Corso di Statistica - Prof. Fabio Zucca		Appello 1 - 4 luglio 2018	
Nome e cognome:		**Matricola:**	

 8994

Esercizio 1 Si considerino $k \in \mathbb{N}$ ($k > 0$) lanci indipendenti di una moneta in cui la probabilità di avere testa in un lancio è pari a $p \in (0, 1)$.

1. Calcolare la probabilità di avere almeno una croce in k lanci.
2. Se $k \geq 2$, calcolare la probabilità di avere almeno una testa in k lanci dato che è uscita almeno una croce.
3. Ripeto 1500 volte il seguente esperimento: lancio 4 volte una moneta equilibrata e vedo se riesco ad ottenere almeno una croce. Qual è la probabilità che l'esperimento riesca almeno 1300 volte?

Chi non riesce a risolvere l'esercizio in forma astratta, lo risolva nel caso specifico $p = 1/2$ e $k = 4$.

Soluzione. Ovviamente se $\{X_i\}_{i\geq 1}$ sono variabili bernoulliane i.i.d. allora il numero totale di teste ottenute $X := \sum_{i=1}^k X_i$ è una binomiale $B(k,p)$.

1. $\mathbb{P}(X \leq k-1) = 1 - \mathbb{P}(X = k) = 1 - p^k$. Nel caso specifico $15/16$.

2. Si ha $\mathbb{P}(X \geq 1 | X \leq k-1) = \mathbb{P}(1 \leq X \leq k-1)/\mathbb{P}(X \leq k-1) = (1 - \mathbb{P}(X = 0) - \mathbb{P}(X = k))/\mathbb{P}(X \leq k-1) = (1 - p^k - (1-p)^k)/(1-p^k) = 1 - (1-p)^k/(1-p^k)$. Nel caso specifico $14/15$.

3. Siano $\{Z_i\}_{i=1}^{1500}$ i risultati degli esperimenti (1 se esce almeno una croce e 0 altrimenti), sono tutte variabili Bernoulliane indipendenti di parametro $15/16$. Pertanto, essendo $1500 \cdot 15/16 > 1500/16 > 5$ posso utilizzare l'approssimazione gaussiana da cui

$$\begin{aligned}\mathbb{P}(\sum_i Z_i \geq 1300) &= \mathbb{P}((\sum_i Z_i - 1500 \cdot 15/16)/\sqrt{1500 \cdot 15/16^2} \geq (1299.5 - 1500 \cdot 15/16)/\sqrt{1500 \cdot 15/16^2} \\ &\approx 1 - \Phi(-106.75 \cdot 16/150) \approx 1.\end{aligned}$$

Esercizio 2 Si consideri la seguente funzione reale di variabile reale (dipendente da un parametro $c \in \mathbb{R}$)

$$f_c(x) := \begin{cases} c^2/4 & x \in [-1, 1] \\ (2c-1)/3 & x \in (1, 2] \\ (2c-1)/6 & x = [-2, -1) \\ 0 & \text{altrove.} \end{cases}$$

1. Calcolare tutti i valori di c affinché f_c sia una densità di una variabile assolutamente continua.
2. Sia X una variabile aleatoria di densità f_c (per i valori di c calcolati al punto precedente): quanto vale $\mathbb{P}(X \geq 1/2)$?

Soluzione.

1. Le condizioni
$$\begin{cases} f_c(x) \geq 0 & \forall x \in \mathbb{R} \\ \int_{\mathbb{R}} f_c(x)\mathrm{d}x = 1 \end{cases}$$
sono equivalenti a
$$\begin{cases} c \geq 1/2 \\ \frac{c^2}{2} + \frac{2c-1}{3} + \frac{2c-1}{6} = 1 \end{cases}$$
cioè
$$\begin{cases} c \geq 1/2 \\ c^2 + 2c - 3 = 0 \end{cases}$$
che diviene
$$\begin{cases} c \geq 1/2 \\ c = 1 \text{ oppure } c = -3 \end{cases}$$
da cui si ha l'unica soluzione $c = 1$. La densità è
$$f_1(x) := \begin{cases} 1/4 & x \in [-1, 1] \\ 1/3 & x \in (1, 2] \\ 1/6 & x = [-2, -1) \\ 0 & \text{altrove.} \end{cases}$$

2.
$$\begin{aligned} \mathbb{P}(X \geq 1/2) &= \int_{[1/2,+\infty)} f_1(x)\mathrm{d}x = \int_{[1/2,1]} f_1(x)\mathrm{d}x + \int_{(1,2]} f_1(x)\mathrm{d}x + \int_{(2,+\infty)} f_1(x)\mathrm{d}x \\ &= \frac{1}{8} + \frac{1}{3} = \frac{11}{24}; \end{aligned}$$

Esercizio 3 Vogliamo valutare l'efficacia di un metodo di memorizzazione. Allo scopo vengono individuati $n = 10$ soggetti ai quali viene fatto leggere un brano prima e dopo aver seguito una sessione di addestramento della memoria. I dati raccolti, che possiamo considerare provenienti da due popolazioni normali accoppiate $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ e $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, si riferiscono al numero di parole memorizzate da ogni individuo, prima (X) e dopo (Y) la sessione di addestramento. Il metodo viene considerato efficace se il numero di parole del brano memorizzate aumenta dopo l'addestramento.

Condurre tramite l'analisi del p-value un opportuno test di ipotesi per verificare se è ragionevole supporre che il metodo sia efficace.

Allo scopo potrebbero essere utili alcune delle seguenti informazioni:

X	$\sum_{i=1}^n x_i = 3152$	$\sum_{i=1}^n x_i^2 = 1062158$
Y	$\sum_{i=1}^n y_i = 3390$	$\sum_{i=1}^n y_i^2 = 1240820$
$Z = Y - X$	$\sum_{i=1}^n z_i = 238$	$\sum_{i=1}^n z_i^2 = 11160$

Soluzione. Ci è richiesto di condurre un test di ipotesi per il confronto tra le medie di due popolazioni normali accoppiate con varianza incognita.

L'ipotesi da sottoporre a test è $H : \mu_Y > \mu_X$. Vogliamo accettare H solo di fronte ad una evidenza statistica quindi $H_0 : \mu_Y \leq \mu_X$ contro $H_1 : \mu_Y > \mu_X$.
Poiché i campioni sono accoppiati, il test si riduce al test sulla media di una popolazione normale $Z = Y - X$. Dunque le ipotesi semplificate diventano $H_0 : \mu_Z \leq \mu_0$ contro $H_1 : \mu_Z > \mu_0$ con $\mu_0 = 0$.

La statistica test è $t_n = \frac{\bar{z}_n - \mu_0}{s_n}\sqrt{n} = \frac{23.8-0}{24.710}\sqrt{10} \approx 3.046$ dove $n = 10$, $\bar{z}_n = \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{238}{10} = 23.8$, $s_n = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} z_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} z_i\right)^2\right)} = \frac{1}{3}\sqrt{11160 - \frac{1}{10}\left(238\right)^2} \approx 24.710$

Il p-value è $\bar{\alpha} : t_{1-\bar{\alpha}}(n-1) = t_n$ ovvero $1 - \bar{\alpha} = F_{T_{n-1}}(t_n)$ ovvero il complemento a 1 della funzione di ripartizione di una t di student con $n-1$ gradi di libertà valutata in t_n. Sulle tavole dei quantili della t di student, al rigo corrispondente a 9 gradi di libertà , individuiamo gli estremi di contenimento $t_{0.99}(9) = 2.821434 < T_n < t_{0.995}(9) = 3.249843$ da cui $0.99 < 1 - \bar{\alpha} < 0.995$ e infine $0.005 < \bar{\alpha} < 0.01$. Possiamo quindi concludere che l'ipotesi nulla è rifiutata $\forall \alpha \geq 0.01$ ovvero che, per tali valori di significatività, esistono evidenze statistiche per supporre che il metodo sia efficace.

 8994

Esercizio 4 Di un campione $\{x_i\}_{i=1}^n$ costituito da $n = 10$ dati sulla concentrazione di iodio in una soluzione, vengono registrate le seguenti informazioni:
$\sum_{i=1}^n x_i = 49.531$ e $\sum_{i=1}^n (x_i - \bar{x}_n)^2 = 0.000703$

Nell'ipotesi che i dati provengano da una popolazione gaussiana avente media e varianza incognita:

1. si costruisca un intervallo fiduciario bilatero al 90% per σ^2;
2. si stabilisca ad un livello di significatività $\alpha = 0.01$ se è ragionevole supporre che il valore vero della varianza sia inferiore a $\sigma_0^2 = 0.001$.

Soluzione.

1. L'intervallo di confidenza per la varianza di una popolazione normale con media incognita è dato da:

$$I \equiv \left(\frac{(n-1)S_n^2}{\chi^2_{1-\beta}(n-1)}; \frac{(n-1)S_n^2}{\chi^2_{\beta}(n-1)} \right)$$

   dove: $n = 10$; $(n-1)S_n^2 = \sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = 0.000703$; $\beta = \frac{1-\alpha}{2} = \frac{1-0.90}{2} = 0.05$; $1 - \beta = 0.95$.
   Pertanto: $\chi^2_{\beta}(n-1) = \chi^2_{0.05}(9) = 3.325$ e $\chi^2_{1-\beta}(n-1) = \chi^2_{0.95}(9) = 16.92$ e l'intervallo richiesto è:

$$I \equiv \left(\frac{0.000703}{16.92}; \frac{0.000703}{3.325} \right) \equiv (0.0000416; 0.0002115)\,.$$

2. Ci è richiesto di condurre un test di ipotesi per la varianza di una popolazione normale avente media incognita. L'ipotesi da sottoporre a test è $H : \sigma^2 < \sigma_0^2$. Vogliamo accettare H solo di fronte ad una evidenza statistica quindi: $H_0 : \sigma^2 \geq \sigma_0^2$ contro $H_1 : \sigma^2 < \sigma_0^2$.

   La regione di rifiuto è $R = \left[0; \chi^2_{\beta}(n-1)\right]$ dove $n - 1 = 9$, $\beta = \alpha = 0.01$ e il valore critico ricavato dalle tavole è $\chi^2_{0.01}(9) = 4.168$.

   La statistica test è $W_n = \frac{(n-1)S_n^2}{\sigma_0^2} = \frac{0.000703}{0.001} = 0.703$.

   La condizione di rifiuto è : rifiuto se $W_n < \chi^2_{\beta}(n-1)$ ovvero se $0.703 < 4.168$ il che è vero pertanto rifiuto H_0 quindi: è ragionevole supporre che il valore vero della varianza sia inferiore a σ_0^2.