

Corso di Laurea in Informatica - a.a. 2003/04

Calcolo delle Probabilità e Statistica

Prova scritta del III appello - 28/5/2004

Il candidato risolva i tre problemi proposti, motivando le proprie risposte.

1. In un dato anno, gli studenti di Informatica che hanno superato l'esame di CPS sono così ripartiti tra i tre corsi A, B, C: il 30% degli studenti ha frequentato il corso A, il 36% il corso B e i restanti il corso C. A fine anno, la proporzione di studenti che ha superato l'esame di CPS con voto maggiore o uguale a 27 è del 10 %. Tuttavia, tra gli studenti del corso A, questa proporzione è solo dell'8%.
 - (a) Qual è la probabilità che uno studente che ha frequentato il corso B o il corso C abbia preso un voto maggiore o uguale a 27?
 - (b) Qual è la probabilità che uno studente che ha preso un voto maggiore o uguale a 27 abbia frequentato uno dei corsi B o C?
2. La durata di vita (misurata in anni) di un certo componente può essere modellizzata con una v.a. avente densità

$$f(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{1}{2} & \text{se } 0 \leq x < 1 \\ \frac{1}{x^3} & \text{se } x \geq 1 \end{cases}$$

- (a) Calcolare la vita media del componente.
 - (b) Se dopo un anno il componente è ancora funzionante, qual è la probabilità che lo sia dopo 2 anni?
 - (c) Un certo impianto contiene 5 di questi componenti, che funzionano indipendentemente l'uno dall'altro. Qual è la probabilità che dopo due anni esattamente 3 componenti siano funzionanti?
3. La quantità di principio attivo presente in un certo farmaco è soggetta ad una certa variabilità, ma la ditta che lo produce dichiara una quantità media di 20mg per pillola. Supponendo che questa quantità possa essere modellizzata come una v.a. normale, vengono analizzate 25 pillole, dalle quali risulta una media campionaria di $\bar{x} = 19.7$ mg e una somma degli scarti quadratici pari a $\sum_{i=1}^{25} (x_i - \bar{x})^2 = 40.5$.
 - (a) Calcolare una stima della varianza.
 - (b) Si effettui il test di ipotesi $H_0 : \mu = \mu_0$. Si può affermare, a livello 5%, che la dichiarazione della ditta è falsa?
 - (c) Si supponga di conoscere la varianza reale $\sigma^2 = 1.56$ e di aver ottenuto la stessa media campionaria $\bar{x} = 19.7$ su un campione di n pillole. Qual è il minimo n per poter rifiutare l'ipotesi del test $H_0 : \mu = \mu_0$ a livello 5%?

Calcolo delle Probabilità e Statistica

Soluzioni della prova scritta del III appello - 28/5/2004

1. Denotiamo con A l'evento { lo studente ha frequentato il corso A } e con B e C gli eventi definiti in modo analogo. Sia poi V l'evento { lo studente ha preso un voto maggiore o uguale di 27 }. I dati in nostro possesso sono: $P(A) = 0.30, P(B) = 0.36, P(V) = 0.10, P(V|A) = 0.08$.

- (a) Dobbiamo calcolare $P(V|B \cup C)$. Posto $D = B \cup C$, osserviamo che A e D formano una partizione. Quindi per la legge delle probabilità totali, si ha:

$$P(V) = P(V|A)P(A) + P(V|D)P(D).$$

Dato che $P(D) = 1 - P(A) = 0.7$, si ricava $P(V|D) = \frac{0.10 - 0.08 \cdot 0.30}{0.7} = 0.108$

- (b) Dobbiamo calcolare $P(B \cup C|V) = P(D|V)$ con le notazioni di prima. Per la formula di Bayes, avremo

$$P(D|V) = \frac{P(V|D)P(D)}{P(V)},$$

da cui ricaviamo $P(D|V) = \frac{0.108 \cdot 0.7}{0.1} = 0.756$. In modo alternativo, si può osservare che, dato che D è il complementare di A , si ha: $P(D|V) = 1 - P(A|V)$ e, sempre per la formula di Bayes, si ha $P(A|V) = \frac{P(V|A)P(A)}{P(V)}$.

2. Denotiamo con X la v.a. che modella la durata di vita di un generico componente.

- (a) Dobbiamo calcolare $E[X]$. Per cominciare, dovremmo provare che X ha speranza matematica finita, cioè che $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$. D'altra parte, dato che la densità di X è non nulla solo per $x \geq 0$, avremo che $\int_{-\infty}^{\infty} |x|f(x)dx = \int_0^{\infty} x|f(x)dx = \int_0^{\infty} xf(x)dx = E[X]$ e quindi possiamo contestualmente calcolare la speranza e verificare che sia finita. Quindi:

$$\begin{aligned} E[X] &= \int_0^1 x \frac{1}{2} dx + \int_1^{\infty} x \frac{1}{x^3} dx = \frac{1}{2} \int_0^1 x dx + \int_1^{\infty} \frac{1}{x^2} dx \\ &= \frac{1}{2} \left[\frac{x^2}{2} \right]_0^1 + [-x^{-1}]_1^{\infty} = \frac{1}{4} + 1 = \frac{5}{4}. \end{aligned}$$

- (b) La probabilità richiesta è $P(X > 2|X > 1)$ che, per definizione di probabilità condizionata, è pari a $\frac{P(X > 2, X > 1)}{P(X > 1)}$. Dato che l'evento $(X > 2)$ è contenuto nell'evento $(X > 1)$, si ha $P(X > 2, X > 1) = P(X > 2)$, da cui ricaviamo:

$$P(X > 2|X > 1) = \frac{P(X > 2)}{P(X > 1)}$$

(si poteva anche osservare che $P(X > 2|X > 1) = \frac{P(X > 1|X > 2)P(X > 2)}{P(X > 1)}$ per la formula di Bayes e che banalmente $P(X > 1|X > 2) = 1$). Ora abbiamo:

$$P(X > 1) = 1 - P(X \leq 1) = 1 - \int_0^1 \frac{1}{2} dx = 1 - \frac{1}{2} = \frac{1}{2}$$

$$P(X > 2) = \int_2^\infty \frac{1}{x^3} dx = \left[-\frac{x^{-2}}{2} \right]_2^\infty = \frac{1}{8}$$

da cui ricaviamo $P(X > 2|X > 1) = \frac{1}{4}$.

- (c) Denotiamo con X_i ($i = 1, \dots, 5$) la durata di vita dell' i -esimo componente. Ogni X_i ha la stessa densità di X e le X_i sono fra loro indipendenti. Definiamo per ogni i una nuova variabile aleatoria Y_i nel modo seguente:

$$Y_i = \begin{cases} 1 & \text{se } X_i > 2 \\ 0 & \text{altrimenti} \end{cases}$$

È evidente che ogni Y_i è una v.a. bernoulliana di parametro $p = P(X_i > 2) = P(X > 2) = 1/8$ e che le Y_i sono indipendenti. Definiamo la v.a. $N = Y_1 + \dots + Y_5$ che conta il numero di componenti funzionanti dopo 2 anni. La v.a. N ha distribuzione $\mathcal{B}(5, 1/8)$, quindi la probabilità da calcolare è

$$P(N = 3) = \binom{5}{3} \left(\frac{1}{8}\right)^3 \left(1 - \frac{1}{8}\right)^2 = 0.015$$

3. (a) Una stima corretta della varianza è $S^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / (n - 1) = 40.5/24 = 1.69$.
 (b) Dobbiamo effettuare un test bilatero sulla media di un campione gaussiano. Poiché non conosciamo la varianza del campione, dobbiamo usare il test di Student. La regione critica sarà della forma

$$\left\{ \left| \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \right| > t_{1-\alpha/2}(n-1) \right\}.$$

Con i dati a disposizione, $n = 25$, $\bar{x} = 19.7$, $\mu_0 = 20$, $S^2 = 1.69$, $\alpha = 0.05$, si verifica che: $\left| \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \right| = 1.15 < t_{0.975}(24) = 2.06$ quindi non si può rifiutare l'ipotesi. In modo alternativo, si può osservare che la regione di accettazione è

$$D^c = \left[\mu_0 - \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1), \mu_0 + \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1) \right] = [19.46, 20.54]$$

quindi l'osservazione $\bar{x} = 19.7$ appartiene alla regione di accettazione. Non possiamo dunque affermare che la dichiarazione della ditta è falsa.

- (c) In questo caso conosciamo la varianza, dunque possiamo usare la distribuzione normale (e i suoi quantili) per calcolare la regione critica del test che sarà della forma:

$$\left\{ \left| \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \right| > \phi_{1-\alpha/2} \right\}.$$

Per poter rifiutare l'ipotesi è dunque necessario che la disuguaglianza di sopra sia verificata, cioè:

$$\frac{n(\bar{X} - \mu_0)^2}{\sigma^2} > \phi_{1-\alpha/2}^2,$$

da cui si ricava $n > \frac{\sigma^2 \phi_{1-\alpha/2}^2}{(\bar{X} - \mu_0)^2}$ che, con i dati a disposizione (ricordando che $\phi_{0.975} = 1.96$) dà $n > 66.59$. Quindi il minimo n che permette il rifiuto dell'ipotesi è $n = 67$.