Corso di Laurea in Ingegneria Informatica
Anno Accademico 2014/2015
Calcolo delle Probabilità e Statistica Matematica

Nome ....

N. Matricola .... Ancona, 14 luglio 2015

1. Tre macchine producono gli stessi pezzi di ricambio. La macchina A produce 600 pezzi al giorno, la macchina B ne produce 400 e la macchina C ne produce 200. Si sa che i pezzi difettosi prodotti dalla macchina A sono il 4% del totale, quelli difettosi della macchina B sono il 3% e quelli della C l'1 %. Infine, i pezzi prodotti vengono raccolti in un unico contenitore. Se, estraendo un pezzo a caso dal contenitore, questo risulta difettoso, qual è la probabilità che sia stato prodotto dalla macchina B?

   Soluzione. Siano A, B e C gli eventi "il pezzo preso a caso è prodotto dalla macchina A, B o C". Sia inoltre D l'evento "il pezzo è difettoso". Allora

   $$P(A) = \frac{600}{1200} = \frac{1}{2}, \quad P(B) = \frac{1}{3}, \quad \frac{1}{6}$$
   $$P(D|A) = 0.04, \quad P(D|B) = 0.03, \quad P(D|C) = 0.01$$

   In base al teorema delle probabilità totali

   $$P(D) = P(D|A)\,P(A) + P(D|B)\,P(B) + P(D|C)\,P(C) = 0.03167$$

   La domanda cui si chiede di rispondere è $P(B|D)$. Con Bayes abbiamo

   $$P(B|D) = \frac{P(D|B)\,P(B)}{P(D)} = 0.3158$$

2. Sia X una variabile casuale distribuita uniformemente nell'intervallo $[0, 20]$. Siano $Y = [X]$ (con $[x]$ la parte intera di x) e $Z = [X + 0.5]$. Calcolare

   - $E[Y]$ ed $E[Z]$;
   - $E[|X - Y|]$ ed $E[|X - Z|]$.

   Soluzione. X è una variabile continua di densità

   $$f(x) = \begin{cases} 1/20 & \text{per } 0 \leq x \leq 20 \\ 0 & \text{altrimenti} \end{cases}$$

La variabile $Y = [X]$ assume i valori interi $0, 1, ..., 20$, con

$$\begin{aligned}
&P(Y = 0) = P(0 \leq X < 1) = 1/20 \\
&P(Y = 1) = P(1 \leq X < 2) = 1/20 \\
&P(Y = 2) = P(2 \leq X < 3) = 1/20 \\
&... \\
&P(Y = 19) = P(19 \leq X < 20) = 1/20 \\
&P(Y = 20) = P(X = 20) = 0
\end{aligned}$$

Si verifica facilmente che la somma delle probabilità fa 1. La media è

$$E[Y] = 0\,\frac{1}{20} + 1\,\frac{1}{20} + 2\,\frac{1}{20} + ... + 19\,\frac{1}{20} + 20 \times 0 = \frac{19}{2} = 9.5$$

Anche la variabile $Z = [X + 0.5]$ assume i valori interi $0, 1, ..., 20$, ma con probabilità

$$\begin{aligned}
&P(Z = 0) = P(0 \leq X < 0.5) = 0.5/20 = 1/40 \\
&P(Z = 1) = P(0.5 \leq X < 1.5) = 1/20 \\
&P(Z = 2) = P(1.5 \leq X < 2.5) = 1/20 \\
&... \\
&P(Z = 19) = P(18.5 \leq X < 19.5) = 1/20 \\
&P(Y = 20) = P(19.5 \leq X \leq 20) = 0.5/20 = 1/40
\end{aligned}$$

Si verifica facilmente che la somma delle probabilità fa 1. La media è

$$E[Z] = 0\,\frac{1}{40} + 1\,\frac{1}{20} + 2\,\frac{1}{20} + ... + 19\,\frac{1}{20} + 20\,\frac{1}{40} = 10$$

3. Un insieme di misurazioni della temperatura corporea di un gruppo di bambini affetti da otite fornisce i seguenti dati, in gradi centigradi:

37.2, 37.6 36.4 36.5 37.5 37.1 36.9 37.5 36.6 36.8

Si sospetta che l'otite in questione abbia l'effetto di alzare la temperatura corporea dei bambini colpiti rispetto alla temperatura normale di una persona sana, che è di 36.7 gradi centigradi. Verificare tale ipotesi al livello del 95 %.

Soluzione. L'ipotesi nulla è che la temperatura corporea sia distribuita secondo una normale di media $\mu_0 = 36.7$ (gradi centigradi) e varianza incognita; l'ipotesi alternativa è che la media sia più alta:

$$H_0 : \mu = \mu_0; \qquad H_1 : \mu > \mu_0.$$

Il rango del campione è $n = 10$ ed il test è quindi un test di Student unilaterale (ad una coda) con 9 gradi di libertà. La media campionaria è $\overline{X}_n = 37.01$ e la varianza campionaria

$S_n^2 = 0.192$, quindi $S_n = 0.438$. Il valore del quantile di Student di riferimento è in questo caso $t_{0.05}(9) = 1.833$. La regione di rigetto (dell'ipotesi nulla) del test è

$$\overline{X}_n > \mu_0 + t_{0.05}(9)\frac{S_n}{n} = 36.7 + 1.833\frac{0.438}{10} = 36.78$$

Siccome $\overline{X}_n = 37.01 > 36.78$ L'ipotesi nulla è rifiutata e si può dire che i dati del campione confermano al 95 % che la temperatura si sia alzata per effetto dell'otite.

4. Si fa una rilevazione statistica sul numero di telefonate al minuto che arrivano ad un centralino. Un campione di 55 telefonate fornisce la seguente tabella:

telefonate/min.	frequenza oss.
0	1
1	4
2	6
3	9
4	12
5	9
6	7
7	4
8	3
9	2
10	1
> 10	0

Utilizzando il test del χ^2, possiamo affermare, con un livello di confidenza del 5% che il numero di telefonate al minuto che arrivano segua una legge normale?

Soluzione. Indichiamo con X la variabile casuale che rappresenta il numero di telefonate al minuto, con x_i i valori che può assumere e con ω_i ($i = 1, 2, ..., k$, con $k = 12$) la loro frequenza (osservata). Dobbiamo, quindi, calcolare preventivamente le probabilità che competono a ciascun valore x_i, vale a dire $p_i = \omega_i/n$. La media μ delle telefonate al minuto si calcola quindi come

$$\mu = \bar{X}_n = \sum_{i=1}^{k} p_i\, x_i = \sum_{i=1}^{k} \frac{\omega_i\, x_i}{n} = 4.25$$

mentre per la varianza usiamo come al solito lo stimatore non distorto S^2:

$$\sigma^2 = S^2 = \frac{n}{n-1}\sum_{i=1}^{k} p_i\,(x_i-\mu)^2 = \frac{1}{n-1}\sum_{i=1}^{k} \omega_i\,(x_i-\mu)^2 = 4.42$$

e quindi $\sigma = 2.1$.

Prestiamo ora attenzione agli intervalli in cui si deve suddividere il dominio della variabile X. Siccome X è una variabile discreta, la scelta più conveniente è quella di introdurre intervalli centrati attorno ai valori x_i, aggiungendo poi due intervalli illimitati di coda. La suddivisione in intervalli è raffigurata in Figura 1: $A_i = [a_i, b_i] \equiv [x_{i-1} - 0.5, x_{i-1} + 0.5]$ per $i = 2, 3, ..., 12$, mentre $A_1 = (-\infty, b_1] \equiv (-\infty, x_1 - 0.5]$ ed $A_{k+2} = [a_{k+2}, \infty) \equiv [x_k + 0.5, \infty)$ (con $k + 2 = 13$). Per calcolare le frequenze attese, dobbiamo prima calcolare la probabilità teorica che compete

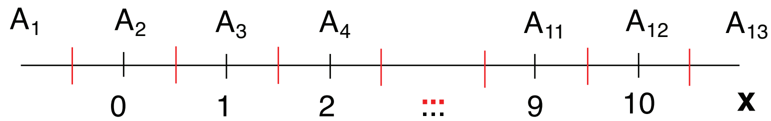

Figura 1: La suddivisione in intervalli centrati.

a ciascun valore x_i, che identifichiamo con la probabilità $\bar{p}_i$ da attribuire a ciascun intervallo, sulla base della distribuzione normale ipotizzata. Introducendo la variabile standardizzata

$$Z = \frac{X - \mu}{\sigma},$$

con i valori calcolati sopra per μ e σ, abbiamo per l'$i-$esimo intervallo:

$$\bar{p}_i = P(\{X \in A_i\}), \qquad i = 1, ..., k+2$$

Per gli intervalli limitati abbiamo:

$$\bar{p}_i = P(\{X \in A_i\}) = P\left(\left\{\frac{a_i - \mu}{\sigma} < Z < \frac{b_i - \mu}{\sigma}\right\}\right) = \Phi\left(\frac{b_i - \mu}{\sigma}\right) - \Phi\left(\frac{a_i - \mu}{\sigma}\right)$$

dove Φ è la funzione di ripartizione della distribuzione normale standard, da calcolare usando le tavole. Per gli intervalli di coda invece abbiamo:

$$\bar{p}_1 = P(\{X \in A_1\}) = P\left(\left\{Z < \frac{b_1 - \mu}{\sigma}\right\}\right) = \Phi\left(\frac{b_1 - \mu}{\sigma}\right)$$
$$\bar{p}_{k+2} = P(\{X \in A_{k+2}\}) = P\left(\left\{\frac{a_{k+2} - \mu}{\sigma} < Z\right\}\right) = 1 - \Phi\left(\frac{a_{k+2} - \mu}{\sigma}\right)$$

Le frequenze attese Ω_i sono quindi date da

$$\Omega_i = \bar{p}_i \, n$$

Riassumiamo i risultati ottenuti nella seguente tabella:

Intervallo	x_i	freq. oss. (ω_i)	$\bar{p}_i$	freq. attese (Ω_i)
A_1	< 0	0	0.012	0.65
A_2	0	1	0.025	1.38
A_3	1	4	0.058	3.19
A_4	2	6	0.11	5.88
A_5	3	9	0.16	8.68
A_6	4	12	0.19	10.27
A_7	5	9	0.18	9.72
A_8	6	7	0.13	7.37
A_9	7	3	0.081	4.48
A_{10}	8	2	0.040	2.18
A_{11}	9	1	0.015	0.85
A_{12}	10	1	0.048	0.26
A_{13}	> 10	0	0.00154	0.081

Prima di procedere con il calcolo della grandezza

$$w = \sum_i \frac{(\omega_i - \Omega_i)^2}{\Omega_i}$$

dobbiamo raggruppare le classi in modo che le frequenze attese siano ≥ 5. Raggruppando gli intervalli A_1, A_2 ed A_3 e gli intervalli A_9, A_{10}, A_{11}, A_{12} ed A_{13} otteniamo la seguente tabella, in cui omettiamo l'indicazione degli intervalli:

x_i	freq. oss. (ω_i)	freq. attese (Ω_i)
< 2	5	5.22
2	6	5.88
3	9	8.68
4	12	10.27
5	9	9.72
6	7	7.37
> 6	7	7.84

Il numero delle classi è ora uguale ad 7. Abbiamo usato i dati per determinare media e varianza, quindi il numero di gradi di libertà è 7-1-2=4. Il calcolo esplicito sui dati dell'ultima tabella dà $w = 0.48$, mentre $\chi^2_{0.05}(4) = 9.488$ (dalle tavole). L'ipotesi nulla è pertanto accettata.