Prova scritta di "Analisi statistica dei dati sperimentali"
5/12/2003

Problema 1

In un urna ci sono b palline blu, r palline rosse e g palline gialle, per un totale di N palline. Vengono effettuate n estrazioni e si ottengono k palline blu (dopo ogni estrazione la pallina viene rimessa nell'urna).

- Quale e' la probabilita' $P_1(k)$ da associare a questa operazione?
- Nell'ipotesi che la pallina estratta non venga riposta nell'urna, come si puo' calcolare la probabilita' $P_2(k)$ che in n estrazioni si abbiano k palline blu?
- Si dimostri che per $N \to \infty$ la seconda distribuzione diventa uguale alla prima, e giustificare il risultato.
- Si abbiano 9 palline blu, 3 rosse e 5 gialle. Determinare le probabilita' $P_1(k)$ e $P_2(k)$ che su cinque estrazioni si ottengano 2 palline blu.

Soluzione

a) Nel primo caso siamo ovviamente di fronte ad una probabilita' di tipo binomiale, con n, $p = b/N$ e $q = 1 - p = (r + g)/N$, ovvero

$$P_1(k) = P_1(k; n, p) = \binom{n}{k} p^k q^{n-k}$$

Infatti, la domanda implica che vi siano in realta' due casi, l'estrazione della pallina "blue" e della pallina "non-blue". Quindi a tutti gli effetti possiamo considerare il "non-blue" come la somma di rosso e giallo.

b) Le considerazioni svolte un attimo fa sono immediatamente applicabili al caso successivo. Infatti ora dobbiamo utilizzare la distribuzione ipergeometrica, la cui forma generale e'

$$P_2(k_1, k_2, ..., k_i; a_1, a_2, ..., a_i) = \frac{\binom{a_1}{k_1}\binom{a_2}{k_2}\cdots\binom{a_i}{k_i}}{\binom{a_1+a_2+...+a_i}{k_1+k_2+...+k_i}}$$

che nel caso dell'esercizio si riduce a

$$P_2(k, n-k; b, r+g) = \frac{\binom{b}{k}\binom{r+g}{n-k}}{\binom{b+r+g}{n}}$$

visto che possiamo definire palline "blue" e "non-blue" (quest'ultimo uguale alla somma delle palline rosse e gialle).

c) Nel caso in cui le palline diventino molto numerose ($N = b + r + g \to \infty$) , pur mantendendo costanti le proporzioni, possiamo calcolare la distribuzione ipergeometrica esplicitandone i coefficienti binomiali

$$P_2(k, n-k; b, r+g) = \left(\frac{b!}{k!(b-k)!}\frac{(r+g)!}{(n-k)!(r+g-n+k)!}\right) : \left(\frac{(N)!}{n!(N-n)!}\right)$$

e nell'ipotesi di $N \to \infty$ (che implica anche che b, r e g tendono all'infinito) possiamo approssimare

$$\frac{b!}{(b-k)!} = b^k \quad ; \quad \frac{(r+g)!}{(r+g-n+k)!} = (r+g)^{n-k} \quad ; \quad \frac{(N)!}{(N-n)!} = (N)^n$$

che ci permette di riscrivere l'espressione finale

$$\begin{aligned} P_2(k, n-k; b, r+g) &= \left(\frac{b^k}{k!} \frac{(r+g)^{n-k}}{(n-k)!} \right) : \left(\frac{(b+r+g)^n}{n!} \right) \\ &= \frac{n!}{k!(n-k)!} \frac{b^k (r+g)^{n-k}}{N^n} \\ &= \frac{n!}{k!(n-k)!} \left(\frac{b}{N} \right)^k \left(\frac{r+g}{N} \right)^{n-k} \\ &= \binom{n}{k} p^k q^{n-k} \\ &= P_1(k; n, p) \end{aligned}$$

d) Esplicitiamo le espressioni $P_1(k=2)$ e $P_2(k=2)$ nel caso in cui $n=5$, $b=9$, $r=3$ e $g=5$. Ne deriva che $N=17$, $p=b/N=9/17$ e quindi

$$P_1(2; 5, 9/17) = \binom{5}{2} (9/17)^2 (8/17)^3 = 0.292$$

$$P_2(2, 3; 9, 8) = \frac{\binom{9}{2}\binom{8}{3}}{\binom{17}{5}} = 0.326$$

Problema 2

Un'equipe di ricercatori vuole valutare una possibile relazionelineare tra la latitudine θ e l'accelerazione di gravita' g, del tipo $g = a + b\theta$, ed a tal scopo effettua la misura di g a cinque latitudini diverse (0, 20, 40, 60 ed 80 gradi di latitudine nord), trovando rispettivamente i seguenti valori (9.84, 9.82, 9.80, 9.77, 9.76 m/s^2) tutti con la stessa incertezza (0.01). Discutere il risultato.

Soluzione

In questo esercizio dobbiamo innanzitutto verificare che i dati raccolti siano effettivamente correlati da una relazione lineare; successivamente dobbiamo determinare i parametri a e b con la loro incertezza, ed infine discutere sul risultato trovato, facendo il test di χ^2 che ci permette di rigettare l'eventuale ipotesi di linearita' dei dati. Si noti che stabilire una qualche correlazione lineare tra gli stessi non implica necessariamente che i dati siano rappresentabili al meglio da una retta. Infatti, stabilito che i dati hanno tutti la stessa incertezza assoluta, il coefficiente di correlazione lineare r avrebbe comunque sempre lo stesso valore indipendente dalla stessa, cosa che non succede nel test di χ^2.

a) In virtu' dell'uguaglianza delle incertezze assolute, calcolo le seguenti quantita' separatamente utili per il calcolo di r:

$$N = 5; \ \sum x_j = 200; \ \sum x_j^2 = 12000;$$

$$\sum y_j = 48.99; \ \sum y_j^2 = 480.01; \ \sum x_j y_j = 1955.4$$

$$r = \frac{N\sum x_j y_j - \sum x_j \sum y_j}{[N\sum x_j^2 - (\sum x_j)^2]^{1/2}[N\sum y_j^2 - (\sum y_j)^2]^{1/2}} = -0.9922$$

Considero ora l'ipotesi "i dati non sono correlati", e calcolo la probabilita' che pescando un campione di N coppie (x, y) da una popolazione di dati non correlati si possa ottenere un valore di $|r|$ maggiore di quello sperimentale

$$P_N(|r| > 0.9922) \simeq 0.0005 < 0.01$$

Il risultato implica che l'ipotesi puo' essere rigettata con un alto livello di significativita' (maggiore del 99%).

b) I parametri a e b sono determinati dalla "regressione lineare", ed il risultato ci da $a = 9.8400$ e $b = -0.00105$. Dobbiamo valutare l'incertezza sulla determinazione di questi due valori prima di poterli scrivere con il corretto numero di cifre significative. Il calcolo delle incertezze sui parametri a e b viene fatto tramite la propagazione delle incertezze casuali. Nel caso di incertezza $s_y = s$ costante per tutti i valori le espressioni diventano particolarmente semplici. Valutato $\Delta = N\sum x_j^2 - (\sum x_j)^2 = 5 \times 12000 - (200)^2 = 20000$ si ottiene

$$\begin{aligned} s_a &= s\sqrt{\sum x_j^2/\Delta} = 0.01\sqrt{12000/20000} = 0.008 \\ s_b &= s\sqrt{N/\Delta} = 0.01\sqrt{5/20000} = 0.00016 \end{aligned}$$

Per cui $s_a = 9.840 \pm 0.008$ ed $s_b = (-1.05 \pm 0.16) \cdot 10^{-3}$

c) Il calcolo del χ^2_ν (chi-quadro ridotto)

$$
\begin{aligned}
\chi^2_{nu} &= \frac{1}{N-2} \sum_{j=1}^{5} \frac{(y_j - a - bx_j)^2}{s_j^2} \\
&= \frac{1}{3s^2} \sum_{j=1}^{5} (y_j - a - bx_j)^2 \\
&= \frac{1}{3(0.01)^2} [0 + 0.1 + 0.4 + 4.9 + 1.6] \cdot 10^{-5} \\
&= 3.33 \cdot 10^3 \times 7.00 \cdot 10^{-5} = 0.23
\end{aligned}
$$

$$P(\chi^2_3 > 0.23) \simeq 0.85 > 0.05$$

Il valore trovato non permette di rigettare l'ipotesi che i dati seguano un andamento lineare.

Problema 3

Un monitor di radiazione di piccole dimensioni posto a 10 cm di distanza da una sorgente, misura 1500 conteggi in un'ora. Viene tolta la sorgente e si effettua una misura del fondo di 312 conteggi in un'ora. Ponendo il monitor a 20 cm di distanza si ripete la misura e si ottengono 640 e 290 conteggi, rispettivamente in presenza ed in assenza della sorgente (e sempre nello stesso tempo di un'ora).

1. Determinare la frequenza di conteggio (in s^{-1}) a fondo sottratto, nei due casi.
2. Le due misure, nell'ipotesi di sorgente puntiforme ed isotropa, sono compatibili?
3. La sorgente viene schermata da una parete che ha una fenditura orizzontale. Che tipo di distribuzione dell'intensita' ci aspettiamo di misurare dall'altra parte della fenditura?

Soluzione

a) Siano $M_1 = 1500c/h$, $F_1 = 312c/h$, $M_2 = 640c/h$ ed $F_2 = 290c/h$. I quattro conteggi sono caratterizzati da una incertezza di tipo poissoniano. Valutiamo pertanto l'incertezza e riscriviamo i valori dei conteggi correttamente (in questo caso manteniamo conservativamente due cifre significative per l'incertezza; arrotonderemo correttamente piu' avanti)

$$\begin{aligned}
\sigma_{M_1} &= \sqrt{1500} \simeq 39 \quad \rightarrow M_1 = 1500 \pm 39 \\
\sigma_{F_1} &= \sqrt{312} \simeq 18 \quad \rightarrow F_1 = 312 \pm 18 \\
\sigma_{M_2} &= \sqrt{640} \simeq 25 \quad \rightarrow M_2 = 640 \pm 25 \\
\sigma_{F_2} &= \sqrt{290} \simeq 17 \quad \rightarrow F_2 = 290 \pm 17
\end{aligned}$$

Prima di procedere dobbiamo fare alcune considerazioni. La misura del conteggio dipende dalla distanza dalla sorgente (che nel testo viene assunta puntiforme ed isotropa), mentre la misura del fondo in assenza di sorgente dipende dalle caratteristiche dello strumento, ma anche forse dell'ambiente circostante (potrebbe esserci una qualche altra sorgente di radiazione non nota...). Quindi dobbiamo innanzitutto verificare se le due misure di fondo sono compatibili. Per fare questo possiamo applicare l'ipotesi nulla.

$$\Delta = F_1 - F_2 = 22 \; ; s_\Delta = \sqrt{F_1 + F_2} = 25$$

Non serve certamente calcolare l'integrale per vedere se l'ipotesi e' valida (i due valori sono distanti tra loro meno di σ). Possiamo quindi procedere in quanto abbiamo stabilito che le due misure di fondo sono compatibili. Essendo compatibili, i conteggi possono essere tra loro sommati per ottenere una statistica migliore

$$F = (F_1 + F_2)/2 = 301 \; ; \; s_F = \sqrt{F_1 + F_2}/2 = 12$$

Possiamo quindi valutare i conteggi a fondo sottratto $C_1 = M_1 - F$ e $C_2 = M_2 - F$

$$C_1 = M_1 - F = 1500 - 301 = 1199 \quad , \quad s_{C_1} = \sqrt{M_1 + s_F^2} = 40$$
$$C_2 = M_2 - F = 640 - 301 = 350 \quad , \quad s_{C_2} = \sqrt{M_2 + s_F^2} = 28$$

calcoliamo ora le frequenze di conteggio (in s^{-1})

$$f_1 = (3.33 \pm 0.11) \cdot 10^{-1} \ ; \ f_2 = (0.97 \pm 0.08) \cdot 10^{-1}$$

b) Per verificare se le due misure sono compatibili, dobbiamo prima normalizzarle tra loro. Nell'ipotesi di sorgente puntiforme ed isotropa, l'intensita' della radiazione diminuisce in modo inverso con il quadrato della distanza, pertanto nel secondo caso dovremmo avere un'intensita' ridotta di un quarto. Normalizzando f_2 ad f_1 dobbiamo calcolare $f_2^{'} = 4 \times f_2$ e confrontare i due risultati.

$$\Delta f = f_1 - f_2^{'} = (0.333 - 4 \times 0.097) = -0.055$$
$$s_{\Delta f} = \sqrt{s_{f_1}^2 + s_{f_2^{'}}^2} = \sqrt{((0.011^2 + (4 \times 0.008)^2)} = 0.034$$
$$\frac{|\Delta f - 0|}{s_{\Delta f}} = 1.6 < 2$$

Quindi non possiamo rigettare l'ipotesi che i dati siano compatibili.

c) Una sorgente puntiforme ed isotropa ha un'intensita' di radiazione costante in funzione dell'angolo di emissione. Passando attraverso una fenditura, ed evitando di produrre fenomeni di diffrazione apprezzabili, si puo' determinare l'intensita' della luce lungo la fenditura. Se $f(\theta) = cost$, la corrispondente distribuzione di intensita' lungo la fenditura sara' tale che $f(\theta)\, d\theta = g(x)\, dx$. Rimane da determinare $g(x) = f(\theta(x)\, d\theta/dx$. Ma $x = r\tan\theta$, dove r e' la distanza della sorgente dalla fenditura, ed assumendo unitaria tale distanza, otteniamo $\theta = \arctan x$ e $d\theta/dx = 1/(1+x^2)$. Alla fine si ottiene, tenendo conto della normalizzazione della densita' di probabilita',

$$g(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

che e' la distribuzione di Cauchy, vista a lezione