Corso di Statistica - Prof. Fabio Zucca	Appello 0 - 1 luglio 2016		
Nome e cognome:		**Matricola:**	

 8994

Esercizio 1 Il numero di auto che transitano da un singolo ingresso di un casello autostradale in 1 minuto è descritto da una variabile di Poisson di valore atteso 4. Gli ingressi del casello sono 10 ed i numeri delle auto che transitano da ciascun ingresso in un fissato intervallo di tempo sono indipendenti tra loro.

1. Quanto vale la probabilità che non transitino auto in 30 secondi in un fissato ingresso?
2. Calcolare la probabilità che, complessivamente, non transitino auto in 30 secondi attraverso l'intero casello.
3. Calcolare la probabilità che in almeno due ingressi su 10 non passino auto in 30 secondi.
4. Qual è il numero atteso della variabile aleatoria che conta il numero di auto che transitano complessivamente nei 10 ingressi in 1 ora? Quanto vale la varianza?
5. Quanti minuti dovremo attendere, come minimo, per vedere transitare complessivamente nei 10 ingressi almeno 4000 auto con probabilità maggiore di 0.9?
6. Sapendo che complessivamente sono transitate almeno 2400 auto in un'ora, quanto vale la probabilità che ne siano transitate al massimo 2500?

Soluzione. Sia $X_a(i)$ la variabile che conta il numero di auto che transitano dall'ingresso i in un intervallo temporale di lunghezza a (espresso in minuti). Chiaramente, al variare di i, le variabili sono indipendenti e $X_a(i) \sim \mathcal{P}(4a)$. Vista l'indipendenza tra i vari ingressi, il numero totale di auto $Y_a = \sum_{i=1}^{10} X_a(i)$ che transitano al casello in un intervallo temporale di lunghezza a è una variabile di legge $\mathcal{P}(40a)$.

1. $\mathbb{P}(X_{0.5} = 0) = \exp(-2) \approx 0.1353352832$.

2. $\mathbb{P}(Y_{0,5} = 0) = \exp(-2)^{10} = \exp(-20) \approx 2.06115362243856 \cdot 10^{-9}$.

3. Sia Z la variabile che conta il numero di caselli in cui non transitano auto in 30 secondi. Chiaramente $Z \sim B(10, \exp(-2))$; da cui

$$\mathbb{P}(Z \geq 2) = 1-\mathbb{P}(Z=0)+\mathbb{P}(Z=1) = 1-(1-\exp(-2))^2-10\cdot\exp(-2)(1-\exp(-2))^9 = 1-0.233602441-0.365629$$

4. Poiché $Y_a \sim \mathcal{P}(40a)$ si ha $\mathbb{E}[Y_{60}] = 2400$ e $\mathrm{var}(Y_{60}) = 2400$.

5. In generale $\mathbb{E}[Y_a] = \mathrm{var}(Y_a) = 40a$. Se $40a \geq 5$ possiamo applicare il TCL ottenendo $Y_a \approx H_a \sim \mathcal{N}(40a, 40a)$ e quindi

$$\begin{aligned} 0.9 \leq \mathbb{P}(Y_a \geq 4000) = \mathbb{P}(Y_a > 3999.5) &= \mathbb{P}\Big(\frac{Y_a - 40a}{\sqrt{40a}} > \frac{3999,5 - 40a}{\sqrt{40a}}\Big) \\ &\approx \mathbb{P}\Big(\frac{H_a - 40a}{\sqrt{40a}} > \frac{3999,5 - 40a}{\sqrt{40a}}\Big) = 1 - \Phi\Big(\frac{3999,5 - 40a}{\sqrt{40a}}\Big). \end{aligned}$$

Questo equivale a

$$\frac{3999,5 - 40a}{\sqrt{40a}} \leq q_{0.1} \approx -1.2815515655$$

cioè

$$\begin{cases} x = \sqrt{40a} \geq 0 \\ x^2 - 1.2815515655x - 3999.5 \geq 0 \end{cases}$$

che ha come unica soluzione accettabile $x \geq 63.8856221626$ che implica $a \geq (63.8856221626)^2/40 = 102.0343179775$ e le condizione di applicabilità del TCL sono soddisfatte.

6. Essendo $2400 > 5$, utilizziamo l'approssimazione gaussiana data dal TCL; pertanto $Y_{60} \approx H_{60} \sim \mathcal{N}(2400, 2400)$ da cui

$$\begin{aligned} \mathbb{P}(X \leq 2500 | X \geq 2400) &= 1 - \mathbb{P}(X > 2500.5 | X \geq 2400) = 1 - \mathbb{P}(X > 2500.5 | X > 2399.5) \\ &= 1 - \frac{\mathbb{P}(X > 2500.5)}{\mathbb{P}(X > 2399.5)} = 1 - \frac{1 - \Phi((2500.5 - 2400)/\sqrt{2400})}{1 - \Phi((2399.5 - 2400)/\sqrt{2400})} \\ &\approx 1 - \frac{1 - 0.9798883141}{1 - 0.4959283831} = 1 - 0.0398984692 = 0.9601015308. \end{aligned}$$

Esercizio 2 Un dado viene lanciato 100 volte ottenendo i seguenti risultati

n	1	2	3	4	5	6
$f_{ass}(n)$	15	18	15	17	16	19

1. È plausibile che il dado sia truccato ad un livello di significatività del 10%?
2. Stimare (specificando un intervallo) il P-value del precedente test.
3. Supponiamo ora che il dado venga lanciato 1000 volte e si ottengano gli stessi risultati del caso precedente moltiplicati per 10. È plausibile che il dado sia truccato ad un livello di significatività del 5%?
4. Stimare il P-value del test precedente.
5. Calcolare un intervallo di confidenza al 99% per la probabilità che esca 1 nel caso dei 1000 lanci descritto nei due punti precedenti. Quanti lanci occorrono come minimo per essere certi che l'intervallo di confidenza abbia ampiezza totale non superiore a 0.05 (non conoscendo a priori l'esito dei lanci)?
6. Supponiamo di lanciare ora un secondo dado per 200 volte ottenendo 32 volte il numero 1. Utilizzando i dati originali su 100 lanci del primo dado, possiamo ragionevolmente concludere che le probabilità che esca 1 sono differenti per i due dadi?

Soluzione. La regione di rifiuto del test di adattamento a livello α è $Q > \chi^2_{1-\alpha}(k-1)$ dove k è il numero di classi coinvolte e

$$Q = n \sum_{i^1}^{k} \frac{(p_i - f_{ass}(i)/n)^2}{p_i} = n \sum_{i^1}^{k} \frac{(p_i - f_{rel}(i))^2}{p_i}$$

(p_i probabilità teorica di appartenenza di un dato alla classe i-esima ed n ampiezza del campione). Nel caso specifico $p_i = 1/6$, $k = 6$ mentre $n = 100$ nel primo caso e $n = 1000$ nel secondo. Il P-value del test risolve l'equazione $Q = \chi^2_{1-\bar{\alpha}}(k-1)$. Il tast è applicabile in quanto $np_i \geq 5$.

1. Utilizzando i dati a disposizione si ottiene $q = 0.8$. Dalle tabelle $\chi^2_{0.9}(5) = 9.236349 > q$ quindi non si può rifiutare l'ipotesi nulla; pertanto non ci sono evidenze statistiche per affermare che il dado sia truccato (conclusione debole).

2. Dalle disuguaglianze $\chi^2_{0.025}(5) = 0.831209 > \chi^2_{1-\bar{\alpha}}(5) \equiv 0.8 > 0.554297 = \chi^2_{0.01}(5)$ e dalla monotonia del quantile si ha $0.025 > 1 - \bar{\alpha} > 0.01$ cioè $0.99 > \bar{\alpha} > 0.975$. Il calcolatore restituirebbe un valore $\bar{\alpha} \simeq 0.9770333438$.

3. Osserviamo che $f_{rel}(i) = f_{ass}(i)/n$ sono identiche al caso precedente (tutte le grandezze assolute sono state moltiplicate per 10. Rimane il fattore moltiplicativo n prima della somma che definisce Q; questo comporta che la stima q risulta pari a quella precedente moltiplicata per 10, quindi $q = 8$. Essendo ancora $\chi^2_{0.9}(5) = 9.236349 > q$, non si può rifiutare l'ipotesi nulla $\chi^2_{0.9}(5) = 9.236349 < q$ al 10%: non ci sono quindi ragioni sufficienti per affermare che il dado sia truccato nemmeno al 5% (in ogni caso $\chi^2_{0.95}(5) = 11.07048$).

4. Dalle disuguaglianze $\chi^2_{0.9}(5) = 9.236349 > \chi^2_{1-\bar{\alpha}}(5) \equiv 8 > 1.610309 = \chi^2_{0.1}(5)$ e dalla monotonia del quantile si ha $0.9 > 1 - \bar{\alpha} > 0.1$ cioè $0.9 > \bar{\alpha} > 0.1$. Il calcolatore restituirebbe un valore $\bar{\alpha} \simeq 0.1562356276$.

5. Il numero di volte che esce 1 è descritto da una Binomiale $\mathcal{B}(1000, p)$ di parametro p incognito. La stima di p è $\bar{p}_{1000} = 150/1000$. Si osservi che $n\bar{p}_{1000} \geq 5$ e $n(1 - \bar{p}_{1000}) \geq 5$ pertanto l'approssimazione gaussiana è buona. Un'intervallo di confidenza a livello $\alpha = 0.99$, ha come estremi

$$\begin{aligned} p^{\pm} &= \bar{p}_{1000} \pm z_{(1+0.99)/2} \cdot \sqrt{\frac{\bar{p}_{1000}(1-\bar{p}_{1000})}{1000}} = 0.15 \pm 2.5758293035 \cdot \sqrt{0.0001131} \\ &= 0.15 \pm 2.5758293035 \cdot 0.0106348484 = \begin{cases} 0.1773935541 \\ 0.1226064459 \end{cases} \end{aligned}$$

   L'ampiezza massima di un intervallo di confidenza a livello α di una Bernoulli si calcola come

$$2 \cdot z_{(1+\alpha)/2} \sqrt{\bar{p}(1-\bar{p})/n} \leq 2 \cdot z_{(1+\alpha)/2} \sqrt{1/4n} = z_{(1+\alpha)/2}/\sqrt{n}$$

   da cui $n \geq (z_{0.975}/0.05)^2 = 2653.9586404085$ cioè $n \geq 2654$.

6. Si tratta di un test del tipo $H_0 : p_1 = p_2$ contro $H_1 : p_1 \neq p_2$. Notiamo che il test si può condurre, essendo $100\bar{p}_{1,100} = 15 \geq 5$, $100(1 - \bar{p}_{1,100}) = 100 - 15 \geq 5$, $200\bar{p}_{2,200} = 32 \geq 5$ e $200(1 - \bar{p}_{2,200}) = 200 - 32 \geq 5$.

   La zona di rifiuto è

$$|t| > t_{1-\alpha/2}(298) \approx q_{1-\alpha/2}$$

   dove $t = (\bar{p}_{1,100} - \bar{p}_{2,200})/\sqrt{\hat{p}(1-\hat{p})(1/100 + 1/200)}$ e $\hat{p} = (15 + 32)/(100 + 200) = 47/300 = 0.1566666667$. Pertanto $t = (0.15 - 0.16)/\sqrt{0.1321222222 \cdot (1/100 + 1/200)} = -0.01/\sqrt{0.0019818333} = -0.2246293165$. Non essendo fornito alcun livello di significatività, calcoliamo il P-value (utilizzando il fatto che i gradi di libertà sono 298 si può passare all'uso di una normale $\mathcal{N}(0,1)$) $\bar{\alpha} = 2 - 2\Phi(|t|) = 2 - 2\Phi(0.2246293165) = 2 - 2 \cdot 0.5888661718 = 0.8222676564$. Un P-value così grande è una chiara indicazione dell'impossibilità di rifiutare H_0, quindi i due dadi sembrano possedere la stessa probabilità di far uscire il numero 1.

   Per chi volesse utilizzare un test meno preciso, $t = (\bar{p}_{1,100} - \bar{p}_{2,200})/\sqrt{p_{1,100}(1-p_{1,100})/100 + p_{2,200}(1-p_{2,200})/20}$ si ha $t = -0.01/0.045271404 = -0.226629803$. Pertanto $\bar{\alpha} = 2 - 2\Phi(|t|) = 0.820711615$ che è una chiara indicazione dell'impossibilità di rifiutare H_0, quindi i due dadi sembrano possedere la stessa probabilità di far uscire il numero 1.