



UNIVERSITÀ DI PERUGIA
Dipartimento di Matematica e Informatica



Appunti *Simulazione*

Formulario

Anno Accademico 2021-2022

Last Update: January 18, 2023

Contents

| | | |
|----------|---|----------|
| 1 | Distribuzioni | 4 |
| 1.1 | Stimare la Distribuzione | 4 |
| 1.2 | Calcolare la Probabilità di una Distribuzione | 5 |
| 1.2.1 | Esponenziale | 5 |
| 1.2.1.1 | Positiva | 5 |
| 1.2.1.2 | Negativa | 5 |
| 1.2.2 | Poisson | 5 |
| 1.2.3 | Geometrica | 5 |
| 2 | Goodness of Fit | 6 |
| 2.1 | Test χ^2 | 6 |
| 2.1.1 | Dati senza Intervalli | 6 |
| 2.1.2 | Dati con Intervalli | 7 |
| 2.2 | Test Kolmogorov | 9 |
| 2.2.1 | Dati Senza Intervalli | 9 |
| 2.2.2 | Dati Con Intervalli | 10 |
| 2.3 | Informazioni utili su Formule | 11 |
| 2.3.1 | Komorov | 11 |
| 2.3.1.1 | <i>cumsum</i> | 11 |



*”Oi, con quanto sentimento
defeco sul tuo naso,
così che ti coli sul mento.”*

Wolfgang Amadeus Mozart

Chapter 1

Distribuzioni

1.1 Stimare la Distribuzione

Per stimare una distribuzione avendo solo i dati iniziali del problema effettua le seguenti operazioni:

1. $n = \sum f_i$: assicurati di aver calcolato la somma totale delle osservazioni
2. Calcola la **Media**:
 - (a) Aggiungi *Colonna Totale*: $\text{categoria}_i * f_i$
 - (b) Calcola la media effettiva con: $\text{media} = \frac{\sum \text{totale}}{n}$
3. Calcola la **Varianza** σ^2 :
 - (a) Aggiungi *Colonna ris*: $(\text{categoria}_i - \text{media})^2 * f_i$
 - (b) Calcola la varianza effettiva con: $\sigma^2 = \frac{\sum \text{ris}}{n-1}$
4. Calcola la **Deviazione Standard** σ :
 - (a) $\sigma = \sqrt{\sigma^2}$
5. Calcola $V = \frac{\sigma}{\text{media}}$

Una volta completati tutti i calcoli controlla se il coefficiente V è vicino ad 1 e:

- se lo è allora utilizza l'**Esponenziale**,
- se non lo è è **Poissoniana** ma per avere una verifica, controlla che la media e la varianza siano uguali.



Note:

Si può avere una prima idea del tipo di distribuzione anche osservando le frequenze per categoria:

- Se le frequenze hanno valori alti per le prime categorie e poi decrescono, probabilmente è esponenziale negativa.
- Se le frequenze hanno valori bassi per le prime categorie e poi crescono, probabilmente è esponenziale...
- Se le frequenze hanno valori alti nelle categorie centrali e bassi verso le categorie agli estremi, probabilmente è Poissoniana
- Nel caso in cui sia geometrica solitamente viene esplicitato.

1.2 Calcolare la Probabilità di una Distribuzione

1.2.1 Esponenziale

1.2.1.1 Positiva

$$p(i) = \frac{e^{\frac{-\text{categoria}_i}{\text{media}}}}{\text{media}}$$

1.2.1.2 Negativa

$$p(i) = 1 - e^{\frac{-\text{categoria}_i}{\text{media}}}$$

1.2.2 Poisson

$$p(i) = \frac{e^{-\text{media}} * \text{media}^{\text{categoria}_i}}{\text{categoria}_i!}$$

1.2.3 Geometrica

$$p(i) = \rho * (1 - \rho)^{\text{categoria}_i}$$



Chapter 2

Goodness of Fit

2.1 Test χ^2

Devi utilizzare questa sezione solo se il numero delle **osservazioni totale** $n > 30$.

2.1.1 Dati senza Intervalli

Devi utilizzare questa sezione solo quando hai dei dati **Senza Intervalli**, devi anche fare attenzione che il **numero di osservazioni** $n > 30!!$

Operazioni da effettuare:

1. Riportare i dati in una tabella in Calc:
 - *Colonna 1: categorie*
 - *Colonna 2: f_i*
2. Raggruppare le categorie se $\exists \text{ categoria} < 5$:
 - Parti dall'ultimo a salire (dal basso verso l'alto delle categorie)
 - Raggruppare tutte nell'ultima categoria che le faccia diventare maggiori di 5 sommando le frequenze.
 - *Esempio:*



| | A | B | C | D | E |
|----|--------|-----------|------------------|---|---|
| 1 | VALORI | frequenze | f(i) raggruppate | | |
| 2 | 0 | 59 | 59 | | |
| 3 | 1 | 26 | 26 | | |
| 4 | 2 | 24 | 24 | | |
| 5 | 3 | 18 | 18 | | |
| 6 | 4 | 12 | 12 | | |
| 7 | 5 | 5 | 5 | | |
| 8 | 6 | 4 | 12 | | |
| 9 | 7 | 3 | | | |
| 10 | 9 | 3 | | | |
| 11 | 11 | 2 | | | |
| 12 | | | | | |

| | A | B | C | D |
|----|--------|-----------|------------------|---|
| 1 | VALORI | frequenze | f(i) raggruppate | |
| 2 | 0 | 59 | 59 | |
| 3 | 1 | 26 | 26 | |
| 4 | 2 | 24 | 24 | |
| 5 | 3 | 18 | 18 | |
| 6 | 4 | 12 | 12 | |
| 7 | 5 | 5 | 9 | |
| 8 | 6 | 1 | | |
| 9 | 7 | 1 | | |
| 10 | 9 | 1 | | |
| 11 | 11 | 1 | | |

3. Calcolare:

- $n = \sum(f_i)$
- $f(i) = f_i/n$: non serve
- Capire la distribuzione se non è data (vedi 1.1)
- $p(i)$: dipende dalla distribuzione (vedi 1.2)
- $F_i = n * p(i)$: numero di intervalli unitari teorici con i arrivi
- $G_i = \frac{(f_i - F_i)^2}{F_i}$
- $V = \sum G_i$: sommare tutti i valori di G
- $df = \text{Numero Categorie} - 1 - \text{Numero Parametri Distribuzione}$

Una volta terminati i calcoli devi guardare la riga nella tabella del χ^2 (AGGIUNGERE REF) con lo stesso valore di df : devi controllare che il valore V ricada negli intervalli che non superino il P_{95} .

2.1.2 Dati con Intervalli

Devi utilizzare questa sezione solo quando hai dei dati divisi in **Intervalli**, devi anche fare attenzione che il **numero di osservazioni** $n > 30$!!

Calcoli da effettuare:

- Riportare i dati in una tabella in Calc:
 - Colonna 1: categorie*, probabilmente devi aggiungerle tu, parti da 0 in poi
 - Colonna 2: intervallo*, del tipo $x_1 - x_2$. Fai sempre attenzione che $x_2 \geq x_1$!!! In caso li inverti.
 - Colonna 3: frequenza* f_i



2. Aggiungere *Colonna* x_1 (intervallo più piccolo)
3. Aggiungere *Colonna* x_2 (intervallo più grande)
4. Aggiungere *Colonna media-intervalli* tra x_2 e x_1
5. Calcolare:
 - (a) frequenze pesate = media-intervallo _{i} * f_i
 - (b) $n = \sum(f_i)$
 - (c) media = $\sum(\text{frequenza pesata}_i)/n$
 - (d) differenza medie = (media-intervalli _{i} - media)²
 - (e) frequenze pesate 2 = differenza media _{i} * f_i
 - (f) frequenze relative = $f_i * n$
 - (g) varianza $\sigma^2 = \sum(\text{frequenza pesata } 2_i)/n - 1$
 - (h) capire la distribuzione se non è data (vedi 1.1)
 - (i) ~~$f(i) = f_i/n$: non serve~~
 - (j) $p(i) = p(x_2) - p(x_1)$ = calcolare secondo la distribuzione (vedi 1.2)
 - (k) $F_i = n * p(i)$: numero di intervalli unitari teorici con i arrivi
 - (l) $G_i = \frac{(f_i - F_i)^2}{F_i}$
 - (m) $V = \sum G_i$: sommare tutti i valori di G
 - (n) df = Numero Categorie - 1 - Numero Parametri Distribuzione
6. Raggruppare le categorie se $\exists \text{ categoria} < 5$:
 - Parti dall'ultimo a salire (dal basso verso l'alto delle categorie)
 - Raggruppare tutte nell'ultima categoria che le faccia diventare maggiori di 5 sommando le frequenze.
 - *Esempio:*

| | A | B | C | D | E |
|----|--------|-----------|------------------|----|---|
| 1 | VALORI | frequenze | f(i) raggruppate | | |
| 2 | | 0 | 59 | 59 | |
| 3 | | 1 | 26 | 26 | |
| 4 | | 2 | 24 | 24 | |
| 5 | | 3 | 18 | 18 | |
| 6 | | 4 | 12 | 12 | |
| 7 | | 5 | 5 | 5 | |
| 8 | | 6 | 4 | 12 | |
| 9 | | 7 | 3 | | |
| 10 | | 9 | 3 | | |
| 11 | | 11 | 2 | | |
| 12 | | | | | |

| | A | B | C | D |
|----|--------|-----------|------------------|----|
| 1 | VALORI | frequenze | f(i) raggruppate | |
| 2 | | 0 | 59 | 59 |
| 3 | | 1 | 26 | 26 |
| 4 | | 2 | 24 | 24 |
| 5 | | 3 | 18 | 18 |
| 6 | | 4 | 12 | 12 |
| 7 | | 5 | 5 | 9 |
| 8 | | 6 | 1 | |
| 9 | | 7 | 1 | |
| 10 | | 9 | 1 | |
| 11 | | 11 | 1 | |



Una volta terminati i calcoli devi guardare la riga nella tabella del χ^2 (AGGIUNGERE REF) con lo stesso valore di df : devi controllare che il valore V ricada negli intervalli che non superino il P_{95} .

2.2 Test Kolmogorov

Devi utilizzare questa sezione solo se il numero delle **osservazioni totale** $n < 30$.

2.2.1 Dati Senza Intervalli

Devi utilizzare questa sezione solo quando hai dei dati **Senza Intervalli**, devi anche fare attenzione che il **numero di osservazioni totali** $n < 30$!!

Operazioni da effettuare:

1. Riportare i dati in una tabella in Calc:
 - *Colonna categorie*
 - *Colonna frequenze f_i*
2. Calcolare:
 - (a) $f(i) = f_i/n$: frequenze osservate
 - (b) Individuare la distribuzione di probabilità adatta (vedi 1.1)
 - (c) $p(i)$: probabilità teorica (vedi 1.2)
 - (d) $d_i = cumsum(f(i))$: somma cumulativa delle $f(i)$
 - (e) $D_i = cumsum(p(i))$: somma cumulativa delle $p(i)$
 - (f) $D = |d_i - D_i|$: la differenza assoluta
 - (g) $D_{max} = \max(D)$: il massimo valore tra le differenze assolute D

Una volta completati tutti i calcoli, cercare nella tabella di *Kolmogorov-Smirnov* la riga corrispondente al valore delle osservazioni totali n : se il valore D_{max} è sotto il $D_{0,10}$ la distribuzione è accettata, altrimenti no.



2.2.2 Dati Con Intervalli

Devi utilizzare questa sezione solo quando hai dei dati **Senza Intervalli**, devi anche fare attenzione che il **numero di osservazioni totali** $n < 30!!$

N.B.: *non abbiamo trovato esercizi con cui testare questa sezione !*

Operazioni da effettuare:

1. Riportare i dati in una tabella in Calc:
 - *Colonna categorie*: probabilmente devi aggiungerle tu, parti da 0 in poi
 - *Colonna intervallo*: del tipo $x_1 - x_2$. Fai sempre attenzione che $x_2 \geq x_1$!!! In caso li inverti.
 - *Colonna frequenze* f_i
2. Aggiungere *Colonna* x_1 (estremo più piccolo dell'intervallo)
3. Aggiungere *Colonna* x_2 (estremo più grande dell'intervallo)
4. Calcolare:
 - (a) $f(i) = f_i/n$: frequenze osservate
 - (b) Individuare la distribuzione di probabilità adatta (vedi 1.1)
 - (c) $p(i) = p(x_2) - p(x_1)$: probabilità teorica per ogni intervallo (vedi 1.2)
 - (d) $d_i = \text{cumsum}(f(i))$: somma cumulativa delle $f(i)$
 - (e) $D_i = \text{cumsum}(p(i))$: somma cumulativa delle $p(i)$
 - (f) $D = |d_i - D_i|$: la differenza assoluta
 - (g) $D_{max} = \max(D)$: il massimo valore tra le differenze assolute D

Una volta completati tutti i calcoli, cercare nella tabella di *Kolmogorov-Smirnov* la riga corrispondente al valore delle osservazioni totali n : se il valore D_{max} è sotto il $D_{0,10}$ la distribuzione è accettata, altrimenti no.



2.3 Informazioni utili su Formule

2.3.1 Komorov

2.3.1.1 *cumsum*

Per calcolare *cumsum* (somma cumulativa) va eseguito il seguente procedimento:

- La prima cella resta uguale alla prima cella della colonna di riferimento (es. $f(i)$ o $p(i)$)
- Dalla seconda cella in poi si blocca la prima cella della somma cumulativa (quella calcolata al punto precedente) e si somma fino alla cella i di riferimento (vedi Figura 2.1)

| f_i | =C2 | | | | | | |
|-----------|------|--------|-----|-------|----------|------|--|
| B | C | D | E | F | G | | |
| FREQUENZE | f(i) | totale | ris | p(i) | d_i | | |
| 0 | 3 | 0,15 | 0 | 10,83 | 0,149569 | 0,15 | |
| 1 | 6 | 0,3 | 6 | 4,86 | 0,28418 | 0,45 | |
| 2 | 5 | 0,25 | 10 | 0,05 | 0,269971 | 0,7 | |
| 3 | 3 | 0,15 | 9 | 3,63 | 0,170982 | 0,85 | |
| 4 | 2 | 0,1 | 8 | 8,82 | 0,081216 | 0,95 | |
| 5 | 1 | 0,05 | 5 | 9,61 | 0,030862 | 1 | |

| =SOMMA(\$C\$2:C3) | | | | | | | |
|-------------------|------|--------|-------|----------|------|--|--|
| B | C | D | E | F | G | | |
| FREQUENZE | f(i) | totale | ris | p(i) | d_i | | |
| 3 | 0,15 | 0 | 10,83 | 0,149569 | 0,15 | | |
| 6 | 0,3 | 6 | 4,86 | 0,28418 | 0,45 | | |
| 5 | 0,25 | 10 | 0,05 | 0,269971 | 0,7 | | |
| 3 | 0,15 | 9 | 3,63 | 0,170982 | 0,85 | | |
| 2 | 0,1 | 8 | 8,82 | 0,081216 | 0,95 | | |
| 1 | 0,05 | 5 | 9,61 | 0,030862 | 1 | | |

Figure 2.1: Esempio di calcolo della funzione *cumsum*