Lezione del 7/4.
Alcuni esercizi da Blitzstein, 4. (1)

- Codifichiamo i giorni della settimana da 1 a 7, e scegliamo a caso uno tra questi, che chiamiamo X. Sia quindi Y il numero del giorno successivo ad X.

a) Determinare le PMF di X e di Y.

Evidentemente

$$\mathbb{P}(X=k) = 1/7 \quad k=1,2,\dots,7.$$

Ora Y è una funzione bijettiva su $\{1,\dots,7\}$ (in effetti una permutazione) e quindi ogni elemento del dominio $\{1,\dots,7\}$ è messo in corrispondenza con uno ed un solo elemento dell'immagine $\{1,\dots,7\}$ "trasmettendogli" la sua massa da cui

$$\mathbb{P}(Y=k) = 1/7, \quad k=1,2,\dots,7.$$

b) Determinare $\mathbb{P}(X<Y)$

Evidentemente $\{X<Y\} = \{X \neq 7\}$ e quindi

$$\mathbb{P}(X<Y) = \mathbb{P}(X \neq 7) = \frac{6}{7}$$

Osservazione 1. E' chiaro che $\{X=Y\} = \emptyset$,

quindi $\mathbb{P}(X=Y)=0$, anche se $X \stackrel{\mathcal{L}}{=} Y$.

Riflessione. Considerando un mese X dell'anno scelto a caso, e Y il mese successivo. E' evidente che

$$\mathbb{P}(X=k) = \mathbb{P}(Y=k) = \frac{1}{12} \quad , k=1,2,\dots,12$$

e quanto vale $\mathbb{P}(X<Y)$

$$\mathbb{P}(X<Y) = \mathbb{P}(X \neq 12) = \frac{11}{12}.$$

Osservazione 2. In questo modo, costruiamo per ogni N variabili aleatorie X e Y tali che

$$X \stackrel{\mathcal{L}}{=} Y \qquad \mathbb{P}(X<Y) = \frac{N-1}{N}$$

E' possibile che $X \stackrel{\mathcal{L}}{=} Y$ e $\mathbb{P}(X<Y) = 1$?

Se il supporto della PMF di X è finito è facile rispondere.

- È possibile che $\mathbb{E}(X) \geq 10^n \mathbb{E}(Y) \geq 0$ e $\mathbb{P}(Y > X) \geq 1 - 10^{-m}$?

Sì. Sia Y una variabile aleatoria positiva di media finita e sia Z una variabile di Bernoulli indipendente da Y con $\mathbb{P}(Z=1) = 10^{-m}$ e $X = 10^l YZ$. Allora

$$\mathbb{P}(Y > X) \geq \mathbb{P}(Z=0) = 1 - 10^{-m}$$

mentre per LOTP

$$\mathbb{E}(X) = \mathbb{P}(Z=0)\,\mathbb{E}(X|Z=0) + \mathbb{P}(Z=1)\,\mathbb{E}(X|Z=1)$$

$$= \quad 0 \quad + 10^{-m} \cdot 10^l\, \mathbb{E}(Y)^*$$

Basta quindi scegliere $l - m \geq n$.

* Si noti $\mathbb{E}(X|Z=1) = \mathbb{E}(10^l Y | Z=1)$

$$= 10^l \underbrace{\mathbb{E}(Y|Z=1) = 10^l \mathbb{E}(Y)}$$

$\hookrightarrow$ dato che Y è indipendente da Z.

C'è una formuletta utilissima che è opportuno provare (la spesa vale l'impresa...).

Supponiamo che X sia una variabile aleatoria a valori interi non negativi.

Allora $\mathbb{E}(X) = \sum_{k=1}^{\infty} k \, \mathbb{P}(X=k) =$

$$
\begin{array}{c|cccc}
n & & & & \\
4 & & & & \mathbb{P}(X=4) \\
3 & & & \mathbb{P}(X=3) & \mathbb{P}(X=4) \\
2 & & \mathbb{P}(X=2) & \mathbb{P}(X=3) & \mathbb{P}(X=4) \\
1 & \mathbb{P}(X=1) & \mathbb{P}(X=2) & \mathbb{P}(X=3) & \mathbb{P}(X=4) \\
\hline
 & 1 & 2 & 3 & 4 \quad k
\end{array}
$$

rappresenta la somma colonna per colonna della matrice (che potrebbe essere anche infinita) riportata qui sopra. Sommando invece sulle righe (e questo si può giustificare finché si sommano quantità non negative)

$$= \mathbb{P}(X \geq 1) + \mathbb{P}(X \geq 2) + \mathbb{P}(X \geq 3) + \dots = \sum_{n=1}^{\infty} \mathbb{P}(X \geq n)$$

Chi aspira ad una dimostrazione più formale

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k\,\mathbb{P}(X=k) = \sum_{k=1}^{\infty} \sum_{n=1}^{k} 1 \cdot \mathbb{P}(X=k) =$$

$$= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mathbb{P}(X=k) = \sum_{n=1}^{\infty} \mathbb{P}(X \geq n)$$

Attenzione che la stessa formula si può trovare scritta nella forma

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} \mathbb{P}(X > n) = \sum_{n=0}^{\infty} (1 - F_X(n))$$

dove F_X è la CDF di X.

Esempio: $X \sim$ Geometrica (p)

$\mathbb{P}(X > n) = \mathbb{P}(\text{primi } n \text{ insuccessi}) = q^n$, quindi

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} q^n = \frac{1}{1-q} = \frac{1}{p}, \text{ come già visto.}$$

La formula vale anche quando $\mathbb{E}(X) = +\infty$.

Esempio: $\mathbb{P}(X=n) = \frac{1}{n} - \frac{1}{n+1}$, $n = 1, 2, \ldots$ (Mengoli)

$$\mathbb{P}(X \geq n) = \frac{1}{n} \Rightarrow \mathbb{E}(X) = \sum_{n=1}^{\infty} \frac{1}{n} = +\infty.$$

Se $\mathbb{P}(X \leq M) = 1$, allora

$$X = \sum_{k=1}^{M} 1_{\{X \geq k\}} \Rightarrow \mathbb{E}(X) = \sum_{k=1}^{M} \mathbb{E}(1_{\{X \geq k\}}) = \sum_{k=1}^{M} \mathbb{P}(X \geq k)$$

Il risultato precedentemente ottenuto è conseguenza della linearità del valore atteso. Se non esiste M con le proprietà di cui sopra, cioè il supporto di X è infinito, serve anche qualche risultato di passaggio al limite

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{k=1}^{\infty} 1_{\{X \geq k\}}\right) = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k)$$

Passiamo ora alla risoluzione degli esercizi proposti per il fine settimana. Per ciascuno di essi cercherò di dare indicazioni sulla appropriatezza di un'approssimazione di Poisson e, nei casi in cui è possibile dato lo stato delle nostre conoscenze, calcolare anche la varianza (nel caso, quindi, di somme di variabili indipendenti).

Lezione 7/4.

Altri esercizi sul paradigma di Poisson (pag. 178, Blitzstein)

1. Ci sono k palline e n scatole (distinguibili)

Le palline sono piazzate a caso nelle scatole, con le n^k configurazioni ugualmente probabili.

Siano $n = 1000$, $k = 5806$. Determinare:

a) il numero medio di scatole <u>vuote</u>;

b) la probabilità che <u>almeno</u> una scatola sia vuota (lasciare n e k generici)

c) una buona approssimazione per la probabilità in b).

Risoluzione: Questo rientra nella tipologia studiata in precedenza 1 b) $(k_n = O(n \log n))$

a) $$n\left(1-\frac{1}{n}\right)^k = 1000\left(\frac{999}{1000}\right)^{5806} \simeq 3$$

$$10^{3 - 5806\,(\cdot 4345 \times 10^{-7})} = 10^{3-2,52271} \simeq 3$$

b) Per il principio di inclusione-esclusione:

$$\sum_{j=1}^{n} (-1)^{j+1} \binom{n}{j} \left(1-\frac{j}{n}\right)^k$$

c) Approssimazione Poisson (3) $= 1-e^{-3} \simeq 1-\dfrac{1}{20} = 0.95$

2. Torniamo al problema dei compleanni:
Qual'è la probabilità che con $k = 23$ persone ci sia almeno un compleanno in comune?

Risoluzione. Si tratta di un problema della tipologia 1 f) con $k = 23$, $n = 365$.
Ovviamente la risposta esatta è

$$1 - \frac{364 \cdot 363 \cdot \ldots \cdot 343}{365^{22}}$$

ma dato che $k^2 = 529$ è dell'ordine di grandezza di $n = 365$ si può approssimare la legge del numero di coppie con un compleanno in comune con una Poisson di media $\frac{k(k-1)}{2n} = \frac{23 \cdot 11}{365} = \frac{253}{365}$

La probabilità che una variabile con questa legge sia non nulla è $1 - e^{-253/365} \cong 1 - e^{-0.69315} \cong \frac{1}{2}$

che conferma la sensibilità di questa approssimazione.

3. Ci sono 1600 studenti iscritti ad un corso di laurea. Calcolare la probabilità (approssimata) che ci siano almeno 2 studenti che non solo hanno lo stesso compleanno, ma sono nati alla stessa ora e allo stesso minuto.

Risoluzione. Stavolta $k = 1600$ ma $n = 365 \times 24 \times 60 = 525600$. Il numero atteso di coincidenze è

$$\frac{1600 \cdot 1599}{2 \times 525600} = 2,434.$$

Per una variabile aleatoria di Poisson con questa media la probabilità che non sia nulla è $1 - e^{-2,434} \cong 0.9122$

4. Infine, per il problema delle coincidenze di Montmort, vi invito a visitare il sito

ttps://demonstrations.wolfram.com/Montmorts/problem