Statistica I. Ingegneria Gestionale. Scritto del 20/07/2010

Esercizio 1. i) Il rischio di un evento è calcolato come prodotto tra la sua probabilità ed il suo costo. Un impianto produttivo può essere costruito perfettamente, oppure avere dei difetti di costruzione, cosa che avviene solo il 5% delle volte. Se è costruito perfettamente, nel primo anno di lavoro accade uno stop per malfunzionamento il 3% delle volte; se ha un difetto di costruzione, il 10% delle volte. Lo stop costa 5.000 euro. Calcolare il rischio di stop.

167.5

ii) Supponiamo che, se l'impianto è costuito perfettamente, l'istante in cui avviene il primo stop sia una v.a. esponenziale di media 8 mesi. Calcolare la probabilità che lo stop avvenga dopo i primi 10 mesi.

0.286

iii) Date $X \sim B(2, 0.3)$, $Y \sim N(0, 1)$ indipendenti, calcolare $P(X + Y \leq 0)$.

0.3137

iv) Calcolare inoltre $E\left[e^{3X-2Y}\right]$.

334.24

v) La sezione sviluppo delle Ferrovie di un certo paese deve monitorare il traffico di persone su una certa linea in un certo orario, per valutare l'eventualità di un aumento della dimensione del treno. Fa registrare quindi il numero N di persone che viaggia su quella linea a quell'ora, per 20 giorni. I risultati sono i numeri $x_1, ..., x_{20}$. Vale $\overline{x} = 123.87$, $x_1^2 + ... + x_{20}^2 = 314322.8$. Nella stima del numero medio giornaliero, che errore *relativo* si commette, al 95% (usare per semplicità i quantili gaussiani)?

7%

vi) Accettando per buoni i parametri trovati sopra e l'ipotesi che il numero sia gaussiano, quanti posti a sedere deve avere il treno per coprire le esigenze il 95% delle volte?

157

vii) Descriviamo ora il numero N di persone che viaggia su quella linea a quell'ora con una v.a. esponenziale. In questa ipotesi, quanti posti a sedere deve avere il treno per coprire le esigenze il 95% delle volte?

372

viii) I due modelli producono risultati assai diversi, per cui è necessario fare delle verifiche circa la loro bontà. Sia sulla base dei dati già in nostro possesso, sia utilizzando (eventualmente) i seguenti dati, valutare quale sia il modello più appropriato. Nel campione $x_1, ..., x_{20}$ ci sono 5 valori inferiori a 110 e 4 superiori a 140.

Risposta:

ix) Accettato il modello gaussiano, dimensionato il treno con 160 posti, le cose vanno bene per alcuni anni. Poi però comincia a verificarsi un certo sovraffollamento. Una verifica su 20 settimane mostra un valore medio sperimentale pari a 138.4. Se questo fosse il vero valor medio, e la varianza fosse rimasta immutata, quante volte in percentuale troveremmo il treno pieno?

$$0.138$$

x) L'aumento a 138.4 può anche essere un caso. Per valutare questa eventualità, calcolare il valore p unilaterale (cioè dando per scontato che se c'è un cambiamento esso è in crescita).

$$0.0005$$

1 Soluzioni

Esercizio 1.

$$\begin{aligned} P\left(Mal\right) &= P\left(Mal|per\right)P\left(per\right) + P\left(Mal|nonper\right)P\left(nonper\right) \\ &= 0.03\cdot 0.95 + 0.1\cdot 0.05 = 0.0335 \end{aligned}$$

e quindi il rischio è $0.0335 \cdot 5000 = 167.5$.

ii) Se la media μ è 8 (in mesi), il parametro λ della v.a. esponenziale T è $\lambda = \frac{1}{\mu} = \frac{1}{8}$. Siccome $P\left(T > 10\right) = e^{-\lambda\cdot 10} = e^{-\frac{10}{8}}$ troviamo $P\left(T > 10\right) = 0.286$.

iii)

$$\begin{aligned} P\left(X+Y\leq 0\right) = {} & P\left(X+Y\leq 0|X=0\right)P\left(X=0\right) + P\left(X+Y\leq 0|X=1\right)P\left(X=1\right) \\ & + P\left(X+Y\leq 0|X=2\right)P\left(X=2\right) \end{aligned}$$

$$\begin{aligned} &= P\left(Y\leq 0\right)P\left(X=0\right) + P\left(1+Y\leq 0\right)P\left(X=1\right) + P\left(2+Y\leq 0\right)P\left(X=2\right) \\ &= 0.5\cdot 0.7^2 + \Phi\left(-1\right)\cdot 2\cdot 0.3\cdot 0.7 + \Phi\left(-2\right)\cdot 0.3^2 \\ &= 0.5\cdot 0.7^2 + \left(1-0.8413\right)\cdot 2\cdot 0.3\cdot 0.7 + \left(1-0.9772\right)\cdot 0.3^2 = 0.3137. \end{aligned}$$

iv)

$$E\left[e^{3X-2Y}\right] = E\left[e^{3X}\right]E\left[e^{-2Y}\right] = \varphi_X\left(3\right)\varphi_Y\left(-2\right) = \left(0.3e^3+0.7\right)^2 e^{\frac{2^2}{2}} = 334.24$$

v) Vale

$$S^2 = \frac{1}{19}\left(314322.8 - 20\cdot 123.87^2\right) = 391.96$$

quindi

$$S = 19.798.$$

L'errore assoluto è (usando i quantili gaussiani)

$$\delta = \frac{19.798\cdot 1.96}{\sqrt{20}} = 8.677.$$

L'errore relativo è $\frac{\delta}{\mu}$, che approssimiamo a

$$\frac{\delta}{\overline{x}} = \frac{8.677}{123.87} = 0.070.$$

vi) Cerchiamo n (= numero di posti) tale che

$$P\left(N\leq n\right) = 0.95$$

con $N \sim N\left(123.87, 19.798^2\right)$, quindi

$$n = 123.87 + 19.798\cdot 1.64 = 156.34.$$

vii) Il parametro λ di un'esponenziale è il reciproco della media, quindi lo stimiamo con $\frac{1}{\overline{x}} = \frac{1}{123.87} = 8.073 \times 10^{-3}$. Cerchiamo n (= numero di posti) tale che

$$P(N \leq n) = 0.95$$

con $N \sim Exp\left(8.073 \times 10^{-3}\right)$, quindi

$$1 - e^{-8.073 \times 10^{-3} n} = 0.95$$

$$e^{-8.073 \times 10^{-3} n} = 0.05$$

$$8.073 \times 10^{-3} n = -\log 0.05 = 2.996$$

$$n = \frac{2.996}{8.073 \times 10^{-3}} = 371.11$$

quindi $n = 372$.

viii) Una prima osservazione è che la deviazione standard di un'esponenziale è pari alla media. Ma nel nostro campione vale $S = 19.798$ contro $\overline{x} = 123.87$. Questo è un grosso sintomo di non esponenzialità. Poi, confrontiamo le tre proporzioni relative ai dati inferiori a 110, tra 110 e 140, superiori a 140:

$$\text{sperimentali: } 0.25, 0.55, 2$$

$$\text{gaussiane: } 0.242, 0.549, 0.209$$

$$\text{esponenziali: } 0.588, 0.089, 0.323$$

essendo $\frac{5}{20} = 0.25$, $\frac{11}{20} = 0.55$, $\frac{4}{20} = .2$,

$$P(N_{gauss} \leq 110) = \Phi\left(\frac{110 - 123.87}{19.798}\right) = \Phi(-0.70) = 1 - 0.758 = 0.242$$

$$P(N_{gauss} \leq 140) = \Phi\left(\frac{140 - 123.87}{19.798}\right) = \Phi(0.8147) = 0.791$$

$$P(N_{\mathrm{exp}} \leq 110) = 1 - e^{-8.073 \times 10^{-3} \cdot 110} = 0.588$$

$$P(N_{\mathrm{exp}} \leq 140) = 1 - e^{-8.073 \times 10^{-3} \cdot 140} = 0.677.$$

Si vede ad occhio che le proporzioni gaussiane sono molto più vicine. Volendo quantificare questa impressione con un test statistico, si può calcolare il valore di chi-quadro nei due casi, trovando che il valore gaussiano è molto inferiore a quello esponenziale. Addirittura, il valore gaussiano supera il test chi quadro: il test non rifiuta l'ipotesi gaussiana. Prendiamo quindi come ragionevole la stima del numero di posti trovata con l'ipotesi gaussiana.

ix) Avremmo $N \sim N\left(138.4, 19.798^2\right)$, quindi

$$P(N > 160) = 1 - \Phi\left(\frac{160 - 138.4}{19.798}\right) = 1 - \Phi(1.091) = 1 - 0.8621 = 0.138.$$

x) Si tratta di un test unilaterale destro; si calcola $z = \frac{138.4-123.87}{19.798}\sqrt{20} = 3.282$ e lo si confronta con un quantile del tipo $q_{1-\alpha}$. Il valore p è la probabilità che Z assuma valori più estremi di quello sperimentale, ovvero

$$P\left(Z > 3.282\right) = 1 - \Phi\left(3.282\right) = 1 - 0.9995 = 0.0005.$$