Metodi Matematici e Statistici, Modulo di Statistica, Ing. Gestionale prova scritta del 29/6/04

1 Esercizio 1

Due eventi A e B possono influire sulla possibile rottura di un certo impianto (es. A=allentamento di certi bulloni, B=ossidazione di certi contatti). Quando l'impianto si rompe, le cause possono essere solamente A o B (per semplicità). Supponiamo (sempre per semplicità) che le cause A e B siano alternative (non accadono mai contemporaneamente).

Le cause A e B sono molto difficili e costose da osservare direttamente, per cui, quando si rompe l'impianto, si cerca innanzi tutto di intuire tramite ragionamenti probabilistici se è più probabile che la rottura sia dovuta ad A oppure a B.

Come dicevamo, A e B possono influire sulla rottura ma non la causano necessariamente. Si ritiene che quando si avvera A, nel 50% dei casi l'impianto continui a funzionare. Invece quando si verifica B, l'impianto continua a funzionare solo il 20% delle volte.

i) Supponiamo a priori equiprobabili le cause A, B. Quando l'impianto si rompe, propendiamo per credere che la causa sia B. Cosa giustifica quantitativamente questa propensione?

ii) Dopo un certo tempo di sperimentazione, si sa che A accade con probabilità p, B con probabilità $1-p$. Per quali valori di p, quando l'impianto si rompe, propendiamo per credere che la causa sia A?

2 Esercizio 2

Siano $X_1, ..., X_n$ v.a. discrete indipendenti, ciascuna che vale 1 con probabilità p, -1 con probabilità $q = 1-p$. Sia Z la somma delle X_i. Calcolare media, varianza e funzione generatrice di Z. Dopo aver eseguito il calcolo in modo elementare (a scelta), cercare anche di eseguirlo riconducendosi alle binomiali.

3 Esercizio 3

Un grossista riceve varie proteste dai suoi clienti circa il funzionamento di una marca di batterie: la confezione dichiara una durata media di 20 ore, mentre troppo spesso si osservano durate decisamente inferiori. Il grossista decide di valutare egli stesso l'eventuale malfunzionamento di quella marca e prova 10 batterie, riscontrando in esse una durata media di 18 ore. Sulla confezione è anche dichiarata una deviazione standard di 2 ore. Con che significatività può il grossista contestare al produttore la bontà delle batterie? Oltre ai calcoli, discutere bene le ipotesi adottate per la risoluzione.

4 Soluzioni

Soluzione Es. 1.

i) Indichiamo con R l'evento rottura dell'impianto. Per ipotesi $P(R|A) = 0.5$, $P(R|B) = 0.8$, ed inoltre $P(A) = 0.5$, $P(B) = 0.5$. Pertanto,

$$P(R) = P(R|A)\,P(A) + P(R|B)\,P(B) = 0.5 \cdot 0.5 + 0.8 \cdot 0.5 = 0.65$$

$$P(A|R) = \frac{P(R|A)\,P(A)}{P(R)} = \frac{0.5 \cdot 0.5}{0.65} = 0.384\,6$$

$$P(B|R) = \frac{P(R|B)\,P(B)}{P(R)} = \frac{0.8 \cdot 0.5}{0.65} = 0.615\,3$$

(che hanno somma uno), quindi è più probabile B. Questo giustifica matematicamente la propensione verso B.

ii) Se ora $P(A) = p$, $P(B) = 1 - p$, vale

$$\begin{aligned} P(R) &= P(R|A)\,P(A) + P(R|B)\,P(B) \\ &= 0.5 \cdot p + 0.8 \cdot (1 - p) = 0.8 - 0.3 \cdot p \end{aligned}$$

$$P(A|R) = \frac{P(R|A)\,P(A)}{P(R)} = \frac{0.5 \cdot p}{0.8 - 0.3 \cdot p}$$

quindi deve essere

$$\frac{0.5 \cdot p}{0.8 - 0.3 \cdot p} > \frac{1}{2}$$

ovvero

$$0.5 \cdot p > 0.4 - 0.15 \cdot p$$

da cui

$$p > \frac{0.4}{0.65} = 0.615\,38.$$

Soluzione Es. 2. Ciascuna X_i ha media

$$E[X_i] = p - q = 2p - 1$$

varianza

$$\begin{aligned} Var\left[X_i\right] &= E\left[X_i^2\right] - E\left[X_i\right]^2 = 1 - (2p-1)^2 = 4p - 4p^2 \\ &= 4pq = 1 - p^2 - q^2 + 2pq \end{aligned}$$

e generatrice

$$\phi_i\left(t\right) = e^t p + e^{-t} q$$

e quindi per linearità

$$E\left[Z\right] = n\left(2p-1\right)$$

mentre usando i teoremi sull'indipendenza

$$Var\left[Z\right] = n4pq$$

$$\phi_Z\left(t\right) = \left(e^t p + e^{-t} q\right)^n.$$

E' possibile ricondursi alla binomiale nel seguente modo: $Y_i = \frac{X_i+1}{2}$ è una Bernoulli di parametro p, quindi

$$\begin{aligned} Z &= X_1 + ... + X_n = 2Y_1 - 1 + ... + 2Y_n - 1 \\ &= 2B - n \end{aligned}$$

dove B è una $B(n,p)$. Già sapendo che $E\left[B\right] = np$, $Var\left[B\right] = npq$, $\phi_B\left(t\right) = \left(e^t p + q\right)^n$, vale

$$E\left[Z\right] = 2E\left[B\right] - n = 2np - n = n\left(2p-1\right)$$

$$Var\left[Z\right] = Var\left[2B\right] = 4Var\left[B\right] = 4npq$$

$$\begin{aligned} \phi_Z\left(t\right) &= E\left[e^{tZ}\right] = E\left[e^{t(2B-n)}\right] = E\left[e^{2tB}e^{-tn}\right] \\ &= e^{-tn}\phi_B\left(2t\right) = e^{-tn}\left(e^{2t}p + q\right)^n = \left(e^t p + e^{-t} q\right)^n. \end{aligned}$$

Soluzione Es. 3. L'ipotesi nulla è che la media sia $\mu_0 = 20$ ore. Il campione di numerosità $n = 10$ ha dato media sperimentale $\overline{x} = 18$. La deviazione standard dichiarata è $\sigma = 2$. Lo studente può eseguire a sua scelta un test bilaterale o unilaterale (sono entrambi ragionevoli).

La durata di un dispositivo è in genere una v.a. poco gaussiana, come insegna la teoria delle v.a. esponenziali. Mettiamoci però ugualmente nell'ipotesi che la durata sia gaussiana (aggiungendo un commento sotto), e che $\sigma = 2$ sia la vera deviazione standard, mentre dubitiamo della media. Calcoliamo

$$z = \sqrt{n}\frac{\overline{x} - \mu_0}{\sigma} = \sqrt{10}\frac{-2}{2} = -3.1623$$

quindi il valore p, che è la miglior significatività che permette di rifiutare l'ipotesi nulla, è

$$p = 2 - 2\Phi(|z|) = 2 - 2\Phi(3.1623) = 2 - 2 \cdot 0.9992 = 0.0016.$$

Con altro linguaggio, essendo $1 - 0.0016 = .9984$, al 99,84% si può contestare la bontà del prodotto. Per essere anche più precisi, si può dire che per qualsiasi $\alpha > 0.0016$ si può rifiutare l'ipotesi.

Osserviamo che se omettessimo l'ipotesi di gaussianità della v.a. durata, potremmo appellarci al teorema limite centrale per ritenere approssimativamente valide le stesse conclusioni, ma va osservato che la numerosità del campione è piuttosto bassa e quindi l'approssimazione può essere scarsa. Infine, non sappiamo come omettere l'ipotesi che $\sigma = 2$ sia la vera deviazione standard, in quanto non abbiamo dati circa la deviazione standard sperimentale.