# Corso Integrato di Statistica Informatica e Analisi dei Dati Sperimentali

A.A 2009-2010

## Esercitazione D

# 1 Introduzione

**Scopo dell'esercitazione**
Calcolo di Probabilit e Confronto qualitativo tra dati sperimentali e distribuzioni teoriche. Rappresentazione dei dati con la loro incertezza.

## 1.1 Uso di EXCEL e Gnuplot

Alla pag. web del corso, nella sezione *Guide rapide:Excel* ci sono esempi di funzioni predefinite da utilizzare con Excel o Open office per il calcolo di funzioni statistiche.
Il file **Es.10.funzioni_utili.plt** nella sezione **Dati Esercizi In aula**
Excel ha alcune funzioni predefinite utili per il calcolo delle probabilitá, usando Gnuplot queste devono essere definite.

**Istogrammi di frequenza**
Per costruire un istogramma di frequenze bisogna calcolare la probabilitá di osservare un valore della variabile X in un dato intervallo (classe):

$$P(x_a < X \leq x_b) = F(x_b) - F(x_a)$$

Nel caso di distribuzioni discrete, se le classi contengono un solo valore, si calcola direttamente le $P(x_a)$. Se le classi contengono piú valori si devono sommare le probabilitá associate adi diversi valori inclusi nella classe. Questo puó essere fatto sommando esplicitamente le probabilitá oppure come differenza tra le distribuzioni integrate calcolate per i valori estremi della classe. Quindi per una distribuzione binomiale:

$$P(x_a < X \leq x_b) = \mathrm{DISTRIB.BINOM(x_b, N, p, 1) - DISTRIB.BINOM(x_a, N, p, 1)}$$

E per una distribuzione di Poisson:

$$P(x_a < X \leq x_b) = \mathrm{Poisson(x_b, \lambda, 1) - DISTRIB.BINOM(x_a, \lambda, 1)}$$

Per una distribuzione di probabilitá di una variabile continua bisogna sempre utilizzare le funzioni di distribuzione cumulative. Quindi:

$$P(x_a < X \leq x_b) = \int_{-x_a}^{x_b} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

ovvero:

$$P(x_a < x \leq x_b) = \mathrm{DISTRIB.NORM(x_b, \mu, \sigma, 1) - DISTRIB.NORM(x_a, \mu, \sigma, 1)}$$

In Gnuplot é definita la funzione: `norm(x)` che calcola la funzione di distribuzione integrata per una variabile che segue una distribuzione normale standard ($\mu = 0$, $\sigma = 1$), ovvero:

$$norm(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = F(x)$$

Per il calcolo: si puó definire in Gnuplot una funzione `PNin`:

```
PNin(z1,z2) = norm(z2) - norm(z1)
```

Il comando:

```
 pr PNin(z1,z2)
```

calcola la probabilitá che una variabile aleatoria $Z$ (che segue una distribuzione normale standard) assuma un valore compreso tra $z_1$ e $z_2$: $PNin(z1, z2) = P(z_1 \leq Z \leq z_2)$

Se X é una variabile aleatoria che segue una distribuzione Normale con valore atteso $\mu$ e deviazione standard $\sigma$ si puó calcolare la probabilitá di osservare un valore di X nell'intervallo a-b effettuando un cambio di variabili: la variabile

$$z = \frac{x - \mu}{\sigma}$$

segue una distribuzione normale standard, quindi si puó definire la funzione z in Gnuplot:

```
Z(x,m,s) = (x-m)/s
```

La funzione:

```
Pin(x1,x2,m,s) = norm(Z(x2,m,s)) - norm(Z(x1,m,s))
```

calcola la probabilitá di osservare un valore della X tra $x_1$ e $x_2$: $P(x_1 \leq X \leq x_2)$
La funzione:

```
Pmin(x1,m,s) = norm(Z(x1,m,s))
```

calcola la probabilitá di osservare un valore della X minore di $x_1$: $P(X \leq x_1)$
La funzione:

```
Pmag(x2,m,s) = 1-norm(Z(x2,m,s))
```

calcola la probabilitá di osservare un valore della X maggiore di $x_2$: $P(x_2 \leq X)$

# 2 Esercizi ed esperimenti pratici proposti

Risolvere gli esercizi e inviarli, insieme ad una breve relazione, all'indirizzo: **ci.biologia@gmail.com** specificando l'esercizio svolto e gli studenti coinvolti.

## 2.1 Esercizi

**Esercizio D1**
**Scopo**: calcolo di probabilitá per variabili che seguono la distribuzione binomiale.

A. Sapendo che la probabilitá di osservare un individuo con il carattere A nella popolazione é p=0.15 calcolare:
- La probabilitá di osservare almeno 3 individui con il carattere A in un campione di 20 individui. (R. 0.595)
- La probabilitá di non osservare nessun individuo on il carattere A su un campione di 10 individui (R. 0.80)
- La probabilitá di osservare almeno 10 individui che non presentano il carattere A in un campione di 15 individui (R. 0.98)

B. Calcolare e riportare su un grafico gli istogrammi di frequenza per distribuzioni di probabilitá binomiale con:
- N=10, p=0.5
- N=12, p=0.8
- N=8, p=0.2
- N=18, p=0.2

Utilizzare come esempio i fogli Excel `Distribuzioni.xls`. Provare ad utilizzare la funzioni di Gnuplot.
C. La probabilit che un insetto di una data specie abbia il carattere A : p=0.6 - Quale la probabilit che su tre insetti scelti a caso due abbiano il carattere A? ($P(k = 2|3, 0.6) = 0.432$) - Quale la probabilit che su tre insetti scelti a caso almeno due abbiano il carattere A? ($P(k \geq 2|3, 0.6) = 0.712$)
Selezionando 12 triplette di insetti, quale la probabilit che: - almeno 7 abbiano almeno 2 insetti con la caratteristica A? (P=0.9) - si osservino meno di 4 casi in cui solo due insetti hanno la caratteristica A? (P=0.16)

**Esercizio D2**
**Scopo**: calcolo di probabilitá per variabili che seguono la distribuzione di Poisson.

A. Osservando un rubinetto che perde si osservano in media 4.5 gocce al minuto. Calcolare:
- La probabilitá di osservare almeno 2 gocce in un minuto (R. 0.94)
- La probabilitá di osservare un numero x di gocce compreso nell'intervallo $\lambda - \sigma \leq x \leq \lambda + \sigma$ intorno al valore atteso. (R. 0.66)
- La probabilitá di osservare un numero x di gocce compreso nell'intervallo $\lambda - 2\sigma \leq x \leq \lambda + 2\sigma$ intorno al valore atteso. Verificare la validitá del teorema di Chebyshev (R. 0.95, P(Ch)> 0.75)

B. In una data regione popolazione di piante della specie rara A é di 600 esemplari per km$^2$.
- Quale é la probablitá di osservare 5 esemplari in un quadrato di 100 m di lato? (R. 0.16).
- Quale é la probablitá di osservare meno di 3 esemplari in un quadrato di 100 m di lato? (R. 0.062).
- Quale é la probablitá di osservare almeno 6 esemplari in un quadrato di 100 m di lato? (R. 0.39).
C. Calcolare e riportare su un grafico gli istogrammi di distribuzioni di probabilitá di Poisson con: $\lambda$=3; $\lambda$=6; $\lambda$=9; $\lambda$=13
Utilizzare come esempio i fogli Excel `Distribuzioni.xls`.

**Esercizio D3**
A. Viene selezionato un campione di fili metallici di lunghezza compresa tra 2.5 e 7.5 mm con distribuzione uniforme. Calcolare la probabilitá di estrarre un filo di lunghezza:
- compresa tra 4 e 5 mm. (R. 0.2);
- maggiore di 5 mm. (R. 0.5);
- minore di 3 mm. (R. 0.1);
B. Una qualitá di semi viene dispersa uniformemente su un campo quadrato di superficie $S_o = 4$ km$^2$. - Quale  la probabilit che un seme cada in una zona quadrata di lato $L_o = 600m^2$ all'interno di $S_o$? - Quale  la probabilit che un seme cada sulla superficie compresa tra $r_1 = 400$ m e $r_2 = 600$ m al centro del campo? - Quale  la probabilit che un seme cada sulla superficie compresa tra $r_1 = 500$ m e $r_2 = 700$ m al centro del campo?

**Esercizio D4**
**Scopo**: calcolo di probabilitá per variabili che seguono la distribuzione Normale.
Il peso di un frutto maturo della specie A segue una distribuzione normale con valor medio $\mu = 54.5$ g e varianza $\mu = 9.5$ g. Calcolare:
- la probabilitá di osservare un frutto maturo che pesa esattamente 54.5 g (R. 0.0);
- la probabilitá di osservare un frutto maturo il cui peso é tra 40 g e 50 g (R. 0.25);
- la probabilitá di osservare un frutto maturo il cui peso é tra 53 g e 58 g (R. 0.20);
- la probabilitá di osservare un frutto maturo che pesa meno di 35 g (R. 0.02);
- la prbabilitá di osservare un frutto maturo che pesa meno di 35 g oppure piú di 70 g (R. 0.07);
- la probabilitá di osservare un frutto maturo il cui peso sia tra 45 g e 64 g (R. 68.2);
- la probabilitá che prendendo 10 frutti meno di 4 pesino meno di 50 g (R. usare distribuzione normale per calcolare la probabilitá di pesare medo di 50 g e quindi la distribuzione binomiale per la risposta: p=0.32, P($< 4|10) = 0.6$)

Calcolare e riportare su un grafico distribuzioni di densitá di probabilitá Normale con: ($\mu = 9.3$, $\sigma = 2.5$); ($\mu = 9.3$, $\sigma = 3.5$); ($\mu = 9.3$, $\sigma = 5.2$)

Calcolare l'istogramma di frequenza calcolati per una variabile che segua la distribuzione di Poisson con $\lambda = 6.5$ $P(x, 6.5)$ nell'intervallo $[0 \leq x \leq 20]$ e confrontarli con:
- una distribuzione di Gauss con $\mu = 6.5$ e $\sigma = 2.5$ $G(x, 6.5, 2.5)$
- per una distribuzione di Gauss con $\mu = 6.5$ e $\sigma = 1.5$ $G(x, 6.5, 1.5)$
- per una distribuzione di Gauss con $\mu = 6.5$ e $\sigma = 3.5$ $G(x, 6.5, 3.5)$
Utilizzare come esempio i fogli Excel `Distribuzioni.xls`.

**Esercizio D5** I valori di un parametro fisiologico A in persone sane seguono una distribuzione normale con valore atteso $\mu_s = 84.5$ ng/ml e deviazione standard $\sigma_s = 7.8$ ng/ml. In persone affette da una patologia Y i valori seguono una distribuzione normale con valore atteso $\mu_p = 72.5$ ng/ml e deviazione standard $\sigma_p = 8.3$ ng/ml.
- Riportare su un grafico le due distribuzioni a confronto.
- Stabilire la frazione di falsi negativi in un test se si usa come discriminante il valore $x_o = 80$ ng/ml.
- Calcolare la probabilit di falsi positivi, la specificitá e la sensibilitá del test.
Discutere i risultati ottenuti in una breve relazione.

## 2.2 Esperimenti

**Esperimento D1**
**Scopo**: Riportare su un grafico dati sperimentali con la propria incertezza. Confrontare dati sperimentali con distribuzioni teoriche.
I files: **binomiale.dat**, **Poisson.dat**, **uniforme.dat**, **normale.dat** contengono dati sperimentali diversi. Confrontare graficamente i dati sperimentali con le distribuzioni modello riportando le incertezze negli istogrammi di frequenza.
- **binomiale.dat**: confrontare con una distribuzione binomiale $N = 12$, $p = 0.31$.

- **Poisson.dat**: confrontare con una distribuzione di Poisson con valore atteso $\lambda = 4$.
- **uniforme.dat**: confrontare con una distribuzione di uniforme tra 8 e 11 mm.
- **normale.dat**: confrontare con una distribuzione normale con valore atteso $\mu = 81.2$ mmHg e $\sigma = 5.5$ mmHg.

Discutere i risultati ottenuti in una breve relazione.

**Esperimento D2**
**Scopo**: Riportare su un grafico dati sperimentali con l'errore. Confrontare dati sperimentali con distribuzioni teoriche (distribuzioni discrete).

Per un numero $N_t$ di volte: estrarre a caso 12 fagioli da un contenitore e riportare il numero di fagioli neri in un file di dati (eventualmente usare i dati del file: `dati_exp_discreti.txt`).
Calcolare un istogramma di frequenze utilizzando classi di ampiezza 2: 0,2,4,6...
- calcolare le incertezze sui valori ottenuti.
- Confrontare i dati sperimentali con i valori teorici calcolati utilizzando una funzione di distribuzione binomiale con N=12 e p=0.5
- Confrontare i dati sperimentali con i valori teorici calcolati utilizzando una funzione di distribuzione di Poisson il cui valor medio sia quello ottenuto dai dati sperimentali.

Riassumere i risultati in una relazione da inviare come nelle esercitazioni precedenti
Si puó seguire, come esempio, l'esercizio svolto nel file `Confronto_2.xls`.

**Esperimento D3**
**Scopo**: Riportare su un grafico dati sperimentali con l'errore. Confrontare dati sperimentali con distribuzioni teoriche (distribuzioni continue).

Il file `dati_exp_continui.txt` contiene il peso di un campione di alcuni frutti.
- Importare i dati in un foglio elettronico.
- Calcolare le tabelle frequenze assolute e relative usando classi di ampiezza 2.
- Calcolare le incertezze sui valori ottenuti.
- Confrontare i dati sperimentali con i valori teorici calcolati utilizzando una funzione di distribuzione Normale (Gauss) con $\mu = 54$ [g] e $\sigma = 3$[g].

Riassumere i risultati in una relazione da inviare come nelle esercitazioni precedenti
Si puó seguire, come esempio, l'esercizio svolto nel file `Confronto_3.xls`.

**Esperimento D4**
Utilizzando i dati sulle dimensioni dei liposomi (**Es.04.liposoma**) confrontare la distribuzione ottenuta con una distribuzione Normale. Utilizzare il valore medio e la deviazione standard calcolati come parametri della distribuzione normale. Commentare i risultati ottenuti in una breve relazione.
Il file **Es.04.liposoma_02.xls** contiene un esempio di calcolo.

**Esperimento D5**
Ripetere l'esperienza descritta nell'esercizio (**RelazionePoisson.zip**). Utilizzare i dati disponibili o, meglio, raccogliere i dati personalmente.
Commentare i risultati ottenuti in una breve relazione.