Prova scritta di "Analisi statistica dei dati sperimentali"
9/1/2004

Problema 1

La distribuzione di Lorentz e' definita dalla seguente espressione:

$$\frac{1}{\pi}\frac{\Gamma/2}{(x-\mu)^2+(\Gamma/2)^2}$$

a) Qual e' il significato di Γ?
b) Qual e' la frazione dell'area di questa distribuzione nell'intervallo $\mu \pm 3(\Gamma/2)$?
c) In base ai risultati ottenuti, qule e' la differenza tra la distribuzione di Lorentz e quella di Gauss centrata su μ con $\sigma = \Gamma/2.345$?

Soluzione

a) Γ rappresenta la larghezza a meta' altezza ($FWHM$) della densita' di probabilita' lorentziana

$$L(x;\mu,\Gamma) = \frac{1}{\pi}\frac{\Gamma/2}{(x-\mu)^2+(\Gamma/2)^2}$$

Per verificarlo, devo innanzitutto valutare il massimo della distribuzione, che per motivi di simmetria della funzione (pari) e' sicuramente in corrispondenza di $x = \mu$.

$$L_{max} = L(\mu;\mu,\Gamma) = \frac{1}{\pi}\frac{\Gamma/2}{(\Gamma/2)^2} = \frac{2}{\pi\Gamma}$$

Determino ora i (due) valori di x che rendono uguale a $L_{max}/2$ il valore della distribuione

$$L_{max}/2 = \frac{1}{\pi}\frac{\Gamma/2}{(x-\mu)^2+(\Gamma/2)^2} = \frac{1}{\pi\Gamma}$$

$$(x-\mu)^2+(\Gamma/2)^2 = \Gamma^2/2$$

$$(x-\mu)^2 = \Gamma^2/2 - \Gamma^2/4 = \Gamma^2/4$$

$$x-\mu = \pm\Gamma/2 \Rightarrow x_{1,2} = \mu \pm \Gamma/2$$

ed alla fine

$$FWHM = |x_2 - x_1| = \Gamma \qquad \text{c.v.d}$$

b) Calcolo l'integrale della distribuzione che corrisponde alla probabilita'

$$P(\mu - 3\Gamma/2 \le x \le \mu + 3\Gamma/2) = \int_{\mu-3\Gamma/2}^{\mu+3\Gamma/2} L(x;\mu,\Gamma)dx$$

operando, per semplicita', un cambio di variabile $y = (x-\mu)/(\Gamma/2)$

$$P(-3 \le y \le 3) = \frac{1}{\pi}\int_{-3}^{+3}\frac{1}{y^2+1}dy = \frac{1}{\pi}\left[\arctan y\right]_{-3}^{+3} = 0.7952$$

c) Nella distribuzione di Gauss $FWHM = 2.345\sigma$, per cui le due distribuzioni cosi' identificate hanno la stessa $FWHM$. Per quanto riguarda l'area della distribuione compresa tra gli stessi valori

$$P(\mu - 3\Gamma/2 \leq x \leq \mu + 3\Gamma/2) = \int_{\mu-3\Gamma/2}^{\mu+3\Gamma/2} \frac{1}{\sigma\sqrt{2\pi}} e^{(x-\mu)^2/2\sigma^2} dx$$

posso calcolarla in funzione di σ, ovvero riducendo l'integrale a quello di una gaussiana standard, attraverso il cambio di variabile $t = (x - \mu)/\sigma$ e sapendo che $\sigma = \Gamma/2.345$. Ottengo alla fine

$$P(-3.52 \leq t \leq 3.52) \simeq 1$$

che dimostra come la distribuzione di Gauss sia piu' stretta della corrispondente distribuzione di Lorentz (ed infatti quest'ultima non converge per il calcolo della varianza).

Problema 2

In una "slot machine" con tre finestre ci sono due simboli di limone su dieci in ogni finestra. La macchina paga somme diverse nel caso appaiano un limone, una coppia di limoni o una terna.

a) Quanto dovrebbero essere le somme pagate nei tre casi affinche' la macchina fosse onesta?
b) Se dopo mille giocate non si ottiene nemmeno una terna, si puo' gia' pensare che la slot machine sia truccata?
c) Ripetere i calcoli di "a)" e "b)" nel caso in cui la finestra centrale abbia invece un solo simbolo di limone.

Soluzione

a) Poiche' le tre finestre hanno la stessa frazione di simboli di limone (2 su 10), la probabilita' che compaia quel simbolo e' la stessa su ogni finestra. Assumendo infine che gli eventi nelle tre finestre siano totalmente indipendenti, posso applicare la distribuione Binomiale

$$P_B(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

con $n = 3$ e $p = 2/10 = 0.2$. Devo pertanto valutare le diverse probabilita'

$$\begin{aligned}
P_B(1) &= \binom{3}{1} (0.2)^1 (0.8)^2 = 0.384 \\
P_B(2) &= \binom{3}{2} (0.2)^2 (0.8)^1 = 0.096 \\
P_B(3) &= \binom{3}{3} (0.2)^3 (0.8)^0 = 0.008
\end{aligned}$$

Una slot machine "onesta" dovrebbe pagare una somma equivalente all'inverso della probabilita' di ottenere l'evento, e quindi rispettivamente 2, 10 e 125 volte la moneta inserita (in questo caso l'arrotondamento e' a discapito del giocatore...)

b) Per valutare se la slot machine e' truccata, debbo calcolare la probabilita' di ottenere una terna dalla slot machine (che ho appena trovato al punto "a"), ed applicare la binomiale, con $k = 0$ ed $n = 1000$. In questo particolare caso il calcolo e' semplice, in quanto il termine binomiale $\binom{1000}{0} = 1$, e quindi

$$P_B(0) = \binom{1000}{0} (0.008)^0 (0.992)^{1000} = 3.25 \cdot 10^{-4} \ll 0.01$$

Quindi se dopo 1000 lanci non ho ottenuto nemmeno una terna, posso rigettare l'ipotesi che la slot machine sia onesta in modo altamente significativo (C.L. 99%). In generale si puo' applicare a questi casi la formula di Poisson, in quanto $p << 1$ ed $n >> 1$, e $\mu = np = 1000 \times 0.008 = 8$, e poi calcolare

$$P_P(k; \mu) = \frac{\mu^k}{k!} e^{-\mu} \Rightarrow P_P(0; 8) = e^{-8} = 3.36 \cdot 10^{-4} \ll 0.01$$

Il risultato e' leggermente diverso in quanto la formula di Poisson e' pur sempre un'approssimazione.

c) In questo caso non posso agire come al punto "a" in quanto la probabilita' dell'evento non e' la stessa per tutte le finestre. Devo pertanto calcolare la probabilita' applicando le regole base della probabilita', ipotizzando, anche in questo caso, come nel caso precedente, che gli eventi siano indipendenti. Dette $p_1 = 0.2$, $p_2 = 0.1$ e $p_3 = 0.2$ le tre probabilita' di ottenere il simbolo del limone nelle tre finestre, la probabilita' di ottenere uno, due o tre simboli sulle tre finestre sara' data da

$$\begin{aligned} P(1) &= (1-p_1)(1-p_2)(p_3) + (1-p_1)(p_2)(1-p_3) + (p_1)(1-p_2)(1-p_3) \\ P(2) &= (1-p_1)(p_2)(p_3) + (p_1)(1-p_2)(p_3) + (p_1)(p_2)(1-p_3) \\ P(3) &= (p_1)(p_2)(p_3) \end{aligned}$$

e numericamente

$$\begin{aligned} P(1) &= 0.144 + 0.064 + 0.144 &&= 0.352 \\ P(2) &= 0.016 + 0.036 + 0.016 &&= 0.068 \\ P(3) &= &&= 0.004 \end{aligned}$$

Analogamente a quanto visto prima, si applica Poisson con $\mu = np = 4$. In questo caso la probabilita' di non ottenere alcuna terna e'

$$P_P(k;\mu) = \frac{\mu^k}{k!} e^{-\mu} \Rightarrow P_P(0;4) = e^{-4} = 0.018 < 0.05$$

Si puo' rigettare l'ipotesi che la slot machine sia onesta in modo significativo (C.L. 95%) ma non in modo altamente significativo (C.L. 99%).

Problema 3

Si consideri la seguente espressione:

$$z = R \ln t/t_0$$

a) se t ha le dimensioni di un tempo, quali sono le dimensioni di t_0 e perche'? Quali sono le dimensioni di z?
b) assumendo le seguenti incertezze statistiche

$$t = 10 \pm 1,\ t_0 = 100 \pm 2,\ R = 7.4 \pm 0.3$$

calcolare z e la sua incertezza.
c) come nel caso precedente, ma assumendo che l'incertezza su R sia un errore di sensibilita' (valutare entrambe le soluzioni)

Soluzione

a) La funzione "ln" e' trascendente, pertanto sia l'argomento della funzione, che il valore della funzione stessa sono adimensionati. Questo implica che $[t_0] = [t]$ e che $[z] = [R]$.

b) Innanzitutto calcolo il valore di z

$$z = R \ln t/t_0 = 7.4 \ln 10/100 = -17.039$$

e tengo un numero di cifre "adeguato", non avendo ancora avuto modo di determinare l'incertezza statistica su z. Per farlo, applico la formula di propagazione delle incertezze statistiche (somma in quadratura).

$$\begin{aligned}
s^2 &= \left(\frac{\partial z}{\partial R}\right)^2 s_R^2 + \left(\frac{\partial z}{\partial t}\right)^2 s_t^2 + \left(\frac{\partial z}{\partial t_0}\right)^2 s_{t_0}^2 \\
&= \left(\ln \frac{t}{t_0}\right)^2 s_R^2 + \left(\frac{R}{t}\right)^2 s_t^2 + \left(\frac{R}{t_0}\right)^2 s_{t_0}^2 \\
&= (-2.303)^2 \times (0.3)^2 + (0.74)^2 \times (1)^2 + (0.074)^2 \times (2)^2 \\
&= 5.302 \times 0.09 + 0.548 \times 1 + 5.476 \cdot 10^{-3} \times 10^4 \\
&= 0.477 + 0.548 + 0.022 \\
&= 1.047 \\
s &= 1.023
\end{aligned}$$

Possiamo ora correttamente rappresentare il valore di z con la sua incertezza statistica

$$z = -17.0 \pm 1.0 \qquad \text{C.L.68\%}$$

c) In questo caso $\Delta R = 0.3$ e' un'incertezza massima (errore di sensibilita'). Posso seguire due strade diverse:

c1) Calcolo l'incertezza massima su z; per farlo devo prima "trasformare" le incertezze statistiche su t e t_0 in incertezze massime, ponendo $\Delta t = 3 s_t = 3$ e

$\Delta t_0 = 3 s_{t_0} = 6$, e procedendo con la formula di propagazione delle incertezze massime

$$\begin{aligned}
\Delta z &= \left|\frac{\partial z}{\partial R}\right| \Delta R + \left|\frac{\partial z}{\partial t}\right| \Delta t + \left|\frac{\partial z}{\partial t_0}\right| \Delta t_0 \\
&= \left|\ln \frac{t}{t_0}\right| \Delta R + \left|\frac{R}{t}\right| \Delta t + \left|\frac{R}{t_0}\right| \Delta t_0 \\
&= |-2.303| \times 0.3 + |0.74| \times 1 + |0.074| \times 2 \\
&= 1.579
\end{aligned}$$

$$z = -17.0 \pm 1.6 \qquad \text{C.L.}100\%$$

c2) calcolo l'incertezza statistica su z trasformando l'incertezza massima su R in un'inceretezza statistica con un'ipotesi di distribuione uniforme $\sigma_R = \Delta R/\sqrt{3} = 0.173$, e rifacendo il calcolo del punto “b”

$$\begin{aligned}
s^2 &= (-2.303)^2 \times (0.173)^2 + (0.74)^2 \times (1)^2 + (0.074)^2 \times (2)^2 \\
&= 0.159 + 0.548 + 0.022 \\
&= 0.729 \\
s &= 0.854
\end{aligned}$$

$$z = -17.0 \pm 0.8 \qquad \text{C.L.}68\%$$

Anche se nella distribuzione uniforme solo il 58% e' contenuto entro $\pm\sigma_R$ (abbiamo in questo caso applicato il teorema del limite centrale).