Compitino di Probabilità e Statistica del 19 marzo 2005 (Corso di Laurea Triennale in Matematica, Università degli Studi di Padova).

Cognome Nome Matricola

	Es. 1	Es. 2	Es. 3	Es. 4	Somma	Voto parziale
Prima parte				XXX XXX XXX		+
Seconda parte	XXX XXX XXX					=
	VOTO FINALE:					

Il compitino è in 2 parti:

- **la prima parte** è costituita dall'esercizio 1, dal punto 1 dell'esercizio 2 e dal punto 1 dell'esercizio 3, e sostituisce il primo compitino;
- **la seconda parte** è costituita dai punti 2 e 3 dell'esercizio 2, dal punto 2 dell'esercizio 3 e dall'esercizio 4.

Tutti devono fare la seconda parte, mentre deve fare la prima parte solo chi vuole sostituire i voti del compitino (chi vuole tenere il voto del compitino per la prima parte scrive "vedi compitini" al posto delle soluzioni).

Attenzione: si consegnano SOLO i fogli di questo fascicolo.

Esercizio 1. **(1^a parte)** Un aereo è scomparso e si presume che sia finito, con uguale probabilità, in una zona fra 3 possibili. Sia $1 - \beta_i$ la probabilita' di rintracciare l'aereo nella zona i, se l'aereo si trova effetivamente nella zona i, $i = 1, 2, 3$ (le costanti β_i sono le probabilità di non recuperare l'aereo nella zona i; esse dipendono generalmente da condizioni geografiche ed ambientali).

a) Qual è la probabilità condizionata che l'aereo si trovi nella zona i, $i = 1, 2, 3$, sapendo che le ricerche nella zona 1 hanno dato esito negativo?

b) Rispondere alla stessa domanda sapendo che anche le ricerche nella zona 2 hanno dato esito negativo.

Esercizio 2. Supponiamo che il numero annuale di vittime del traffico ad un dato incrocio segua una distribuzione di Poisson di parametro $\lambda = 6$.

a) **(1^a parte)** Qual è la probabilità di aver osservato esattamente 6 morti nell'anno 2003?

b) **(1^a parte)** Qual è la probabilità di aver osservato esattamente 13 morti nel biennio 2003–2004?

c) **(2^a parte)** Supponiamo che l'incrocio sia stato rifatto per migliorarne la sicurezza, e nel biennio 2005–2006 si osservino 7 vittime. Questo è un miglioramento significativo? Per rispondere alla domanda, calcolare la probabilità di avere un numero di vittime minore o uguale a 6 con la "vecchia" distribuzione.

Suggerimento: ricordarsi che una variabile aleatoria di Poisson si può approssimare con una normale. Infatti una opportuna binomiale si può approssimare sia con una Poisson che con una normale; allora ...

Esercizio 3. Consideriamo una variabile aleatoria reale X di varianza σ^2, e chiamiamo $m_k := E[X^k]$ il momento di ordine k, $k \in \mathbb{N}$. Definiamo **skewness** di X la quantità

$$\mathrm{skw}(X) := \frac{E[(X - m_1)^3]}{\sigma^3}$$

Mostrare che

a) **(2^a parte)** $\mathrm{skw}(X) = \frac{m_3 - 3m_1m_2 + 2m_1^2}{\sigma^3}$;

b) **(2^a parte)** Se $S_n := \sum_{i=1}^n X_i$, con $(X_i)_i$ i.i.d., allora $\mathrm{skw}(S_n) = \frac{\mathrm{skw}(X_1)}{\sqrt{n}}$;

c) **(1^a parte)** Se $X \sim B(n, p)$, allora $\mathrm{skw}(X) = \frac{1-2p}{\sqrt{np(1-p)}}$;

d) **(1^a parte)** Se $X \sim Po(\lambda)$, allora $\mathrm{skw}(X) = \frac{1}{\sqrt{\lambda}}$.

Suggerimento: ricordare che se γ_X è la funzione generatrice dei momenti di X, allora $E[X^k] = \gamma_X^{(k)}(0)$.

Esercizio 4. **(2^a parte)** Supponiamo di voler reclutare dei soggetti per uno studio sull'ipertensione e di sapere che il 10% della popolazione è ipertensivo. Per lo studio occorrono 100 ipertensivi.

a) Se si reclutano 1000 soggetti, quale sarà la probabilità di avere più di 100 ipertensivi nel nostro campione?

b) Quante persone bisogna reclutare per avere almeno il 90% di probabilità di avere almeno 100 ipertensivi nel nostro campione?

Soluzioni

Esercizio 1 Definiamo gli eventi

$$\begin{aligned} A_i &:= \{\text{l'aereo è nella zona } i\}, \quad i = 1,2,3, \\ B_i &:= \{\text{l'aereo viene ritrovato nella zona } i\}, \quad i = 1,2,3. \end{aligned}$$

Le probabilità fornite nel testo sono quindi $P(A_i) = \frac{1}{3}$ e $P(B_i|A_i) = 1-\beta_i$per ogni $i = 1,2,3$. Chiaramente $P(B_i|A_j) = 0$ per ogni $i \neq j$, poichè $A_j \subseteq B_i^c$ per ogni $i \neq j$.

a) Per la formula di Bayes abbiamo

$$P(A_i|B_1^c) = \frac{P(B_1^c|A_i)P(A_i)}{P(B_1^c)}$$

Per la formula della probabilità totale abbiamo poi che

$$P(B_1^c) = \sum_{i=1}^{3} P(B_1^c|A_i)P(A_i) = \beta_1 \cdot \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = \frac{1}{3}(2+\beta_1)$$

quindi

$$\begin{aligned} P(A_1|B_1^c) &= \frac{\frac{1}{3}\beta_1}{\frac{1}{3}(2+\beta_1)} = \frac{\beta_1}{2+\beta_1} \\ P(A_2|B_1^c) = P(A_3|B_1^c) &= \frac{\frac{1}{3}}{\frac{1}{3}(2+\beta_1)} = \frac{1}{2+\beta_1} \end{aligned}$$

b) Come prima abbiamo

$$P(A_i|B_1^c \cap B_2^c) = \frac{P(B_1^c \cap B_2^c|A_i)P(A_i)}{P(B_1^c \cap B_2^c)}$$

Per la formula della probabilità totale abbiamo poi che

$$P(B_1^c \cap B_2^c) = \sum_{i=1}^{3} P(B_1^c \cap B_2^c|A_i)P(A_i) = \beta_1 \cdot \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = \frac{1}{3}(2+\beta_1)$$

poichè

$$P(B_1^c \cap B_2^c|A_1) = \frac{P(B_1^c \cap B_2^c \cap A_1)}{P(A_1)} = \frac{P(B_1^c \cap A_1)}{P(A_1)} = P(B_1^c|A_1) = \beta_1$$

e analogamente $P(B_1^c \cap B_2^c|A_2) = \beta_2$. Infine

$$\begin{aligned} P(A_1|B_1^c \cap B_2^c) &= \frac{\frac{1}{3}\beta_1}{\frac{1}{3}(1+\beta_1+\beta_2)} = \frac{\beta_1}{1+\beta_1+\beta_2} \\ P(A_2|B_1^c \cap B_2^c) &= \frac{\frac{1}{3}\beta_2}{\frac{1}{3}(1+\beta_1+\beta_2)} = \frac{\beta_2}{1+\beta_1+\beta_2} \\ P(A_3|B_1^c \cap B_2^c) &= \frac{\frac{1}{3}}{\frac{1}{3}(1+\beta_1+\beta_2)} = \frac{1}{1+\beta_1+\beta_2} \end{aligned}$$

Esercizio 2 Innanzitutto definiamo le variabili aleatorie

$$X_i = \text{numero di morti all'incrocio per l'anno } i$$

allora possiamo considerare le $(X_i)_i$ i.i.d. con legge $Po(6)$.

a) Calcoliamo

$$P\{X_{2003} = 6\} = e^{-6}\frac{6^6}{6!} = 0.1606$$

b) Dato che le $(X_i)_i$ sono indipendenti, abbiamo che $X_{2003} + X_{2004} \sim Po(12)$. Allora

$$P\{X_{2003} + X_{2004} = 13\} = e^{-12}\frac{12^{13}}{13!} = 0.1056$$

c) Abbiamo visto che, se $\lambda > 5$, la legge di Poisson $Po(\lambda)$ è approssimabile con una legge $N(\lambda, \lambda)$. Se definiamo $X := X_{2003} + X_{2004}$ e $Z := \frac{X-12}{\sqrt{12}}$, allora

$$\begin{aligned} P\{X \leq 7\} &= P\{X < 7.5\} = P\left\{Z < \frac{7.5 - 12}{\sqrt{12}}\right\} \simeq \\ &\simeq \Phi\left(-\frac{4.5}{\sqrt{12}}\right) = \Phi(-1.29) = 1 - \Phi(1.29) = 1 - 0.90147 = 0.09853 \end{aligned}$$

dove Φ è la funzione di distribuzione di una legge normale standard, e abbiamo usato la correzione di continuità.

La conclusione è che, se abbiamo ancora $X_{2003} + X_{2004} \sim Po(12)$, è abbastanza improbabile che ci siano stati solo 7 morti: sembra quindi che ci sia stato un miglioramento significativo nella sicurezza dell'incrocio.

Esercizio 3

a) Per la linearità del valor medio si ha

$$\begin{aligned} \text{skw}(X) &= \frac{E[X^3 - 3X^2m_1 + 3X^2m_1^2 - m_1^3]}{\sigma^3} = \\ &= \frac{m_3 - 3m_1m_2 + 3m_1^3 - m_1^3}{\sigma^3} = \frac{m_3 - 3m_1m_2 + 2m_1^2}{\sigma^3} \end{aligned}$$

b) Innanzitutto continuiamo a chiamare $m_k = E[X_i^k]$. Allora $E[S_n] = nm_1$ e Var $[S_n] = n$Var $[X_i] = n\sigma^2$. Per calcolare $E[(S_n - nm_1)^3] = E[(\sum_{i=1}^n (X_i - m_1))^3]$ si può procedere in due modi: uno è fare il conto direttamente sviluppando il cubo: dopo vari passaggi algebrici si arriva a

$$\text{skw}(S_n) = \frac{E[(S_n - nm_1)^3]}{n^{3/2}\sigma^3} = \frac{\sum_{i=1}^n E[(X_i - m_1)^3]}{n^{3/2}\sigma^3} = \frac{n}{n^{3/2}}\text{skw}(X_1) = \frac{\text{skw}(X_1)}{\sqrt{n}}$$

Per questi passaggi algebrici si usa l'indipendenza delle $(X_i)_i$ e il fatto che $E[(X_i - m_1)(X_j - m_1)^2] = 0$ e $E[(X_i - m_1)(X_j - m_1)(X_k - m_1)] = 0$ per ogni $i \neq j \neq k$. Un altro modo è sfruttare il suggerimento sulla funzione $\gamma_{S_n - nm_1}$. Infatti, chiamando per brevità $\gamma := \gamma_{X_i - m_1}$, si ha

$$\begin{aligned}
\gamma_{S_n-nm_1}(t) &= \gamma^n(t)\\
\gamma'_{S_n-nm_1}(t) &= n\gamma'(t)\gamma^{n-1}(t)\\
\gamma''_{S_n-nm_1}(t) &= n\gamma''(t)\gamma^{n-1}(t) + n(n-1)(\gamma')^2(t)\gamma^{n-2}(t)\\
\gamma'''_{S_n-nm_1}(t) &= n\gamma'''(t)\gamma^{n-1}(t) + 3n(n-1)\gamma'(t)\gamma''(t)\gamma^{n-2}(t) +\\
&\quad + n(n-1)(n-2)(\gamma')^3(t)\gamma^{n-2}(t)
\end{aligned}$$

e quindi, poichè $\gamma(0) = 1$, $\gamma'(0) = 0$, si ha che

$$E[(S_n - nm_1)^3] = \gamma'''_{S_n-nm_1}(0) = n\gamma'''(0) = nE[(X_i - m_1)^3]$$

da cui segue facilmente la tesi.

c) Se $X \sim Be(p)$, allora

$$\begin{aligned}
E[(X - m_1)^3] &= E[(X-p)^3] = p(1-p)^3 + (1-p)(-p)^3 =\\
&= p(1-p)(1-2p+p^2-p^2) = p(1-p)(1-2p)
\end{aligned}$$

Se $X \sim B(n,p)$, allora ha la stessa legge di $S_n := \sum_{i=1}^n X_i$, con $X_i \sim Be(p)$ indipendenti. Allora per il punto a)

$$\text{skw}(X) = \text{skw}(S_n) = \frac{\text{skw}(X_1)}{\sqrt{n}} = \frac{p(1-p)(1-2p)}{\sqrt{n}p(1-p)\sqrt{p(1-p)}} = \frac{1-2p}{\sqrt{np(1-p)}}$$

d) Se $X \sim Po(\lambda)$, allora abbiamo

$$\begin{aligned}
\gamma_X(t) &= \exp(\lambda(e^t - 1)),\\
\gamma'_X(t) &= \lambda\exp(t + \lambda(e^t-1)),\\
\gamma''_X(t) &= \lambda(1+\lambda e^t)\exp(t+\lambda(e^t-1)),\\
\gamma'''_X(t) &= \lambda(1+\lambda e^t)^2\exp(t+\lambda(e^t-1)) + \lambda^2\exp(2t+\lambda(e^t-1)).
\end{aligned}$$

Allora si ottiene facilmente che $E[X] = \lambda$, $E[X^2] = \lambda + \lambda^2$, $E[X^3] = \lambda + 3\lambda^2 + \lambda^3$, e infine

$$\text{skw}(X) = \frac{\lambda + 3\lambda^2 + \lambda^3 - 3\lambda^2 - 3\lambda^3 + 2\lambda^3}{\lambda\sqrt{\lambda}} = \frac{1}{\sqrt{\lambda}}$$

Esercizio 4 Per ogni persona reclutata, definiamo la variabile aleatoria

$$X_i = \begin{cases} 1 & \text{se l'}i\text{-esima persona è ipertesa}\\ 0 & \text{altrimenti}\end{cases}$$

Allora $X_i \sim Be(p)$, con $p = 0.1$, e le $(X_i)_i$ sono i.i.d. Definiamo poi $S_n = \sum_{i=1}^n X_i$ = n. di persone ipertese nel nostro campione; allora $S_n \sim B(n,p)$.

a) Dobbiamo calcolare

$$\begin{aligned}
\mathbb{P}\{S_{1000} > 100\} &= \mathbb{P}\{S_{1000} > 100.5\} = \mathbb{P}\left\{\frac{S_{1000} - 0.1 \cdot 1000}{\sqrt{1000 \cdot 0.1 \cdot 0.9}} > \frac{100.5 - 100}{\sqrt{90}}\right\} \simeq \\
&\simeq P\left\{Z > \frac{0.5}{\sqrt{90}}\right\} = 1 - P\left\{Z \leq \frac{0.5}{\sqrt{90}}\right\} = \\
&= 1 - F_Z(0.05) = 1 - 0.51994 = 0.48006
\end{aligned}$$

dove abbiamo usato la correzione di continuità e l'approssimazione normale. È stato possibile utilizzare quest'ultima poichè $np = 100 > 5$ e $n(1-p) = 900 > 5$.

b) Dobbiamo cercare gli n tali che $P\{S_n \geq 100\} \geq 0.9$. Abbiamo che

$$\begin{aligned}
\mathbb{P}\{S_n \geq 100\} &= \mathbb{P}\{S_n > 99.5\} = \mathbb{P}\left\{\frac{S_n - 0.1n}{\sqrt{0.09n}} > \frac{99.5 - 0.1n}{0.3\sqrt{n}}\right\} \simeq \\
&\simeq P\left\{Z > \frac{99.5 - 0.1n}{0.3\sqrt{n}}\right\} = 1 - P\left\{Z \leq \frac{99.5 - 0.1n}{0.3\sqrt{n}}\right\} = \\
&= 1 - F_Z\left(\frac{99.5 - 0.1n}{0.3\sqrt{n}}\right) \geq 0.9
\end{aligned}$$

Vogliamo ora utilizzare l'approssimazione normale. Dato che non sappiamo a priori quanto vale n, la utilizzeremo supponendo che n sia abbastanza grande, salvo poi verificare che lo sia. Abbiamo quindi che

$$F_Z\left(\frac{99.5 - 0.1n}{0.3\sqrt{n}}\right) \leq 0.1$$

e quindi

$$\frac{99.5 - 0.1n}{0.3\sqrt{n}} \leq q_{0.1} = -q_{0.9} = -1.28$$

dove $Z \sim N(0,1)$ e abbiamo indicato con q_α il suo quantile di livello α, ed abbiamo applicato la correzione di continuità. Dobbiamo quindi risolvere la disequazione

$$0.1n - 0.384\sqrt{n} - 99.5 \geq 0$$

La sua soluzione è l'intervallo

$$\sqrt{n} \geq \frac{0.384 + \sqrt{0.384^2 + 4 \cdot 0.1 \cdot 99.5}}{2 \cdot 0.1} = 33.52$$

e si ottiene infine

$$n = 33.52^2 = 1123.72$$

Poichè n deve essere intero, la soluzione si ha per $n \geq 1124$. Per questo numero, si ha che $np > 5$ ed $n(1-p) > 5$, ed è stato quindi lecito utilizzare l'approssimazione normale.

Compitino di Probabilità e Statistica del 19 marzo 2005 (Corso di Laurea Triennale in Matematica, Universitá degli Studi di Padova) (docente: Tiziano Vargiolu)

Hanno superato la prova:

	17/2	19/3	Voto finale
Alessio Davide	—	22	22
Barichello Marica	9	13,5	22,5
Beccaro Giulia	—	21	21
Bettanini Jacopo	—	19	19
Bettin Sandro	13,5	13	26,5
Birolo Giovanni	—	24,5	24,5
Bortoletto Giulio	—	29	29
Braidotti Angela	—	17	17
Cammarata Agata	—	22	22
Campalani Laura	—	25,5	25,5
Cavaliere Elena	—	17,5	17,5
Church Jon Matteo	—	19	19
Corsi Lara	10	13,5	23,5
Deolmi Giulia	—	25	25
De Rossi Giulia	—	21	21
Dimasi Gianpaolo	—	34,5	34,5
Gabelli Lucia	—	25	25
Galesso Giorgia	—	21,5	21,5
Garofalo Matteo	—	22	22
Garonzi Martino	—	18	18
Gonzato Francesco	9	15	24
Maino Paola	13,5	15,5	29
Maranzan Giorgia	—	22	22
Morra Stefano	—	32	32
Pasetto Damiano	—	20	20
Patron Elena	—	22	22
Pinton Stefano	—	17	17
Rebellato Marta	—	23	23
Rossignolo Eleonora	—	18,5	18,5
Santella Gabriele	—	17	17
Trevisan Marco	—	24,5	24,5
Valente Giulia Erica	—	25,5	25,5
Zannol Manuel	—	17,5	17,5
Zumbo Alessandra	—	27,5	27,5

Visione compiti corretti e orale: mercoledì 23 febbraio ore 10.00 aula P20