Statistica I. Ingegneria Gestionale. Scritto del 14/02/2012

Cerchiare, su questo foglio, le risposte corrette e risolvere per esteso gli esercizi sui fogli assegnati.

Esercizio 1. Si analizzano statisticamente i guasti settimanali che avvengono in un impianto. Si considerano tutti i guasti, senza fare distinizioni di tipologia. Osservando i guasti per dieci settimane si sono osservate le seguenti numerosità:

$$4, \quad 4, \quad 8, \quad 5, \quad 7, \quad 4, \quad 8, \quad 3, \quad 5, \quad 3.$$

Indichiamo con N la v.a. numero di guasti alla settimana.

1. (3 pt) Supponiamo N gaussiana. Se escludiamo il 5% delle settimane, che numero massimo di guasti ci si deve aspettare settimanalmente?

..., ..., ..., ..., 8.245.

2. (3 pt) Per quale percentuale di settimane i guasti sono ≤ 2 (spiegare il legame rigoroso tra la risposta data ed il concetto di percentuale di settimane)?

..., ..., ..., ..., 0.052.

3. (4 pt) Coi dati a disposizione, si ha una certa precisione nella stima del numero medio di guasti alla settimana. Se volessimo raddoppiare la precisione (dimezzare l'errore), quante ulteriori settimane di osservazione servirebbero?

..., ..., ..., ..., 30.

4. (4 pt) Tenendo conto dell'imprecisione sulla media (relativamente ai 10 dati noti) relativa ad un livello di confidenza del 90%, che valore cautelativo si potrebbe dare per il numero massimo di riparazioni ad eccezioni del 5% delle settimane?

..., ..., ..., ..., 9.24.

5. (4 pt) Ogni tanto viene controllato se il numero di guasti settimanali è statisticamente simile a quello descritto sopra oppure no, senza pregiudizi circa il fatto che aumenti o diminuisca. Il controllo viene eseguito registrando i dati relativi ad 8 settimane consecutive ed eseguendo un test al 90%. Se in un certo periodo storico il vero numero medio di guasti fosse diventato pari a 6, che probabilità avremmo di accorgerci che la media è cambiata? Scrivere la probabilità richiesta motivando la scrittura e da lì calcolare la probabilità richiesta.

..., ..., ..., ..., 0.3785.

6. (4 pt) Supponiamo (nelle condizioni dell'inizio dell'esercizio) che l'impianto sia composto da 20 sottosistemi identici che si possono guastare in modo indipendente (una volta sola a settimana, per semplicità; e supponiamo che non possano esserci altri tipi di guasto, per semplicità). Che variabile aleatoria si dovrebbe

usare ora per N? Con essa, calcolare la probabilità di avere al massimo un guasto.

..., ..., ..., ..., 0.022.

Esercizio 2. i) (3 pt) Siano X, Y due v.a. di Poisson indipendenti di parametri 2 e 3 rispettivamente.

i) (3 pt) Calcolare $E\left[X\left(1-X\right)Y\right]$.

..., ..., ..., ..., -12.

ii) (3 pt) Calcolare $P\left(X+Y<2\right)$.

..., ..., ..., ..., $=0.04$.

iii) (2 pt) Se $X_1,...,X_{40}$ sono distribuite come X e sono indipendenti, calcolare $P\left(X_1+...+X_{40}>80\right)$.

..., ..., ..., ..., $=0.5$.

Quantili utili per questo compito:

$x=$	-2.976	-1.621	0.314	1.645	2	2.8996	3.297	3.894
$\Phi\left(x\right)=$	0.0015	0.052	0.623	0.95	0.977	0.998	0.9995	0.99995

Formule e teoremi potenzialmente utili. 1) Se X è $B\left(n,p\right)$, allora $P\left(N=k\right)=\binom{n}{k}p^k\left(1-p\right)^{n-k}$, $E\left[X\right]=np$, $Var\left[X\right]=np\left(1-p\right)$; se X è $\mathcal{P}\left(\lambda\right)$, allora $P\left(X=k\right)=e^{-\lambda}\frac{\lambda^k}{k!}$, $E\left[X\right]=Var\left[X\right]=\lambda$; se X è esponenziale di parametro λ, $F\left(x\right)=1-e^{-\lambda x}$, $E\left[X\right]=\sigma\left[X\right]=\frac{1}{\lambda}$. 2) Se $X_1,...,X_n$ sono $N\left(\mu,\sigma^2\right)$, allora $P\left(\left|\overline{X}-\mu\right|\leq\frac{\sigma q_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right)=1-\alpha$. 3) Se $X_1,...,X_n$ sono $N\left(\mu,\sigma^2\right)$, allora $Z=\frac{\overline{X}-\mu_0}{\sigma}\sqrt{n}$ è una $N\left(d,1\right)$ con $d=\frac{\mu-\mu_0}{\sigma}\sqrt{n}$.

1 Soluzioni

Esercizio 1. i) $\overline{x} = 5.1$, $s = 1.912$. Come illustrato anche con un disegno (qui non riportato) si cerca quel valore $\lambda_{\max}$ tale che $P(N \geq \lambda_{\max}) = 0.05$, valore che quindi sta a destra della media. Esso vale

$$\lambda_{\max} = 5.1 + 1.912 \cdot q_{0.95} = 5.1 + 1.912 \cdot 1.645 = 8.2452.$$

ii) Dobbiamo calcolare

$$P(N \leq 2) = \Phi\left(\frac{2 - 5.1}{1.912}\right) = \Phi(-1.6213) = 0.052.$$

Il legame rigoroso è dato dalla legge dei grandi numeri, come descritto in altri compiti.

iii) La precisione (o errore), relativamente ad un livello di confidenza $1 - \alpha$, è data da

$$\delta = \frac{1.912 \cdot q_{-\frac{\alpha}{2}}}{\sqrt{10}}.$$

Se vogliamo che diventi pari a $\frac{\delta}{2}$, dobbiamo avere un campione di numerosità 40: infatti esso produrrebbe un errore pari a

$$\frac{1.912 \cdot q_{-\frac{\alpha}{2}}}{\sqrt{40}} = \frac{1}{2}\frac{1.912 \cdot q_{-\frac{\alpha}{2}}}{\sqrt{10}}.$$

Servono quindi altre 30 settimane.

iv) La stima della media al 90% è

$$\begin{aligned}\mu &= 5.1 \pm \frac{1.912 \cdot q_{0.95}}{\sqrt{10}} = 5.1 \pm \frac{1.912 \cdot 1.645}{\sqrt{10}} \\ &= 5.1 \pm 0.995.\end{aligned}$$

Quindi la situazione peggiore nella stima del numero massimo di guasti è che sia

$$\mu = 5.1 + 0.995 = 6.095.$$

In questo caso, con le notazioni usate sopra,

$$\lambda_{\max} = 6.095 + 1.912 \cdot q_{0.95} = 6.095 + 1.912 \cdot 1.645 = 9.24.$$

v) Eseguiamo un test bilaterale al 90%, con ipotesi nulla $\mathcal{H}_0 =$ la media è 5.1, confrontando $z = \frac{\overline{x} - 5.1}{1.912}\sqrt{8}$ con $q_{1-\frac{\alpha}{2}} = q_{0.95} = 1.645$: se $|z| > 1.645$ si rifiuta l'ipotesi. Supponiamo che la media vera sia diventata pari a 6. Ci accorgiamo che la media è cambiata se $|z| > 1.645$, quindi l'esercizio chiede di calcolare

$$P_{\mu=6}(|Z| > 1.645)$$

dove $Z = \frac{\overline{X}-5.1}{1.912}\sqrt{8}$ ed il campione da cui è tratto $\overline{X}$ ha media $\mu = 6$. La v.a. Z è una $N(d,1)$ con $d = \frac{6-5.1}{1.912}\sqrt{8} = 1.331$. Quindi

$$\begin{aligned}
P_{\mu=6}(|Z| > 1.645) &= 1 - P_{\mu=6}(|Z| \leq 1.645) \\
&= 1 - P_{\mu=6}(Z \leq 1.645) + P_{\mu=6}(Z \leq -1.645) \\
&= 1 - \Phi(1.645 - 1.331) + \Phi(-1.645 - 1.331) \\
&= 1 - \Phi(0.314) + \Phi(-2.976) \\
&= 1 - 0.623 + 0.0015 = 0.3785.
\end{aligned}$$

vi) Indichiamo con X_1 la v.a. che vale 1 se il primo sottosistema si rompe, 0 altrimenti, e poniamo $p = P(X_1 = 1)$; indichiamo poi con $X_2, ..., X_{20}$ le analoghe variabili, indipendenti, per gli altri sottosistemi. Vale $N = X_1 + ... + X_{20}$ quindi N è una $B(20, p)$, per un noto teorema. Serve stimare p per il seguito. Essendo $E[N] = 20 \cdot p$ ed avendo la stima 5.1 per la media, stimiamo

$$p = \frac{5.1}{20} = 0.255.$$

La probabilità richiesta è $P(N \leq 1)$ quindi

$$\begin{aligned}
P(N \leq 1) &= P(N = 0) + P(N = 1) \\
&= \binom{20}{0} 0.255^0 (1 - 0.255)^{20-0} + \binom{20}{1} 0.255^1 (1 - 0.255)^{20-1} \\
&= 0.022.
\end{aligned}$$

Esercizio 2. i)

$$\begin{aligned}
E[X(1-X)Y] &= E\left[X - X^2\right] E[Y] = \left(E[X] - E\left[X^2\right]\right) E[Y] \\
&= \left(E[X] - Var[X] - E[X]^2\right) E[Y] \\
&= \left(2 - 2 - 2^2\right) 3 = -12.
\end{aligned}$$

ii) Si può fare in vari modi. Il più veloce utilizza il fatto che $Z = X + Y$ è Poisson di parametro 2+3=5, per cui

$$\begin{aligned}
P(X + Y < 2) &= P(Z < 2) = P(Z = 0) + P(Z = 1) \\
&= e^{-5}\frac{5^0}{0!} + e^{-5}\frac{5^1}{1!} = 0.04.
\end{aligned}$$

iii)

$$P(X_1 + ... + X_{40} > 80) = P\left(\frac{X_1 + ... + X_{40} - 40 \cdot 2}{\sqrt{40}\sqrt{2}} > \frac{80 - 40 \cdot 2}{\sqrt{40}\sqrt{2}}\right)$$

circa uguale a $P(Z > 0)$ con $Z \sim N(0,1)$, quindi circa uguale a $1/2$.