Tutorato di Probabilità 1, X
a.a. 2003/2004

Esercizio 1 **a)** Una moneta equa viene tirata 1000 volte. Qual è la probabilità di ottenere un numero di teste compreso tra 490 e 520?

b) Una moneta equa viene tirata 1000 volte. Determinare k affinché la probabilità di ottenere un numero di teste compreso tra 490 e k sia circa 0.8.

c) Una moneta equa viene tirata più di 800 volte. Quante volte dev'essere tirata perché con probabilità maggiore di 0.6 escano più di 400 teste?

d) Trovare informazioni sulla probabilità che in un lancio singolo esca testa se, su 1000 lanci,

d1) più del 50% delle volte il numero di teste uscite è minore di 200;

d2) meno del 20% delle volte il numero di teste uscite è minore di 200.

Esercizio 2 **a)** Siano $Y_1 \sim \mathrm{N}(0, \sigma_1^2)$ e $Y_2 \sim \mathrm{N}(0, \sigma_2^2)$ indipendenti. Dimostrare[1] che $Y_1 + Y_2 \sim \mathrm{N}(0, \sigma_1^2 + \sigma_2^2)$. Dedurre che se $Y_1, \dots, Y_n$ sono gaussiane indipendenti, di media nulla e varianza $\sigma_1^2, \dots, \sigma_n^2$ rispettivamente, allora $Y_1 + \cdots Y_n \sim \mathrm{N}(0, \sigma_1^2 + \cdots + \sigma_n^2)$.

b) Dimostrare che se $Z \sim \mathrm{N}(m, \sigma^2)$ allora $\mu + \rho Z \sim \mathrm{N}(\mu + \rho m, \rho^2 \sigma^2)$.

c) Dimostrare che se $X_1, \dots, X_n$ sono gaussiane indipendenti, con $X_i \sim \mathrm{N}(m_i, \sigma_i^2)$ allora $X_1 + \cdots + X_n \sim \mathrm{N}(m_1 + \cdots + m_n, \sigma_1^2 + \cdots + \sigma_n^2)$.

d) Dimostrare che se $X_1, \dots, X_n$ sono gaussiane indipendenti, con $X_i \sim \mathrm{N}(m_i, \sigma_i^2)$, allora

$$\frac{\sum_{i=1}^n (X_i - m_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \sim \mathrm{N}(0, 1).$$

Dedurre che il Teorema del Limite Centrale è **esatto** per v.a. gaussiane indipendenti: se $X_1, \dots, X_n$ sono i.i.d. gaussiane di media m e varianza σ^2 allora **per ogni** n

$$\frac{\sum_{i=1}^n (X_i - m)}{\sqrt{n\sigma^2}} \sim \mathrm{N}(0, 1).$$

Esercizio 3 In una popolazione femminile (adulta) è noto che la statura media è 164 cm e che nell'88.5% dei casi la statura è inferiore a 170 cm. Si suppone che la variabile "statura" sia normalmente distribuita.

(a) Calcolare la deviazione standard.

(b) Consideriamo un campione di 25 donne, prese a caso nella popolazione femminile. Calcolare la probabilità che la media aritmetica delle stature osservate nel campione sia maggiore di 166 cm.

Esercizio 4 Sia X una v.a. con densità di probabilità

$$p(x) = \frac{3}{x^4} \mathbb{1}_{x>1}.$$

(a) Dire per quali valori di β esiste $\mathbb{E}[X^\beta]$ e, se possibile, calcolare $\mathbb{E}[X]$ e $\mathrm{Var}(X)$.

Sia $\{X_n\}_n$ una successione di copie indipendenti di X e $\{S_n\}_n$ la successione delle somme parziali:

$$S_n = \sum_{k=1}^n X_k.$$

(b) Usando la disuguaglianza di Chebycev, stimare n affinché $\mathrm{P}\left(\left|\frac{1}{n} S_n - \frac{3}{2}\right| > 0.1\right) \le 10^{-3}$.

(c) Calcolare $\lim_{n\to\infty} \mathrm{P}\left(\frac{2 S_n - 3n}{\sqrt{n}} > 0\right)$.

[1] **Sugg.:** scrivere la densità di $Y_1 + Y_2$ usando la convoluzione delle due densità e manipolare la forma quadratica che compare all'esponente scrivendola come il quadrato di un binomio più un resto. Usare poi le proprietà della distribuzione gaussiana, in particolare il fatto che $\int_{\mathbb{R}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz = \sqrt{2\pi\sigma^2}$.

Soluzioni

1) Indichiamo con S_n il numero di teste uscite su n lanci e con F_n la sua f.d.: l'approssimazione gaussiana con correzione dà

$$F_n(x) \simeq \Phi\Big(\frac{x+0.5-n\,p}{\sqrt{n\,p(1-p)}}\Big) \tag{1}$$

dove, al solito, Φ denota la f.d. di una gaussiana standard e p la probabilità che in un singolo lancio esca testa. Ricordiamo inoltre che, per $\xi > 0$,

$$\Phi(-\xi) = 1 - \Phi(\xi).$$

a) Qui $p = \frac{1}{2}$ e $n = 1000$. Si chiede $\mathrm{P}(490 \le S_{1000} \le 510)$. Da (1),

$$\mathrm{P}(490 \le S_{1000} \le 520) = F_{1000}(520) - F_{1000}(489) \simeq \Phi\Big(\frac{520.5-500}{\sqrt{250}}\Big) - \Phi\Big(\frac{489.5-500}{\sqrt{250}}\Big)$$
$$= \Phi\Big(1.2965\Big) - \Phi\Big(-0.6641\Big) = \Phi\Big(1.2965\Big) - 1 + \Phi\Big(0.6641\Big) \simeq 0.90147 - 1 + 0.7454 = 0.64687,$$

quindi $\mathrm{P}(490 \le S_{1000} \le 520) \simeq 0.64687$.

b) Qui $p = \frac{1}{2}$ e $n = 1000$. Si chiede k affinché $\mathrm{P}(490 \le S_{1000} \le k) = 0.7$. Da (1),

$$0.7 = \mathrm{P}(490 \le S_{1000} \le k) = F_{1000}(k) - F_{1000}(489) \simeq \Phi\Big(\frac{k+0.5-500}{\sqrt{250}}\Big) - \Phi\Big(\frac{489.5-500}{\sqrt{250}}\Big).$$

Allora

$$\Phi\Big(\frac{k+0.5-500}{\sqrt{250}}\Big) \simeq 0.7 + \Phi\Big(\frac{489.5-500}{\sqrt{250}}\Big) = 0.7 + \Phi\Big(-0.6641\Big) = 0.7 + 1 - \Phi\Big(0.6641\Big) \simeq 0.9546.$$

Ora, uno sguardo alle tavole dà $\Phi^{-1}(0.9546) \simeq 1.7$, quindi

$$\frac{k+0.5-500}{\sqrt{250}} \simeq 1.7, \qquad \text{cioè} \qquad k \simeq 499.5 + 1.7 \cdot \sqrt{250} = 526.3794.$$

c) Qui $p = \frac{1}{2}$. Si chiede $n > 800$ affinché $\mathrm{P}(S_n > 400) > 0.6$. Da (1),

$$0.6 < \mathrm{P}(S_n > 400) = 1 - F_n(400) \simeq 1 - \Phi\Big(\frac{400.5-n/2}{\sqrt{n}/2}\Big).$$

Basta allora cercare n affinché

$$\Phi\Big(\frac{801-n}{\sqrt{n}}\Big) < 0.4$$

Ora, poiché $n > 800$, $\frac{801-n}{\sqrt{n}} \le 0$, quindi

$$\Phi\Big(\frac{801-n}{\sqrt{n}}\Big) = 1 - \Phi\Big(\frac{n-801}{\sqrt{n}}\Big) < 0.4$$

cioè

$$\Phi\Big(\frac{n-801}{\sqrt{n}}\Big) > 0.6.$$

Ma $\Phi^{-1}(0.6) \simeq 0.26$ e poiché Φ è crescente possiamo scrivere

$$\frac{n-801}{\sqrt{n}} > 0.26.$$

Risolvendo la disequazione, otteniamo $n > 808.5$, quindi basta tirare la moneta 809 volte.

d1) Qui $n = 1000$ e si cerca p affinché $\mathrm{P}(S_n < 200) > 0.5$. Da (1),

$$0.5 < \mathrm{P}(S_{1000} < 200) = \mathrm{P}(S_{1000} \leq 201) = F_{1000}(201) \simeq \Phi\Big(\frac{201.5 - 1000\,p}{\sqrt{1000\,p(1-p)}}\Big).$$

Ma allora $201.5 - 1000\,p > 0$, cioè $p < 0.2015$.

d2) Se invece si avesse $\mathrm{P}(S_n < 200) < 0.2$, allora

$$0.2 > \mathrm{P}(S_{1000} < 200) = \mathrm{P}(S_{1000} \leq 201) = F_{1000}(201) \simeq \Phi\Big(\frac{201.5 - 1000\,p}{\sqrt{1000\,p(1-p)}}\Big).$$

Poiché $\Phi^{-1}(0.2) = -\Phi^{-1}(1-0.2) = -\Phi^{-1}(0.8) \simeq -0.84$, basta che

$$\frac{201.5 - 1000\,p}{\sqrt{1000\,p(1-p)}} < -0.84$$

cioè $1000\,p - 201.5 > 0.84\,\sqrt{1000\,p(1-p)}$. Quindi, $1000\,p - 201.5 > 0$, cioè $p > 0.2105$ e

$$10^6(p - 0.2015)^2 > (0.84)^2 \cdot 10^3 \cdot p(1-p).$$

Risolvendo, si ottiene $p < 0.19106$ oppure $p > 0.21236$. Poiché p dev'essere maggiore di 0.2015, ne segue che $p > 0.21236$.

2) a) Calcoliamo la densità di $Y_1 + Y_2$:

$$f_{Y_1+Y_2}(z) = \int_{\mathbb{R}} f_{Y_1}(y) f_{Y_2}(z-y)\,dy = \frac{1}{2\pi\,\sigma_1\,\sigma_2}\int_{\mathbb{R}} e^{-\frac{1}{2\sigma_1^2\sigma_2^2}\big(\sigma_2^2 y^2 + \sigma_1^2(z-y)^2\big)}\,dy.$$

Lavoriamo sulla forma quadratica all'esponente:

$$\begin{aligned}\sigma_2^2 y^2 + \sigma_1^2(z-y)^2 &= y^2(\sigma_1^2+\sigma_2^2) - 2\sigma_1^2 yz + \sigma_1^2 z^2 = y^2(\sigma_1^2+\sigma_2^2) - 2\sigma_1^2 yz + \sigma_1^2 z^2 \\ &+ \frac{\sigma_1^4}{\sigma_1^2+\sigma_2^2} z^2 - \frac{\sigma_1^4}{\sigma_1^2+\sigma_2^2} z^2 = (\sigma_1^2+\sigma_2^2)\Big(y - \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2} z\Big)^2 + \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\,z^2.\end{aligned}$$

Posto $\mu_z = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2} z$, possiamo scrivere

$$f_{Y_1+Y_2}(z) = \frac{1}{2\pi\,\sigma_1\,\sigma_2}\int_{\mathbb{R}} e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}}\, e^{-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}(y-\mu_z)^2}\,dy = \frac{1}{2\pi\,\sigma_1\,\sigma_2}\, e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}}\int_{\mathbb{R}} e^{-\frac{(y-\mu_z)^2}{2\sigma_1^2\sigma_2^2/(\sigma_1^2+\sigma_2^2)}}\,dy.$$

Ora, la funzione $y \mapsto e^{-\frac{(y-\mu_z)^2}{2\sigma_1^2\sigma_2^2/(\sigma_1^2+\sigma_2^2)}}$ dà, a meno della costante, una densità gaussiana, quindi il suo integrale su tutto $\mathbb{R}$ è pari alla costante stessa:

$$\int_{\mathbb{R}} e^{-\frac{(y-\mu_z)^2}{2\sigma_1^2\sigma_2^2/(\sigma_1^2+\sigma_2^2)}}\,dy = \sqrt{2\pi\,\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}},$$

quindi

$$f_{Y_1+Y_2}(z) = \frac{1}{2\pi\,\sigma_1\,\sigma_2}\, e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}} \cdot \sqrt{2\pi\,\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\, e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}}$$

che è una densità gaussiana, di media 0 e varianza $\sigma_1^2 + \sigma_2^2$, quindi $Y_1 + Y_2 \sim \mathrm{N}(0, \sigma_1^2 + \sigma_2^2)$.

Ora, se $Y_3 \sim \mathrm{N}(0, \sigma_3^2)$ è indipendente da Y_1 e Y_2 allora è indipendente dalla v.a. gaussiana $Y_1 + Y_2$, quindi, per quanto visto sopra, $Y_1 + Y_2 + Y_3 \sim \mathrm{N}(0, \sigma_1^2 + \sigma_2^2 + \sigma_3^2)$. Iterando questo procedimento, segue immediatamente se $Y_1, \dots, Y_n$ sono gaussiane indipendenti, di media nulla e varianza $\sigma_1^2, \dots, \sigma_n^2$ rispettivamente, allora $Y_1 + \cdots Y_n \sim \mathrm{N}(0, \sigma_1^2 + \cdots + \sigma_n^2)$

b) Indichiamo con W la v.a. $\mu + \rho Z$, con $Z \sim \mathrm{N}(m, \sigma^2)$ e $\rho \neq 0$, e calcoliamo la f.d. di W:

$$F_W(w) = \mathrm{P}(\rho Z \le w - m) = \begin{cases} \mathrm{P}\Big(Z \le \frac{w-\mu}{\rho}\Big) = F_Z\Big(\frac{w-\mu}{\rho}\Big) & \text{se } \rho > 0 \\ \mathrm{P}\Big(Z \ge \frac{w-\mu}{\rho}\Big) = 1 - F_Z\Big(\frac{w-\mu}{\rho}\Big) & \text{se } \rho < 0 \end{cases}$$

Derivando,

$$f_W(w) = \begin{cases} f_Z\Big(\frac{w-\mu}{\rho}\Big) \cdot \frac{1}{\rho} & \text{se } \rho > 0 \\ -f_Z\Big(\frac{w-\mu}{\rho}\Big) \cdot \frac{1}{\rho} & \text{se } \rho < 0, \end{cases} \qquad \text{quindi} \quad f_W(w) = \frac{1}{\sqrt{2\pi\, \rho^2 \sigma^2}}\, e^{-\frac{(w-\mu-\rho\, m)^2}{2\rho^2\sigma^2}}$$

cioè $W = \mu + \rho Z \sim \mathrm{N}(\mu + \rho m, \rho^2 \sigma^2)$.

c) Poniamo $Y_i = X_i - m_i$: le Y_i sono indipendenti e, per **b)**, gaussiane di media 0 e varianza σ_i^2. Allora, per **a)**, $Y = Y_1 + \cdots + Y_n \sim \mathrm{N}(0, \sigma^2)$, con $\sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2$. Ora, $X_1 + \cdots + X_n = Y + m$, con $m = m_1 + \cdots + m_n$, e, da **b)**, $Y + m \sim \mathrm{N}(m, \sigma^2)$. Ricapitolando, $X_1 + \cdots + X_n \sim \mathrm{N}(m_1 + \cdots + m_n, \sigma_1^2 + \cdots + \sigma_n^2)$.

d) Si noti che

$$\frac{\sum_{i=1}^n (X_i - m_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}} = \frac{1}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \sum_{i=1}^n X_i - \frac{\sum_{i=1}^n m_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}}.$$

Ora, abbiamo visto che $\sum_{i=1}^n X_i \sim \mathrm{N}(\sum_{i=1}^n m_i, \sum_{i=1}^n \sigma_i^2)$ quindi, usando **b)** (con $\mu = -\frac{\sum_{i=1}^n m_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}}$ e $\rho = \frac{1}{\sqrt{\sum_{i=1}^n \sigma_i^2}}$) otteniamo che $\frac{\sum_{i=1}^n (X_i - m_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}}$ è gaussiana, di media

$$-\frac{\sum_{i=1}^n m_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} + \frac{1}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \sum_{i=1}^n m_i = 0$$

e varianza

$$\Big(\frac{1}{\sqrt{\sum_{i=1}^n \sigma_i^2}}\Big)^2 \sum_{i=1}^n \sigma_i^2 = 1.$$

Se $m_1 = \cdots = m_n$ e $\sigma_1^2 = \cdots = \sigma_n^2 = \sigma^2$ allora ovviamente

$$\frac{\sum_{i=1}^n (X_i - m_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}} = \frac{\sum_{i=1}^n (X_i - m)}{\sqrt{n\, \sigma^2}} \sim \mathrm{N}(0,1)$$

per ogni n. Quindi il TLC vale **non solo** asintoticamente (per n grande) ma è vero per ogni n.

3) Indichiamo con X la variabile (aleatoria) "statura delle donne". Si ipotizza che $X \sim \mathrm{N}(164, \sigma^2)$, con σ^2 a priori non noto. È però noto che

$$\mathrm{P}(X \le 170) = \frac{88.5}{100} = 0.885$$

a) Si ha, ricordando che $\frac{X-164}{\sigma} \sim \mathrm{N}(0,1)$,

$$0.885 = \mathrm{P}(X \le 170) = \mathrm{P}\Big(\frac{X - 164}{\sigma} \le \frac{170 - 164}{\sigma}\Big) = \Phi\Big(\frac{6}{\sigma}\Big).$$

Ora, $0.885 \simeq \Phi^{-1}(1.2)$, quindi

$$\frac{6}{\sigma} = 1.2, \qquad \text{cioè} \qquad \sigma \simeq 5.$$

Ciò significa che mediamente la statura delle donne è 164 cm, con un errore in gran parte compreso tra $+5$ e -5 cm, ovvero per lo più la statura è compresa tra 159 cm e 169 cm:

$$\mathrm{P}(159 \leq X \leq 169) = \mathrm{P}\Big(-1 \leq \frac{X-164}{5} \leq 1\Big) = 2\Phi(1) - 1 = 0.6826.$$

b) Siano $X_1, \ldots, X_{25}$ le stature osservate nel campione, che possiamo supporre indipendenti. La media aritmetica è $\bar{X}_{25} = \frac{1}{25}\sum_{i=1}^{25} X_i$. Ora, poiché $X_i \sim \mathrm{N}(164, 5)$ per ogni i, sappiamo che $\sum_{i=1}^{25} X_i \sim \mathrm{N}(25 \cdot 164, 25 \cdot 5)$, quindi $\bar{X}_{25} = \frac{1}{25}\sum_{i=1}^{25} X_i \sim \mathrm{N}(\frac{1}{25} \cdot 25 \cdot 164, \frac{1}{25^2} \cdot 25 \cdot 5) = \mathrm{N}(164, 1/5)$, quindi $\frac{\bar{X}_{25}-164}{\sqrt{1/5}} \sim \mathrm{N}(0,1)$. Allora,

$$\mathrm{P}(\bar{X}_{25} \leq 166) = \mathrm{P}\Big(\frac{\bar{X}_{25}-164}{\sqrt{1/5}} \leq \frac{166-164}{\sqrt{1/5}}\Big) = \Phi\Big(2\sqrt{5}\Big) = \Phi\Big(4.4721\Big) \simeq 1.$$

4) a) Perché esista $\mathbb{E}(X^\beta)$ deve esistere $\int_{\mathbb{R}} |x|^\beta p(x)\, dx$. Ora, $p(x) \equiv 0$ per $x \leq 1$, basta quindi lavorare con $\int_1^{+\infty} x^\beta p(x)\, dx$. Il problema, se c'è un problema, qui è ovviamente all'infinito: studiamo il limite, per $\alpha \to +\infty$, di $\int_1^\alpha x^\beta p(x)\, dx$:

$$\int_1^\alpha x^\beta p(x)\, dx = 3\int_1^\alpha x^{\beta-4}\, dx = \begin{cases} 3\,\frac{x^{\beta-3}}{\beta-3}\,\big|_1^\alpha = \frac{3}{3-\beta}(1-\alpha^{\beta-3}) & \text{se } \beta - 4 \neq -1 \\ 3\ln x\,\big|_1^\alpha = 3\ln\alpha & \text{se } \beta-4=-1 \end{cases}$$

Quindi, esiste finito il limite per $\alpha \to +\infty$ di $\int_1^\alpha x^\beta p(x)\, dx$ se e solo se $\beta - 3 < 0$, cioè $\beta < 3$. In particolare quindi l'integrale $\int_{\mathbb{R}} |x|^\beta p(x)\, dx$ esiste per $\beta = 1, 2$, dunque esistono sia media che varianza di X e

$$\mathbb{E}(X) = \int_1^{+\infty} x\, p(x)\, dx = \frac{3}{2}, \qquad \mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \int_1^{+\infty} x^2\, p(x)\, dx - \frac{9}{4} = 3 - \frac{9}{4} = \frac{3}{4}.$$

b) Poiché $\frac{3}{2} = \mathbb{E}(X_i)$ per ogni i, allora $\frac{3}{2} = \mathbb{E}(\frac{1}{n}S_n)$, quindi la disuguaglianza di Chebycev si può effettivamente usare e

$$\mathrm{P}\Big(\Big|\frac{1}{n}S_n - \frac{3}{2}\Big| > 0.1\Big) \leq \frac{\mathrm{Var}\Big(\frac{1}{n}\,S_n\Big)}{0.1^2} = 100\,\frac{1}{n^2}\cdot n\,\mathrm{Var}(X_i) = \frac{300}{4\,n}.$$

Quindi, perché $\mathrm{P}\Big(\Big|\frac{1}{n}S_n - \frac{3}{2}\Big| > 0.1\Big) \leq 10^{-3}$ basta che $\frac{300}{4\,n} \leq 10^{-3}$, cioè $n \geq 13334$.

c) Osserviamo che

$$\frac{2\,S_n - 3\,n}{\sqrt{n}} = 2\,\frac{S_n - \frac{3}{2}n}{\sqrt{n}} = 2\sqrt{\frac{3}{4}}\,\frac{S_n - \frac{3}{2}n}{\sqrt{3n/4}}$$

e, dal TLC, la legge di $\frac{S_n - \frac{3}{2}n}{\sqrt{3n/4}}$ converge, per $n \to +\infty$ ad una gaussiana standard. Allora,

$$\lim_{n\to\infty} \mathrm{P}\Big(\frac{2\,S_n - 3\,n}{\sqrt{n}} > 0\Big) = \lim_{n\to\infty} \mathrm{P}\Big(\sqrt{3}\,\frac{S_n - \frac{3}{2}n}{\sqrt{3n/4}} > 0\Big) = \lim_{n\to\infty} \mathrm{P}\Big(\frac{S_n - \frac{3}{2}n}{\sqrt{3n/4}} > 0\Big) = 1 - \Phi(0) = 1 - \frac{1}{2} = 0.5.$$