Statistica Matematica A

Docente: Dott. F. Zucca

I recupero

Lecco 14/2/2004

Candidato

Cognome: .... **Nome:** ....

Corso di Laurea: ....

Matricola: .... **Anno di corso:** ....

Scegliere l'opzione che vi interessa

☐ **Recupero I parte** $\Longrightarrow$ Esercizi **1, 2, 3**; tempo a disposizione **2h**.

☐ **Recupero II parte** $\Longrightarrow$ Esercizi **4, 5, 6**; tempo a disposizione **2h**.

☐ **Appello** $\Longrightarrow$ Svolgere gli esercizi **1, 2, 4, 5**; tempo a disposizione **2h40m**.

1 Esercizi

Esercizio 1 Il tempo che intercorre tra il momento in cui un principiante della boxe riceve un pugno ed il pugno successivo è governato da una legge esponenziale.

1. Sapendo che la probabilità di ricevere un pugno prima che siano trascorsi dieci secondi da quello precedente è $1/2$, calcolare la probabilità che debba attendere almeno 2 secondi dal primo pugno per ricevere il secondo.

2. Supponiamo che anche il tempo che intercorre tra l'inizio del match e l'arrivo del primo pugno abbia legge esponenziale con lo stesso parametro calcolato al punto precedente. Se non ha ancora incassato alcun pugno dopo 10 secondi dall'inizio, qual è la probabilità che non ne incassi nessuno nei primi 20 secondi del match?

3. Supponendo che i tempi che intercorrono tra i vari pugni siano indipendenti ed identicamente distribuiti e supponendo che dopo 10 pugni il nostro malcapitato vada KO, calcolare (utilizzando un'opportuna approssimazione) che vada KO nei primi 3 minuti. Qual è il minimo numero di pugni che deve saper incassare per non finire KO nei primi 3 minuti con probabilità pari almeno al 90%?

Soluzione. Sia X la **v.a.** che misura in secondi il tempo di attesa tra due pugni. Il testo ci dice che $X \sim Esp(\lambda)$. Innanzi tutto ricaviamo il valore del parametro. Sappiamo che $\frac{1}{2} = 1 - \frac{1}{2} = 1 - \mathbb{P}(X < 10) = \mathbb{P}(X > 10) = e^{-10\lambda}$ da cui ricaviamo: $\lambda = \frac{\ln 2}{10}$

1. $\mathbb{P}(X > 2) = e^{-2\lambda} = e^{-\frac{\ln 2}{5}} = 2^{-1/5}$
2. Per l'assenza di memoria: $\mathbb{P}(X > 20 | X > 10) = \mathbb{P}(X > 10) = \frac{1}{2}$
3. Sia $S_n = \sum_{i=1}^{n} X_i$ dove $X_i \sim Esp(\lambda)$ rappresenta il tempo di attesa per l'i-esimo pugno. (Osserviamo che $S_n \sim \Gamma(n, \lambda)$ e dunque la condizione di applicabilità del TLC può ritenersi "abbastanza" soddisfatta anche per $n = 10$). In ogni caso applicando il Teorema Centrale del Limite si ha $S_n \simeq N\left(\frac{n}{\lambda}, \frac{n}{\lambda^2}\right)$ e $\mathbb{P}(S_n \leq 180) \simeq \Phi\left(\frac{180 - \frac{n}{\lambda}}{\frac{\sqrt{n}}{\lambda}}\right) \simeq \Phi\left(\frac{18 \ln 2 - 10}{\sqrt{10}}\right) \simeq 0.7832$.
   Ora troviamo $n : 1 - \Phi\left(\frac{180 - \frac{n}{\lambda}}{\frac{\sqrt{n}}{\lambda}}\right) \geq 0.9$. Siamo condotti alla risoluzione in $\mathbb{N}$ dell'equazione: $n - z_{0.9}\sqrt{n} - 18 \ln 2 \geq 0$ dalla quale ricaviamo $\sqrt{n} \geq 4.231$ e quindi $n \geq \lceil 17.9 \rceil = 18$.

Esercizio 2 Si consideri la seguente funzione reale

$$f_c(x) := \begin{cases} \frac{1}{2}\sin(x) & \text{se } x \in [0, c] \\ 0 & \text{altrimenti} \end{cases}$$

dipendente dal parametro reale $c \geq 0$.

1. Calcolare per quali valori di $c \geq 0$ la funzione f_c è una densità di probabilità.
2. In corrispondenza al valore (o ai valori) trovati al punto precedente, detta X_c una variabile aleatoria avente densità f_c, si calcolino media e varianza di X_c.
3. Quanto vale $\mathbb{P}(X_c \geq \pi/2)$?

Soluzione.

1. Ovviamente $f_c \geq 0$ se e solo se $c \leq \pi$. D'altro canto
$$1 = \int_{\mathbb{R}} f_c(x)\mathrm{d}x = \frac{1}{2}\int_0^c \sin(x)\mathrm{d}x$$
$$\frac{1}{2}\cos(x)|_0^c = \frac{1}{2}(1 - \cos(c))$$
se e solo se $c = \pi(2k+1)$ con $k \in \mathbb{N}$. Pertanto l'unica soluzione è $c = \pi$.

2. Per la simmetria di f_π rispetto a $\pi/2$ si ha che
$$\mathbb{E}(X_\pi) = \int_0^\pi x f_\pi(x)\mathrm{d}x = \frac{\pi}{2};$$
alternativamente basta calcolare per parti la primitiva di f_c che risulta $(\sin(x) - x\cos(x))/2$.

   Similmente
$$\begin{aligned}\mathrm{var}(X_\pi) &= \int_0^\pi x^2 f_\pi(x)\mathrm{d}x - \frac{\pi^2}{4} \\ &= \frac{1}{2}(2x\sin(x) - x^2\cos(x) + 2\cos(x))|_0^\pi - \frac{\pi^2}{4} \\ &= \frac{\pi^2}{4} - 2.\end{aligned}$$

3. Ancora per simmetria, o per calcolo diretto dell'integrale della densità, si ha $\mathbb{P}(X_\pi \geq \pi/2) = 1/2$.

Esercizio 3 Un'azienda produttrice di pompe idrauliche rifornisce un importante cantiere navale che utilizza tali pompe per lo spurgo dei servizi igienici di bordo. Poiché invertendo la polarità del motore (che funziona a corrente continua) la pompa espelle anziché aspirare, è assai importante che sulla presa di alimentazione sia stampigliato un simbolo che indichi correttamente il polo positivo. La macchina stampigliatrice sbaglia a indicare la polarità l'1% delle volte e, conseguentemente, l'1% della produzione risulta difettosa. La produzione è a linea e i pezzi prodotti vengono immagazzinati e poi venduti in lotti di 10 pezzi ognuno. Sia W la v.a. geometrica che conta le pompe non difettose prodotte prima di averne una difettosa. Sia X_i la v.a. bernoulliana che indica lo stato dell'i-esima pompa di un lotto (assumendo valore 1 se è difettosa e 0 se non lo è). Sia $Y = \sum_{i=1}^{10} X_i$ la v.a. che conta il numero di pompe difettose presenti in un lotto. Siamo all'inizio di un giorno di lavoro qualsiasi. La linea viene avviata e inizia a produrre.

1. Calcolare la probabilità che il primo pezzo prodotto dalla linea sia difettoso.

2. Calcolare la probabilità che il centesimo pezzo prodotto dalla linea sia difettoso.

3. Calcolare la probabilità che il centounesimo pezzo prodotto dalla linea sia sano sapendo che i primi cento lo erano.

4. Osservando il legame tra Y e le X_i desumere la legge di Y e i suoi parametri.

5. Qual è la probabilità che un lotto contenga almeno una pompa difettosa?

Soluzione. Poiché $X_i \sim B(p)$ avremo che $W \sim G(p)$. Dunque:

1. $\mathbb{P}(W = 0) = p = 0.01$

2. Ci sono stati problemi di interpretazione: se avete inteso la domanda come la probabilità, che il 100-esimo pezzo prodotto sia difettoso indipendentemente dagli altri allora la risposta è $1 - p$; se invece avete inteso la probabilità che il 100-esimo pezzo sia il primo della catena ad essere difettoso allora la risposta è $\mathbb{P}(W = 99) = q^{99}p \simeq 0.0037$

3. Osserviamo che il pezzo n-esimo è sano se sono sani almeno (e non esattamente) i primi n pezzi dunque, per l'assenza di memoria della geometrica: $\mathbb{P}(W \geq 101|W \geq 100) = \mathbb{P}(W \geq 1) = 1 - \mathbb{P}(W = 0) = 1 - p = 0.99$

4. Sotto l'ipotesi di indipendenza delle X_i abbiamo che $Y \sim B(10, 0.01)$

5. $\mathbb{P}(Y \geq 1) = 1 - \mathbb{P}(Y = 0) = 1 - (1 - p)^{10} \simeq 0.0956$

Esercizio 4 Il famoso produttore di cioccolato Von Dent afferma che le sue tavolette contengono almeno il 90% di cacao purissimo. Da un'indagine a campione basata su 15 confezioni è risultata una concentrazione media pari a 0.89, con una deviazione standard di 0.04. Sulla base dell'indagine, supponendo che la concentrazione segua legge Normale, Von Dent può considerare veritiera la sua affermazione ad un livello di significatività $\alpha = 0.05$?

Soluzione. Siamo nel caso di test di ipotesi asimmetrico per la media di una popolazione normale con varianza incognita.

Formuliamo le ipotesi:

$$H_0 : \mu \geq \mu_0$$

$$H_1 : \mu < \mu_0$$

La regione di rifiuto corrisponde alla coda di sinistra quindi il test deve essere rifiutato se:

$$\frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}} < t_\alpha(n-1)$$

Poiché $\bar{x}_n = 0.89$, $s_n = 0.04$, $n = 15$ e $\mu_0 = 0.9$ ricaviamo:

$$\frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}} = \frac{0.89 - 0.9}{0.04/\sqrt{15}} \simeq -0.96825$$

mentre dalle tavole:

$$t_{0.05}(14) = -t_{0.95}(14) \simeq -1.761310$$

Il test non può essere rifiutato e pertanto le affermazioni del produttore non possono essere smentite.

In maniera alternativa, mediante l'uso delle tavole si può valutare il p-value

$$\bar{\alpha} = F_{t(n-1)}\left(\frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}}\right) = F_{t(14)}\left(-0.96825\right) = 1 - F_{t(14)}\left(0.96825\right) \in (0.15, 0.2);$$

pertanto non possiamo rifiutare l'ipotesi nulla al 5% essendo $\bar{\alpha} > 0.05$.

Esercizio 5 Il test d'ingresso alla facoltà di ingegneria è preceduto da una prova che può essere valutata A, B, C o D (A=molto buono, D=insufficiente). Su un numero molto elevato di anni si sono presi in esame 50000 studenti ottenendo la distribuzione di voti riassunta dalla seguente tabella:

VOTO	A	B	C	D
N.Studenti	9000	21000	12000	8000

che tradizionalmente viene considerata la distribuzione teorica di riferimento.

Il celebre istituto di "Monkey Island" famoso in tutto il mondo per le sue scimmie sapienti, sceglie 100 tra i migliori esemplari e sottopone loro il test di ammissione ottenendo la seguente tabella

VOTO	A	B	C	D
N.Scimmie	26	30	16	28

1. C'è accordo tra la distribuzione dei risultati ottenuti dalle scimmie e quella di riferimento ad un livello del 10%?
2. Stimare il P-value.
3. Cosa cambierebbe nei due punti precedenti se dimezzassimo il numero di scimmie mantenendo le stesse proporzioni?

Soluzione.

1. Facilmente calcoliamo la quantità

$$Q = \sum_{i=1}^{4} 100\left(\frac{f_i^2}{p_i} - 1\right)$$

dove

VOTO	A	B	C	D
f_i	0.26	0.3	0.16	0.28
N.Studenti	0.18	0.42	0.24	0.16

ottenendo $Q = 18.65079$ che cade nella regione di rifiuto $Q > \chi^2_{0.9}(3) = 6.25$ pertanto rifiutiamo l'ipotesi nulla.

2. La stima del P-value è $\overline{\alpha} < 0.001$ (il computer restituisce il valore $\overline{\alpha} \simeq 0.000323$).

3. Dimezzando il numero di scimmie senza cambiare le proporzioni, si dimezza anche Q pertanto nel secondo caso $Q = 9.325395$ che ancora non permette di accettare l'ipotesi nulla al 10%, ed $\overline{\alpha} \simeq 0.025$(il computer restituisce il valore $\overline{\alpha} \simeq 0.02526329$).

Esercizio 6 Nel laghetto vicino alla centrale nucleare di Springfield nel 2004 c'è stata un'impennata di nascite di "pesci con tre occhi": su un campione di 30000 pesci ben 1200 possiedono questa buffa mutazione contro i 1100 del 2003, i 1160 del 2002, i 1240 del 2001 ed i 1000 del 2000 (con egual campione).

Il signor Smithers, noto segretario personale del padrone della centrale, tranquillizza la popolazione dicendo che non c'è aumento dell'inquinamento da radiazioni in quanto non si può affermare che il numero di mutazioni dell'ultimo anno sia superiore alla media degli ultimi 4 anni.

1. Si scelga l'opportuna ipotesi nulla (ricordando che il signor Smithers ha tutto l'interesse a non far chiudere la centrale e quindi a mostrare che la media non sia superiore) e si discuta se al livello di significatività del 5% il signor Smithers può fare una simile affermazione.
   (Sugg: si prenda la media degli ultimi anni come valore vero di riferimento.)

2. (**Domanda facoltativa priva di punteggio**) Come si chiama il padrone della centrale?

Soluzione.

1. Si consideri un test di ipotesi sulla percentuale della popolazione p con ipotesi nulla

$$H_0 : p \leq (1100 + 1160 + 1240 + 1000)/(4 \cdot 30000)$$

Il valore stimato è $\overline{p} = 0.04$ mentre il valore di confronto $p_0 = 1125/30000 \approx 0.0375$. La regione di rifiuto ed il p-value per questo test sono

$$\overline{p} > p_0 + \sqrt{p_0(1-p_0)}\,\frac{q_{1-\alpha}}{\sqrt{n}}$$

$$\bar{\alpha} = 1 - \phi\left(\frac{\overline{p} - p_0}{\sqrt{p_0(1-p_0)/n}}\right)$$

Dalle tabelle $q_{0.95} = 1.6449$ pertanto la regione di rifiuto ed il p-value risultano

$$\overline{p} > 0.03930 \qquad \bar{\alpha} = 1 - \phi(2.2792) = 1 - 0.9886 = 0.0114,$$

pertanto l'ipotesi nulla va rifiutata e quindi il signor Smithers non poteva affermare ciò che ha detto.

2. Mr. Montgomery Burns (ref. The Simpsons).