

Corso di Laurea in Informatica Applicata
Esame di Calcolo delle Probabilità e Statistica
Prova scritta dell' 11 gennaio 2007

Primo esercizio. Per una certa stampante S_1 , la probabilità che un generico foglio si inceppi durante la stampa, indipendentemente dagli altri, vale $p = 0.05$.

- (a) Qual è la probabilità di riuscire a stampare un documento di 10 pagine?
- (b) Di quante pagine deve essere composto, al massimo, un documento, affinché la probabilità di riuscire a portarne a termine la stampa non sia inferiore a 0.90?
- (c) Supponiamo ora di avere a disposizione, oltre a S_1 , una seconda stampante S_2 , di prestazioni identiche alla prima, e che lavora indipendentemente da essa. Se S_1 si inceppa prima di aver finito di stampare un documento, la stampa della parte rimanente viene trasferita su S_2 . Qual è ora la probabilità di riuscire a completare la stampa di un documento di 10 pagine?
- (d) Sapendo che la stampa del documento del punto (c) è stata completata, quanto vale la probabilità che S_1 si sia inceppata?
- (e) Ancora nella situazione del punto (c), sia N il numero di stampanti impegnate nella stampa del documento. Calcolare la densità di N e la sua media.

Secondo esercizio. Sia X una v. a. di legge uniforme sull'intervallo $[-1, 1]$.

- (a) Calcolare la legge della v. a. $Y = |X|$. Si tratta di una legge nota?
- (b) Posto $Z = X^2 + |X|$, mostrare che $P(0 \leq Z \leq 2) = 1$.
- (c) Calcolare la legge della v. a. Z .

SUGGERIMENTO PER IL PUNTO (c): si consiglia di abbreviare i calcoli sfruttando i risultati dei punti (a) e (b).

Terzo esercizio. In un determinato tratto di autostrada con limite di velocità 120 km/h, la polizia stradale ha rilevato la velocità di un campione casuale di 81 automobili. La velocità media nel campione è stata di $\bar{x} = 132$ km/h, con $s = 20$ (s è la deviazione standard ricavata dai dati).

- (a) Calcolare un intervallo di fiducia bilaterale di livello 0.95 per la velocità media delle automobili. Cosa conclude la polizia?
 (b) Quale dovrebbe essere l'ampiezza del campione per ottenere un intervallo di ampiezza non superiore a 8 (km/h)?

SOLUZIONI

- (a) Per $i = 1, 2, \dots, 10$ consideriamo gli eventi

$$A_i = \{\text{il foglio } i\text{-esimo viene stampato}\}.$$

Gli $(A_i)_{1 \leq i \leq 10}$ costituiscono una famiglia di eventi indipendenti, ed inoltre, per ogni i , si ha $P(A_i) = 0.95$.

Posto allora

$$E = \{\text{la stampa del documento viene completata}\},$$

si ha evidentemente $E = A_1 \cap A_2 \cap \dots \cap A_{10}$, e quindi, per l'indipendenza

$$P(E) = \prod_{i=1}^{10} P(A_i) = (0.95)^{10} = 0.5987.$$

OSSERVAZIONE 1. Proponiamo una soluzione alternativa, basata su tecniche di variabili aleatorie. Per ogni $i = 1, 2, 3, \dots$ consideriamo la v. a.

$$X_i = \begin{cases} 0 & \text{se } S_1 \text{ stampa l}'i\text{-esimo foglio} \\ 1 & \text{in caso contrario.} \end{cases}$$

Le v. a. (X_i) costituiscono evidentemente una successione di v. a. indipendenti e tutte di legge $B(1, p)$, con $p = 0.05$, e descrivono una successione di prove indipendenti e bernoulliane, in cui il “successo” corrisponde all'inzeppamento della stampante.

Poniamo

$$Y = \text{numero di fogli del documento stampati fino all'inzeppamento.}$$

Si vede facilmente che Y è una v. a. che assume i valori interi $0, 1, 2, \dots, 10$; inoltre $P(Y = 0) = P(X_1 = 1) = p$; invece, per $k = 1, 2, \dots, 9$ si ha, per l'indipendenza delle (X_i) ,

$$\begin{aligned}
P(Y = k) &= P(X_1 = 0, X_2 = 0, \dots, X_k = 0, X_{k+1} = 1) \\
&= P(X_1 = 0)P(X_2 = 0) \dots P(X_k = 0)P(X_{k+1} = 1) = (1 - p)^k p.
\end{aligned}$$

Infine

$$\begin{aligned}
P(Y = 10) &= P(X_1 = 0, X_2 = 0, \dots, X_{10} = 0) \\
&= P(X_1 = 0)P(X_2 = 0) \dots P(X_{10} = 0) = (1 - p)^{10}.
\end{aligned}$$

In conclusione

$$P(Y = k) = \begin{cases} (1 - p)^k p & \text{per } k = 0, 1, \dots, 9 \\ (1 - p)^{10} & \text{per } k = 10. \end{cases}$$

Si osservi che il punto (a) dell'esercizio chiede proprio $P(Y = 10)$.

(b) Ragionando come nel punto (a) precedente (primo o secondo metodo), si vede che la probabilità di riuscire a completare la stampa di un documento di n pagine vale $(0.95)^n$. Si tratta dunque di risolvere in termini di n la disuguaglianza $(0.95)^n \geq 0.90$. Passando ai logaritmi si trova facilmente (attenzione, $\log 0.95 < 0$!)

$$n \leq \frac{\log 0.90}{\log 0.95} = 2.058$$

Dunque si può stampare un documento di sole 2 pagine.

(c) Per $i = 1, 2, \dots, 10$, consideriamo ora gli eventi

$$A_i = \{\text{il foglio } i\text{-esimo viene stampato da } S_1\}.$$

$$B_i = \{\text{il foglio } i\text{-esimo viene stampato da } S_2\}.$$

Si tratta di nuovo di eventi tutti tra loro indipendenti, ed inoltre, per ogni i , si ha $P(A_i) = P(B_i) = 0.95$.

Per ogni $k = 1, 2, \dots, 10$ poniamo

$$C_k = \{S_1 \text{ si inceppa al } k\text{-esimo foglio e la stampa viene completata da } S_2\};$$

Poniamo poi

$$\begin{aligned} C_{11} &= \{S_1 \text{ non si inceppa mai nel corso della stampa del documento}\} \\ &= \{\text{la stampa viene completata da } S_1\}; \end{aligned}$$

L'evento di cui si cerca la probabilità è evidentemente

$$C = \left(\bigcup_{k=1}^{11} C_k \right);$$

gli eventi $C_1, C_2, \dots, C_{11}, E$ sono due a due disgiunti, e quindi

$$P(C) = \sum_{k=1}^{11} P(C_k);$$

dal punto (a) sappiamo che $P(C_{11}) = 0.5987$; non resta che calcolare $P(C_k)$ per ogni $k = 1, 2, \dots, 10$. È facile vedere che

$$C_k = A_1 \cap A_2 \cap \dots \cap A_{k-1} \cap A_k^c \cap B_1 \cap \dots \cap B_{10-k+1},$$

(spiegazione: siccome S_1 si inceppa al k -esimo foglio, allora S_1 ha stampato tutti i primi $k-1$ fogli; siccome la stampa viene completata da S_2 , allora S_2 ha stampato i rimanenti $10 - (k-1) = 10 - k + 1$ fogli). Per l'indipendenza di tutti gli eventi in questione, si deduce

$$\begin{aligned} P(C_k) &= \left(\prod_{i=1}^{k-1} P(A_i) \right) \times P(A_k^c) \times \left(\prod_{i=1}^{10-k+1} P(B_i) \right) \\ &= (0.95)^{k-1} \times 0.05 \times (0.95)^{10-k+1} = (0.95)^{10} \times 0.05 = 0.0299. \end{aligned}$$

Si trova quindi

$$P(C) = 10 \times 0.0299 + 0.5987 = 0.299 + 0.5987 = 0,898.$$

OSSERVAZIONE 2. In alternativa, il calcolo si può fare così: $P(C) = 10 \times (0.95)^{10} \times 0.05 + (0.95)^{10} = (0.95)^{10} \times 1.5 = 0,898$.

OSSERVAZIONE 3. È interessante osservare che la quantità ottenuta è del tipo $R_n = nq^n(1 - q) + q^n = q^n(n(1 - q) + 1)$ (con $q = 0.95$ e $n = 10$). Eventualmente, per controllo, si può verificare che questa quantità è effettivamente ≤ 1 . La dimostrazione più semplice è per induzione su n : la cosa è banalmente vera per $n = 1$, dato che $R_1 = q$. Per fare il passo induttivo, supponiamo $R_n \leq 1$ e osserviamo che

$$\begin{aligned} R_{n+1} &= q^{n+1}((n+1)(1-q) + 1) = q(R_n + q^n(1-q)) \\ &\leq q(1 + q^n(1-q)) = q + q^{n+1}(1-q) \leq q + (1-q) = 1. \end{aligned}$$

(d) Utilizziamo i simboli introdotti nel punto precedente. Si chiede di calcolare

$$\begin{aligned} P(\cup_{k=1}^{10} C_k | C) &= \frac{P((\cup_{k=1}^{10} C_k) \cap C)}{P(C)} \\ &= \frac{P(\cup_{k=1}^{10} C_k)}{P(C)} = \frac{10 \times (0.95)^{10} \times 0.05}{(0.95)^{10} \times 1.5} = \frac{10 \times 0.05}{1.5} = \frac{1}{3} = 0.333. \end{aligned}$$

(e) N può prendere i soli valori 1 e 2. Dal punto (a) si ricava $P(N = 1) = (0.95)^{10} = 0.5987$.

Per differenza si ottiene $P(N = 2) = 1 - 0.5987 = 0.4013$.

Pertanto $\mathbf{E}[N] = 1 \times 0.5987 + 2 \times 0.4013 = 1.4013$.

Secondo esercizio. (a) Calcoleremo la funzione di ripartizione di Y . Dato che $Y = |X|$ è una v. a. non negativa, per $t < 0$ intanto si ha $P(Y \leq t) = 0$. Per $t \geq 0$ si può scrivere

$$P(Y \leq t) = P(|X| \leq t) = P(-t \leq X \leq t) = F(t) - F(-t),$$

dove con F si indica la f. d. r. della v. a. X , e cioè, per una formula nota

$$F(u) = \begin{cases} 0 & \text{per } u < -1 \\ \frac{u+1}{2} & \text{per } -1 \leq u < 1 \\ 1 & \text{per } u \geq 1. \end{cases}$$

Continuando il calcolo si trova facilmente, per $t \geq 0$,

$$F(t) - F(-t) = \begin{cases} \frac{t+1}{2} - \frac{-t+1}{2} = t & \text{per } 0 \leq t \leq 1 \\ 1 & \text{per } t > 1. \end{cases}$$

In conclusione

$$G(t) = P(Y \leq t) = \begin{cases} 0 & \text{per } t < 0, \\ \frac{t+1}{2} - \frac{-t+1}{2} = t & \text{per } 0 \leq t \leq 1 \\ 1 & \text{per } t > 1, \end{cases}$$

e si riconosce quindi che $Y \sim \mathcal{U}[0, 1]$ (uniforme sull'intervallo $[0, 1]$).

(b) Dato che

$$\{-1 \leq X \leq 1\} = \{|X| \leq 1\} = \{X^2 \leq 1\},$$

si ha

$$\{-1 \leq X \leq 1\} = \{|X| \leq 1\} \cap \{X^2 \leq 1\} \subseteq \{Z \leq 2\}$$

e quindi

$$1 = F(1) - F(-1) = P(-1 \leq X \leq 1) \leq P(Z \leq 2) \leq 1.$$

(c) Calcoliamo la f. d. r. di Z . Per quanto detto nel punto precedente, si ha subito

$$P(Z \leq t) = \begin{cases} 0 & \text{per } t < 0, \\ 1 & \text{per } t > 2. \end{cases}$$

Per $0 \leq t \leq 2$, osservando che $X^2 = Y^2$ si può scrivere

$$P(Z \leq t) = P(Y^2 + Y \leq t) = P(Y^2 + Y - t \leq 0).$$

Risolvendo la disequazione $Y^2 + Y - t \leq 0$, si trova facilmente

$$\{Y^2 + Y - t \leq 0\} = \left\{ \frac{-1 - \sqrt{1+4t}}{2} \leq Y \leq \frac{-1 + \sqrt{1+4t}}{2} \right\}.$$

Dato che $0 \leq Y \leq 1$, l'evento precedente si può scrivere anche nella forma

$$\begin{aligned} & \left\{ \frac{-1 - \sqrt{1+4t}}{2} \leq Y \leq \frac{-1 + \sqrt{1+4t}}{2}, 0 \leq Y \leq 1 \right\} \\ & = \left\{ 0 \leq Y \leq \frac{-1 + \sqrt{1+4t}}{2}, Y \leq 1 \right\} \end{aligned}$$

poiché $\frac{-1 - \sqrt{1+4t}}{2} < 0$.

Osserviamo che il numero $\frac{-1+\sqrt{1+4t}}{2}$ è non negativo per ogni valore di t , mentre la disuguaglianza

$$\frac{-1+\sqrt{1+4t}}{2} \leq 1$$

è verificata esattamente per $t \leq 2$. Possiamo allora concludere che (ricordiamo che stiamo considerando esattamente i valori di t minori o uguali a 2)

$$\{Y^2 + Y - t \leq 0\} = \left\{0 \leq Y \leq \frac{-1+\sqrt{1+4t}}{2}\right\},$$

e di conseguenza (ricordare che $Y \sim \mathcal{U}[0, 1]$)

$$\begin{aligned} P(Z \leq t) &= P(Y^2 + Y - t \leq 0) = P\left(0 \leq Y \leq \frac{-1+\sqrt{1+4t}}{2}\right) \\ &= G\left(\frac{-1+\sqrt{1+4t}}{2}\right) = \frac{-1+\sqrt{1+4t}}{2}. \end{aligned}$$

Ricapitolando

$$P(Z \leq t) = \begin{cases} 0 & \text{per } t < 0, \\ \frac{-1+\sqrt{1+4t}}{2} & \text{per } 0 \leq t \leq 2 \\ 1 & \text{per } t > 2. \end{cases}$$

Terzo esercizio. (a) Poiché la varianza è ricavata dai dati, quindi non nota, utilizzeremo l'intervallo

$$\left(\bar{X} - t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}}\right).$$

I dati sono $\bar{x} = 132$, $s = 20$, $n = 81$, $1 - \alpha = 0.95$, da cui $1 - \alpha/2 = 0.975$ e (dalle tavole dei quantili della t di Student a 80 gradi di libertà) $t_{0.975}(80) = 1.9901$. Sostituendo nella formula precedente, si trova l'intervallo $(127.58, 132.42)$, e la polizia conclude con fiducia del 95% che in quel tratto di autostrada il limite di 120 km/hr è superato.

(b) La lunghezza dell'intervallo precedente è data da

$$\left(\bar{X} + t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}}\right) - \left(\bar{X} - t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}}\right) = 2t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}},$$

e quindi si tratta di trovare n in modo che valga la disuguaglianza

$$2t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}} \leq 8,$$

che equivale a

$$(*) \quad \frac{t_{1-\alpha/2}(n-1)}{\sqrt{n}} \leq \frac{8}{40} = 0.2.$$

I dati sono ancora $1 - \alpha/2 = 0.975$, $s = 20$, ma questa volta non possiamo usare come valore del quantile $t_{0.975}(n-1) = 1.9901$, perché tale valore corrisponde a $n = 81$, mentre adesso n è incognito.

Usando le tavole dei quantili della t di Student a nostra disposizione con diversi valori di n , si vede che la disuguaglianza precedente è verificata per $n = 121$, ma non lo è ad esempio per $n = 81$. Per valori intermedi ($81 \leq n \leq 121$), non abbiamo le tavole a disposizione ma, osservando che il quantile è una funzione decrescente di n , si può ragionare così

$$1.9799 = t_{0.975}(120) < t_{0.975}(n-1) < t_{0.975}(80) = 1.9901,$$

e quindi

$$\frac{1.9799}{\sqrt{n}} < \frac{t_{0.975}(n-1)}{\sqrt{n}} < \frac{1.9901}{\sqrt{n}}.$$

La disuguaglianza (*) sarà dunque verificata se

$$\frac{1.9901}{\sqrt{n}} \leq 0.2,$$

cioè, risolvendo, se $n \geq 99.01$. Pertanto la (*) vale per $n \geq 100$. D'altra parte, la stessa (*) non può essere valida se

$$\frac{1.9799}{\sqrt{n}} \geq 0.2,$$

cioè, risolvendo, per $n \leq 98.0001$. Dunque (*) non vale per $n \leq 98$. A questo punto, l'unico valore su cui rimane incertezza (e per il quale dovremmo avere tavole più raffinate od usare al computer un pacchetto statistico) è $n = 99$. E qui dobbiamo fermarci.