Corso di Statistica - Prof. Fabio Zucca	Appello 4 - 4 febbraio 2016		
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 $\mathbb{P}(X_i > n, \forall i = 1, \ldots, m) = (1-p)^{mn}$. 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 $\mathbb{P}(X_i \leq n, \forall i = 1, \ldots, m) = (1-(1-p)^n)^m$. 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 $1-(1-1/2^5)^3 = 1-(31/32)^5 \approx 0.146784812$.

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 $3(1/32)(31/32)^4 \approx 0.082569212$.

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 $3(1/2)^5(31/32)^4/(1-(31/32)^5) \approx 0.562518769$.

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 Siamo di nuovo alle elezioni ed i segretari di due partiti, quello "dei pinguini" e quello "delle mele morsicate", si rivolgono a noi per un sondaggio. Intervistiamo 1000 persone delle quali 320 si dicono intenzionate a votare per il partito dei pinguini. In un altra intervista fatta a 500 persone 140 si dicono intenzionate a votare per il partito delle mele mezze morsicate.

1. Sapendo che alle ultime elezioni, il partito dei pinguini ha preso il 28% dei voti, è ragionevole concludere che l'intenzione di voto al momento porterebbe ad un risultato migliore?

2. Cosa possiamo concludere scambiando tra loro le ipotesi del precedente test?

3. Indicare un intervallo di confidenza al 90% per la percentuale di voto a favore di ciascuno dei due partiti.

4. Si può concludere a livello del 10% che la percentuale di voti che il partito dei pinguini prederebbe al momento sia superiore dell'1% a quelli che prenderebbe il partito delle mele mezze morsicate?

Soluzione. Chiamiamo $\{x_i\}_{i=1}^{1000}$ e $\{y_i\}_{i=1}^{500}$ i dati dei campioni relativi ai partiti "dei pinguini" e "delle mele morsicate" rispettivamente. Si tratta di un modello Bernoulliano in cui, essendo $1000 - 320 > 320 \geq 5$ e $500 - 140 > 140 \geq 5$, si può applicare un'approssimazione gaussiana e le condizioni usuali di applicazioni dei test ed intervalli di confidenza per popolazioni di Bernoulli sono soddisfatte. Ovviamente $\bar{x}_n = 0.32$ ($n = 1000$) e $\bar{y}_m = 0.28$ ($m = 500$). Chiamiamo p e q le probabilitá vere che un elettore preso a caso voti per il partito "dei pinguini" e quello "delle mele morsicate" rispettivamente (si tratta in effetti delle vere percentuali di voti che prenderebbero i partiti allo stato attuale; infatti (approccio classico) l'estrazione equidistribuita in una popolazione, avviene all'interno di una sottopopolazione di peso p esattamente con probabilità p).

1. Il test è per una popolazione di Bernoulli con $H_0\colon p \leq 0.28 =: p_0$ contro $H_1\colon p > 0.28$; La statistica test è $(\bar{x}_n - p_0)/\sqrt{p_0(1-p_0)/n}$. La regione di rifiuto a livello α ed il P-value sono

$$\frac{\bar{x}_n - p_0}{p_0(1-p_0)/n} > z_{1-\alpha}, \qquad \bar{\alpha} = 1 - \Phi\Big(\frac{\bar{x}_n - p_0}{p_0(1-p_0)/n}\Big).$$

Dai dati

$$\bar{\alpha} = \Phi\Big(\frac{0.32 - 0.28}{\sqrt{0.28 \cdot 0.72/1000}}\Big) = 1 - \Phi(2.8171808491) = 1 - 0.9975776381 = 0.0024223619.$$

Ci sono quindi evidenze statistiche per supportare un risultato migliore.

2. Se scambiamo tra loro le ipotesi del precedente teorema si vede subito dai dati che saremmo costretti ad accettare l'ipotesi nulla per ogni livello $\alpha < 0.5$. Quantitativamente, il nuovo P-value $\bar{\alpha}' = 1 - \bar{\alpha}$ dove $\bar{\alpha}$ è il P-value calcolato al punto precedente; da cui $\bar{\alpha}' = 0.9975776381$. I due test in questo caso sono concordi.

3. L'usuale formula per il calcolo di un intervallo di confidenza per una popolazione di Bernoulli in questo caso è

$$[\bar{x}_n - q_{(1+\alpha)/2}\sqrt{\frac{\bar{x}_n(1-\bar{x}_n}{n}}, \bar{x}_n + q_{(1+\alpha)/2}\sqrt{\frac{\bar{x}_n(1-\bar{x}_n}{n}}], \qquad [\bar{y}_m - q_{(1+\alpha)/2}\sqrt{\frac{\bar{y}_m(1-\bar{y}_m}{m}}, \bar{y}_m + q_{(1+\alpha)/2}\sqrt{\frac{\bar{y}_m(1-\bar{y}_m}{m}}]$$

In questo caso $n = 1000$, $\bar{x}_{1000} = 0.32$, $m = 500$, $\bar{y}_{500} = 0.28$, $\alpha = 0.9$ da cui $q_{0.95} = 1.644853627$. Quindi le due semiampiezze sono rispettivamente

$$1.644853627\sqrt{0.32 \cdot 0.64}1000 = 0.0235392289 \qquad \text{e} \qquad 1.644853627\sqrt{0.28 \cdot 0.72}500 = 0,0330283987$$

da cui gli intervalli

$$[0.2964607711, 0.3435392289] \qquad \text{e} \qquad [0.2469716013, 0.3130283987].$$

4. Il test in questione è un test a due popolazioni in cui $H_0 : p - q \leq 0.01 =: \delta$ e $H_1 : p - q > 0.01$. La statistica test è $(\bar{x}_n - \bar{y}_m - \delta)/\sqrt{\hat{p}(1-\hat{p})(1/n + 1/m)}$ dove $\hat{p} = (n\bar{x}_n + m\bar{y}_m)/(m+n) = (320 + 140)/(1000 + 500) = 23/75$. La regione di rifiuto a livello α ed il P-value sono

$$\frac{\bar{x}_n - \bar{y}_m - \delta}{\hat{p}(1-\hat{p})(1/n + 1/m)} < z_\alpha, \qquad \bar{\alpha} = \Phi\Big(\frac{\bar{p}_n - \bar{q}_m - \delta}{\sqrt{\hat{p}(1-\hat{p})(1/n + 1/m)}}\Big).$$

Dai dati

$$\frac{\bar{x}_n - \bar{y}_m - \delta}{\hat{p}(1-\hat{p})(1/n + 1/m)} = \frac{0.32 - 0.28 - 0.01}{\sqrt{(23/75)(52/75) \cdot 0.003}} = 1.1878355009.$$

$$\bar{\alpha} = 1 - \Phi(1.1878355009) = 1 - 0.8825508876 = 0.1174491124.$$

Non ci sono quindi evidenze statistiche per rifiutare l'ipotesi nulla al 10%.