Lezione del 2/4. (1)

Abbiamo presentato, nell'ultima parte della scorsa lezione, una importante variabile aleatoria $X_r$ su di uno spazio di Bernoulli $(p)$: la prova in cui si verifica l'$r$-esimo successo. Abbiamo fatto vedere che il numero delle prove successive all'$(i-1)$-esimo successo, prima di vedere l'$i$-esimo, è la variabile $X_i - X_{i-1}$ ha sempre distribuzione geometrica $(p)$, qualunque sia $i$. Di conseguenza la legge di $X_r$ è la $r$-esima potenza di convoluzione della legge geometrica: in altre parole

$$X_r = (X_r - X_{r-1}) + (X_{r-1} - X_{r-2}) + \dots + (X_2 - X_1) + X_1$$

è somma di $r$ variabile con distribuzione geometrica $(p)$, che per di più sono anche INDIPENDENTI.
Abbiamo quindi dedotto che $\mathbb{E}(X_r) = r\,\mathbb{E}(X_1) = r/p$.
Cerchiamo ora di determinare la PMF di $X_r$.
Quanto ottenuto finora ci fornisce un modo difficile di arrivarci, cui accenniamo di seguito.

Supponiamo per il momento $r=2$. Allora

$$\mathbb{P}(X_2 = \ell) = \sum_{k=1}^{\ell-1} \mathbb{P}(X_1 = k) \cdot \mathbb{P}(X_2 - X_1 = \ell - k)$$

(la convoluzione delle PMF di $X_1$ e $X_2 - X_1$, dato che queste sono indipendenti).

Inoltre sono entrambe geometriche, quindi

$$= \sum_{k=1}^{\ell-1} q^{k-1} p \, q^{\ell-k-1} p = p^2 q^{\ell-2} \sum_{k=1}^{\ell-1} 1 = (\ell-1) p^2 q^{\ell-2}$$

dove $\ell = 2, 3, \ldots$ (il secondo successo almeno alla seconda prova!)

Con difficoltà crescenti al crescere di $r$ si può procedere su questa strada per induzione su $r$: ad esempio, per $r=3$, $X_3 = (X_3 - X_2) + X_2$ è la convoluzione di una geometrica $(p)$ con la PMF or ora ottenuta, e così via...

Tuttavia proprio la forma della PMF di $X_2$ ci suggerisce che il calcolo diretto è tutt'altro che impossibile. Infatti

$$\mathbb{P}(X_2=\ell)=\mathbb{P}(\underbrace{TCC\cdots CT}_{\ell})+\mathbb{P}(\underbrace{CTC\cdots CT}_{\ell})+\dots+\mathbb{P}(\underbrace{C\cdots CTT}_{\ell})=$$

$= (\ell-1)\,p^2 q^{\ell-2}$ che avevamo già calcolato.

In generale l'evento $\{X_r=\ell\}$ è realizzato da tutte e sole le stringhe di lunghezza $\ell$ di Teste e Croci che finiscono con una Testa e contengono altre $(r-1)$ Teste. Tutte queste $\ell$-ple hanno probabilità $p^r q^{\ell-r}$ e sono tante quanti sono i modi di scegliere, tra le $(\ell-1)$ prove il cui risultato non è fissato, i posti dove si verificano le $r-1$ teste. Riassumendo:

$$\mathbb{P}(X_r=\ell)=\binom{\ell-1}{r-1}p^r q^{\ell-r}\,,\quad \ell=r,r+1,\dots$$

detta PMF Binomiale Negativa $(r,p)$.

Ovviamente per $r=1$ si ritrova la PMF geometrica $(p)$, visto che $\binom{\ell-1}{0}=1$.

Trovo a questo punto opportuno mettervi in guardia rispetto ad una spiacevole circostanza: a seconda dei testi, gli autori si riferiscono alle leggi binomiali negative (inclusa la geometrica) per le PMF delle variabili $X_r$ che abbiamo definito all'inizio della lezione. Altrettanti autori danno invece questi nomi alle PMF delle variabili $X'_r = X_r - r$, che contano gli insuccessi prima dell'r-esimo successo (sottrarre $r$ significa sottrarre le prove in cui si sono verificati i successi), quindi

$$\mathbb{P}(X'_r = k) = \mathbb{P}(X_r = r+k) = \binom{r+k-1}{r-1} p^r q^k$$

$k = 0, 1, 2, \ldots$ (notate che per queste variabili la PMF ha l'insieme degli interi non-negativi quale supporto).

Ad esempio, per alcuni autori (incluso Blitzstein!) la PMF geometrica è la seguente

$$P(X_1' = k) = p\,q^k, \quad k = 0, 1, 2, \dots$$

Un modo di visualizzare la trasformazione della PMF di $X_r$ a quella di $X_r'$ è che le masse vengono traslate all'indietro di $r$ unità, in modo che il supporto sia sempre costituito dagli <u>interi non negativi</u>.

Non possiamo esimerci dal verificare, per via analitica, che quelle che abbiamo scritto sono proprio delle PMF, ad esempio, prendendo l'espressione alla fine di pagina 4

$$\sum_{k=0}^{\infty} \binom{r+k-1}{r-1} q^k = p^{-r} = (1-q)^{-r}$$

Questo discende dal fatto che la serie al membro di sinistra è una serie convergente per $q < 1$, detta <u>serie binomiale</u>, che potete memorizzare nel modo seguente:

Per $|x|<1$, vale lo sviluppo in serie

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k$$

Sembra la formula del binomio di Newton, ma $\alpha \in \mathbb{R}$ qualsiasi, nel qual caso

$$\binom{\alpha}{k} = \frac{\alpha(\alpha-1)\dots(\alpha-k+1)}{k!}$$

Ora se $\alpha = -r$

$$\binom{-r}{k} = \frac{-r(-r-1)\dots(-r-k+1)}{k!} = (-1)^k \binom{r+k-1}{k}$$

e quindi possiamo riscrivere quando $x=-q$

$$(1-q)^{-r} = \sum_{k=0}^{\infty} \binom{r+k-1}{k} (-1)^k (-q)^k = \sum_{k=0}^{\infty} \binom{r+k-1}{r-1} q^k$$

che è quello che volevamo dimostrare, e che spiega anche il nome "binomiale negativa" che viene data a questa PMF.

Si noti che, come per la binomiale, anche per la binomiale negativa la convoluzione della PMF Bineg$(r,p)$ con la Bineg$(s,p)$ dà la Bineg$(r+s,p)$ (ovvio, con l'interpretazione del numero delle prove per ottenere un certo numero di successi).

Possiamo allora rispondere alla seguente domanda: Sapendo che $X_{r+s} = l$ (l'$(r+s)$-esimo successo alla $l$-esima prova) qual'è la PMF di $X_r$? (la stessa questione per la binomiale dava l'ipergeom.)

$$\mathbb{P}(X_r = k \mid X_{r+s} = l) = \frac{\mathbb{P}(X_r = k)\,\mathbb{P}(X_{r+s} - X_r = l-k)}{\mathbb{P}(X_{r+s} = l)} =$$

$$= \frac{\cancel{p^r}\,\cancel{q^{k-r}}\binom{k-1}{r-1}\cancel{p^s}\,\cancel{q^{l-k-s}}\binom{l-k-1}{s-1}}{\cancel{p^{r+s}}\,\cancel{q^{l-r-s}}\binom{l-1}{r+s-1}} =$$

$$= \frac{\binom{k-1}{r-1}\binom{l-k-1}{s-1}}{\binom{l-1}{r+s-1}}$$

non dipende da $p$! $k = r, \dots, l-s$ (altrimenti i coefficienti binomiali a numeratore sono nulli)

Da dove esce fuori questa distribuzione? Vi invito a pensarci per vostro conto, anche se nella prossima pagina vi do la soluzione, quindi fermate l'ascolto a questo punto, saltate pagina, e andate a pag. 9 (prossimo spezzone audio).

Si consideri l'insieme degli interi da 1 a $l-1$, da dove estraiamo a caso un sottoinsieme di $r+s-1$ elementi, che possiamo ordinare in ordine crescente

$$1 \leq X_{(1)} < X_{(2)} < \dots < X_{(r+s-1)} \leq l-1.$$

Qual è la probabilità che $X_{(r)} = k$?

Tra gli $\binom{l-1}{r+s-1}$ sottoinsiemi possibili quelli che realizzano l'evento si ottengono moltiplicando il numero dei modi di scegliere $r-1$ elementi da $k-1$ per il numero dei modi di scegliere i rimanenti

$s-1 = r+s-1-(r-1)-1$ = dei

$l-k-1 = l-1-(k-1)-1$, quindi

$$\mathbb{P}\left(X_r = k \mid X_{r+s} = l\right) = \mathbb{P}\left(X_{(r)} = k\right)$$

la $r$-esima STATISTICA D' ORDINE tra gli estratti $X_{(1)}, X_{(2)}, \dots, X_{(r+s-1)}$

A questo punto non so resistere alla tentazione di introdurre subito la PMF più importante che ci resta da presentare. Per farlo, trovo istruttivo derivarla come limite della PMF binomiale: si manda il numero delle prove all'infinito e si mantiene costante la media $np = \lambda > 0$, prendendo quindi $p = \frac{\lambda}{n}$. Teniamo $k$ fisso.

$$\binom{n}{k} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!} \, \frac{n(n-1)\cdots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n} \left(1 - \frac{\lambda}{n}\right)^{-k}$$

Ora il fattore $\lambda^k/k!$ non dipende da $n$,

$$\frac{n(n-1)\cdots(n-k+1)}{n \; n \; \cdots \; n} = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{k-1}{n}\right) \longrightarrow 1$$

Il termine $\left(1 - \frac{\lambda}{n}\right)^{-k}$ tende a $1$ (non si tratta di una forma indeterminata perché $k$ è fisso), mentre dovrebbe essere noto il limite notevole $\left(1 + \frac{x}{n}\right)^n \longrightarrow e^x$ qualunque sia $x$, per $n \to \infty$. Quindi se $X_n \sim Bin\left(n, \frac{\lambda}{n}\right)$

$$\lim_{n \to \infty} \mathbb{P}(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad k = 0, 1, 2, \ldots$$

l'espressione al membro di destra è la PMF di Poisson, con media $\lambda$. Un calcolo appena più complicato mostra che lo stesso risultato vale con $p = p_n = \frac{\lambda_n}{n}$, con $\lim_{n \to \infty} \lambda_n = \lambda$.

Naturalmente questo precorre i tempi, dato che non abbiamo garanzia che il limite di una PMF sia una PMF né che la sua media sia effettivamente $\lambda$. (questa necessità del passaggio del limite all'interno di una somma infinita). La verifica non è tuttavia problematica

$$\sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} \quad e^{-\lambda}e^{\lambda} = 1, \qquad \forall\lambda$$

$$\sum_{k=1}^{\infty} k\, e^{-\lambda}\frac{\lambda^k}{k!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!} = \lambda e^{-\lambda}e^{\lambda} = \lambda$$

È anche facile vedere che l'andamento, in funzione dell'intero $k$, della PMF di Poisson, ricalca quello della binomiale

$$e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} < e^{-\lambda}\frac{\lambda^k}{k!} \iff k<\lambda$$

per cui il massimo della PMF si ottiene quando $k=\lfloor\lambda\rfloor$ (quando $\lambda$ è intero gli interi $\lambda$ e $\lambda-1$ hanno uguale massa).

C'è un'altra proprietà importante che la PMF di Poisson eredita dalla binomiale, la chiusura rispetto alla convoluzione. In altre parole, se $X \sim \mathbb{P}_0(\lambda)$ e $Y \sim \mathbb{P}_0(\mu)$, e $X$ e $Y$ sono indipendenti, allora

$$\mathbb{P}(X+Y=\ell) = \sum_{k=0}^{\ell} \mathbb{P}(X=k)\,\mathbb{P}(Y=\ell-k) =$$

$$= \sum_{k=0}^{\ell} e^{-\lambda}\frac{\lambda^k}{k!}\, e^{-\mu}\frac{\mu^{\ell-k}}{\ell-k!} = \frac{e^{-\lambda-\mu}}{\ell!} \sum_{k=0}^{\ell} \binom{\ell}{k} \lambda^k \mu^{\ell-k} =$$

$$= \frac{e^{-\lambda-\mu}}{\ell!} (\lambda+\mu)^{\ell}, \quad \ell = 0, 1, 2, \ldots$$

---

Questo ci permette anche di determinare la PMF di $X$ condizionata a $X+Y$.

$$\mathbb{P}(X=k \mid X+Y=\ell) = \frac{\mathbb{P}(X=k)\,\mathbb{P}(Y=\ell-k)}{\mathbb{P}(X+Y=\ell)} =$$

$$= \frac{\cancel{e^{-\lambda}}\frac{\lambda^k}{k!}\,\cancel{e^{-\mu}}\frac{\mu^{\ell-k}}{(\ell-k)!}}{\cancel{e^{-\lambda-\mu}}\frac{(\lambda+\mu)^{\ell}}{\ell!}} = \binom{\ell}{k}\left(\frac{\lambda}{\lambda+\mu}\right)^{k}\left(\frac{\mu}{\lambda+\mu}\right)^{\ell-k}$$

$$k = 0, 1, \ldots, \ell$$

Quindi $X \mid X+Y=\ell \sim \mathrm{Bin}\left(\ell, \frac{\lambda}{\lambda+\mu}\right)$.

Questo ci permette di costruire variabili aleatorie $X' \sim Po(\lambda)$ e $Y' \sim Po(\mu)$ indipendenti in questo modo: se prendiamo $Z (= X'+Y')$ una variabile di Poisson di media $\nu = \lambda+\mu$ e

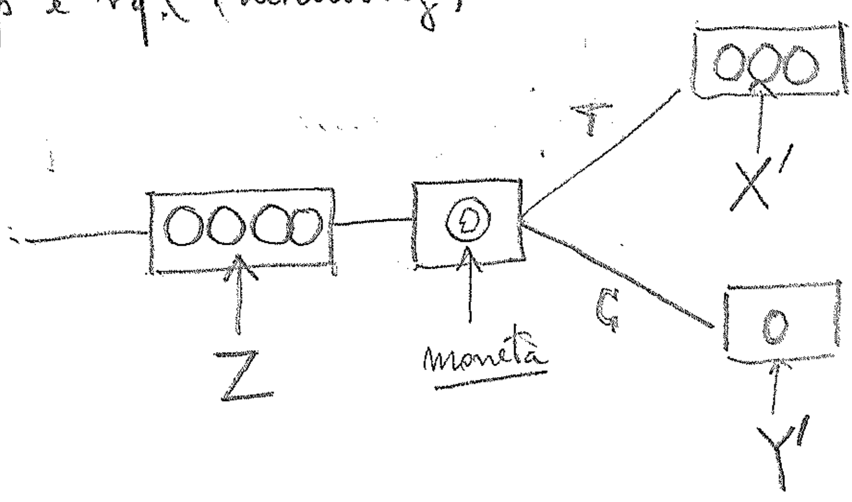
$$X' \mid Z=\ell \sim \mathrm{Bin}(\ell, p), \quad \text{con } p = \frac{\lambda}{\lambda+\mu}$$

allora $X'$ e $Y' = Z - X'$ sono INDIPENDENTI e

$$X' \sim Po(\lambda), \quad Y' \sim Po(\mu), \quad q = 1-p$$

(nota che $\lambda = \nu p$, $\mu = \nu q$).

Ancora più esplicitamente se ciascuno degli accadimenti "registrati" da una variabile di Poisson $Z$ di media $\nu$ lanciamo una moneta con probabilità $p$ di testa e attribuiamo l'accadimento a $X'$ se esce testa e a $Y'$ se esce croce, INDIPENDENTEMENTE per ciascuno degli accadimenti, otteniamo variabili di Poisson $X'$ e $Y'$ INDIPENDENTI, di media $\nu p$ e $\nu q$ (thinning)



Attenzione: qualunque sia il numero intero $l$, le variabili aleatorie $X$ e $l-X$ <u>non possono</u> essere indipendenti (tranne nel caso che $X$ è costante), ma la situazione cambia se $l$ è sostituita da una VARIABILE ALEATORIA.

"Da un punto di vista modellistico il fatto che la legge di Poisson($\lambda$) si ottiene come limite di una opportuna successione di binomiali legittima il suo uso per il conteggio di accadimenti nel tempo e/o nello spazio, quando la probabilità di vederne prodursi uno di essi in una porzione microscopica di tempo e/o spazio è infinitesima, indipendentemente per ciascuna porzione in cui pensiamo diviso il tempo e/o lo spazio. In questi casi è comodo riferire il parametro $\lambda$ ad un intervallo di tempo unitario e regione dello spazio di volume unitario, che va poi moltiplicato per la lunghezza dell'intervallo di osservazione cui siamo interessati (analogamente per lo spazio).

In questo modo, se diciamo che in media ci arrivano 3 messaggi di posta elettronica all'ora durante l'orario di lavoro, possiamo assumere che il numero di messaggi che ci arrivano in una giornata lavorativa di 8 ore abbia legge Po ($8 \times 3 = 24$)

Riflettete: la probabilità di ricevere un messaggio diciamo in un minuto è p, molto piccola, ma in 8 ore ci sono $60 \times 8 = 480$ minuti. Se per ciascuno di essi ho una prova di Bernoulli indipendenti, allora il numero di messaggi in una giornata è Bin (480, p): se voglio riprodurre la media di 24, devo scegliere $p = \frac{24}{480} = \frac{1}{20}$, ma allora tanto vale approssimare questa distribuzione con Po (24). (d'altra parte questa "discretizzazione" del problema su ciascun minuto sembra piuttosto arbitraria...). Considerazioni analoghe valgono per fenomeni che si svolgono nello spazio, come i difetti in una lamina prodotta in un'acciaieria.

Più in generale si può approssimare con una distribuzione di Poisson la somma di variabili Bernoulliane indipendenti (o addirittura "poco" dipendenti tra loro) con probabilità di successo piccole (rispetto alla media), anche se questa non è costante.

Serve più altro strumento per potere chiudere la parte del corso relativa ai modelli discreti e passare a quelli continui. Abbiamo visto che la media (se esiste) è un indice di "dove" si trova una PMF sull'asse reale, in particolare coincide con il centro di simmetria nel caso di PMF simmetrica. Ma naturalmente ci sono tante PMF con la stessa media (a cominciare da quella che assegna a questa una massa unitaria) ed è opportuno aver un altro indice per distinguere tra queste.

Questo indice è la varianza, che misura la dispersione intorno alla media. Come posso fare questo: ci sono buoni motivi per misurare quanto la variabile $X$ è distante dalla media $m = \mathbb{E}(X)$ mediante $(X-m)^2$. Naturalmente questa è una variabile aleatoria: prendiamone quindi la media

$$\mathbb{E}(X-m)^2 = \mathrm{var}(X)$$

è la VARIANZA della PMF (distribuzione di $X$).

Naturalmente l'esistenza della varianza presuppone quella della media, non solo: anche se la media esiste, la varianza può essere infinita. Esempio

$$\mathbb{P}(X = \pm k) = \frac{1}{2}\left(\frac{1}{k^2} - \frac{1}{(k+1)^2}\right) = \frac{2k+1}{2k^2(k+1)^2}, \quad k = 1, 2, \dots$$

è una PMF incontrata in precedenza, che ha media nulla, quindi (raggruppando le masse su $k$ e $-k$ e usando LOTUS)

$$\mathrm{var}(X) = \mathbb{E}(X^2) = \sum_{k=1}^{\infty} k^2 \frac{(2k+1)}{k^2(k+1)^2} = \sum_{k=1}^{\infty} \frac{2k+1}{(k+1)^2} = +\infty$$

dato che $\frac{2k+1}{(k+1)^2} = O\left(\frac{1}{k}\right)$ $\left(\text{visto che } k\frac{(2k+1)}{(k+1)^2} \to 2\right)$

e quindi la serie di interesse si maggiora, da un certo $k$ in poi, con un multiplo della serie armonica.

Supponiamo ora che la varianza (e la media) di $X$ siano finite: allora dato che $\mathbb{E}(aX+b) = a\mathbb{E}(X) + b$

$$\mathrm{var}(aX+b) = \mathbb{E}\left(aX + \cancel{b} - a\mathbb{E}X - \cancel{b}\right)^2 = a^2 \mathrm{var}(X)$$

quindi la varianza non dipende dall'origine del sistema di riferimento in cui è misurata $X$, dipende invece dall'unità di misura in cui è misurata $X$, al quadrato.

Per questo in pratica è conveniente (ad esempio per confrontarla con la media) esprimerla nella stessa unità di misura di X prendendone la radice quadrata e ottenendo la deviazione standard.

Lasciamo al seguito far vedere che la varianza controlla gli scarti grandi dalla media e per ora osserviamo solo che la varianza di X è nulla se e solo se X è costante (a meno di esiti $s \in S$ di probabilità nulla).

Infatti, se $X = c$ allora $\mathbb{E}(X) = c$, quindi $(X - \mathbb{E}X)^2$ è identicamente nulla ed ha quindi media nulla.

Viceversa, se $var(X) = \mathbb{E}(X - \mathbb{E}X)^2 = 0$ allora $(X - \mathbb{E}X)^2$ è una variabile non negativa di media nulla, quindi non possono esistere esiti $s \in S$ con massa positiva dove $X \neq \mathbb{E}(X)$.

Per calcolare la varianza, spesso non è conveniente il calcolo diretto: si pensi che X prende magari solo valori interi mentre $m$ può essere anche irrazionale: lo stesso quindi avviene per $(X - \mathbb{E}X)^2$ rendendo difficile calcolare la media di questa nuova variabile.

Continuiamo a indicare $\mathbb{E}(X)$ con $m$. Allora

$$(X-m)^2 = X^2 - 2Xm + m^2$$

Prendendo il valore atteso a destra e a sinistra, abbiamo

$$\mathrm{var}(X) = \mathbb{E}(X-m)^2 = \mathbb{E}(X^2) - 2m\,\mathbb{E}(X) + m^2 = \mathbb{E}(X^2) - 2m^2 + m^2$$
$$= \mathbb{E}(X^2) - (\mathbb{E}X)^2$$

una formula talmente comoda che gli studenti sono spesso portati a confonderla con la definizione.

<u>Avvertenza 1.</u> Ricordate che $\mathbb{E}(g(X)) \geqslant g(\mathbb{E}X)$ quando $g$ è convessa, come è certamente la funzione $g(x) = x^2$.

<u>Avvertenza 2.</u> Come si vede $\mathrm{var}(X) < +\infty$ se e solo se $\mathbb{E}(X^2) < +\infty$. Ma avevamo presupposto che esistesse la media $\mathbb{E}(X) < +\infty$ (ricordate che equivale a $\mathbb{E}(|X|) < +\infty$). Ora si vede facilmente per il teorema di confronto tra serie che $\mathbb{E}(X^2) < +\infty$ implica che $\mathbb{E}(|X|) < +\infty$. Abbiamo stabilito quindi

$$\mathrm{var}(X) < +\infty \iff \mathbb{E}(X^2) < +\infty.$$

Calcoli di varianze di distribuzioni "importanti"

- $X$ Bernoulli $(p)$.

Dato che $X$ assume solo valore 0 o 1, $X^2 = X$

e quindi $\mathbb{E}(X^2) = \mathbb{E}(X) = p$, da cui

$$\operatorname{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \mathbb{E}(X) - (\mathbb{E}X)^2 = p - p^2 = p(1-p)$$

Alternativamente

$$\operatorname{var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}(X-p)^2$$

La variabile $(X-p)^2$ vale $p^2$ quando $X=0$ (che ha probabilità $q$) e $(1-p)^2 = q^2$ quando $X=1$ (che ha prob. $p$)

per cui $\operatorname{var}(X) = p^2 q + q^2 p = pq(p+q) = pq = p(1-p)$

- $X$ uniforme su $\{1, \dots, N\}$, $\mathbb{E}(X) = \frac{N+1}{2}$

$$\mathbb{E}(X^2) = \frac{1}{N} \sum_{i=1}^{N} i^2 = \frac{N(N+1)(2N+1)}{6N} = \frac{(N+1)(2N+1)}{6}$$

quindi $\operatorname{var}(X) = \frac{(N+1)(2N+1)}{6} - \frac{(N+1)^2}{4} = \frac{N+1}{12}\left(2(2N+1) - 3(N+1)\right)$

$= \frac{N^2-1}{12}$. Quando $N=6$, $\operatorname{var}(X) = \frac{35}{12}$, $\operatorname{stdev}(X) \cong 1.71$

- $X$ Poisson $\mathbb{E}(X)=\lambda$

Per calcolare $\mathbb{E}(X^2)$ conviene utilizzare l'identità

$$\mathbb{E}[X(X-1)] = \mathbb{E}(X^2) - \mathbb{E}(X) \Rightarrow \mathbb{E}(X^2) = \mathbb{E}[X(X-1)] + \mathbb{E}(X).$$

$$\mathbb{E}[X(X-1)] = \sum_{k=2}^{\infty} k(k-1)\, e^{-\lambda} \frac{\lambda^k}{k!} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2$$

e quindi $\mathbb{E}(X^2) = \lambda^2 + \lambda \Rightarrow \mathrm{var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$

La Poisson ha media e varianza uguali (e questo fornisce un rapido test per l'adattamento di dati sperimentali a questa distribuzione), ma attenzione

$$\mathrm{st\,dev}(X) = \sqrt{\mathbb{E}(X)}.$$

- $X$ geometrica $(p)$

Adottiamo il solito metodo del condizionamento al primo passo per calcolare $\mathbb{E}(X^2)$.

$$\mathbb{E}(X^2) = 1\cdot p + \mathbb{E}(X^2 | X>1)\cdot q$$

$$X^2 = (X-1)^2 + 2X - 1 = (X-1)^2 + 2(X-1) + 1$$

Dato che $X-1 | X>1 \approx \mathrm{Geom}(p)$

$\mathbb{E}(X^2|X>1) = \mathbb{E}(X^2) + 2\cdot\frac{1}{p} + 1$, quindi

$$\mathbb{E}(X^2) = 1 + 2\frac{q}{p} + q\,\mathbb{E}(X^2) \Rightarrow \mathbb{E}(X^2) = \frac{1}{p} + 2\frac{q}{p^2} = \frac{1}{p^2} + \frac{q}{p^2}$$

$$\mathrm{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \frac{q}{p^2}$$

Per calcolare la varianza di binomiale e binomiale negativa anticipiamo un risultato fondamentale

Se $X_1, \dots, X_n$ sono variabili <u>indipendenti</u> con varianza finita, allora la varianza è additiva, cioè

$$var(X_1 + \dots + X_n) = var(X_1) + \dots + var(X_n)$$

Se $X_1 \dots X_n$ sono identicamente distribuite, allora

$$var(X_1 + \dots + X_n) = n\, var(X_1) \quad \left(\text{ma } stdev(X_1 + \dots + X_n) = \sqrt{n}\, stdev(X_1)\right)$$

- $X$ Binomiale $(n, p)$

  $X$ Somma di $n$ Bernoulliane indipendenti $p$.

  $$var(X) = np(1-p)$$

- $X$ Binomiale negativa $(r, p)$

  $X$ Somma di $r$ Geometriche indipendenti $p$

  $$var(X) = \frac{rq}{p^2}$$

- Per l'ipergeometrica $(M, K, n)$ avremo bisogno di una formula più generale per la varianza di una somma con addendi non indipendenti.
  Diamo tuttavia la formula

$$var(Hypg(M, K, n)) = var\left(Bin\left(n, \frac{K}{M}\right)\right) \cdot \frac{M-n}{M-1}$$