Politecnico di Milano
Facoltà di Ingegneria Industriale

I Prova in Itinere di Statistica Matematica A per Ingegneria ENG
03 Maggio 2004

Nome e cognome:
Numero di matricola:

Problema 1. La ditta Chip&Chip sta progettando una nuova struttura elettronica composta da 4 componenti identici ed indipendenti collegati in parallelo. Ciascun componente è a sua volta composto da n micro-componenti identici ed indipendenti collegati in serie, ognuno di affidabilità pari ad a. Si stabilisca il valore da attribuire ad a in funzione di n affinché l'affidabilità totale dell'intera apparecchiatura sia superiore a 0.9.

Soluzione L'affidabilità di ognuno dei 4 componenti è $f(a) = a^n$, da cui l'affidabilità complessiva del sistema risulta $A(a) = 1 - (1 - f(a))^4 \equiv 1 - (1 - a^n)^4$. Impostiamo quindi la disequazione $A(a) \geq 0.9$ da risolversi nell'intervallo $[0, 1]$. La soluzione finale è $a \geq (1 - \sqrt[4]{0.1})^{1/n}$.

Problema 2. La seguente tabella sintetizza i dati relativi alla variazione percentuale tra le calorie osservate e quelle realmente presenti in un campione di 40 cibi differenti.

Classi	Freq. Assoluta
[-60;-30]	0
(-30;0]	10
(0;30]	20
(30;60]	6
(60;90]	1
(90;120]	1
(120;150]	1
(150;180]	0
(180;210]	0
(210;240]	0
(240;270]	1

L'istogramma associato alla tabella è il seguente:

(a) Si calcoli il valore medio delle variazioni tra calorie osservate e calorie dichiarate e la loro varianza.

(b) Si calcoli la mediana delle variazioni tra calorie osservate e calorie dichiarate.

(c) In questo campione di 40 cibi, sappiamo che alcuni sono di produzione nazionale (N), alcuni di produzione regionale (R), ed altri locale (L). Analizzando i seguenti boxplot affiancati costruiti sulla base di tali dati, quali conclusioni possiamo trarre per quanto riguarda la distribuzione della variazione tra calorie ossevate e calorie dichiarate nelle tre sottopopolazioni di cibi (Nazionale, Regionale e Locale)?

Soluzione

a) Siano m e S^2 la media e la varianza campionarie relative al campione riassunto nella tabella di distribuzione di frequenza. Allora, detti $\{x_i\}$ i valori medi delle classi,

$$m = \frac{1}{40}\sum_{i=1}^{11} F^{ass}(i)x_i = \frac{10\cdot(-15)+20\cdot(15)+6\cdot 45+75+95+105+135+255}{40} = \frac{99}{4} = 24.75$$

$$S^2 = \frac{1}{39}\left(\sum_{i=1}^{11} x_i^2 F^{ass}(i) - 40m^2\right) \approx 2417.9.$$

b) La mediana evidentemente appartiene alla classe $(0,30]$ essendo la frequenza cumulativa campionaria $F(30) = 3/4 \geq 1/2 > 1/4 = F(0)$. Un'approssimazine del valore della mediana si ha interpolando linearmente tra i valori estremi dell'intervallo ottenendo 15.

c) Facciamo solo alcune delle considerazioni possibili. La variabilità della variazione percentuale tra calorie osservate e calorie dichiarate è grande per i prodotti locali, e modesta per quelli di produzione nazionale. Il valore mediano della variazione percentuale per i prodotti Nazionali è prossima 0, mentre è positiva per quelli a produzione Regionale, e positiva e grande (circa 75%) per quelli a prouzione locale; ciò evidenzia che per più del 50% dei prodotti Locali, le calorie osservate sono più del 75% di quelle dichiarate. Più del 75% dei prodotti Locali e Regionali dichiarano meno calorie di quelle realmente presenti. Infine la distribuzione della variazione percentuale per i prodotti Nazionali e Regionali risultano abbastanza simmetriche, mentre la distribuzione della variazione percentuale per i prodotti Locali presenta un asimmetria visualizzata da una coda a destra. Si noti la presenza di un outlier inferiore nel boxplot relativo alla variazione percentuale dei prodotti Nazionali.

Problema 3. Sia X una variabile aleatoria continua di cui si conosce la funzione di densità a meno di una costante $k > 0$:

$$f(x) = \begin{cases} k\left(4 - x^2\right), & \text{se } -2 < x < 2 \\ 0, & \text{altrove} \end{cases}$$

(a) Si determini $k > 0$ tale che f sia effettivamente una funzione di densità.

(b) Si calcolino media e varianza di X.

(c) Si determini per quali valori di $a > 0$, $b \in \mathbb{R}$ la variabile aleatoria Y definita come $Y = aX + b$ ha media e varianza pari a 1

Soluzione

(a) f_X deve soddisfare le seguenti condizioni: $f_X(x) \geq 0$ per ogni $x \in \mathbb{R}$ e $\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$.
La prima condizione è sempre soddisfatta mentre per la seconda

$$\int_{-\infty}^{+\infty} f_X(x)\,dx = \int_{-2}^{2} k\left(4 - x^2\right) dx = k\left(4x - \frac{x^3}{3}\right)\Bigg|_{-2}^{2} = \frac{32}{3}k,$$

quindi $\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$ se e solo se $k = \frac{3}{32}$.

(b) Il calcolo della media è di fatto inutile se si osserva che la densità è una funzione pari, pertanto $E(X) = 0$. In ogni caso il calcolo esplicito è $E(X) = \int_{-\infty}^{+\infty} x f_X(x)\,dx = \frac{3}{32}\int_{-2}^{2} x\left(4 - x^2\right) dx = \frac{3}{32}\left(2x^2 - \frac{x^4}{4}\right)\Big|_{-2}^{2} = 0$

$V(X) = \int_{-\infty}^{+\infty} \left(x^2 - E(X)\right) f_X(x)\,dx = \frac{3}{32}\int_{-2}^{2} x^2\left(4 - x^2\right) dx = \frac{3}{32}\left(\frac{4}{3}x^3 - \frac{x^5}{5}\right)\Big|_{-2}^{2} = \frac{4}{5}$

(c) Abbiamo che $Y = aX + b$ ha media

$$E(Y) = aE(X) + b = b$$

e varianza

$$V(Y) = a^2 V(X) = \frac{4}{5}a^2.$$

Affinché Y abbia media e varianza pari a 1, a e b devono valere $a = \frac{\sqrt{5}}{2}$ (siccome per ipotesi $a > 0$) e $b = 1$.

Problema 4. In un colorificio viene prodotta una vernice speciale ottenuta come miscela di colorante e diluente che vengono versati in quantità opportune nei barattoli da mettere in commercio. Le proporzioni desiderate di colorante e diluente sono 1:2; cioè ogni barattolo dovrebbe contenere 1/3 di colorante che chiamaremo C e 2/3 di diluente che chiamaremo D. In pratica le quantità C e D versate in ogni barattolo (espresse in litri) sono variabili aleatorie indipendenti e $C \sim N(\mu, \sigma^2)$ mentre $D \sim N(2\mu, \sigma^2)$.

(a) Determinare la legge di $2C - D$.

(b) Qual è il valore massimo ammissibile per σ^2 affinchè, con probabilità maggiore o uguale a 0.9, la differenza tra $2C$ e D non superi in valore assoluto 0.1.

Soluzione Ricordiamo che se $\{X_i\}_{i=1}^n$ sono variabili indipendenti gaussiane tali che $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ allora

$$\sum_{i=1}^n (a_i X_i + b_i) \sim \mathcal{N}\left(\sum_{i=1}^n (a_i \mu_i + b_i), \sum_{i=1}^n a_i^2 \sigma_i^2\right).$$

a) $2C - D \sim \mathcal{N}(0, 5\sigma^2)$.

b)

$$0.9 \leq \mathbb{P}(|2C - D| \leq 0.1) = \mathbb{P}\left(\frac{|2C - d|}{\sqrt{5}\sigma} \leq \frac{0.1}{\sqrt{5}\sigma}\right)$$

che equivale a

$$\Phi\left(\frac{0.1}{\sqrt{5}\sigma}\right) \geq \frac{1 + 0.9}{2}$$

da cui

$$\sigma \leq \frac{0.1}{\sqrt{5} z_{0.95}} \approx 0.0272.$$

Problema 5. Nella fascia oraria 22-24, il numero di connessioni ad Internet alla linea di un provider è descritto da un processo di Poisson con una media di 16 connessioni al minuto.

(a) Qual è la probabilità che, in quella fascia oraria, vi siano più di 3 connessioni alla linea in due minuti?

(b) Qual è la probabilià che trascorrano più di 5 secondi tra una connessione e la successiva?

(c) Supponiamo ora che il provider gestisca 50 linee ognuna delle quali, indipendentemente dalle altre, nella fascia oraria 22-24 riceve un numero di connessioni descritto da un processo di Poisson con media di 16 connessioni al minuto.

Si calcoli (in modo approssimato) la probabilità che, nella fascia oraria 22-24, il numero totale di connessioni al provider in un minuto sia superiore a 1000.

Soluzione

(a) Se indichiamo con X la v.a. che conta il numero di connessioni in due minuti, allora X ha una distribuzione di Poisson di parametro pari a 32. Quindi la probabilità richiesta è data da:

$$\begin{aligned} P(X \quad > \quad 3) = 1 - P(X \leq 3) = 1 - \sum_{k=0}^{3} e^{-32} \frac{(32)^k}{k!} = \\ = \quad 1 - e^{-32}\left(1 + 32 + \frac{32^2}{2} + \frac{32^3}{6}\right) \simeq 1 \end{aligned}$$

(b) Siccome il numero di connessioni segue un processo di Poisson, allora il tempo di attesa T tra una connessione e la successiva segue una legge esponenziale di parametro λ. Dato che il numero medio di connessioni in un minuto è pari a 16, allora la media del tempo di attesa (calcolato in minuti) è pari a 1/16, quindi in secondi 60/16=4/15. Quindi $\lambda = 4/15$ per il calcolo in secondi.

La probabilità cercata è data quindi da

$$P(T_{\text{sec}} > 5) = 1 - P(T_{\text{sec}} \leq 5) = e^{-5\lambda} = e^{-4/3} = 0.264$$

(c) Indicando con X_i il numero di connessioni in un minuto alla i-esima linea del provider, $i = 1, ..., 50$, abbiamo che $E(X_i) = 16$; $V(X_i) = 16$:

$$P(X_1 + ... + X_{50} > 1000) = P\left(\frac{X_1 + ... + X_{50} - 50 \cdot 16}{4\sqrt{50}} > \frac{1000 - 50 \cdot 16}{4\sqrt{50}}\right)$$

che, utilizzando il Teorema Centrale del Limite, possiamo approssimare con:

$$\begin{aligned} P\left(\frac{X_1 + ... + X_{50} - 50 \cdot 16}{4\sqrt{50}} > \frac{1000 - 50 \cdot 16}{4\sqrt{50}}\right) \quad &\simeq \quad 1 - \Phi\left(\frac{1000 - 50 \cdot 16}{4\sqrt{50}}\right) = \\ &= \quad 1 - \Phi\left(\sqrt{50}\right) \simeq 0 \end{aligned}$$