Calcolo delle Probabilità,
corso di Laurea in INFORMATICA, canale 1 (A-L), Prof. Nappo, a.a. 2024-25
SCHEMA SULLE VARIABILI ALEATORIE

ATTENZIONE: SI TRATTA SOLO DI UNO SCHEMA!!
IN PARTICOLARE
PER QUANTO RIGUARDA LE VARIABILI ALEATORIE DISCRETE SU SPAZI DI PROBABILITÀ FINITI MOLTE DELLE DIMOSTRAZIONI SONO SUGLI APPUNTI Spizzichino Nappo [SN]

PER LE VARIABILI ALEATORIE DISCRETE, MA CHE ASSUMO UN'INFINITA' NUMERABILE DI VALORI, C'È BUONA PARTE DEL PROGRAMMA CHE DOVREMO SVOLGERE;

PER LE VARIABILI ASSOLUTAMENTE CONTINUE, OSSIA CON DENSITA', FORSE C'È ANCHE QUALCOSA IN PIÙ
GUARDARE I CHIARIMENTI NEL DIARIO DELLE LEZIONI e NEL PROGRAMMA

Indice

1 Richiami generali per variabili aleatorie in spazi di probabilità finiti

1.1 Introduzione alle variabili aleatorie

Nella prima parte del corso ci siamo ripetutamente imbattuti in oggetti quali: somma dei punteggi nel lancio di due dadi, numero di votanti per uno schieramento in un sondaggio elettorale, numero di successi su n prove bernoulliane, massimo fra i cinque numeri risultanti da un'estrazione del lotto, etc....
Si è trattato in ogni caso di situazioni in cui viene considerata una grandezza aleatoria X e valgono le seguenti condizioni:
- **il valore che X assumerà sarà connesso (in qualche preciso modo) al risultato elementare di un qualche esperimento aleatorio;**
- **possiamo elencare i valori che possono essere assunti da X;**
- **sussiste una situazione di incertezza relativamente allo specifico valore che X effettivamente assume**.
- **In base alla misura di probabilità assegnata sullo spazio campione in tale esperimento, potremo valutare la probabilità che si presentino i vari possibili valori per la grandezza X.**

ESEMPIO: LANCIO DI DUE DADI A 6 FACCE:
X_1 rappresenta il valore/punteggio del primo dado, X_2 il valore/punteggio del secondo dado, $X = X_1 + X_2$ la somma dei punteggi ottenuti
QUI $\Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\} = \big\{(i,j),\ i,j \in \{1,2,3,4,5,6\}\big\}$
e se esce il risultato (i,j) allora X_1 assume il valore i, X_2 assume il valore j ed $X = X_1 + X_2$ assume il valore $i+j$.
Dal punto di vista matematico si tratta quindi di funzioni:

$$(i,j) \mapsto X_1(i,j) = i$$
$$(i,j) \mapsto X_2(i,j) = j$$
$$(i,j) \mapsto X(i,j) = i+j$$

ESEMPIO: LANCIO DI DUE DADI A 6 FACCE:
L'insieme dei valori che può assumere X_1 è l'insieme $\{1,2,3,4,5,6\}$, lo stesso vale per X_2, invece l'insieme dei valori che può assumere $X = X_1 + X_2$ è l'insieme $\{2,3,4,5,6,7,8,9,10,11,12\}$.
In altre parole $X_1(\Omega) = \{1,2,3,4,5,6\}$, dove $X_1(\Omega)$ denota l'immagine di X_1, anche $X_2(\Omega) = \{1,2,3,4,5,6\}$, e invece $X(\Omega) = \{2,3,4,5,6,7,8,9,10,11,12\}$, e potremmo anche considerare, sempre con $\Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$,

$$\begin{aligned} X_1 :\ \Omega &\to X_1(\Omega) = \{1,2,3,4,5,6\} \subset \mathbb{R} \\ (i,j) &\mapsto X_1(i,j) = i \\ X_2 :\ \Omega &\to X_2(\Omega) = \{1,2,3,4,5,6\} \subset \mathbb{R} \\ (i,j) &\mapsto X_2(i,j) = j \\ X :\ \Omega &\to X(\Omega) = \{2,3,4,5,6,7,8,9,10,11,12\} \subset \mathbb{R} \\ (i,j) &\mapsto X(i,j) = i+j \end{aligned}$$

ESEMPIO: LANCIO DI DUE DADI A 6 FACCE: Se i dadi sono ben equilibrati e prendiamo come misura di probabilità $\mathbb{P}_c$, la probabilità classica, e valutiamo quindi $\mathbb{P}_c(X_1 = k) = \frac{1}{6}$, per ogni $k \in \{1,2,3,4,5,6\}$, similmente $\mathbb{P}_c(X_2 = k) = \frac{1}{6}$, per ogni $k \in \{1,2,3,4,5,6\}$, e, tenendo conto che

$$\{X=2\} = \{(1,1)\}, \quad \{X=3\} = \{(1,2),(2,1)\}, \quad \{X=4\} = \{(1,3),(2,2),(3,1)\},$$
$$\{X=5\} = \{(1,4),(2,3),(3,2),(4,1)\}, \quad \{X=6\} = \{(1,5),(2,4),(3,3),(4,2),(5,1)\},$$
$$\{X=7\} = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\},$$
$$\{X=8\} = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}, \quad \{X=9\} = \{(3,6),(4,5),(5,4),(6,3)\},$$
$$\{X=10\} = \{(4,6),(5,5),(6,4)\}, \quad \{X=11\} = \{(5,6),(6,5)\}, \quad \{X=12\} = \{(6,6)\},$$

otteniamo

$$\mathbb{P}_c(X=2) = \frac{1}{36}, \mathbb{P}_c(X=3) = \frac{2}{36}, \mathbb{P}_c(X=4) = \frac{3}{36},$$
$$\mathbb{P}_c(X=5) = \frac{4}{36}, \mathbb{P}_c(X=6) = \frac{5}{36}, \mathbb{P}_c(X=7) = \frac{6}{36},$$
$$\mathbb{P}_c(X=8) = \frac{5}{36}, \mathbb{P}_c(X=9) = \frac{4}{36}, \mathbb{P}_c(X=10) = \frac{3}{36},$$
$$\mathbb{P}_c(X=11) = \frac{2}{36}, \mathbb{P}_c(X=12) = \frac{1}{36},$$

Se invece i due dadi fossero truccati e si supponesse che la probabilità che il risultato (i,j) abbia probabilità proporzionale a $i \cdot j$, ossia si usasse la probabilità $\mathbb{Q}$ tale che $\mathbb{Q}(\{(i,j)\}) = \frac{i \cdot j}{21^2}$ si avrebbe, come visto in precedenza $\mathbb{Q}(X_1 = k) = \frac{k}{21}$, per $k \in \{1,2,3,4,5,6\}$, similmente $\mathbb{Q}(X_2 = k) = \frac{k}{21}$, per $k \in \{1,2,3,4,5,6\}$, e, sempre tenendo conto che

$$\{X=2\} = \{(1,1)\}, \quad \{X=3\} = \{(1,2),(2,1)\}, \quad \{X=4\} = \{(1,3),(2,2),(3,1)\},$$
$$\{X=5\} = \{(1,4),(2,3),(3,2),(4,1)\}, \quad \{X=6\} = \{(1,5),(2,4),(3,3),(4,2),(5,1)\},$$
$$\{X=7\} = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\},$$
$$\{X=8\} = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}, \quad \{X=9\} = \{(3,6),(4,5),(5,4),(6,3)\},$$
$$\{X=10\} = \{(4,6),(5,5),(6,4)\}, \quad \{X=11\} = \{(5,6),(6,5)\}, \quad \{X=12\} = \{(6,6)\},$$

otteniamo

$$\mathbb{Q}(X=2) = \frac{1 \cdot 1}{21^2} = \frac{1}{21^2},$$
$$\mathbb{Q}(X=3) = \frac{1 \cdot 2 + 2 \cdot 1}{21^2} = \frac{4}{21^2},$$
$$\mathbb{Q}(X=4) = \frac{1 \cdot 3 + 2 \cdot 2 + 3 \cdot 1}{21^2} = \frac{10}{21^2},$$
$$\mathbb{Q}(X=5) = \frac{1 \cdot 4 + 2 \cdot 3 + 3 \cdot 2 + 4 \cdot 1}{21^2} = \frac{20}{21^2},$$
$$\mathbb{Q}(X=6) = \frac{1 \cdot 5 + 2 \cdot 4 + 3 \cdot 3 + 4 \cdot 2 + 5 \cdot 1}{21^2} = \frac{35}{21^2},$$
$$\mathbb{Q}(X=7) = \frac{1 \cdot 6 + 2 \cdot 5 + 3 \cdot 4 + 4 \cdot 3 + 5 \cdot 2 + 6 \cdot 1}{21^2} = \frac{56}{21^2},$$
$$\mathbb{Q}(X=8) = \frac{2 \cdot 6 + 3 \cdot 5 + 4 \cdot 4 + 5 \cdot 3 + 6 \cdot 2}{21^2} = \frac{70}{21^2},$$
$$\mathbb{Q}(X=9) = \frac{3 \cdot 6 + 4 \cdot 5 + 5 \cdot 4 + 6 \cdot 3}{21^2} = \frac{76}{21^2},$$
$$\mathbb{Q}(X=10) = \frac{4 \cdot 6 + 5 \cdot 5 + 6 \cdot 4}{21^2} = \frac{73}{21^2},$$
$$\mathbb{Q}(X=11) = \frac{5 \cdot 6 + 6 \cdot 5}{21^2} = \frac{60}{21^2},$$
$$\mathbb{Q}(X=12) = \frac{6 \cdot 6}{21^2} = \frac{36}{21^2}$$

Osserviamo che la famiglia degli eventi $H_k^X := \{X = k\}$, per $k \in X(\Omega) = \{2,3,4,5,6,7,8,9,10,11,12\}$ è una partizione, che chiameremo la **partizione generata da** X, e quindi, sia per $\mathbb{P} = \mathbb{P}_c$ che per $\mathbb{P} = \mathbb{Q}$ si ha

$$\mathbb{P}(H_2^X) + \mathbb{P}(H_3^X) + \cdots + \mathbb{P}(H_{11}^X) + \mathbb{P}(H_{12}^X) = \sum_{k=2}^{12} \mathbb{P}(H_k^X) = 1$$

ovvero $\sum_{k=2}^{12} \mathbb{P}(\{X = k\}) = 1$. Quindi la funzione definita da:

$$p_X : X(\Omega) = \{2,3,4,5,6,7,8,9,10,11,12\} \to [0,1]; \quad k \mapsto p_X(k) := \mathbb{P}(\{X = k\})$$

definisce, nel solito modo, una probabilità sull'insieme delle parti di $X(\Omega) = \{2,3,4,5,6,7,8,9,10,11,12\}$:

$$\mathbf{P}_X : \mathcal{P}\big(X(\Omega)\big) = \mathcal{P}\big(\{2,3,4,5,6,7,8,9,10,11,12\}\big) \to [0,1];$$

$$I \mapsto \mathbf{P}_X(I) := \sum_{k \in I} p_X(k) \Big(= \sum_{k \in I} \mathbb{P}(\{X = k\}) \Big)$$

Tale probabilità prende il nome di **distribuzione (di probabilità) di** X e invece la funzione $k \mapsto p_X(k)$, viene detta **densità discreta di** X.
Osserviamo infine che $\mathbf{P}_X(I)$ coincide con $\mathbb{P}(X \in I)$, infatti

$$\mathbf{P}_X(I) := \sum_{k \in I} p_X(k) = \sum_{k \in I} \mathbb{P}(\{X = k\}) = \mathbb{P}\big(\cup_{k \in I} \{X = k\}\big) = \mathbb{P}(X \in I)$$

Ad esempio, se $I = \{2,3,4,5\}$ allora possiamo scrivere l'evento $\{X \in I\}$

$$\big\{X \in \{2,3,4,5\}\big\} = \{X = 2\} \cup \{X = 3\} \cup \{X = 4\} \cup \{X = 5\}$$

e quindi

$$\begin{aligned} \mathbb{P}(X \in \{2,3,4,5\}) &= \mathbb{P}(\{X = 2\} \cup \{X = 3\} \cup \{X = 4\} \cup \{X = 5\}) \\ &= \mathbb{P}(\{X = 2\}) + \mathbb{P}(\{X = 3\}) + \mathbb{P}(\{X = 4\}) + \mathbb{P}(\{X = 5\}) \\ &= p_X(2) + p_X(3) + p_X(4) + p_X(5). \end{aligned}$$

Nel caso in cui $X(\Omega)$ abbia cardinalità grande, non è possibile elencare tutti i valori di $p(x_k)$, per ogni $x_k \in X(\Omega)$: ad esempio nel caso precedente in cui X è uguale alla somma ottenuta lanciando due dadi ben equilibrati abbiamo trovato che

$$p_X(2) = 1/36,\ p_X(3) = 2/36,\ p_X(3) = 4/36,\ p_X(4) = 3/36,$$

$$p_X(5) = 4/36,\ p_X(6) = 5/36,\ p_X(7) = 6/36,$$

$$p_X(8) = 5/36,\ p_X(9) = 4/36,\ p_X(10) = 3/36,\ p_X(11) = 2/36,\ p_X(12) = 1/36,$$

e, per $k = 2,3,4,5,6,7$, è immediato osservare che il numeratore di $p_X(k)$ cresce di volta in volta di 1 e coincide con $k-1$, mentre, per $k = 8,9,10,11,12$, il numeratore decresce di volta in volta di 1 e quindi potremo sicuramente scriverlo come $a-1$, per un opportuno intero a: non è difficile convincersi che $a = 13$, osservando ad esempio che $8+5 = 9+4 = 10+3 = 11+2 = 12+1 = 13$. Potremo quindi scrivere sinteticamente

$$p_X(k) = \begin{cases} \dfrac{k-1}{36} & \text{per } k = 2,3,...,7 \\ \dfrac{13-k}{36} & \text{per } k = 8,9,...,12 \end{cases}$$

NOTA BENE: è chiaro che, per $k = 2,3,...,7$ *si ha che* $\mathbb{P}(X = k)$ *aumenta di* $1/36$ *e quindi* $\mathbb{P}(X = k) = \frac{k+c_1}{36}$*, per qualche valore* c_1*, mentre per gli altri valori di* k *ogni volta diminuisce di* $1/36$ *e quindi, per* $k = 8,9,...,12$ $\mathbb{P}(X = k) = \frac{c_2-k}{36}$*. Per trovare* c_1 *e* c_2 *basta uguagliare, ad esempio,* $\frac{1}{36} = \mathbb{P}(X = 2) = \frac{1+c_1}{36}$ *(da cui* $2+c_1 = 1$*, ossia* $c_1 = -1$*) e* $\frac{1}{36} = \mathbb{P}(X = 12) = \frac{c_2-12}{36}$ *(da cui* $1 = c_2 - 12$ *ossia* $c_2 = 13$*).*

1.2 Definizione di variabile aleatoria

Sia $\Omega = \{\omega_1, ..., \omega_N\}$ un insieme finito che rappresenta l'insieme tutti i "casi" (eventi elementari) possibili.

Definizione 1.1. *Una variabile aleatoria X (a valori reali) è una funzione definita*[1] *su* Ω

$$X : \Omega \to X(\Omega) \subset \mathbb{R}; \qquad \omega \mapsto X(\omega)$$

NOTA BENE: Sarebbe forse più ragionevole chiamare la funzione X "numero aleatorio" invece di "variabile aleatoria", ma questo è l'uso in probabilità. Inoltre per denotare le variabili aleatorie si usano le lettere MAIUSCOLE e in STAMPATELLO, e di solito si usano le ultime lettere dell'alfabeto inglese (S, T, U, V, X, Y, W e Z), invece che le lettere f, g, h, (o anche φ, ψ) in quanto in probabilità vengono usate per denotare le funzioni definite su $\mathbb{R}$ o su $\mathbb{R}^d$ e a valori reali (ossia in $\mathbb{R}$) o a valori vettoriali (ossia in $\mathbb{R}^\ell$).
Ovviamente, dato che Ω è finito[2] anche $X(\Omega)$, ossia l'immagine di Ω tramite la funzione X, è un insieme finito e viene detto *l'insieme dei valori che può assumere la variabile aleatoria X*.

In genere si indicano con x_i gli elementi di $X(\Omega)$ e quindi si scrive $X(\Omega) = \{x_1, x_2, \cdots, x_n\}$ (dove chiaramente $n \leq N$).

Ad ogni variabili aleatoria su uno spazio finito si associa una partizione di Ω, detta la **partizione generata da** X:

$$\left\{H_1^X, H_2^X, \cdots, H_n^X\right\}, \quad \text{dove} \quad H_k^X := \{\omega \in \Omega : X(\omega) = x_k\}, \quad k = 1, 2, ..., n.$$

Più sinteticamente si scrive

$$H_k^X := \{X = x_k\}, \quad k = 1, 2, ..., n.$$

oppure, per mettere più in evidenza la dipendenza da x_k a volte si scrive

$$H_x^X = \{X = x\}, \quad \text{per } x \in X(\Omega) = \{x_1, x_2, ..., x_n\}.$$

Non è difficile convincersi che, a partire dalla partizione generata da X e da $X(\Omega) = \{x_1, x_2, \cdots, x_n\}$, è possibile riscrivere

$$X(\omega) = \sum_{k=1}^{n} x_k \mathbf{1}_{H_k^X}(\omega), \quad \text{dove } \mathbf{1}_{H_k^X}(\omega) = 1 \Leftrightarrow \omega \in H_k^X, \text{ ossia } X(\omega) = x_k, \quad \text{e} \quad \mathbf{1}_{H_k^X}(\omega) = 0, \text{ altrimenti.}$$

1.2.1 Densità discreta, distribuzione e funzione di distribuzione

Supponiamo ora che sia dato lo Spazio di Probabilità finito $\left(\Omega, \mathcal{P}(\Omega), \mathbb{P}\right)$.

Definizione 1.2. *La funzione* $p_X : X(\Omega) \to [0,1]; \quad x \in X(\Omega) \mapsto p_X(x) := \mathbb{P}(\{X = x\}) = \mathbb{P}(H_x^X)$ *è detta* ***densità discreta*** *di X, e quindi si ha*

$$\boxed{p_X(x) = \sum_{\omega \in \Omega : X(\omega) = x} p(\omega), \quad x \in X(\Omega)}$$

NOTA BENE: è chiaro che[3] $p_X(x_k) \geq 0$, per ogni $k = 1, 2, ..., n$, e

$$\sum_{k=1}^{n} p_X(x_k) = 1 \quad \text{in quanto} \quad \sum_{k=1}^{n} p_X(x_k) = \sum_{k=1}^{n} \mathbb{P}(\{X_{x_k}\}) = \sum_{k=1}^{n} \mathbb{P}(H_{x_k}^X) \quad \text{e } \left\{H_{x_1}^X, H_{x_2}^X, ..., H_{x_n}^X\right\} \text{ è una partizione}$$

e che per ogni sottoinsieme J di $X(\Omega)$

$$\sum_{x \in J} p_X(x) = \sum_{x \in J} \mathbb{P}(\{X = x\}) = \mathbb{P}\Big(\bigcup_{x \in J} \{X = x\}\Big) = \mathbb{P}(\{X \in J\})$$

[1] ATTENZIONE, ricordare che data una funzione $f : A \to B$; $a \mapsto f(a)$, è diversa dalla funzione $\widetilde{f} : A \to f(A)$; $a \mapsto f(a)$, dove $f(A) = \{b \in B : \exists a \in A, \text{ tale che } f(a) = b\}$ è l'immagine di A tramite f, in quanto le due funzioni pur avendo lo stesso dominio A, ed associando ad ogni $a \in A$ lo stesso elemento $b \in B$, hanno codominio (in genere) diverso. Tuttavia, per semplicità di notazione, non terremo conto di questa precisazione.

[2] Come vedremo è possibile considerare anche il caso in cui Ω non è finito e nemmeno numerabile, e in tale caso è possibile che anche $X(\Omega)$ non sia finito, e nemmeno numerabile.

[3] Ricordiamo che una partizione è una famiglia di eventi incompatibili ed esaustivi, ovvero gli insiemi che li rappresentano sono disgiunti a due a due e la loro unione è Ω, che rappresenta l'evento certo.
In questo caso: se $x_i \neq x_j$ allora $H_{x_i}^X := \{X = x_i\} \neq \{X = x_j\} =: H_{x_j}^X$ ossia $H_{x_i}^X$ e $H_{x_j}^X$ non si possono verificare contemporaneamente e almeno uno tra $H_{x_i}^X$ si deve verificare. In altre parole se ne verifica uno e uno solo tra $H_{x_1}^X$, $H_{x_2}^X$, ..., $H_{x_n}^X$.

NOTA BENE: per semplicità di notazione si scrive $\mathbb{P}(X=x)$ invece di $\mathbb{P}(\{X=x\})$ e analogamente $\mathbb{P}(X\in J)$ invece di $\mathbb{P}(\{X\in J\})$.
Il fatto che $p_X(x_k)\geq 0$ e $\sum_{k=1}^n p_X(x_k)=1$, ci permette di affermare che

$$\mathbf{P}_X:\ \mathcal{P}\big(X(\Omega)\big)\to[0,1];\quad J\subseteq X(\Omega)\mapsto \mathbf{P}_X(J):=\sum_{x\in J}p_X(x)=\mathbb{P}(X\in J)$$

definisce una probabilità su $\big(X(\Omega),\mathcal{P}\big(X(\Omega)\big)\big)$. La probabilità $\mathbf{P}_X$ è detta **distribuzione della variabile aleatoria** X.
Chiaramente la distribuzione di X, ossia la misura di probabilità $\mathbf{P}_X$ è univocamente individuata dalla densità discreta p_X, ossia dall'insieme $X(\Omega)$ dei valori che può assumere X e da $p_X(x)=\mathbb{P}(X=x)$, per $x\in X(\Omega)$, esattamente come la probabilità $\mathbb{P}$ è univocamente individuata dalla densità $\omega_i\mapsto p(\omega_i)$ (anche detta funzione di massa) tramite la "regola" $\mathbb{P}(A)=\sum_{\omega\in A}p(\omega)$, per ogni $A\subseteq\Omega$.

Come vedremo meglio per ogni variabile aleatoria X (anche a valori infiniti numerabili o anche infiniti non numerabili) si può definire la **funzione di distribuzione**

$$F_X:\ \mathbb{R}\to[0,1];\quad x\mapsto F_X(x):=\mathbb{P}(X\leq x)$$

Nel caso delle variabili aleatorie definite su uno spazio di probabilità finito, se supponiamo di aver numerato gli elementi di $X(\Omega)$ in ordine crescente, ossia se supponiamo che $x_1<x_2<\cdots<x_n$ si ha che

$$\{X\leq x\}=\bigcup_{k:x_k\leq x}\{X=x_k\},\quad \text{e quindi}\quad \{X\leq x\}=\bigcup_{1\leq k\leq \ell}\{X=x_k\},\quad \text{per } x_\ell\leq x<x_{\ell+1}$$

e quindi

$$F_X(x):=\mathbb{P}(X\leq x)=\begin{cases}0 & \text{per } x<x_1\\ p_X(x_1) & \text{per } x_1\leq x<x_2\\ p_X(x_1)+p_X(x_2) & \text{per } x_2\leq x<x_3\\ \cdots & \cdots\\ \sum_{k=1}^{\ell}p_X(x_k) & \text{per } x_\ell\leq x<x_{\ell+1}\\ \cdots & \cdots\\ \\ \sum_{k=1}^{n-1}p_X(x_k) & \text{per } x_{n-1}\leq x<x_n\\ 1=\sum_{k=1}^{n}p_X(x_k) & \text{per } x\geq x_n\end{cases}$$

ESEMPIO (dal testo di ROSS) Se X è una variabile aleatoria con $X(\Omega)=\{1,2,3,4\}$ e con

$$p_X(1)=\frac{1}{4},\quad p_X(2)=\frac{1}{2},\quad p_X(3)=\frac{1}{8},\quad p_X(4)=\frac{1}{8},$$

allora

$$F_X(x):=\mathbb{P}(X\leq x)=\begin{cases}0 & \text{per } x<1\\ p_X(1)=\frac{1}{4} & \text{per } 1\leq x<2\\ p_X(1)+p_X(2)=\frac{1}{4}+\frac{1}{2}=\frac{3}{4} & \text{per } 2\leq x<3\\ p_X(1)+p_X(2)+p_X(3)=\frac{1}{4}+\frac{1}{2}+\frac{1}{8}=\frac{7}{8} & \text{per } 3\leq x<4\\ 1=\sum_{k=1}^{4}p_X(k) & \text{per } x\geq 4\end{cases}$$

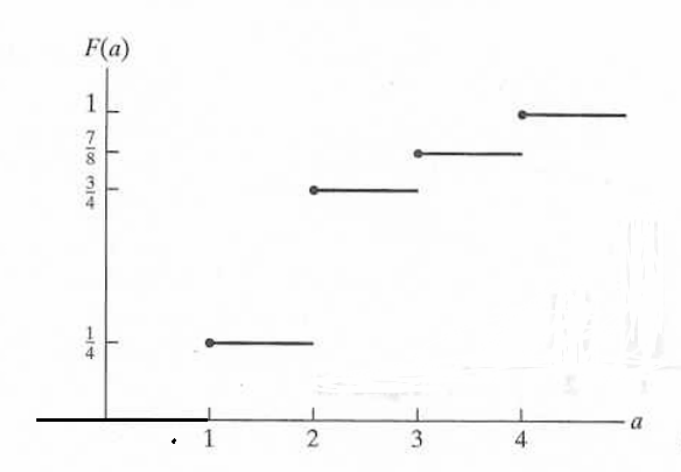


Figura 1: Grafico di $F(x) = \mathbb{P}(X \leq x)$ del precedente esempio

Notiamo che, per le variabili aleatorie discrete, F_X è una funzione costante a tratti, e che dal grafico di F_X è possibile ricavare l'insieme $X(\Omega)$ come l'insieme dei punti di discontinuità di F_X, ed è possibile ricavare anche i valori $p_X(x_k) = F_X(x_k) - F_X(x_k-)$, come l'ampiezza dei salti di F_X nei punti di discontinuità.
Ad esempio, anche se non conoscessimo la densità discreta di X, da questa figura, ossia il grafico di $F(x) = \mathbb{P}(X \leq x)$, o anche solo dalla descrizione della funzione $F(x)$

$$F_X(x) := \mathbb{P}(X \leq x) = \begin{cases} 0 & \text{per } x < 1 \\ \frac{1}{4} & \text{per } 1 \leq x < 2 \\ \frac{3}{4} & \text{per } 2 \leq x < 3 \\ \frac{7}{8} & \text{per } 3 \leq x < 4 \\ 1 & \text{per } x \geq 4 \end{cases}$$

potremmo ricavare la si vede che la funzione $F(x)$ è costante a tratti e ha come punti di discontinuità 1, 2, 3 e 4 e quindi $X(\Omega) = \{1,2,3,4\}$. Inoltre

$$F(1) - F(1-) = \frac{1}{4} - 0 = \frac{1}{4} \qquad \text{e quindi } p_X(1) = \mathbb{P}(X = 1) = \tfrac{1}{4},$$

$$F(2) - F(2-) = \frac{3}{4} - \frac{1}{4} = \frac{2}{4} = \frac{1}{2} \qquad \text{e quindi } p_X(2) = \mathbb{P}(X = 2) = \tfrac{1}{2},$$

$$F(3) - F(3-) = \frac{7}{8} - \frac{3}{4} = \frac{7}{8} - \frac{6}{8} = \frac{1}{8} \qquad \text{e quindi } p_X(3) = \mathbb{P}(X = 3) = \frac{1}{8},$$

$$F(4) - F(4-) = 1 - \frac{7}{8} = \frac{1}{8} \qquad \text{e quindi } p_X(4) = \mathbb{P}(X = 4) = \frac{1}{8},$$

1.3 Valore atteso e sue proprietà

Per le dimostrazioni vedere gli APPUNTI Spizzichino-Nappo

Definizione 1.3 (valore atteso)**.** *Sia* $X:\ \Omega \equiv \{\omega_1,...,\omega_N\} \to X(\Omega) \subseteq \mathbb{R}$ *una variabile aleatoria definita su* $(\Omega, \mathcal{P}(\Omega))$. *Sia inoltre definita la probabilità* $\mathbb{P}$, *su* $(\Omega, \mathcal{P}(\Omega))$. *Si definisce* **valore atteso** *di X (rispetto a* $\mathbb{P}$*) il numero*

$$\boxed{\mathbb{E}(X) \equiv \sum_{i=1}^{N} p(\omega_i) X(\omega_i),}$$

dove, come al solito, $p(\omega_i) = \mathbb{P}(\{\omega_i\})$.

Ricordiamo la relazione tra la funzione p la probabilità: $\boxed{\mathbb{P}(E) = \sum_{i:\omega_i \in E} p(\omega_i)}$

e che la funzione $\omega_i \in \Omega, \mapsto p(\omega_i)$ ha la proprietà che

$$\boxed{p(\omega_i) \geq 0,\ \forall \omega_i \in \Omega, \quad \text{e} \quad \sum_{i=1}^{N} p(\omega_i) = 1.}$$

Si dimostrano facilmente (vedere [SN]) che per il valore atteso valgono le seguenti proprietà

LINEARITÀ: $\boxed{\mathbb{E}(aX+bY) = a\mathbb{E}(X) + b\mathbb{E}(Y), \quad \forall\, a,b \in \mathbb{R} \text{ e } X \text{ ed } Y \text{ variabili aleatorie}}$

(che si estende al caso di n variabili aleatorie)

MONOTONIA: $\boxed{X(\omega_i) \leq Y(\omega_i) \quad \forall i = 1,2,\ldots,N \quad \Rightarrow \quad \mathbb{E}(X) \leq \mathbb{E}(Y)}$

IL VALORE ATTESO DIPENDE SOLO DALLA SUA DENSITÀ DISCRETA ossia dalla funzione

$$p_X:\ X(\Omega) \to [0,1] \subset \mathbb{R}, \quad x_k \mapsto p_X(x_k) := \mathbb{P}(X = x_k)$$

dove $X(\Omega) = \{x_1,...,x_n\}$ è l'insieme dei valori che può assumere la variabile aleatoria.
Infatti si dimostra facilmente (vedere [SN]) che vale la seguente espressione alternativa del valore atteso:

$$\boxed{\mathbb{E}(X) = \sum_{k=1}^{n} x_k p_X(x_k) = \sum_{k=1}^{n} x_k \mathbb{P}(X = x_k)}$$

NOTA BENE: come conseguenza: due variabili aleatorie con la stessa densità discreta hanno lo stesso valore atteso.

Per variabili aleatorie X a valori in $\{0,1,...,m\}$ si ha anche (vedere [SN] o anche qui sotto la prima nota a pagina 24)

$$\boxed{\mathbb{E}(X) = \sum_{k=0}^{m-1} \mathbb{P}(X > k).}$$

Infine vale la **formula del valore atteso totale** (vedere [SN]): data una partizione H_1,...,H_m con $\mathbb{P}(H_j) > 0$, $j = 1,2,...,m$

$$\boxed{\mathbb{E}(X) = \sum_{j=1}^{m} \mathbb{E}(X|H_j)\mathbb{P}(H_j)}, \quad \text{dove } \mathbb{E}(X|H_j) := \sum_{k=1}^{n} x_k \mathbb{P}(X = x_k|H_j).$$

1.4 Trasformazioni di variabili aleatorie discrete

Sia h una funzione nota e X una variabile aleatoria di cui è nota la densità discreta. Allora la variabile aleatoria $\omega \in \Omega \mapsto Z(\omega) := h(X(\omega))$ è una trasformazione di X. Vale inoltre che (per la dimostrazione vedere [SN])

se $Z = h(X)$, allora $\quad \boxed{\mathbb{E}(Z) = \mathbb{E}(h(X)) = \sum_{k=1}^{n} h(x_k) p_X(x_k) = \sum_{k=1}^{n} h(x_k) \mathbb{P}(X = x_k)}$

e la densità discreta di Z è data da

$$p_Z:\ Z(\Omega)\to[0,1]\subset\mathbb{R},\quad z_j\mapsto \boxed{p_Z(z_j)=\mathbb{P}(Z=z_j)=\sum_{k:h(x_k)=z_j}p_X(x_k)=\sum_{k:h(x_k)=z_j}\mathbb{P}(X=x_k).}$$

in quanto $\{Z=z_j\}=\cup_{k:h(x_k)=z_j}\{X=x_k\}$

1.4.1 Varianza

DEFINIZIONE: la ***varianza*** di una variabile aleatoria è definita come

$$\boxed{Var(X):=\mathbb{E}\big[\big(X-\mathbb{E}(X)\big)^2\big]}\quad\text{e si può calcolare come}\quad\boxed{Var(X)=\mathbb{E}\big(X^2\big)-\big(\mathbb{E}(X)\big)^2}$$

(per la verifica vedere [SN], ma la verifica è immediata: basta sviluppare l'espressione $\big(X-\mathbb{E}(X)\big)^2$ e usare la proprietà di linearità del valore atteso)
PROPRIETA' della VARIANZA

$$\boxed{Var(X)\geq 0}\quad\boxed{Var(X)=\mathbb{E}(X^2)-\big(\mathbb{E}(X)\big)^2}\quad\boxed{Var(aX+b)=a^2Var(X),\ \text{per ogni } a,b\in\mathbb{R}}$$

Ne diamo rapidamente la verifica:
Chiaramente $0\leq(X(\omega)-E(X))^2$ per ogni $\omega\in\Omega$ e quindi, per la proprietà di monotonia del valore atteso si ha: $0=\mathbb{E}(0)\leq\mathbb{E}\big[(X(\omega)-E(X))^2\big]=:Var(X)$.
Chiaramente, per ogni $\omega\in\Omega$ si ha $\big(X(\omega)-\mathbb{E}(X)\big)^2=X(\omega)^2+\big(\mathbb{E}(X)\big)^2-2\cdot X(\omega)\mathbb{E}(X)$ da cui

$$Var(X)=\mathbb{E}\big[\big(X-\mathbb{E}(X)\big)^2\big]=\mathbb{E}\big[X^2+\big(\mathbb{E}(X)\big)^2-2\cdot X\mathbb{E}(X)\big]=\mathbb{E}\big[X^2\big]+\big(\mathbb{E}(X)\big)^2-2\cdot\mathbb{E}(X)\mathbb{E}(X)\big]=\mathbb{E}\big(X^2\big)-\big(\mathbb{E}(X)\big)^2$$

Chiaramente, per la proprietà di linearità del valore atteso, $\mathbb{E}(aX+b)=a\mathbb{E}(X)+b$, inoltre per ogni $\omega\in\Omega$ si ha $\big(aX(\omega)+b-(a\mathbb{E}(X)+b)\big)^2=\big(aX(\omega)+\not{b}-a\mathbb{E}(X)-\not{b}\big)^2=a^2\big(X(\omega)-\mathbb{E}(X)\big)^2$, da cui

$$Var(aX+b)=\mathbb{E}\big[\big(aX+b-(a\mathbb{E}(X)+b)\big)^2\big]=\mathbb{E}\big[a^2\big(X)-\mathbb{E}(X)\big)^2\big]=a^2\,\mathbb{E}\big[\big(X)-\mathbb{E}(X)\big)^2\big]=a^2\,Var(X)$$

IMPORTANTE: la varianza di X spesso viene denotara come σ_X^2 è un indicatore della dispersione della distribuzione di X, ma come indicatore è più opportuno considerare la **deviazione standard** (anche detta **scarto quadratico medio**) ossia

$$\boxed{\sigma_X:=\sqrt{Var(X)}}.$$

1.5 Densità congiunta, Covarianza di due variabili aleatorie e Varianza della somma

Date due variabili aleatorie X e Y, definite su uno spazio di probabilità finito, la loro DENSITÀ DISCRETA CONGIUNTA è la funzione

$$p_{X,Y}:\ X(\Omega)\times Y(\Omega)\to[0,1]\subset\mathbb{R},\quad(x_k,y_h)\mapsto\boxed{p_{X,Y}(x_k,y_h):=\mathbb{P}(X=x_k,Y=y_h)}$$

Per calcolare il valore atteso di una trasformazione delle due variabili aleatorie, ossia di $W=g(X,Y)$, dove g è una funzione nota, allora, analogamente a quanto accade per le funzioni di una sola variabile, si ha:

$$\boxed{\mathbb{E}(W)=\mathbb{E}\big[g(X,Y)\big]=\sum_{k=1}^{n}\sum_{h=1}^{m}g(x_k,y_h)p_{X,Y}(x_k,y_h)=\sum_{k=1}^{n}\sum_{h=1}^{m}g(x_k,y_h)\mathbb{P}(X=x_k,Y=y_h)}$$

Inoltre, per calcolare la sua densità discreta, basta osservare che

$$\{W=w\}=\big\{\omega\in\Omega:\ g(X(\omega),Y(\omega)=w\big\}=\bigcup_{(x_k,y_h):g(x_k,y_h)=w}\{X=x_k,Y=y_h\}$$

per cui:

$$p_W(w):=\mathbb{P}(W=w)=\mathbb{P}(g(X,Y)=w)=\sum\sum_{x_k,y_h):g(x_k,y_h)=w}\mathbb{P}(X=x_k,Y=y_h)=\sum\sum_{x_k,y_h):g(x_k,y_h)=w}p_{X,Y}(x_k,y_h)$$

DEFINIZIONE: ***Covarianza di X e Y:***

$$\boxed{Cov(X,Y) := \mathbb{E}\big[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\big],} \quad \text{e si può calcolare come } \boxed{Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}.$$

Per la verifica vedere [SN], ma la verifica è immediata:
basta sviluppare l'espressione $\big(X-\mathbb{E}(X)\big)\big(Y-\mathbb{E}(Y)\big)$ e usare la proprietà di linearità del valore atteso

PROPRIETÀ: $\boxed{Cov(X,Y) = Cov(Y,X)}$, $\boxed{Cov(aX_1+bX_2,Y) = aCov(X_1,Y)+bCov(X_2,Y)}$

Se $Cov(X,Y) > 0$ si dice che X e Y sono correlate positivamente;

Se $Cov(X,Y) < 0$ si dice che X e Y sono correlate negativamente;

Se $Cov(X,Y) = 0$ si dice che X e Y sono scorrelate, e si dice anche che X e Y non sono correlate.

Si dimostra facilmente che

$$\boxed{Var(X+Y) = Var(X)+Var(Y)+2Cov(X,Y)}$$

e più in generale (per la dimostrazione vedere [SN]))

$$\boxed{Var\left(\sum_{k=1}^{n} X_k\right)} = \sum_{h=1}^{n}\sum_{k=1}^{n} Cov(X_h,X_k) \boxed{= \sum_{k=1}^{n} Var(X_k) + \sum_{h\neq k}^{1\leq h,k\leq n} Cov(X_h,X_k)} = \sum_{k=1}^{n} Var(X_k) + 2\sum_{h=1}^{n-1}\sum_{k=h+1}^{n} Cov(X_h,X_k).$$

QUINDI se le variabili aleatorie sono scorellate, allora la varianza della somma è la somma delle varianze, ossia

$$\boxed{\text{se } Cov(X_h,X_k) = 0, \text{ per ogni } h \neq k \text{ allora } Var\Big(\sum_{k=1}^{n} X_k\Big) = \sum_{k=1}^{n} Var(X_k).}$$

1.6 Indipendenza per variabili aleatorie

1.6.1 Indipendenza di due variabili aleatorie e relazione con la non correlazione

DEFINIZIONE: Due variabili aleatorie X e Y **discrete**[4]sono ***indipendenti*** (brevemente $X \perp\!\!\!\perp Y$) se e solo se sono indipendenti le partizioni $\mathcal{P}_X$ e $\mathcal{P}_Y$ generate da X e da Y rispettivamente, dove

$$\mathcal{P}_X = \big\{H_x^X = \{X = x\},\, x \in X(\Omega)\big\}, \quad \mathcal{P}_Y = \big\{H_y^Y = \{Y = xy\},\, y \in Y(\Omega)\big\}$$

ossia

$$\mathbb{P}(H_x^X \cap H_y^Y) = \mathbb{P}(H_x^X)\mathbb{P}(H_y^Y), \quad \text{per ogni } x \in X(\Omega) \text{ e } y \in Y(\Omega)$$

o equivalentemente se e solo seù

$$\boxed{\mathbb{P}(X = x_k, Y = y_h) = \mathbb{P}(X = x_k)\mathbb{P}(Y = y_h)} \quad \text{per ogni } x_k \in X(\Omega) \text{ e } y_h \in Y(\Omega),$$

ovvero se e solo se

$$\boxed{p_{X,Y}(x_k,y_h) = p_X(x_k)p_Y(y_h)} \quad \text{per ogni } x_k \in X(\Omega) \text{ e } y_h \in Y(\Omega).$$

Si dimostra facilmente che ciò equivale a chiedere che

$$\boxed{\mathbb{P}(X \in I, Y \in J) = \mathbb{P}(X \in I)\mathbb{P}(Y \in J)} \quad \text{per ogni } I \subset X(\Omega) \text{ e } J \subset Y(\Omega).$$

Vale inoltre la seguente proprietà:
Proposizione: *Le variabili aleatorie discrete*[5] *X e Y sono indipendenti SE E SOLO SE, per ogni coppia di funzioni reali a valori reali h_1 e h_2, si ha*

$$\mathbb{E}\big[h_1(X)h_2(Y)\big] = \mathbb{E}\big[h_1(X)\big] \cdot \mathbb{E}\big[h_2(Y)\big]$$

[4]Per definizione si dice che una variabile aleatoria X è discreta se e solo se può assumere valori in un insieme finito o infinito numerabile, ossia $|X(\Omega)| \leq |\mathbb{N}|$. In questo contesto siamo in spazi di probabilità finiti (Ω finito) e quindi tutte le variabili sono necessariamente discrete, e lo stesso vale se Ω è infinito, ma numerabile, ossia $|\Omega| = |\mathbb{N}|$, ma la definizione vale anche in spazi di probabilità più generali.

[5]Qui siamo nel caso Ω finito e non ci sono problemi tecnici. Tuttavia, nel caso in cui $X(\Omega)$ e/o $Y(\Omega)$ siano infiniti numerabili, vanno aggiunte delle condizione che assicurino che i valori attesi abbiano senso e sian finiti. Vedere la Definizione 3.1

Come conseguenza della precedente proprietà, prendendo in particolare $h_1(x) = x$ e $h_2(y) = y$ di ottiene che

$$\boxed{\text{se } X \perp\!\!\!\perp Y \text{ allora } \mathbb{E}(XY) = \mathbb{E}(X)\cdot\mathbb{E}(Y) \text{ ossia } Cov(X,Y) = 0} \quad \text{e quindi} \quad \boxed{Var(X+Y) = Var(X) + Var(Y)}.$$

IMPORTANTE: non vale il viceversa, ossia $\boxed{Cov(X,Y) = 0 \text{ NON IMPLICA } X \perp\!\!\!\perp Y}$

(vedere il controesempio nell'Osservazione 10.4 degli Appunti [SN] con X uniforme in $\{-1,0,1\}$ e $Y = X^2$, in modo che $XY = X^3$ e$X^3 = X$, in quanto $x^3 = x$ SE E SOLO SE $x \in \{-1,0,1\}$.)

1.6.2 Indipendenza di n variabili aleatorie

DEFINIZIONE: Date n variabili aleatorie **discrete** $X_1, X_2, ..., X_n$, si dice che sono ***(completamente/globalmente indipendenti*** se e solo se sono indipendenti le partizione $\mathcal{P}_i$ generate da X_i, ossia

$$\mathcal{P}_i = \{H_x^{X_i} = \{X = x\}, \text{per} x \in X_i(\Omega)\}$$

Ciò equivale a chiedere che per ogni $x^{(1)} \in X_1(\Omega)$, $x^{(2)} \in X_2(\Omega), ..., x^{(n)} \in X_n(\Omega)$,

$$\mathbb{P}(X_1 = x^{(1)}, X_2 = x^{(2)}, ..., X_n = x^{(n)}) = \mathbb{P}(X_1 = x^{(1)})\cdot\mathbb{P}(X_2 = x^{(2)})\cdots\mathbb{P}(X_n = x^{(n)}).$$

Inoltre si dimostra che ciò equivale a chiedere che

$$\mathbb{P}(X_1 \in I_1, X_2 \in I_2, ..., X_n \in I_n) = \mathbb{P}(X_1 \in I_1)\cdot\mathbb{P}(X_2 \in I_2)\cdots\mathbb{P}(X_n \in I_n) \quad \text{per ogni } I \subset X(\Omega) \text{ e } J \subset Y(\Omega),$$

Vale inoltre la seguente proprietà:
Proposizione *le variabile aleatorie $X_1, X_2, ..., X_n$, sono (completamente) indipendenti se e solo se, per ogni scelta di funzioni h_1, h_2,..., h_n si ha*

$$\mathbb{E}\big[h_1(X_1)h_2(X_2)\cdots h_n(X_n)\big] = \mathbb{E}\big[h_1(X)\big]\cdot\mathbb{E}\big[h_2(X_2)\big]\cdots\mathbb{E}\big[h_n(X_n)\big].$$

Osservazione: Se $X_1, X_2, ..., X_n$, sono (completamente) indipendenti, scegliendo in particolare $h_\ell(x) = 1$ per $\ell \neq k, j$ si ottiene che, per qualunque scelta di h_k e h_j sia ha

$$\mathbb{E}\big[h_k(X_k)h_j(X_j)\big] = \mathbb{E}\big[h_k(X_k)\big]\mathbb{E}\big[h_j(X_j)\big] \quad \Leftrightarrow \quad X_k \perp\!\!\!\perp X_j \quad \text{e quindi} \quad Cov(X_k, X_j) = 0.$$

Di conseguenza si ottiene la seguente proprietà: Se $X_1, X_2, ..., X_n$, sono (completamente) indipendenti, allora

$$Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n).$$

1.6.3 Indipendenza di una successione di variabili aleatorie

NOTA BENE: si tratta di un'anticipazione, in quanto negli spazi di probabilità finiti, non può esistere una successione di variabili aleatorie indipendenti

DEFINIZIONE: Una successione di variabili aleatorie $\{X_n, n \geq 1\}$ è una successione di variabili aleatorie (**completamente/globalmente**) indipendenti se per ogni $J \subset \mathbb{N}$ finito le variabili aleatorie $\{X_j,\, j \in J\}$ sono indipendenti.

2 Variabili aleatorie discrete con un numero finito di valori: distribuzioni notevoli

Degenere in $\hat{x}$: $X:\omega\mapsto X(\omega)=\hat{x}$, allora $X(\Omega)=\{\hat{x}\}$ e $p_X(\hat{x}):=\mathbb{P}(X=\hat{x})=1$

$$\mathbb{E}(X)=\mathbb{E}(\hat{x})=\hat{x},\qquad Var(X)=\mathbb{E}\big(\overbrace{(X-\hat{x})^2}^{=0}\big)=0.$$

Binaria o funzione indicatrice di un evento A, con $\mathbb{P}(A)=p$, anche detta di Bernoulli di parametro p, $Bernoulli(p)$:

$$X(\omega)=\mathbf{1}_A(\omega)=\begin{cases}1 & \text{se } \omega\in A\\ 0 & \text{se } \omega\in\overline{A}\end{cases},\quad\Leftrightarrow\quad \{X=1\}=A,\ \{X=0\}=\overline{A}$$

per cui $X(\Omega)=\{0,1\}$, $\{X=0\}=\overline{A}$, $\{X=1\}=A$, e $p_X(0)=1-\mathbb{P}(A)=1-p$, $p_X(1)=\mathbb{P}(A)=p$.

> Si può scrivere rapidamente
> $$p_X(h)=p^h\cdot(1-p)^{1-h}.\quad h=0,1$$
> infatti si ha
> $$p_X(0)=p^0\cdot(1-p)^{1-0}=1-p,\quad \text{e}\quad p_X(1)=p^1\cdot(1-p)^{1-1}=p.$$

Osservazione: $(\mathbf{1}_A)^2=\mathbf{1}_A$ in quanto $1^2=1$ e $0^2=0$ e si ha

$$\mathbb{E}(\mathbf{1}_A)=\mathbb{P}(A),\quad Var(\mathbf{1}_A)=\mathbb{P}(A)\big(1-\mathbb{P}(A)\big).$$

Infatti $\mathbb{E}(\mathbf{1}_A)=1\cdot\mathbb{P}(X=1)+0\cdot\mathbb{P}(X=0)=\mathbb{P}(A)$, *e*

$$Var(\mathbf{1}_A)=\mathbb{E}\big[\overbrace{(\mathbf{1}_A)^2}^{=\mathbf{1}_A}\big]-\big(\mathbb{E}(\mathbf{1}_A)\big)^2=\mathbb{E}(\mathbf{1}_A)-\big(\mathbb{E}(\mathbf{1}_A)\big)^2=\mathbb{P}(A)-[\mathbb{P}(A)]^2=\mathbb{P}(A)\big(1-\mathbb{P}(A)\big).$$

Somma di due variabili binarie: $X:=\mathbf{1}_A+\mathbf{1}_B$ allora $X(\Omega)=\{0,1,2\}$ e

$$H_0^X=\{X=0\}=\overline{A}\cap\overline{B},\quad H_1^X=\{X=1\}=A\triangle B=(A\cap\overline{B})\cup(\overline{A}\cap B),\quad H_2^X=\{X=2\}=A\cap B,$$

e quindi

$$p_X(0)=\mathbb{P}(\overline{A}\cap\overline{B}),\quad p_X(1)=\mathbb{P}(A\cap\overline{B})+\mathbb{P}(\overline{A}\cap B),\quad p_X(2)=\mathbb{P}(A\cap B).$$

Per la proprietà di linearità del valore atteso si ha

$$\mathbb{E}(\mathbf{1}_A+\mathbf{1}_B)=\mathbb{E}(\mathbf{1}_A)+\mathbb{E}(\mathbf{1}_B)=\mathbb{P}(A)+\mathbb{P}(B)$$

> La proprietà di linearità ci permette di arrivare subito al valore atteso, senza bisogno di fare un calcolo del tipo
> $$\mathbb{E}(X)=0\cdot p_X(0)+1\cdot p_X(1)+2\cdot p_X(2)=0\cdot\mathbb{P}(\overline{A}\cap\overline{B}),+1\cdot\big[\mathbb{P}(A\cap\overline{B})+\mathbb{P}(\overline{A}\cap B)\big]+\mathbb{P}(A\cap B).$$

$$Var(\mathbf{1}_A+\mathbf{1}_B)=Var(\mathbf{1}_A)+Var(\mathbf{1}_B)+2Cov(\mathbf{1}_A,\mathbf{1}_B)=\mathbb{P}(A)\big(1-\mathbb{P}(A)\big)+\mathbb{P}(B)\big(1-\mathbb{P}(B)\big)+2\big[\mathbb{P}(A\cap B)-\mathbb{P}(A)\mathbb{P}(B)\big]$$

Infatti $\mathbf{1}_A\cdot\mathbf{1}_B=\mathbf{1}_{A\cap B}$:

$$\mathbf{1}_A(\omega)\cdot\mathbf{1}_B(\omega)=\begin{cases}1 & \mathbf{1}_A(\omega)=1\text{ e }\mathbf{1}_B(\omega)=1\Leftrightarrow\omega\in A\text{ e }\omega\in B\Leftrightarrow\omega\in A\cap B\\ 0 & \text{altrimenti}\end{cases}\qquad \mathbf{1}_{A\cap B}(\omega)=\begin{cases}1 & \omega\in A\cap B\\ 0 & \text{altrimenti}\end{cases}$$

e quindi

$$Cov(\mathbf{1}_A,\mathbf{1}_B)=\underbrace{\mathbb{E}(\mathbf{1}_A\cdot\mathbf{1}_B)}_{=\mathbb{E}(\mathbf{1}_{A\cap B})}-\mathbb{E}(\mathbf{1}_A)\cdot\mathbb{E}(\mathbf{1}_B)=\mathbb{P}(A\cap B)-\mathbb{P}(A)\mathbb{P}(B).$$

Uniforme in $[n] := \{1,2,...,n\}$: $X \sim Unif([n])$ se e solo se $p_X(k) = \mathbb{P}(X=k) = \frac{1}{n}$, $k = 1,2,...,n$.
Come visto in Esempio 9.1 di [SN] (per il valore atteso)

$$\mathbb{E}(X) = \frac{n+1}{2}, \quad Var(X) = \frac{n^2-1}{12}.$$

Per il valore atteso si ricorda che

$$\mathbb{E}(X) = \sum_{k=1}^{n} k\,\mathbb{P}(X=k) = \sum_{k=1}^{n} k\frac{1}{n} = \frac{\not{n}(n+1)}{2}\cdot\frac{1}{\not{n}} = \frac{n+1}{2}$$

Per la varianza si ricorda che si può calcolare $Var(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2$ *e che*

$$\mathbb{E}(X^2) = \sum_{k=1}^{n} k^2\frac{1}{n} = \frac{1}{n}\,\frac{n(n+\frac{1}{2})(n+1)}{3} = \frac{(2n+1)(n+1)}{6}\left(= \frac{2n^2+3n+1}{6}\right),$$

da cui

$$Var(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \frac{(2n+1)(n+1)}{6} - \frac{(n+1)^2}{4} = \frac{(n+1)[2(2n+1)-3(n+1)]}{12} = \frac{(n+1)[4n+2-3n-3]}{12}$$
$$= \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12}.$$

Ipergeometrica: $X \sim Hyp(M;m,n)$ se e solo se

$$p_X(k) = \mathbb{P}(X=k) = \frac{\binom{m}{k}\binom{M-m}{n-k}}{\binom{M}{n}}, \quad k, \text{ tale che } 0 \le k \le m,\ 0 \le n-k \le M-m$$

e $p_X(k) = 0$ per tutti gli altri valori di k.
prototipo di variabile aleatoria ipergeometrica: in n estrazioni **senza reinserimento** da un'urna che contiene M oggetti di cui $m = m_A$ di tipo A, ed $m_B = M - m_A$ di tipo B, allora $S_n := \mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \cdots + \mathbf{1}_{A_n}$, dove $A_i = \{$*all'i-sima estrazione esce un oggetto di tipo A*$\}$ ha distribuzione ipergeometrica.
Come visto nell'Esempio 9.4 (si veda anche l'Osservazione 9.1) e nell'Esempio 10.4 di [SN] si ha che:

$$\mathbb{E}(X) = n\frac{m}{M}, \quad Var(X) = n\frac{m}{M}\left(1-\frac{m}{M}\right)\left(1-\frac{n-1}{M-1}\right).$$

Per il valore atteso: prima di tutto calcoliamo il valore atteso di S_n*, e ciò permette di ottenere il valore atteso di una qualunque variabile aleatoria X con distribuzione ipergeometrica*[6] *con gli stessi paramentri di* S_n*.*
Infatti, sempre grazie alla proprietà di linearità, tenendo conto che $\mathbb{P}(A_k) = \mathbb{P}(A_1) = \frac{m_A}{M}$*,*

$$\mathbb{E}(X) = \mathbb{E}(S_n) = \mathbb{E}(\mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \cdots + \mathbf{1}_{A_n}) = \mathbb{E}(\mathbf{1}_{A_1}) + \mathbb{E}(\mathbf{1}_{A_2}) + \cdots + \mathbb{E}(\mathbf{1}_{A_n}) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots \mathbb{P}(A_n)$$
$$= \frac{m_A}{M} + \frac{m_A}{M} + \cdots \frac{m_A}{M} = n\frac{m_A}{M}$$

Per la varianza: tenendo conto della formula della varianza della somma e che $\mathbb{P}(A_h \cap A_k) = \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\,\mathbb{P}(A_2|A_1) = \frac{m_A}{M}\frac{m_A-1}{M-1}$ *e che* $Cov(\mathbf{1}_A, \mathbf{1}_B) = \mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)$*,*

$$Var(X) = Var(S_n) = \sum_{k=1}^{n} \overbrace{Var(\mathbf{1}_{A_k})}^{=\mathbb{P}(A_k)(1-\mathbb{P}(A_k))} + \sum_{h\neq k}^{1\le h,k\le n}\sum \overbrace{Cov(\mathbf{1}_{A_h},\mathbf{1}_{A_k})}^{=\mathbb{P}(A_h\cap A_k)-\mathbb{P}(A_h)\mathbb{P}(A_k)}$$
$$= n\,\mathbb{P}(A_1)\,(1-\mathbb{P}(A_1)) + n(n-1)\,\big(\mathbb{P}(A_1\cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)\big)$$
$$= n\frac{m_A}{M}\left(1-\frac{m_A}{M}\right) + n(n-1)\left(\frac{m_A}{M}\frac{m_A-1}{M-1} - \frac{m_A}{M}\frac{m_A}{M}\right)$$
$$= n\,\vartheta(1-\vartheta)\,\big(1-\frac{n-1}{M}\big), \quad \text{dove } \vartheta := \frac{m_A}{M}.$$

[6] Infatti il valore atteso di una variabile aleatoria discreta dipende SOLO dalla sua distribuzione e quindi se X è ipergeometrica $Hyp(N,m;n)$ ed $m_A = m$, allora $\mathbb{E}(X) = \mathbb{E}(S_n)$. Una osservazione analoga vale per la varianza.

È immediato rendersi conto che se n è fissato, m_A ed M sono grandi, in modo che $\frac{n-1}{M}$ è piccola, sempre con $\vartheta := \frac{m_A}{M}$, allora $Var(X)$ è approssimata da $n\vartheta(1-\vartheta)$. Ricordiamo che si dimostra (vedere l'Osservazione 6.3 in [SN]) allora, per ogni $k \in \{0,1,...,n\}$, $\mathbb{P}(X=k)$ è approssimata da $\binom{n}{k}\vartheta^k(1-\vartheta)^{n-k}$. Come vedremo la varianza di una v.a. $Bin(n,\vartheta)$ vale appunto $n\vartheta(1-\vartheta)$.

Binomiale: $X \sim Bin(n,\vartheta)$ se e solo se $p_X(k) = \mathbb{P}(X=k) = \binom{n}{k}\vartheta^k(1-\vartheta)^{n-k}, \quad k=0,1,...,n$

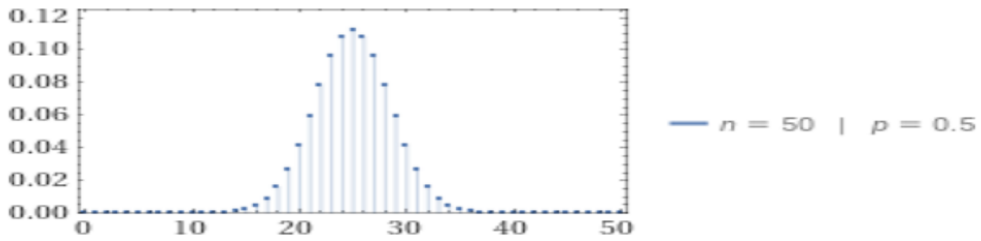


Figura 2: $X \sim Bin(50,\frac{1}{2})$, con $\mathbb{E}(X) = 50\cdot\frac{1}{2} = 25, \quad Var(X) = 50\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{25}{2} = 12,5$

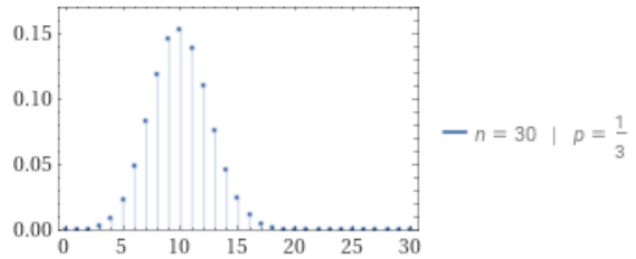


Figura 3: $X \sim Bin(30,\frac{1}{3})$, con $\mathbb{E}(X) = 30\cdot\frac{1}{3} = 10, \quad Var(X) = 30\cdot\frac{1}{3}\cdot\frac{2}{3} = \frac{20}{3} \approx 6,6667$

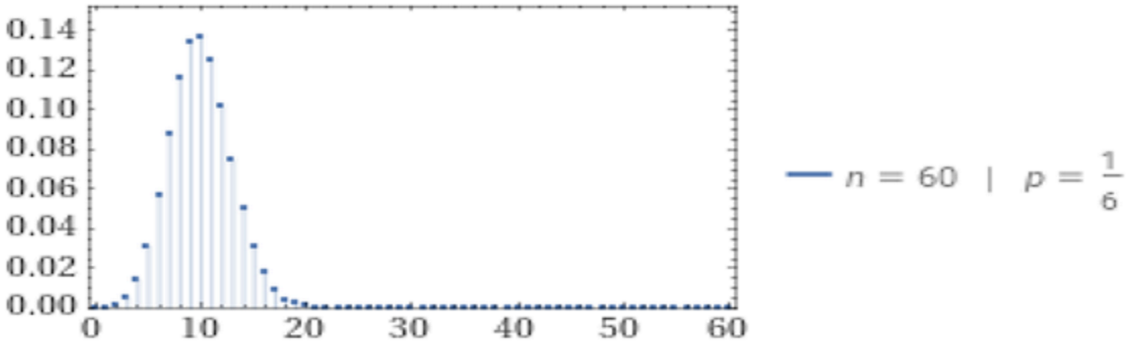


Figura 4: $X \sim Bin(60,\frac{1}{6})$, con $\mathbb{E}(X) = 60\cdot\frac{1}{6} = 10, \quad Var(X) = 60\cdot\frac{1}{6}\cdot\frac{5}{6} = \frac{50}{6} \approx 8,3333$

prototipo: numero di successi in uno schema di Bernoulli, ossia su *n* prove indipendenti e con le stesse probabilità, ovvero la variabile aleatoria $S_n = \mathbf{1}_{E_1} + \cdots + \mathbf{1}_{E_n}$, dove E_1,...,E_n sono eventi (completamente/globalmente) indipendenti e con $\mathbb{P}(E_i) = \vartheta$. Come visto nell'Esempio 9.3 (si veda anche l'Osservazione 9.1) e nell'Esempio 10.3 di [SN] si ha:

$$\mathbb{E}(X) = n\vartheta, \quad Var(X) = n\vartheta(1-\vartheta).$$

Per il valore atteso: sempre per la proprietà di linearità si ha

$$\begin{aligned}\mathbb{E}(X) = \mathbb{E}(S_n) &= \mathbb{E}(\mathbf{1}_{E_1} + \mathbf{1}_{E_2} + \cdots + \mathbf{1}_{E_n}) = \mathbb{E}(\mathbf{1}_{E_1}) + \mathbb{E}(\mathbf{1}_{E_2}) + \cdots + \mathbb{E}(\mathbf{1}_{E_n}) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \cdots \mathbb{P}(E_n) \\ &= \vartheta + \vartheta + \cdots + \vartheta = n\vartheta\end{aligned}$$

Per la varianza, l'idea è: sfruttare la formula del valore atteso e l'indipendenza degli eventi E_k, che implica che $Cov(\mathbf{1}_{E_h}, \mathbf{1}_{E_k}) = 0$, per ogni $k \neq h$, di modo che si ha

$$Var(X) = \sum_{k=1}^{n} \overbrace{Var(\mathbf{1}_{E_k})}^{=\mathbb{P}(E_k)(1-\mathbb{P}(E_k))} + \sum_{h\neq k}^{1\leq h,k\leq n} \overbrace{Cov(\mathbf{1}_{E_h}, \mathbf{1}_{E_k})}^{=0} = n\cdot\vartheta(1-\vartheta).$$

Tempo di primo successo troncato

Sia X_n il tempo di primo successo troncato in n prove di Bernoulli, definito come segue:
ossia se $A_1, A_2, ..., A_n$ sono n eventi indipendenti e tutti con la stessa probabilità p, la variabile X_n è definita come segue:

$$X_n := 1 \quad \text{se e solo se } \omega \in A_1,$$

da cui $\mathbb{P}(X_n = 1) = p$

$$X_n := 2 \quad \text{se e solo se } \omega \in \overline{A}_1 \cap A_2,$$

da cui

$$\mathbb{P}(X_n = 2) = \mathbb{P}(\overline{A}_1 \cap A_2) = \mathbb{P}(\overline{A}_1)P(A_2) = (1-p)p$$
$$.....$$
$$X_n := k \quad \text{se e solo se } \omega \in \overline{A}_1 \cap \overline{A}_2 \cap .... \cap \overline{A}_{k-1} \cap A_k,$$

da cui

$$\mathbb{P}(X_n = k) = \mathbb{P}(\overline{A}_1 \cap \overline{A}_2 \cap .... \cap \overline{A}_{k-1} \cap A_k) = \mathbb{P}(\overline{A}_1)\mathbb{P}(\overline{A}_2) \cdots \mathbb{P}(\overline{A}_{k-1})\mathbb{P}(A_k) = (1-p)^{k-1}p$$
$$.....$$
$$X_n := n \quad \text{se e solo se } \omega \in \overline{A}_1 \cap \overline{A}_2 \cap .... \cap \overline{A}_{n-1} \cap A_n,$$

da cui

$$\mathbb{P}(X_n = n) = \mathbb{P}(\overline{A}_1 \cap \overline{A}_2 \cap .... \cap \overline{A}_{n-1} \cap A_n) = \mathbb{P}(\overline{A}_1)\mathbb{P}(\overline{A}_2) \cdots \mathbb{P}(\overline{A}_{n-1})\mathbb{P}(A_n) = (1-p)^{n-1}p$$

infine poniamo (in modo arbitrario)

$$X_n := n+1 \quad \text{se e solo se } \omega \in \overline{A}_1 \cap \overline{A}_2 \cap .... \cap \overline{A}_{n-1} \cap \overline{A}_n,$$

da cui

$$\mathbb{P}(X_n = n+1) = \mathbb{P}(\overline{A}_1 \cap \overline{A}_2 \cap .... \cap \overline{A}_{n-1} \cap \overline{A}_n) = \mathbb{P}(\overline{A}_1)\mathbb{P}(\overline{A}_2) \cdots \mathbb{P}(\overline{A}_{n-1})\mathbb{P}(\overline{A}_n) = (1-p)^n$$

Usando il fatto che l'insieme dei valori che assume X_n è contenuto in $\{0, 1, ..., n, n+1\}$, possiamo affermare che

$$\mathbb{E}(X_n) = \mathbb{P}(X_n > 0) + \mathbb{P}(X_n > 1) + \mathbb{P}(X_n > 2) + ... + \mathbb{P}(X_n > n)$$

e usando il fatto che

$$\mathbb{P}(X_n > k) = \mathbb{P}(\overline{A}_1 \cap \overline{A}_2 \cap \cdots \cap \overline{A}_{k-1} \cap \overline{A}_k) = (1-p)^k,$$

si ottiene che (supponiamo $0 < p < 1$, per evitare casi banali)

$$\begin{aligned}\mathbb{E}(X_n) &= \mathbb{P}(X_n > 0) + \mathbb{P}(X_n > 1) + \mathbb{P}(X_n > 2) + ... + \mathbb{P}(X_n > n) \\ &= 1 + (1-p) + ... + (1-p)^k + ... + (1-p)^n = \frac{1-(1-p)^{n+1}}{1-(1-p)} = \frac{1-(1-p)^{n+1}}{p}.\end{aligned}$$

Numero delle Concordanze: Consideriamo Ω come l'insieme di tutte le permutazioni $\sigma = \big(\sigma(1), \sigma(2), ..., \sigma(n)\big)$ di n oggetti, che supponiamo numerati da 1 a n (quindi $|\Omega| = n!$) e mettiamoci nell'ipotesi che tutte le permutazioni siano equiprobabili. Posto

$$A_k = \{\text{concordanza al posto } k\} = \{\sigma \in \Omega : \ \sigma(k) = k\}, \quad k = 1, 2, ..., n,$$

e posto $X = S_n$ il numero di indici che sono punti fissi, ossia il numero di indici $i \in \{1, 2, ..., n\}$ tali che $\sigma(i) = i$, si ha

$$S_n = \mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \cdots + \mathbf{1}_{A_n}, \quad \text{e} \quad \mathbb{E}(S_n) = 1, \quad Var(S_n) = 1.$$

Infatti, da tale espressione, in modo simile a quanto fatto per la distribuzione ipergeometrica, possiamo calcolare valore atteso e varianza:
tenendo conto che

$$\mathbb{P}(A_k) = \frac{|A_k|}{|\Omega|} = \frac{(n-1)!}{n!} = \frac{1}{n}, \quad \forall\, k = 1, 2, ..., n,$$

otteniamo che

$$\mathbb{E}(S_n) = \mathbb{E}\big(\mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \cdots + \mathbf{1}_{A_n}\big) = \mathbb{E}\big(\mathbf{1}_{A_1}\big) + \mathbb{E}\big(\mathbf{1}_{A_2}\big) + \cdots + \mathbb{E}\big(\mathbf{1}_{A_n}\big) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_n) = n\frac{1}{n} = 1,$$

e, tenendo conto che

$$\mathbb{P}(A_h \cap A_k) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}, \quad \forall\, h \neq k \ \in \{1, 2, ..., n\},$$

$$Var(S_n) = \sum_{k=1}^{n} \overbrace{Var(\mathbf{1}_{A_k})}^{=\mathbb{P}(A_k)(1-\mathbb{P}(A_k))} + \sum_{h \neq k}^{1 \leq h,k \leq n} \overbrace{Cov(\mathbf{1}_{A_h}, \mathbf{1}_{A_k})}^{=\mathbb{P}(A_h \cap A_k) - \mathbb{P}(A_h)\mathbb{P}(A_k)}$$

$$= n\,\mathbb{P}(A_1)\,(1 - \mathbb{P}(A_1)) + n(n-1)\,\big(\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)\big)$$

$$= n\,\frac{1}{n}\left(1 - \frac{1}{n}\right) + n(n-1)\,\left[\frac{1}{n}\frac{1}{n-1} - \frac{1}{n}\frac{1}{n}\right]$$

$$= \left(1 - \frac{1}{n}\right) + n(n-1)\,\frac{1}{n}\,\frac{n-(n-1)}{(n-1)n} = \left(1 - \frac{1}{n}\right) + \frac{1}{n} = 1.$$

Osserviamo che per calcolare valore atteso e varianza del numero delle concordanze **non è stato necessario calcolare la sua densità discreta**, che invece andiamo a calcolare ora. Nella soluzione del problema delle concordanze abbiamo visto che, da una parte, con l'ausilio della formula di inclusione/esclusione si ottiene che

$$\mathbb{P}(X = 0) = \mathbb{P}(S_n = 0) = p(n) := 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} + \cdots + (-1)^h\,\frac{1}{h!} + \cdots + (-1)^n\,\frac{1}{n!} = \sum_{h=0}^{n} (-1)^h\,\frac{1}{h!},$$

dall'altra però si potrebbe anche scrivere che

$$\mathbb{P}(X = 0) = \mathbb{P}(S_n = 0) = p(n) = \frac{\kappa_n(0)}{n!},$$

dove $\boxed{\kappa_n(0)}$ denota il ***numero delle permutazioni*** $(i_1, i_2, ...i_n)$ ***di*** n ***elementi che non presentano punti fissi***, ossia per le quali, qualunque sia $h \in \{1, 2, ..., n\}$, il valore i_h è diverso da h, ovvero $\kappa_n(0)$ è il numero delle permutazioni di n elementi senza concordanze, e dedurre che

$$\boxed{\kappa_n(0) = n! \cdot p(n) = n!\,\sum_{h=0}^{n} (-1)^h\,\frac{1}{h!}.}$$

Per rendere queste note autocontenute ricordiamo brevemente come si arriva al calcolo di $\mathbb{P}(X=0)=\mathbb{P}(S_n=0)=p(n)$ con la formula di inclusione/esclusione: posto $q(n):=\mathbb{P}(S_n>0)=1-\mathbb{P}(S_n=0)=1-p(n)$, si ha

$$\begin{aligned}\mathbb{P}(S_n>0)=q(n)&=\mathbb{P}(A_1\cup A_2\cup\cdots\cup A_n)\\&=\sum_{k=1}^{n}(-1)^{k-1}\sum_{\{i_1,i_2,\cdots,i_k\}}\mathbb{P}(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k})\end{aligned}$$

dove $\sum_{\{i_1,i_2,\cdots,i_k\}}$ va fatta su tutte le $\binom{n}{k}=\dfrac{n!}{k!(n-k)!}$ combinazioni di n elementi di classe k, ovvero su tutti i sottoinsiemi $\{i_1,i_2,\cdots,i_k\}\subset\{1,2,...,n\}$ di cardinalità k.
Inoltre, per ogni scelta del sottoinsieme/combinazione $\{i_1,i_2,\cdots,i_k\}\subseteq\{1,2,...,n\}$

$$\mathbb{P}(A_{i_1}\cap A_{i_2}\cap...\cap A_{i_k})=\mathbb{P}(A_1\cap A_2\cap...\cap A_k)=\frac{(n-k)!}{n!}$$

e quindi la somma

$$\sum_{\{i_1,i_2,\cdots,i_k\}}\mathbb{P}(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k})=\binom{n}{k}\mathbb{P}(A_1\cap A_2\cap...\cap A_k)=\frac{n!}{k!(n-k)!}\,\frac{(n-k)!}{n!}=\frac{1}{k!}$$

e di conseguenza

$$\begin{aligned}p(n)=1-q(n)&=1-\frac{1}{1!}+\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}+\cdots+(-1)^k\frac{1}{k!}+\cdots+(-1)^n\frac{1}{n!}\\&=\sum_{k=0}^{n}(-1)^k\frac{1}{k!}.\end{aligned}$$

L'osservazione che, per ogni $m\geq 1$,

$$\kappa_m(0)=m!\cdot p(m)=m!\sum_{k=0}^{m}(-1)^k\frac{1}{k!},$$

dove $\kappa_m(0)$ è il ***numero delle permutazioni*** $(i_1,i_2,....i_m)$ ***di*** m ***elementi che non presentano punti fissi***, ossia senza concordanze, ci permette di calcolare $\mathbb{P}(S_n=k)$: infatti i casi favorevoli a $\{S_n=k\}$ si contano tenendo presente che si possono scegliere in $C_k^n=\binom{n}{k}$ modi possibili quali sono le k lettere nella busta giusta (cioè i k valori j per i quali $i_j=j$), e che, per ognuna di queste scelte, ci sono poi tante scelte quante sono le permutazioni dei rimanenti $n-k$ elementi che non hanno punti fissi, ossia ci sono $\kappa_{n-k}(0)$ scelte. Per quanto osservato precedentemente queste ultime sono $\kappa_{n-k}(0)=(n-k)!p(n-k)$. In conclusione,

$$\begin{aligned}p_{S_n}(k)=\mathbb{P}(S_n=k)&=C_k^n\sharp\{\text{permutazioni di } n-k \text{ elementi che non presentano punti fissi}\}\,\frac{1}{n!}=\frac{\binom{n}{k}\cdot\kappa_{n-k}(0)}{n!}\\&=\frac{n!}{k!(n-k)!}\,(n-k)!\,p(n-k)\,\frac{1}{n!}=\frac{p(n-k)}{k!}\end{aligned}$$

Tenendo conto che $p(n-k)=\sum_{h=0}^{n-k}(-1)^h\,\frac{1}{h!}$, si ha dunque

$$p_{S_n}(k)=\mathbb{P}(S_n=k)=\frac{p(n-k)}{k!}=\frac{1}{k!}\left(\sum_{h=0}^{n-k}(-1)^h\frac{1}{h!}\right),\quad k=0,1,...,n-2$$

$$p_{S_n}(n-1)=\mathbb{P}(S_n=n-1)=0,\quad p_{S_n}(n)=\mathbb{P}(S_n=n)=\frac{1}{n!}$$

in quanto è chiaro non ci possono essere esattamente $n-1$ concordanze, perché se ce ne sono almeno $n-1$ allora ce ne sono necessariamente n e che solo la permutazione $\sigma=(1,2,...,n)$ ammette n concordanze.
Vale, però la pena osservare che la precedente formula $\frac{p(n-k)}{k!}$ vale anche per $k=n-1$ e $k=n$ in quanto

$$\frac{p(n-(n-1))}{(n-1)!}=\frac{1}{(n-1)!}\sum_{h=0}^{n-(n-1)}(-1)^h\frac{1}{h!}=\frac{1}{(n-1)!}\left[1-\frac{1}{1!}\right]=0,$$

e

$$\frac{p(n-n)}{n!}=\frac{1}{n!}\sum_{h=0}^{n-n}(-1)^h\frac{1}{h!}=\frac{1}{n!}(-1)^0\frac{1}{0!}=\frac{1}{n!}.$$

2.1 Verso le variabili aleatorie con infiniti valori.

Per iniziare vediamo come si arriva in modo naturale a variabili aleatorie che possono assumere un'infinità di valori. Nella prossima Sezione vedremo come si formalizza questo approccio prendendo uno spazio $\Omega = \{\omega_i,\ i \geq 1\}$ numerabile.

A partire da variabili aleatorie Binomiali verso le variabili di Poisson
Partiremo dal caso del limite di una variabile aleatoria S_n di tipo binomiale, ossia come modello limite del numero dei successi in n prove ripetute (ossia indipendenti e tutte con la stessa probabilità) quando però la probabilità di successo è piccola rispetto al numero delle prove.

Qui diamo solo l'enunciato (per la dimostrazione vedere il Teorema 14.1 in [SN] o equivalentemente, in queste note, il Teorema 4.1 a pagina 30)

Si dimostra che, se n è grande, se ϑ è piccolo, e, posto

$$\boxed{\lambda := n\vartheta,}$$

se λ non è troppo grande (empiricamente se $\vartheta < 0.05$ e $\lambda < 20$)
allora, per ogni $k \geq 0$ non troppo grande,

$$\boxed{\mathbb{P}(S_n = k) \text{ è approssimata da } \frac{\lambda^k}{k!}e^{-\lambda}.}$$

Qui sotto ne riportiamo l'enunciato:

TEOREMA DI APPROSSIMAZIONE DI POISSON: se $S_n \sim Bin(n, \vartheta_n)$, con $\vartheta_n = \frac{\lambda}{n}$ allora vale

$$\lim_{n\to\infty} \mathbb{P}(S_n = k) = \mathbb{P}(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad \text{per ogni } k \geq 0.$$

Questo risultato porta alla definizione di variabili aleatorie X con distribuzione di Poisson di parametro $\lambda > 0$ come variabili aleatorie a valori in $\mathbb{N} \cup \{0\}$ e tali che

$$\boxed{\mathbb{P}(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad \text{per ogni } k \geq 0.}$$

Vedremo inoltre che $\mathbb{E}(X) = \mathbb{E}(S_n) = n\vartheta_n = n \cdot \frac{\lambda}{n} = \lambda$ e che $Var(X) = \lambda = \lim_{n\to\infty} Var(S_n) = \lim_{n\to\infty} n \cdot \theta_n(1-\theta_n) = \lambda\left(1 - \frac{\lambda}{n}\right) = \lambda$.

Per una formulazione migliore del Teorema di approssimazione di Poisson vedere l'enunciato del Teorema di Le Cam in queste note, a pagina 28, dopo la distribuzione di Poisson di parametro λ, che assicura che l'errore di approssimazione è minore di $\lambda\vartheta = n\vartheta^2$.

A partire da variabili aleatorie Tempo di primo successo troncato verso le variabili aleatorie Geometriche
Ricordiamo che se $A_1, A_2, ..., A_n$ sono n eventi indipendenti e tutti con la stessa probabilità p, la variabile X_n tempo di primo successo troncato è definita come

$$\{X_n = k\} = \overline{A}_1 \cap \overline{A}_2 \cap .... \cap \overline{A}_{k-1} \cap A_k \qquad k = 1, 2, ..., n, \qquad \{X_n = n+1\} = \overline{A}_1 \cap \overline{A}_2 \cap .... \cap \overline{A}_{n-1} \cap \overline{A}_n$$

di modo che

$$\mathbb{P}(X_n = k) = \mathbb{P}(\overline{A}_1)\mathbb{P}(\overline{A}_2)\cdots\mathbb{P}(\overline{A}_{k-1})\mathbb{P}(A_k) = (1-p)^{k-1}p, \qquad k = 1, 2, ..., n, \quad \mathbb{P}(X_n = n+1) = \mathbb{P}(\overline{A}_1 \cap \overline{A}_2 \cap .... \cap \overline{A}_{n-1} \cap \overline{A}_n)$$

Mandando n all'infinito, si ottiene che

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = (1-p)^{k-1}p, \quad k \geq 1.$$

Inoltre, da

$$\begin{aligned}\mathbb{E}(X_n) &= \mathbb{P}(X_n > 0) + \mathbb{P}(X_n > 1) + \mathbb{P}(X_n > 2) + ... + \mathbb{P}(X_n > n) \\ &= 1 + (1-p) + ... + (1-p)^k + ... + (1-p)^n = \frac{1-(1-p)^{n+1}}{1-(1-p)} = \frac{1-(1-p)^{n+1}}{p}.\end{aligned}$$

si ottiene che

$$\lim_{n\to\infty}\mathbb{E}(X_n)=\lim_{n\to\infty}\sum_{k=0}^{n}\mathbb{P}(X_n>k)=\lim_{n\to\infty}\frac{1-(1-p)^{n+1}}{p}=\frac{1}{p}\qquad\text{[in quanto } (1-p)^{n+1}\text{ tende a zero]}$$

Da queste osservazioni nascono le v.a. Geometriche di parametro p (vedere la Sezione successiva)

A partire dal Numero delle Concordanze verso una variabile aleatoria di Poisson di parametro 1
Abbiamo visto che, se S_n indica il numero delle concordanze, la sua densità discreta è data da

$$p_{S_n}(k)=\mathbb{P}(S_n=k)=\frac{p(n-k)}{k!}=\frac{1}{k!}\left(\sum_{h=0}^{n-k}(-1)^h\frac{1}{h!}\right),\quad k=0,1,...,n-2$$

$$p_{S_n}(n-1)=\mathbb{P}(S_n=n-1)=0,\quad p_{S_n}(n)=\mathbb{P}(S_n=n)=\frac{1}{n!}$$

Si verifica immediatamente che se n è grande, e k è piccolo rispetto ad n (o meglio se n è grande rispetto a k) allora $\mathbb{P}(S_n=k)$ *è approssimata da* $\frac{1}{k!}e^{-1}$*, che, come vedremo dopo aver introdotto le v.a. di Poisson, coincide con la densità discreta di una v.a. X di Poisson di parametro 1: infatti*[7]

$$\lim_{n\to\infty}\frac{1}{k!}\sum_{h=0}^{n-k}(-1)^h\frac{1}{h!}=\frac{1}{k!}\sum_{h=0}^{\infty}(-1)^h\frac{1}{h!}=\frac{1}{k!}e^{-1}$$

Inoltre, come abbiamo visto, $\mathbb{E}(S_n)=Var(S_n)=1$*, e, come vedremo nella prossima sezione,* $\mathbb{E}(S_n)=\mathbb{E}(X)=1$ *e* $Var(S_n)=1=Var(X)=1$.

[7] Basta ricordare che la serie esponenziale $\sum_{k=0}^{\infty}\frac{x^k}{k!}$ converge per ogni $x\in\mathbb{R}$, che $\sum_{k=0}^{\infty}\frac{x^k}{k!}=e^x$ (in quanto è la serie di Mac Laurin di e^x) e infine prendere $x=-1$.

3 Spazi di probabilità numerabili

Richiamiamo qui la definizione di spazio di probabilità generale.

> Uno spazio di probabilità numerabile con la probabilità numerabilmente additiva (anche detta probabilità σ-additiva) è una terna $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ con $\Omega = \{\omega_i, i \in \mathbb{N}\}$ un insieme numerabile, cioè che si può mettere in corrispondenza biunivoca con l'insieme $\mathbb{N}$ dei numeri naturali. e $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ è una funzione tale che
>
> ***i)*** per ogni $A \subset \Omega$, $\mathbb{P}(A) \geq 0$ (è una condizione ridondante: è inclusa nella richiesta che $\mathbb{P}(A) \in [0,1]$)
>
> ***ii)*** $\mathbb{P}(\Omega) = 1$ (normalizzazione)
>
> ***iii)*** se A_k, per $k \geq 1$, è una successione di eventi incompatibili a due a due, ossia $A_h \cap A_k = \emptyset$, allora
>
> $$\mathbb{P}\big(\cup_{k=1}^{\infty} A_k\big) = \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$
>
> La proprietà ***iii)*** è detta additività numerabile, o σ-additività (che si legge "*sigma-additività*")

In tali spazi vale anche la proprietà di additività finita e quindi tutte le proprietà delle probabilità viste negli spazi finiti.
Ad esempio si ha che ogni $\mathbb{P}$ è caratterizzata dalla funzione $p(\omega_i) = \mathbb{P}(\{\omega_i\})$, $i \geq 1$, per la quale vale che

$$\boxed{\mathbb{P}(E) = \sum_{i:\omega_i \in E} p(\omega_i)} \qquad \text{e inoltre} \qquad \boxed{p(\omega_i) \geq 0, \quad \forall\, i \geq 1, \quad \text{e} \quad \sum_{i=1}^{\infty} p(\omega_i) = 1.}$$

Viceversa data una funzione $\omega_i \in \Omega \mapsto p(\omega_i)$, tale che $p(\omega_i) \geq 0, \quad \forall\, i \geq 1$, e $\sum_{i=1}^{\infty} p(\omega_i) = 1$, tale funzione definisce una probabilità numerabilmente additiva su Ω.
In particolare continuano a valere la formula delle probabilità totali per partizioni finite, ma valgono anche per partizioni infinite numerabili.

3.1 Variabili aleatorie con un'infinità numerabile di valori

Analogamente per le variabili aleatorie X che possono assumere un'infinità di valori, ossia funzioni $X : \Omega \to X(\Omega) \subset \mathbb{R}$, $\omega_i \mapsto X(\omega_i)$, per $i \geq 1$, con $X(\Omega) = \{x_k, k \geq 1\}$ numerabile, la distribuzione di X è individuata da una funzione

$$p_X : X(\Omega) \to [0,1]; \quad x_k \mapsto p_X(x_k) = \mathbb{P}(X = x_k)$$

La funzione p_X deve soddisfare le condizioni $\boxed{p_X(x_k) \geq 0, \qquad \sum_{k=1}^{\infty} p_X(x_k) = 1.}$

3.1.1 Esempi

Una variabile aleatoria discreta con infiniti valori possibili nasce naturalmente nel seguente esempio

Esempio 3.1 (Tempo di primo successo)**.** *Sia $\{E_1, E_2, ...\}$ una successione di eventi e siano $X_1 = \mathbf{1}_{E_1}, X_2 = \mathbf{1}_{E_2}, ...$ le corrispondenti variabili indicatrici.*

Consideriamo la variabile aleatoria

$$T \equiv \inf\{n \geq 1 : X_n = 1\},$$

per la quale si ha

$$\{T = n\} = \{X_1 = 0, X_2 = 0, ..., X_{n-1} = 0, X_n = 1\} = \overline{E}_1 \cap \overline{E}_2 \cap \cdots \cap \overline{E}_{n-1} \cap E_n$$

e possiamo interpretare T come il numero (aleatorio) di prove necessarie fino ad ottenere il primo successo, nella successione di prove $\{E_1, E_2, ...\}$.

L'insieme dei valori possibili per T coincide con l'insieme dei numeri naturali $\mathbb{N} = \{1, 2, ...\}$.

Esempio 3.2. *In un test di affidabilità un'apparecchiatura elettrica appena prodotta viene testata lasciandola funzionare ininterrottamente fino a quando non smetta di funzionare a causa di qualche guasto. Indicando con T la lunghezza complessiva del tempo di durata realizzato dall'apparecchiatura, se calcoliamo il tempo in giorni otteniamo una variabile aleatoria i cui valori possibili sono in linea di principio tutti i numeri interi.*
*Si osservi che invece, se si misura (*o se si potesse misurare*) il tempo con precisione assoluta, allora i valori possibili sono (*sarebbero*) in linea di principio tutti i valori reali positivi.*

La definizione di valore atteso è del tutto simile al caso di spazi finiti, ma al posto di una somma finita si usa la somma di una serie: di conseguenza si aggiunge una condizione di convergenza assoluta che garantisce che la somma della serie non dipenda dall'ordine in cui viene fatta la somma:

Definizione 3.1. *Il valore atteso di una variabile aleatorie X in uno spazio di probabilità numerabile si definisce come*

$$\boxed{\mathbb{E}(X) \equiv \sum_{i=1}^{\infty} p(\omega_i) X(\omega_i),}$$

purché la serie sia assolutamente convergente, ossia purché $\boxed{\sum_{i=1}^{\infty} p(\omega_i)|X(\omega_i)| < \infty,}$ *in modo che la somma della serie non dipenda dall'ordine in cui viene calcolata.*

Alternativamente (ed equivalentemente) si può definire anche attraverso la densità discreta di X:

$$\boxed{\mathbb{E}(X) = \sum_{k\geq 1} x_k p_X(x_k) = \sum_{k\geq 1} x_k \mathbb{P}(X = x_k)}$$

purché

$$\sum_{k\geq 1} |x_k| p_X(x_k) = \sum_{k\geq 1} |x_k| \mathbb{P}(X = x_k) < \infty.$$

Vale inoltre, che se $Z = h(X)$, allora $\boxed{\mathbb{E}(Z) = \mathbb{E}(h(X)) = \sum_{k\geq 1}^{n} h(x_k) p_X(x_k) = \sum_{k\geq 1} h(x_k) \mathbb{P}(X = x_k),}$

purché

$$\sum_{k\geq 1} |h(x_k)| p_X(x_k) = \sum_{k\geq 1} |h(x_k)| \mathbb{P}(X = x_k) < \infty.$$

e

$$p_Z : Z(\Omega) \to [0,1] \subset \mathbb{R}, \quad z_j \mapsto \boxed{p_Z(z_j) = \mathbb{P}(Z = z_j) = \sum_{k:h(x_k)=z_j} p_X(x_k) = \sum_{k:h(x_k)=z_j} \mathbb{P}(X = x_k).}$$

Per variabili aleatorie X a valori in $\{0\} \cup \mathbb{N}$ si ha anche (vedere la nota alla pagina seguente)

$$\boxed{\textbf{se } \mathbb{E}(X) < \infty \quad \textbf{allora} \quad \mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k).}$$

Infine, come visto in precedenza per varibaili aleatorie in spazi di probabilità finiti, anche in spazi di probabilità numerabili, data una partizione finite, vale la formula del valore atteso totale. Inoltre vale anche anche per partizioni numerabili:
data una partizione numerabile $\{H_j, j \geq 1\}$ con $\mathbb{P}(H_j) > 0$, $j \geq 1$ e una variabile aleatoria X

$$\boxed{\mathbb{E}(X) = \sum_{j\geq 1} \mathbb{E}(X|H_j)\mathbb{P}(H_j)}, \quad \text{dove } \mathbb{E}(X|H_j) = \sum_{k\geq 1} x_k \mathbb{P}(X = x_k|H_j),$$

sempre sotto la condizione di convergenza assoluta ossia che

$$\boxed{\mathbb{E}(|X|) = \sum_{j\geq 1} \mathbb{E}(|X||H_j)\mathbb{P}(H_j)}, \quad \text{dove } \mathbb{E}(X|H_j) = \sum_{k\geq 1} |x_k| \mathbb{P}(X = x_k|H_j) < \infty,$$

condizione che assicura che la somma della serie a doppio indice $\sum_{j\geq 1}\sum_{k\geq 1} |x_k|\mathbb{P}(X = x_k|H_j)\mathbb{P}(H_j)$ non dipenda dall'ordine in cui viene fatta la somma.

Se $X(\Omega) \subseteq \{0,1,2,...,n\}$**, allora** $\mathbb{E}(X) = \sum_{k=0}^{n-1} \mathbb{P}(X > k)$

NOTA BENE: è obbligatoria solo la dimostrazione nel caso finito, vedere anche la Lezione 7 in [SN]
Procedendo come nella dimostrazione nel caso finito (vedere anche la ***Proposizione* 9.8 in [SN]** e l'osservazione in questa nota) si ottiene che, per ogni variabile aleatoria X a valori in $\mathbb{Z}^+$, e quindi anche per $X(\Omega) \subseteq \{0,1,2,...,n\}$,

$$\sum_{k=0}^{n} k\,\mathbb{P}(X=k) = \sum_{k=0}^{n} k\,[\mathbb{P}(X>k-1) - \mathbb{P}(X>k)] = \sum_{k=1}^{n} k\,\mathbb{P}(X>k-1) - \sum_{k=1}^{n} k\,\mathbb{P}(X>k)$$
$$= \sum_{h=0}^{n-1} (h+1)\,\mathbb{P}(X>h) - \sum_{k=0}^{n-1} k\,\mathbb{P}(X>k) - n\,\mathbb{P}(X>n) = \sum_{k=0}^{n-1} \mathbb{P}(X>k) - n\,\mathbb{P}(X>n).$$

Osservazione: ***nel caso di una variabile aleatoria valori in*** $\{0,1,2,...,n\}$ ***ovviamente*** $\mathbb{P}(X>n) = 0$ ***e quindi si ha***

$$\boxed{\mathbb{E}(X) = \sum_{k=0}^{n} k\,\mathbb{P}(X=k) = \sum_{k=0}^{n-1} \mathbb{P}(X>k)}$$

e ciò conclude la dimostrazione del caso finito.

Se $X(\Omega) \subseteq \{0\} \cup \mathbb{N}$**, allora** $\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k)$

NOTA BENE: la dimostrazione nel caso infinito è facoltativa.
Sia X una variabile aleatoria a valori in $\mathbb{Z}^+ = \{0\} \cup \mathbb{N}$, Ovviamente in tale caso $\mathbb{E}(X) = \sum_{k=0}^{\infty} k\mathbb{P}(X=k)$, e la serie $\sum_{k=0}^{\infty} k\mathbb{P}(X=k)$ è a termini tutti non negativi e quindi o converge o diverge. Lo stesso vale per la serie $\sum_{k=0}^{\infty} \mathbb{P}(X>k)$.
La condizione $\mathbb{E}(X) < \infty$ equivale alla convergenza a zero della successione della serie resto:

$$\sum_{k=0}^{\infty} k\mathbb{P}(X=k) < \infty \quad \Leftrightarrow \quad \lim_{n\to\infty} \sum_{k=n}^{\infty} k\mathbb{P}(X=k) = 0,$$

e implica la condizione

$$\lim_{n\to\infty} n\,\mathbb{P}(X>n) = 0,$$

in quanto

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

e ciò mostra che se $\mathbb{E}(X) < \infty$ allora $\lim_{n\to\infty} n\,\mathbb{P}(X>n) = 0$.
A questo punto, per ottenere che $\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X>k)$, basta osservare che, se $\mathbb{E}(X) < \infty$ allora, per quanto visto nella nota precedente,

$$\mathbb{E}(X) = \lim_{n\to\infty} \sum_{k=0}^{n} k\,\mathbb{P}(X=k) = \lim_{n\to\infty} \Big[\sum_{k=0}^{n-1} \mathbb{P}(X>k) - n\,\mathbb{P}(X>n) \Big]$$
$$= \sum_{k=0}^{\infty} \mathbb{P}(X>k) - \underbrace{\lim_{n\to\infty} n\,\mathbb{P}(X>n)}_{\longrightarrow_{n\to\infty} 0} = \sum_{k=0}^{\infty} \mathbb{P}(X>k),$$

e ciò conclude la dimostrazione.
Tuttavia, vale la pena osservare che, sempre per variabili aleatorie a valori in $\mathbb{Z}^+$, la condizione $\mathbb{E}(X) < \infty$ non è necessaria affinché $\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X>k)$ infatti (sia quando $\mathbb{E}(X) < \infty$ sia quando $\mathbb{E}(X) = \infty$)

$$\mathbb{E}(X) \geq \sum_{k=0}^{n} k\,\mathbb{P}(X=k) + \sum_{k=n+1}^{\infty} n\,\mathbb{P}(X>k)$$
$$= \Big[\sum_{k=0}^{n-1} \mathbb{P}(X>k) - n\,\mathbb{P}(X>n) \Big] + n\,\mathbb{P}(X>n) = \sum_{k=0}^{n-1} \mathbb{P}(X>k).$$

e ciò mostra che se $\sum_{k=0}^{\infty} \mathbb{P}(X>k) = \infty$ allora anche $\mathbb{E}(X) = \infty$.

4 Variabili aleatorie discrete con un numero infinito di valori: casi notevoli

4.1 Variabili aleatorie Geometriche

Si dice che X è una variabile aleatoria Geometrica di parametro $p \in (0,1)$, in simboli $X \sim Geom(p)$ (a partire da 1) se e solo se

$$p_X(k) = \mathbb{P}(X = k) = (1-p)^{k-1} p, \quad k = 1,2,3,...$$

Il prototipo delle variabili aleatorie Geometriche di parametro p si ottiene come nel precedente esempio delle variabili aleatorie geometriche troncate, come tempo di primo successo, con l'aggiunta della condizione che invece di un numero finito n di eventi indipendenti tutti con probabilità p, si ha una successione di eventi E_ℓ, $\ell \geq 1$ ***indipendenti***,

ossia | comunque fissato $n \geq 2$ gli eventi E_1, E_2,...,E_n sono indipendenti | e inoltre | $\mathbb{P}(E_i) = p, \forall\, i \geq 1$ |

Come vedremo a breve (vedere le note a pagina 27) utilizzando le proprietà delle serie di potenze, si ottiene che $\mathbb{E}(X) = \frac{1}{p}, \quad Var(X) = \frac{1-p}{p^2},$ ma prima ne vedremo qui sotto un'altra dimostrazione che utilizza la formula del valore atteso totale e la proprietà di mancanza di memoria delle variabili aleatorie geometriche.

Inoltre $\mathbb{P}(X > n) = (1-p)^n,\ n \geq 0, \quad \Leftrightarrow\ \{X > n\} = \cap_{i=1}^n \overline{E_i}$

da cui $\mathbb{P}(X > n) = \mathbb{P}\Big(\bigcap_{i=1}^n \overline{E_i}\Big) \overset{indip}{=} \prod_{i=1}^n \mathbb{P}(\overline{E}_i) = (1-p)^n.$

4.1.1 Valore atteso di $X \sim Geom(p)$ con le probabilità di sopravvivenza: ossia $\mathbb{P}(X > k)$, $k \geq 0$.

Vale la pena ricordare che si può calcolare il valore atteso di X anche usando il fatto che, poiché X è una variabile aleatoria a valori in $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$ il suo valore atteso si può calcolare nel seguente modo

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k) = \sum_{k=0}^{\infty} (1-p)^k = \frac{1}{1-(1-p)} = \frac{1}{p}$$

dove abbiamo usato il fatto che $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ per ogni x tale che $|x| < 1$.

4.1.2 Le variabili geometriche hanno la proprietà della mancanza di memoria

La proprietà di mancanza di memoria per una variabile aleatorie X a valori in $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$ significa che

$$\boxed{\mathbb{P}(X > h+k | X > k) = \mathbb{P}(X > h), \quad h,k \geq 0,} \quad \Leftrightarrow \quad \boxed{\mathbb{P}(X - k > h | X > k) = \mathbb{P}(X > h), h,k \geq 0}$$

ossia la distribuzione di $X - k$, condizionatamente a sapere che $X > h$, è la stessa di X:

$$\boxed{\mathbb{P}(X - k = j | X > k) = \mathbb{P}(X = j), \quad j,k \geq 0,} \quad \Leftrightarrow \quad \boxed{\mathbb{P}(X - k > h | X > k) = \mathbb{P}(X > h), h,k \geq 0},$$

L'equivalenza tra queste due proprietà si deduce immediatamente ricordando che per una variabile aleatoria Y a valori interi si ha

$$\mathbb{P}(Y > h) = \sum_{j>h} \mathbb{P}(Y = j) \qquad \text{e} \qquad \mathbb{P}(Y = j) = \mathbb{P}(Y > j-1) - \mathbb{P}(Y = j).$$

In altre parole la proprietà di mancanza di memoria significa: **sapendo non c'è stato alcun successo nelle prime k prove**, il numero $X - k$ di prove che si devono ancora aspettare per ottenere il primo successo, ha la stessa distribuzione di X, cioè **è come se si ricominciasse daccapo** (come vedremo per $X \sim Geom(p)$, è essenziale che gli eventi E_i, $i \geq 1$ siano indipendenti e tutti con la stessa probabilità).
Per $X \sim Geom(p)$, la proprietà di mancanza di memoria si deduce immediatamente. Infatti:

$$\mathbb{P}(X > h+k | X > k) = \frac{\mathbb{P}(X > h+k, X > k)}{\mathbb{P}(X > k)} = \frac{\mathbb{P}(X > h+k)}{\mathbb{P}(X > k)} = \frac{(1-p)^{k+h}}{(1-p)^k} = (1-p)^h = \mathbb{P}(X > h).$$

L'altra versione della proprietà di mancanza di memoria si può dedurre anche direttamente:

$$\mathbb{P}(X = j+k | X > k) = \frac{\mathbb{P}(X = j+k, X > k)}{\mathbb{P}(X > k)} = \frac{\mathbb{P}(X = j+k)}{\mathbb{P}(X > k)} = \frac{(1-p)^{k+j-1}}{(1-p)^k} = (1-p)^{j-1} p = \mathbb{P}(X = j).$$

4.1.3 Valore atteso e varianza di $X \sim Geom(p)$ con la proprietà di mancanza di memoria e la formula del valore atteso totale.

In particolare per $k=1$, $\mathbb{P}(X > h+1|X>1) = \mathbb{P}(X>h)$, ossia

$$\mathbb{P}(X-1>h|X>1) = \mathbb{P}(X>h), \quad \forall\, h \geq 0, \quad \Leftrightarrow \quad \mathbb{P}(X-1=j|X>1) = \mathbb{P}(X=j), \quad \forall\, j \geq 0.$$

da cui, **condizionatamente a** $\{X>1\}$, la variabile aleatoria $\boxed{X' := X-1}$, ha la stessa distribuzione di X. Da questa proprietà si ottiene un altro modo per ottenere il valore atteso di X (oltre che con le proprietà delle serie): posto $H = \{X=1\} = E_1$ e quindi $\overline{H} = \{X>1\} = \overline{E}_1$, grazie alla formula del valore atteso totale

$$\mathbb{E}(X) = \mathbb{P}(H)\mathbb{E}(X|H) + \mathbb{P}(\overline{H})\mathbb{E}(X|\overline{H}) = p\,\mathbb{E}(X|X=1) + (1-p)\,\mathbb{E}(X|X>1)$$

e quindi, tenendo conto che, ovviamente, $\mathbb{E}(X|X=1)=1$, e che, avendo posto $X' := X-1$,

$$\mathbb{E}(X|X>1) = \mathbb{E}(1+X'|X>1) = 1+\mathbb{E}(X'|X>1) = 1+\mathbb{E}(X) = 1+\mu,$$

dove l'ultima uguaglianza dipende dal fatto che, come osservato prima, **condizionatamente a** $\{X>1\}$, la variabile aleatoria $X' := X-1$, ha la stessa distribuzione di X, e quindi anche il valore atteso condizionato $\mathbb{E}(X'|X>1)$ coincide con $\mathbb{E}(X)$. Alla fine, posto $\boxed{\mu := \mathbb{E}(X) = \mathbb{E}(X'|X>1)}$ si ha che

$$\boxed{\mathbb{E}(X) = p\,\mathbb{E}(X|X=1) + (1-p)\,\mathbb{E}(X|X>1) \quad \Leftrightarrow \quad \mu = p\cdot 1 + (1-p)(1+\mu) \quad \Leftrightarrow \quad \mu = \frac{1}{p}.}$$

In modo simile si può calcolare $\mathbb{E}(X^2)$, infatti, usando di nuovo la formula del valore atteso totale, e il fatto che $X = (X-1)+1$ e quindi $X^2 = (X-1)^2 + 2(X-1) + 1 = (X')^2 + 2X' + 1$, si ottiene

$$\begin{aligned}
\mathbb{E}(X^2) &= \mathbb{P}(H)\mathbb{E}(X^2|H) + \mathbb{P}(\overline{H})\mathbb{E}(X^2|\overline{H}) = p\,\mathbb{E}(X^2|X=1) + (1-p)\,\mathbb{E}(X^2|X>1) \\
&= p\,\mathbb{E}(X^2|X=1) + (1-p)\,\mathbb{E}((X'+1)^2|X>1) = p\cdot 1 + (1-p)\,\mathbb{E}((X')^2 + 2X' + 1|X>1) \\
&= p + (1-p)\,\left[\mathbb{E}((X')^2|X>1) + 2\mathbb{E}(X'|X>1) + 1\right], \quad \text{e quindi, posto } \boxed{\mu_2 := \mathbb{E}(X^2) = \mathbb{E}((X')^2|X>1)}
\end{aligned}$$

$$\mu_2 = p + (1-p)(\mu_2 + 2\mu + 1) \Leftrightarrow \mu_2 = p + \mu_2 - \mu_2 p + 2(1-p)\frac{1}{p} + 1 - p \Leftrightarrow \cancel{\mu_2} - \cancel{\mu_2} + \mu_2 p = \frac{2}{p} - 2 + 1 \Leftrightarrow \mu_2 = \frac{2-p}{p^2}$$

e quindi $$\boxed{Var(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}}$$

Si noti che, sia nel caso del calcolo del valore atteso μ di X che per il calcolo del valore atteso μ_2 di X^2, abbiamo dato per scontato che μ e μ_2 fossero finiti: solo in questo caso possiamo scrivere, ad esempio $\mu(1-p) = \mu - p\mu$. In un senso più preciso, e tenendo conto che per ogni variabile aleatoria Y a valori in $\mathbb{Z}_+ = \mathbb{N}\cup\{0\}$ si ha che $\mathbb{E}(Y)$ esiste sempre, ma può essere infinito, abbiamo solo dimostrato che se $X \sim Geom(p)$ ammette valore atteso finito, allora $\mathbb{E}(X) = 1/p$.

4.1.4 Valore atteso di $X \sim Geom(p)$ con le serie di potenze.

Infine il valore atteso di X e la sua varianza si possono calcolare anche usando la definizione e le proprietà delle serie di potenze: ossia se

$$\sum_{k=0}^{\infty} a_k x^k$$

ha raggio di convergenza r, allora posto $f(x)$ la funzione $x \mapsto f(x) := \sum_{k=0}^{\infty} a_k x^k$, definita per $|x| < r$, ammette le derivate di ogni ordine e si ha, sempre per $|x| < r$,

$$\frac{d^k}{dx^k} f(x) = \frac{d^k}{dx^k} \sum_{k=0}^{\infty} a_k x^k = \sum_{k=0}^{\infty} \frac{d^k}{dx^k} a_k x^k.$$

Valore atteso di una v.a. $T \sim Geom(\vartheta)$ con le serie di potenze

$$\mathbb{E}(T) = \sum_{k=1}^{\infty} k\mathbb{P}(T=k) = \sum_{k=1}^{\infty} k\vartheta(1-\vartheta)^{k-1}.$$

Considerando che, ovviamente, per $k=0$ si ha $k\vartheta(1-\vartheta)^{k-1}=0$, possiamo considerare la somma della serie con $k=0$ incluso: ossia

$$\mathbb{E}(T) = \sum_{k=0}^{\infty} k\vartheta(1-\vartheta)^{k-1} = \sum_{k=0}^{\infty} \vartheta \frac{d}{dx}x^k\Big|_{x=1-\vartheta} = \vartheta \sum_{k=0}^{\infty} \frac{d}{dx}x^k\Big|_{x=1-\vartheta}$$

Per le proprietà delle serie di potenze, sappiamo che $\sum_{k=0}^{\infty} \frac{d}{dx}x^k = \frac{d}{dx}\sum_{k=0}^{\infty} x^k$ *[essendo* $1-\vartheta \in (0,1)$ *possiamo supporre che* x *vari in un intervallo chiuso strettamente contenuto nell'intervallo* $(-1,1)$*],* da cui

$$\mathbb{E}(T) = \vartheta \frac{d}{dx}\frac{1}{(1-x)}\Big|_{x=1-\vartheta} = \vartheta \frac{1}{(1-x)^2}\Big|_{x=1-\vartheta} = \vartheta \frac{1}{(1-(1-\vartheta))^2} = \frac{1}{\vartheta}.$$

Varianza di una v.a. $T \sim Geom(\vartheta)$

$$Var(T) = \mathbb{E}\Big[\big(T-\mathbb{E}(T)\big)^2\Big] = \cdots = \mathbb{E}(T^2) - \big(\mathbb{E}(T)\big)^2 = \frac{2(1-\vartheta)+\vartheta}{\vartheta^2} - \frac{1}{\vartheta^2} = \frac{2(1-\vartheta)+\vartheta-1}{\vartheta^2} = \frac{(1-\vartheta)}{\vartheta^2}$$

Infatti, per calcolare $\mathbb{E}(T^2)$ osserviamo che $T^2 = T(T-1)+T$ ossia che $\mathbb{E}(T^2) = \mathbb{E}[T(T-1)+T] = \mathbb{E}[T(T-1)] + \mathbb{E}[T]$:

$$\begin{aligned}\mathbb{E}(T^2) &= \sum_{k=1}^{\infty} k^2\mathbb{P}(T=k) = \sum_{k=1}^{\infty} [k(k-1)+k]\vartheta(1-\vartheta)^{k-1} \\ &= \overbrace{\sum_{k=1}^{\infty} k(k-1)\vartheta(1-\vartheta)^{k-1}}^{\mathbb{E}[T(T-1)]} + \overbrace{\sum_{k=1}^{\infty} k\vartheta(1-\vartheta)^{k-1}}^{=\mathbb{E}(T)=\frac{1}{\vartheta}},\end{aligned}$$

in quanto $\sum_k (a_k+b_k) = \sum_k a_k + \sum_k b_k$*, [inoltre, se* $a_k, b_k \geq 0$*, non ci sono problemi di convergenza].*
Considerando che, ovviamente, per $k=0$ si ha $k\vartheta(1-\vartheta)^{k-1}=0$, possiamo considerare la somma della serie con $k=0$ incluso: ossia

$$\begin{aligned}\mathbb{E}(T^2) &= \sum_{k=0}^{\infty} k(k-1)\vartheta(1-\vartheta)^{k-1} + \frac{1}{\vartheta} = \sum_{k=0}^{\infty} k(k-1)\vartheta(1-\vartheta)(1-\vartheta)^{k-2} + \frac{1}{\vartheta} \\ &= \sum_{k=0}^{\infty} \vartheta(1-\vartheta)\frac{d^2}{dx^2}x^k\Big|_{x=1-\vartheta} + \frac{1}{\vartheta} = \vartheta(1-\vartheta)\sum_{k=0}^{\infty} \frac{d^2}{dx^2}x^k\Big|_{x=1-\vartheta} + \frac{1}{\vartheta}.\end{aligned}$$

Di nuovo, per le proprietà delle serie di potenze, sappiamo che $\sum_{k=0}^{\infty} \frac{d^2}{dx^2}x^k = \frac{d^2}{dx^2}\sum_{k=0}^{\infty} x^k$ *[essendo* $1-\vartheta \in (0,1)$ *possiamo supporre che* x *vari in un intervallo chiuso strettamente contenuto nell'intervallo* $(-1,1)$*],* da cui

$$\begin{aligned}\mathbb{E}(T^2) &= \vartheta(1-\vartheta)\frac{d^2}{dx^2}\frac{1}{(1-x)}\Big|_{x=1-\vartheta} + \frac{1}{\vartheta} = \vartheta(1-\vartheta)\frac{2}{(1-x)^3}\Big|_{x=1-\vartheta} + \frac{1}{\vartheta} \\ &= \vartheta(1-\vartheta)\frac{2}{(1-(1-\vartheta))^3} = \frac{2(1-\vartheta)}{\vartheta^2} + \frac{1}{\vartheta} = \frac{2(1-\vartheta)+\vartheta}{\vartheta^2}.\end{aligned}$$

4.2 Variabili aleatorie di Poisson

Una variabile aleatoria di dice che ha distribuzione di Poisson di parametro λ (lambda), dove $\lambda > 0$, in breve

$\boxed{X \sim Poiss(\lambda)}$ (dove $\lambda > 0$) se e solo se $\boxed{p_X(k) = \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},\ k = 0, 1, 2, ...}$

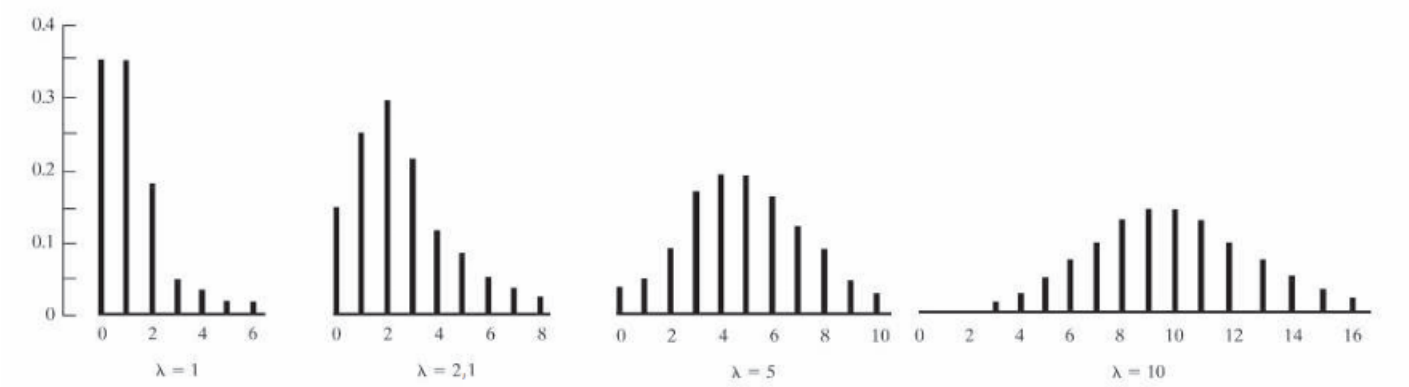
Come visto in Esercizio proposto 14.5 in [SN], e riportato nella pagina seguente,

$$\boxed{\mathbb{E}(X) = \lambda, \quad Var(X) = \lambda.}$$

Vale la seguente formula ricorsiva, di immediata verifica $\boxed{\mathbb{P}(X = k+1) = \frac{\lambda}{k+1} \mathbb{P}(X = k).}$

per cui $\boxed{\mathbb{P}(X = k) = \frac{\lambda}{k} \mathbb{P}(X = k-1) \geq \mathbb{P}(X = k-1) \ \Leftrightarrow\ \frac{\lambda}{k} \geq 1 \ \Leftrightarrow\ k \leq \lambda.}$

ossia, la densità discreta di una v.a. di Poisson cresce prima del valore λ e poi decresce (in modo simile alla densità discreta di una v.a. $Bin(n, p)$. Se $\lambda < 1$ allora la densità discreta decresce sempre, se $\lambda = 1$ allora $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = e^{-1}$ e poi decresce.



Vale inoltre il **Teorema di POISSON (Teorema 14.1 in [SN])**: se $S_n \sim Bin(n, \vartheta_n)$, con $\vartheta_n = \frac{\lambda}{n}$ allora, per ogni $k \geq 0$ vale

$$\boxed{\lim_{n\to\infty} \mathbb{P}(S_n = k) = \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}}$$

(dimostrazione negli Appunti [SN], ma che riportiamo nella pagina seguente)
INOLTRE VALE IL SEGUENTE **TEOREMA DI LE CAM** (senza dimostrazione)
Sia $S_n \sim Bin(n, \vartheta)$ e sia $X \sim Poisson(n\vartheta)$ (si noti che S_n ed X hanno lo stesso valore atteso) allora

$$\sum_{k=0}^{\infty} \left|\mathbb{P}(S_n = k) - \mathbb{P}(X = k)\right| = \sum_{k=0}^{\infty} \left|\mathbb{P}(S_n = k) - \frac{(n\vartheta)^k}{k!} e^{-n\vartheta}\right| \leq n\vartheta^2.$$

Il motivo per cui è interessante è il seguente: qualunque sia I intervallo reale

$$\left|\mathbb{P}(S_n \in I) - \mathbb{P}(X \in I)\right| = \left|\sum_{k\in I} \mathbb{P}(S_n = k) - \sum_{k\in I} \mathbb{P}(X = k)\right| = \left|\sum_{k\in I} \left(\mathbb{P}(S_n = k) - \mathbb{P}(X = k)\right)\right|$$

(per la disuguaglinza triangolare: $|a+b| \leq |a| + |b|$)

$$\leq \sum_{k\in I} \left|\mathbb{P}(S_n = k) - \mathbb{P}(X = k)\right| \leq \sum_{k=0}^{\infty} \left|\mathbb{P}(S_n = k) - \mathbb{P}(X = k)\right| \leq n\vartheta^2$$

ovvero approssimando $\mathbb{P}(S_n \in I)$ con $\mathbb{P}(X \in I)$ si commette un errore che al massimo vale $n\vartheta^2 = \lambda\vartheta$:

$$\mathbb{P}(X \in I) - n\vartheta^2 \leq \mathbb{P}(S_n \in I) \leq \mathbb{P}(X \in I) + n\vartheta^2$$

e ciò ci dà un'indicazione di quanto buona sia l'approssimazione.

4.3 Proprietà delle distribuzioni di Poisson

Date due v.a. X ed Y indipendenti di Poisson di parametri λ e μ, rispettivamente, la somma di X e Y è ancora una v.a. di Poisson di parametro $\lambda+\mu$: del resto, si vede subito che

$$\mathbb{E}(X+Y) = \mathbb{E}(X)+\mathbb{E}(Y) = \lambda+\mu,$$

e quindi $\lambda+\mu$ è l'unico parametro possibile. (Per la definizione di indipendenza per variabili aleatorie vedere la Sezione 1.6, in queste note, mentre per la verifica formale di questa proprietà vedere più sotto l'ESEMPIO 3, a pagina 32, dove si parla della somma di variabili aleatorie indipendenti, o vedere l'Esempio 15.4 in [SN]).

Questo risultato si estende anche al caso di n variabili aleatorie completamente (o globalmente) indipendenti tra loro: in particolare se le variabili aleatorie X_i hanno tutte la stessa distribuzione $Poiss(\lambda)$ allora S_n ha distribuzione $Poiss(n\cdot\lambda)$.

Tuttavia va osservato che per calcolare $\mathbb{P}(S_n \le x)$, per $x \ge 0$, pur avendo a disposizione una formula esatta, ovvero

$$\mathbb{P}(S_n \le x) = \sum_{0\le k\le \lfloor x\rfloor} \mathbb{P}(S_n = k) = \sum_{k=0}^{\lfloor x\rfloor} \frac{(n\cdot\lambda)^k}{k!}\, e^{-n\cdot\lambda},$$

se n è "grande", gli elementi della precedente sommatoria sono composti da fattori molto grandi ($(n\cdot\lambda)^k$) moltiplicati per fattori molto piccoli ($e^{-n\cdot\lambda}$), e che quindi possono essere "scomodi" da calcolare.

Successivamente (Lezione 15 in [SN] o anche il file sull'Approssimazione Normale) vedremo come ottenere un valore approssimato per $\mathbb{P}(S_n \le x)$ anche in questo esempio (approssimazione normale o gaussiana).

Valore atteso e varianza di una v.a. di Poisson:

Mostriamo che, per $X \sim Poiss(\lambda)$, si ha $\mathbb{E}(X) = \lambda$ e $\mathbb{E}(X^2) = \lambda^2+\lambda$, da cui

$$Var(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \lambda^2+\lambda-\lambda^2 = \lambda.$$

In entrambi i casi è fondamentale ricordare che

$$\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x, \quad \text{per ogni } x\in\mathbb{R}.$$

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{\infty} k\,\mathbb{P}(X=k) = \sum_{k=0}^{\infty} k\,\frac{\lambda^k}{k!}\,e^{-\lambda} = \sum_{k=1}^{\infty} k\,\frac{\lambda^k}{k!}\,e^{-\lambda} = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}\,e^{-\lambda}, \qquad \text{(posto } h=k-1\text{)}\\
&= \sum_{h=0}^{\infty} \frac{\lambda^{h+1}}{h!}\,e^{-\lambda} = \sum_{h=0}^{\infty} \frac{\lambda^h\lambda}{h!}\,e^{-\lambda} = \sum_{h=0}^{\infty} \frac{\lambda^h}{h!}\,\lambda e^{-\lambda} = \lambda e^{-\lambda}\sum_{h=0}^{\infty}\frac{\lambda^h}{h!}\\
&= \lambda e^{-\lambda}e^{\lambda} = \lambda.
\end{aligned}$$

In modo del tutto analogo si ottiene che $\mathbb{E}(X^2) = \lambda^2+\lambda$:
considerando che $X^2 = X+X^2-X = X+X(X-1)$, si ha

$$\begin{aligned}
\mathbb{E}\big(X^2\big) &= \mathbb{E}\big(X+X(X-1)\big) = \mathbb{E}\big(X\big)+\mathbb{E}\big(X(X-1)\big)\\
&= \lambda + \sum_{k=0}^{\infty} k(k-1)\,\mathbb{P}(X=k) = \lambda + \sum_{k=0}^{\infty} k(k-1)\,\frac{\lambda^k}{k!}\,e^{-\lambda} = \lambda + \sum_{k=2}^{\infty} k(k-1)\,\frac{\lambda^k}{k!}\,e^{-\lambda}\\
&= \lambda + \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!}\,e^{-\lambda} = \lambda + \sum_{h=0}^{\infty} \frac{\lambda^{h+2}}{h!}\,e^{-\lambda} \qquad \text{(avendo posto } h=k-2\text{)}\\
&= \lambda + \sum_{h=0}^{\infty} \frac{\lambda^h\lambda^2}{h!}\,e^{-\lambda} = \lambda + \sum_{h=0}^{\infty} \frac{\lambda^h}{h!}\,\lambda^2 e^{-\lambda} = \lambda + \lambda^2 e^{-\lambda}\sum_{h=0}^{\infty}\frac{\lambda^h}{h!}\\
&= \lambda + \lambda^2 e^{-\lambda}e^{\lambda} = \lambda+\lambda^2.
\end{aligned}$$

Terminiamo con la dimostrazione del Teorema di Approssimazione di Poisson, ovvero la Legge dei piccoli numeri.

Teorema 4.1 (**Teorema di approssimazione di Poisson**). *Sia* $\lambda > 0$ *un numero reale. Per ogni* $n > \lambda$*, sia* S_n *una variabile aleatoria binomiale* $Bin(n, \lambda/n)$*. Allora si ha*

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

Dimostrazione. Basta osservare che

$$\begin{aligned}\mathbb{P}(S_n = k) &= \binom{n}{k} \vartheta^k (1-\vartheta)^{n-k} = \frac{n!}{k!\,(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!}\, \frac{n!}{(n-k)!}\, \frac{1}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \underset{n\to\infty}{\longrightarrow} \frac{\lambda^k}{k!}\, e^{-\lambda}\end{aligned}$$

in quanto valgono le seguenti tre relazioni

$$(i) \qquad \frac{n!}{(n-k)!}\, \frac{1}{n^k} = \frac{n(n-1)\cdots\big(n-(k-1)\big)}{n^k} = \frac{n}{n} \times \frac{n-1}{n} \times \cdots \times \frac{n-(k-1)}{n} \underset{n\to\infty}{\longrightarrow} 1^k = 1,$$

$$(ii) \qquad \left(1 - \frac{\lambda}{n}\right)^n \underset{n\to\infty}{\longrightarrow} e^{-\lambda},$$

ed infine

$$(iii) \qquad \left(1 - \frac{\lambda}{n}\right)^{-k} \underset{n\to\infty}{\longrightarrow} 1.$$

□

5 Somma di variabili aleatorie discrete indipendenti

Siano X ed Y due v.a. discrete con $X(\Omega)=\{x_k, k\geq 1\}$ e $Y(\Omega)=\{y_h, h\geq 1\}$, allora

$$\mathbb{P}(X+Y=z)=\sum_{h,k\geq 1: x_k+y_h=z}\mathbb{P}(X=x_k,Y=y_h)=\sum_{k\geq 1}\mathbb{P}(X=x_k,Y=z-x_k)=\sum_{h\geq 1}\mathbb{P}(X=z-y_h,Y=y_h)$$

Inoltre X ed Y sono indipendenti se e solo se per ogni scelta degli intervalli I e J vale

$$\mathbb{P}(\{X\in I\}\cap\{Y\in J\})=\mathbb{P}(X\in I,Y\in J)=\mathbb{P}(X\in I)\mathbb{P}(Y\in J)$$

Nel caso di variabili aleatorie discrete ciò equivale a

$$\mathbb{P}(X=x_k,Y=y_h)=\mathbb{P}(X=x_k)\mathbb{P}(Y=y_h)\quad\Leftrightarrow\quad p_{X,Y}(x_k,y_h)=p_X(x_k)p_Y(y_h),\qquad\forall\, k,h\geq 1.$$

e quindi se X ed Y sono indipendenti allora

$$\mathbb{P}(X+Y=z)=\sum_{h,k\geq 1: x_k+y_h=z}\mathbb{P}(X=x_k)\mathbb{P}(Y=y_h)=\sum_{k\geq 1}\mathbb{P}(X=x_k)\mathbb{P}(Y=z-x_k)=\sum_{h\geq 1}\mathbb{P}(X=z-y_h)\mathbb{P}(Y=y_h)$$

ovvero

$$p_{X+Y}(z)=\sum_{k\geq 1}p_X(x_k)p_Y(z-x_k)=\sum_{h\geq 1}p_X(z-y_h)p_Y(y_h)$$

5.1 Somma di due variabili aleatorie uniformi indipendenti

ESEMPIO 1: La somma di due variabili aleatorie discrete X ed Y, indipendenti ed entrambe uniformi in $\{1,2,...,n\}$ ha distribuzione triangolare discreta in $\{2,3,\cdots,2n\}$, ossia (si tratta della generalizzazione dell'Esempio 7.1, nel caso in cui $n=6$). Infatti (scrivendo per semplicità di notazione $X\perp\!\!\!\perp Y$, per indicare che X ed Y sono indipendenti)

$$\mathbb{P}(X+Y=k)=\sum_j\mathbb{P}(X=j,Y=k-j)\overset{X\perp\!\!\!\perp Y}{=}\sum_j\mathbb{P}(X=j)\mathbb{P}(Y=k-j),$$

dove le somme vanno fatte per j tale che $1\leq j\leq n$ e $1\leq k-j\leq n$ (ossia per $k-n\leq j\leq k-1$), ovvero

$$\max\{1,k-n\}\leq j\leq\min\{n,k-1\};$$

quindi, se $k=2,...,n+1$, $\max\{1,k-n\}=1$ e $\min\{n,k-1\}=k-1$, si tratta di fare la somma per $j\in\{1,...,k-1\}$ e di conseguenza:

$$\mathbb{P}(X+Y=k)=\sum_{j=1}^{k-1}\mathbb{P}(X=j)\mathbb{P}(Y=k-j)=\frac{k-1}{n^2},$$

mentre, se $k=n+2,...,2n$, $\max\{1,k-n\}=k-n$ e $\min\{n,k-1\}=n$, si tratta di fare la somma per $j\in\{k-n,...,n\}$ e di conseguenza su $n-(n-k)+1=2n-(k-1)$ addendi:

$$\mathbb{P}(X+Y=k)=\sum_{j=k-n}^{n}\mathbb{P}(X=j)\mathbb{P}(Y=k-j)=\frac{2n-(k-1)}{n^2}.$$

5.2 Somma di due variabili aleatorie Binomiali (con lo stesso parametro θ) indipendenti

ESEMPIO2: La somma di due v.a. X e Y indipendenti, con $X\sim Bin(n,\vartheta)$ e $Y\sim Bin(m,\vartheta)$ è una v.a. $Bin(n+m,\vartheta)$. Per la dimostrazione *analitica*, ossia con i conti, e per approfondimenti, vedere la Sezione 8.1.1 dal titolo **Distribuzione della somma di X e Y e distribuzione di X dato il valore della somma** in [SN], e in particolare la soluzione dell'**Esercizio proposto 8.5** e l'***Osservazione 8.3***, che riportiamo qui, adattata alle notazioni usate in queste note:

Osservazione *Per ottenere che $X+Y$ è una v.a. $Bin(n+m,\vartheta)$ si potrebbe anche ragionare come segue: siano E_j, per $j=1,2,\ldots,m+n$, eventi globalmente indipendenti di probabilità ϑ, cioè tale da formare uno schema di*

Bernoulli. Consideriamo le variabili $X' = \sum_{j=1}^{n} 1_{E_j}$ *e* $Y' = \sum_{j=n+1}^{n+1} 1_{E_j}$. *Per tali variabili aleatorie valgono le seguenti proprietà :*

(a) X' *ha la stessa distribuzione di* X, *cioè* $Bin(n,\vartheta)$,

(b) Y' *ha la stessa distribuzione di* Y, *cioè* $Bin(m,\vartheta)$,

(c) X' *ed* Y' *sono indipendenti*

e quindi (X',Y') **ha la stessa distribuzione congiunta di** (X,Y).

Di conseguenza,

• *da una parte* $Z' := X' + Y' = \sum_{i=1}^{n+m} 1_{E_j}$ *ha la stessa distribuzione di* $Z = X + Y$,

• *mentre dall'altra parte* Z' *ha chiaramente distribuzione binomiale* $Bin(n+m,\vartheta)$*;*

e con ciò si ottiene che anche Z *ha distribuzione binomiale* $Bin(n+m,\vartheta)$.

5.3 Somma di due variabili aleatorie di Poisson indipendenti

ESEMPIO 3: La somma di due v.a. di Poisson INDIPENDENTI $X \sim Poisson(\lambda)$ e $Y \sim Poisson(\mu)$, è ANCORA una v.a. di Poisson e precisamente $X + Y \sim Poisson(\lambda+\mu)$, ossia

$$\mathbb{P}(X+Y=n) = \frac{(\lambda+\mu)^n}{n!} e^{-(\lambda+\mu)}, \quad n \geq 0.$$

INFATTI (si consiglia prima di controllare a mano i casi $k=0$, $k=1$ e $k=2$, e di vedere anche l'Esempio 15.4 in [SN])

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

quindi, tenendo conto che, se $k > n$, *allora* $\mathbb{P}(Y=n-k)=0$, *si ha*

$$= \sum_{k=0}^{n} \mathbb{P}(X=k)\mathbb{P}(Y=n-k) = \sum_{k=0}^{n} \frac{\lambda^k}{k!} e^{-\lambda} \frac{\mu^{n-k}}{(n-k)!} e^{-\mu}$$

tenendo conto che $e^{-\lambda}e^{-\mu} = e^{-(\lambda+\mu)}$, *e moltiplicando e dividento per* $n!$

$$= e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{k=0}^{n} n! \frac{\lambda^k}{k!} \frac{\mu^{n-k}}{(n-k)!} = e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \lambda^k \mu^{n-k} =$$

tenendo conto della formula della potenza del binomio $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$ *si ottiene la tesi, cioè*

$$\mathbb{P}(X+Y=n) = \frac{(\lambda+\mu)^n}{n!} e^{-(\lambda+\mu)}, \quad n \geq 0.$$

Questo risultato si generalizza immediatamente al caso della somma di n variabili aleatorie di Poisson indipendenti, più precisamente

Generalizzazione: se $X_i \sim Poiss(\lambda_i)$, $i = 1,2,..,n$ sono indipendenti allora $X_1 + X_2 + \cdots + X_n$ è ancora una variabile di Poisson di parametro $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.
Tale risultato si dimostra per induzione.

Prima di dare un'idea della dimostrazione della generalizzazione nel caso $n=3$, ricordiamo che l'indipendenza di $X_1,...,X_n$ significa che, comunque scelti J_i intervalli

$$\mathbb{P}\Big(\bigcap_{i=1}^{n} \{X_i \in J_i\}\Big) = \prod_{i=1}^{n} \mathbb{P}\big(\{X_i \in J_i\}\big)$$

o equivalentemente (essendo le X_i a valori interi), comunque scelti k_i interi

$$\mathbb{P}\Big(\bigcap_{i=1}^{n} \{X_i = k_i\}\Big) = \prod_{i=1}^{n} \mathbb{P}\big(\{X_i = k_i\}\big)$$

Idea della dimostrazione nel caso $n = 3$:
• Se X_1, X_2, X_3 sono indipendenti allora anche $X_1 + X_2$ e X_3 sono indipendenti.
• Se X_1, X_2, X_3 sono variabili aleatorie indipendenti di Poisson di parametri $\lambda_1, \lambda_2, \lambda_3$, rispettivamente, allora $X_1 + X_2$ è di Poisson di parametro $\lambda_1 + \lambda_2$.

Basta allora solo osservare che la somma

$$X_1 + X_2 + X_3 = (X_1 + X_2) + X_3$$

è la somma di due variabili aleatorie indipendenti e di Poisson, e quindi è di Poisson, per il caso con $n = 2$ dimostrato in precedenza.

5.4 Somma di n variabili aleatorie Geometriche (con lo stesso parametro p) indipendenti e tempi di n successo per eventi indipendenti e tutti con la stessa probabilità

ESEMPIO 4: Somma di n variabili geometriche Δ_i, $i = 1, ..., n$, indipendenti tutte di parametro p (vedere l'Esempio 15.3 in [SN], e riportato qui sotto) e il tempo T_n di n-simo successo in prove ripetute (ovvero uno schema di Bernoulli infinito con probabilità p.)
Si dimostra che la somma di n variabili aleatorie indipendenti $X_1, ..., X_n$, tutte Geom(p) e il tempo T_n, definito come il tempo di n-simo successo in uno schema di Bernoulli infinito con probabilità p, hanno la stessa distribuzione, che è detta di Pascal di parametri n e p, cioè vale che, posto

$$\boxed{\Delta_1 = T_1, \cdots, \Delta_i = T_i - T_{i-1}, \quad i = 1, ..., n., \quad \text{sono indipendenti e tutte } Geom(p)}$$

$$\boxed{\mathbb{P}(\Delta_1 + \cdots + \Delta_n = k) = \mathbb{P}(T_n = k) = \binom{k-1}{n-1} (1-p)^{k-n} p^n, \quad k \geq n.}$$

Come conseguenza potremo calcolare facilmente valore atteso e varianza del tempo T_n di n-simo successo:

$$\boxed{\mathbb{E}(T_n) = \mathbb{E}(\Delta_1 + \cdots + \Delta_n) = \mathbb{E}(\Delta_1) + \cdots + \mathbb{E}(\Delta_n) = \frac{n}{p},}$$

$$\boxed{Var(T_n) = Var(\Delta_1 + \cdots + \Delta_n) = Var(\Delta_1) + \cdots + Var(\Delta_n) = n\frac{1-p}{p^2}.}$$

Tuttavia è possibile ottenere il risultato anche facendo i calcoli. Ma questo è lasciato per esercizio.
Esempio 15.3** in [SN] **(Somma di n variabili indipendenti, geometriche di parametro p e distribuzione di Pascal)
Iniziamo con il caso $n = 2$. Dimostreremo che se X_1 ed X_2 sono due variabili aleatorie Geom(p) (a partire da 1) indipendenti, ossia

$$\mathbb{P}(X_i = k) = p(1-p)^{k-1}, \text{ per } k = 1, 2, \ldots, \ i = 1, 2$$

e

$$\mathbb{P}(X_1 = k_1, X_2 = k_2) = p(1-p)^{k_1-1} p(1-p)^{k_2-1}, \text{ per } k_i = 1, 2, \ldots, \ i = 1, 2$$

allora $Z = X_1 + X_2$ è una v.a. con **distribuzione di Pascal** *di parametri 2 e p, ossia*

$$\mathbb{P}(Z = k) = \binom{k-1}{1} p^2 (1-p)^{k-2}, \quad k = 2, 3, \ldots.$$

Consideriamo uno schema di Bernoulli infinito (ossia una successione di eventi $\{E_n\}_{n \geq 1}$ globalmente indipendenti) e siano T_1 uguale al tempo di primo successo e T_2 il tempo di secondo successo. Poniamo

$$\Delta_1 = T_1 \quad e \quad \Delta_2 = T_2 - T_1.$$

Dimostreremo che la distribuzione congiunta di Δ_1 e Δ_2 è la stessa di X_1 e X_2. Quindi Δ_1 e Δ_2 hanno la stesse marginali di X_1 e X_2, e sono indipendenti. Di conseguenza Z ha la stessa distribuzione di $\Delta_1 + \Delta_2 = T_1 + (T_2 - T_1) = T_2$.
Infatti è facile convincersi che la distribuzione congiunta di Δ_1, Δ_2, è la stessa di X_1, X_2, cioè che

$$\mathbb{P}(\Delta_1 = k_1, \Delta_2 = k_2) = (1-p)^{k_1-1} p\,(1-p)^{k_2-1} p = \mathbb{P}(X_1 = k_1)\mathbb{P}(X_2 = k_2), \quad k_i \geq 1, e \ i = 1, 2,$$

in quanto

$$\{\Delta_1 = k_1, \Delta_2 = k_2\} = \overline{E}_1 \cap \cdots \overline{E}_{k_1-1} \cap E_{k_1} \cap \overline{E}_{k_1+1} \cap \cdots \overline{E}_{k_1+k_2-1} \cap E_{k_1+k_2},$$

e inoltre che

$$\mathbb{P}(T_2 = k) = \binom{k-1}{1} p^2 (1-p)^{k-2}, \quad k = 2,3,\ldots,$$

in quanto l'evento $\{T_2 = k\}$ *coincide con l'evento*

$$\{\textit{la k-sima prova è un successo, e tra le prime } \ k-1 \ \textit{prove c'è esattamente un successo}\} = E_k \cap \{S_{k-1} = 1\}$$

(qui $S_m = \sum_{k=1}^m \mathbf{1}_{E_k}$*) da cui*

$$\mathbb{P}(T_2 = k) = \mathbb{P}(E_k \cap \{S_{k-1} = 1\}) = \mathbb{P}(E_k)\mathbb{P}(S_{k-1} = 1) = p\binom{k-1}{1} p(1-p)^{(k-1)-1} = \binom{k-1}{1} p^2 (1-p)^{k-2}.$$

Consideriamo ora il caso n. Dimostreremo che se X_1, X_2, $\ldots$, X_n *sono variabili aleatorie* $Geom(p)$ *indipendenti, ossia*

$$\mathbb{P}(X_i = k) = p(1-p)^{k-1}, \quad \textit{per } k = 1,2,\ldots, \ i = 1,2,\ldots$$

e

$$\mathbb{P}(X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n) = p(1-p)^{k_1-1} p(1-p)^{k_2-1} \cdots p(1-p)^{k_n-1}, \quad \textit{per } k_i = 1,2,\ldots, \ i = 1,2,\ldots$$

allora, posto

$$Z_n = X_1 + X_2 + \cdots X_n$$
$$\mathbb{P}(Z_n = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}, \quad k = n, n+1, \ldots.$$

ossia Z_n *ha una* **distribuzione di Pascal** *di parametri n e p. Iniziamo considerando uno schema di Bernoulli infinito (ossia una successione di eventi globalmente indipendenti e tutti di probabilità p) e poniamo* T_1 *uguale al* **tempo di primo successo** *e* T_2 *il* **tempo di secondo successo**, $\ldots$, T_n *uguale al* **tempo di** *n*-**simo successo**. *Poniamo inoltre*

$$\Delta_1 = T_1, \quad \Delta_2 = T_2 - T_1, \ \ldots, \ \Delta_n = T_n - T_{n-1}.$$

Dimostreremo che la distribuzione congiunta di Δ_1, $\Delta_2, \ldots \Delta_n$ *è la stessa di* $X_1, X_2 \ldots, X_n$. *In altre parole* Δ_1, $\Delta_2, \ldots, \Delta_n$ *hanno la stesse marginali di* X_1, $X_2, \ldots, X_n$, *e sono indipendenti e perciò* Z_n *ha la stessa distribuzione di* $\Delta_1 + \Delta_2 + \cdots + \Delta_n = T_1 + (T_2 - T_1) + \cdots + (T_n - T_{n-1}) = T_n$.
Infatti è facile convincersi che

$$\mathbb{P}(\Delta_1 = k_1, \Delta_2 = k_2, \ldots, \Delta_n = k_n) = (1-p)^{k_1-1} p\,(1-p)^{k_2-1} p \cdots (1-p)^{k_n-1} p, \ k_i = 1,2,\ldots, i = 1,2,\ldots,n,$$

(cioè la distribuzione congiunta di Δ_1, $\Delta_2, \ldots \Delta_n$ *è la stessa di* $X_1, X_2 \ldots, X_n$*): infatti*

$$\begin{aligned}\{\Delta_1 = k_1, \Delta_2 = k_2, \ldots, \Delta_n = k_n\} = &\left(\overline{E}_1 \cap \overline{E}_2 \cap \cdots \overline{E}_{k_1-1} \cap E_{k_1}\right) \bigcap \left(\overline{E}_{k_1+1} \cap \overline{E}_{k_1+2} \cap \cdots \overline{E}_{k_1+k_2-1} \cap E_{k_1+k_2}\right) \\ &\bigcap \left(\overline{E}_{k_1+k_2+1} \cap \overline{E}_{k_1+k_2+2} \cap \cdots \overline{E}_{k_1+k_2+k_3-1} \cap E_{k_1+k_2+k_3}\right) \bigcap \cdots \\ &\cdots \bigcap \left(\overline{E}_{k_1+k_2+\ldots+k_{n-1}+1} \cap \overline{E}_{k_1+k_2+\ldots+k_{n-1}+2} \cap \cdots \overline{E}_{k_1+k_2+\ldots+k_{n-1}+k_n-1} \cap E_{k_1+k_2+\ldots+k_{n-1}+k_n}\right)\end{aligned}$$

Inoltre è facile convincersi che

$$\mathbb{P}(T_n = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}, \quad k = n, n+1, \ldots,$$

in quanto

$$\{T_n = k\} = \{\textit{la k-sima prova è un successo, e tra le prime } k-1 \textit{ prove ci sono esattamente } n-1 \textit{ successi}\}$$

ossia, essendo $S_m = \sum_{k=1}^m \mathbf{1}_{E_k}$,

$$\{T_n = k\} = E_k \cap \{S_{k-1} = n-1\}.$$

Essendo gli eventi E_k *e* $\{S_{k-1} = n-1\}$ *indipendenti, si ottiene quindi che*

$$\begin{aligned}\mathbb{P}(T_n = k) = \mathbb{P}(E_k \cap \{S_{k-1} = n-1\}) = \mathbb{P}(E_k)\mathbb{P}(S_{k-1} = n-1) &= p\binom{k-1}{n-1} p^{n-1} (1-p)^{(k-1)-(n-1)} \\ = \binom{k-1}{n-1} p^n (1-p)^{k-n}, &\quad k \geq n.\end{aligned}$$

5.4.1 NUMERO DI INSUCCESSI PRIMA DEL PRIMO SUCCESSO e NUMERO DI INSUCCESSI PRIMA DELL'n-simo SUCCESSO.

Se T_1 è il tempo di primo successo in prove indipendenti di probabilità p, e $\tilde{T}_1$ è il numero di insuccessi prima del primo successo, allora

$$\tilde{T}_1 := T_1 - 1, \quad \text{e quindi} \quad \mathbb{P}(\tilde{T}_1 = k) = \mathbb{P}(T_1 = k+1) = p(1-p)^{k+1-1} = p(1-p)^k, \quad \forall k \geq 0.$$

Una v.a. a valori in $\mathbb{Z}^+ = \{0\} \cup \mathbb{N}$ e con $\mathbb{P}(X = k) = p(1-p)^k$ viene detta *Geom*(p) *a partire da 0*. Quindi $\tilde{T}_1$ è una v.a. *Geom*(p) a partire da 0.

In modo simile se T_n è il tempo di n-simo successo in prove indipendenti di probabilità p, e $\tilde{T}_n$ è il numero di insuccessi prima dell'n-simo successo, allora

$$\tilde{T}_n := T_n - n, \quad \text{e quindi} \quad \mathbb{P}(\tilde{T}_n = k) = \mathbb{P}(T_n = k+n) = \binom{k+n-1}{n-1} p^n (1-p)^{n+k-n} = \binom{k+n-1}{n-1} p^n (1-p)^k, \quad \forall k \geq 0.$$

Se invece si considerano n v.a. X_1', X_2',..., X_n', indipendenti e *Geom*(p), ma a partire da 0, ossia se $\mathbb{P}(X_1' = k_1, X_2' = k_2, ..., X_n' = k_n) = p(1-p)^{k_1} p(1-p)^{k_2} \cdots p(1-p)^{k_n}$, per $k_i = 0, 1, 2, \ldots$, $i = 1, 2, ..., n$, allora $Z' = X_1' + X_2' + \cdots + X_n'$ è una v.a. con ***distribuzione binomiale negativa*** di parametri n e p, ossia

$$\mathbb{P}(Z' = k) = \binom{k+n-1}{n-1} p^n (1-p)^k, \quad k = 0, 1, 2, \ldots.$$

Per convincersene, basta pensare che le $X_i' = X_i - 1$, dove le v.a. X_i, per $i \geq 1$, sono come nell'**Esempio 15.3**, e che quindi

$$Z' = \sum_{i=1}^{n} X_i' = \sum_{i=1}^{n} (X_i - 1) = \sum_{i=1}^{n} X_i - n = Z - n,$$

da cui

$$\begin{aligned} \mathbb{P}(Z' = k) = \mathbb{P}(Z - n = k) = \mathbb{P}(Z = k+n) &= \binom{(k+n)-1}{n-1} p^n (1-p)^{(k+n)-n} \\ &= \binom{n+k-1}{k} p^n (1-p)^{(k+n)-n} \; k \geq 0. \end{aligned}$$

Come curiosità: il nome di binomiale negativa si spiega ricordando la definizione del coefficiente binomiale esteso ai numeri reali $\binom{\alpha}{k} = \dfrac{\alpha(\alpha-1)\cdots(\alpha-(k-1))}{k!}$ e osservando che, per $\alpha = -n$ si ha

$$\binom{-n}{k} = \frac{-n(-n-1)(-n-2)\cdots(-n-(k-1))}{k!} = (-1)^k \frac{n(n+1)\cdots(n+k-1)}{k!}$$

e quindi

$$\binom{n+k-1}{k} = \frac{(n+k-1)(n+k-2)\cdots(n+\cancel{k-1}-\cancel{(k-1)}}{k!} = \binom{-n}{k}(-1)^k$$

Altra curiosità: i coefficienti binomiali negativi entrano in gioco naturalmente quando si considera la serie di Mac Laurin di $(1+x)^\alpha$

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

Si osservi che quando $\alpha = n$ si ottiene che

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

in quanto per $k > n$ si ottiene che il prodotto $n(n-1)\cdots(n-(k-1)) = 0$, in quanto contiene un fattore nullo.

6 Variabili Continue come limiti di variabili discrete

In precedenza abbiamo visto come, in modo del tutto naturale, è sorta l'esigenza di passare da variabili aleatorie a valori in un insieme finito a variabili alaeatorie a valori in un insieme discreto numerabile: si pensi al caso delle variabili aleatorie con distribuzione binomiale $Bin(n,p)$ che sono approssimate con variabili aleatorie di Poisson (si veda il Teorema di Approssimazione di Poisson) o anche le variabili aleatorie geometriche come tempi di primo successo: ovviamente il numero delle prove che si possono effettuare nella realtà è finito, ma dal punto di vista matematico è più semplice pensare ad un infinito potenziale, cioè non mettere alcun limite al numero delle prove che si potranno effettuare.

In modo del tutto simile sorge l'esigenza di passare da variabili aleatorie discrete a variabili aleatorie più generali, che possono assumere valori in un intervallo reale.

In particolare nel seguente esempio, vedremo come sorge naturale considerare una variabile aleatorie a valori nell'intervallo $(0,1)$ della retta reale.

Esempio 6.1 (**variabili aleatorie uniformi come limite di variabili discrete uniformi**). *Supponiamo che n sia un numero intero "grande", e che X_n sia variabile aleatoria uniforme sull'insieme $\{x_i^{(n)} = \frac{i}{n}, \quad i = 1,2,\ldots,n\}$, ossia*

$$\mathbb{P}(X_n = \frac{i}{n}) = \frac{1}{n}, \qquad i \in \{1,2,\ldots,n\}$$

Equivalentemente, possiamo supporre che

$$X_n = \frac{U^{(n)}}{n} \quad \text{dove } U^{(n)} \text{ è una variabile aleatoria uniforme in } \{1,2,...,n\}.$$

e quindi, qualunque sia $n \geq 1$ si ha che l'insieme dei valori che può assumere X_n è contenuto in $(0,1]$, ovvero $X_n(\Omega)) \subset (0,1]$.
Mostreremo che, per una generica funzione g continua,

$$\lim_{n\to\infty} \mathbb{E}[g(X_n)] = \int_0^1 g(x)\,dx. \tag{1}$$

e inoltre, mostreremo che il valore della probabilità che X_n sia in un intervallo $(a,b] \subset (0,1]$ converge a $b-a$, ossia

$$\lim_{n\to\infty} \mathbb{P}(a < X_n \leq b) = b-a, \qquad \text{con } 0 < a < b < 1, \tag{2}$$

Il calcolo di $\mathbb{E}[g(X_n)]$ in (1) è relativamente semplice, infatti,

$$\mathbb{E}[g(X_n)] = \sum_{i=1}^n g(\tfrac{i}{n})\mathbb{P}(X_n = \tfrac{i}{n}) = \sum_{i=1}^n g(\tfrac{i}{n})\,\frac{1}{n}.$$

Per mostrare che esiste finito (e quanto vale) il limite per n che tende ad infinito di tale espressione cerchiamo di capire quanto vale questa l'espressione di tale valore per n "grande".

Non è difficile vedere che

$$\lim_{n\to\infty} \mathbb{E}[g(X_n)] = \lim_{n\to\infty} \sum_{i=1}^n g(\tfrac{i}{n})\,\frac{1}{n} = \int_0^1 g(x)\,dx,$$

infatti, posto $x_i^{(n)} := \frac{i}{n}$, per $i = 0,1,\ldots,n$, notiamo che $\frac{1}{n} = \frac{i}{n} - \frac{i-1}{n} = x_i^{(n)} - x_{i-1}^{(n)}$, e riscriviamo

$$\sum_{i=1}^n g(\tfrac{i}{n})\,\frac{1}{n} = \sum_{i=1}^n g(x_i^{(n)})\,\left(x_i^{(n)} - x_{i-1}^{(n)}\right) = \sum_{i=1}^n g(x_i^{(n)})\,\Delta x_i.$$

A questo punto, il lettore avrà riconosciuto che la precedente somma è una somma di Riemann, che serve per ottenere/definire il valore dell'integrale di $g(x)$ sull'intervallo $(0,1)$. Finiamo con l'osservare che il limite della precedente espressione esiste e vale appunto $\int_0^1 g(x)\,dx$, perché abbiamo assunto che g sia una funzione continua.

Per aiutare il lettore ricordiamo che per calcolare approssimativamente l'integrale definito di una funzione continua $f(x)$ in un intervallo $[a,b]$, basta suddividere l'intervallo $[a,b]$ in n intervalli $[x_{i-1}^{(n)}, x_i^{(n)}]$, $i = 1,2,\cdots,n$, di ampiezza $\Delta x_i := x_i^{(n)} - x_{i-1}^{(n)}$ "piccola".
In pratica basta assumente che, l'ampiezza non dipenda da i, ossia basta prendere $\Delta x_i = \Delta_x := \frac{b-a}{n}$, ovvero considerare

$$x_0^{(n)} := a,\ x_1^{(n)} := x_0^{(n)} + \Delta_x, \cdots x_i^{(n)} := x_{i-1}^{(n)} + \Delta_x = a + i\frac{b-a}{n}, \cdots x_n^{(n)} := x_{n-1}^{(n)} + \Delta_x = b$$

allora, comunque scelto $\xi_i^{(n)} \in [x_{i-1}^{(n)}, x_i^{(n)}]$

$$\int_a^b f(x)\,dx \approx \sum_{i=1}^{n} f(\xi_i^{(n)})\Delta x_i$$

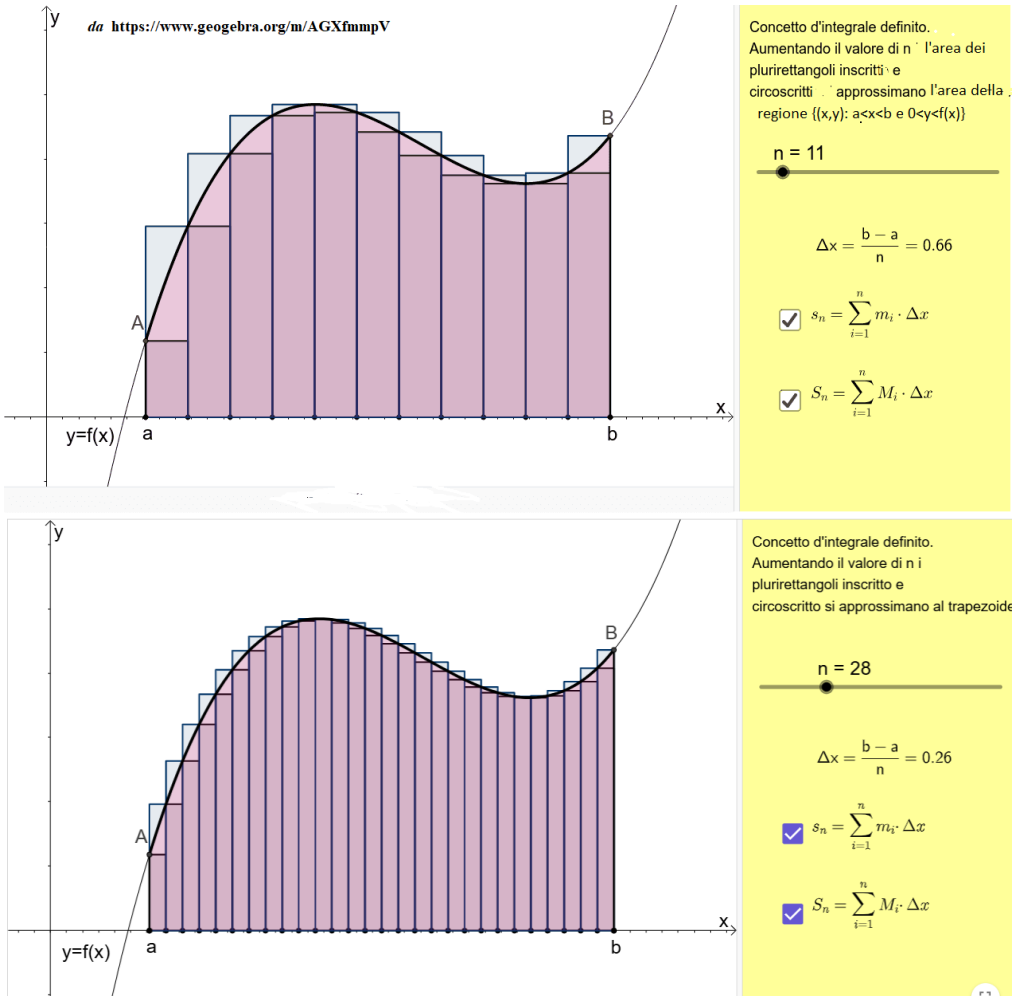


Figura 5: Le figure sono prese da https://www.geogebra.org/m/AGXfmmpV

Nelle figure precedenti la funzione per "trapezoide" si intende la figura compresa tra l'asse delle x, con $a \le x \le b$ e il grafico della funzione $f(x)$, ossia, essendo in questo caso la funzione f positiva, l'insieme

$$\{(x,y) :\ a \le x \le b,\ 0 \le y \le f(x)\}$$

si ha

$$m_i := \min_{x \in [x_{i-1}^{(n)}, x_i^{(n)}]} f(x) \quad e \quad M_i := \max_{x \in [x_{i-1}^{(n)}, x_i^{(n)}]} f(x).$$

Chiaramente qualunque sia $\xi_i^{(n)} \in [x_{i-1}^{(n)}, x_i^{(n)}]$ si ha

$$m_i \le f(\xi_i^{(n)}) \le M_i \quad \Leftrightarrow \quad \sum_{i=1}^{n} m_i \Delta x_i \le \sum_{i=1}^{n} f(\xi_i^{(n)})\Delta x_i \le \sum_{i=1}^{n} M_i \Delta x_i.$$

Vale la pena osservare che posto

$$f(x)=\begin{cases}0 & \text{per } x<0\\ 1 & \text{per } 0<x<1\\ 0 & \text{per } x>1\end{cases}$$

abbiamo ottenuto che

$$\lim_{n\to\infty}\mathbb{E}\big(g(X_n)\big)=\int_0^1 g(x)\,dx=\int_{-\infty}^{\infty} g(x)\,f(x)\,dx$$

Si noti l'analogia con la formula

$$\mathbb{E}\big(g(X_n)\big)=\sum_{x_k\in X(\Omega)} g(x_k)p_{X_n}(x_k)=\sum_x g(x)\,p_X(x)$$

Al posto di una somma c'è l'integrale, e al posto della densità discreta p_{X_n} c'è la funzione $f(x)$ che ha proprietà che

$$f(x)\geq 0,\quad \int_{-\infty}^{\infty} f(x)\,dx=1.$$

Anche il calcolo di $\mathbb{P}(a<X_n\leq b)$ in (2) è relativamente semplice, infatti, prima di tutto, per $a<b$,

$$\mathbb{P}(a<X_n\leq b)=\mathbb{P}(X_n\leq b)-\mathbb{P}(X_n\leq a).$$

(per la verifica della precedente uguaglianza, vedere la seguente nota)

Funzione di distribuzione di una v.a. X

Osserviamo che, per ogni variabile aleatoria X vale

$$\{X\leq b\}=\{a<X\leq b\}\cup\{X\leq a\},\quad \text{da cui}\quad \mathbb{P}(X\leq b)=\mathbb{P}(a<X\leq b)+\mathbb{P}(X\leq a).$$

Da cui ancora

$$\mathbb{P}(a<X\leq b)=\mathbb{P}(X\leq b)-\mathbb{P}(X\leq a).$$

Per il calcolo della probabilità che X sia nell'intervallo $(a,b]$, per ogni a, b, con $a<b$, basta quindi conoscere la funzione

$$x\mapsto F_X(x):=\mathbb{P}(X\leq x),$$

di modo che la precedente uguaglianza diviene

$$\mathbb{P}(a<X\leq b)=F_X(b)-F_X(a)$$

La funzione $F_X:\ \mathbb{R}\to[0,1]$, così definita ha un ruolo molto importante nella definizione delle variabili aleatorie continue (e non solo continue) viene detta **funzione di distribuzione** *(o* **di ripartizione***) della variabile aleatoria X.*

Occupiamoci quindi, al variare di x, della probabilità di

$$\mathbb{P}(X_n\leq x)=\frac{\#\{i:\ 0<i\leq n,\ \frac{i}{n}\leq x\}}{n}=\frac{\#\{i:\ 0<i\leq n,\quad i\leq xn\}}{n}\tag{3}$$

Tale probabilità è chiaramente nulla per $x\leq 0$ ed è chiaramente uguale ad 1, quando invece $x\geq 1$, in formule

$$\mathbb{P}(X_n\leq x)=0\qquad \text{per } x\leq 0,\qquad\qquad \mathbb{P}(X_n\leq x)=1\qquad \text{per } x\geq 1,$$

e quindi banalmente

$$\lim_{n\to\infty}\mathbb{P}(X_n\leq x)=0\qquad \text{per } x\leq 0,\qquad\qquad \lim_{n\to\infty}\mathbb{P}(X_n\leq x)=1\qquad \text{per } x\geq 1,$$

Il caso interessante è quindi il caso in cui $x \in (0,1)$. Per tali valori di x, dall'espressione (3) si vede subito[8] *che, denotando con $\lfloor \alpha \rfloor$ la parte intera inferiore di α,*

$$\mathbb{P}(X_n \le x) = \frac{\#\{i:\ 0 < i \le xn\}}{n} = \frac{\#\{i:\ 0 < i \le \lfloor xn \rfloor\}}{n} = \frac{\lfloor xn \rfloor}{n}, \qquad \text{per } x \in (0,1).$$

Abbiamo detto che ci interessa il calcolo per n "grande", e quindi vogliamo vedere se, per n che tende ad infinito, c'è un limite per tale probabilità. Si vede facilmente (vedere la nota successiva) che

$$\lim_{n\to\infty} \frac{\lfloor xn \rfloor}{n} = x, \quad \text{per ogni } x \in \mathbb{R}.$$

Di conseguenza

$$\lim_{n\to\infty} \mathbb{P}(X_n \le x) = \lim_{n\to\infty} \frac{\lfloor xn \rfloor}{n} = x, \qquad \text{per } x \in (0,1).$$

Verifica di: qualunque sia $x \in \mathbb{R}$ ***si ha*** $\lim_{n\to\infty} \frac{\lfloor xn \rfloor}{n} = x$.

Infatti, per definizione della parte intera inferiore di un numero reale y, si ha che $k = \lfloor y \rfloor$ se e solo se $k \le y < k+1$, ossia $\lfloor y \rfloor \le y < \lfloor y \rfloor + 1$. Quindi per $y = nx$ si ha

$$\lfloor nx \rfloor \le nx < \lfloor nx \rfloor + 1. \quad \Leftrightarrow \quad 0 = \lfloor nx \rfloor - \lfloor nx \rfloor \le nx - \lfloor nx \rfloor < \lfloor nx \rfloor + 1 - \lfloor nx \rfloor = 1$$

Quindi valgono le seguenti disuguaglianze

$$0 \le nx - \lfloor nx \rfloor \le 1 \quad \Leftrightarrow \quad 0 \le \frac{nx - \lfloor nx \rfloor}{n} = x - \frac{\lfloor nx \rfloor}{n} \le \frac{1}{n}.$$

, da cui immediatamente segue il limite

Riassumendo si ha

$$\lim_{n\to\infty} \mathbb{P}(X_n \le x) = \begin{cases} \lim_{n\to\infty} 0 = 0, & \text{per } x \le 0. \\ \lim_{n\to\infty} \frac{\lfloor nx \rfloor}{n} = x & \text{per } 0 < x < 1. \\ \lim_{n\to\infty} 1 = 1, & \text{per } x \ge 1. \end{cases}$$

A questo punto appare del tutto naturale dare la seguente definizione[9] di variabile aleatoria Uniforme nell'intervallo $(0,1)$.

Definizione 6.1. *Si dice che una variabile aleatoria continua U ha* **distribuzione uniforme** *in $(0,1)$ se la funzione $x \in \mathbb{R} \mapsto F_U(x) := \mathbb{P}(U \le x)$ è data da*

$$F_U(x) := \mathbb{P}(U \le x) = \begin{cases} 0 & \text{per } x \le 0 \\ x & \text{per } 0 < x < 1 \\ 1 & \text{per } x \ge 1 \end{cases}$$

e, per questa variabile aleatoria, qualunque sia la funzione g continua, definiremo

$$\mathbb{E}\left[g(U)\right] = \int_0^1 g(x)\,dx = \int_{-\infty}^{+\infty} g(x)\, f_U(x)\,dx.$$

dove

$$f_U(x) := \begin{cases} 0 & \text{per } x < 0 \\ 1 & \text{per } 0 < x < 1 \\ 0 & \text{per } x \ge 1 \end{cases}$$

[8] La prima uguaglianza è ovvia, in quanto $nx < n$, per $x \in (0,1)$. La terza uguaglianza è una banalità, mentre un pochino meno ovvia è la seconda uguaglianza. Per questo motivo invitiamo il lettore a considerare un esempio

$$\{i:\ 0 < i \le \sqrt{5}\} = \{i:\ 0 < i \le 2\} \qquad (\text{e } \lfloor \sqrt{5} \rfloor = 2)$$

e a ricordare che la definizione di parte intera inferiore di α è il massimo degli interi i tali che $i \le \alpha$.

[9] Vedere anche la successiva generalizzazione di variabile aleatoria Uniforma in un intervallo nella Sezione 9.

***Osservazione* 6.1.** Ovviamente una variabile aleatoria uniforme in $(0,1)$ può essere definita anche senza alcun riferimento all'approssimazione di variabili aleatorie X_n discrete e uniformi in $\{i/n,\, i=1,2,...,n\}$:
Si tratta della scelta "*a caso*" di un punto X nell'intervallo $(0,1)$, dove "*a caso*" in $(0,1)$ significa che la probabilità che si trovi nell'intervallo $(0,1)$ vale 1, e che si trovi in un intervallo I contenuto in $(0,1)$, è uguale all'ampiezza dell'intervallo stesso. Più in generale, preso un qualunque intervallo J la probabilità che si trovi in J è uguale all'ampiezza dell'intervallo $J\cap(0,1)$.

Notiamo inoltre che le anche il fatto che $\mathbb{E}[g(X_n)]$ converge a $\mathbb{E}[g(U)]$ per ogni funzione continua ci permette di ottenere che anche il valore atteso e la varianza di X_n convergono al valore atteso e la varianza di U, e infatti, ricordando che, se U_n ha distribuzione uniforme in $\{1,2,...,n\}$, allora

$$\mathbb{E}[U_n]=\frac{n+1}{2},\qquad Var(U_n)=\frac{n^2-1}{12}$$

e quindi, tenendo conto che $X_n=U_n/n$, otteniamo che

$$\mathbb{E}[X_n]=\mathbb{E}\left[\tfrac{U_n}{n}\right]=\frac{n+1}{2n}\underset{n\to\infty}{\longrightarrow}\frac{1}{2},\qquad Var(X_n)=Var\left(\tfrac{U_n}{n}\right)=\frac{n^2-1}{12n^2}\underset{n\to\infty}{\longrightarrow}\frac{1}{12},$$

inoltre, come è facile ottenere,

$$\mathbb{E}[U]=\frac{1}{2},\qquad Var(U)=\frac{1}{12}.$$

Consideriamo ora un altro esempio.

Esempio 6.2 (variabili aleatorie esponenziali come limite di variabili aleatorie geometriche). *Sia $T^{(n)}$ una variabile aleatoria Geometrica di parametro $p=p_n=\lambda/n$, cioè*

$$\mathbb{P}(T^{(n)}=k)=p(1-p)^{k-1}=\frac{\lambda}{n}\left(1-\frac{\lambda}{n}\right)^{k-1},\quad k=1,2,....$$

Come sappiamo, $T^{(n)}$ si ottiene come tempo (calcolato in numero di prove) di primo successo in una successione di prove ripetute (ossia uno schema di Bernoulli infinito) e quindi

$$\mathbb{P}(T^{(n)}>k)=(1-p)^k=\left(1-\frac{\lambda}{n}\right)^k,\qquad k=0,1,2,....$$

Consideriamo ora la variabile aleatoria

$$X_n=\frac{T^{(n)}}{n}.$$

Possiamo pensare sempre X_n come tempo di primo successo, nella situazione in cui invece di fare una sola prova in ciascuna unità di misura, in ogni unità di misura si effettuano n prove, e il tempo è un tempo fisico e NON il numero di prove effettuate. Ad esempio, se misuriamo il tempo in ore, e se si effettuano le prove a distanza di un minuto una dall'altra, allora in ogni ora si effettuano 60 prove. Assumiamo che la probabilità di successo sia uguale a $5/60$ in modo che il valore atteso del numero dei successi in un'ora sia $5=60\cdot(5/60)$. Mentre $T^{(60)}$ indica il numero di prove effettuate fino al primo successo (o equivalentemente il tempo di primo successo, quando però il tempo è misurato in minuti), invece $X^{(60)}=T^{(60)}/60$ calcola il tempo, misurato in ore, che si deve aspettare fino al primo successo incluso. Analogamente, se invece le prove vengono effettuate a distanza di un secondo l'una dall'altra allora in un'ora si effettuano 3600 prove e $X^{(3600)}=T^{(3600)}/3600$ rappresenta il tempo, misurato in ore, che si deve aspettare per ottenere il primo successo, invece $T^{(3600)}$ rappresenta il numero di prove effettuate fino al primo successo, (o equivalentemente il tempo di primo successo, ma misurato in secondi invece che in ore). Il riscalamento prevede che la probabilità di successo però sia stavolta più piccola e valga $5/3600$, in modo che il valore atteso del numero dei successi in un'ora sia sempre lo stesso, cioè valga $5=3600\cdot(5/3600)$.

Vogliamo calcolare analogamente al caso precedente la sua funzione di distribuzione e vedere se ammette limite per n che tende ad infinito.

$$\mathbb{P}(X_n\le x)=0\qquad\qquad per\ x\le 0$$

mentre

$$\mathbb{P}(X_n\le x)=1-\mathbb{P}(X_n>x)\qquad per\ x>0.$$

Inoltre, sempre per $x > 0$,

$$\mathbb{P}(X_n > x) = \mathbb{P}(T^{(n)} > nx) = \mathbb{P}(T^{(n)} > \lfloor nx \rfloor)$$

in quanto $T^{(n)}$ è a valori interi[10]*; di conseguenza, essendo $T^{(n)}$ una variabile aleatoria Geom$(\frac{\lambda}{n})$, si ha*

$$\mathbb{P}(X_n > x) = \mathbb{P}(T^{(n)} > \lfloor nx \rfloor) = \left(1 - \frac{\lambda}{n}\right)^{\lfloor nx \rfloor} = \left(1 - \frac{\lambda}{n}\right)^{n \frac{\lfloor nx \rfloor}{n}} = \left[\left(1 - \frac{\lambda}{n}\right)^{n}\right]^{\frac{\lfloor nx \rfloor}{n}} \underset{n\to\infty}{\longrightarrow} (e^{-\lambda})^{x} = e^{-\lambda x},$$

in quanto

$$e^{-\lambda} = \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n}, \quad e \quad x = \lim_{n\to\infty} \frac{\lfloor nx \rfloor}{n}.$$

Riassumendo, abbiamo dimostrato che

$$\lim_{n\to\infty} \mathbb{P}(X_n \le x) = \begin{cases} 0 & per\ x \le 0 \\ 1 - e^{-\lambda x} & per\ x > 0. \end{cases} \tag{4}$$

Definizione 6.2. *Si dice che una variabile aleatoria X ha **distribuzione esponenziale di parametro** λ se la sua funzione di dsitribuzione $x \in \mathbb{R} \mapsto F_X(x) := \mathbb{P}(X \le x)$ è data da*

$$F_X(x) := \mathbb{P}(X \le x) = \begin{cases} 0 & per\ x \le 0, \\ 1 - e^{-\lambda x} & per\ x > 0. \end{cases}$$

Possiamo quindi riassumere quanto visto nel procedente Esempio 6.2 dicendo che la variabile aleatoria X_n, che è una Geometrica di parametro λ/n riscalata **converge "in distribuzione"** ad una variabile esponenziale X di parametro λ.

***Osservazione* 6.2.** Anche in questo caso, nelle stesse condizioni e con le stesse notazioni del precedente Esempio 6.2, è possibile verificare che valore atteso e varianza di X_n, convergono al valore atteso e alla varianza di X: infatti

$$\mathbb{E}\left(T^{(n)}\right) = \frac{1}{p}, \qquad Var\left(T^{(n)}\right) = \frac{1-p}{p^2},$$

da cui, tenendo conto che $X_n = T^{(n)}/n$ e $p = \lambda/n$, otteniamo

$$\mathbb{E}(X_n) = \frac{1}{n}\frac{1}{\frac{\lambda}{n}} = \frac{1}{\lambda}, \qquad Var(X_n) = \frac{1}{n^2}\frac{1-\frac{\lambda}{n}}{\left(\frac{\lambda}{n}\right)^2} = \frac{1-\frac{\lambda}{n}}{\lambda^2} \underset{n\to\infty}{\longrightarrow} \frac{1}{\lambda^2},$$

inoltre è facile verificare che

$$\mathbb{E}(X) = \frac{1}{\lambda}, \qquad Var(X) = \frac{1}{\lambda^2}.$$

Le analogie tra variabili aleatorie geometriche ed esponenziali non finiscono qui: un'altra importante analogia che accomuna queste due famiglie di variabili aleatorie è la cosìdetta proprietà di mancanza di memoria.

***Osservazione* 6.3 (proprietà di mancanza di memoria per variabili aleatorie $Geom(p)$ ed $EXP(\lambda)$).** Si osservi che sia le variabili aleatorie Geometriche, che le variabili aleatorie Esponenziali hanno la proprietà della MANCANZA DI MEMORIA, ossia godono della seguente proprietà: Sia T una variabile aleatoria Geometrica di parametro $p \in (0,1)$. Si ha

$$\mathbb{P}(T > h+k) = \mathbb{P}(T > h)\mathbb{P}(T > k), \quad \forall h,k \in \{0,1,2,\cdots\},$$

infatti

$$\mathbb{P}(T > h+k) = (1-p)^{h+k} = (1-p)^h(1-p)^k = \mathbb{P}(T > h)\mathbb{P}(T > k), \quad \forall h;k \in \{0,1,2,\cdots\}$$

ossia, $\forall h,k \in \{0,1,2,\cdots\}$,

$$\mathbb{P}(T > h+k | T > k) = \frac{\mathbb{P}(T > h+k, T > k)}{\mathbb{P}(T > k)} = \frac{\mathbb{P}(T > h+k)}{\mathbb{P}(T > k)}$$

[10] Per capire questo punto, basta pensare, ad esempio, al fatto che $\{T^{(n)} > \pi\} = \{T^{(n)} > 3\}$

(si noti che $\{T > h+k, T > k\} = \{T > h+k\}$, in quanto $h \geq 0$)

$$= \frac{(1-p)^{h+k}}{(1-p)^k} = (1-p)^h = \mathbb{P}(T > h)$$

cioè

$$\mathbb{P}(T-k > h|T > k) = \mathbb{P}(T > h), \forall h,k \in \{0,1,2,\cdots\}.$$

Quest'ultima formulazione mette in evidenza perché tale proprietà è detta di mancanza di memoria: infatti, se sappiamo che $T > k$ (cioè che non ho avuto successi fino alla k-sima prova inclusa) allora sappiamo che $T-k$ rappresenta il tempo che devo ancora aspettare per ottenere il primo successo, e $\mathbb{P}(T-k > h|T > k)$ rappresenta la probabilità di dover aspettare più di altre h prove prima del primo successo, sapendo che ho già effettuato k prove e che sono tutte fallite.

Il fatto che $\mathbb{P}(T-k > h|T > k) = \mathbb{P}(T > h)$ ci dice che la funzione distribuzione, condizionata a $T > k$, del numero di prove che devo ancora aspettare è la stessa funzione di distribuzione della variabile aleatoria tempo di primo successo (cioè è come se iniziassi a contare da capo il numero di prove, senza avere alcuna informazione su quanto è accaduto nelle prove precedenti).

Analogamente se X ha distribuzione esponenziale di parametro $\lambda > 0$ si ha

$$\mathbb{P}(X > s+t) = \mathbb{P}(X > s)\mathbb{P}(X > t), \quad \forall s,t > 0$$

infatti

$$\mathbb{P}(X > s+t) = e^{-\lambda(s+t)} = e^{-\lambda s}e^{-\lambda t} = \mathbb{P}(X > s)\mathbb{P}(X > t), \quad \forall s;t > 0;$$

ossia, $\forall s,t > 0$

$$\mathbb{P}(X > s+t|X > t) = \frac{\mathbb{P}(X > s+t, X > t)}{\mathbb{P}(X > t)} = \frac{\mathbb{P}(X > s+t)}{\mathbb{P}(X > t)}$$

in quanto, essendo $s > 0$, si ha $\{X > s+t\} \supseteq \{X > t\}$, si ha $\{X > s+t, X > t\} = \{X > s+t\}$, e quindi, in conclusione

$$\mathbb{P}(X > s+t|X > t) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = \mathbb{P}(X > s)$$

cioè

$$\mathbb{P}(X-t > s|X > t) = \mathbb{P}(X > s), \quad \forall s,t > 0$$

L'interpretazione è analoga al caso di una variabile aleatoria Geometrica, se si pensa ad X come il tempo di guasto di un'apparecchiatura (o ad un tempo di attesa, ad esempio, di una prima telefonata, etc.).

7 Spazi di probabilità generali

Richiamiamo qui la definizione di spazio di probabilità generale. Ossia qui non supponiamo che Ω sia numerabile. Prima di tutto va richiamato il concetto di σ-algebra:

Definizione 7.1. *Una famiglia di insiemi di Ω, ossia $\mathcal{F} \subset \mathcal{P}(\Omega)$ (dove, come al solito, $\mathcal{P}(\Omega)$ indica l'insieme delle parti di Ω) è una* ***σ-algebra****, se e solo se è tale che*

i) $\Omega \in \mathcal{F}$

ii) *se $A \in \mathcal{F}$, allora $\overline{A} \in \mathcal{F}$*

iii) *se $A_k \in \mathcal{F}$, per $k \geq 1$, allora $\cup_{k=1}^{\infty} A_k \in \mathcal{F}$*

Osservazione: se $E_1, E_2, \ldots \in \mathcal{F}$ è una successione di eventi allora $\bigcap_{j=1}^{\infty} E_j \in \mathcal{F}$, come si vede subito applicando la formula di de Morgan e le proprietà ***ii)*** e ***iii)***:

$$\bigcap_{j=1}^{\infty} E_j = \overline{\bigcup_{j=1}^{\infty} \overline{E}_j} \in \mathcal{F} \tag{5}$$

in quanto se $E_j \in \mathcal{F}$ allora $\overline{E}_j \in \mathcal{F}$, e quindi anche $\bigcup_{j=1}^{\infty} \overline{E}_j \in \mathcal{F}$ e lo stesso vale per il suo complementare.

Spazio di probabilità

Uno spazio di probabilità, con la probabilità numerabilmente additiva (anche detta σ-additiva) è una terna $(\Omega, \mathcal{F}, \mathbb{P})$ con $\mathcal{F}$ una σ-algebra, e $\mathbb{P} : \mathcal{F} \to [0,1]$ una funzione tale che

i) per ogni $A \in \mathcal{F}$, $\mathbb{P}(A) \geq 0$

ii) $\mathbb{P}(\Omega) = 1$

iii) se $A_k \in \mathcal{F}$, per $k \geq 1$, e $A_h \cap A_k = \emptyset$, allora

$$\mathbb{P}\big(\cup_{k=1}^{\infty} A_k\big) = \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$

Ricordiamo che **solo** gli insiemi che appartengono alla σ-algebra $\mathcal{F}$ vengono detti **eventi**. Tali eventi rappresentano gli insiemi per i quali è possibile controllare se si sono verificati o no, e inoltre solo per gli insiemi/eventi di $\mathcal{F}$ è possibile calcolare la probabilità.

INTERPRETAZIONE della sigma-algebra $\mathcal{F}$

Una possibile interpretazione della sigma-algebra degli eventi $\mathcal{F}$ è considerare i suoi sottinsieme comegli insie per i quali si può affermare che si sono verificati oppure no, e che ciò dipende dal nostro stato di informazione

Esempio del lancio di un dado in cui sono stati coperti di rosso i numeri pari e di blu i numeri dispari: in tale caso è impossibile verificare se è uscito il numero 2, ma solo se è uscito un numero pari o no (ossia se è uscito un numero dispari)

Come nel caso di Ω numerabile di dimostra che vale anche la proprietà di **additività finita** e quindi **tutte le proprietà fondamentali delle probabilità**, come ad esempio la **monotonia**, o **la formula delle probabilità totali**.

DIFFERENZA IMPORTANTE

L'unica differenza fondamentale è che bisogna sempre aggiungere che si tratta di eventi, ossia che i sottoinsiemi di cui vogliamo valutare la probabilità devono appartenere a $\mathcal{F}$, ad esempio

Proprietà di monotonia:

In uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ *si ha che*

Se A *e* B *sono eventi, cioè se* $A, B \in \mathcal{F}$*, allora* $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Chiaramente la condizione che $A, B \in \mathcal{F}$ è necessaria perché altrimenti non sarebbe possibile calcolare $\mathbb{P}(A)$ e $\mathbb{P}(B)$, in quanto la probabilità è definita solo per i sottoinsiemi di Ω che appartengono a $\mathcal{F}$.

Analogamente per la **formula delle Probabilità Totali** bisogna richiedere che $E \in \mathcal{F}$ e che $H_i \in \mathcal{F}$, per ogni elemento della partizione.

Negli spazi di probabilità generali (e comunque quando vale l'additività numerabile) vale anche la proprietà di continuità della probabilità: ossia

Proposizione **(Proprietà di continuità delle probabilità)** In uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ siano dati, per $n = 1, 2, ..., A_n \in \mathcal{F}, B_n \in \mathcal{F}$ tali che

$$A_n \subseteq A_{n+1}; \qquad B_n \supseteq B_{n+1}$$

e poniamo

$$A := \bigcup_{n=1}^{\infty} A_n; \qquad B := \bigcap_{n=1}^{\infty} B_n.$$

Se $\mathbb{P}$ è σ-additiva allora risulta

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n); \qquad \mathbb{P}(B) = \lim_{n\to\infty} \mathbb{P}(B_n)$$

Si tratta della Proposizione 15.1 (Proprietà di continuità delle probabilità) in [SN] oppure la Proposizione 6.1 nella Sezione 2.6 del ROSS, il cui enunciato è riportato nel file "dal libro di Ross.pdf" che contiene esempi ed esercizi presi dal Ross e suggeriti dalla prof.ssa Faggionato.

Il motivo per cui tale proprietà si chiama continuità delle probabilità risiende nel fatto che, se $A_n \subseteq A_{n+1}$ e $A := \bigcup_{n=1}^{\infty} A_n$ allora si scrive $A = \lim_{n\to\infty} A_n$.

Analogamente se $B_n \supseteq B_{n+1}$ e $B := \bigcap_{n=1}^{\infty} B_n$ allora si scrive $B = \lim_{n\to\infty} B_n$.

La dimostrazione non è richiesta, tuttavia va detto che basta dimostrare il caso di A_n crescente (poi basta usare la formula di Morgan per ottenere il caso di B_n decrescente, ponendo $A_n := \overline{B}_n$.
Inoltre, posto

$$C_1 = A_1, \qquad C_2 = A_2 \setminus A_1 = A_2 \cap \overline{A}_1, \qquad C_3 = A_3 \setminus A_2 = A_3 \cap \overline{A}_2, ...$$

È facile verificare che $\forall n \in \mathbb{N}$, $C_1, ..., C_n$ sono a due a due disgiunti e che $A_n = \bigcup_{i=1}^{n} C_i$ da cui $\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(C_i)$ e $A = \bigcup_{i=1}^{\infty} C_i$, da cui $\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(C_i)$, e quindi, per definizione di somma di una serie,

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(C_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(C_i) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

8 Funzione di distribuzione per variabili aleatorie in spazi generali

Sia X una variabile aleatoria, in uno spazio generale $(\Omega, \mathcal{F})$, ossia

$$X:\ \Omega \to \mathbb{R}, \quad \omega \mapsto X(\omega)$$

con la proprietà che, per ogni intervallo I, ossia $I = (\alpha, \beta]$, oppure $I = (\alpha, \beta)$, o $I = [\alpha, \beta)$, o $I = [\alpha, \beta]$, con $-\infty \le \alpha \le \beta \le \infty$, vale che

$$\{X \in I\} = \{\omega \in \Omega : X(\omega) \in I\} \in \mathcal{F}.$$

In realtà basta che si abbia

$$\text{per ogni } x \in \mathbb{R} \quad \{X \le x\} \in \mathcal{F},$$

infatti, ad esempio, per ogni $-\infty \le \alpha \le \beta \le \infty$,

$$\{X \le \beta\} = \{X \le \alpha\} \cup \{\alpha < X \le \beta\} \quad \Leftrightarrow \quad \{\alpha < X \le \beta\} = \{X \le \beta\} \setminus \{X \le \alpha\} = \{X \le \beta\} \cap \overline{\{X \le \alpha\}}$$

Sia $\mathbb{P}$ una probabilità numerabilmente additiva, su $(\Omega, \mathcal{F})$, allora per calcolare la distribuzione di X, ossia per calcolare per ogni I intervallo finito o infinito di $\mathbb{R}$, basta conoscere la funzione di distribuzione di X (anche detta funzione di ripartizione di X), ossia

$$F:\ \mathbb{R} \to [0,1], \quad x \mapsto F_X(x) := \mathbb{P}(X \le x),$$

come vedremo a breve.
Prima di vedere le proprietà delle funzioni di distribuzione, vediamo come sono fatte per alcune classi di v.a. In questa sezione vedremo solo v.a. discrete e v.a. assolutamente continue.

Si dice che X è discreta se $X(\Omega)$ ha cardinalità finita o numerabile, ovvero

$$X(\Omega) = \{x_1, ..., x_m\} \quad \text{oppure} \quad X(\Omega) = \{x_k,\ k \in \mathbb{N}\}.$$

La densità discreta di X è definita come

$$p_X(x_k) = \mathbb{P}(X = x_k), \quad x_k \in X(\Omega),$$

e gode delle proprietà

$$p_X(x_k) \ge 0, \quad \forall x_k \in X(\Omega), \qquad \sum_k p_X(x_k) = 1$$

La funzione di distribuzione di una v.a. discreta si calcola come

$$\boxed{F_X(x) = \sum_{k: x_k \le x} p_X(x_k)} \quad \text{infatti} \quad \boxed{\mathbb{P}(X \le x) = \mathbb{P}\big(\cup_{k: x_k \le x} \{X = x_k\}\big) = \sum_{k: x_k \le x} \mathbb{P}(X = x_k)}$$

Nel caso in cui X è discreta finita la $F_X(x)$ è una funzione costante a tratti, i punti di salto coincidono con i valori che può assumere la v.a. X, e inoltre

$$F_X(x) - F_X(x^-) = \sum_{k: x_k \le x} p_X(x_k) - \sum_{k: x_k < x} p_X(x_k)$$

e quindi

$$\mathbb{P}(X = x) = F_X(x) - F_X(x^-) > 0, \quad \text{se e solo se } x = x_h, \quad \text{ossia } x \in X(\Omega).$$

Ad esempio per $X(\omega) \equiv 0$ per ogni ω, allora $\quad F_X(x) = \begin{cases} 0 & \text{se } x < 0, \\ 1 & \text{se } x \ge 0. \end{cases}$

Ad esempio, $X = \mathbf{1}_A$ con A un evento, ossia con $A \in \mathcal{F}$, e $p := \mathbb{P}(A)$ allora

$$F_X(x) = \begin{cases} 0 & \text{per } x < 0, \\ \mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A) = 1 - p & \text{per } 0 \le x < 1, \\ 1 & \text{per } x \ge 1. \end{cases}$$

Si noti che i punti di discontinuità di F_X sono 0 e 1 e $\mathbb{P}(X=0)=\mathbb{P}(\overline{A})=F_X(0)-F_X(0^-)=(1-\mathbb{P}(A))-0$, e $\mathbb{P}(X=1)=\mathbb{P}(A)=F_X(1)-F_X(1^-)=1-(1-\mathbb{P}(A))$.

Più in generale, come già osservato, gli unici punti di discontinuità della funzione di distribuzione di una variabile aleatoria X discreta con $X(\Omega)$ finito, sono i valori x_k, e si ha $F_X(x_k)-F_X(x_k^-)=\mathbb{P}(X=x_k)$.

Si dice invece che X è **assolutamente continua** se esiste una funzione $f_X(x)$, tale che

$$f_X(x)\geq 0 \qquad \int_{-\infty}^{\infty} f_X(x)\,dx=1$$

e per la quale, qualunque sia $I=I(\alpha,\beta)$ un intervallo di estremi α e β, con $-\infty\leq\alpha\leq\beta\leq\infty$, si ha

$$\mathbb{P}(X\in I(\alpha,\beta))=\int_{\alpha}^{\beta} f_X(x)\,dx=F_X(\beta)-F_X(\alpha)$$

In tale caso la funzione di distribuzione è continua (e quindi $\mathbb{P}(X=x)=0$ per ogni $x\in\mathbb{R}$) e vale

$$\boxed{F_X(x)=\int_{-\infty}^{x} f_X(t)\,dt,} \qquad \text{per ogni } x\in\mathbb{R}.$$

Inoltre se $F_X(x)$ è continua e ammette derivata continua, tranne eventualmente in un numero finito di punti, allora vale la relazione

$$f_X(x)=F_X'(x), \quad \text{esclusi al massimo un numero finito di punti.}$$

Ricordiamo, infine, che invece di dire che X è assolutamente continua con densità $f_X(x)$, si dice anche che X ammette densità $f_X(x)$.

LE PROPRIETÀ DELLA FUNZIONE DI DISTRIBUZIONE
P0. $F_X(x)\in[0,1]$, per ogni $x\in\mathbb{R}$
P1. $F_X(x)$ è **monotona non decrescente** (si dice anche **crescente in senso lato**), ossia

$$\text{se } x'\leq x'', \quad \text{allora } \ F_X(x')\leq F_X(x'')$$

P2. $F_X(x)$ è **continua a destra e ammette limite a sinistra** (ossia $F_X(x^+)=F_X(x)$ ed esiste $F_X(x^-)$)
P3. $F_X(x)$ è **normalizzata**, cioè $\lim_{x\to-\infty}F_X(x)=0$ e $\lim_{x\to+\infty}F_X(x)=1$.

Dimostriamo solo
la proprietà **P1**: se $x'\leq x''$ allora $\{X\leq x'\}\subseteq\{X\leq x''\}$ e quindi $F_X(x')=\mathbb{P}(X\leq x')\leq\mathbb{P}(X\leq x'')=F_X(x'')$
e la seconda parte della proprietà **P3**, ossia $\lim_{x\to+\infty}F_X(x)=1$:

$$\lim_{x\to+\infty}F_X(x)=\lim_{n\to+\infty}F_X(n)=\lim_{n\to+\infty}\mathbb{P}(X\leq n)$$

E inoltre **(qui senza dimostrazione)**, in analogia con il caso discreto vale

$$F_X(x)-F_X(x^-)=\mathbb{P}(X=x), \qquad \text{per ogni } x\in\mathbb{R}.$$

INFINE (qui senza dimostrazione) se $F(x)$ ha le precedenti proprietà, allora è la funzione di distribuzione di una variabile aleatoria.

Per verificare che la funzione di ripartizione, $F_X(x)$ effettivamente permette di ottenere le probabilità $\mathbb{P}(X \in I)$, dove I è un intervallo di estremi α e β, osserviamo che, essendo, per ogni $-\infty \leq \alpha \leq \beta \leq \infty$,

$$\{X \leq \beta\} = \{X \leq \alpha\} \cup \{\alpha < X \leq \beta\}$$

e quindi

$$\mathbb{P}(X \leq \beta) = \mathbb{P}(X \leq \alpha) + \mathbb{P}(\alpha < X \leq \beta) \quad \Leftrightarrow \quad F_X(\beta) = F_X(\alpha) + \mathbb{P}(\alpha < X \leq \beta)$$

da cui

$$\mathbb{P}(\alpha < X \leq \beta) = F_X(\beta) - F_X(\alpha)$$

Di conseguenza

$$\mathbb{P}(\alpha \leq X \leq \beta) = \mathbb{P}(\alpha < X \leq \beta) + \mathbb{P}(X = \alpha) = F_X(\beta) - F_X(\alpha) + \mathbb{P}(X = \alpha)$$
$$\mathbb{P}(\alpha < X < \beta) = \mathbb{P}(\alpha < X \leq \beta) - \mathbb{P}(X = \beta) = F_X(\beta) - F_X(\alpha) - \mathbb{P}(X = \beta)$$
$$\mathbb{P}(\alpha \leq X < \beta) = \mathbb{P}(\alpha < X \leq \beta) + \mathbb{P}(X = \alpha) - \mathbb{P}(X = \beta) = F_X(\beta) - F_X(\alpha) + \mathbb{P}(X = \alpha) - \mathbb{P}(X = \beta)$$

Basta inoltre ricordare che $\mathbb{P}(X = x) = F_X(x) - F_X(x^-)$ per calcolare $\mathbb{P}(X = \alpha)$ e $\mathbb{P}(X = \beta)$.

Infine, sempre in analogia con il caso discreto:

$$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x)\, f_X(x)\, dx, \quad \text{purché} \quad \int_{-\infty}^{+\infty} |g(x)|\, f_X(x)\, dx < \infty.$$

Analogie tra caso discreto e caso continuo

caso discreto $p_X(x_k) \geq 0 \qquad \sum_{x_k} p_X(x_k) = 1$

caso assol. continuo $f_X(x) \geq 0 \qquad \int_{-\infty}^{\infty} f_X(x)\, dx = 1$

caso discreto $F(x) := \mathbb{P}(X \leq x) = \sum_{x_k 1leqx} p_X(x_k)$

caso assol. continuo $F_X(x) := \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt = -F_X(\alpha)$

caso discreto $\mathbb{P}(X \in (\alpha, \beta]) = \sum_{x_k \in (\alpha,\beta]} p_X(x_k) = F_X(\beta) - F_X(\alpha)$ per ogni $-\infty \leq \alpha \leq \beta \leq \infty$,

caso assol. continuo $\mathbb{P}(X \in (\alpha, \beta]) = \int_{\alpha}^{\beta} f_X(x)\, dx = F_X(\beta) - F_X(\alpha)$ per ogni $-\infty \leq \alpha \leq \beta \leq \infty$,

caso discreto $\mathbb{E}(g(X)) = \sum_{-\infty < x_k < x_k} g(x_k)\, p_X(x_k)$ purché $\sum_{-\infty < x_k < x_k} |g(x_k)|\, p_X(x_k) < \infty.$

caso assol. continuo $\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x)\, f_X(x)\, dx,$ purché $\int_{-\infty}^{+\infty} |g(x)|\, f_X(x)\, dx < \infty.$

9 Variabili aleatorie con densità di probabilità: casi notevoli

9.1 Variabili aleatorie uniformi in (a,b)

Una variabile aleatoria X è detta **UNIFORME in** (a,b) e si scrive $X \sim Unif(a,b)$, o anche $X \sim R(a,b)$ (la R sta per *Random*) se e e solo se ammette densità $f_X(x)$ e vale

$$f_X(x) = \begin{cases} 0 & \text{per } x < a \\ c & \text{per } a < x < b \\ 0 & \text{per } x > b \end{cases}$$

e quindi, poiché $1 = \int_{-\infty}^{+\infty} f_X(x)\,dx = \int_a^b c\,dx = c(b-a)$, necessariamente $c = \frac{1}{b-a}$. Inoltre

$$F_X(x) = \begin{cases} 0 & \text{per } x < a \\ \frac{x-a}{b-a} & \text{per } a \le x < b \\ 1 & \text{per } x \ge b \end{cases}$$

(Attenzione: in Esempio 14.10 ed Esempio 14.11 in [SN] viene data prima la funzione di distribuzione e poi la densità) Il valore atteso e la varianza valgono:

$$\mathbb{E}(X) = \frac{a+b}{2} \qquad Var(X) = \frac{(b-a)^2}{12}.$$

Infatti

$$\mathbb{E}(X) = \int_a^b x \frac{1}{b-a}\,dx = \frac{1}{b-a}\,\frac{x^2}{2}\Big|_a^b = \frac{1}{b-a}\,\frac{b^2-a^2}{2} = \frac{1}{b-a}\,\frac{(b-a)(b+a)}{2} = \frac{a+b}{2}$$

$$\begin{aligned} Var(X) &= \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \int_a^b x^2 \frac{1}{b-a}\,dx - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{4a^2+4ab+4b^2-3a^2-6ab-3b^2}{12} = \frac{(b-a)^2}{12}. \end{aligned}$$

in quanto

$$\mathbb{E}(X^2) = \frac{1}{b-a}\,\frac{x^3}{3}\Big|_a^b = \frac{1}{b-a}\,\frac{b^3-a^3}{3} = \frac{1}{b-a}\,\frac{(b-a)(a^2+ab+b^2)}{3} = \frac{a^2+ab+b^2}{3}$$

Data una v.a. $U \sim Unif(0,1)$, si può ottenere una variabile aleatoria $X \sim Unif(a,b)$ definendo

$$X(\omega) = a + (b-a)U(\omega).$$

È immediato che $X = a+(b-a)U$ assume solo valori tra a e b, ma per la verifica che è una v.a. uniforme vedere[11] la Sezione 14.6 Trasformazioni di variabili aleatorie e il caso delle trasformazioni affini in [SN] (a pagina 220).

Tutto è basato sul fatto che $\{X \le x\} = \{a+(b-a)U \le x\} = \{(b-a)U \le x-a\} = \{U \le \frac{x-a}{b-a}\}$ per cui

$$F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}\left(U \le \frac{x-a}{b-a}\right) = F_U\left(\frac{x-a}{b-a}\right) = \begin{cases} 0 & \text{per } \frac{x-a}{b-a} < 0 \\ \frac{x-a}{b-a} & \text{per } 0 \le \frac{x-a}{b-a} < 1 \\ 1 & \text{per } \frac{x-a}{b-a} \ge 1 \end{cases} = \begin{cases} 0 & \text{per } x < a \\ \frac{x-a}{b-a} & \text{per } a \le x < b \\ 1 & \text{per } x \ge b \end{cases}$$

[11] Attenzione nella versione di settembre 2024 di [SN] c'è stato un problema per cui la Lezione 11 è stata divisa erroneamente in due parti che sono diventate Lezione 11 Campionamento da popolazioni etc.. e 12 Descrizione formale del modello. Come conseguenza la numerazione delle Lezioni successive è cambiata: ad esempio la Lezione 14 ora è la Lezione 15 e quindi ora la Sezione 14.6 Trasformazioni di variabili aleatorie e il caso delle trasformazioni affini appare come la sezione 15.6

9.2 Variabili aleatorie Esponenziali

Una variabile aleatoria X si dice **esponenziale di parametro** $\lambda > 0$, e si scrive $X \sim EXP(\lambda)$, se e solo se ammette densità $f_X(x)$

$$f_X(x) = \begin{cases} 0 & \text{per } x < 0, \\ \\ \lambda e^{-\lambda x} & \text{per } x > 0. \end{cases}$$

Si osservi che $f_X(x) \geq 0$ per ogni $x \in \mathbb{R}$ e che

$$\int_{-\infty}^{+\infty} f_X(x)\,dx = \int_0^{\infty} \lambda e^{-\lambda x}\,dx = \int_0^{\infty} (-de^{-\lambda x}) = -e^{-\lambda x}\Big|_0^{\infty} = -0 + e^{-\lambda 0} = 1.$$

Quindi

$$F_X(x) = \begin{cases} 0 & \text{per } x < 0, \\ \\ 1 - e^{-\lambda x} & \text{per } x \geq 0. \end{cases}$$

Come anticipato nell'Esercizio proposto 14.6 in [SN], ma non dimostrato, il valore atteso e la varianza di una v.a. $EXP(\lambda)$ valgono

$$\mathbb{E}(X) = \frac{1}{\lambda}, \qquad Var(X) = \frac{1}{\lambda^2}$$

INFATTI

$$\mathbb{E}(X) = \int_0^{\infty} x\lambda e^{-\lambda x}\,dx$$

essendo $\lambda e^{-\lambda x}\,dx = d(-e^{-\lambda x})$*, ed utilizzando la formula di integrazione per parti:* $\int h(x)\,dg(x) = h(x)\,g(x) - \int g(x)\,dh(x)$*, si ottiene che*

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{\infty} x\,d(-e^{-\lambda x}) = x\,(-e^{-\lambda x})\Big|_0^{\infty} - \int_0^{\infty} (-e^{-\lambda x})\,dx \\ &= -0 + 0 + \int_0^{\infty} e^{-\lambda x}\,dx = \frac{1}{\lambda}\int_0^{\infty} \lambda e^{-\lambda x}\,dx = \frac{1}{\lambda}\cdot 1 = \frac{1}{\lambda}. \end{aligned}$$

Analogamente

$$\mathbb{E}(X^2) = \int_0^{\infty} x^2\,\lambda e^{-\lambda x}\,dx$$

essendo $\lambda e^{-\lambda x}\,dx = d(-e^{-\lambda x})$*, ed utilizzando la formula di integrazione per parti:* $\int h(x)\,dg(x) = h(x)\,g(x) - \int g(x)\,dh(x)$*, si ottiene che*

$$\begin{aligned} \mathbb{E}(X^2) &= \int_0^{\infty} x^2\,d(-e^{-\lambda x}) = x^2\,(-e^{-\lambda x})\Big|_0^{\infty} - \int_0^{\infty} (-e^{-\lambda x})\,d(x^2) \\ &= -0 + 0 + \int_0^{\infty} e^{-\lambda x}\,2x\,dx = \frac{2}{\lambda}\int_0^{\infty} x\lambda e^{-\lambda x}\,dx = \frac{2}{\lambda}\cdot \mathbb{E}(X) = \frac{2}{\lambda^2}. \end{aligned}$$

da cui

$$Var(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

9.3 Variabili aleatorie di Cauchy

Una variabile aleatoria X si dice che ha **distribuzione di Cauchy** (standard)[12] se ammette densità

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \quad x \in \mathbb{R},$$

e funzione di distribuzione

$$F_X(x) = \tfrac{1}{\pi}\arctan(x) + \tfrac{1}{2}, \quad x \in \mathbb{R}.$$

(in [SN] viene data prima la funzione di distribuzione e poi la densità, vedere l'Esempio 14.9, e ossservando poi che la derivata di $\frac{1}{\pi}\arctan(x)+\frac{1}{2}$, è $\frac{1}{\pi}\frac{1}{1+x^2}$)
Una variabili aleatoria di Cauchy, non ammette valore atteso (vedere Esercizio proposto 14.8 in [SN]) in quanto

$$\int_{-\infty}^{+\infty} |x| f_X(x)\,dx = \int_{-\infty}^{+\infty} |x| \frac{1}{\pi}\frac{1}{1+x^2}\,dx = 2\int_0^{\infty} \frac{x}{1+x^2}\,dx = \int_0^{\infty} d\log(1+x^2) = +\infty.$$

9.4 Variabili aleatorie gaussiane

Una variabile aleatoria X si dice **gaussiana standard** o **normale standard** e si scrive $X \sim N(0,1)$ se ammette densità (vedere Esempio 14.13 in [SN])

$$f_X(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}}\exp\{-\tfrac{x^2}{2}\}, x \in \mathbb{R}.$$

Come visto in Esercizio proposto 14.7 si ha

$$\mathbb{E}(X) = 0, \qquad Var(X) = 1$$

Per mostrare che $\mathbb{E}(X) = 0$, basta osservare che la funzione $x \mapsto g(x) := x f_X(x)$ è dispari, cioè $g(-x) = g(x)$ e che, essendo $d\left(\frac{1}{2}x^2\right) = x\,dx$, si ha

$$\mathbb{E}(|X|) = \int_{-\infty}^{\infty} |x|\, f_X(x)\,dx = 2\int_0^{\infty} x\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 2\frac{1}{\sqrt{2\pi}}\int_0^{\infty} e^{-\frac{1}{2}x^2}\, d\left(\frac{1}{2}x^2\right)$$

con il cambio di variabile $y = \frac{1}{2}x^2$

$$\mathbb{E}(|X|) = 2\frac{1}{\sqrt{2\pi}}\int_0^{\infty} e^{-y}\,dy = 2\frac{1}{\sqrt{2\pi}}\left(-e^{-y}\right)_0^{\infty} = 2\frac{1}{\sqrt{2\pi}}\left(-0+1\right) < \infty$$

e quindi, il valore atteso vale 0, di conseguenza la varianza coincide con $\mathbb{E}(X^2)$ e sia ha

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x e^{-\frac{1}{2}x^2}\, d\left(\frac{1}{2}x^2\right)$$

essendo $e^{-\frac{1}{2}x^2}\, d\left(\frac{1}{2}x^2\right) = -d\left(e^{-\frac{1}{2}x^2}\right)$, ed integrando per parti

$$= \frac{1}{\sqrt{2\pi}}\left\{\left[-x e^{-\frac{1}{2}x^2}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx\right\} = \frac{1}{\sqrt{2\pi}}\left\{[-0+0] + \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx\right\} = \int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

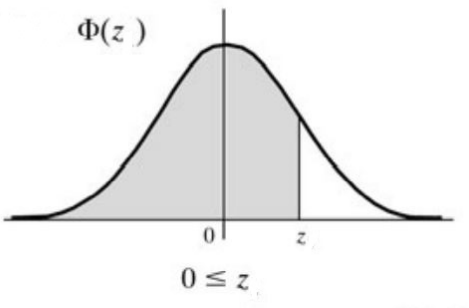
[12] Si dice che il caso standard, anche se la variabile aleatoria non è standard nel senso usuale

La funzione di distribuzione di una gaussiana standard è di solito indicata con $\Phi(x)$ ed è tabulata (vedere qui sotto o anche il paragrafo **Spiegazione dell'uso della tavola della gaussiana standard** in [SN]).

La funzione

$$\Phi(z) = \mathbb{P}(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy,$$

dove $Z \sim N(0,1)$ è una variabile aleatoria guassiana standard. In altre parole calcola l'area al di sotto del grafico della densità di probabilità gaussiana standard $y = \varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$ come descritto nella figura qui accanto.



Una variabile aleatoria Y si dice gaussiana di valore atteso μ e varianza $\sigma^2 \in (0,\infty)$ e si scrive $Y \sim N(\mu,\sigma^2)$ se ammette densità

$$f_Y(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\tfrac{(x-\mu)^2}{2\sigma^2}\}, \quad x \in \mathbb{R}.$$

e funzione di distribuzione $F_Y(y) = \Phi\big(\frac{y-\mu}{\sigma}\big)$, dove $\sigma = \sqrt{\sigma^2} > 0$ (vedere la sezione 14.6 Trasformazioni di variabili aleatorie e il caso delle trasformazioni affini in [SN]). I parametri μ e σ^2 sono rispettivamente il valore atteso e la varianza di una v.a. $Y \sim N(\mu,\sigma^2)$. Inoltre, a partire da una v.a. $X \sim N(0,1)$, si ottiene una v.a. $Y \sim N(\mu,\sigma^2)$ definendo

$$Y = \mu + \sigma X.$$

Infatti dato che $Y = \mu + \sigma Z$ dove $Z \sim N(0,1)$ e preso $\sigma > 0$ si ha

$$\{Y \le y\} = \{\mu + \sigma Z \le y\} = \{\sigma Z \le y - \mu\} = \{Z \le \tfrac{y-\mu}{\sigma}\}$$

e quindi

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}\big(Z \le \tfrac{y-\mu}{\sigma}\big) = \Phi(\tfrac{y-\mu}{\sigma})$$

da cui

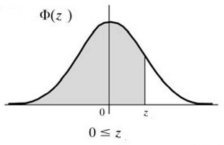
$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}\Phi(\tfrac{y-\mu}{\sigma}) = \phi(\tfrac{y-\mu}{\sigma})\, \frac{d}{dy}\frac{y-\mu}{\sigma} = \frac{1}{\sqrt{\pi}}\, e^{-\left(\frac{y-\mu}{\sigma}\right)^2} \frac{1}{\sigma} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\tfrac{(y-\mu)^2}{2\sigma^2}\}$$

Inoltre è immediato capire che $\mathbb{E}(Y) = \mu$ e $Var(Y) = \sigma^2$: infatti

$$\mathbb{E}(Y) = \mathbb{E}\big(\mu + \sigma X\big) = \mu + \sigma \overbrace{\mathbb{E}(X)}^{=0} = \mu, \qquad Var(Y) = Var(\mu + \sigma X) = \sigma^2 \overbrace{Var(X)}^{=1} = \sigma^2$$

9.5 Tavola della funzione di distribuzione gaussiana standard

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy$$



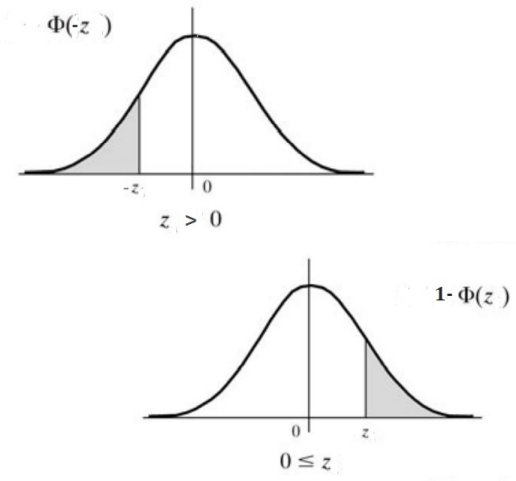
x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Spiegazione dell'uso della tavola della gaussiana standard:

NOTA BENE: *Ci sono delle tavole che calcolano invece* $\mathbb{P}(0 < Z \le z) = \int_0^z \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$. *Le due tavole si distinguono facilmente: basta osservare il valore che hanno nel punto* 0.00*, la nostra mette il valore* $\mathbb{P}(Z \le 0) = 0.5 = 1/2$*, mentre l'altro tipo mette il valore 0, in quanto* $\mathbb{P}(0 < Z \le 0) = 0$.

Per iniziare si noti che gli indici di riga sono i 35 numeri $\{0.0,\ 0.1,\ \ldots,\ 3.3,\ 3.4\}$ che vanno da 0 a 3.4 e che differiscono tra loro di un decimo, mentre gli indici di colonna sono i 10 numeri $\{0.00,\ 0.01, \ldots, 0.09\}$, che vanno da 0 a 0.09 e differiscono tra loro di un centesimo. Sommando un numero di riga, con uno di colonna si può ottenere uno tra i 350 valori di x che vanno da 0 a 3.49, e che differiscono tra loro di un centesimo. Viceversa ognuno di tali valori x, ad esempio $x = 1.43$, si può considerare come la somma della parte fino ai decimi più la parte dei centesimi, nell'esempio $x = 1.43 = 1.4 + 0.03$, individuando così un indice di riga, nell'esempio 1.4, ed uno di colonna, nell'esempio 0.03. Nella tavola, al posto di riga 1.4 e di colonna 0.03 si trova il valore di $\Phi(1.43) = 0.9236$, ovvero della funzione di distribuzione di una gaussiana standard, calcolata in $1.4 + 0.03$, e approssimato alla quarta cifra decimale.

I valori di $\Phi(x)$ per $x \ge 3.50$ si possono[13] approssimare con 1. Per quanto riguarda i valori di $\Phi(x)$ per valori negativi si usa la seguente relazione tra $\Phi(-x)$ e $1 - \Phi(x)$, dovuta alla simmetria della densità, come illuatrato qui sotto:



$$\boxed{\Phi(-x) = 1 - \Phi(x)}$$

ed in questo modo si può ottenere la funzione di distribuzione in[14] 699 valori tra $-3,49$ e $3,49$, equispaziati di un centesimo, ossia in

$$x = \frac{k}{100}, \qquad \text{per } -349 \le k \le 349.$$

Dalla relazione precedente si ottiene ad esempio che $\Phi(-1.43) = 1 - \Phi(1.43) = 1 - 0.9236 = 0.0764$

Infine la tavola ci permette di calcolare la funzione di distribuzione di una variabile aleatoria Y con distribuzione $N(\mu, \sigma^2)$, previo una trasformazione affine di Φ, ossia

$$\mathbb{P}(Y \le y) = \Phi\left(\tfrac{y-\mu}{\sigma}\right).$$

Ad esempio se $W \sim N(1,4)$ e si vuole calcolare $\mathbb{P}(W \le 3.86)$, dalla precedente espressione si ottiene che, essendo $W \sim N(\mu, \sigma^2)$, con $\mu = 1$ e $\sigma^2 = 4$,

$$\mathbb{P}(W \le 3.86) = \Phi\left(\tfrac{3.86-1}{2}\right) = \Phi(1.43) = 0.9236$$

Si noti infine che anche in questo la tavola ci permette di calcolare la funzione di distribuzione di una variabile aleatoria con distribuzione $N(\mu, \sigma^2)$ in 699 valori y, ossia i valori per i quali

$$-3.49 \le \tfrac{y-\mu}{\sigma} \le 3.49 \qquad \Leftrightarrow \qquad \mu - 3.49\,\sigma \le y \le \mu + 3.49\,\sigma$$

o più precisamente per $\frac{y-\mu}{\sigma} = \frac{k}{100}$, per $-349 \le k \le 349$, cioè

$$y = \mu + \frac{k}{100}\,\sigma, \qquad \text{per } -349 \le k \le 349.$$

[13] Ovviamente approssimare con 1 numeri maggiori o uguali a 0.9998 ha senso solo in problemi in cui la precisione non è fondamentale.

[14] Si noti che $\Phi(0) = \Phi(-0) = 1/2$ e quindi non si tratta di 700 valori, ma solo di 699

10 Somma di variabili aleatorie indipendenti con densità

Anche per le v.a. indipendenti con densità vale una formula (senza dimostrazione) analoga a quella delle v.a. discrete e che ricordiamo qui

$$p_{X+Y(z)} = \mathbb{P}(X+Y=z) = \sum_{k\geq 1} \mathbb{P}(X=x_k, Y=z-x_k) = \sum_{h\geq 1} \mathbb{P}(X=z-y_h, Y=y_h) \overset{X \perp Y}{=} \sum_{k\geq 1} p_X(x_k) p_Y(z-x_k) = \sum_{h\geq 1} p_X(z-y_h) p_Y(y_h)$$

Se X ed Y sono indipendenti e hanno densità $f_X(x)$ ed $f_Y(y)$ rispettivamente allora anche la somma $X+Y$ ha densità e vale

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy$$

ESEMPIO 1: siano X ed Y due v.a. **uniformi in** $(0,1)$ e indipendenti, allora, poiché $0 < X+Y < 2$, in quanto $0 < X < 1$ e $0 < Y < 1$, chiaramente si ha che

$$f_{X+Y}(z) = 0, \quad \text{per } z < 0 \text{ e per } z > 2$$

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx \quad \text{per } z \in (0,2)$$

l'integrtale va fatto solo per i valori di x per i quali sia $f_X(x) \neq 0$, dove vale $c = 1$, ossia per $x \in (0,1)$, sia per i quali $f_Y(z-x) \neq 0$, dove vale $c = 1$, ossia per $0 < z-x < 1$, o equivalentemente per $z-1 < x < z$. In altre parole l'integrale va calcolato per $x \in (0,1) \cap (z-1, z)$.
È facile verificare che se $z \in (0,1)$ allora $(0,1) \cap (z-1,z) = (0,z)$ e invece se $z \in (1,2)$ allora $(0,1) \cap (z-1,z) = (z-1,1)$ e quindi

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx = \int_0^z 1 \cdot 1\, dx = z, \quad \text{se } z \in (0,1),$$

e che

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx = \int_{z-1}^1 1 \cdot 1\, dx = 1-(z-1) = 2-z, \quad \text{se } z \in (1,2)$$

Disegnando il grafico di questa funzione si capisce come mai questa distribuzione di chiama **distribuzione triangolare**.
Per il calcolo della funzione di distribuzione vedere l'Esempio 14.15 in [SN].
Da notare l'analogia con la distribuzione della somma di due v.a., entrambe uniformi discrete in $\{1,2,...,n\}$ e indipendenti.

ESEMPIO 2: La somma di due v.a. $X \sim EXP(\lambda)$ e $Y \sim EXP(\lambda)$, indipendenti ha come densità e come funzione di distribuzione

$$f_{X+Y}(z) = \begin{cases} 0 & \text{per } z < 0, \\ \lambda^2 z e^{-\lambda z} & \text{per } z > 0. \end{cases} \qquad F_{X+Y}(z) = \begin{cases} 0 & \text{per } z < 0, \\ 1 - e^{-\lambda z} - \lambda z e^{-\lambda z} & \text{per } z \geq 0. \end{cases}$$

Prima di verificare che la funzione di densità di $X+Y$ ha effettivamente l'espressione vista, verifichiamo che la precedente funzione è effettivamente la funzione di distribuzione di f_{X+Y}: chiaramente $F_{X+Y}(z)$è derivabile (tranne in $z = 0$) e si ha

$$F'_{X+Y}(z) = \begin{cases} 0 & \text{per } z < 0, \\ -(-\lambda) e^{-\lambda z} - \lambda e^{-\lambda z} - \lambda z (-\lambda) e^{-\lambda z} = \lambda^2 z e^{-\lambda z} & \text{per } z > 0. \end{cases}$$

Rimane da ricavare la funzione di densità di $X+Y$.
Chiaramente $X+Y$ assume solo valori maggiori di 0 e quindi $f_{X+Y}(z) = 0$, per $z < 0$, mentre

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)\, dx, \quad \text{per } z > 0$$

dove l'integrale va calcolato dove $f_X(x) f_Y(z-x) > 0$ e quindi, poiché sia X che Y assumono solo valori positivi, va calcolato per $x > 0$ e $z-x > 0$, ossia per $0 < x < z$, e quindi

$$= \int_0^z f_X(x) f_Y(z-x)\, dx = \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)}\, dx = \lambda^2 \int_0^z e^{-\lambda x} e^{-\lambda z + \lambda x}\, dx$$

$$= \lambda^2 \int_0^z \cancel{e^{-\lambda x}} e^{-\lambda z} \cancel{e^{\lambda x}}\, dx = \lambda^2 e^{-\lambda z} z.$$

11 Funzione di distribuzione del massimo e del minimo di due v.a. indipendenti

In generale per calcolare la funzione di distribuzione del massimo e del minimo di due variabili aleatorie indipendenti con funzione di distribuzione $F_X(x)$ ed $F_Y(y)$ rispettivamente si procede nel seguente modo:
per il **massimo** $X \vee Y = \max(X,Y)$

$$F_{X\vee Y}(z) = \mathbb{P}(X \le z, Y \le z) \overset{X \perp Y}{=} \mathbb{P}(X \le z)\mathbb{P}(Y \le z) = F_X(z)\,F_Y(z)$$

Se le due variabili X ed Y hanno la stessa distribuzione, ossia se $F_X = F_Y = F$, allora

$$F_{X\vee Y}(z) = F(z)^2$$

e se INOLTRE X ed Y hanno densità $f_X = f_Y = f = F' = f$ allora

$$f_{X\vee Y}(z) = \big(F_{X\vee Y}\big)'(z) = (F(z)^2)' = 2F(z)\,F'(z) = 2f(z)F(z).$$

per il **minimo** $X \wedge Y = \min(X,Y)$

$$\begin{aligned} F_{X\wedge Y}(z) &= \mathbb{P}(X \wedge Y \le z) = 1 - \mathbb{P}(X \wedge Y > z) = 1 - \mathbb{P}(X > z, Y > z) \overset{X \perp Y}{=} 1 - \mathbb{P}(X > z)\mathbb{P}(Y > z) \\ &= 1 - \big(1 - F_X(z)\big)\big(1 - F_Y(z)\big) = F_X(z) + F_Y(z) - F_X(z)F_Y(z) \end{aligned}$$

e se INOLTRE X ed Y hanno densità $f_X = f_Y = f = F' = f$ allora

$$F_{X\wedge Y}(z) = 2F(z) - F(z)^2$$

e se, OLTRE ad essere indipendenti, X ed Y hanno densità $f_X = f_Y = f = F' = f$ allora

$$f_{X\wedge Y}(z) = \big(F_{X\wedge Y}\big)'(z) = (2F(z) - F(z)^2)' = 2f(z) - 2F(z)\,F'(z) = 2f(z)\big(1 - F(z)\big).$$

ESEMPIO Se $X \sim EXP(\lambda)$ ed $Y \sim EXP(\mu)$ sono due v.a. indipendenti allora

$$\mathbb{P}(\min(X,Y) > x) = \mathbb{P}(X > x)\mathbb{P}(Y > x) = \begin{cases} 1 & \text{per } x < 0 \\ e^{-\lambda x}e^{-\mu x} = e^{-(\lambda+\mu)x} & \text{per } x \ge 0 \end{cases}$$

$$\Updownarrow$$

$$F_{X\wedge Y}(x) = 1 - \mathbb{P}(\min(X,Y) > x) = \begin{cases} 0 & \text{per } x < 0 \\ 1 - e^{-(\lambda+\mu)x} & \text{per } x \ge 0 \end{cases}$$

$$F_{X\vee Y}(x) = \mathbb{P}(\max(X,Y) \le x) = \mathbb{P}(X \le x)\mathbb{P}(Y \le x) = \begin{cases} 0 & \text{per } x < 0 \\ \big(1 - e^{-\lambda x}\big)\big(1 - e^{-\mu x}\big) = 1 - e^{-\lambda x} - e^{-\mu x} + e^{-(\lambda+\mu)x} & \text{per } x \ge 0 \end{cases}$$

$$\Updownarrow$$

$$f_{X\vee Y}(x) = \frac{d}{dx}F_1(x) = \begin{cases} 0 & \text{per } x < 0 \\ \lambda e^{-\lambda x} + \mu e^{-\mu x} - (\lambda+\mu)\,e^{-(\lambda+\mu)x} & \text{per } x \ge 0 \end{cases}$$

Di conseguenza, ricordando che $\mathbb{E}(X) = \int_0^\infty x f_X(x)\,dx = \int_0^\infty x\lambda e^{-\lambda x}\,dx = \frac{1}{\lambda}$, si ottiene immediatamente che

$$\mathbb{E}(X \vee Y) = \int_0^\infty x\lambda e^{-\lambda x}\,dx + \int_0^\infty x\mu e^{-\mu x}\,dx - \int_0^\infty x(\lambda+\mu)\,e^{-(\lambda+\mu)x}\,dx = \frac{1}{\lambda} + \frac{1}{\mu} - \frac{1}{\lambda+\mu}.$$

Un caso particolarmente interessante riguarda invece le variabili aleatorie Geometriche:
Siano $X_A \sim Geom(p_A)$ ed $X_B \sim Geom(p_B)$ due variabili aleatorie indipendenti.

In modo del tutto analogo al caso delle variabili aleatorie esponenziali si ottiene che, posto $q_A = 1 - p_A$ e $q_B = 1 - p_B$, si ha

$$\boxed{\min(X_A, X_B) \sim Geom(p), \quad \text{con } p = p_A + p_B - p_A p_B = 1 - q_A q_B.}$$

Inoltre, si ottiene che

$$\mathbb{P}(X_A < X_B) = \frac{p_A(1-p_B)}{p_A + p_B - p_A p_B}, \quad \mathbb{P}(X_A > X_B) = \frac{p_B(1-p_A)}{p_A + p_B - p_A p_B}, \quad \mathbb{P}(X_A = X_B) = \frac{p_A p_B}{p_A + p_B - p_A p_B}, \tag{6}$$

Si osservi che, come deve essere,

$$\mathbb{P}(X_A < X_B) + \mathbb{P}(X_A > X_B) + \mathbb{P}(X_A = X_B) = \frac{p_A(1-p_B)}{p_A + p_B - p_A p_B} + \frac{p_B(1-p_A)}{p_A + p_B - p_A p_B} + \frac{p_A p_B}{p_A + p_B - p_A p_B} = 1$$

Questi risultati si possono ottenere con dei conti abbastanza semplici, e come vedremo si possono ottenere abbastanza sinteticamente, con un opportuno modello.

Iniziamo descrivendo il modello: possiamo considerare X_A e X_B siano i tempi di primo successo relativi a due successioni di eventi

$$\{A_k, k \geq 1\} \quad \text{e} \quad \{B_k, k \geq 1\}$$

tali che

- $\{A_k, k \geq 1\}$ sono eventi indipendenti e tutti con la stessa probabilità: $\mathbb{P}(A_k) = p_A$, $k \geq 1$,
- $\{B_h, h \geq 1\}$ sono eventi indipendenti e tutti con la stessa probabilità: $\mathbb{P}(B_h) = p_B$, $h \geq 1$,
- La famiglia di eventi $\{A_k, B_h, k \geq 1, h \geq 1\}$ è una famiglia di eventi indipendenti

(come al solito *indipendenti* significa **globalmente indipendenti**, ed essendo una famiglia infinita di eventi, ciò significa che, comunque presi un numero finito di eventi dalla famiglia $\{A_k, B_h, k \geq 1, h \geq 1\}$, tali eventi sono eventi indipendenti).

L'indipendenza garantisce che il tempo di primo successo X_A relativo alla successione $\{A_k, k \geq 1\}$ sia indipendente dal tempo di primo successo X_B relativo alla successione $\{B_h, h \geq 1\}$.

L'idea è che possiamo pensare di *sincronizzare* le prove in modo che le prove avvengano contemporaneamente, nel senso che, ad esempio l'evento $X_A = 3$ e $X_B = 5$ avviene quando si verifica una situazione del seguente tipo

$$\begin{pmatrix} \overline{A}_1 \\ \overline{B}_1 \end{pmatrix}, \begin{pmatrix} \overline{A}_2 \\ \overline{B}_2 \end{pmatrix}, \begin{pmatrix} A_3 \\ \overline{B}_3 \end{pmatrix}, \begin{pmatrix} \overline{A}_4 \cup A_4 \\ \overline{B}_4 \end{pmatrix}, \begin{pmatrix} \overline{A}_5 \cup A_5 \\ B_5 \end{pmatrix}, \dots$$

ovvero

$$\left(\overline{A}_1 \cap \overline{B}_1\right) \cap \left(\overline{A}_2 \cap \overline{B}_2\right) \cap \left(A_3 \cap \overline{B}_3\right) \cap \left((\overline{A}_4 \cup A_4) \cap \overline{B}_4\right) \cap \left((\overline{A}_5 \cup A_5) \cap B_5\right)$$

Si noti che in questo caso (oltre a $X_A = 3, X_B = 5$) si ha $\min(X_A, X_B) = 3$

In modo del tutto analogo, l'evento $X_A = 5$ e $X_B = 3$ avviene quando si verifica una situazione del seguente tipo

$$\begin{pmatrix} \overline{A}_1 \\ \overline{B}_1 \end{pmatrix}, \begin{pmatrix} \overline{A}_2 \\ \overline{B}_2 \end{pmatrix}, \begin{pmatrix} \overline{A}_3 \\ B_3 \end{pmatrix}, \begin{pmatrix} \overline{A}_4 \\ \overline{B}_4 \cup B_4 \end{pmatrix}, \begin{pmatrix} A_5 \\ \overline{B}_4 \cup B_4 \end{pmatrix}, \dots$$

ovvero

$$\left(\overline{A}_1 \cap \overline{B}_1\right) \cap \left(\overline{A}_2 \cap \overline{B}_2\right) \cap \left(\overline{A}_3 \cap B_3\right) \cap \left(\overline{A}_4 \cap (\overline{B}_4 \cup B_4)\right) \cap \left(A_5 \cap (\overline{B}_5 \cup B_5)\right)$$

Si noti che in questo caso (oltre a $X_A = 5$, $X_B = 3$) si ha $\min(X_A, X_B) = 3$ Infine, sempre in modo del tutto analogo, l'evento $X_A = 3$ e $X_B = 3$ avviene quando si verifica una situazione del seguente tipo

$$\begin{pmatrix} \overline{A}_1 \\ \overline{B}_1 \end{pmatrix}, \begin{pmatrix} \overline{A}_2 \\ \overline{B}_2 \end{pmatrix}, \begin{pmatrix} A_3 \\ B_3 \end{pmatrix}, \dots$$

ovvero

$$\left(\overline{A}_1 \cap \overline{B}_1\right) \cap \left(\overline{A}_2 \cap \overline{B}_2\right) \cap \left(A_3 \cap B_3\right)$$

Si noti che in questo caso (oltre a $X_A = 3, X_B = 3$) si ha $\min(X_A, X_B) = 3$

A questo punto forse è chiaro che l'evento $\min(X_A, X_B) = 3$ coincide con l'evento

$$\left(\overline{A}_1 \cap \overline{B}_1\right) \cap \left(\overline{A}_2 \cap \overline{B}_2\right) \cap \left(A_3 \cup B_3\right) = \left(A_3 \cup B_3\right)^c \cap \left(A_3 \cup B_3\right)^c \cap \left(A_3 \cup B_3\right)$$

Ovviamente questa osservazione si generalizza mettendo k al posto di 3 e quindi è chiaro che, posto

$$C_k := A_k \cup B_k, \quad k \geq 1$$

si ha che, tali eventi sono indipendenti e tutti con probabilità

$$\mathbb{P}(C_k) = \mathbb{P}(A_k \cup B_k) = \mathbb{P}(A_k) + \mathbb{P}(B_k) - \mathbb{P}(A_k \cap B_k) = \mathbb{P}(A_k) + \mathbb{P}(B_k) - \mathbb{P}(A_k)\mathbb{P}(B_k) = p_A + p_B - p_A p_B$$

e che

$$\min(X_A, X_B) = X_C = \text{tempo di primo successo, relativo alla successione } \{C_k,\ k \geq 1\},$$

e quindi

$$\min(X_A, X_B) \sim Geom(p_A + p_B - p_A p_B).$$

(Per una dimostrazione alternativa e una dimostrazione analitica vedere le note successive a pagina 57 e 58)

Inoltre è abbastanza chiaro anche che, qualunque sia $k \geq 1$

$$\begin{aligned}\mathbb{P}(X_A < X_B | \min(X_A, X_B) = k) &= \frac{\mathbb{P}\left((\cap_{h=1}^{k-1}\overline{C}_h) \cap (A_k \cap \overline{B}_k)\right)}{\mathbb{P}\left((\cap_{h=1}^{k-1}\overline{C}_h) \cap (A_k \cup B_k)\right)} = \frac{\left[\prod_{h=1}^{k-1}\mathbb{P}(\overline{C}_h)\right]\mathbb{P}(A_k)\mathbb{P}(\overline{B}_k)}{\left[\prod_{h=1}^{k-1}\mathbb{P}(\overline{C}_h)\right]\mathbb{P}(A_k \cup B_k)} \\ &= \frac{\mathbb{P}(A_k)\mathbb{P}(\overline{B}_k)}{\mathbb{P}(A_k \cup B_k)} = \frac{p_A(1-p_B)}{p_A + p_B - p_A p_B}\end{aligned}$$

e di conseguenza si ottiene la prima uguaglianza in (6):

$$\begin{aligned}\mathbb{P}(X_A < X_B) &= \sum_{k=1}^{\infty} \mathbb{P}(\min(X_A, X_B) = k)\mathbb{P}(X_A < X_B | \min(X_A, X_B) = k) \\ &= \sum_{k=1}^{\infty} \mathbb{P}(\min(X_A, X_B) = k)\frac{p_A(1-p_B)}{p_A + p_B - p_A p_B} = \frac{p_A(1-p_B)}{p_A + p_B - p_A p_B}.\end{aligned}$$

Le altre uguaglianze si ottengono in modo simile.

Esercizio proposto 11.1. *Dimostrare che* $\mathbb{P}(X_A \leq X_B) = p_A/(p_A + p_B - p_A p_B)$ *e* $\mathbb{P}(X_A \geq X_B) = p_B/(p_A + p_B - p_A p_B)$.

Dimostrazione alternativa del fatto che $\min(X_A, X_B) \sim Geom(p_A + p_B - p_A p_B)$

In realtà sarebbe stato più semplice osservare che per dimostrare che una variabile aleatoria Y a valori in $\mathbb{N}$ abbia distribuzione $Geom(p)$, basterebbe dimostrare che

$$\mathbb{P}(Y > k) = (1-p)^k, \text{ per ogni } k \geq 0,$$

in quanto ciò equivale a chiedere che, per ogni $k \geq 1$,

$$\mathbb{P}(Y = k) = \mathbb{P}(Y > k-1) - \mathbb{P}(Y > k) = (1-p)^{k-1} - (1-p)^k = (1-p)^{k-1}\left(1 - (1-p)\right) = (1-p)^{k-1}p.$$

Sarebbe stato quindi ancora più semplice osservare che

$$\{\min(X_A, X_B) > k\} = \bigcap_{h=1}^{k} (\overline{A}_h \cap \overline{B}_h)$$

e quindi

$$\begin{aligned}\mathbb{P}(\min(X_A, X_B) > k) &= \prod_{h=1}^{k} \mathbb{P}(\overline{A}_h \cap \overline{B}_h) = \prod_{h=1}^{k} \mathbb{P}(\overline{A}_h)\mathbb{P}(\overline{B}_h) \\ &= \prod_{h=1}^{k} (1-p_A)(1-p_B) = \left((1-p_A)(1-p_B)\right)^k \\ &= \left(1 - p_A - p_B + p_A p_B\right)^k = (1-p)^k, \quad \text{con } p = p_A + p_B - p_A p_B.\end{aligned}$$

Dimostrazione analitica del fatto che $\min(X_A, X_B) \sim Geom(p_A + p_B - p_A p_B)$ **e delle relazioni** (6)

Utilizzando sempre il fatto che ci basta dimostrare che

$$\mathbb{P}(\min(X_A, X_B) > k) = (1-p)^k, \quad \text{per ogni } k \geq 0,$$

con $p = p_A + p_B - p_A p_B$ osserviamo che

$$\begin{aligned}\mathbb{P}(\min(X_A, X_B) > k) &= \mathbb{P}(X_A > k, X_B > k) = \mathbb{P}(X_A > k)\mathbb{P}(X_B > k) = (1-p_A)^k (1-p_B)^k \\ &= \big((1-p_A)(1-p_B)\big)^k = \cdots = (1-p)^k\end{aligned}$$

scon $p = p_A + p_B - p_A p_B$, (si tratta di conti identici ai precedenti)

Per calcolare, ad esempio, $\mathbb{P}(X_A < X_B)$ si potrebbe procedere anche come segue

$$\begin{aligned}\mathbb{P}(X_A < X_B) &= \sum_{k=1}^{\infty} \mathbb{P}(X_A = k, X_A < X_B) = \sum_{k=1}^{\infty} \mathbb{P}(X_A = k, k < X_B) \\ &= \sum_{k=1}^{\infty} \mathbb{P}(X_A = k)\mathbb{P}(k < X_B) = \sum_{k=1}^{\infty} p_A (1-p_A)^{k-1} (1-p_B)^k \\ &= p_A (1-p_B) \sum_{k=1}^{\infty} (1-p_A)^{k-1} (1-p_B)^{k-1} = p_A (1-p_B) \sum_{h=0}^{\infty} [(1-p_A (1-p_B)]^h \\ &= \frac{p_A (1-p_B)}{1 - 1 - p_A (1-p_B)}\end{aligned}$$

12 Trasformazioni di variabili aleatorie

Sia X una variabile aleatoria, e sia $h : \mathbb{R} \to \mathbb{R}$ una funzione reale. L'applicazione

$$\omega \mapsto Y(\omega) := h\big(X(\omega)\big)$$

è una trasformazione della variabile aleatoria X.
Nel caso degli spazi finiti Y è sempre una variabile aleatoria e, supponendo che sia nota la densità discreta di X, ossia è nota la funzione $x \mapsto p_X(x) := \mathbb{P}(X = x)$, per $x \in X(\Omega) = \{x_1, x_2, ..., x_n\}$, la distribuzione di Y è individuata da $Y(\Omega) = h\big(X(\Omega)\big) = \{y_1, y_2, ..., y_m\}$ e da

$$\mathbb{P}(Y = y_j) = \sum_{k: h(x_k) = y_j} \mathbb{P}(X = x_k), \qquad j = 1, 2, ..., m$$

(si veda la ***Proposizione 12*** della Lezione 9 in [SN]).
Inoltre, come già ricordato

$$\mathbb{E}(Y) = \mathbb{E}\big[h(X)\big] = \sum_{k=1}^{n} h(x_k)\mathbb{P}(X = x_k)$$

Nel caso degli spazi generali ci sono due differenze:

(i) non è sempre vero che Y sia una variabile aleatoria;

(ii) può accadere che la distribuzione di Y non si possa calcolare nel modo precedente.

Il problema si pone in particolare se la variabile aleatoria X non è discreta, mentre se la variabile aleatoria X è discreta numerabile, allora si generalizza immediatamente la precedente relazione, in quanto necessariamente Y è discreta (finita o numerabile), infatti la formula è la stessa, ma bisogna solo precisare che in

$$\mathbb{P}(Y = y_k) = \mathbb{P}(h(X) = y_k) = \sum_{i: h(x_i) = y_k} \mathbb{P}(X = x_i),$$

invece di una somma finita, potrebbe trattarsi della somma di una serie, e che l'insieme $Y(\Omega)$ dei i valori y_k che Y assume potrebbe essere infinito numerabile.

Per variabili aleatorie generali, va detto (senza dimostrazione) che

se h è continua, allora $Y = h(X)$ è sicuramente una variabile aleatoria.

Tuttavia la condizione che h sia una funzione continua, è solo una condizione sufficiente e non è necessaria.
Un'altra **condizione sufficiente** è che **la funzione h sia continua a tratti**, ossia esistano un numero finito o numerabile di intervalli J_k, $k = 1, 2, ...$, che formano una partizione di $\mathbb{R}$ e tali che h è continua su ciascun intervallo J_k. In particolare se h è costante a tratti.

Nel caso particolare in cui h sia continua e strettamente monotona (ossia strettamente crescente o strettamente decrescente) e quindi invertibile si può sempre ricavare la funzione di distribuzione di Y.
In particolare, se è strettamente crescente, si ha che

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(h(X) \le y) = \mathbb{P}\big(X \le h^{-1}(y)\big) = F_X\big(h^{-1}(y)\big)$$

Se invece h è continua e strettamente decrescente (e quindi invertibile) si ha che

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(h(X) \le y) = \mathbb{P}\big(X \ge h^{-1}(y)\big) = \mathbb{P}\big(X > h^{-1}(y)\big) + \mathbb{P}\big(X = h^{-1}(y)\big) \\ &= 1 - F_X\big(h^{-1}(y)\big) + \mathbb{P}\big(X = h^{-1}(y)\big) \end{aligned}$$

SI NOTI che basta che h sia invertibile su un intervallo I tale che $\mathbb{P}(X \in I) = 1$

ESEMPIO 1 Sia X uniforme in $(0,1)$, sia $h(x) = -\dfrac{\log(x)}{\lambda}$, con $\lambda > 0$, e sia $\boxed{Y = h(X) = \dfrac{-\log(X)}{\lambda}.}$

Si osservi che $h(x)$ è definita solo per $x > 0$, dove è continua e decrecente. Ciò non comporta però nessun problema, in

quanto la variabile aleatoria X assume valori sono in $(0,1)$ e quindi, ricordando che, per $x \in (0,1)$, la funzione $\log(x)$ assume solo valori in $(-\infty,0)$ e $\lambda > 0$, chiaramente Y assume solo valori strettamente positivi: quindi

$$F_Y(y) = \mathbb{P}(Y \le y) = \begin{cases} 0 & \text{per } y \le 0 \\ = \mathbb{P}\big(-\frac{\log(X)}{\lambda} \le y\big) = \mathbb{P}\big(-\log(X) \le \lambda y\big) = \mathbb{P}\big(\log(X) \ge -\lambda y\big) & \\ = \mathbb{P}\big(\exp\{\log(X)\} \ge \exp\{-\lambda y\}\big) = \mathbb{P}(X \ge e^{-\lambda y}) = 1 - e^{-\lambda y} & \text{per } y > 0 \end{cases}$$

ovvero $Y \sim EXP(\lambda)$.

Consideriamo ora il caso in cui h è una trasformazione affine (spesso indicata anche come trasformazione lineare, però): dati due numeri reali $\alpha \neq 0$ e β

$$\boxed{h(x) = \alpha x + \beta} \quad \text{e quindi} \quad \boxed{h^{-1}(y) = \frac{y-\beta}{\alpha}, \qquad (\alpha \neq 0)}$$

e in cui X ammetta densità di probabilità $f_X(x) = F_X'(x)$.

Allora anche Y ammette densità di probabilità e si ha $\boxed{f_Y(y) = f_X\big(\frac{y-\beta}{\alpha}\big)\,\frac{1}{|\alpha|}}$

Infatti, nel caso h crescente, ossia $\alpha > 0$,

$$f_Y(y) = F_Y'(y) = \frac{d}{dy} F_X\big(\tfrac{y-\beta}{\alpha}\big) = F_X'\big(\tfrac{y-\beta}{\alpha}\big)\,\frac{d}{dy}\frac{y-\beta}{\alpha} = f_X\big(\tfrac{y-\beta}{\alpha}\big)\,\frac{1}{\alpha} = f_X\big(\tfrac{y-\beta}{\alpha}\big)\,\frac{1}{|\alpha|}$$

mentre nel caso h decrescente, ossia $\alpha < 0$, ricordando che, essendo X assolutamente continua, $\mathbb{P}\big(X = h^{-1}(y)\big) = 0$,

$$\begin{aligned} f_Y(y) = F_Y'(y) &= \frac{d}{dy}\Big[1 - F_X\big(\tfrac{y-\beta}{\alpha}\big)\Big] = -F_X'\big(\tfrac{y-\beta}{\alpha}\big)\,\frac{d}{dy}\frac{y-\beta}{\alpha} \\ &= -f_X\big(\tfrac{y-\beta}{\alpha}\big)\,\frac{1}{\alpha} = f_X\big(\tfrac{y-\beta}{\alpha}\big)\,\frac{1}{|\alpha|} \end{aligned}$$

ESEMPIO 2 In particolare se X è $Unif(0,1)$, e $a < b$, allora la v.a. $Y := a + (b-a)X$ è una variabile aleatoria con distribuzione $Unif(a,b)$, infatti

$$F_X(x) = \begin{cases} 0 & \text{per } x < 0, \\ x & \text{per } 0 \le x < 1, \\ 1 & \text{per } x \ge 1, \end{cases} \text{ e quindi } F_Y(y) = F_X\big(\tfrac{y-a}{b-a}\big) = \begin{cases} 0 & \text{per } \frac{y-a}{b-a} < 0, \Leftrightarrow\ y < a, \\ \frac{y-a}{b-a} & \text{per } 0 \le \frac{y-a}{b-a} < 1, \Leftrightarrow\ 0 \le y-a < b-a \Leftrightarrow\ a \le y < b, \\ 1 & \text{per } \frac{y-a}{b-a} \ge 1, \Leftrightarrow\ y-a \ge b-a \Leftrightarrow\ y \ge b. \end{cases}$$

ESEMPIO 3 Se invece $X \sim N(0,1)$ e $Y = \mu + \sigma X$, con $\sigma \neq 0$, allora $Y \sim N(\mu, \sigma^2)$, infatti

$$f_Y(y) = \frac{1}{|\sigma|}\, f_X\big(\tfrac{y-\mu}{\sigma}\big) = \frac{1}{|\sigma|}\,\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}}.$$

Va anche notato che viceversa, se $Y \sim N(\mu, \sigma^2)$ allora la sua standardizzata $Y^* = \frac{Y-\mu}{\sigma}$ è una variabile aleatoria gaussiana standard.
Inoltre, nel caso in cui $\sigma > 0$

$$F_Y(y) = \mathbb{P}(\mu + \sigma X \le y) = \mathbb{P}\big(X \le \tfrac{y-\mu}{\sigma}\big) = \Phi\big(\tfrac{y-\mu}{\sigma}\big).$$

Una formula analoga vale quando $\sigma < 0$, infatti, in tal caso si ha

$$F_Y(y) = \mathbb{P}(\mu + \sigma X \le y) = \mathbb{P}\big(X \ge \tfrac{y-\mu}{\sigma}\big) = \mathbb{P}\big(X > \tfrac{y-\mu}{\sigma}\big) = 1 - \Phi\big(\tfrac{y-\mu}{\sigma}\big) = \Phi\big(-\tfrac{y-\mu}{\sigma}\big) = \Phi\big(\tfrac{y-\mu}{|\sigma|}\big).$$

Un caso particolarmente interessante è anche il caso in cui X è uniforme in $(0,1)$: se $h(x)$ è una funzione strettamente crescente e continua e $Y = h(X)$, allora, ricordando che $F_X(x) = 0$, per $x < 0$, $F_X(x) = x$, per $0 \le x \le 1$, $F_X(x) = 1$, per $x > 1$, si ottiene

$$F_Y(y) = F_X(h^{-1}(y)) = \begin{cases} 0 & \text{se } h^{-1}(y) < 0 \\ h^{-1}(y) & \text{se } 0 \le h^{-1}(y) \le 1 \\ 1 & \text{se } h^{-1}(y) > 1 \end{cases}$$

Sappiamo che è in molti linguaggi di programmazione esiste un'istruzione denotata come *random* che come output fornisce un valore tra $(0,1)$ e che possiamo pensare come la realizzazione/simulazione di una variabile aleatoria X uniforme in $(0,1)$.

(anche se, per vari motivi, ciò non è del tutto vero, ad esempio l'output può assumere solo un numero finito, per quanto grande, di valori in $(0,1)$, ma c'è anche il problema di poter considerare effettivamente l'output come un valore casuale)

Nasce quindi il problema: ***data una funzione di distribuzione $G(x)$, come possiamo simulare una variabile aleatoria con tale funzione di distribuzione?***

Se $G(x)$ è continua e strettamente crescente, e a valori in $[0,1]$ e quindi anche la sua inversa G^{-1} è continua e strettamente crescente, possiamo prendere $h(x) = G^{-1}(x)$ in modo che $h^{-1}(y) = G(y)$. Allora, osservando che $h^{-1}(y) = G(y)$ assume solo valori in $[0,1]$, si ottiene che, per $Y = G^{-1}(X)$

$$F_Y(y) = h^{-1}(y) = G(y).$$

ESEMPIO 4 Vogliamo simulare una variabile aleatoria di Cauchy, che ha come funzione di distribuzione

$$G(x) = \frac{1}{\pi}\arctan(x) + \frac{1}{2}$$

Osserviamo che $G(x)$ assume solo i valori tra $(0,1)$. Per quanto visto prima dobbiamo solo trovare la sua funzione inversa $G^{-1}(x)$ e definire $\boxed{Y = G^{-1}(X) = \tan\left(\pi(X - \frac{1}{2})\right),}$ dove X è uniforme in $(0,1)$. In questo caso si ha

$$G^{-1}(x) = \tan\left(\pi(x - \tfrac{1}{2})\right), \qquad x \in (0,1).$$

Infatti dato $x \in (0,1)$ dobbiamo trovare $y \in \mathbb{R}$ tale che

$$G(y) = x \quad \Leftrightarrow \quad \frac{1}{\pi}\arctan(y) + \tfrac{1}{2} = x \quad \Leftrightarrow \quad \frac{1}{\pi}\arctan(y) = x - \tfrac{1}{2} \quad \Leftrightarrow \quad \arctan(y) = \pi\,(x - \tfrac{1}{2}) \quad \Leftrightarrow \quad y = \tan\left(\pi(x - \tfrac{1}{2})\right).$$

ESEMPIO 5 Nel caso di una variabile $EXP(\lambda)$ possiamo anche usare il risultato dell'ESEMPIO 1, e otterremo che, data una variabile aleatoria X è uniforme in $(0,1)$ la variabile aleatoria $\boxed{Y' = -\dfrac{\log(X)}{\lambda},}$ è una v.a. $EXP(\lambda)$, ossia per simulare una v.a. $EXP(\lambda)$ basta un'istruzione del tipo $-\log(Random)/\lambda$.
Se usassimo invece la funzione inversa della funzione di distribuzione otterremmo che $G^{-1}(x) = -\frac{\log(1-x)}{\lambda}$, e quindi $\boxed{Y = G^{-1}(X) = -\dfrac{\log(1-X)}{\lambda},}$ è una v.a. $EXP(\lambda)$.
Osservazione: i due modi di ottenere una v.a. $EXP(\lambda)$ non sono in contraddizione, in quanto se X è uniforme in $(0,1)$ allora anche $1 - X$ è uniforme in $(0,1)$.

SUGGERIMENTO: Simulare una sequenza finita di variabili aleatorie X_i di Cauchy, per $i = 1, 2, \ldots, n$, e, al variare di n, calcolare la media aritmetica $\frac{1}{n}\left(X_1 + \cdots X_n\right)$.
e fare lo stesso per una sequenza finita di v.a. $EXP(\lambda)$, per qualche valore fissato di λ.

Ricordando che la legge dei grandi numeri NON si applica a tale tipo di variabili aleatorie, che non hanno valore atteso, verificare che tale media aritmetica varia in modo incontrollato.

Invece se si simulano le variabili aleatorie X_i di tipo $EXP(\lambda)$ (per un valore fissato di $\lambda > 0$), verificare che la media aritmetica si avvicina sempre più al valore atteso delle X_i, ossia a $\frac{1}{\lambda}$.

PER LE SIMULAZIONI: può essere utile osservare che, posto come al solito

$$S_n = \sum_{k=1}^{n} X_k, \quad \text{per la somma e } Y_n = \frac{S_n}{n} \text{ per la media aritmetica,}$$

vale la seguente formula ricorsiva

$$Y_{n+1} = \frac{n}{n+1} Y_n + \frac{X_{n+1}}{n+1},$$

in quanto

$$Y_{n+1} = \frac{S_{n+1}}{n+1} = \frac{n}{n+1} \frac{S_n + X_{n+1}}{n} = \frac{n}{n+1} \frac{S_n}{n} + \frac{X_{n+1}}{n+1} = \frac{n}{n+1} Y_n + \frac{X_{n+1}}{n+1}.$$