Laboratorio di Informatica

Dr Carlo Meneghini

Dip. di Fisica "E. Amaldi"
via della Vasca Navale 84
st. – 83 – I piano

meneghini@fis.uniroma3.it

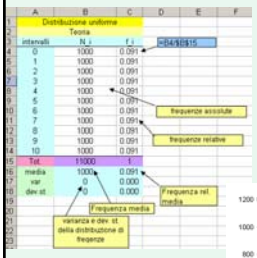
tel.: 06 55177217

http://www.fis.uniroma3.it/~meneghini

Esercitazione Vb

popolazioni, campionamento, inferenza statistica

Popolazione: un'urna contiene 11 000 palline numerate da 0 a 11, in tutto 1000, palline per ogni valore.

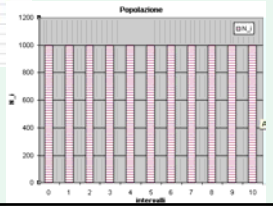


La tabella mostra le caratteristiche della **popolazione**:

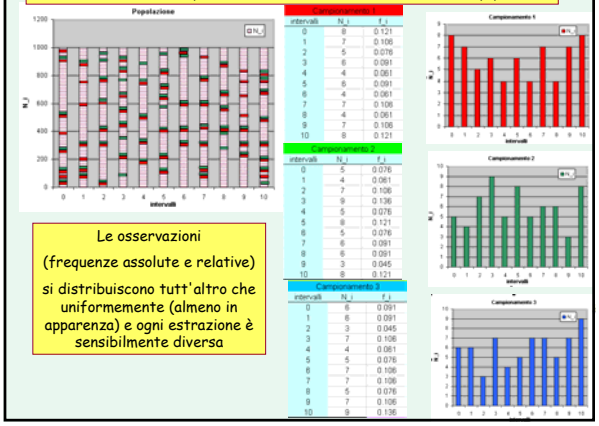
a) ogni valore compare 1000 volte (frequenza assoluta)

b) la frequenza relativa di ogni valore è 0.091

Problema: Cosa mi posso aspettare estraendo a caso M palline dall'urna?



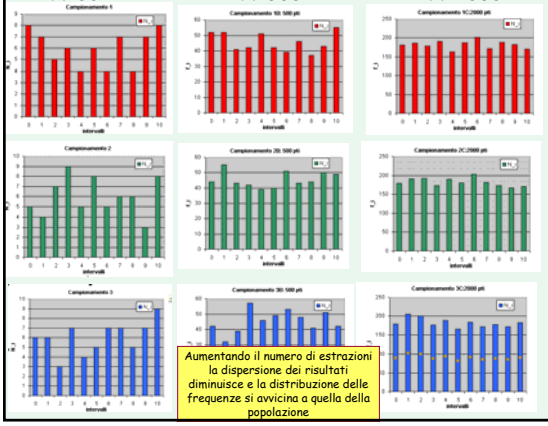
In rosso e verde sono due possibili estrazioni casuali di **66** elementi dalla popolazione



Le osservazioni (frequenze assolute e relative) si distribuiscono tutt'altro che uniformemente (almeno in apparenza) e ogni estrazione è sensibilmente diversa

---

N=66 N=500 N=2000



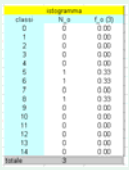
Aumentando il numero di estrazioni la dispersione dei risultati diminuisce e la distribuzione delle frequenze si avvicina a quella della popolazione

---

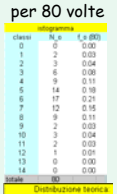
Il numero di osservazioni di un dato valore si distribuisce seguendo la distribuzione di Poisson!

N=66 estrazioni

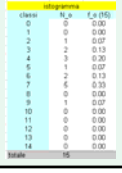
per 3 volte

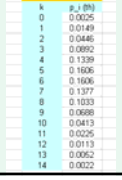


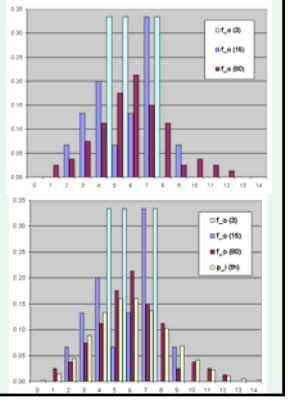
per 80 volte



per 15 volte



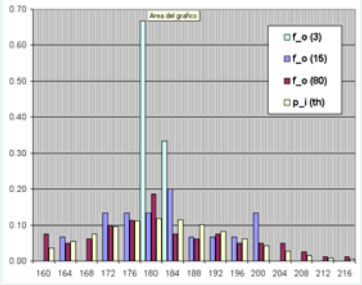




Il numero di osservazioni di un dato valore si distribuisce seguendo la distribuzione di Poisson!

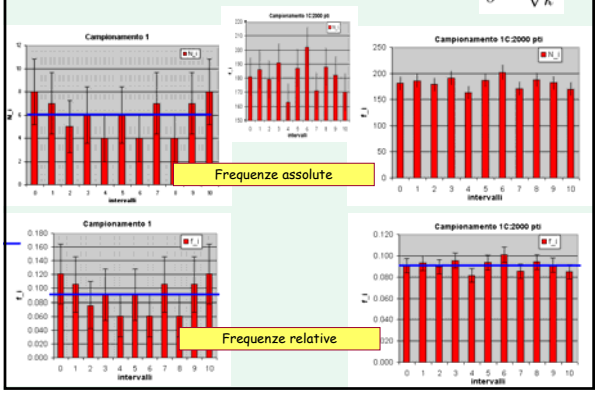
N=2000 estrazioni

Ripetendo l'esperimento di estrazione di 2000 valori per: 3, 15 e 80 volte la distribuzione dei successi si avvicina sempre più alla sistribuzione di Poisson



Poisson

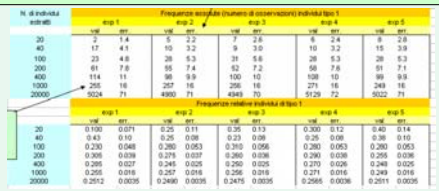
$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad \lambda = \bar{k} \quad \sigma^2 = \lambda \qquad \frac{\bar{k}}{\sigma} = \frac{1}{\sqrt{\bar{k}}}$$



Il file popolazione.dat contiene i dati di nua popolazione di individui di tipo 1, 2, 3, 4

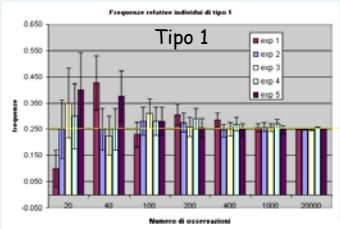
Si vuole conoscere la frazione di individui di un tipo dato

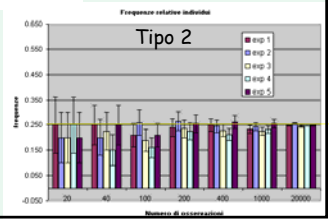
Gli esperimenti consistono nello stimare la frazione di individui di tipo dato prelevando campioni di diversa numerosità dalla popolazione



Confrontare tra loro i risultati di campionamenti con la stessa numerosità e con numerosità differenti

Per una possibile soluzione: **popolazione_soluzione.xls**

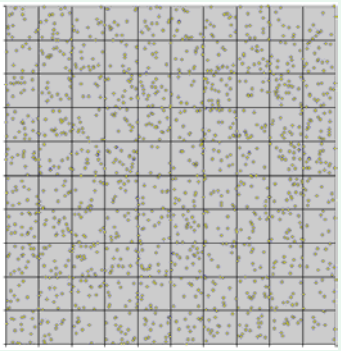




Risultati per esperimenti effettuati con un numero crescente di campionamenti, ripetuti 5 volte.

Problema: in una data regione sono distribuiti N esemplari di una specie di piante A.

Determinare la densità di piante: D



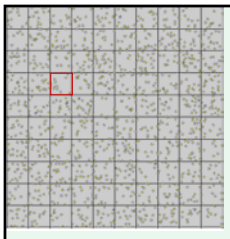
<u>Soluzione esatta</u>
- Misuro la superficie: S
- Conto tutte le piante in S: N
- calcolo: D=N/S

Lungo, costoso, a volte impossibile (es. densità di globuli rossi nel sangue!!!)

<u>Approssimata</u>
- Scelgo una superficie: s
- Conto le piante in s: n
- calcolo: D ~ n/s

Come scegliere s? Di quanto mi posso sbagliare? Che confidenza dare al risultato?

Statistica inferenziale



<u>Approssimata</u>
- Scelgo una superficie: s
- Conto le piante in s: n
- calcolo: D ~ n/s

Problemi:
Come scegliere s?
Di quanto mi posso sbagliare nella stima di D?
Come ottimizzare il rapporto Costi/benefici?

Statistica inferenziale

Assumendo una distribuzione di Poisson per i conteggi in ogni areola s, l'errore associato ad una misura n è $n^{1/2}$, quindi:

$$D = \frac{n}{s} \qquad \sigma_D = \frac{\sqrt{n}}{s} \qquad \epsilon = \frac{\sigma_D}{D} = \frac{1}{\sqrt{n}}$$

Densità stimata $= D \pm \frac{D}{\sqrt{n}} = D\left(1 \pm \frac{1}{\sqrt{n}}\right)$

Su quale area campionare in modo da avere un errore relativo ε% (es. ε=2%)

Densità stimata

$$= D \pm \frac{D}{\sqrt{n}} = D\left(1 \pm \frac{1}{\sqrt{n}}\right)$$

err. relativo

$$\epsilon = \frac{\sigma_D}{D} = \frac{1}{\sqrt{n}}$$

Per ottenere un errore del ε=5% deve essere: n=400

1) effettuo una prima misura n_o su una superficie s_o in modo da stimare D_o

2) Usando D_o posso stimare la superficie necessaria per ottenere n = 400 conteggi:

$s_1 \sim n/D_o$

3) Effettuo una misura sulla superficie s_1

err. relativo

$$\epsilon = \frac{\sigma_D}{D} = \frac{1}{\sqrt{n}}$$

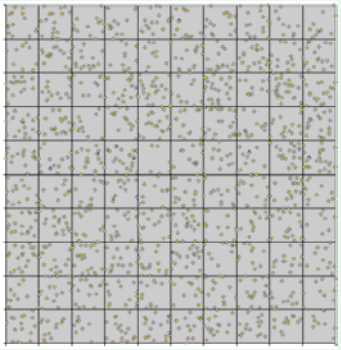
effettuo una prima misura n_o su una superficie s_o ottenendo un errore relativo ε_o

$$\epsilon_o = \frac{1}{\sqrt{n_o}} = \frac{1}{\sqrt{D_o s_o}}$$

essendo:

$$\frac{\epsilon_o}{\epsilon} = \frac{\sqrt{D_o s}}{\sqrt{D_o s_o}}$$

La superficie su cui effettuare il campionamento è:

$$s = \left(\frac{\epsilon_o}{\epsilon}\right)^2 s_o$$

Esercizio: stimare la densità di individui nella griglia in figura con un errore del 10%