II Prova in Itinere di Statistica Matematica A - 27 Gennaio 2003

Allievi Meccanici, II anno, sez. N-Z

Docente: L. Valdettaro

Tempo a disposizione: 2 ore

I. Domande a risposta multipla

Ogni domanda ha una sola risposta esatta. Le risposte sbagliate sono contate negativamente per un terzo del valore della risposta esatta.

1. **Se una variabile aleatoria X continua ha la proprietà seguente:**
$P(X < -t + a) = P(X > t + a)\ \forall t$**, allora detto z_α il quantile della variabile aleatoria si ha:**

☒ $z_\alpha + z_{1-\alpha} = 2a$
☐ nessuna delle altre risposte è corretta
☐ $z_\alpha + z_{1-\alpha} = 0$
☐ $z_\alpha - z_{1-\alpha} = 0$

2. **Sia X una variabile aleatoria gaussiana: $X \sim \mathcal{N}(\mu, \sigma^2)$.**
La probabilità $P(X < \mu + \sigma)$:

☐ dipende solo da σ
☐ dipende solo da μ
☐ dipende sia da μ che da σ
☒ non dipende nè da σ nè da μ

3. **Si effettua un test chi-quadrato d'indipendenza per due variabili aleatorie su un campione di dimensione n. Si considera poi un nuovo test su un campione raddoppiato in cui si suppone che le frequenze relative rimangono invariate; allora si può affermare che:**

☒ il p-value diminuisce
☐ il p-value aumenta
☐ il p-value non cambia
☐ non si hanno sufficienti informazioni per poter stabilire come cambia il p-value

4. **Consideriamo n variabili aleatorie gaussiane $X_i \sim \mathcal{N}(\mu, \sigma^2)$ indipendenti e identicamente distribuite. Sia $Y = \sum_{i=1}^{n} X_i/2^i$. Al tendere di n all'infinito la varianza di Y:**

☐ tende all'infinito
☒ tende al valore $\sigma^2/3$
☐ tende al valore $3\sigma^2$
☐ tende a zero

II. Esercizi

Riportare oltre al risultato anche (sinteticamente) il procedimento con cui esso è stato trovato. Gli esercizi dove compare solo il risultato verranno considerati nulli, anche se questo fosse esatto.

---

1. **Un campionamento su una serie di** 400 **dati ha fornito una media campionaria** $\overline{x}_n$ **e una deviazione standard campionaria** s_n**.**

1) Calcolare il livello di confidenza α **in modo che l'ampiezza dell'intervallo di confidenza per la media sia pari a** $s_n/5$**.**

2) Si può affermare che la varianza vera σ^2 **sia uguale a** $1.2s_n^2$ **con livello di significatività** $1-\alpha$**? (si usi il valore di** α **trovato al punto 1).**

Soluzione.
1.1) La soluzione si ricava invertendo la formula per il calcolo dell'ampiezza dell'intervallo di confidenza per la media con varianza incognita: $t_{\frac{1+\alpha}{2}}(n-1) = \frac{A\sqrt{n}}{2s_n} = 2$. Per il computo del quantile $t_{\frac{1+\alpha}{2}}(n-1)$, essendo $n > 120$, possiamo consultare le tavole della normale standard ottenendo $\alpha = 2\Phi(2) - 1 \simeq 0.9545$

1.2) L'ipotesi nulla da testare è $H_0 : \sigma^2 = \sigma_0^2$ con un livello di significatività $\beta = 1-\alpha$ dove $\sigma_0^2 = 1.2s_n^2$. La regione di rifiuto del test è data da: $\chi^2 > \chi^2_{1-\beta/2}(n-1) \cup \chi^2 < \chi^2_{\beta/2}(n-1)$ dove $\chi^2 = \frac{(n-1)s_n^2}{1.2s_n^2} = \frac{399}{1.2} \simeq 332.5$. Per il calcolo dei quantili, che non troviamo sulle tavole essendo $n > 30$, possiamo usare l'approssimazione $\chi^2(m) \simeq \mathcal{N}(m, 2m)$ da cui $\chi^2_\alpha(m) = q_\alpha\sqrt{2m} + m$ ottenendo

$$\chi^2_{1-\beta/2}(n-1) \simeq \chi^2_{0.9772}(399) \simeq z_{0.9772}\sqrt{798} + 399 \simeq 2 \cdot 28.2489 + 399 \simeq 455.4977$$

$$\chi^2_{\beta/2}(n-1) \simeq \chi^2_{0.0228}(399) \simeq z_{0.0228}\sqrt{798} + 399 \simeq -z_{0.9772}\sqrt{798} + 399 \simeq 342.5023$$

Oppure possiamo usare la migliore approssimazione: $\chi^2_\alpha(m) \simeq \frac{1}{2}\left(q_\alpha + \sqrt{2m-1}\right)^2$ ottenendo

$$\chi^2_{0.9772}(399) \simeq \frac{1}{2}\left(z_{0.9772} + \sqrt{797}\right)^2 \simeq \frac{1}{2}\left(2 + \sqrt{797}\right)^2 \simeq 458.0331$$

$$\chi^2_{0.0228}(399) \simeq \frac{1}{2}\left(z_{0.0228} + \sqrt{797}\right)^2 \simeq \frac{1}{2}\left(-2 + \sqrt{797}\right)^2 \simeq 344.9669$$

Da notare che con l'uso del calcolatore, senza approssimazione, avremmo ottenuto:
$\chi^2_{0.9772}(399) \simeq 457.4734$ e $\chi^2_{0.0228}(399) \simeq 344.5250$.
In tutti e tre i casi l'ipotesi nulla è da ritenersi rifiutata.

---

2. **Uno scienziato sostiene che nell'universo il 9% delle stelle ammette un sistema planetario.**

1) Supponendo vera l'ipotesi dello scienziato calcolare, facendo la dovuta approssimazione e motivandola, la probabilità che su 10000 stelle almeno un decimo di esse abbia dei pianeti.

2) Sulle 80 stelle più vicine a noi ne sono state trovate 3 con un sistema planetario. Si sottoponga a verifica l'ipotesi dello scienziato al livello di significatività del 10%.

Soluzione.
2.1) Siano $n = 10000$ e $p = 0.09$. Definiamo la v.a. X che conta quante stelle su n hanno pianeti; allora $X \sim B(n, p)$. Poiché il calcolo di

$$P(X \geq 1000) = 1 - P(X \leq 999) = 1 - \sum_{k=0}^{999} \binom{n}{k} p^k (1-p)^{n-k}$$

è impraticabile, vediamo se possiamo usare il TLC per approssimare. Le condizioni $np = 900 > 5$ e $n(1-p) = 9100 > 5$ sono soddisfatte e quindi possiamo applicare il teorema. Ricordandoci della correzione di continuità otteniamo

$$P(X \leq 999) = P\left(Z \leq \frac{999.5 - 900}{\sqrt{900 \cdot 0.91}}\right) \simeq \Phi\left(\frac{99.5}{30\sqrt{0.91}}\right) \simeq \Phi(3.4768) \simeq 0.9997$$

essendo $Z \sim \mathcal{N}(0, 1)$ da cui $P(X \geq 1000) \simeq 0.0003$

2.2) L'ipotesi nulla da testare è $H_0 : p = p_0 = 0.09$ con un livello di significatività $\alpha = 0.1$ sapendo che $n = 50$ e $\overline{x}_n = \frac{3}{80} = 0.0375$. Rifiuto il test se $|z| > z_{1-\alpha/2}$ dove $z = \frac{\overline{x}_n - p_0}{\sqrt{p_0(1-p_0)/n}}$

Poiché $z_{1-\alpha/2} = z0.95 \simeq 1.645$ e

$$|z| = \frac{|0.0375 - 0.09|}{\sqrt{0.09 \cdot 0.91/80}} \simeq \frac{0.0525}{0.0320} \simeq 1.6408$$

l'ipotesi nulla non può essere rifiutata.

---

3. **La lunghezza delle rotaie prodotte da una ditta segue approssimativamente una legge normale con media 100 metri e deviazione standard 50 centimetri.**

1) Calcolare la probabilità che una rotaia abbia una lunghezza inferiore a 99 metri.

2) La ditta ha in appalto la costruzione di una linea ferroviaria di 99.98 Km. Qual è la probabilità che 1000 rotaie siano sufficienti a coprire l'intero percorso?

Soluzione.
Sia X_i la v.a. aleatoria che misura in metri la lunghezza di una rotaia allora $X_i \sim \mathcal{N}(\mu, \sigma^2)$ con $\mu = 100$ e $\sigma = 0.5$. Inoltre indichiamo con $Z \sim \mathcal{N}(0, 1)$ una v.a. normale standard e definiamo le v.a. $S_n = \sum_{i=1}^{n} X_i$ e $\overline{X}_n = \frac{1}{n} S_n$. Ora possiamo calcolare:
3.1)

$$P(X_1 < 99) = P\left(Z < \frac{99 - 100}{0.5}\right) = \Phi(-2) = 1 - \Phi(2) \simeq 0.0228$$

3.2)

$$P(S_n \geq 99980) = P(\overline{X}_n \geq 99.98) = P\left(Z \geq \frac{99.98 - 100}{0.5/\sqrt{1000}}\right) = 1 - \Phi(-1.2649) \simeq 0.897$$

---

4. **Si vuole stabilire se il flusso di automobili che transitano attraverso un casello autostradale in un determinato intervallo di tempo può essere bene approssimato mediante una legge di Poisson. A tale fine si effettuano 100 osservazioni diverse supposte indipendenti in cui si conta il numero di vetture che transitano dal casello in un minuto, e si ottiene la seguente tabella:**

Numero di automobili	0	1	2	3	4	5	≥ 6
frequenza assoluta	9	28	31	14	11	7	0

Stimare il parametro della Poissoniana dai dati della tabella, e discutere la validità dell'adattamento dei dati alla legge di Poisson con un livello di significatività del 5%.

Soluzione.
Per verificare l'adattamento dei dati ad una distribuzione con legge poissoniana usiamo il test chi-quadrato di buon adattamento. La regione di rifiuto è data da $Q > \chi^2_{1-\alpha}(N_c - 1 - r)$ dove r è il numero di parametri eventualmente da stimare, N_c è il numero di classi rimanenti dopo l'eventuale raggruppamento per soddisfare la condizione di capienza e $\alpha = 0.05$ è il livello di significatività. La statistica test è $Q = \sum_{k=1}^{N_c} \frac{(np_k - N_k)^2}{np_k}$ dove $n = 100$ è l'ampiezza del campione a nostra disposizione mentre np_k e N_k sono, rispettivamente, le frequenze assolute attese e osservate della k-esima classe. Procediamo quindi al calcolo delle frequenze attese usando la consueta formula $np_k = ne^{-\lambda} \frac{\lambda^k}{k!}$ (tranne per l'ultima classe la cui frequenza si ottiene come differenza tra 1 e la somma delle frequenze precedenti). Poiché il parametro λ della distribuzione poissoniana non è dato dobbiamo prima effettuarne la stima usando come stimatore la media campionaria:

$$\lambda = \overline{x}_n = \frac{1}{100}(0 \cdot 9 + 1 \cdot 28 + 2 \cdot 31 + 3 \cdot 14 + 4 \cdot 11 + 5 \cdot 7 + 6 \cdot 0) = 2.11$$

Ricordiamoci, inoltre, di porre $r = 1$.

k	0	1	2	3	4	5	≥ 6
np_k	12.1238	25.5812	26.9882	18.9817	10.0128	4.2254	2.0869

Le ultime due classi devono essere accorpate in quanto non soddisfano la condizione di capienza $np_k \geq 5$:

k	0	1	2	3	4	≥ 5
np_k	12.1238	25.5812	26.9882	18.9817	10.0128	6.3123

Ora possiamo fissare il numero di classi $N_c = 6$ ed effettuare il test: sulle tavole troviamo $\chi^2_{1-\alpha}(N_c - 1 - r) = \chi^2_{0.95}(4) \simeq 9.4877$ e, con un po' di conti, ricaviamo

$$Q = \sum_{k=1}^{N_c} \frac{(np_k - N_k)^2}{np_k} = 3.1096$$

Quindi l'ipotesi di buon adattamento non può essere rifiutata.