Esame di Probabilità e Statistica del 19 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. Una compagnia di assicurazioni classifica la popolazione in tre classi di rischio (basso, medio, alto) a seconda della loro probabilità di provocare un incidente in un anno, rispettivamente pari a 0.05, 0.15, 0.30. Se la distribuzione della popolazione tra queste tre classi è pari rispettivamente al 20%, 50% e 30%:

1. Quale percentuale della popolazione causerà un incidente nel corso di un anno?
2. Se un assicurato non ha causato incidenti nel 2006, qual è la probabilità che si tratti di una persona a rischio basso?
3. Nelle stesse ipotesi del punto 2., qual è la probabilità che si tratti di una persona a rischio medio?

Esercizio 2. Consideriamo un processo di Bernoulli $(X_n)_n$ di parametro p, e definiamo gli istanti dei primi due successi:

$$\begin{aligned} T_1 &:= \inf\{n > 0 \mid X_n = 1\}, \\ T_2 &:= \inf\{n > T_1 \mid X_n = 1\} \end{aligned}$$

1. Calcolare la funzione di ripartizione di T_2 e la sua densità discreta.
2. Supponiamo che $U := T_2 - T_1$ sia una variabile aleatoria indipendente da T_1 e con uguale legge $Ge(p)$. Usando questa ipotesi, calcolare la densità discreta di T_2.
3. Calcolare la densità congiunta di (T_1, T_2) senza usare il punto 2.
4. Calcolando la densità congiunta di (T_1, U), dimostrare che U è indipendente da T_1 e con legge $Ge(p)$.

Esercizio 3.

1. Siano $X, Y \sim N(0,1)$ indipendenti. Calcolare la distanza media tra le due variabili aleatorie (cioè $\mathbb{E}[|X - Y|]$).
2. Siano $X = (X_1, X_2)$, $Y = (Y_1, Y_2)$ vettori aleatori sul piano con $X_1, X_2, Y_1, Y_2 \sim N(0,1)$ indipendenti. Che legge ha il quadrato della distanza euclidea

$$D^2 := (X_1 - Y_1)^2 + (X_2 - Y_2)^2$$

(o meglio, $\frac{D^2}{2}$)?

3. Calcolare $\mathbb{E}[|D|]$.
4. Rispondere alle domande 2. e 3. nel caso in cui $X = (X_1, \ldots, X_d)$, $Y = (Y_1, \ldots, Y_d)$ siano vettori aleatori d-dimensionali con $X_1, \ldots, X_d, Y_1, \ldots, Y_d \sim N(0,1)$ indipendenti.

Esercizio 4. Il punteggio di un test si distribuisce come una variabile aleatoria di media 55.

1. Dare una limitazione alla probabilità che il punteggio superi 70.

Supponiamo di sapere anche che la varianza sia 25.

2. Dare di nuovo una limitazione alla probabilità che il punteggio superi 70, e alla probabilità che sia compreso tra 40 e 70.

3. In una classe di n studenti vorremmo che la votazione media sia compresa tra 53 e 57. Supponendo di poter utilizzare l'approssimazione normale, quanti studenti sono necessari affinchè la probabilità che questo accada sia almeno 0.99?

4. Se ci sono invece 50 studenti, qual è la probabilità che la votazione media sia compresa tra 53 e 57?

Soluzioni

Esercizio 1. Definiamo gli eventi

$$\begin{aligned} I &:= \{ \text{ causare un incidente nel corso dell'anno } \}, \\ A &:= \{ \text{ essere della classe di rischio alta } \}, \\ M &:= \{ \text{ essere della classe di rischio media } \}, \\ B &:= \{ \text{ essere della classe di rischio bassa } \} \end{aligned}$$

allora abbiamo che

$$\mathbb{P}(B) = 0.2, \qquad \mathbb{P}(M) = 0.5, \qquad \mathbb{P}(A) = 0.3$$

$$\mathbb{P}(I|B) = 0.05, \qquad \mathbb{P}(I|M) = 0.15, \qquad \mathbb{P}(I|A) = 0.30$$

1. Poichè Ω è unione disgiunta degli eventi B, M e A e questi sono tutti non trascurabili, possiamo applicare la formula della probabilità totale:

$$\mathbb{P}(I) = \mathbb{P}(I|B)\mathbb{P}(B)+\mathbb{P}(I|M)\mathbb{P}(M)+\mathbb{P}(I|A)\mathbb{P}(A) = 0.05\cdot 0.2+0.15\cdot 0.5+0.3\cdot 0.3 = 0.175$$

2. Dato che sia B che I^c sono non trascurabili, possiamo usare la formula di Bayes:

$$\mathbb{P}(B|I^c) = \frac{\mathbb{P}(I^c|B)\mathbb{P}(B)}{\mathbb{P}(I^c)} = \frac{(1-\mathbb{P}(I|B))\mathbb{P}(B)}{1-\mathbb{P}(I)} = \frac{0.95\cdot 0.2}{0.825} = 0.230303$$

3. Dato che M è non trascurabile, possiamo usare anche qui la formula di Bayes:

$$\mathbb{P}(M|I^c) = \frac{\mathbb{P}(I^c|M)\mathbb{P}(M)}{\mathbb{P}(I^c)} = \frac{(1-\mathbb{P}(I|M))\mathbb{P}(M)}{1-\mathbb{P}(I)} = \frac{0.85\cdot 0.5}{0.825} = 0.515151$$

Esercizio 2.

1. Poichè T_2 assume solo valori naturali, è sufficiente calcolare la funzione di ripartizione per $t \in \mathbb{N}$, $t \geq 2$: per come è definita T_2, si ha che

$$\begin{aligned} F_{T_2}(t) &= \mathbb{P}\{T_2 \leq t\} = \mathbb{P}\left\{\sum_{i=1}^{t} X_i \geq 2\right\} = 1 - \mathbb{P}\left\{\sum_{i=1}^{t} X_i = 0\right\} - \mathbb{P}\left\{\sum_{i=1}^{t} X_i = 1\right\} = \\ &= 1-(1-p)^t - tp(1-p)^{t-1} \end{aligned}$$

La densità discreta si può poi ottenere come

$$\begin{aligned} p_{T_2}(t) &= \mathbb{P}\{T_2 = t\} = F_{T_2}(t) - F_{T_2}(t-1) = \\ &= 1-(1-p)^t - t(1-p)^{t-1} - 1 + (1-p)^{t-1} + (t-1)(1-p)^{t-2} = \\ &= (t-1)p^2(1-p)^{t-2} \end{aligned}$$

oppure dal punto 3. sommando sulla densità congiunta.

2. Supponendo verificate le ipotesi del punto 2., si ha $T_2 = T_1 + U$, con T_1, U indipendenti e di legge $Ge(p)$. Allora, poichè $T_1, U \geq 1$,

$$\begin{aligned}
\mathbb{P}\{T_2 = k\} &= \mathbb{P}\{T_1 + U = k\} = \sum_{h=1}^{k-1} \mathbb{P}\{T_1 = h, U = k-h\} = \\
&= \sum_{h=1}^{k-1} \mathbb{P}\{T_1 = h\}\mathbb{P}\{U = k-h\} = \sum_{h=1}^{k-1} p(1-p)^{h-1}p(1-p)^{k-h-1} = \\
&= \sum_{h=1}^{k-1} p^2(1-p)^{k-2} = (k-1)p^2(1-p)^{k-2}
\end{aligned}$$

3. Per $1 \leq h < k$ abbiamo

$$\begin{aligned}
\mathbb{P}\{T_1 = h, T_2 = k\} &= \mathbb{P}\{X_1 = \ldots = X_{h-1} = 0, X_h = 1, X_{h+1} = \ldots = 0, X_k = 1\} = \\
&= (1-p)^{h-1}p(1-p)^{k-1-h-1+1}p = p^2(1-p)^{k-2}
\end{aligned}$$

e chiaramente $\mathbb{P}\{T_1 = h, T_2 = k\} = 0$ altrimenti.

4. Per $h, k \geq 1$ abbiamo

$$\begin{aligned}
\mathbb{P}\{T_1 = h, U = k\} &= \mathbb{P}\{T_1 = h, T_2 - T_1 = k\} = \mathbb{P}\{T_1 = h, T_2 = h+k\} = \\
&= p^2(1-p)^{h+k-2} = p(1-p)^{h-1}p(1-p)^{k-1} = p_{T_1}(h)p_U(k)
\end{aligned}$$

che dimostra contemporaneamente che $U \sim Ge(p)$ e che è indipendente da T_1.

Esercizio 3.

1. $X - Y \sim N(0,2)$, quindi $X - Y = \sqrt{2}Z$, con $Z \sim N(0,1)$; poichè $|X - Y| \geq 0$ q.c., possiamo calcolare

$$\mathbb{E}[|X-Y|] = \sqrt{2}\mathbb{E}[|Z|] = \sqrt{2} \cdot 2 \int_0^{+\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz = \frac{2}{\sqrt{\pi}}[-e^{-\frac{z^2}{2}}]_0^{+\infty} = \frac{2}{\sqrt{\pi}}$$

2. $X_i - Y_i \sim N(0,2)$, quindi $X_i - Y_i = \sqrt{2}Z_i$, con $Z_i \sim N(0,1)$. Quindi $D^2 = 2(Z_1^2 + Z_2^2) = 2Z$, con $Z \sim \chi^2(2) = \Gamma(1, \frac{1}{2})$.

3. Poichè $|D| \geq 0$ q.c., possiamo calcolare

$$\begin{aligned}
\mathbb{E}[|D|] &= \mathbb{E}[\sqrt{2Z}] = \sqrt{2}\mathbb{E}[\sqrt{Z}] = \sqrt{2}\int_0^{+\infty} z^{1/2} \frac{\left(\frac{1}{2}\right)^1}{\Gamma(1)} z^{1-1} e^{-z/2}\, dz = \\
&= \frac{1}{\sqrt{2}} \int_0^{+\infty} z^{1/2} e^{-z/2}\, dz = \frac{1}{\sqrt{2}} \frac{\Gamma(3/2)}{\left(\frac{1}{2}\right)^{3/2}} = 2\Gamma\left(\frac{3}{2}\right) = 2 \cdot \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}
\end{aligned}$$

dove abbiamo usato il fatto che $z^{1/2}e^{-z/2}$ è, a meno della costante di normalizzazione $\frac{\left(\frac{1}{2}\right)^{3/2}}{\Gamma(3/2)}$, la densità di una $\Gamma(\frac{3}{2}, \frac{1}{2})$.

4. Abbiamo ancora $X_i - Y_i \sim N(0,2)$, quindi $X_i - Y_i = \sqrt{2} Z_i$, con $Z_i \sim N(0,1)$. Quindi $D^2 = 2(Z_1^2 + \ldots + Z_d^2) = 2Z$, con $Z \sim \chi^2(d) = \Gamma(\frac{d}{2}, \frac{1}{2})$. Anche qui $|D| \geq 0$ q.c., quindi

$$\begin{aligned} \mathbb{E}[|D|] &= \mathbb{E}[\sqrt{2Z}] = \sqrt{2}\mathbb{E}[\sqrt{Z}] = \sqrt{2} \int_0^{+\infty} z^{1/2} \frac{\left(\frac{1}{2}\right)^{d/2}}{\Gamma(\frac{d}{2})} z^{\frac{d}{2}-1} e^{-z/2}\, dz = \\ &= \frac{1}{2^{\frac{d}{2}-\frac{1}{2}}\Gamma(\frac{d}{2})} \int_0^{+\infty} z^{\frac{d}{2}-\frac{1}{2}} e^{-z/2}\, dz = \frac{2\Gamma(\frac{d}{2}+\frac{1}{2})}{\Gamma(\frac{d}{2})} \end{aligned}$$

dove abbiamo usato il fatto che $z^{\frac{d}{2}-\frac{1}{2}} e^{-z/2}$ è, a meno della costante di normalizzazione $\frac{\left(\frac{1}{2}\right)^{\frac{d}{2}+\frac{1}{2}}}{\Gamma(\frac{d}{2}+\frac{1}{2})}$, la densità di una $\Gamma(\frac{d}{2}+\frac{1}{2}, \frac{1}{2})$.

Esercizio 4.

1. Chiamiamo X la variabile aleatoria corrispondente al punteggio del test. Si può ragionevolmente supporre che $X \geq 0$. Allora, per la disuguaglianza di Markov, si ha

$$\mathbb{P}\{X > 70\} \leq \frac{\mathbb{E}[X]}{70} = \frac{55}{70} = 0.78571$$

2. Per la disuguaglianza di Chebichev si ha

$$\mathbb{P}\{40 < X < 70\} = \mathbb{P}\{|X - 55| < 15\} \geq 1 - \frac{\text{Var}\,[X]}{15^2} = \frac{8}{9} = 0.88888$$

e quindi

$$\mathbb{P}\{X > 70\} \leq \mathbb{P}\{|X - 55| > 15\} \leq \frac{\text{Var}\,[X]}{15^2} = \frac{1}{9} = 0.111111$$

3. Chiamiamo ora X_i il risultato dell'i-esimo studente, e $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$. Stiamo supponendo che la legge di $\bar{X}_n$ si possa approssimare con una $N(55; \frac{25}{n})$. Allora, centrando e riducendo $\bar{X}_n$, si ha

$$\begin{aligned} 0.99 &\leq \mathbb{P}\{53 < \bar{X}_n < 57\} = \mathbb{P}\{|\bar{X}_n - 55| < 2\} = \mathbb{P}\left\{ \frac{|\bar{X}_n - 55|}{\frac{5}{\sqrt{n}}} < \frac{2}{5}\sqrt{n} \right\} = \\ &\simeq \Phi\left(\frac{2}{5}\sqrt{n}\right) - \Phi\left(-\frac{2}{5}\sqrt{n}\right) = 2\Phi\left(\frac{2}{5}\sqrt{n}\right) - 1 \end{aligned}$$

quindi

$$\Phi\left(\frac{2}{5}\sqrt{n}\right) \geq 0.995$$

che significa

$$\frac{2}{5}\sqrt{n} \geq q_{0.995} = 2.57$$

che, svolgendo i calcoli, diventa $n \geq 41.28$, ossia $n \geq 42$.

4. Calcoliamo

$$\begin{aligned} \mathbb{P}\{53 < \bar{X}_{50} < 57\} &= \mathbb{P}\left\{ \frac{|\bar{X}_n - 55|}{\frac{5}{\sqrt{50}}} < \frac{2}{5}\sqrt{50} \right\} \simeq \Phi\left(2\sqrt{2}\right) - \Phi\left(-2\sqrt{2}\right) = \\ &= 2\Phi\left(2\sqrt{2}\right) - 1 = 2\Phi(2.82) - 1 = 2 \cdot 0.99760 - 1 = 0.99520 \end{aligned}$$

Esame di Calcolo delle Probabilità del 19 marzo 2009
(Corso di Laurea Triennale in Matematica, Universitá degli Studi di Padova)
(docente: Tiziano Vargiolu)

Hanno superato la prova:

Balbinot Sarah	$22 + 3^-$
Basei Matteo	$22.5 + 3^+$
Boarotto Francesco	$16.5 + 3^-$
Chiarini Alberto	$28 + 3^+$
Cielo Francesco	19.5
De Anna Francesco	$20.5 + 3^+$
Donini Michele	$18 + 3^-$
Galliani Davide	$23 + 3^-$
Griggio Jacopo	$22.5 + 3^+$
Guarniero Pieralberto	19
Lovato Ilaria	$16.5 + 3^-$
Mirandola Diego	$21 + 3^-$
Paolini Alessandro	$22 + 3^+$
Piccininno Alessandro	$19 + 3^+$
Pieropan Marta	20
Piovesan Teresa	$17 + 3^-$
Zocca Alessandro	$21 + 3^+$
Zoppello Marta	$15 + 3^+$

Visione compiti, registrazione voti e orali: mercoledì 25 marzo ore 9.30 aula 1AD/50.
Verrà data precedenza alla registrazione voti a chi accetta il voto dello scritto e ha un bonus.