Esercitazioni di Statistica Matematica A
Lezione 6

Applicazioni della legge dei grandi numeri e della formula di Chebicev

1.1) Sia $\{X_i\}_{i\in\mathbb{N}}$ una successione di variabili aleatorie i.i.d. (**i**ndipendenti ed **i**denticamente **d**istribuite) tali che $E(X_i) = \mu \in \mathbb{R}$ e $\mathrm{Var}(X) = \sigma^2 \in \mathbb{R}$. Mostrare che:

a) se $\lambda < \mu$ allora $\lim_{n\to+\infty} \mathbb{P}(\sum_{i=1}^n X_i \le n\lambda) = 0$;

b) se $\lambda > \mu$ allora $\lim_{n\to+\infty} \mathbb{P}(\sum_{i=1}^n X_i \le n\lambda) = 1$;

c) se $\mathrm{Var}(X_i) < +\infty$ e $\lambda = \mu$ allora $\lim_{n\to+\infty} \mathbb{P}(\sum_{i=1}^n X_i \le n\lambda) = 1/2$.

d)

$$\lim_{n\to+\infty} e^{-n\mu} \sum_{i=0}^{[\lambda n]} \frac{(n\mu)^i}{i!} = \begin{cases} 0 & \text{se } \lambda < \mu \\ 1 & \text{se } \lambda > \mu \\ \frac{1}{2} & \text{se } \lambda = \mu. \end{cases}$$

1.2) Due amici fanno una gara su una pista circolare. Il primo dei due (chiamiamolo A) si muove con una velocità media di 10 km/h, mentre il secondo (sia esso B) ha una velocità media 9.99 km/h (attenzione, le velocità non sono costanti!). Il concorrente A concede un vantaggio di c km a B; mostrare che, a patto di aspettare un tempo adeguato, il primo concorrente raggiungerà il secondo.

1.3) Le confezioni di pasta alimentare di una certa linea di produzione hanno un peso che può essere assimilato ad una variabile aleatoria X avente media $\mu = 0.5$ kg e deviazione standard $\sigma = 0.003$ kg. Si determini:

a) il limite inferiore della probabilità che, estraendo a sorte una confezione, il peso della confezione sia compreso nell'intervallo di estremi $0.5 \pm 2 \times 0.003$;

b) il limite superiore della probabilità che X sia esterno all'intervallo $(0.491, 0.509)$;

c) il limite inferiore per $\mathbb{P}(0.495 < X < 0.505)$;

d) l'intervallo intorno alla media in cui è compresa la variabile aleatoria X con probabilità almeno uguale al 95%;

e) il limite superiore della probabilità che X sia esterno all'intervallo $(0.45, 0.56)$.

Variabili discrete

2.1) Una variabile X ha la funzione di ripartizione così definita

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ a_1 & \text{se } 0 \le x < 1 \\ a_2 & \text{se } 1 \le x < 3 \\ 1 & \text{se } 3 \le x \end{cases}$$

dove $0 < a_1 < a_2 < 1$. Si calcoli la densità discreta ed il valore atteso di X e $\sqrt{X}$. Quando i due valori attesi coincidono?

2.2) Si costruisce un gioco basato sul lancio di un dado non truccato. Sia X il risultato del dado e Y la somma vinta così definita

$$\begin{aligned} X = 1 &\Longrightarrow Y = -20 \\ X \in \{2,3\} &\Longrightarrow Y = -\alpha \\ X \in \{4,5\} &\Longrightarrow Y = 0 \\ X = 6 &\Longrightarrow Y = 40 \end{aligned}$$

o, equivalentemente, $Y = 40\mathbb{1}_{\{6\}}(X) - \alpha\mathbb{1}_{\{2,3\}}(X) - 20\mathbb{1}_{\{1\}}(X)$. Trovare $\alpha \in \mathbb{R}$ affinché il gioco sia equo e, per tale valore, determinare la densità discreta di X e la sua moda. Si consideri ora una partita fatta da due lanci del dado e sia Z la somma vinta; descrivere densità discreta, media e varianza di Z.

2.3) Si consideri la seguente funzione dipendente dal parametro reale k

$$F(x) := \begin{cases} 0 & \text{se } x \leq -1 \\ k(x+1) & \text{se } x \in (-1,0) \\ k(x+1) + 1/4 & \text{se } x \in [0,1) \\ 2k + 3/8 & \text{se } x \geq 1. \end{cases}$$

Determinare $k \in \mathbb{R}$ affinché F sia una funzione di ripartizione. Calcolare quindi media e varianza. Sia X una variabile aleatoria con funzione di ripartizione F, quanto valgono $\mathbb{P}(X \in (0,1))$ e $\mathbb{P}(X \in [0,1])$?

Soluzioni

1.1)

a) Dalla legge debole dei grandi numeri si ha che, per ogni $\epsilon > 0$,

$$\lim_{n\to+\infty} \mathbb{P}\left(\left|\sum_{i=1}^n X_i - n\mu\right| \geq n\epsilon\right) = 0$$

pertanto, essendo $\{\sum_{i=1}^n X_i - n\lambda < 0\} \subseteq \{|\sum_{i=1}^n X_i - n\mu| > (\mu-\lambda)n\}$, si ottiene

$$0 \leq \lim_{n\to+\infty} \mathbb{P}\left(\sum_{i=1}^n X_i < n\lambda\right) \leq \lim_{n\to+\infty} \mathbb{P}\left(\left|\sum_{i=1}^n X_i - n\mu\right| > (\mu-\lambda)n\right) = 0.$$

b) È del tutto analogo al caso (a).

c) Dal Teorema Centrale del Limite, se $\sigma^2 := \mathrm{Var}(X_i)$,

$$\lim_{n\to+\infty} \mathbb{P}\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq x\right) = \Phi(x).$$

Considerando che

$$\left\{\sum_{i=1}^n X_i - m\mu \leq 0\right\} = \left\{\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq 0\right\}$$

si ha l'asserto.

d) Si osservi che

$$\exp(-n\mu) \sum_{=0}^{[\lambda n]} \frac{(n\mu)^i}{i!} = \mathbb{P}(Y_n \leq n\lambda)$$

dove $Y_n \sim P(n\mu)$; ricordando che la somma di n variabili aleatorie i.i.d. con distribuzione $P(\mu)$ ha distribuzione $P(n\mu)$ si ottiene il risultato cercato.

1.2) Sia X_i (risp. Y_i) il numero di km percorsi dal primo concorrente nella (risp. dal secondo concorrente) nell'i-esima ora e sia Z_n il distacco tra i due. L'insieme di variabili aleatorie $\{X_1, Y_1, X_2, Y_2, \ldots\}$ piò essere considerato un insieme di variabili indipendenti. In tal caso è evidente che

$$Z_n = c - \sum_{i=1}^n (X_i - Y_i).$$

Dalle ipotesi si ha che $\{X_i - Y_i\}_i$ sono indipendenti ed identicamente distribuite con media $E(X_i - Y_i) = E(X_i) - E(Y_i) = 0.01$ e varianza finita (quanto vale la varianza?). Dalla legge debole dei grandi numeri si ha che

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{i=1}^n (X_i - Y_i) = 0.01 > 0$$

in probabilità. Pertanto per ogni $\lambda < 0.01$ si ha che

$$\lim_{n \to +\infty} \mathbb{P}(\sum_{i=1}^{n}(X_i - Y_i) \geq n\lambda) = 1;$$

si fissi $\lambda = 0.005$. Ora se n_0 è scelto in modo che $0.005n_0 \geq c$ e se A è l'evento "il primo concorrente raggiunge il secondo" allora si ha che $A \supseteq \cup_{n \in \mathbb{N}}\{\sum_{i=1}^{n}(X_i - Y_i) \geq 0.005n\}$ pertanto

$$\mathbb{P}(A) \geq \mathbb{P}(\cup_{n \geq n_0}\{\sum_{i=1}^{n}(X_i - Y_i) \geq c\}) \geq \mathbb{P}(\sum_{i=1}^{k}(X_i - Y_i) \geq 0.005k), \qquad \forall k \geq n_0$$

da cui l'asserto passando al limite per $k \to +\infty$.
Oss1) Il problema si svolge più rapidamente facendo uso della legge forte dei grandi numeri.

bf 1.3)

a) Si tratta di una applicazione diretta della formula:

$$\mathbb{P}\left(|X - \mu| < k\sigma\right) \geq 1 - \frac{1}{k^2}$$

dalla quale risulta evidente che $k = 2$; pertanto l'estremo inferiore cercato è dato da:

$$\mathbb{P}\left(|X - 0.5| < 0.006\right) \geq 1 - \frac{1}{2^2} = 0.75.$$

b) Per utilizzare ancora la precedente formula, dobbiamo prima ricavare k. Dalla relazione $0.5 - k(0.003) = 0.491$ otteniamo $k = 3$. Pertanto l'estremo superiore cercato è dato da:

$$\mathbb{P}\left[(X \leq 0.491) \cap (X \geq 0.509)\right] \leq \frac{1}{3^2} = 0.11.$$

c) Dalla relazione $0.5 - k(0.003) = 0.495$ si trova $k = 1.7$. Ne consegue che il limite inferiore cercato é 0.65.

d) Anche in questo caso si tratta di trovare k; si ha allora

$$1 - \frac{1}{k^2} = 0.95$$

da cui

$$k = \frac{1}{\sqrt{1 - 0.95}} = 4.47.$$

Pertanto l'intervallo richiesto è: $(\mu - 4.47\sigma, \mu + 4.47\sigma)$ ovvero $(0.487, 0.513)$.

e) Una variante della formula di Chebichev afferma che

$$\mathbb{P}(|X-\beta| \geq \alpha) \leq \frac{\sigma^2 + (\beta-\mu)^2}{\alpha^2}$$

per ogni $\beta \in \mathbb{R}$ (dove μ è la media di X e σ^2 è la varianza. In questo caso $\beta = 5.05$ e $\alpha = 0.55$ pertanto il limite superiore risulta

$$\frac{9 \cdot 10^{-6} + 2.5 \cdot 10^{-3}}{0.5^2} = 0.00829.$$

2.1) Il range di X è l'insieme finito $\{0, 1, 3\}$ e la densità discreta è

$$\rho_X(t) := \mathbb{P}(X = t) = \begin{cases} a_1 & \text{se } t = 0 \\ a_2 - a_1 & \text{se } t = 1 \\ 1 - a_2 & \text{se } t = 3 \\ 0 & \text{altrimenti.} \end{cases}$$

Analogamente il range di $\sqrt{X}$ è l'insieme finito $\{0, 1, \sqrt{3}\}$ e la sua densità discreta è

$$\rho_{\sqrt{(X)}}(t) := \mathbb{P}(X = t) = \begin{cases} a_1 & \text{se } t = 0 \\ a_2 - a_1 & \text{se } t = 1 \\ 1 - a_2 & \text{se } t = \sqrt{3} \\ 0 & \text{altrimenti.} \end{cases}$$

La media $E(X) = \sum_{i \in \mathrm{Rg}(X)} i \rho_X(i) = 3 - 2a_2 - a_1$ e $E(\sqrt{X}) = \sum_{i \in \mathrm{Rg}(\sqrt{X})} i \rho_{\sqrt{X}}(i) = \sqrt{3} - (\sqrt{3} - 1)a_2 - a_1$ da cui, facilmente, si ha $E(X) > E(\sqrt{X})$ per ogni $a_2 < 1$.

2.2) Il gioco è equo se e solo se la media $E(Y) = 0$; pertanto

$$0 = E(Y) = 40\mathbb{P}(X = 6) - \alpha\mathbb{P}(X \in \{2, 3\}) - 20\mathbb{P}(X = 1) = 10/3 - \alpha/3,$$

da cui $\alpha = 10$. A questo punto $\mathrm{Rg}(Y) = \{-20, -10, 0, 40\}$ e la sua densità discreta è

$$\rho_Y(t) := \mathbb{P}(Y = t) = \begin{cases} 1/6 & \text{se } t = -20 \\ 1/3 & \text{se } t = -10 \\ 1/3 & \text{se } t = 0 \\ 1/6 & \text{se } t = 40 \\ 0 & \text{altrimenti.} \end{cases}$$

Si vede immediatamente che la distribuzione è bimodale con mode -10 e 0.

Siano ora X_1 e X_2 le somme vinte rispettivamente nel primo e nel secondo lancio; le due variabili sono i.i.d. con distribuzione equivalente a quella di X. Poiché $Z = X_1 + X_2$ si ha che $E(Z) = E(X_1) + E(X_2) = 2E(X) = 0$ e, dall'indipendenza, $\mathrm{Var}(Z) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = 2\mathrm{Var}(X) = 2(400/6 + 100/3 + 1600/6) = 2200/3$. Ovviamente $\mathrm{Rg}(Z) = \mathrm{Rg}(X_1) + \mathrm{Rg}(X_2) := \{x + y : x \in \mathrm{Rg}(X_1), y \in \mathrm{Rg}(X_2)\}$ il che significa $\mathrm{Rg}(Z) =$

$\{-40, -30, -20, -10, 0, 20, 30, 40, 80\}$ inoltre, essendo X_1 ed X_2 puramente discrete si ha che la densità di Z

$$\begin{aligned}\mathbb{P}(Z = n) &= \sum_{i \in \mathrm{Rg}(X_1), j \in \mathrm{Rg}(X_2): i+j=n} \mathbb{P}(X_1 = i, X_2 = j) \\ &= \sum_{i \in \mathrm{Rg}(X_1), j \in \mathrm{Rg}(X_2): i+j=n} \mathbb{P}(X_1 = i)\mathbb{P}(X_2 = j),\end{aligned}$$

dove si è utilizzata anche la proprietà di indipendenza. In definitiva

$$\mathbb{P}(Z = n) = \begin{cases} 1/36 & \text{se } n = -40 \\ 1/9 & \text{se } n = -30 \\ 2/9 & \text{se } n = -20 \\ 2/9 & \text{se } n = -10 \\ 1/9 & \text{se } n = 0 \\ 1/18 & \text{se } n = 20 \\ 1/9 & \text{se } n = 30 \\ 1/9 & \text{se } n = 40 \\ 1/36 & \text{se } n = 80. \end{cases}$$

2.3) La funzione F è sempre continua da destra; inoltre è non decrescente se e solo se $k \geq 0$ e per tali valori essa è anche non negativa. Per ogni valore di k vale

$$\lim_{x \to -\infty} F(x) = 0$$

mentre

$$\lim_{x \to +\infty} F(x) = 1 \qquad \Longleftrightarrow \qquad k = 5/16.$$

Pertanto condizione necessaria e sufficiente affinché F sia una funzione di ripartizione è che $k = 3/16$. F non è continua, pertanto la variabile non ammetterà densità; F non è nemmeno costante a tratti, pertanto la variabile X non sarà puramente discreta. Per calcolare la densità mista (ammesso che esista) procediamo come segue. Sia

$$h(x) := F(x) - \lim_{t \to x^-} F(t) = \begin{cases} 1/4 & \text{se } x = 0 \\ 1/8 & \text{se } x = 1 \\ 0 & \text{altrimenti} \end{cases}$$

e si consideri

$$H(x) := \sum_{t \leq x} h(t);$$

si vede immediatamente che H è continua da destra e che

$$G(x) := F(x) - H(x) = \begin{cases} 0 & \text{se } x \leq -1 \\ 5(x+1)/16 & \text{se } x \in (-1, 1] \\ 5/8 & \text{se } x \geq 1. \end{cases}$$

è continua. Inoltre la sua derivata esiste ovunque tranne che in -1 e 1 e vale

$$G'(x) = \begin{cases} 0 & \text{se } x \in (-\infty, -1) \cup (1, +\infty) \\ 5/16 & \text{se } x \in (-1, 1), \end{cases}$$

pertanto, per il teorema fondamentale del calcolo,

$$G(a) - G(b) = \int_a^b G'(x) \mathrm{d}x$$

(il valore di $G'(-1)$ e $G'(+1)$ può essere arbitrariamente fissato). Quindi

$$\mathbb{P}(X \in A) = \int_A G'(x) \mathrm{d}x + \sum_{x \in A} h(x)$$

in particolare, per ogni $a < b$,

$$\mathbb{P}(X \in (a, b]) = F(b) - F(a) = \int_a^b h(x) \mathrm{d}x + \sum_{a < x \leq b} h(x).$$

La media si calcola come segue

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x G'(x) \mathrm{d}x + \sum x h(x) \\ &= \frac{5}{16} \int_{-1}^{1} x \mathrm{d}x + h(1) = \frac{1}{8}, \end{aligned}$$

mentre la varianza vale

$$\begin{aligned} \mathrm{Var}(X) &= \int_{-\infty}^{+\infty} x^2 G'(x) \mathrm{d}x + \sum x^2 h(x) - E(X)^2 \\ &= \frac{5}{16} \int_{-1}^{1} x^2 \mathrm{d}x + h(1) - \frac{1}{64} = \frac{1}{3} - \frac{1}{64} = \frac{61}{192}. \end{aligned}$$

Facilmente

$$\mathbb{P}(X \in (0, 1)) = G(b) - G(a) + \sum_{0 < x < 1} h(x) = 5/16,$$

mentre

$$\mathbb{P}(X \in [0, 1]) = G(b) - G(a) + \sum_{0 \leq x \leq 1} h(x) = 11/16 \; (\equiv \mathbb{P}(X \in (0, 1)) + \mathbb{P}(X = 0) + \mathbb{P}(X = 1)\,).$$