**Statistica I. Ingegneria Gestionale. Scritto del 29/06/2010**

**Esercizio 1.** i) In un sistema di lettura automatica dei testi scritti a mano, le lettere $n$ ed $u$ sono facili da confondere. In una certa lingua, la $n$ compare con frequenza $1/15$, mentre la $u$ con frequenza $1/25$. Se la lettera è davvero $n$, il sistema legge $n$ il 90% delle volte. Se la lettera è $u$, il sistema la legge $n$ il 20% delle volte. Se la lettera è diversa sia da $n$ sia da $u$, il sistema la legge $n$ il 2% delle volte. Se il sistema legge $n$, con che probabilità ha sbagliato?

0.225, 0.302, 0.421

ii) Se il sistema, nel leggere un testo, legge 12 volte la lettera $n$, con che probabilità ne sbaglia almeno una?

0.746, 0.812, 0.987

iii) Date $X$, $Y$ indipendenti, entrambe Poisson di parametro 2, calcolare $P(X+Y \leq 2)$.

0.127, 0.238, 0.318

iv) Calcolare inoltre $E\left[(X-2)^2 e^Y\right]$.

62.16, 153.2, 15.21

v) Se $f$ è la funzione che vale $f(x) = x$ per $x \leq 2$ e $f(x) = 2$ per $x > 2$, detta $Z$ la v.a. $Z = f(X)$ (sempre con $X \sim \mathcal{P}(2)$) calcolare $E[Z]$.

1.459, 2.942, 3.12

**Esercizio 2.** Si vuole monitorare il livello $L$ di rumore ($L$ va considerato come una variabile aleatoria) di un sistema meccanico, dando l'allarme quando $L$, mediato su un certo numero di osservazioni, supera una certa soglia; si ritiene infatti che il livello di rumore rifletta eventuali mal funzionamenti. Non interessa quindi quando $L$ è troppo basso, ma solo quando è troppo alto. Si misura $L$ in $n$ istanti di tempo ravvicinati, ottenendo i numeri $l_1, ..., l_n$ e si calcola la media $\bar{l}_n$.

i) Innanzi tutto bisogna conoscere le proprietà statistiche di $L$. Si misura $L$ per 100 volte in condizioni note di buon funzionamento. Si ottengono i valori $l_1 + ... + l_{100} = 315.726$, $l_1^2 + ... + l_{100}^2 = 1124.92$. Stimare $E[L]$ al 95%.

$3.157 \pm 0.123$, $3.157 \pm 0.323$, $3.157 \pm 0.223$

ii) Prendiamo per buoni i valori di media e deviazione ottenuti con l'esperimento precedente. Misuriamo un singolo valore di $L$ in condizioni di buon funzionamento. Tale valore, di quale numero $l_{\max}$ sarà minore, con probabilità 0.95?

$l_{\max} = 5.023$, 4.023, 6.023

iii) Ora bisogna eseguire i controlli descritti all'inizio. Scegliamo $n = 10$. Ogni tanto si calcola $\bar{l}_{10}$ e si vede se è ammissibile o no. Fissando una significatività pari a 95%, con che soglia va confrontato $\bar{l}_{10}$?

$$3.747, \qquad 2.747, \qquad 4.747$$

iv) I falsi allarmi (superamenti della soglia quando la media non è cambiata) non ci preoccupano eccessivamente, perche' appena si verifica un allarme, come prima cosa si ripete il controllo, confidando che così facendo, se quello era un falso allarme, il secondo valore di $\bar{l}_{10}$ sia rientrato nei limiti. Il vero problema invece è che il sistema non si accorga di un cambiamento sistematico considerevole. I tecnici ci dicono che un aumento della media del rumore di un'unità corrisponde ad un pericolo. Riesce il nostro sistema a scoprire un tale aumento? Se accorgersi di un tale aumento viene ritenuto vitale, per cui si può accettare di sbagliare al massimo una volta su 1000, cosa si deve fare? Discutere il problema.

# 1 Soluzioni

**Esercizio 1.** i) Usiamo la scrittura $L = n$ per dire che il sistema ha letto $n$, $L \neq n$ per dire il contrario, ed usiamo semplicemente $n$, $u$, *altro* per dire che la lettera da decifrare era una $n$, una $u$, o altro. Allora

$$P(n) = \frac{1}{15}, \qquad P(u) = \frac{1}{25}, \qquad P(altro) = 1 - \frac{1}{15} - \frac{1}{25}$$
$$P(L = n|n) = 0.9, \qquad P(L = n|u) = 0.2, \qquad P(L = n|altro) = 0.02.$$

Quindi

$$P(L = n) = 0.9 \cdot \frac{1}{15} + 0.2 \cdot \frac{1}{25} + 0.02 \cdot \left(1 - \frac{1}{15} - \frac{1}{25}\right) = 0.086$$

$$P(n|L = n) = \frac{P(L = n|n) \cdot P(n)}{P(L = n)} = \frac{0.9 \cdot \frac{1}{15}}{0.086} = 0.698$$

quindi la probabilità richiesta è $1 - 0.698 = 0.302$.

ii) Il numero $N$ di errori su 12 letture della lettera $n$ è una $B(12, p)$ con $p$ = probabilità di errore in una lettura, quindi $p = 0.302$. Dobbiamo calcolare

$$P(N \geq 1) = 1 - P(N < 1) = 1 - P(N = 0) = 1 - (1 - p)^n = 1 - (1 - 0.302)^{12} = 0.987.$$

iii)

$$\begin{aligned}
P(X + Y \leq 2) &= \sum_{k=0}^{2} P(X + Y \leq 2|Y = k) P(Y = k) = \sum_{k=0}^{2} P(X \leq 2 - k) P(Y = k) \\
&= P(X \leq 2) P(Y = 0) + P(X \leq 1) P(Y = 1) + P(X \leq 0) P(Y = 2) \\
&= e^{-2}\left(1 + 2 + \frac{4}{2}\right) e^{-2} + e^{-2}(1 + 2) e^{-2} 2 + e^{-2} e^{-2} \frac{4}{2} = 0.238.
\end{aligned}$$

iv)

$$E\left[(X - 2)^2 e^Y\right] = E\left[(X - 2)^2\right] E\left[e^Y\right] = Var[X] \varphi_Y(1) = 2 \cdot e^{2(e-1)} = 62.16.$$

v) La v.a. $Z$ assume il valore 0 quando $X$ vale 0, quindi con probabilità $e^{-2}$; assume il valore 1 quando $X$ vale 1, quindi con probabilità $e^{-2} \cdot 2$; assume il valore 2 quando $X$ vale un qualsiasi valore $\geq 2$, quindi con probabilità $1 - e^{-2}(1 + 2)$; pertanto

$$E[Z] = 1 \cdot e^{-2} \cdot 2 + 2 \cdot \left(1 - e^{-2}(1 + 2)\right) = 1.459.$$

**Esercizio 2.** i)

$$\bar{l}_{100} = \frac{l_1 + ... + l_{100}}{100} = \frac{315.726}{100} = 3.157$$

$$S_L^2 = \frac{1}{99} \sum_{k=1}^{100} \left(l_k - \bar{l}_{100}\right)^2 = \frac{1}{99} \sum_{k=1}^{100} \left(l_k^2 - \bar{l}_{100}^2\right) = \frac{l_1^2 + ... + l_{100}^2}{99} - \frac{100 \cdot \bar{l}_{100}^2}{99} = 1.295$$

$$S_L = 1.138$$

$$\mu = 3.157 \pm \frac{1.138 \cdot q_{0.975}}{\sqrt{100}} = 3.157 \pm \frac{1.138 \cdot 1.96}{10} = 3.157 \pm 0.223.$$

ii) Dobbiamo trovare il numero $l_{\max}$ che soddisfa

$$P(L < l_{\max}) = 0.95.$$

Con un disegno (e confrontando con la gaussiana canonica) si vede che

$$l_{\max} = \mu + \sigma \cdot q_{0.95} = 3.157 + 1.138 \cdot 1.64 = 5.023.$$

iii) Il problema è si affronta con la teoria degli intervalli si confidenza, o dei test, oppure osservando che è lo stesso problema dell'esecizio precedente. La v.a. $\overline{L}_{10}$ è una gaussiana di media $\mu_0 = 3.157$ e varianza $\frac{\sigma^2}{10}$ e la soglia $\lambda$ con cui va confrontato $\overline{l}_{10}$ con significatività 0.95 è definita dall'equazione

$$P\left(\overline{L}_{10} < \lambda\right) = 0.95$$

da cui si trova

$$\lambda = \mu_0 + \frac{\sigma}{\sqrt{10}} \cdot q_{0.95} = 3.157 + \frac{1.138}{\sqrt{10}} \cdot 1.64 = 3.747.$$

iv) Un valore $\mu = \mu_0 + 1 = 4.157$ viene considerato pericoloso; il sistema deve accorgersi di esso. Vediamo con che probabilità se ne accorge (potenza):

$$\begin{aligned} P_\mu\left(\overline{L}_{10} > \lambda\right) &= 1 - P_\mu\left(\overline{L}_{10} < \lambda\right) = 1 - \Phi\left(\frac{\lambda - \mu}{\sigma}\sqrt{10}\right) \\ &= 1 - \Phi\left(\frac{3.747 - 4.157}{1.138}\sqrt{10}\right) = 1 \cdot -\Phi(-1.139) = \Phi(1.139) = 0.8729 \end{aligned}$$

dove abbiamo usato il fatto che, quando la media vera è $\mu$, $\overline{L}_{10}$ è una gaussiana di media $\mu = 4.157$ e varianza $\frac{\sigma^2}{10}$. E' una buona potenza.

Se però si vuole sbagliare al massimo una volta su mille, la potenza deve essere $\geq 0.999$. L'unico modo di ottenerla è aumentare $n$. Dev'essere $P_\mu\left(\overline{L}_{10} > \lambda\right) \geq 0.999$ (seguono i calcoli per trovare $n$).