Esame di Probabilità e Statistica del 30 marzo 2009
(Corso di Laurea Triennale in Matematica, Università degli Studi di Padova).

Cognome Nome Matricola

Es. 1	Es. 2	Es. 3	Es. 4	Somma	Voto finale

Attenzione: si consegnano SOLO i fogli di questo fascicolo.

Esercizio 1. Il cosiddetto "test del DNA" non fa altro che misurare la lunghezza di K geni, senza controllare le basi azotate che li compongono. Per ognuno di tali geni, la probabilità che due dati individui presentino una lunghezza uguale viene assunta come pari a $1/10$. Un'altra ipotesi che viene comunemente fatta è che le lunghezze di geni diversi siano indipendenti l'una dall'altra.
Supponiamo di misurare la lunghezza di $K = 10$ geni da un campione di DNA trovato su una "scena del crimine".

1. Calcolare la probabilità che un dato individuo abbia la lunghezza dei suoi geni uguali a quella del campione.
2. Supponendo di avere un database di $n = 101905$ individui, calcolare la probabilità che almeno uno di questi individui abbia le lunghezze dei suoi geni uguali a quelle del campione incriminato.
3. Calcolare la probabilità che almeno due di questi individui abbiano le lunghezze dei loro geni uguali a quelle del campione incriminato.
4. Supponendo di aver trovato un individuo con le lunghezze dei geni uguali a quelle del campione incriminato, calcolare la probabilità che ce ne sia almeno un altro.

Esercizio 3.

1. Sia $X \sim U(a, b)$. Dimostrare che, per ogni $c > 0$, la variabile aleatoria cX ha legge $U(ca, cb)$.
2. Sia $X \sim Exp(\lambda)$. Dimostrare che, per ogni $c > 0$, la variabile aleatoria cX ha legge $Exp(\frac{\lambda}{c})$.
3. Sia $X \sim \Gamma(a, \lambda)$, e definiamo $Y := cX$, con $c > 0$. Esprimere F_{cX} in funzione di F_X.
4. Calcolare la densità di cX, e concludere che $cX \sim \Gamma(a, \frac{\lambda}{c})$.

Esercizio 4. Una clinica ha la stessa probabilità di avere 2, 3 o 4 dottori di turno in un dato giorno. Indipendentemente dal numero di dottori presenti, il numero di pazienti visitati da ognuno di questi dottori è una variabile aleatoria di Poisson di media 30, e queste variabili aleatorie sono indipendenti fra loro. Si denoti con X il numero di pazienti visitati nella clinica in un dato giorno.

1. Si trovi $\mathbb{E}[X]$.
2. Si trovi $\mathrm{Var}\,[X]$.
3. Supponendo di poter approssimare una legge $Po(\lambda)$ con una legge $N(\lambda, \lambda)$, si calcoli $\mathbb{P}\{X > 80\}$.
4. Dimostrare che, se $X \sim Po(\lambda)$, allora si può approssimare la legge di X con $N(\lambda, \lambda)$ (suggerimento: X ha la stessa distribuzione di $Y = \sum_{i=1}^{n} Y_i^n$, con $(Y_i^n)_i$ i.i.d. $\sim Po(\lambda/n)$).

Valori utili della funzione di ripartizione di una normale standard:

$$\Phi(0.36) = 0.64058, \qquad \Phi(1.00) = 0.84134, \qquad \Phi(2.64) = 0.99585, \qquad \Phi(3.60) = 0.99984$$

Soluzioni

Esercizio 1.

1. Definiamo gli eventi

$$\begin{aligned} A_j &:= \{ \text{ la lunghezza del } j\text{-esimo gene dell'individuo è uguale al campione}\}, \quad j = 1, \ldots, K, \\ B &:= \{ \text{ le lunghezze di tutti i geni dell'individuo sono uguali al campione}\} \end{aligned}$$

Allora l'evento B ha probabilità pari a

$$\mathbb{P}(B) = \mathbb{P}\left(\bigcap_{j=1}^{K} A_j\right) = \prod_{j=1}^{K} \mathbb{P}(A_j) = \frac{1}{10^K} = 10^{-7}$$

dove abbiamo usato l'indipendenza delle lunghezze dei geni, e quindi degli eventi $(A_j)_j$.

2. Definiamo, per ogni individuo $i = 1, \ldots, n$, l'evento $B_i :=$ { le lunghezze di tutti i geni dell'i-esimo individuo sono uguali a quelle del campione }. Allora $\mathbb{P}(B_i) = 10^{-7}$ dal punto 1. Definiamo poi le variabili aleatorie di Bernoulli $X_i := \mathbf{1}_{B_i}$, che si possono supporre indipendenti tra di loro, e la variabile aleatoria $S_n := \sum_{i=1}^{n} X_i$, che per l'indipendenza delle X_i ha legge $B(n, p)$, con $p := 10^{-7}$. Poichè nel nostro caso $n = 101905 > 100$ e $p = 10^{-7} < 0.01$, possiamo utilizzare l'approssimazione di Poisson, e quindi $S_n \approx Po(\lambda)$, con $\lambda = np = 0.00001019 = 1.019 \cdot 10^{-5}$. Allora bisogna calcolare

$$\mathbb{P}\{S_n \geq 1\} = 1 - \mathbb{P}\{S_n < 1\} = 1 - \mathbb{P}\{S_n = 0\} \simeq 1 - e^{-\lambda} \simeq 1.019 \cdot 10^{-5} (\simeq \lambda)$$

3. Bisogna calcolare

$$\mathbb{P}\{S_n \geq 2\} = 1 - \mathbb{P}\{S_n = 0\} - \mathbb{P}\{S_n = 1\} \simeq 1 - e^{-\lambda} - \lambda e^{-\lambda} = 5.192 \cdot 10^{-11} (\simeq \lambda^2/2)$$

4. Bisogna calcolare

$$\mathbb{P}\{S_n \geq 2 \mid S_n \geq 1\} = \frac{\mathbb{P}(\{S_n \geq 2\} \cap \{S_n \geq 1\})}{\mathbb{P}\{S_n \geq 1\}} = \frac{\mathbb{P}\{S_n \geq 2\}}{\mathbb{P}\{S_n \geq 1\}} \simeq 5.09 \cdot 10^{-6} (\simeq \lambda/2)$$

Esercizio 3.

1. Ricordando che una variabile aleatoria di legge $U(a, b)$ ha funzione di ripartizione

$$F(t) = \begin{cases} 0 & \text{se } t \leq a, \\ \frac{t-a}{b-a} & \text{se } a \leq t \leq b, \\ 1 & \text{se } b \leq t, \end{cases}$$

abbiamo che

$$F_{cX}(t) = \mathbb{P}\{cX \leq t\} = \mathbb{P}\left\{X \leq \frac{t}{c}\right\} = \begin{cases} 0 & \text{se } t/c \leq a, \\ \frac{t/c-a}{b-a} & \text{se } a \leq t/c \leq b, \\ 1 & \text{se } b \leq t/c, \end{cases} = \begin{cases} 0 & \text{se } t \leq ca, \\ \frac{t-ca}{cb-ca} & \text{se } ca \leq t \leq cb, \\ 1 & \text{se } cb \leq t, \end{cases}$$

e segue la tesi.

2. Vedi punto 4.
3. Abbiamo $F_{cX}(t) = \mathbb{P}\{cX \leq t\} = \mathbb{P}\{X \leq \frac{t}{c}\} = F_X(\frac{t}{c})$.
4. Siccome F_X, e quindi F_{cX}, è C^1 a tratti, la densità si ottiene derivando:

$$f_{cX}(t) = F'_{cX}(t) = F'_X\left(\frac{t}{c}\right)\frac{1}{c} = \frac{\lambda^a}{\Gamma(a)}\left(\frac{t}{c}\right)^{a-1} e^{-\lambda\frac{t}{c}} \cdot \frac{1}{c}\mathbf{1}_{\{t/c\geq 0\}} = \frac{\left(\frac{\lambda}{c}\right)^a}{\Gamma(a)} t^{a-1} e^{-\frac{\lambda}{c}t}\mathbf{1}_{\{t\geq 0\}}$$

e segue la tesi.

Esercizio 4. Innanzitutto scriviamo $X = \sum_{i=1}^N X_i$, con X_i i.i.d. $\sim Po(30)$ e N variabile aleatoria uniforme sull'insieme $\{2,3,4\}$, chiaramente tale che $\mathbb{P}\{N = n\} = 1/3$ per $n = 2,3,4$. Se definiamo le probabilità $\mathbb{P}_n := \mathbb{P}(\cdot|\{N = n\})$, allora X sotto $\mathbb{P}_n$ ha legge $Po(30n)$; in particolare, $X \geq 0$ quasi certamente.

1. Usando la formula della disintegrazione su X, abbiamo

$$\mathbb{E}[X] = \sum_{n=2}^{4} \mathbb{E}_n[X]\mathbb{P}\{N = n\} = \frac{1}{3}(60 + 90 + 120) = 90$$

2. Siccome Var $[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, possiamo calcolare

$$\mathbb{E}[X^2] = \sum_{n=2}^{4} \mathbb{E}_n[X^2]\mathbb{P}\{N = n\} = \frac{1}{3}(60 + 3600 + 90 + 8100 + 120 + 14400) = 8790$$

e quindi Var $[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 8790 - 8100 = 690$.

3. Usando la formula della probabilità totale, abbiamo

$$\mathbb{P}\{X > 80\} = \sum_{n=2}^{4} \mathbb{P}\{X > 80 \mid N = n\}\mathbb{P}\{N = n\} = \sum_{n=2}^{4} \mathbb{P}_n\{X > 80\}\mathbb{P}\{N = n\}$$

Ricordandosi ora che $X \sim Po(30n)$ sotto $\mathbb{P}_n$ ed usando il suggerimento con la correzione di continuità, calcoliamo:

$$\begin{aligned}
\mathbb{P}_2\{X > 80\} &= \mathbb{P}_2\left\{\frac{X-60}{\sqrt{60}} > \frac{80.5-60}{\sqrt{60}}\right\} \simeq 1 - \Phi(2.64) = 0.00415 \\
\mathbb{P}_3\{X > 80\} &= \mathbb{P}_3\left\{\frac{X-90}{\sqrt{90}} > \frac{80.5-90}{\sqrt{90}}\right\} \simeq 1 - \Phi(-1.00) = \Phi(1.00) = 0.84134 \\
\mathbb{P}_4\{X > 80\} &= \mathbb{P}_4\left\{\frac{X-120}{\sqrt{120}} > \frac{80.5-120}{\sqrt{120}}\right\} \simeq 1 - \Phi(-3.60) = \Phi(3.60) = 0.99984
\end{aligned}$$

Abbiamo quindi

$$\mathbb{P}\{X > 80\} = \frac{1}{3}(0.00415 + 0.84134 + 0.99984) = 0.61511$$

Nota: se avessimo approssimato direttamente $X \sim N(90, \sqrt{690})$ (cosa che non ha alcun fondamento teorico, poichè X NON ha legge di Poisson), avremmo ottenuto:

$$\mathbb{P}\{X > 80\} = \mathbb{P}\left\{\frac{X-90}{\sqrt{690}} > \frac{80.5-90}{\sqrt{690}}\right\} \simeq \Phi(0.36) = 0.64058$$

risultato diverso.

4. Seguendo il suggerimento, siccome per un n fissato le $(Y_i^n)_i$ sono i.i.d $\in L^2$, sappiamo che vale l'approssimazione normale e che quindi possiamo approssimare

$$Y^* := \frac{Y - \mathbb{E}[Y]}{\sqrt{\text{Var}\,[Y]}} = \frac{Y - \lambda}{\sqrt{\lambda}}$$

con una normale $N(0,1)$ se n è "grande": tuttavia, $X \sim \sum_{i=1}^n Y_i$ per ogni n, quindi possiamo supporre che n sia "sufficientemente grande" (attenzione: Y^* rimane una variabile aleatoria discreta, concentrata sull'insieme ottenuto trasformando $\mathbb{N}$ con la funzione affine $n \rightarrow \frac{n-\lambda}{\sqrt{\lambda}}$). Ma allora possiamo approssimare $Y = \sqrt{\lambda}Y^* + \lambda$ con una legge $N(\lambda, (\sqrt{\lambda})^2) = N(\lambda, \lambda)$.

Notiamo infine che NON è possibile dimostrare una convergenza in legge col teorema limite centrale, e che quindi questa rimane un'approssimazione: si può dimostrare che la sua bontà dipende da quanto è grande λ; in particolare, per $\lambda > 5$ si ottengono già dei buoni risultati.

Esame di Calcolo delle Probabilità del 30 marzo 2009
(Corso di Laurea Triennale in Matematica, Universitá degli Studi di Padova)
(docente: Tiziano Vargiolu)

Hanno superato la prova:

Boarotto Francesco	$27 + 3^-$
Cimetta Leone Cesare	20
Di Gianbattista Francesco	$25.5 + 3^-$
Donini Michele	$20 + 3^-$
Grandi Riccardo	$23.5 + 3^+$
Maso Claudia	21
Mirandola Diego	$24 + 3^-$
Morino Laura	$14.5 + 3^+$
Paolini Alessandro	$18.5 + 3^+$
Piccininno Alessandro	$22.5 + 3^+$
Pieropan Marta	26.5
Tasinato Alessia	18.5
Tasinato Davide	$18 + 3^-$
Tessari Alessandro	19.5
Tormen Adriana	$18 + 3^-$
Vitturi Marco	24
Zanin Giorgia	$17.5 + 3^-$
Zoppello Marta	$21 + 3^+$

Visione compiti, registrazione voti e orali: giovedì 2 aprile ore 10.00 aula 1AD/50.
Verrà data precedenza alla registrazione voti a chi accetta il voto dello scritto e ha un bonus.