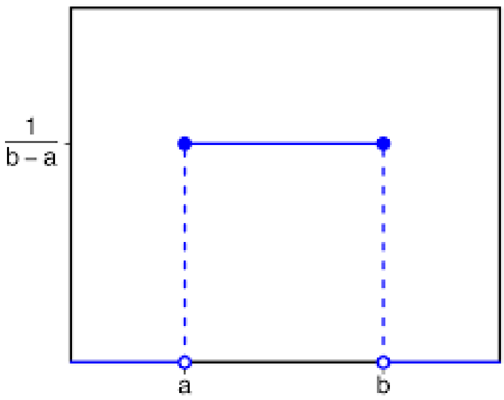

Figure 2.1: Distribuzione Continua Uniforme

3 Trasformazione di una variabile aleatoria

Introduciamo un esperimento casuale su un certo spazio campionario e con misura di probabilità P.
Supponiamo di avere una variabile casuale X relativa all'esperimento, ed una seconda Y da essa derivata per mezzo della relazione $y = f(x)$.
Nel caso in cui $f(x)$ sia monotona non decrescente, e indicando con $x = g(y)$ la corrispettiva funzione inversa, la caratterizzazione statistica di Y nei termini della sua funzione di densità di probabilità $p_Y(y)$ può essere ottenuta a partire da quella di X, nei termini della funzione di distribuzione di Y, come:

$$F_Y(y) = P(Y \leq y) = P(X \leq g(y))$$

e calcolando poi $p_Y(y) = \frac{dF_Y(y)}{dy}$.
Se invece la $f(x)$ è monotona ma decrescente, consideriamo semplicemente che le probabilità

$$P(x \leq X \leq x + dx) = p_X(x)dx \quad e \quad P(y \leq Y \leq y + dy)|_{y=f(x)} = p_Y(y)dy$$

devono essere uguali, ma dato che con $f(x)$ decrescente ad un dx positivo corrisponde un dy negativo, prendiamo il valore assoluto di entrambi:

$$p_X(x)|dx| = p_Y(y)|dy$$

e sostituiamo x con la sua funzione inversa $g(y)$.

$$p_Y(y) = p_X(g(y))\frac{|dg(y)|}{|dy|}$$

D'altra parte, se la trasformazione $f(x)$ non è monotona due o più valori di X producono lo stesso valore di Y, e non esiste una funzione inversa $x = g(y)$ univoca.
In tal caso si suddivide la variabilità di X in più intervalli i, in modo che per ciascuno di essi possa definirsi una $f_i(x)$ monotona, e dunque si possa calcolare il lato destro di ogni funzione inversa $g_i(y) = f_i^{-1}(x)$ e sommare i risultati per ottenere $p_Y(y)$.

4 Problema

Nel quadrato unitario si traccia una retta passante per il punto ($\frac{1}{2}$,$\frac{1}{2}$) che fa angolo ϕ tra 0 e π radianti (in senso antiorario) con l'asse dell'ordinata (x).
Questo angolo e' scelto secondo la distribuzione $U(0,\ \pi)$.
Sia L la lunghezza del segmento di questa retta che sta all'interno del quadrato unitario.
Trovare distribuzione e media di L.

Esistono un numero infinito di rette passanti per il punto ($\frac{1}{2}$,$\frac{1}{2}$) nel quadrato unitario, quindi la variabile aleatoria L è definita nel continuo.
Quindi trovare la distribuzione di L significa definire la funzione di densità di probabilità di L.
Come si vede in Figura 4.1 la simmetria del quadrato ci permette di ridurre la scelta dell'angolo ϕ tra 0 e $\frac{\pi}{4}$, poichè tutte le lunghezze (che variano tra 1 e $\sqrt{2}$) si ripetono negli intervalli successivi ($[\frac{\pi}{4}$,$\frac{\pi}{2}]$,$[\frac{\pi}{2}$,$\frac{3\pi}{4}]$,$[\frac{3\pi}{4}$,$\pi]$).

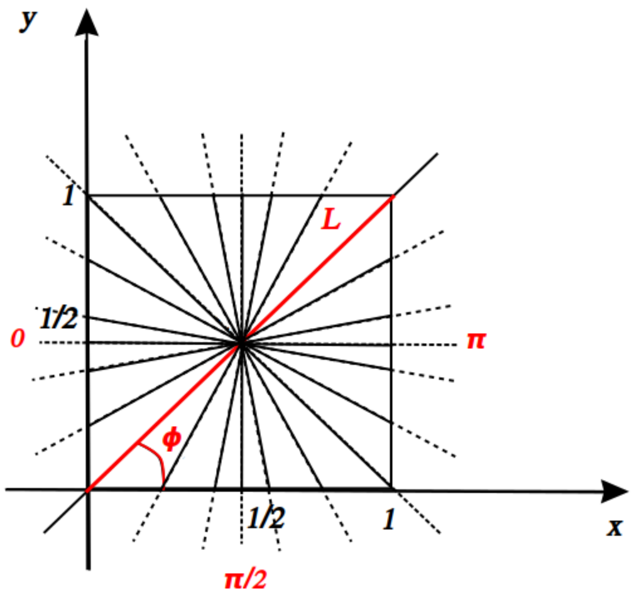

Figure 4.1: Rappresentazione geometrica del problema preso in esame

Si noti inoltre che la lunghezza del segmento all'interno del quadrato unitario è in realtà la retta secante dell'angolo ϕ, rappresentata in blu in Figura 4.2.
Dove

$$sec(\phi) = \frac{1}{cos(\phi)} \tag{5}$$

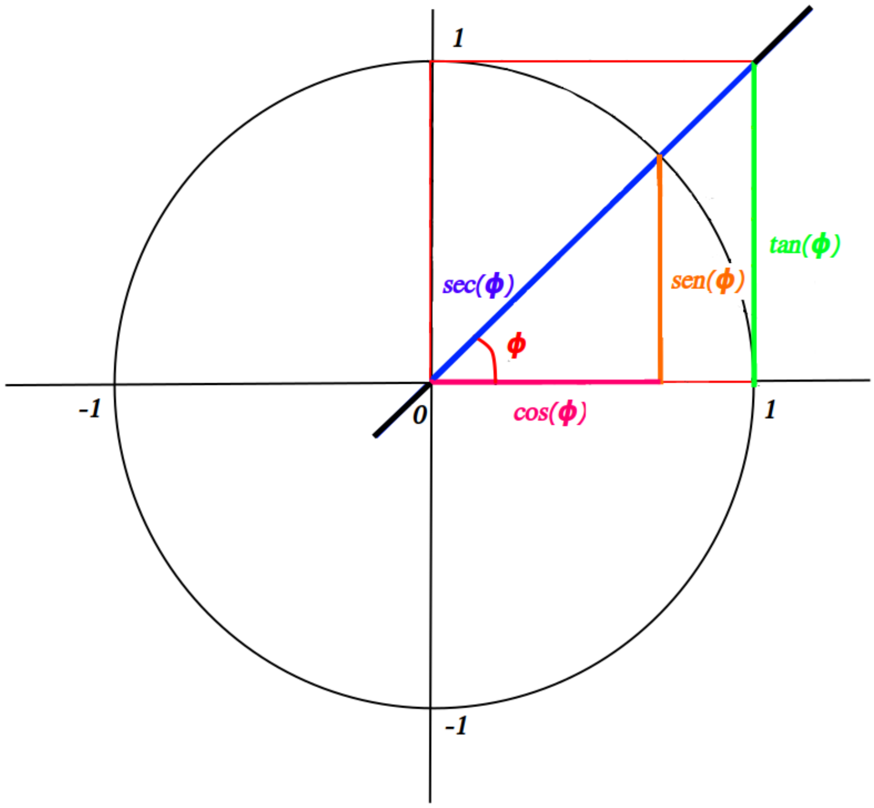

Figure 4.2: Geometricamente, la secante può essere vista anche come l'ipotenusa del triangolo rettangolo avente come cateti il raggio della circonferenza unitaria e la tangente dell'angolo

Sappiamo quindi che $L = sec(\Phi)$ con $\Phi \sim U(0, \frac{\pi}{4})$.
Si tratta di una trasformazione di una variabile aleatoria, dove $sec(\phi)$ con $0 \leq \phi \leq \frac{\pi}{4}$ è una funzione monotona non decrescente.
La funzione di ripartizione di Φ è:

$$F_\Phi(\phi) = \begin{cases} 0 & se\ \phi < 0 \\ \frac{\phi}{\pi/4} & se\ 0 \leq \phi < \frac{\pi}{4} \\ 1 & se\ \phi \geq \frac{\pi}{4} \end{cases}$$

Calcoliamo $F_L(l)$:

$$F_L(l) = P(L \leq l) = P(sec(\Phi) \leq l) = P(\Phi \leq arcsec(l)) = F_\Phi(arcsec(l)) = \frac{arcsec(l)}{\pi/4}$$

Quindi la funzione di ripartizione di L sarà:

$$F_L(l) = \begin{cases} 0 & se\ l < 1 \\ \frac{arcsec(l)}{\pi/4} & se\ 1 \leq l < \sqrt{2} \\ 1 & se\ l \geq \sqrt{2} \end{cases}$$

Calcoliamo la derivata di $F_L(l)$ per trovare la funzione di densità di probabilità:

$$f_L(l) = \frac{-\left(-\frac{4}{l^2\sqrt{1-\frac{1}{l^2}}}\right)\pi}{\pi^2} = \frac{4}{\pi l^2\sqrt{1-\frac{1}{l^2}}}$$

Calcoliamo il valore atteso $E[L]$:

$$E[L] = \int l f_L(l)dl = \int_1^{\sqrt(2)} l\frac{4}{\pi l^2\sqrt{1-\frac{1}{l^2}}}dl = \int_1^{\sqrt(2)} \frac{4}{\pi l\sqrt{1-\frac{1}{l^2}}}dl$$

$$= \frac{2\ln\left(\sqrt{2}+1\right)}{\pi} - \frac{2\ln\left(\sqrt{2}-1\right)}{\pi} = 1.12219$$

4.1 Simulazione in R

Calcoliamo la lunghezza del segmento con $l = 1/cos(\phi)$ con ϕ scelto in modo uniforme tra 0 e $\frac{\pi}{4}$ (con la funzione $runif()$). Il programma esegue 1000000 simulazioni e ne calcola la media. In questo modo otteniamo una stima molto precisa, infatti vediamo che il valore ottenuto (in basso in Figura 4.3) è molto vicino al valore atteso calcolato in precedenza.

```
ripetizioni = 1000000;
Simulazione <- function(){
  lunghezze <- list();
  u <- runif(ripetizioni, min = 0, max = 45);
  for (i in 1:ripetizioni){
    l <- 1/cospi(((pi*u[i])/180)/pi);
    lunghezze[length(lunghezze)+1] = l;
  }
  return(lunghezze);
}

Media <- function(){
  l <- Simulazione();
  sum <- 0;

  for (i in 1:ripetizioni){
    sum <- sum + as.numeric(l[i]);
  }
  media = sum/ripetizioni;
  print(media);
}
Media()
```

```
## [1] 1.122358
```

Figure 4.3: Simulazione della media della lunghezza del segmento all'interno del quadrato unitario della retta passante per il punto $(\frac{1}{2},\frac{1}{2})$ che interseca l'asse delle x con angolo ϕ scelto uniformemente a caso tra 0 e $\frac{\pi}{4}$