Corso di Statistica - Prof. Fabio Zucca	I Prova in itinere - 7 maggio 2015		
Nome e cognome:		**Matricola:**	
©I diritti d'autore sono riservati. Ogni sfruttamento commerciale non autorizzato sarà perseguito.			12957

Esercizio 1 Si consideri la seguente famiglia di funzioni reali di variabili reali a due parametri reali $c, \lambda \in \mathbb{R}$

$$f_{\lambda,c}(x) := \begin{cases} \lambda e^{-xc} & \text{se } x \geq 0, \\ 0 & \text{altrimenti.} \end{cases}$$

1. Per quali valori di λ, c la funzione $f_{\lambda,c}$ rappresenta la densità di una variabile assolutamente continua? *Suggerimento: si supponga c fissato e si ricavi λ come funzione di c.*

2. Al variare delle coppie (λ, c) ammissibili calcolate in precedenza e denotato con X la variabile, se ne calcoli il valore atteso e la funzione di ripartizione.
Suggerimento: si utilizzi l'integrazione per parti.

3. Quanto vale $\mathbb{P}(X \geq t + s | X \geq t)$ con $t, s \geq 0$?

4. Sia $\{X_i\}_{i=1}^{100}$ una successione i.i.d. di variabili con densità $f_{1,1}$. Assumendo le uguaglianze $\mathbb{E}(X_i) = \text{var}(X_i) = 1$, calcolare approssimativamente $\mathbb{P}\Big(\sum_{i=1}^{100} X_i \geq 100\Big)$.

Soluzione.

1. Una primitiva di $f_{\lambda,c}$ è $x \mapsto -\lambda e^{-xc}/c$ se $c \neq 0$ (mentre è λx se $c = 0$), inoltre la condizione $f_{\lambda,c} \geq 0$ (quasi ovunque) è equivalente a $\lambda \geq 0$. Se $c \leq 0$ la funzione e^{-xc} non ha integrale finito e non esiste alcun valore di λ che verifichi $\int_{\mathbb{R}} f_{\lambda,c}(x)\mathrm{d}x = 1$. Quindi d'ora in poi supporremmo $c > 0$. Pertanto

$$1 = \int_{-\infty}^{+\infty} f_{\lambda,c}(x)\mathrm{d}x = \int_0^{+\infty} \lambda e^{-xc}\mathrm{d}x = (-\lambda e^{-xc}/c)|_0^{+\infty} = \lambda/c$$

da cui $c = \lambda$. Quindi i valori ammissibili sono tutte le coppie (λ, λ) al variare di $\lambda > 0$.

2. Calcoliamo, con $c = \lambda$,

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_{\lambda,\lambda}(x)\mathrm{d}x = \int_0^{+\infty} x\lambda e^{-x\lambda}\mathrm{d}x = -e^{-x\lambda}(1/\lambda + x)|_0^{+\infty} = 1/\lambda.$$

Analogamente si sarebbe potuto procedere tramite cambio di variabile $z = \lambda x$ ottenendo

$$\begin{aligned}\mathbb{E}[X] &= \int_{\mathbb{R}} x f_{\lambda,\lambda}(x)\mathrm{d}x = \int_0^{+\infty} x\lambda e^{-x\lambda}\mathrm{d}x = \\ &= \lambda^{-1}\int_0^{+\infty} z e^{-z}\mathrm{d}z = -\lambda^{-1}e^{-z}(1+z)|_0^{+\infty} = 1/\lambda.\end{aligned}$$

Calcoliamo ora,

$$F_X(t) := \mathbb{P}(X \leq t) = \int_{-\infty}^{t} f_{\lambda,\lambda}(x)\mathrm{d}x = \begin{cases} 0 & \text{se } t \leq 0, \\ -e^{-x\lambda}|_0^t = 1 - e^{-t\lambda} & \text{se } t > 0. \end{cases}$$

da cui, in forma compatta, $F_X(t) = \mathbb{1}_{(0,+\infty)}(t) \cdot (1 - e^{-t\lambda})$.

3. Dalla definizione, considerando che $t, s \geq 0$ ed utilizzando la forma esplicita della funzione di ripartizione calcolata in precedenza,

$$\begin{aligned}\mathbb{P}(X \geq t+s | X \geq t) &= \frac{\mathbb{P}(X \geq t+s, X \geq t)}{\mathbb{P}(X \geq t)} = \frac{\mathbb{P}(X \geq t+s)}{\mathbb{P}(X \geq t)} \\ &= \frac{1 - F_X(t+s)}{1 - F_X(t)} = \frac{e^{-(t+s)\lambda}}{e^{-t\lambda}} = e^{-s\lambda} = \mathbb{P}(X \geq s).\end{aligned}$$

La variabile X gode quindi della proprietà di assenza di memoria.

4. Si è visto al punto (2) che $\mathbb{E}(X_i) = 1/\lambda = 1$. Si potrebbe calcolare, come al punto (2), la varianza di una variabile di densità $f_{\lambda,\lambda}$ ottenendo $1/\lambda^2$. Quindi le due assunzioni sono consistenti. Utilizzando l'approssimazione derivante dal TCL e ricordando che $\mathbb{E}(X_i) = 1/\lambda = 1$, si ha

$$\begin{aligned}\mathbb{P}\Big(\sum_{i=1}^{100} X_i \geq 100\Big) &= \mathbb{P}\Big(\frac{\sum_{i=1}^{100} X_i - 100}{\sqrt{100}} \geq 0\Big) \\ &\approx 1 - \Phi(0) = \frac{1}{2}.\end{aligned}$$

Oss. La famiglia di variabili assolutamente continue (ad un parametro) di densità $f_{\lambda,\lambda}$ prende il nome di famiglia di variabili esponenziali.

Esercizio 2 In una certa regione, i terremoti si susseguono secondo un processo di Poisson di intensità pari a 6 all'anno.

1. Qual è la probabilità che vi siano almeno due terremoti nella seconda metà del 2015?

2. Qual è la probabilità che esattamente in 9 anni su 10 vi siano dei terremoti?

3. Sapendo che ci sono stati $N = 12$ terremoti nell'arco di 3 anni, qual è la probabilità che ci siano stati $N_a = 5$ terremoti nel primo dei tre anni (e quindi $N_b = 7$ nei successivi due anni)?

4. **(Domanda bonus)** Sapete riconoscere la legge di probabilità del numero di terremoti che accadono nel primo anno sapendo che nei primi tre ne sono accaduti complessivamente 12?
Sugg.: può essere utile ricordare che $\{N_a = 5, N_b = 7\} \subseteq \{N_a + N_b = 5 + 7\} = \{N = 12\}$.

Soluzione. Per ogni tempo $t \geq 0$, dove t è misurato in anni, sia

$$N_t = \text{numero di terremoti in un intervallo di tempo di ampiezza } t.$$

Sappiamo che $(N_t)_{t\geq 0}$ è un processo di Poisson di intensità $\lambda = 6$ anni^{-1}.

1. $N_{1/2} \sim \mathcal{P}(3)$, quindi

$$\begin{aligned}\mathbb{P}(N_{1/2} \geq 2) &= 1 - \mathbb{P}(N_{1/2} = 0) - \mathbb{P}(N_{1/2} = 1) \\ &= 1 - e^{-3} - 3e^{-3} \approx 0.80085.\end{aligned}$$

2. Siano $\{T_i\}_{i=1}^{10}$ le variabili i.i.d. tali che T_i è il numero di terremoti nell'anno i. Chiaramente T_i è una variabile di Poisson di parametro 6. Sia X_i la variabile che vale 1 se c'è almeno un terremoto nell'anno i e 0 altrimenti. Sono tutte variabili di Bernoulli i.i.d. di parametro $\mathbb{P}(X_i = 1) = \mathbb{P}(T_i \geq 1) = 1 - \mathbb{P}(T_i = 0) = 1 - e^{-6} \approx 0.9975$. Quindi $\mathbb{P}(\sum_{i=1}^{10} X_i = 9) = 10(1 - e^{-6})^9 e^{-6} \approx 0.02424$.

3. Si veda la soluzione generale del punto seguente.

4. Sia $N := N_a + N_b$ e $k := k_a + k_b$, dove $N_a \sim Poi(\lambda_a)$ e $N_b \sim Poi(\lambda_b)$ (qui non c'è la notazione N_t numero di terremoti in intervallo di ampiezza t). Osserviamo che l'intersezione di una qualsiasi coppia di eventi tra $\{N_a = k_a\}$, $\{N_b = k_b\}$ e $\{N = k\}$ implica il terzo evento, pertanto $\{N_a = k_a, N_b = k_b\} \cap \{N = k\} = \{N_a = k_a, N_b = k_b\}$; ciò che a noi serve in realtà à la seguente catena di uguaglianze tra eventi $\{N_a = k_a, N_b = k_b\} = \{N_a = k_a, N_a + N_b = k\} = \{N_a = k_a, N = k\}$. Notiamo quindi che l'evento $\{N_a = k_a, N = k\}$ è in realtà riscrivibile come intersezione di due eventi indipendenti $\{N_a = k_a, N_b = k_b\}$, da cui

$$\begin{aligned}\mathbb{P}(N_a = k_a | N = k) &= \frac{\mathbb{P}(N_a = k_a, N = k)}{\mathbb{P}(N = k)} = \frac{\mathbb{P}(N_a = k_a, N_b = k_b)}{\mathbb{P}(N = k)} \\ &= \frac{e^{\lambda_a}\lambda_a^{k_a}/k_a! e^{\lambda_b}\lambda_b^{k_b}/k_b!}{e^{\lambda}\lambda^k/k!} = \binom{k}{k_a}\Big(\frac{\lambda_a}{\lambda}\Big)^{k_a}\Big(\frac{\lambda_b}{\lambda}\Big)^{k_b} \\ &= \binom{k}{k_a}\Big(\frac{\lambda_a}{\lambda}\Big)^{k_a}\Big(1 - \frac{\lambda_a}{\lambda}\Big)^{k-k_a}\end{aligned}$$

La distribuzione condizionata della variabile N_a rispetto a $\{N = k\}$ è la distribuzione binomiale $B(k, \lambda_a/\lambda)$.

La risposta alla domanda del punto precedente, essendo in questo caso $\lambda_b = 2\lambda_a = 2 \cdot 6$ è quindi

$$\mathbb{P}(N_a = 5, N_b = 7 | N = 12) = \binom{12}{5}\frac{1}{3^5}\Big(\frac{2}{3}\Big)^7 \approx 0.190756829.$$