Compito di "Analisi statistica dei dati sperimentali"
14/01/2005

Problema 1

Un fotomoltiplicatore a tre stadi moltiplica il segnale iniziale di fotoelettroni secondo la seguente relazione

$$G = \delta_1 \delta_2 \delta_3$$

Le grandezze δ_i hanno una distribuzione poissoniana, e sono tra loro indipendenti.
a) Si calcolino la media e la varianza relativa di G in modo esatto.
b) Si calcoli la varianza relativa con la regola generale della propagazione delle incertezze statistiche.
c) Si valuti l'errore sistematico, dovuto all'approssimazione "b)", nel caso in cui $< \delta_i >= 4 \quad \forall i$.
N.B. si elabori l'esercizio per il calcolo piu' semplice di $G' = \delta_1 \delta_2$, e facoltativamente per il calcolo di $G = G' \delta_3$.

Soluzione

a) Procediamo per passi successivi, calcolando inizialmente la media e varianza di $G' = \delta_1 \delta_2$, e successivamente quella di $G = G'\delta_3$. Le grandezze δ_i hanno distribuzione poissoniana, per cui $E[\delta_i] = \mu_i$, e $Var[\delta_i] = \mu_i$ (nel calcolo dobbiamo inoltre ricordare l'espressione generale per la varianza $Var[\delta_i] = E[\delta_i^2] - E^2[\delta_i]$, da cui si ricava $E[\delta_i^2] = Var[\delta_i] + E^2[\delta_i]$)

$$E[G'] = E[\delta_1 \delta_2] = E[\delta_1]E[\delta_2] = \mu_1 \mu_2$$

$$\begin{aligned} Var[G'] &= E[(\delta_1\delta_2)^2] - E^2[(\delta_1\delta_2)] = E[\delta_1^2\delta_2^2] - \mu_1^2\mu_2^2 \\ &= E[\delta_1^2]E[\delta_2^2] - \mu_1^2\mu_2^2 = (Var[\delta_1] + \mu_1^2)(Var[\delta_2] + \mu_2^2) - \mu_1^2\mu_2^2 \\ &= Var[\delta_1]Var[\delta_2] + Var[\delta_1]\mu_2^2 + \mu_1^2 Var[\delta_2] + \mu_1^2\mu_2^2 - \mu_1^2\mu_2^2 = \mu_1\mu_2 + \mu_1\mu_2^2 + \mu_1^2\mu_2 \end{aligned}$$

$$E[G] = E[G'\delta_3] = E[G']E[\delta_3] = \mu_1\mu_2\mu_3$$

$$\begin{aligned} Var[G] &= E[(G'\delta_3)^2] - E^2[(G'\delta_3)] = E[G'^2\delta_3^2] - G'^2\mu_3^2 \\ &= E[G'^2]E[\delta_3^2] - E^2[G']\mu_3^2 = (Var[G'] + E^2[G'])(Var[\delta_3] + \mu_3^2) - E^2[G']\mu_3^2 \\ &= Var[G']Var[\delta_3] + Var[G']\mu_3^2 + E^2[G']Var[\delta_3] + E^2[G']\mu_3^2 - E^2[G']\mu_3^2 \\ &= (\mu_1\mu_2 + \mu_1\mu_2^2 + \mu_1^2\mu_2)\mu_3 + (\mu_1\mu_2 + \mu_1\mu_2^2 + \mu_1^2\mu_2)\mu_3^2 + \mu_1^2\mu_2^2\mu_3 \\ &= \mu_1\mu_2\mu_3 + \mu_1\mu_2^2\mu_3 + \mu_1^2\mu_2\mu_3 + \mu_1\mu_2\mu_3^2 + \mu_1\mu_2^2\mu_3^2 + \mu_1^2\mu_2\mu_3^2 + \mu_1^2\mu_2^2\mu_3 \end{aligned}$$

Nell'ipotesi che le tre grandezze siano uguali $\mu_1 = \mu_2 = \mu_3 = \mu$, possiamo semplificare ed otteniamo

$$\frac{Var[G']}{E^2[G']} = \frac{2\mu^3 + \mu^2}{\mu^4} = \frac{2}{\mu} + \frac{1}{\mu^2}$$

$$\frac{Var[G]}{E^2[G]} = \frac{3\mu^5 + 3\mu^4 + \mu^3}{\mu^6} = \frac{3}{\mu} + \frac{3}{\mu^2} + \frac{1}{\mu^3}$$

b) Per il calcolo della varianza sperimentale di G' e G (denotate con $s^2_{G'}$ ed s^2_G), utilizziamo la formula generale della propagazione delle incertezze con le derivate parziali calcolate in corrispondenza dei valori medi delle grandezze misurate $< \delta_i >= \mu_i$, e $s^2_{\delta_i} = \mu_i \quad \forall i$;

$$s^2_{G'} = \left|\frac{\partial G'}{\partial \delta_1}\right|^2_{\mu_1,\mu_2} s^2_{\delta_1} + \left|\frac{\partial G'}{\partial \delta_2}\right|^2_{\mu_1,\mu_2} s^2_{\delta_2} = \mu_2^2\mu_1 + \mu_1^2\mu_2$$

da cui otteniamo

$$\frac{s^2_{G'}}{< G' >^2} = \frac{\mu_2^2\mu_1 + \mu_1^2\mu_2}{(\mu_1\mu_2)^2} = \frac{\mu_1\mu_2(\mu_2 + \mu_1)}{(\mu_1\mu_2)^2} = \frac{\mu_1 + \mu_2}{\mu_1\mu_2}$$

analogamente per G si ricava:

$$s^2_G = \left|\frac{\partial G}{\partial \delta_1}\right|^2_{\mu_1,\mu_2,\mu_3} s^2_{\delta_1} + \left|\frac{\partial G}{\partial \delta_2}\right|^2_{\mu_1,\mu_2,\mu_3} s^2_{\delta_2} + \left|\frac{\partial G}{\partial \delta_3}\right|^2_{\mu_1,\mu_2,\mu_3} s^2_{\delta_3} = \mu_2^2\mu_3^2\mu_1 + \mu_1^2\mu_3^2\mu_2 + \mu_1^2\mu_2^2\mu_3$$

da cui otteniamo

$$\frac{s^2_G}{< G >^2} = \frac{\mu_2^2\mu_3^2\mu_1 + \mu_1^2\mu_3^2\mu_2 + \mu_1^2\mu_2^2\mu_3}{(\mu_1\mu_2\mu_3)^2} = \frac{\mu_1\mu_2 + \mu_1\mu_3 + \mu_2\mu_3}{\mu_1\mu_2\mu_3}$$

c) Nell'ipotesi che $\mu_1 = \mu_2 = \mu_3 = \mu$, le due espressioni diventano

$$\frac{s^2_{G'}}{< G' >^2} = \frac{2}{\mu} \quad ; \quad \frac{s^2_G}{< G >^2} = \frac{3}{\mu}$$

Si nota immediatamente che la formula della propagazione "in quadratura", e' un'approssimazione al primo ordine dell'espressione generale dell'incertezza relativa (del resto proprio in questo modo l'abbiamo trovata!). Nell'ipotesi in cui $\mu >> 1$, i termini di grado superiore al primo, al denominatore, diventano meno rilevanti e l'approssimazione si avvicina all'espressione esatta. L'approssimazione comporta un'errore sistematico (per difetto) nel calcolo della varianza sperimentale di G' e di G (relativo alla varianza esatta) dato da:

$$\frac{Var[G'] - s^2_{G'}}{Var[G']} = \frac{\mu^2}{\mu^2(1 + 2\mu)} = \frac{1}{1 + 2\mu}$$

$$\frac{Var[G] - s^2_G}{Var[G]} = \frac{\mu^3 + 3\mu^4}{\mu^3(1 + 3\mu + 3\mu^2)} = \frac{1 + 3\mu}{1 + 3\mu + 3\mu^2}$$

Per $\mu = 4$ le due quantita' valgono 11% e 21%, rispettivamente. (A titolo di confronto possiamo vedere che per $\mu = 10$ le due quantita' diventano $\sim$ 5% e 9% rispettivamente).

Problema 2

Si abbia una sorgente di ^{241}Am che emette radiazione α isotropicamente. Due rivelatori con la stessa superficie attiva, posti alla stessa distanza dalla sorgente, ma in direzioni diverse, hanno contato rispettivamente 5 e 6 particelle nella prima ora.

a) Supponendo che questo rapporto si mantenga costante nel tempo, quanto dovra' durare la misura per poter stabilire con un alto livello di significativita' che i due rivelatori non hanno la stessa efficienza?

b) Qual e' la probabilita' che i due rivelatori contino entrambi 4 particelle nell'ora successiva?

c) Dopo quanto tempo, a partire da un istante qualsiasi, arriva la terza particella nel primo rivelatore (probabilita' massima)?

Soluzione

a) Il testo e' scritto male, perche' porta a fare delle considerazioni ambigue. Infatti non e' possibile "a priori" ipotizzare che il rapporto si mantenga costante in ogni istante della misura. L'ipotesi che va fatta e' che, nonostante i conteggi siano governati dalla distribuzione di Poisson, avviente *casualmente* che il rapporto, al momento della verifica, sia lo stesso della prima ora. Siano r_1 ed r_2 le frequenze orarie di conteggio dei due rivelatori, per cui i conteggi, dopo un numero di ore T sono rispettivamente $c_1 = r_1 T$ e $c_2 = r_2 T$. Dobbiamo confrontare questi conteggi tra loro, ovvero verificare che la differenza sia nulla (ipotesi nulla). Detta $\Delta = c_1 - c_2$, la differenza tra i due valori, devo calcolare la quantita'

$$t = \frac{|\Delta|}{s_\Delta} = \frac{|c_1 - c_2|}{\sqrt{c_1 + c_2}} = \frac{\sqrt{T}|r_1 - r_2|}{\sqrt{r_1 + r_2}}$$

dove abbiamo tenuto conto del fatto che i conteggi $c_i (i = 1, 2)$ seguono la statistica poissoniana con $s^2_{c_i} = c_i$, e propagato le incertezze in quadratura. Affinche' si possa stabilire con un alto livello di significativita', e' necessario che l'ipotesi si discosti dal risultato sperimentale per una quantita' tale che $P(|t| \geq t_0) \leq 0.01$, il che comporta un valore $t_0 = 2.58$. Quindi

$$t_0 = 2.58 = \frac{\sqrt{T}|5 - 6|}{\sqrt{5 + 6}}$$

$$\sqrt{T} = 2.58\sqrt{11} \rightarrow T = 73.22 \text{ hours}$$

b) La probabilita' di contare un certo valore nell'ora successiva non dipende dal conteggio precedente, ma solamente dalla statistica di poisson; inoltre i due rivelatori sono indipendenti, e quindi la probabilita' che entrambi contino 4 particelle e' il prodotto delle rispettive probabilita', ovvero

$$P(4, 4) = P_{r_1}(4, 5)P_{r_2}(4, 6) = \frac{5^4}{4!}e^{-5}\frac{6^4}{4!}e^{-6} = 0.175 \times 0.134 = 0.0234$$

c) Dobbiamo utilizzare la distribuzione degli intervalli, la cui espressione generale e'

$$I_k(T, \mu) = \frac{\mu^k T^{k-1}}{(k-1)!}e^{-\mu T}$$

e determinare il valore di T che la rende massima, per $k = 3$ e $\mu = 5$. Derivo l'espressione rispetto a T, da cui

$$\begin{aligned}
\frac{dI_k(T,\mu)}{dT} &= \frac{(k-1)\mu^k T^{k-2}}{(k-1)!}e^{-\mu T} - \frac{\mu^k T^{k-1}}{(k-1)!}\mu e^{-\mu T} \\
0 &= e^{-\mu T}\left((k-1)T^{k-2}\mu^k - \mu^{k+1}T^{k-1}\right) \\
&= e^{-\mu T}T^{k-2}\mu^k\left((k-1) - \mu T\right) \\
0 &= k - 1 - \mu T \\
0 &= 3 - 1 - 5T \\
T &= 0.4 \ \text{ hours}
\end{aligned}$$

Problema 3

Da una serie di misure della grandezza y si ottengono i seguenti valori medi $\overline{y}$ (e relative incertezze $s_{\overline{y}}$): 4.8(0.3), 4.8(0.3), 5.2(0.3), 5.1(0.4), 5.4(0.4) 5.4(0.4).
a) Determinare il valore finale $< \overline{y} >$ e la sua deviazione standard.
b) Una particolare teoria prevede $\overline{y}_{th} = 4.6$. Determinare se la teoria puo' essere accettata.
c) Nell'ipotesi che le misure siano state effettuate in anni successivi si puo' verificare se vi sia una correlazione temporale?

Soluzione

a) In questo caso, poiche' le misure $\overline{y}_i$ sono affette da incertezze diverse tra loro, devo calcolare la media pesata

$$< \overline{y} >= \frac{\sum \overline{y}_i w_i}{\sum w_i}$$

con $w_i = 1/s^2_{\overline{y}_i}$, e la relativa incertezza finale

$$s^2_{<\overline{y}>} = \frac{1}{\sum_i w_i}$$

Riportando i valori in tabella

i	$\overline{y}_i$	$s_{\overline{y}_i}$	w_i	$\overline{y}_i w_i$
1	4.8	0.3	11.11	53.33
2	4.8	0.3	11.11	53.33
3	5.2	0.3	11.11	57.78
4	5.1	0.4	6.25	31.88
5	5.4	0.4	6.25	33.75
6	5.4	0.4	6.25	33.75
sum			52.08	268.82

da cui $< \overline{y} >= 5.07 \pm 0.14$

b) Applico in questo caso il test di ipotesi nulla, ovvero calcolo la differenza tra il valore misurato e quello teorico, normalizzata all'incertezza sulla misura:

$$t = \frac{| < \overline{y} >_{th} - < \overline{y} > |}{s_{<\overline{y}>}} = \frac{|4.60 - 5.07|}{0.14} = 3.4$$

e $P(|t| \geq 3.4) < 0.01)$, per cui possiamo rigettare l'ipotesi con un CL del 99%

c) Per verificare se vi e' correlazione temporale fra i valori misurati e l'anno della misura, possiamo innanzitutto scegliere che la prima misura sia stata effettuata all'anno $x_1 = 1$ e via dicendo (ma potrebbe essere un anno di partenza qualsiasi, si complicano solo i calcoli).

$$r = \frac{N \sum x_i \overline{y_i} - \sum x_i \sum \overline{y_i}}{[N \sum x_i^2 - (\sum x_i)^2]^{\frac{1}{2}} [N \sum \overline{y}_i^2 - (\sum \overline{y}_i)^2]^{\frac{1}{2}}} = 0.93$$

Dalle tabelle risulta che $P_6(|r| \geq 0.93) < 0.01)$, per cui possiamo rigettare l'ipotesi che i dati non siano correlati con un CL del 99%. Questo risultato inficia completamente i precedenti due punti dell'esercizio in quanto il valore medio viene costruito sull'ipotesi che i dati appartengano alla stessa popolazione, ed il confronto con l'ipotesi dipende dal momento "storico" in cui viene fatta la misura.