Statistica Matematica A

Docente: Dott. F. Zucca

II appello

Lecco 27/07/2005

<u>**Candidato**</u>

Cognome: .... **Nome:** ....

Corso di Laurea: ....

Matricola: .... **Anno di corso:** ....

tempo a disposizione **2h30m**.

1 Esercizi

Esercizio 1 Si consideri la seguente funzione reale di variabile reale (dipendente da un parametro $c \in \mathbb{R}$)

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

1. Calcolare tutti i valori di c affinché f_c sia una densità discreta.
2. Sia X una variabile aleatoria di densità f_c (per i valori di c calcolati al punto precedente): si calcoli la media e la varianza di X. Quanto vale $\mathbb{P}(X \geq 1/2)$? E $\mathbb{P}(X > 1/2)$?
3. Siano $X_1, \ldots, X_{100}$ variabili aleatorie i.i.d. con distribuzione calcolata al punto (1). Calcolare (approssimativamente)

$$\mathbb{P}\left(\frac{1}{100}\sum_{i=1}^{100} X_i \geq \frac{1}{2}\right).$$

Soluzione.

1. Le condizioni
$$\begin{cases} f_c(x) \geq 0 & \forall x \in \mathbb{R} \\ \sum_x f_c(x) = 1 \end{cases}$$
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 $c = 1$.

2.
$$\begin{aligned} \mathbb{E}(X) &= \sum_x x f_1(x) = \frac{1}{6}; \\ \mathrm{var}(X) &= \sum_x x^2 f_1(x) - \frac{1}{6^2} = \frac{43}{72} =: \sigma^2; \\ \mathbb{P}(X \geq 1/2) &= \sum_{x \geq 1/2} f_1(x) = \frac{7}{12}; \\ \mathbb{P}(X > 1/2) &= \sum_{x > 1/2} f_1(x) = \frac{1}{3}. \end{aligned}$$

3. Dal Teorema Centrale del Limite
$$\frac{1}{100} \sum_{i=1}^{100} X_i \approx Y \sim \mathcal{N}\left(\frac{1}{6}, \frac{\sigma^2}{100}\right)$$
pertanto, essendo $\sigma = \sqrt{43/72} \approx 0.7728$,
$$\begin{aligned} \mathbb{P}\left(\frac{1}{100} \sum_{i=1}^{100} X_i \geq 1/2\right) &\approx \mathbb{P}(Y \geq 1/2) = \mathbb{P}\left(\frac{Y - 1/6}{\sigma/10} \geq \frac{1/2 - 1/6}{\sigma/10}\right) \\ &= 1 - \Phi\left(\frac{10}{3\sigma}\right) 1 - \Phi\left(4.3133\right) \approx 0. \end{aligned}$$
In questo caso la correzione di continuità risulta trascurabile.

Esercizio 2 Uno pneumatico per la formula 1 esplode quando 91 bolle si sono formate sulla sua superficie. Durante una gara il numero di bolle che si possono formare su ciascuna ruota è una variabile aleatoria con distribuzione di Poisson di media 100.

1. Calcolare la probabilità che durante una gara un fissato pneumatico non esploda.
2. Calcolare la probabilità che un auto riesca a finire la gara senza dover cambiare i pneumatici (si suppongano i pneumatici indipendenti).
3. Calcolare la probabilità che esplodano esattamente due pneumatici su quattro.
4. Quanti pneumatici dovremmo utilizzare affinché almeno 80 arrivino "sani" alla fine della gara con probabilità almeno pari al 90%?

(Sugg: per il punto (1) si utilizzi un'opportuna approssimazione)

Soluzione.

1. Sia X il numero di bolle, allora $X \approx Y \sim \mathcal{N}(100, 100)$; la probabilità cercata è
$$p = \mathbb{P}(X \leq 90) = \mathbb{P}(X \leq 90.5) \approx \mathbb{P}((Y - 100)/\sqrt{100} \leq -9.5/\sqrt{100}) = \Phi(-0.95) \approx 0.1711$$
2. Per l'indipendenza la probabilità cercata è $p^4 \approx 8.5704 \cdot 10^{-4}$.
3. Sia Z_i la variabile di Bernoulli che controlla l'efficienza dell'i-esimo pneumatico (vale 1 se non esplode). Allora la somma $Z := \sum_{i=1}^4 Z_i \sim B(4, p)$ e
$$\mathbb{P}(Z = 2) = \binom{4}{2} p^2 (1 - p)^2 \approx 6 \cdot 0.1711^2 \cdot 0.8289^2 = 0.1207.$$
4. Come nel caso precedente siano $\{Z_i\}_{i \in \mathbb{N}^*}$ le variabili di Bernoulli relative ai pneumatici. Dal Teorema Centrale del Limite, $N_n := \sum_{i=1}^n Z_i \approx Y_n \sim \mathcal{N}(np, np(1 - p))$ dove p è il valore calcolato al punto (1).

   Con la correzione di continuità
$$0.9 \leq \mathbb{P}(N_n \geq 79.5) \approx 1 - \Phi\left(\frac{79.5 - np}{\sqrt{np(1 - p)}}\right)$$
   che è equivalente a
$$\frac{79.5 - np}{\sqrt{np(1 - p)}} \leq q_{0.1} \approx -1.282.$$
   Con la sostituzione $t = \sqrt{n}$ si risolve
$$\begin{cases} 0.1711t^2 - 0.4828t - 79.5 \geq 0 \\ t \geq 0 \end{cases}$$
   che ha come unica soluzione
$$t = \frac{0.4828 + \sqrt{0.4828^2 + 4 \cdot 79.5 \cdot 0.1711}}{2 \cdot 0.1711} \approx 23.0125$$
   da cui
$$n \geq \lceil 23.0125^2 \rceil = 530.$$

Esercizio 3 Nell'ultima prova scritta di Statistica Matematica A, Arturo Cecato detto *fisheye non è riuscito a risolvere un esercizio riguardante il calcolo di un intervallo di confidenza per una proporzione relativo ad un campione di ampiezza pari a 100. Tuttavia guardando il compito del suo vicino, prima che l'attento docente lo scoprisse e lo scaraventasse fuori dall'aula, era riuscito a leggere la media campionaria della proporzione e l'estremo sinistro dell'intervallo: 0.65 e 0.5715 rispettivamente.*

1. *Calcolare l'estremo destro e il livello di confidenza.*
2. *Testare l'ipotesi nulla* $H_0 : p \geq 0.7$ *a livello* 10%.

Soluzione.

1. Ricordiamo che gli estremi di un intervallo di confidenza per una proporzione si calcolano come
$$p^{\pm} = \overline{p} \pm q_{(1+\alpha)/2} \sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$
dove $n = 100$, $p^- = 0.5715$ ed $\overline{p} = 0.65$. Poiché $\overline{p}$ è il punto medio di tale intervallo si ha che $p^+ = 2\overline{p} - p^- = 0.7285$. Inoltre $\sqrt{\overline{p}(1-\overline{p})}/10 = 0.0477$ da cui
$$q_{(1+\alpha)/2} \sqrt{\overline{p}(1-\overline{p})}/10 = p^+ - \overline{p} = 0.07285$$
pertanto $q_{(1+\alpha)/2} \approx 1.6447$ che implica $(1+\alpha)/2 \approx 0.95$ ed infine $\alpha \approx 0.9$.

2. La regione di accettazione ha come margine destro
$$p^- = p_0 + q_\alpha \sqrt{\frac{p_0(1-p_0)}{n}}$$
dove $p_0 = 0.7$, $q_\alpha = q_{0.1} = -1.282$ ed $n = 100$; pertanto $p^- := 0.7 - 1.282 \cdot 0.0458 \approx 0.6413$. Poichè $\overline{p} \geq 0.6413$ si ha che l'ipotesi nulla non può essere rifiutata.

Esercizio 4 Il numero di difetti presenti in una lastra di 1 mq di alluminio è una variabile aleatoria. I risultati dell'esame di 100 lastre sono riassunti nella seguente tabella:

Numero difetti	0	1	2	3	4	5	6	≥ 7
Numero lastre	15	20	20	17	14	10	4	0

Valutare l'adattamento ad una distribuzione di Poisson mediante un opportuno test di cui sappiate valutare il P-value. Cosa si conclude al livello $\alpha = 0.05$? Ed al livello $\alpha = 0.1$?

Soluzione. Stimiamo dapprima il parametro della Poisson mediante la relazione
$$\lambda = \frac{\sum_{i=0}^{6} i n_i}{n} = \frac{20 + 2 \cdot 20 + 3 \cdot 17 + 4 \cdot 14 + 5 \cdot 10 + 6 \cdot 4}{100} = 2.41$$

completiamo quindi la tabella con i dati teorici utilizzando la formula $nt_i := 100\lambda^i \exp(-\lambda)/i!$:

Numero difetti	0	1	2	3	4	5	6	≥ 7
N. lastre (n_i)	15	20	20	17	14	10	4	0
N. lastre teor. (nt_i)	8.98	21.65	26.08	20.95	12.62	6.08	2.44	1.2

Calcoliamo quindi la quantità pivotale

$$Q := \sum_{i=0}^{7} (n_i - nt_i)^2 / nt_i - 100 = 11.2$$

da confrontare con i quantili di una distribuzione chiquadrato con 6 gradi di libertà (8 classi ed un parametro stimato). Poichè

$$\chi^2_{0.9}(6) = 10.64 < 11.2 = \chi^2_{1-\overline{\alpha}}(6) < 12.59 = \chi^2_{0.95}(6)$$

si ha che $0.05 < \overline{\alpha} < 0.1$, per cui l'adattamento ad una legge di Poisson è in dubbio poiché dipende fortemente dal livello di affidabilità scelto.

Nel caso $\alpha = 0.05$ l'ipotesi di adattamento non può essere rifiutata, mentre nel caso $\alpha = 0.1$ viene rifiutata.