Corso di Statistica - Prof. Fabio Zucca	Appello 4 - 2 febbraio 2017		
Nome e cognome:		**Matricola:**	

 8994

Esercizio 1 Tre monete equilibrate vengono lanciate contemporaneamente. Quelle che mostrano "testa" vengono messe da parte e si continua a lanciare le altre. Il processo ha termine quando ogni moneta ha mostrato "testa" dopo un lancio. Tutti i lanci sono indipendenti.

1. Calcolare la probabilità che nessuna moneta mostri testa al primo lancio.
2. Calcolare la probabilità che dopo 5 lanci ci sia ancora almeno una moneta da lanciare.
3. Calcolare la probabilità che dopo 5 lanci vi sia esattamente una sola moneta da lanciare.
4. Sapendo che dopo 5 lanci ho almeno una moneta da lanciare, quanto vale la probabilità che ne sia rimasta esattamente una sola da lanciare? (Giustificare adeguatamente la risposta)
5. Che legge ha il numero di monete rimaste da lanciare dopo 5 lanci?
6. Calcolare la probabilità che, lanciando contemporaneamente 100 monete, esca almeno 51 volte testa.

Soluzione. Ogni lancio è un processo di Bernoulli. Se la probabilità che esca testa è p (in questo caso $p = 1/2$) allora la variabile X che conta i lanci necessari a far uscire testa per la prima volta ha legge geometrica di parametro p. Quindi se le monete sono m (in questo caso $m = 3$) abbiamo una famiglia $\{X_i\}_{i=1}^m$ di variabili geometriche indipendenti di parametro p. Ricordiamo che $\mathbb{P}(X_i > n) = (1-p)^n$.

1. La probabilità che nessuna moneta mostri testa nei primi n lanci è $\mathbb{P}(X_i > n, \forall i = 1, \dots, m) = (1-p)^{mn}$. In questo caso $p = 1/2$, $n = 1$ e $m = 3$ da cui la probabilità $1/2^3 = 1/8$.

2. La probabilità che tutte le monete abbiano mostrato testa entro il lancio n (compreso) è $\mathbb{P}(X_i \leq n, \forall i = 1, \dots, m) = (1-(1-p)^n)^m$. La probabilità che rimanga almeno una moneta dopo il lancio m è l'evento complementare e ha probabilità $1-(1-(1-p)^n)^m$. In questo caso vale $1-(1-1/2^5)^3 = 1-(31/32)^3 \approx 0.090851$.

3. Questo punto può essere risolto conoscendo il risultato del punto 5. Senza utilizzare il punto 5, scomponiamo l'evento, tramite la formula delle probabilità totali, utilizzando gli eventi "rimane la moneta j" (al variare di j) e otteniamo

$$\begin{aligned}\mathbb{P}(\#\{i\colon X_i > n\} = 1) &= \sum_{j=1}^m \mathbb{P}(X_j > n, X_i \leq n \ \forall i \neq j) \\ &= m\mathbb{P}(X_1 > n, X_i \leq n \ \forall i > 1) = m(1-p)^n(1-(1-p)^n)^{m-1}\end{aligned}$$

   che in questo caso vale $3(1/32)(31/32)^2 \approx 0.087982$.

4. Osserviamo che l'evento $A := \{\#\{i\colon X_i > n\} = 1\}$ (è rimasta una sola moneta da lanciare dopo n lanci) è un sottoinsieme dell'evento $B := \{\#\{i\colon X_i > n\} \geq 1\}$ (è rimasta almeno una moneta da lanciare dopo n lanci), pertanto $\mathbb{P}(A|B) = \mathbb{P}(A)/\mathbb{P}(B)$ che in questo caso vale $3(1/2)^5(31/32)^2/(1-(31/32)^3) \approx 0.968425$.

5. Le monete si comportano in maniera indipendente, la probabilità di "sopravvivere" n lanci vale $(1-p)^n$ per ognuna di esse, quindi il numero di monete (su m) rimaste da lanciare dopo n lanci è una Binomiale $B(m, (1-p)^n)$ che in questo caso è $B(3, 1/32)$.

6. Essendo $100 \cdot 0.5 \geq 5$ possiamo utilizzare una approssimazione gaussiana. Se Z è il numero di teste su 100 lanci, si ha $Z \sim B(100, 1/2)$ da cui $Z \approx Y \sim \mathcal{N}(100 \cdot 0.5, 100 \cdot 0.5 \cdot 0.5) = \mathcal{N}(50, 25)$. Si ha

$$\begin{aligned}\mathbb{P}(Z \geq 51) = \mathbb{P}(Z > 50.5) \approx \mathbb{P}(Y > 50.5) &= \mathbb{P}\Big(\frac{Y-50}{\sqrt{25}} > \frac{50.5-50}{\sqrt{25}}\Big) \\ &= 1 - \Phi(0.1) \approx 1 - 0.5398278373 = 0.4601721627.\end{aligned}$$

Esercizio 2 Il numero di telefonate pubblicitarie giornaliere ricevute da un utente sono state osservate per un periodo di 30 giorni ed i risultati sono riportati nella seguente tabella:

8	0	0	1	3	4	0	2	12	5
1	8	0	2	0	1	9	3	4	5
3	3	4	7	4	0	1	2	1	2

Si può affermare che la legge che governa il numero di telefonate non sia di Poisson? Svolgere un test ad un livello di significatività pari a 1%.

Soluzione. Abbiamo la seguente tabella

N.Tel	0	1	2	3	4	5	6	7	8	9	10	11	12
f_{ass}	6	5	4	4	4	2	0	1	2	1	0	0	1

Uno stimatore per il parametro della legge di Poisson è la media campionaria $\overline{\lambda} = \sum_{i=0}^{12} i \cdot f_{ass}(i)/30 = 95/30$. Essendo $p_i = e^{-\overline{\lambda}}\overline{\lambda}^i/i!$ si ha

N.Tel	0	1	2	3	4	5	6	7	8	9	10	11	12
f_{ass}	6	5	4	4	4	2	0	1	2	1	0	0	1
$30 \cdot p_i$	1.26	4.00	6.34	6.69	5.30	3.35	1.77	0.80	0.32	0.11	0.04	0.01	0.003

Utilizziamo le seguenti classi

N.Tel	≤ 1	2	3	4	≥ 5
f_{ass}	11	4	4	4	7
$30 \cdot p_i$	5.27	6.34	6.69	5.30	6.40

Per il test si utilizza lo stimatore $Q = \sum_{i=1}^{N}(f_{ass}(i) - n \cdot p_i)^2/(n \cdot p_i)$ e la regione di rifiuto a livello α pari a $Q > \chi^2_{1-\alpha}(N-1-r)$ (in questo caso il numero di parametri stimati è $r = 1$ ed il numero delle classi $N = 5$). La stima $q = 6.24 + 0.86 + 1.08 + 0.32 + 0.06 = 8.56$. Dalle tabelle si ha $\chi^2_{0.95}(3) = 7.814725 < 8.56 = \chi^2_{1-\overline{\alpha}}(3) < 9.348404 = \chi^2_{0.975}(3)$ che implica $\overline{\alpha} \in (0.025, 0.05)$. Non possiamo rifiutare l'adattamento ad un livello di significatività pari a 1%. La stessa conclusione si potrebbe trarre dall'uso della regione di rifiuto e da $\chi^2_{0.99}(3) = 11.34488$.

Esercizio 3 Vogliamo capire quale tra le zone A e B sia più inquinata. Da 10 campioni della zona A, e 12 campioni dalla zona B si ottiene una media campionaria della concentrazione di NO_3 pari a $47mg/l$ per la zona A e di $50mg/l$ per la zona B. Se le varianze sono note e pari a $25(mg/l)^2$, si può affermare che la concentrazione media di NO_3 nella zona B è superiore a quella della zona A?

1. Scegliere un test ed un'ipotesi nulla opportuni;
2. scrivere la regione di rifiuto al livello $\alpha = 0.01$;
3. eseguire il test appena individuato e trarre la conclusione riguardo alla concentrazione;
4. quanto vale il p-value?

Soluzione.

1. Si tratta di un test di confronto di due medie a varianze note dove $H_0 : \mu_A - \mu_B \geq 0$ ed $H_1 : \mu_A - \mu_B < 0$.

2. La regione di rifiuto a livello α del test precedente è

$$\frac{\bar{x}_A - \bar{x}_B}{\sqrt{\sigma_X^2/n_X + \sigma_Y^2/n_Y}} < q_\alpha$$

   dove $q_{0.01} = -q_{0.99} = -2.3263479$.

3. Si ha $\bar{x}_A = 47$, $\bar{x}_B = 50$, $\sigma_X^2 = \sigma_Y^2 = 25$ da cui

$$\frac{\bar{x}_A - \bar{x}_B}{\sqrt{\sigma_X^2/n_X + \sigma_Y^2/n_Y}} = -\frac{3}{5\sqrt{1/10 + 1/12}} = -1.4013 > -2.3263479$$

   pertanto non è possibile rifiutare l'ipotesi nulla e non vi sono evidenze statistiche che la zona B sia più inquinata della zona A.

4. Il P-value si calcola come

$$\begin{aligned}\overline{\alpha} &= \phi\Big(\frac{\bar{x}_A - \bar{x}_B}{\sqrt{\sigma_X^2/n_X + \sigma_Y^2/n_Y}}\Big) = \Phi(-1.4013)\\ &= 1 - \Phi(1.4013) = 1 - 0.9194 = 0.0806.\end{aligned}$$